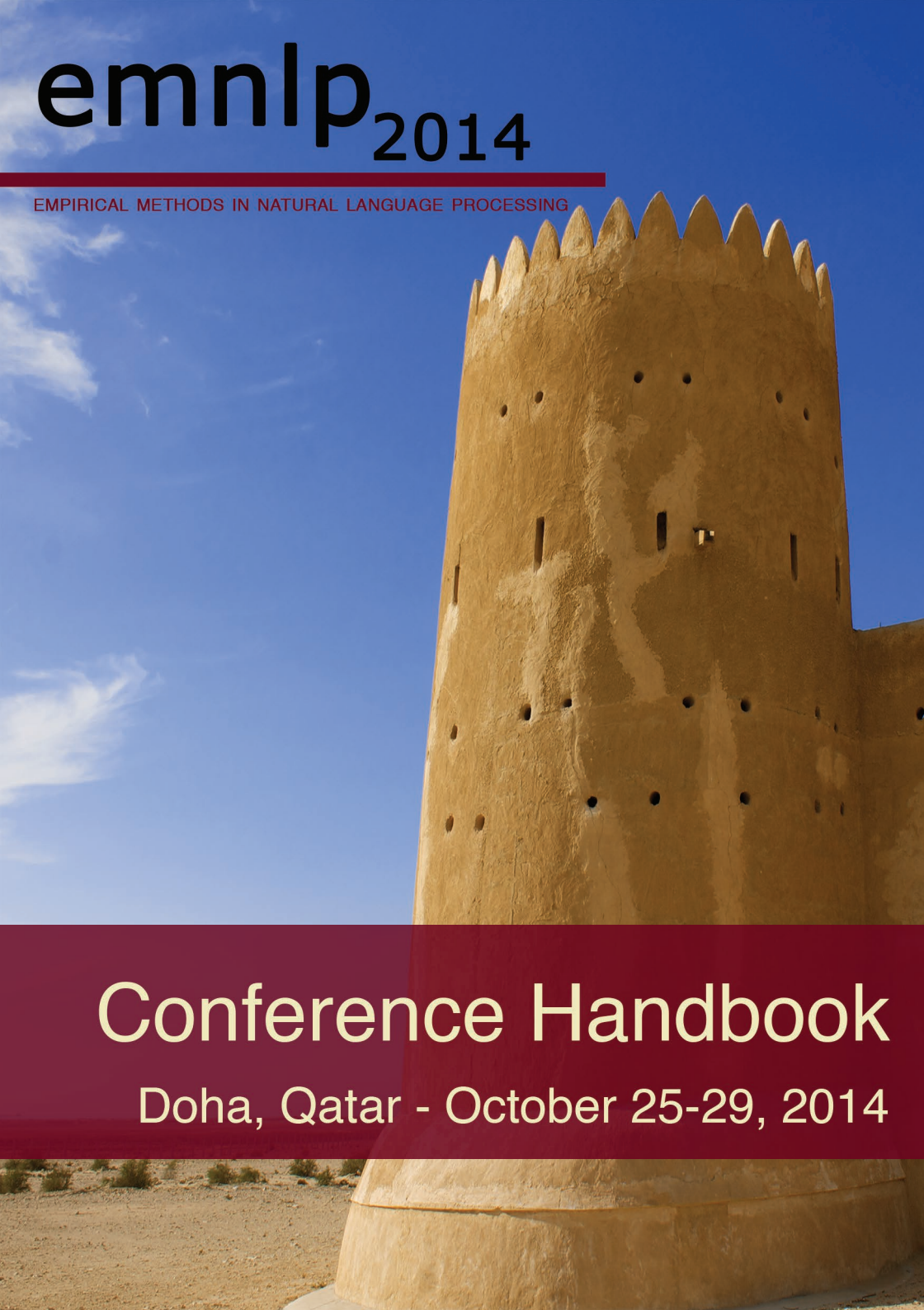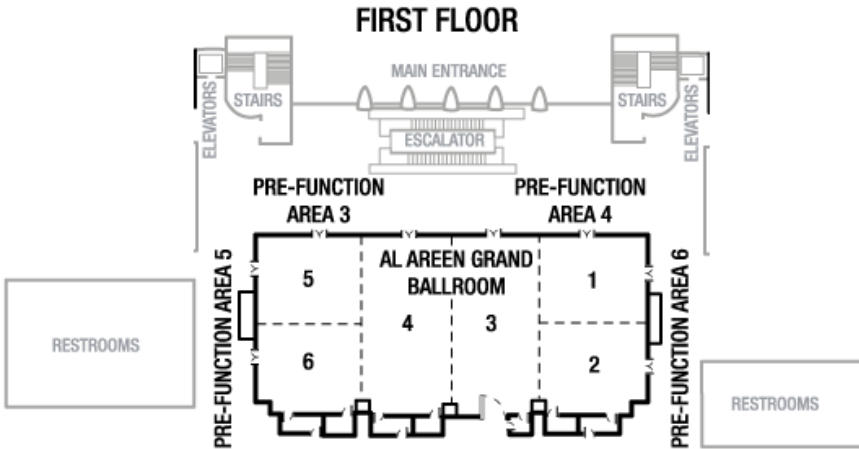# emnlp 2014

EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING

Conference Handbook

Doha, Qatar - October 25-29, 2014

FIRST FLOOR

The EMNLP'2014 main conference and all the associated tutorials and workshops will take place in the *Al Areen Grand Ballroom* located in the first floor of the Marriot hotel. See the picture above. The ballroom can be split in any combination of the six sub-rooms. In the schedule, we will list the location of every event as *Al Areen* followed by the room number(s), e.g., "Al Areen 3" or "Al Areen 1-2". The plenary sessions will occupy rooms 1-4. Lunch, coffee breaks and registration will be in the numbered "Pre-Function Areas" around the Grand Ballroom.

Cover design and Photograph by Francisco Guzmán

Contents

1

Conference Information

Preface by the General Chair

Welcome to the 2014 Conference on Empirical Methods in Natural Language Processing.

The EMNLP conference series is annually organized by SIGDAT, the Association for Computational Linguistics' special interest group on linguistic data and corpus-based approaches to NLP. This year the conference is being held from October 25, 2014 (Sat.) to October 29, 2014 (Wed.) in Doha, Qatar.

In the past five years, the EMNLP conference attendance has been continuously growing, reaching just over 500 paying attendees in 2013, and it is nowadays considered as one of the leading conferences in Computational Linguistics and Natural Language Processing.

Given the growing trend, we believed it was the right time to lead EMNLP into an organization structure typical of large and important conferences. Therefore, we proposed several novelties: first of all, a large organization committee consisting of twenty (plus twenty-six area chairs) well-known members of the ACL community, who carried out several tasks required by the new achieved scale.

Secondly, as this is the first conference edition spanning five days, in addition to six workshops, we also selected and included for the first time an excellent selection of eight tutorials. We defined a registration policy that allows the participants to attend any of the tutorials and workshops (held on October 25th and 29th) by just paying a low flat rate on top of the registration fee for the main conference. We believe this can greatly increase the spread of advanced technology and promote a unified view of the techniques and foundations of our research field.

Thirdly, as a standalone conference, EMNLP required the definition of new administrative procedures and policies, regarding sponsorship booklets, double submission, scholarship assignment, and the joint EACL-ACL-EMNLP call for workshop proposals.

Next, EMNLP is finding new ways to foster the dissemination of research work by facing the increasing number of papers to be presented at the conference. Our new approach consisted in presenting posters in nine sessions each proposing a small numbers of papers: this way poster presentations can receive the space and consideration that they deserve. Then, we are adding a surprise in terms of paper presentation and dissemination, which will be unveiled only few days before the start of the conference.

Finally, this is the first time that an ACL conference is largely supported by a government research foundation. The Qatar National Research Fund (QNRF) has included EMNLP 2014 as one of its local funding events. This enabled EMNLP and SIGDAT to perform unprecedented student scholarship support: more than 30 students were sponsored (partially or entirely) for participating in the confer-

ence. The obtained funds also allowed for offering a social dinner free of charge to all the attendees and still closing the conference budget in active, thus creating additional resources that SIGDAT can use to support the upcoming conferences.

The novelties above as well as the traditional activities that the EMNLP conference series proposes to its members could not have been organized without the work of our large committee. In this respect, I would like to thank our PC co-chairs Walter Daelemans and Bo Pang, who greatly used their large experience with program committees of our community for selecting an excellent program.

Special thanks go to our publication chair Yuval Marton, who did a terrific job in organizing and preparing the proceedings. As a side effect of his proactive action, workshop organizers and future publication chairs using the SoftConf START/ACLPUB systems can now streamline the inclusion of workshops and conference schedules in the proceedings, without heavy manual customization.

We are very grateful to Enrique Alfonseca and Eric Gaussier for selecting interesting and successful workshops and to Lucia Specia and Xavier Carreras, who, for the first time, carried out the new task of selecting tutorials for an EMNLP conference. The workshops and tutorials nicely filled the additional two days of EMNLP, making our conference even more valuable.

Many thanks are due to Katrin Erk and Sebastian Padó, who were challenged by the new activity (for EMNLP) of defining policy for the selection and assignment of participation scholarships to the most deserving students. The uncertainty over the final amount of funds and their diverse nature made this task particularly difficult. Nevertheless, they were able to find appropriate and successful solutions.

As any large conference, we could count on the help of publicity co-chairs to advertise the old and new EMNLP features. We give our gratitude to Mona Diab and Irina Matveeva for their professional work.

Fund hunting is a very important activity for conferences, in this respect, I would like to thank our sponsorship co-chairs, Jochen Leidner, Veselin Stoyanov and Min Zhang, for helping us to look for sponsors in three different continents.

Regarding the SIGDAT side, a special thank is devoted to Noah Smith, who promptly answered any question I came out with. I am also grateful to the other SIGDAT officers (past and new): Eugene Charniak, Mark Johnson, Philipp Koehn, Mark Steedman, who were always there to give suggestions and solutions to critical issues that inevitably arise in any large event.

Many thanks also to Tim Baldwin, Anna Korhonen, Graeme Hirst and David Yarowsky who provided much useful information from past conferences. Last but not least, I would like to thank Priscilla Rasmussen for her help and advice, and her undoubtful qualities of soothsayer regarding the estimation of conference numbers.

Coming back to the sponsor topic, we are enormously thankful to QNRF, for accepting our proposal to fund EMNLP: this has made it possible to sponsor an unprecedented number of students and offer a banquet free of charge to all participants (we needed to create a new level of sponsorship for them, namely, Diamond). We are very grateful to The Qatar Computing Research Institute, which in addition to providing the very valuable Platinum sponsorship, also provided the required man power for organizing the event.

In particular, EMNLP could not be organized in Qatar without the work of Kareem Darwish, the local organization chair. We are also very grateful to Kemal Oflazer, local co-chair and Francisco Guzman Herrera, local sponsorship chair, whose work was determinant to obtain the QNRF sponsorship. We are deeply in debt with the other local organizers, Lluís Màrquez, who also edited the conference booklet, Preslav Nakov, Fabrizio Sebastiani and Stephan Vogel for their help with the daily big and little issues.

Special thanks go to The Carnegie Mellon University in Qatar for helping us with the proposal preparation and management of the QNRF funds and also for supporting us with a Gold sponsorship. Additionally, many thanks go to our silver sponsors, Facebook and Yandex and our bronze sponsor iHorizons, who show the increasing interest of industry in the technology of our community for the design of real-world and high-societal impact applications. In this respect, we sincerely thank Google Inc. and IBM

Watson, New York, for supporting the student participation with their scholarships.

Finally, and foremost, thanks to all the authors and conference attendees who are the main actors of this event, bringing the real value to it and determining its success. My personal thanks also go to the entire SIGDAT committee, for choosing me as the chair of this fantastic conference, held in a fascinating venue.

Alessandro Moschitti
General Chair of EMNLP 2014

Preface by the Program Committee Co-Chairs

We welcome you to the 2014 Conference on Empirical Methods in Natural Language Processing.

As in the previous EMNLP, we invited both long and short papers with a single submission deadline. Short papers encourage the submission of smaller and more preliminary contributions.

We received 790 submissions (after initial withdrawals of unfinished submissions and removal of duplicates), of which 28 were rejected before review for not adhering to the instructions in the call for papers regarding paper length or anonymity. The remaining 510 long and 252 short papers were allocated to one of the fourteen areas. The most popular areas this year were Machine Translation, Semantics, and Syntax (Tagging, Chunking, and Parsing).

Reviewing for a conference of this size involves an army of dedicated professionals volunteering to donate their valuable and scarce time to make sure that the highest possible reviewing standards are reached. We are very grateful to our 26 area chairs and a program committee of more than 500 for their efforts. We accepted 155 long and 70 short papers, representing a global acceptance rate of just under 30%. Nine papers accepted by the ACL journal TACL were added to the program.

Based on the reviews and on nominations by the area chairs, 5 long papers were shortlisted for the best paper award. The best paper will be presented in a plenary best paper award ceremony. We would like to thank Mark Johnson and Claire Cardie for their willingness to serve in the best paper award committee that was set up and for providing excellent advice and motivation for their choice.

We are grateful to the authors for selecting EMNLP as the venue for their work. Congratulations to the authors of accepted submissions. To the authors of rejected submissions, we would like to offer as consolation the fact that because of the competitive nature of the conference and the inevitable time and space limitations, many worthwhile papers could not be included in the program. We hope the feedback of the reviewers will be considered worthwhile by them and lead to successful future submissions.

We are very grateful to our invited speakers Thorsten Joachims and Salim Roukos. Thorsten Joachims is professor at the Computer Science and Information Science departments at Cornell University and shows how integrating microeconomic models of human behavior into the learning process leads to new interaction models and learning algorithms, in turn leading to better-performing systems. Salim Roukos is senior manager of multilingual NLP and CTO of Translation Technologies at the IBM T.J. Watson Research Center and addresses IBM's approach to cognitive computing for building systems and solutions that enable and support richer human-machine interactions, and remaining opportunities in this area for novel statistical models for natural language processing. We thank them for their inspiring talks and presence at the conference.

We would also like to thank our general chair Alessandro Moschitti for his leadership, advice, encouragement, and support, Kareem Darwish and his colleagues for impeccable cooperation from local organization, and Yuval Marton for doing an excellent job assembling these proceedings.

It was an honour to serve as Program Chairs of EMNLP 2014, and we hope that you will enjoy the conference and you will be able to think back later and remember a scientifically stimulating conference and a pleasant time in Doha, Qatar.

Walter Daelemans and Bo Pang

EMNLP 2014 Program Committee Co-Chairs

Conference Committee

General Conference Chair

    Alessandro Moschitti, Qatar Computing Research Institute

Program Committee Co-Chairs

    Walter Daelemans, University of Antwerp
    Bo Pang, Google

Workshops Co-Chairs

    Enrique Alfonseca, Google Research at Zurich
    Eric Gaussier, Université Joseph Fourier (Grenoble I)

Tutorial Co-Chairs

    Lucia Specia, University of Sheffield
    Xavier Carreras, Universitat Politècnica de Catalunya

Publication Chair

    Yuval Marton, Microsoft Corporation

Publicity Co-Chairs

    Mona Diab, George Washington University
    Irina Matveeva, NexLP

Sponsorship Co-Chairs

    Jochen Leidner, Thomson Reuters
    Veselin Stoyanov, Facebook
    Min Zhang, Soochow University

Student Scholarship Co-Chairs

    Katrin Erk, University of Texas at Austin
    Sebastian Padó, University of Stuttgart

Reviewing Coordinators

    Mark Dredze, Johns Hopkins University
    Jiang Guo (Student Volunteer), Harbin Institute of Technology

Area Chairs

    Phonology, Morphology, and Segmentation
        Tomaž Erjavec, Jožef Stefan Institute

    Tagging, Chunking, Syntax and Parsing
        Gosse Bouma, University of Groningen
        Yuji Matsumoto, Nara Institute of Science and Technology

    Discourse, Dialogue, and Pragmatics
        Jennifer Chu-Carroll, IBM Watson Research Center
        Olga Uryupina, University of Trento

Semantics

> Rada Mihalcea, University of Michigan
> Sameer Pradhan, Harvard Medical School

Summarization and Generation

> Anja Belz, University of Brighton
> Dilek Hakkani-Tür, Microsoft Research

NLP-related Machine Learning: theory, methods and algorithms

> Ivan Titov, University of Amsterdam
> Jerry Zhu, University of Wisconsin-Madison

Machine Translation

> Chris Callison-Burch, University of Pennsylvania
> Daniel Gildea, University of Rochester

Information Retrieval, Text Categorization, and Question Answering

> Marie-Francine Moens, Katholieke Universiteit Leuven
> Hinrich Schütze, Ludwig Maximilian University of Munich

Information Extraction

> Doug Downey, Northwestern University
> Marius Pasca, Google

Text Mining and Natural Language Processing Applications

> Massimiliano Ciaramita, Google
> Hwee Tou Ng, National University of Singapore

Sentiment Analysis and Opinion Mining

> Yejin Choi, Stony Brook University
> Minlie Huang, Tsinghua University

NLP for the Web and Social Media

> Irwin King, The Chinese University of Hong Kong
> Qiaozhu Mei, University of Michigan

Spoken Language Processing

> Pascale Fung, Hong Kong University of Science and Technology
> Hugo Van hamme, Katholieke Universiteit Leuven

Computational Psycholinguistics

> Sharon Goldwater, University of Edinburgh

Local Organization

Local Arrangements Co-Chairs

> Kareem Darwish, Qatar Computing Research Institute
> Kemal Oflazer, Carnegie Mellon University – Qatar

Local Sponsorship Chair

> Francisco Guzmán, Qatar Computing Research Institute

Conference Handbook Editor

> Lluís Màrquez, Qatar Computing Research Institute

Local Organizing Committee

Preslav Nakov, Qatar Computing Research Institute
Fabrizio Sebastiani, Qatar Computing Research Institute

Local QCRI Administration

Kimberly Mathern, Qatar Computing Research Institute
Lawrence Tingson, Qatar Computing Research Institute
Jacqueline Caparas, Qatar Computing Research Institute

2

Overview

8:00 – 17:00	**Registration**	(Area 3)
8:00 – 9:00	**Refreshments**	(Area 3-4)
9:00 – 12:30	**Morning Tutorials**	

Sentiment Analysis of Social Media Texts   (Al Areen 4)
Saif Mohammad and Xiaodan Zhu

Spectral Learning Techniques for Weighted Automata, Transducers, and Grammars
(Al Areen 5)
Borja Balle, Ariadna Quattoni, and Xavier Carreras

12:30 – 14:00	**Lunch break**	
14:00 – 17:30	**Afternoon Tutorials**	

Semantic Parsing with Combinatory Categorial Grammars   (Al Areen 4)
Yoav Artzi, Luke Zettlemoyer, and Nicholas FitzGerald

Linear Programming Decoders in Natural Language Processing: From Integer Programming to Message Passing and Dual Decomposition.   (Al Areen 5)
André Martins

Tutorial 1

Sentiment Analysis of Social Media Texts

Saif M. Mohammad and Xiaodan Zhu (National Research Council Canada)

Saturday, October 25, 2014, 9:00 – 12:30pm

Al Areen 4

Automatically detecting sentiment of product reviews, blogs, tweets, and SMS messages has attracted extensive interest from both the academia and industry. It has a number of applications, including: tracking sentiment towards products, movies, politicians, etc.; improving customer relation models; detecting happiness and well-being; and improving automatic dialogue systems. In this tutorial, we will describe how you can create a state-of-the-art sentiment analysis system, with a focus on social media posts.

We begin with an introduction to sentiment analysis and its various forms: term level, message level, document level, and aspect level. We will describe how sentiment analysis systems are evaluated, especially through recent SemEval shared tasks: Sentiment Analysis of Twitter (SemEval-2013 Task 2, SemEval 2014-Task 9) and Aspect Based Sentiment Analysis (SemEval-2014 Task 4).

We will give an overview of the best sentiment analysis systems at this point of time, including those that are conventional statistical systems as well as those using deep learning approaches. We will describe in detail the NRC-Canada systems, which were the overall best performing systems in all three SemEval competitions listed above. These are simple lexical- and sentiment-lexicon features based systems, which are relatively easy to re-implement.

We will discuss features that had the most impact (those derived from sentiment lexicons and negation handling). We will present how large tweet-specific sentiment lexicons can be automatically generated and evaluated. We will also show how negation impacts sentiment differently depending on whether the scope of the negation is positive or negative. Finally, we will flesh out limitations of current approaches and promising future directions.

Saif M. Mohammad is a Research Officer at the National Research Council Canada. His research interests are in Computational Linguistics, especially Lexical Semantics. He develops computational models for sentiment analysis, emotion detection, semantic distance, and lexical-semantic relations such as word-pair antonymy.

Xiaodan Zhu is a Research Officer at the National Research Council Canada. His research interests are in Natural Language Processing, Spoken Language Understanding, and Machine Learning. His recent work focuses on sentiment analysis, emotion detection, speech summarization, and deep learning.

Tutorial 2

Spectral Learning Techniques for Weighted Automata, Transducers, and Grammars

Borja Balle (McGill University), Ariadna Quatoni (Xerox Research Centre Europe), and Xavier Carreras (Xerox Research Centre Europe)

Saturday, October 25, 2014, 9:00 – 12:30pm

Al Areen 5

In recent years we have seen the development of efficient and provably correct algorithms for learning weighted automata and closely related function classes such as weighted transducers and weighted context-free grammars. The common denominator of all these algorithms is the so-called spectral method, which gives an efficient and robust way to estimate recursively defined functions from empirical estimations of observable statistics. These algorithms are appealing because of the of existence of theoretical guarantees (e.g. they are not susceptible to local minima) and because of their efficiency. However, despite their simplicity and wide applicability to real problems, their impact in NLP applications is still moderate. One of the goals of this tutorial is to remedy this situation.

The contents that will be presented in this tutorial will offer a complementary perspective with respect to previous tutorials on spectral methods presented at ICML-2012, ICML-2013 and NAACL-2013. Rather than using the language of graphical models and signal processing, we tell the story from the perspective of formal languages and automata theory (without assuming a background in formal algebraic methods). Our presentation highlights the common intuitions lying behind different spectral algorithms by presenting them in a unified framework based on the concepts of low-rank factorizations and completions of

---

Borja Balle is currently a postdoctoral fellow at McGill University, and prior to that he obtained his PhD from Universitat Politecnica de Catalunya (UPC) in July 2013. His research interests lie on the intersection between automata theory and machine learning, in particular on applications of spectral learning techniques to natural language processing, grammatical inference, and reinforcement learning. He is area chair for NIPS 2014, program committee member for ICGI 2014, and has recently organized three workshops (at ICML 2013, NIPS 2013 and ICML 2014) on methods of moments and spectral learning.

Ariadna Quattoni is currently a researcher at Xerox Research Centre Europe (XRCE), prior to that she was a researcher at the Universitat Politecnica de Catalunya (UPC). She obtained her PhD from MIT in 2009. Her main research focuses on latent variable models for structured prediction with applications to natural language processing and computer vision. On the last years her work has centered on spectral learning techninques for structured prediction problems with applications to sequence tagging, learning general transductions, and parsing.

Xavier Carreras research is in natural language processing and machine learning. He is interested in grammatical induction and parsing methods for syntactic-semantic analysis and translation of natural languages. In 2005 he completed his PhD at the Universitat Politecnica de Catalunya (UPC). From 2006 to 2009 he was a postdoctoral researcher at MIT/CSAIL. From 2009 to 2014 he was a researcher at UPC and since June 2014 he is senior researcher at Xerox Research Centre Europe.

Hankel matrices. In addition, we provide an interpretation of the method in terms of forward and backward recursions for automata and grammars. This provides extra intuitions about the method and stresses the importance of matrix factorization for learning automata and grammars. We believe that this complementary perspective might be appealing for an NLP audience and serve to put spectral learning in a wider and, perhaps for some, more familiar context. Our hope is that this will broaden the understanding of these methods by the NLP community and empower many researchers to apply these techniques to novel problems.

The content of the tutorial will be divided into four blocks of 45 minutes each, as follows. The first block will introduce the basic definitions of weighted automata and Hankel matrices, and present a key connection between the fundamental theorem of weighted automata and learning. In the second block we will discuss the case of probabilistic automata in detail, touching upon all aspects from the underlying theory to the tricks required to achieve accurate and scalable learning algorithms. The third block will present extensions to related models, including sequence tagging models, finite-state transducers and weighted context-free grammars. The last block will describe a general framework for using spectral techniques in more general situations where a matrix completion pre-processing step is required; several applications of this approach will be described.

Tutorial 3

Semantic Parsing with Combinatory Categorial Grammars

Yoav Artzi, Nicholas FitzGerald, and Luke Zettlemoyer (University of Washington)

Saturday, October 25, 2014, 14:00 – 17:30pm

Al Areen 4

Semantic parsers map natural language sentences to formal representations of their underlying meaning. Building accurate semantic parsers without prohibitive engineering costs is a long-standing, open research problem.

The tutorial will describe general principles for building semantic parsers. The presentation will be divided into two main parts: learning and modeling. In the learning part, we will describe a unified approach for learning Combinatory Categorial Grammar (CCG) semantic parsers, that induces both a CCG lexicon and the parameters of a parsing model. The approach learns from data with labeled meaning representations, as well as from more easily gathered weak supervision. It also enables grounded learning where the semantic parser is used in an interactive environment, for example to read and execute instructions. The modeling section will include best practices for grammar design and choice of semantic representation. We will motivate our use of lambda calculus as a language for building and representing meaning with examples from several domains.

The ideas we will discuss are widely applicable. The semantic modeling approach, while implemented in lambda calculus, could be applied to many other formal languages. Similarly, the algorithms for inducing CCG focus on tasks that are formalism independent, learning the meaning of words and estimating parsing parameters. No prior knowledge of CCG is required. The tutorial will be backed by implementation and experiments in the University of Washington Semantic Parsing Framework (UW SPF, http://yoavartzi.com/spf).

Yoav Artzi is a Ph.D. candidate in the Computer Science & Engineering department at the University of Washington. His research studies semantic meaning and machine learning methods for natural language understanding within situated interactive systems. He is currently supported by the 2014 Microsoft Research PhD Fellowship, and is a recipient of the 2012 Yahoo KSC award.

Nicholas FitzGerald is a Ph.D. student at the University of Washington. His research interests are grounded natural language understanding and generation. He is a recipient of an Intel Science and Technology Center Fellowship and an NSERC Postgraduate Scholarship.

Luke Zettlemoyer is an Assistant Professor in the Computer Science & Engineering department at the University of Washington. His research interests are in the intersections of natural language processing, machine learning and decision making under uncertainty. Honors include best paper awards at UAI 2005 and ACL 2009, selection to the DARPA CSSG, and an NSF CAREER Award.

Tutorial 4

Linear Programming Decoders in Natural Language Processing: From Integer Programming to Message Passing and Dual Decomposition

André F. T. Martins (Priberam Labs.)

Saturday, October 25, 2014, 14:00 – 17:30pm

Al Areen 5

This tutorial will cover the theory and practice of linear programming decoders. This class of decoders encompasses a variety of techniques that have enjoyed great success in devising structured models for natural language processing (NLP). Along the tutorial, we provide a unified view of different algorithms and modeling techniques, including belief propagation, dual decomposition, integer linear programming, Markov logic, and constrained conditional models. Various applications in NLP will serve as a motivation. There is a long string of work using integer linear programming (ILP) formulations in NLP, for example in semantic role labeling, machine translation, summarization, dependency parsing, coreference resolution, and opinion mining, to name just a few. At the heart of these approaches is the ability to encode logic and budget constraints (common in NLP and information retrieval) as linear inequalities. Thanks to general purpose solvers (such as Gurobi, CPLEX, or GLPK), the practitioner can abstract away from the decoding algorithm and focus on developing a powerful model. A disadvantage, however, is that general solvers do not scale well to large problem instances, since they fail to exploit the structure of the problem.

This is where graphical models come into play. In this tutorial, we show that most logic and budget constraints that arise in NLP can be cast in this framework. This opens the door for the use of message-passing algorithms, such as belief propagation and variants thereof. An alternative are algorithms based on dual decomposition, such as the subgradient method or AD3. These algorithms have achieved great success in a variety of applications, such as parsing, corpus-wide tagging, machine translation, summarization, joint coreference resolution and quotation attribution, and semantic role labeling. Interestingly, most decoders used in these works can be regarded as structure-aware solvers for addressing relaxations of integer linear programs. All these algorithms have a similar consensus-based architecture: they repeatedly perform certain "local" operations in the graph, until some form of local agreement is achieved. The local operations are performed at each factor, and they range between computing marginals, max-marginals, an optimal configuration, or a small quadratic problem, all of which are commonly tractable and efficient in a wide range of problems.

As a companion of this tutorial, we provide an open-source implementation of some of the algorithms described above, available at http://www.ark.cs.cmu.edu/AD3.

André F. T. Martins is a research scientist at Priberam Labs. He received his dual-degree PhD in Language Technologies in 2012 from Carnegie Mellon University and Instituto Superior Técnico. His PhD dissertation was awarded Honorable Mention in CMU's SCS Dissertation Award competition. Martins' research interests include natural language processing, machine learning, structured prediction, sparse modeling, and optimization. His paper "Concise Integer Linear Programming Formulations for Dependency Parsing" received a best paper award at ACL 2009.

3

Workshops: Saturday, October 25

Overview

Workshop 1: Arabic NLP: Arabic Natural Language Processing Workshop

Organizers: *Nizar Habash, Stephan Vogel, Kareem Darwish, Behrang Mohit, Nadi Tomeh, and Wajdi Zaghouani*

Venue: Al Areen 3

Saturday, October 25, 2014

Session 1: Corpora

9:00–9:20 Using Twitter to Collect a Multi-Dialectal Corpus of Arabic
Hamdy Mubarak and Kareem Darwish

9:20–9:40 The International Corpus of Arabic: Compilation, Analysis and Evaluation
Sameh Alansary and Magdy Nagi

9:45–10:05 Building a Corpus for Palestinian Arabic: a Preliminary Study
Mustafa Jarrar, Nizar Habash, Diyam Akra, and Nasser Zalmout

10:05–10:25 Annotating corpus data for a quantitative, constructional analysis of motion verbs in Modern Standard Arabic
Dana Abdulrahim

10:30–11:00 **Break / Poster setup**

Shared Task

11:00–11:30 The First QALB Shared Task on Automatic Text Correction for Arabic
Behrang Mohit, Alla Rozovskaya, Nizar Habash, Wajdi Zaghouani, and Ossama Obeid

11:30–11:45 Shared Task: 1-minute Summary for Shared Task Participants
Shared Task participants

11:45–12:15 Shared Task: Panel
Group Discussion

12:15–12:30 Main Workshop Poster Teaser 1-minute Summary
Main Workshop participants

12:30–2:00 **Lunch / Main and Shared Task Poster Session**

Main and Shared Task Poster Session

12:30–2:00 Main Workshop Posters
Main Workshop participants

- A Framework for the Classification and Annotation of Multiword Expressions in Dialectal Arabic
Abdelati Hawwari, Mohammed Attia, and Mona Diab

- Al-Bayan: An Arabic Question Answering System for the Holy Quran
Heba Abdelnasser, Maha Ragab, Reham Mohamed, Alaa Mohamed, Bassant Farouk, Nagwa El-Makky, and Marwan Torki

- Automatic Arabic diacritics restoration based on deep nets
Ahmad Al Sallab, Mohsen Rashwan, Hazem Raafat, and Ahmed Rafea

- Combining strategies for tagging and parsing Arabic
Maytham Alabbas and Allan Ramsay

- Named Entity Recognition System for Dialectal Arabic
  Ayah Zirikly and Mona Diab
- Semantic Query Expansion for Arabic Information Retrieval
  Ashraf Mahgoub, Mohsen Rashwan, Hazem Raafat, Mohamed Zahran, and Magda Fayek
- Transliteration of Arabizi into Arabic Orthography: Developing a Parallel Annotated Arabizi-Arabic Script SMS/Chat Corpus
  Ann Bies, Zhiyi Song, Mohamed Maamouri, Stephen Grimes, Haejoong Lee, Jonathan Wright, Stephanie Strassel, Nizar Habash, Ramy Eskander, and Owen Rambow
- Tunisian dialect Wordnet creation and enrichment using web resources and other Wordnets
  Rihab Bouchlaghem, Aymen Elkhlifi, and Rim Faiz

12:30–2:00 Shared Task Posters
Shared Task participants

- A Pipeline Approach to Supervised Error Correction for the QALB-2014 Shared Task
  Nadi Tomeh, Nizar Habash, Ramy Eskander, and Joseph Le Roux
- Arabic Spelling Correction using Supervised Learning
  Youssef Hassan, Mohamed Aly, and Amir Atiya
- Autocorrection of arabic common errors for large text corpus
  Taha Zerrouki, Khaled Alhawiti, and Amar Balla
- Automatic Correction of Arabic Text: a Cascaded Approach
  Hamdy Mubarak and Kareem Darwish
- CMUQ@QALB-2014: An SMT-based System for Automatic Arabic Error Correction
  Serena Jeblee, Houda Bouamor, Wajdi Zaghouani, and Kemal Oflazer
- Fast and Robust Arabic Error Correction System
  Michael Nawar and Moheb Ragheb
- GWU-HASP: Hybrid Arabic Spelling and Punctuation Corrector
  Mohammed Attia, Mohamed Al-Badrashiny, and Mona Diab
- TECHLIMED system description for the Shared Task on Automatic Arabic Error Correction
  Djamel Mostefa, Omar Abayou, and Ramzi Abbes
- The Columbia System in the QALB-2014 Shared Task on Arabic Error Correction
  Alla Rozovskaya, Nizar Habash, Ramy Eskander, Noura Farra, and Wael Salloum

Session 2: Text Mining

2:00–2:20 A Large Scale Arabic Sentiment Lexicon for Arabic Opinion Mining
Gilbert Badaro, Ramy Baly, Hazem Hajj, Nizar Habash, and Wassim El-Hajj

2:20–2:40 Evaluating Distant Supervision for Subjectivity and Sentiment Analysis on Arabic Twitter Feeds
Eshrag Refaee and Verena Rieser

2:45–3:05 Arabic Native Language Identification
Shervin Malmasi and Mark Dras

3:05–3:25 AIDArabic A Named-Entity Disambiguation Framework for Arabic Text
Mohamed Amir Yosef, Marc Spaniol, and Gerhard Weikum

3:30–4:00 **Break**

Session 3: Translation & Transliteration

4:00–4:20 Domain and Dialect Adaptation for Machine Translation into Egyptian Arabic
Serena Jeblee, Weston Feely, Houda Bouamor, Alon Lavie, Nizar Habash, and Kemal Oflazer

4:25–4:45 Unsupervised Word Segmentation Improves Dialectal Arabic to English Machine Translation
Kamla Al-Mannai, Hassan Sajjad, Alaa Khader, Fahad Al Obaidli, Preslav Nakov, and Stephan Vogel

4:50–5:10 Arabizi Detection and Conversion to Arabic
Kareem Darwish

Closing Session

5:10–6:00 Workshop Group Discussion
Group Discussion

Workshop 2: CodeSwitch: Computational Approaches to Linguistic Code Switching

Organizers: *Mona Diab, Pascale Fung, Julia Hirschberg, and Thamar Solorio*

Venue: Al Areen 6

Saturday, October 25, 2014

Session 1: Workshop talks

9:00–9:10 Welcome Remarks
The organizers

9:10–9:30 Foreign Words and the Automatic Processing of Arabic Social Media Text Written in Roman Script
Ramy Eskander, Mohamed Al-Badrashiny, Nizar Habash, and Owen Rambow

9:30–9:50 Code Mixing: A Challenge for Language Identification in the Language of Social Media
Utsab Barman, Amitava Das, Joachim Wagner, and Jennifer Foster

9:50–10:10 Detecting Code-Switching in a Multilingual Alpine Heritage Corpus
Martin Volk and Simon Clematide

10:10–10:30 Exploration of the Impact of Maximum Entropy in Recurrent Neural Network Language Models for Code-Switching Speech
Ngoc Thang Vu and Tanja Schultz

10:30–11:00 **Coffee Break**

Session 2: Workshop Talks and Shared Task Systems

11:00–11:20 Predicting Code-switching in Multilingual Communication for Immigrant Communities
Evangelos Papalexakis, Dong Nguyen, and A. Seza Doğruöz

11:20–11:40 Twitter Users #CodeSwitch Hashtags! #MoltoImportante #wow
David Jurgens, Stefan Dimitrov, and Derek Ruths

11:40–11:50 Overview for the First Shared Task on Language Identification in Code-Switched Data
Thamar Solorio, Elizabeth Blair, Suraj Maharjan, Steven Bethard, Mona Diab, Mahmoud Ghoneim, Abdelati Hawwari, Fahad AlGhamdi, Julia Hirschberg, Alison Chang, and Pascale Fung

11:50–12:10 Word-level Language Identification using CRF: Code-switching Shared Task Report of MSR India System
Gokul Chittaranjan, Yogarshi Vyas, Kalika Bali, and Monojit Choudhury

12:10–12:30 The CMU Submission for the Shared Task on Language Identification in Code-Switched Data
Chu-Cheng Lin, Waleed Ammar, Lori Levin, and Chris Dyer

12:30–2:00 **Lunch break**

Session 3: Shared Task and Next Steps

2:00–2:20 Language Identification in Code-Switching Scenario
Naman Jain and Riyaz Ahmad Bhat

2:20–2:40 AIDA: Identifying Code Switching in Informal Arabic Text
Heba Elfardy, Mohamed Al-Badrashiny, and Mona Diab

2:40–3:00 The IUCL+ System: Word-Level Language Identification via Extended Markov Models
Levi King, Eric Baucom, Timur Gilmanov, Sandra Kübler, Dan Whyatt, Wolfgang Maier, and Paul Rodrigues

3:00–3:30 Panel Discussion: Next Steps in CS Research
Group Discussion

3:30–4:00 **Coffee Break (Posters set up time)**

Session 4: Poster Session

4:00–5:30 Workshop and Shared Task Posters
Multiple presenters

- Mixed Language and Code-Switching in the Canadian Hansard
  Marine Carpuat
- "I am borrowing ya mixing ?" An Analysis of English-Hindi Code Mixing in Facebook
  Kalika Bali, Jatin Sharma, Monojit Choudhury, and Yogarshi Vyas
- DCU-UVT: Word-Level Language Classification with Code-Mixed Data
  Utsab Barman, Joachim Wagner, Grzegorz Chrupała, and Jennifer Foster
- Incremental N-gram Approach for Language Identification in Code-Switched Text
  Prajwol Shrestha
- The Tel Aviv University System for the Code-Switching Workshop Shared Task
  Kfir Bar and Nachum Dershowitz
- The CMU Submission for the Shared Task on Language Identification in Code-Switched Data
  Chu-Cheng Lin, Waleed Ammar, Lori Levin, and Chris Dyer
- Word-level Language Identification using CRF: Code-switching Shared Task Report of MSR India System
  Gokul Chittaranjan, Yogarshi Vyas, Kalika Bali, and Monojit Choudhury
- Language Identification in Code-Switching Scenario
  Naman Jain and Riyaz Ahmad Bhat
- AIDA: Identifying Code Switching in Informal Arabic Text
  Heba Elfardy, Mohamed Al-Badrashiny, and Mona Diab
- The IUCL+ System: Word-Level Language Identification via Extended Markov Models
  Levi King, Eric Baucom, Timur Gilmanov, Sandra Kübler, Dan Whyatt, Wolfgang Maier, and Paul Rodrigues

Workshop 3: MOOC: Modeling Large Scale Social Interaction in Massively Open Online Courses

Organizers: *Carolyn Rosé and George Siemens*

Venue: Al Areen 2

Saturday, October 25, 2014

Session 1

9:00–9:20 Opening Remarks
The organizers

9:20–10:30 Keynote: Data Archeology: A theory informed approach to analyzing data traces of social interaction in large scale learning environments
Alyssa Wise

10:30–10:50 **Coffee Break**

Session 2

10:50–11:15 Your click decides your fate: Inferring Information Processing and Attrition Behavior from MOOC Video Clickstream Interactions
Tanmay Sinha, Patrick Jermann, Nan Li, and Pierre Dillenbourg

11:15–11:40 Identifying Student Leaders from MOOC Discussion Forums through Language Influence
Seungwhan Moon, Saloni Potdar, and Lara Martin

11:40–12:05 Towards Identifying the Resolvability of Threads in MOOCs
Diyi Yang, Miaomiao Wen, and Carolyn Rosé

12:05–12:30 Point-of-View Mining and Cognitive Presence in MOOCs: A (Computational) Linguistics Perspective
Noureddine Elouazizi

12:30–2:00 **Lunch**

Session 3

2:00–3:10 Keynote Talk: Analytics: climbing up the ladder of behavior control
Patrick Jermann

3:10–3:30 Shared Task on Prediction of Dropout Over Time in Massively Open Online Courses
Carolyn Rosé and George Siemens

3:30–4:00 **Coffe Break**

Session 4

4:00–4:20 Capturing "attrition intensifying" structural traits from didactic interaction sequences of MOOC learners
Tanmay Sinha, Nan Li, Patrick Jermann, and Pierre Dillenbourg

4:20–4:40 A Process for Predicting MOOC Attrition
Mike Sharkey and Robert Sanders

4:40–5:00 Predicting Attrition Along the Way: The UIUC Model
Bussaba Amnueypornsakul, Suma Bhat, and Phakpoom Chinprutthiwong

5:00–5:20 Predicting MOOC Dropout over Weeks Using Machine Learning Methods
Marius Kloft, Felix Stiehler, Zhilin Zheng, and Niels Pinkwart

Workshop 4: SSST-8: Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation

Organizers: *Dekai Wu, Marine Carpuat, Xavier Carreras, and Eva Maria Vecchi*

Venue: Al Areen 1

Saturday, October 25, 2014

Session 1: Morning Orals

9:00–9:10 Opening Remarks
Dekai Wu, Marine Carpuat, Xavier Carreras, and Eva Maria Vecchi

9:10–9:30 Vector Space Models for Phrase-based Machine Translation
Tamer Alkhouli, Andreas Guta, and Hermann Ney

9:30–9:50 Bilingual Markov Reordering Labels for Hierarchical SMT
Gideon Maillette de Buy Wenniger and Khalil Sima'an

9:50–10:10 Better Semantic Frame Based MT Evaluation via Inversion Transduction Grammars
Dekai Wu, Chi-kiu Lo, Meriem Beloucif, and Markus Saers

10:10–10:30 Rule-based Syntactic Preprocessing for Syntax-based Machine Translation
Yuto Hatakoshi, Graham Neubig, Sakriani Sakti, Tomoki Toda, and Satoshi Nakamura

10:30–11:00 **Coffee break**

Invited talk by Timothy Baldwin

11:00–12:00 Composed, Distributed Reflections on Semantics and Statistical Machine Translation
Timothy Baldwin

Session 2: Morning Spotlights

12:00–12:05 Applying HMEANT to English-Russian Translations
Alexander Chuchunkov, Alexander Tarelkin, and Irina Galinskaya

12:05–12:10 Reducing the Impact of Data Sparsity in Statistical Machine Translation
Karan Singla, Kunal Sachdeva, Srinivas Bangalore, Dipti Misra Sharma, and Diksha Yadav

12:10–12:15 Expanding the Language model in a low-resource hybrid MT system
George Tambouratzis, Sokratis Sofianopoulos, and Marina Vassiliou

12:15–12:20 Syntax and Semantics in Quality Estimation of Machine Translation
Rasoul Kaljahi, Jennifer Foster, and Johann Roturier

12:20–12:25 Overcoming the Curse of Sentence Length for Neural Machine Translation using Automatic Segmentation
Jean Pouget-Abadie, Dzmitry Bahdanau, Bart van Merrienboer, Kyunghyun Cho, and Yoshua Bengio

12:25–12:30 Ternary Segmentation for Improving Search in Top-down Induction of Segmental ITGs
Markus Saers and Dekai Wu

12:30–2:00 **Lunch break**

Session 3: Afternoon Orals and Spotlights

2:00–2:20 A CYK+ Variant for SCFG Decoding Without a Dot Chart
Rico Sennrich

2:20–2:40 On the Properties of Neural Machine Translation: Encoder–Decoder Approaches
Kyunghyun Cho, Bart van Merrienboer, Dzmitry Bahdanau, and Yoshua Bengio

2:40–3:00 Transduction Recursive Auto-Associative Memory: Learning Bilingual Compositional Distributed Vector Representations of Inversion Transduction Grammars
Karteek Addanki and Dekai Wu

3:00–3:20 Transformation and Decomposition for Efficiently Implementing and Improving Dependency-to-String Model In Moses
Liangyou Li, Jun Xie, Andy Way, and Qun Liu

3:20–3:25 Word's Vector Representations meet Machine Translation
Eva Martinez Garcia, Jörg Tiedemann, Cristina España-Bonet, and Lluís Màrquez

3:25–3:30 Context Sense Clustering for Translation
João Casteleiro, Gabriel Lopes, and Joaquim Silva

3:30–4:00 **Coffee break**

Session 4: Afternoon Spotlights

4:00–4:05 Evaluating Word Order Recursively over Permutation-Forests
Miloš Stanojević and Khalil Sima'an

4:05–4:10 Preference Grammars and Soft Syntactic Constraints for GHKM Syntax-based Statistical Machine Translation
Matthias Huck, Hieu Hoang, and Philipp Koehn

4:10–4:15 How Synchronous are Adjuncts in Translation Data?
Sophie Arnoult and Khalil Sima'an

Poster session

4:15–5:30 Poster session of all workshop papers
All workshop presenters

4

Main Conference: Sunday, October 26

Overview

8:00 – 17:00 **Registration**				(Area 3)
8:00 – 8:40 **Refreshments**				(Area 3-4)
8:40 – 9:00 **Opening session**				(Al Areen 1-4)
9:00 – 10:00 **Invited Talk: Salim Roukos**				(Al Areen 1-4)
10:00 – 10:30 **Coffee Break**				(Area 3-4)

Session 1

10:30 – 12:10	Neural Net Mixer	Discourse, Dialogue and Pragmatics	Segmentation / Spoken Language	Posters: Machine Translation

12:10 – 13:30 **Lunch Break**

Session 2

12:30 – 15:10	Computational Psycholinguistics	Tagging, Chunking, Parsing and Syntax	NLP for the Web and Social Media	Posters: Semantics

15:10 – 15:40 **Coffee Break**   (Area 3-4)

Session 3

15:40 – 17:20	Sentiment Analysis and Opinion Mining	Semantics	Machine Translation	Posters: Information Retrieval, Summarization and Question Answering

25

Invited Talk: Salim Roukos

"IBM Cognitive Computing – An NLP Renaissance!"

Sunday, October 26, 2014, 9:00–10:00

Al Areen 1-4

Abstract: Electronically available multi-modal data (primarily text and meta-data) is unprecedented in terms of its volume, variety, velocity, (and veracity). The increased interest and investment in cognitive computing for building systems and solutions that enable and support richer human-machine interactions presents a unique opportunity for novel statistical models for natural language processing.

In this talk, I will describe a journey at IBM during the past three decades in developing novel statistical models for NLP covering statistical parsing, machine translation, and question-answering systems. Along with a discussion of some of the recent successes, I will discuss some difficult challenges that need to be addressed to achieve more effective cognitive systems and applications.

Biography: Salim Roukos is Senior Manager of Multi-Lingual NLP and CTO for Translation Technologies at IBM T. J. Watson Research Center. Dr. Roukos received his B.E. from the American University of Beirut, in 1976, his M.Sc. and Ph.D. from the University of Florida, in 1978 and 1980, respectively. He joined Bolt Beranek and Newman from 1980 through 1989, where he was a Senior Scientist in charge of projects in speech compression, time scale modification, speaker identification, word spotting, and spoken language understanding. He was an Adjunct Professor at Boston University in 1988 before joining IBM in 1989. Dr. Roukos has served as Chair of the IEEE Digital Signal Processing Committee in 1988.

Salim Roukos currently leads a group at IBM T.J. Watson research Center that focuses on various problems using machine learning techniques for natural language processing. The group pioneered many of the statistical methods for NLP from statistical parsing, to natural language understanding, to statistical machine translation and machine translation evaluation metrics (BLEU metric). Roukos has over a 150 publications in the speech and language areas and over two dozen patents. Roukos was the lead of the group which introduced the first commercial statistical language understanding system for conversational telephony systems (IBM ViaVoice Telephony) in 2000 and the first statistical machine translation product for Arabic-English translation in 2003. He has recently lead the effort to create IBM's offering of IBM Real-Time Translation Services (RTTS) a platform for enabling real-time translation applications such as multilingual chat and on-demand document translation.

Parallel Session 1 Overview – Sunday, October 26, 2014

Track A	Track B	Track C	
Neural Net Mixer	*Discourse, Dialogue and Pragmatics*	*Segmentation / Spoken Language*	
Al Areen 1-2	Al Areen 3	Al Areen 4	
Modeling Interestingness with Deep Neural Networks *Jianfeng Gao, Patrick Pantel, Michael Gamon, Xiaodong He, and Li Deng*	Identifying Argumentative Discourse Structures in Persuasive Essays *Christian Stab and Iryna Gurevych*	Semi-Supervised Chinese Word Segmentation Using Partial-Label Learning With Conditional Random Fields *Fan Yang and Paul Vozila*	10:30
Translation Modeling with Bidirectional Recurrent Neural Networks *Martin Sundermeyer, Tamer Alkhouli, Joern Wuebker, and Hermann Ney*	Policy Learning for Domain Selection in an Extensible Multi-domain Spoken Dialogue System *Zhuoran Wang, Hongliang Chen, Guanchun Wang, Hao Tian, Hua Wu, and Haifeng Wang*	Accurate Word Segmentation and POS Tagging for Japanese Microblogs: Corpus Annotation and Joint Modeling with Lexical Normalization *Nobuhiro Kaji and Masaru Kitsuregawa*	10:55
A Neural Network Approach to Selectional Preference Acquisition *Tim Van de Cruys*	A Constituent-Based Approach to Argument Labeling with Joint Inference in Discourse Parsing *Fang Kong, Hwee Tou Ng, and Guodong Zhou*	Revisiting Embedding Features for Simple Semi-supervised Learning *Jiang Guo, Wanxiang Che, Haifeng Wang, and Ting Liu*	11:20
Learning Image Embeddings using Convolutional Neural Networks for Improved Multi-Modal Semantics *Douwe Kiela and Léon Bottou*	Strongly Incremental Repair Detection *Julian Hough and Matthew Purver*	Combining Punctuation and Disfluency Prediction: An Empirical Study *Xuancong Wang, Khe Chai Sim, and Hwee Tou Ng*	11:45

Track P

Posters: Machine Translation

Al Areen 5-6   10:30-12:10

- Submodularity for Data Selection in Machine Translation
  Katrin Kirchhoff and Jeff Bilmes

- Improve Statistical Machine Translation with Context-Sensitive Bilingual Semantic Embedding Model
  Haiyang Wu, Daxiang Dong, Xiaoguang Hu, Dianhai Yu, Wei He, Hua Wu, Haifeng Wang, and Ting Liu

- Transformation from Discontinuous to Continuous Word Alignment Improves Translation Quality
  Zhongjun He, Hua Wu, Haifeng Wang, and Ting Liu

- Unsupervised Word Alignment Using Frequency Constraint in Posterior Regularized EM
  Hidetaka Kamigaito, Taro Watanabe, Hiroya Takamura, and Manabu Okumura

- Asymmetric Features Of Human Generated Translation
  Sauleh Eetemadi and Kristina Toutanova

- Syntax-Augmented Machine Translation using Syntax-Label Clustering
  Hideya Mino, Taro Watanabe, and Eiichiro Sumita

- Testing for Significance of Increased Correlation with Human Judgment
  Yvette Graham and Timothy Baldwin

- Syntactic SMT Using a Discriminative Text Generation Model
  Yue Zhang, Kai Song, Linfeng Song, Jingbo Zhu, and Qun Liu

- Learning Hierarchical Translation Spans
  Masao Utiyama, Eiichiro Sumita, and Hai Zhao

- Neural Network Based Bilingual Language Model Growing for Statistical Machine Translation
  Rui Wang, Hai Zhao, Bao-Liang Lu, Masao Utiyama, and Eiichiro Sumita

- Better Statistical Machine Translation through Linguistic Treatment of Phrasal Verbs
  Kostadin Cholakov and Valia Kordoni

- Fitting Sentence Level Translation Evaluation with Many Dense Features
  Miloš Stanojević and Sima'an Khalil

- A Human Judgement Corpus and a Metric for Arabic MT Evaluation
  Houda Bouamor, Hanan Alshikhabobakr, Behrang Mohit, and Kemal Oflazer

- Learning to Differentiate Better from Worse Translations
  Francisco Guzmán, Shafiq Joty, Lluís Màrquez, Alessandro Moschitti, Preslav Nakov, and Massimo Nicosia

- Two Improvements to Left-to-Right Decoding for Hierarchical Phrase-based Machine Translation
  Maryam Siahbani and Anoop Sarkar

- Reordering Model for Forest-to-String Machine Translation
  Martin Cmejrek

Parallel Session 1

Session 1a: Neural Net Mixer
Al Areen 1-2 *Chair: Holger Schwenk*

Modeling Interestingness with Deep Neural Networks
Jianfeng Gao, Patrick Pantel, Michael Gamon, Xiaodong He, and Li Deng 10:30–10:55

This paper presents a deep semantic similarity model (DSSM), a special type of deep neural networks designed for text analysis, for recommending target documents to be of interest to a user based on a source document that she is reading. We observe, identify, and detect naturally occurring signals of interestingness in click transitions on the Web between source and target documents, which we collect from commercial Web browser logs. The DSSM is trained on millions of Web transitions, and maps source-target document pairs to feature vectors in a latent space in such a way that the distance between source documents and their corresponding interesting targets in that space is minimized. The effectiveness of the DSSM is demonstrated using two interestingness tasks: automatic high-lighting and contextual entity search. The results on large-scale, real-world datasets show that the semantics of documents are important for modeling interestingness and that the DSSM leads to significant quality improvement on both tasks, outperforming not only the classic document models that do not use semantics but also state-of-the-art topic models.

Translation Modeling with Bidirectional Recurrent Neural Networks
Martin Sundermeyer, Tamer Alkhouli, Joern Wuebker, and Hermann Ney 10:55–11:20

This work presents two different translation models using recurrent neural networks. The first one is a word-based approach using word alignments. Second, we present phrase-based translation models that are more consistent with phrase-based decoding. Moreover, we introduce bidirectional recurrent neural models to the problem of machine translation, allowing us to use the full source sentence in our models, which is also of theoretical interest. We demonstrate that our translation models are capable of improving strong baselines already including recurrent neural language models on three tasks: IWSLT 2013 German->English, BOLT Arabic->English and Chinese->English. We obtain gains up to 1.6% BLEU and 1.7% TER by rescoring 1000-best lists.

A Neural Network Approach to Selectional Preference Acquisition
Tim Van de Cruys 11:20–11:45

This paper investigates the use of neural networks for the acquisition of selectional preferences. Inspired by recent advances of neural network models for NLP applications, we propose a neural network model that learns to discriminate between felicitous and infelicitous arguments for a particular predicate. The model is entirely unsupervised - preferences are learned from unannotated corpus data. We propose two neural network architectures: one that handles standard two-way selectional preferences and one that is able to deal with multi-way selectional preferences. The model's performance is evaluated on a pseudo-disambiguation task, on which it is shown to achieve state of the art performance.

Learning Image Embeddings using Convolutional Neural Networks for Improved Multi-Modal Semantics
Douwe Kiela and Léon Bottou 11:45–12:10

We construct multi-modal concept representations by concatenating a skip-gram linguistic representation vector with a visual concept representation vector computed using the feature extraction layers of a deep convolutional neural network (CNN) trained on a large labeled object recognition dataset. This transfer learning approach brings a clear performance gain over features based on the traditional bag-of-visual-word approach. Experimental results are reported on the WordSim353 and MEN word relatedness evaluation tasks. We use visual features computed using either ImageNet or ESP Game images.

Session 1b: Discourse, Dialogue and Pragmatics
Al Areen 3   *Chair: Giuseppe Carenini*

Identifying Argumentative Discourse Structures in Persuasive Essays
Christian Stab and Iryna Gurevych   10:30–10:55

In this paper, we present a novel approach for identifying argumentative discourse structures in persuasive essays. The structure of argumentation consists of several components (i.e. claims and premises) that are connected with argumentative relations. We consider this task in two consecutive steps. First, we identify the components of arguments using multiclass classification. Second, we classify a pair of argument components as either support or non-support for identifying the structure of argumentative discourse. For both tasks, we evaluate several classifiers and propose novel feature sets including structural, lexical, syntactic and contextual features. In our experiments, we obtain a macro F1-score of 0.726 for identifying argument components and 0.722 for argumentative relations.

Policy Learning for Domain Selection in an Extensible Multi-domain Spoken Dialogue System
Zhuoran Wang, Hongliang Chen, Guanchun Wang, Hao Tian, Hua Wu, and Haifeng Wang   10:55–11:20

This paper proposes a Markov Decision Process and reinforcement learning based approach for domain selection in a multi-domain Spoken Dialogue System built on a distributed architecture. In the proposed framework, the domain selection problem is treated as sequential planning instead of classification, such that confirmation and clarification interaction mechanisms are supported. In addition, it is shown that by using a model parameter tying trick, the extensibility of the system can be preserved, where dialogue components in new domains can be easily plugged in, without re-training the domain selection policy. The experimental results based on human subjects suggest that the proposed model marginally outperforms a non-trivial baseline.

A Constituent-Based Approach to Argument Labeling with Joint Inference in Discourse Parsing
Fang Kong, Hwee Tou Ng, and Guodong Zhou   11:20–11:45

Discourse parsing is a challenging task and plays a critical role in discourse analysis. In this paper, we focus on labeling full argument spans of discourse connectives in the Penn Discourse Treebank (PDTB). Previous studies cast this task as a linear tagging or subtree extraction problem. In this paper, we propose a novel constituent-based approach to argument labeling, which integrates the advantages of both linear tagging and subtree extraction. In particular, the proposed approach unifies intra- and inter-sentence cases by treating the immediately preceding sentence as a special constituent. Besides, a joint inference mechanism is introduced to incorporate global information across arguments into our constituent-based approach via integer linear programming. Evaluation on PDTB shows significant performance improvements of our constituent-based approach over the best state-of-the-art system. It also shows the effectiveness of our joint inference mechanism in modeling global information across arguments.

Strongly Incremental Repair Detection
Julian Hough and Matthew Purver   11:45–12:10

We present STIR (STrongly Incremental Repair detection), a system that detects speech repairs and edit terms on transcripts incrementally with minimal latency. STIR uses information-theoretic measures from n-gram models as its principal decision features in a pipeline of classifiers detecting the different stages of repairs. Results on the Switchboard disfluency tagged corpus show utterance-final accuracy on a par with state-of-the-art incremental repair detection methods, but with better incremental accuracy, faster time-to-detection and less computational overhead. We evaluate its performance using incremental metrics and propose new repair processing evaluation standards.

Session 1c: Segmentation / Spoken Language
Al Areen 4   *Chair: Stephan Vogel*

Semi-Supervised Chinese Word Segmentation Using Partial-Label Learning With Conditional Random Fields
Fan Yang and Paul Vozila   10:30–10:55

There is rich knowledge encoded in online web data. For example, punctuation and entity tags in Wikipedia data define some word boundaries in a sentence. In this paper we adopt partial-label learning with conditional random fields to make use of this valuable knowledge for semi-supervised Chinese word segmentation. The basic idea of partial-label learning is to optimize a cost function that marginalizes the probability mass in the constrained space that encodes this knowledge. By integrating some domain adaptation techniques, such as EasyAdapt, our result reaches an F-measure of 95.98% on the CTB-6 corpus, a significant improvement from both the supervised baseline and a previous proposed approach, namely constrained decode.

Accurate Word Segmentation and POS Tagging for Japanese Microblogs: Corpus Annotation and Joint Modeling with Lexical Normalization
Nobuhiro Kaji and Masaru Kitsuregawa   10:55–11:20

Microblogs have recently received widespread interest from NLP researchers. However, current tools for Japanese word segmentation and POS tagging still perform poorly on microblog texts. We developed an annotated corpus and proposed a joint model for overcoming this situation. Our developed annotated corpus of microblog texts enables not only training of accurate statistical models but also quantitative evaluation of their performance. Our joint model with lexical normalization handles the orthographic diversity of microblog texts. We conducted an experiment to demonstrate that the corpus and model substantially contribute to boosting accuracy.

Revisiting Embedding Features for Simple Semi-supervised Learning
Jiang Guo, Wanxiang Che, Haifeng Wang, and Ting Liu   11:20–11:45

Recent work has shown success in using continuous word embeddings learned from unlabeled data as features to improve supervised NLP systems, which is regarded as a simple semi-supervised learning mechanism. However, fundamental problems on effectively incorporating the word embedding features within the framework of linear models remain. In this study, we investigate and analyze three different approaches, including a new proposed distributional prototype approach, for utilizing the embedding features. The presented approaches can be integrated into most of the classical linear models in NLP. Experiments on the task of named entity recognition show that each of the proposed approaches can better utilize the word embedding features, among which the distributional prototype approach performs the best. Moreover, the combination of the approaches provides additive improvements, outperforming the dense and continuous embedding features by nearly 2 points of F1 score.

Combining Punctuation and Disfluency Prediction: An Empirical Study
Xuancong Wang, Khe Chai Sim, and Hwee Tou Ng   11:45–12:10

Punctuation prediction and disfluency prediction can improve downstream natural language processing tasks such as machine translation and information extraction. Combining the two tasks can potentially improve the efficiency of the overall pipeline system and reduce error propagation. In this work[1], we compare various methods for combining punctuation prediction (PU) and disfluency prediction (DF) on the Switchboard corpus. We compare an isolated prediction approach with a cascade approach, a rescoring approach, and three joint model approaches. For the cascade approach, we show that the soft cascade method is better than the hard cascade method. We also use the cascade models to generate an n-best list, use the bi-directional cascade models to perform rescoring, and compare that with the results of the cascade models. For the joint model approach, we compare mixed-label Linear-chain Conditional Random Field (LCRF), cross-product LCRF and 2-layer Factorial Conditional Random Field (FCRF) with soft-cascade LCRF. Our results show that the various methods linking the two tasks are not significantly different from one another, although they perform better than the isolated prediction method by 0.5–1.5

---

[1]The research reported in this paper was carried out as part of the PhD thesis research of Xuancong Wang at the NUS Graduate School for Integrated Sciences and Engineering.

Session 1-posters: Posters: Machine Translation
Al Areen 5-6 10:30–12:10

Submodularity for Data Selection in Machine Translation
Katrin Kirchhoff and Jeff Bilmes

We introduce submodular optimization to the problem of training data subset selection for statistical machine translation (SMT). By explicitly formulating data selection as a submodular program, we obtain fast scalable selection algorithms with mathematical performance guarantees, resulting in a unified framework that clarifies existing approaches and also makes both new and many previous approaches easily accessible. We present a new class of submodular functions designed specifically for SMT and evaluate them on two different translation tasks. Our results show that our best submodular method significantly outperforms several baseline methods, including the widely-used cross-entropy based data selection method. In addition, our approach easily scales to large data sets and is applicable to other data selection problems in natural language processing.

Improve Statistical Machine Translation with Context-Sensitive Bilingual Semantic Embedding Model
Haiyang Wu, Daxiang Dong, Xiaoguang Hu, Dianhai Yu, Wei He, Hua Wu, Haifeng Wang, and Ting Liu

We investigate how to improve bilingual embedding which has been successfully used as a feature in phrase-based statistical machine translation (SMT). Despite bilingual embedding's success, the contextual information, which is of critical importance to translation quality, was ignored in previous work. To employ the contextual information, we propose a simple and memory-efficient model for learning bilingual embedding, taking both the source phrase and context around the phrase into account. Bilingual translation scores generated from our proposed bilingual embedding model are used as features in our SMT system. Experimental results show that the proposed method achieves significant improvements on large-scale Chinese-English translation task.

Transformation from Discontinuous to Continuous Word Alignment Improves Translation Quality
Zhongjun He, Hua Wu, Haifeng Wang, and Ting Liu

We present a novel approach to improve word alignment for statistical machine translation (SMT). Conventional word alignment methods allow discontinuous alignment, meaning that a source (or target) word links to several target (or source) words whose positions are discontinuous. However, we cannot extract phrase pairs from this kind of alignments as they break the alignment consistency constraint. In this paper, we use a weighted vote method to transform discontinuous word alignment to continuous alignment, which enables SMT systems extract more phrase pairs. We carry out experiments on large scale Chinese-to- English and German-to-English translation tasks. Experimental results show statistically significant improvements of BLEU score in both cases over the baseline systems. Our method produces a gain of +1.68 BLEU on NIST OpenMT04 for the phrase-based system, and a gain of +1.28 BLEU on NIST OpenMT06 for the hierarchical phrase-based system.

Unsupervised Word Alignment Using Frequency Constraint in Posterior Regularized EM
Hidetaka Kamigaito, Taro Watanabe, Hiroya Takamura, and Manabu Okumura

Generative word alignment models, such as IBM Models, are restricted to one-to-many alignment, and cannot explicitly represent many-to-many relationships in a bilingual text. The problem is partially solved either by introducing heuristics or by agreement constraints such that two directional word alignments agree with each other. In this paper, we focus on the posterior regularization framework (Ganchev et al., 2010) that can force two directional word alignment models to agree with each other during training, and propose new constraints that can take into account the difference between function words and content words. Experimental results on French-to-English and Japanese-to-English alignment tasks show statistically significant gains over the previous posterior regularization baseline. We also observed gains in Japanese-to-English translation tasks, which prove the effectiveness of our methods under grammatically different language pairs.

Asymmetric Features Of Human Generated Translation
Sauleh Eetemadi and Kristina Toutanova

Distinct properties of translated text have been the subject of research in linguistics for many year (Baker, 1993). In recent years computational methods have been developed to empirically verify the linguistic theories about translated text (Baroni and Bernardini, 2006). While many characteristics of translated text are more apparent in comparison to the original text, most of the prior research has focused on monolingual features of translated and original text.

The contribution of this work is introducing bilingual features that are capable of explaining differences in translation direction using localized linguistic phenomena at the phrase or sentence level, rather than using monolingual statistics at the document level. We show that these bilingual features outperform the monolingual features used in prior work (Kurokawa et al., 2009) for the task of classifying translation direction.

Syntax-Augmented Machine Translation using Syntax-Label Clustering
Hideya Mino, Taro Watanabe, and Eiichiro Sumita

Recently, syntactic information has helped significantly to improve statistical machine translation. However, the use of syntactic information may have a negative impact on the speed of translation because of the large number of rules, especially when syntax labels are projected from a parser in syntax-augmented machine translation. In this paper, we propose a syntax-label clustering method that uses an exchange algorithm in which syntax labels are clustered together to reduce the number of rules. The proposed method achieves clustering by directly maximizing the likelihood of synchronous rules, whereas previous work considered only the similarity of probabilistic distributions of labels. We tested the proposed method on Japanese-English and Chinese-English translation tasks and found order-of-magnitude higher clustering speeds for reducing labels and gains in translation quality compared with previous clustering method.

Testing for Significance of Increased Correlation with Human Judgment
Yvette Graham and Timothy Baldwin

Automatic metrics are widely used in machine translation as a substitute for human assessment. With the introduction of any new metric comes the question of just how well that metric mimics human assessment of translation quality. This is often measured by correlation with human judgment. Significance tests are generally not used to establish whether improvements over existing methods such as BLEU are statistically significant or have occurred by chance, however. In this paper, we introduce a significance test for comparing correlations of two metrics, along with an open-source implementation of the test. When applied to a range of metrics across seven language pairs, tests show for a high proportion of metrics insufficient evidence to conclude BLEU to be significantly outperformed.

Syntactic SMT Using a Discriminative Text Generation Model
Yue Zhang, Kai Song, Linfeng Song, Jingbo Zhu, and Qun Liu

We study a novel architecture for syntactic SMT. In contrast with the dominant approach in the literature, the system does not rely on translation rules, but treat translation as an unconstrained target sentence generation task, using soft features to capture lexical and syntactic correspondences between the source and target languages. Target syntax features and bilingual translation features are trained consistently in a discriminative model. Experiments using the IWSLT 2010 dataset show that the system achieves BLEU comparable to the state-of-the-art syntactic SMT systems.

Learning Hierarchical Translation Spans
Masao Utiyama, Eiichiro Sumita, and Hai Zhao

We propose a simple and effective approach to learn translation spans for the hierarchical phrase-based translation model. Our model evaluates if a source span should be covered by translation rules during decoding, which is integrated into the translation system as soft constraints. Compared to syntactic constraints, our model is directly acquired from an aligned parallel corpus and does not require parsers. Rich source side contextual features and advanced machine learning methods were utilized for this learning task. The proposed approach was evaluated on NTCIR-9 Chinese-English and Japanese-English translation tasks and showed significant improvement over the baseline system.

Neural Network Based Bilingual Language Model Growing for Statistical Machine Translation
Rui Wang, Hai Zhao, Bao-Liang Lu, Masao Utiyama, and Eiichiro Sumita

Since larger n-gram Language Model (LM) usually performs better in Statistical Machine Translation (SMT), how to construct efficient large LM is an important topic in SMT. However, most of the existing LM growing methods need an extra monolingual corpus, where additional LM adaption technology is necessary. In this paper, we propose a novel neural network based bilingual LM growing method, only using the bilingual parallel corpus in SMT. The results show that our method can improve both the perplexity score for LM evaluation and BLEU score for SMT, and significantly outperforms the existing LM growing methods without extra corpus.

Better Statistical Machine Translation through Linguistic Treatment of Phrasal Verbs
Kostadin Cholakov and Valia Kordoni

This article describes a linguistically informed method for integrating phrasal verbs into statistical machine translation (SMT) systems. In a case study involving English to Bulgarian SMT, we show that our method does not only improve

translation quality but also outperforms similar methods previously applied to the same task. We attribute this to the fact that, in contrast to previous work on the subject, we employ detailed linguistic information. We found out that features which describe phrasal verbs as idiomatic or compositional contribute most to the better translation quality achieved by our method.

Fitting Sentence Level Translation Evaluation with Many Dense Features
Miloš Stanojević and Sima'an Khalil

Sentence level evaluation in MT has turned out far more difficult than corpus level evaluation. Existing sentence level metrics employ a limited set of features, most of which are rather sparse at the sentence level, and their intricate models are rarely trained for ranking. This paper presents a simple linear model exploiting 33 relatively dense features, some of which are novel while others are known but seldom used, and train it under the learning-to-rank framework. We evaluate our metric on the standard WMT12 data showing that it outperforms the strong baseline METEOR. We also analyze the contribution of individual features and the choice of training data, language-pair vs. target-language data, providing new insights into this task.

A Human Judgement Corpus and a Metric for Arabic MT Evaluation
Houda Bouamor, Hanan Alshikhabobakr, Behrang Mohit, and Kemal Oflazer

We present a human judgements dataset and also an adapted metric for evaluation of Arabic machine translation. Our medium-scale dataset is first of its kind for Arabic with high annotation quality. We use the dataset to adapt the BLEU score for Arabic. Our score (AL-BLEU) provides partial credits for stem and morphological matchings of hypothesis and reference words. We evaluate BLEU, METEOR and AL-BLEU on our human judgements corpus and show that AL-BLEU has the highest correlation with human judgements. We plan to release both the dataset and the software to the research community.

Learning to Differentiate Better from Worse Translations
Francisco Guzmán, Shafiq Joty, Lluís Màrquez, Alessandro Moschitti, Preslav Nakov, and Massimo Nicosia

We present a pairwise learning-to-rank approach to machine translation evaluation that learns to differentiate better from worse translations in the context of a given reference. We integrate several layers of linguistic information encapsulated in tree-based structures, making use of both the reference and the system output simultaneously, thus bringing our ranking closer to how humans evaluate translations. Most importantly, instead of deciding upfront which types of features %between hypotheses and references are important, we use the learning framework of preference re-ranking kernels to learn the features automatically. % Thus, instead of learning the importance of a predetermined % set of features, we learn the features themselves. The evaluation results show that learning in the proposed framework yields better correlation with humans than computing the direct similarity over the same type of structures. Also, we show our structural kernel learning (SKL) can be a general framework for MT evaluation, in which syntactic and semantic information can be naturally incorporated.

Two Improvements to Left-to-Right Decoding for Hierarchical Phrase-based Machine Translation
Maryam Siahbani and Anoop Sarkar

Left-to-right (LR) decoding (Watanabe et al., 2006) is promising decoding algorithm for hierarchical phrase-based translation (Hiero) that visits input spans in arbitrary order producing the output translation in left to right order. This leads to far fewer language model calls, but while LR decoding is more efficient than CKY decoding, it is unable to capture some hierarchical phrase alignments reachable using CKY decoding and suffers from lower translation quality as a result. This paper introduces two improvements to LR decoding that make it comparable in translation quality to CKY-based Hiero.

Reordering Model for Forest-to-String Machine Translation
Martin Cmejrek

In this paper, we present a novel extension of a forest-to-string machine translation system with a reordering model. We predict reordering probabilities for every pair of source words with a model using features observed from the input parse forest. Our approach naturally deals with the ambiguity present in the input parse forest, but, at the same time, takes into account only the parts of the input forest used by the current translation hypothesis. The method provides improvement from 0.6 up to 1.0 point measured by (Ter - Bleu)/2 metric.

Parallel Session 2 Overview – Sunday, October 26, 2014

Track A *Computational Psycholinguistics* Al Areen 1-2	**Track B** *Tagging, Chunking, Parsing and Syntax* Al Areen 3	**Track C** *NLP for the Web and Social Media* Al Areen 4	
Aligning context-based statistical models of language with brain activity during reading *Leila Wehbe, Ashish Vaswani, Kevin Knight, and Tom Mitchell*	Go Climb a Dependency Tree and Correct the Grammatical Errors *Longkai Zhang and Houfeng Wang*	A Graph-based Approach for Contextual Text Normalization *Cagil Sonmez and Arzucan Özgür*	13:30
A Cognitive Model of Semantic Network Learning *Aida Nematzadeh, Afsaneh Fazly, and Suzanne Stevenson*	An Unsupervised Model for Instance Level Subcategorization Acquisition *Simon Baker, Roi Reichart, and Anna Korhonen*	[TACL] Entity Linking on Microblogs with Spatial and Temporal Signals *Yuan Fang and Ming-Wei Chang*	13:55
[TACL] The Benefits of a Model of Annotation *Rebecca Passonneau and Bob Carpenter*	Parsing low-resource languages using Gibbs sampling for PCFGs with latent annotations *Liang Sun, Jason Mielens, and Jason Baldridge*	ReNoun: Fact Extraction for Nominal Attributes *Mohamed Yahya, Steven Whang, Rahul Gupta, and Alon Halevy*	14:20
Learning Abstract Concept Embeddings from Multi-Modal Data: Since You Probably Can't See What I Mean *Felix Hill and Anna Korhonen*	Incremental Semantic Role Labeling with Tree Adjoining Grammar *Ioannis Konstas, Frank Keller, Vera Demberg, and Mirella Lapata*	Hierarchical Discriminative Classification for Text-Based Geolocation *Benjamin Wing and Jason Baldridge*	14:45

Track P

Posters: Semantics

Al Areen 5-6   13:30-15:10

- Probabilistic Models of Cross-Lingual Semantic Similarity in Context Based on Latent Cross-Lingual Concepts Induced from Comparable Data
  Ivan Vulić and Marie-Francine Moens

- Multi-Predicate Semantic Role Labeling
  Haitong Yang and Chengqing Zong

- Werdy: Recognition and Disambiguation of Verbs and Verb Phrases with Syntactic and Semantic Pruning
  Luciano Del Corro, Rainer Gemulla, and Gerhard Weikum

- Multi-Resolution Language Grounding with Weak Supervision
  R. Koncel-Kedziorski, Hannaneh Hajishirzi, and Ali Farhadi

- Incorporating Vector Space Similarity in Random Walk Inference over Knowledge Bases
  Matt Gardner, Partha Talukdar, Jayant Krishnamurthy, and Tom Mitchell

- Composition of Word Representations Improves Semantic Role Labelling
  Michael Roth and Kristian Woodsend

- Automatic Domain Assignment for Word Sense Alignment
  Tommaso Caselli and Carlo Strapparava

- Nothing like Good Old Frequency: Studying Context Filters for Distributional Thesauri
  Muntsa Padró, Marco Idiart, Aline Villavicencio, and Carlos Ramisch

- Aligning English Strings with Abstract Meaning Representation Graphs
  Nima Pourdamghani, Yang Gao, Ulf Hermjakob, and Kevin Knight

- A Shortest-path Method for Arc-factored Semantic Role Labeling
  Xavier Lluís, Xavier Carreras, and Lluís Màrquez

- Semantic Kernels for Semantic Parsing
  Iman Saleh, Alessandro Moschitti, Preslav Nakov, Lluís Màrquez, and Shafiq Joty

- [TACL] Multi-Modal Models for Concrete and Abstract Concept Meaning
  Felix Hill, Roi Reichart, and Anna Korhonen

- An I-vector Based Approach to Compact Multi-Granularity Topic Spaces Representation of Textual Documents
  Mohamed Morchid, Mohamed Bouallegue, Richard Dufour, Georges Linares, Driss Matrouf, and Renato de Mori

Parallel Session 2

Session 2a: Computational Psycholinguistics
Al Areen 1-2   *Chair: Massimo Poesio*

Aligning context-based statistical models of language with brain activity during reading
Leila Wehbe, Ashish Vaswani, Kevin Knight, and Tom Mitchell   13:30–13:55

Context-based neural networks (1) model the previously seen context as a latent feature vector, (2) integrate successive words into the context using some learned representation (embedding), and (3) compute output probabilities for incoming words given the context. Brain imaging studies have suggested that during reading, the brain (a) continuously builds a context from the successive words and every time it encounters a word it (b) fetches its properties from memory and (c) integrates it with the previous context with a degree of effort that is inversely proportional to how probable the word is. This hints to a parallelism between the neural networks and the brain in modeling context (1 and a), representing the incoming words (2 and b) and integrating it (3 and c). We study the alignment between the latent vectors used by neural networks and brain activity observed via Magnetoencephalography (MEG) when subjects read a story. We apply the neural networks to the same text the subjects read, and explore the ability of these three vector representations to predict the observed word-by-word brain activity. Our novel results show that: before a new word i is read, brain activity is well predicted by the neural network representation of context and the predictability decreases as the brain integrates the word and changes its own representation of context. Secondly, the neural network embedding of word i can predict the MEG activity when word i is presented to the subject, revealing that it is correlated with the brain's representation of word i. Moreover, the activity is predicted in different regions of the brain with delays that are consistent with the placement of each region on the processing pathway. Finally, we show that the output probability computed by the neural networks agrees with the brain's own assessment of the probability of word i.

A Cognitive Model of Semantic Network Learning
Aida Nematzadeh, Afsaneh Fazly, and Suzanne Stevenson   13:55–14:20

Child semantic development includes learning the meaning of words as well as the semantic relations among words. A presumed outcome of semantic development is the formation of a semantic network that reflects this knowledge. We present an algorithm for simultaneously learning word meanings and gradually growing a semantic network, which adheres to the cognitive plausibility requirements of incrementality and limited computations. We demonstrate that the semantic connections among words in addition to their context is necessary in forming a semantic network that resembles an adult's semantic knowledge.

[TACL] The Benefits of a Model of Annotation
Rebecca Passonneau and Bob Carpenter   14:20–14:45

The quality of annotated data for research in Natural Language Processing is generally assumed to be good enough if a few annotators can be shown to be consistent with one another. However, standard agreement measures for interannotator reliability are neither necessary nor sufficient to ensure a high quality corpus. Probabilistic models of agreement have been used in psychometrics and marketing since the 1950s and in epidemiology since the 1970s. More recently, crowdsourcing has motivated their application to annotated language data for machine learning. Here, a case study of word sense annotation contrasts conventional methods for evaluating labels from trained annotators with a probabilistic annotation model applied to crowdsourced data. The annotation model provides far more information about the data, and in particular, each gold standard label has a confidence measure. Knowledge about the certainty of each label can be helpful in many ways: for example, to support more informed evaluation of machine learned models, or better feature engineering for learning. Further, the crowdsourced data was collected at less than half the cost of the conventional approach.

Learning Abstract Concept Embeddings from Multi-Modal Data: Since You Probably Can't See What I Mean
Felix Hill and Anna Korhonen   14:45–15:10

Models that acquire semantic representations from both linguistic and perceptual input outperform linguistic-only models on various NLP tasks. However, this superiority has only been established when learning concrete concepts, which are usually domain specific and also comparatively rare in everyday language. We extend the scope to more widely applicable abstract representations, and present a multi-modal probabilistic language architecture for learning semantic representations for both concrete and abstract concepts. Our model outperforms alternative approaches in

combining input from two modalities and also in propagating perceptual information on concrete to more abstract concepts. We discuss the implications of our results both for optimizing the performance of multi-modal models and for theories of abstract conceptual representation.

Session 2b: Tagging, Chunking, Parsing and Syntax
Al Areen 3   *Chair: Slav Petrov*

Go Climb a Dependency Tree and Correct the Grammatical Errors
Longkai Zhang and Houfeng Wang   13:30–13:55

State-of-art systems for grammar error correction often correct errors based on word sequences or phrases. In this paper we describe a grammar error correction system which corrects grammatical errors at tree level directly. We cluster all error into two groups and divide our system into two modules correspondingly: the general module and the special module. In the general module, we propose a TreeNode Language Model to correct errors related to verbs and nouns. The TreeNode Language Model is easy to train and the decoding is efficient. In the special module, two extra classification models are trained to correct errors related to determiners and prepositions. Experiments show that our system outperforms the state-of-art systems and improves the F1 score.

An Unsupervised Model for Instance Level Subcategorization Acquisition
Simon Baker, Roi Reichart, and Anna Korhonen   13:55–14:20

Most existing systems for subcategorization frame (SCF) acquisition rely on supervised parsing and infer SCF distributions at type, rather than instance level. These systems suffer from poor portability across domains and their benefit for NLP tasks that involve sentence-level processing is limited. We propose a new unsupervised, Markov Random Field-based model for SCF acquisition which is designed to address these problems. The system relies on supervised POS tagging rather than parsing, and is capable of learning SCFs at instance level. We perform evaluation against gold standard data which shows that our system outperforms several supervised and type-level SCF baselines. We also conduct task-based evaluation in the context of verb similarity prediction, demonstrating that a vector space model based on our SCFs substantially outperforms a lexical model and a model based on a supervised parser.

Parsing low-resource languages using Gibbs sampling for PCFGs with latent annotations
Liang Sun, Jason Mielens, and Jason Baldridge   14:20–14:45

PCFGs with latent annotations have been shown to be a very effective model for phrase structure parsing. We present a Bayesian model and algorithms based on a Gibbs sampler for parsing with a grammar with latent annotations. For PCFG-LA, we present an additional Gibbs sampler algorithm to learn annotations from training data, which are parse trees with coarse (unannotated) symbols. We show that a Gibbs sampling technique is capable of parsing sentences in a wide variety of languages and producing results that are on-par with or surpass previous approaches. Our results for Kinyarwanda and Malagasy in particular demonstrate that low-resource language parsing can benefit substantially from a Bayesian approach.

Incremental Semantic Role Labeling with Tree Adjoining Grammar
Ioannis Konstas, Frank Keller, Vera Demberg, and Mirella Lapata   14:45–15:10

We introduce the task of incremental semantic role labeling (iSRL), in which semantic roles are assigned to incomplete input (sentence prefixes). iSRL is the semantic equivalent of incremental parsing, and is useful for language modeling, sentence completion, machine translation, and psycholinguistic modeling. We propose an iSRL system that combines an incremental TAG parser with a semantically enriched lexicon, a role propagation algorithm, and a cascade of classifiers. Our approach achieves an SRL F-score of 78.38% on the standard CoNLL 2009 dataset. It substantially outperforms a strong baseline that combines gold-standard syntactic dependencies with heuristic role assignment, as well as a baseline based on Nivre's incremental dependency parser.

Session 2c: NLP for the Web and Social Media
Al Areen 4   *Chair: Timothy Baldwin*

A Graph-based Approach for Contextual Text Normalization
Cagil Sonmez and Arzucan Özgür   13:30–13:55

The informal nature of social media text renders it very difficult to be automatically processed by natural language processing tools. Text normalization, which corresponds to restoring the noisy words to their canonical forms, provides a solution to this challenge. We introduce an unsupervised text normalization approach that utilizes not only lexical, but also contextual and grammatical features of social text. The contextual and grammatical features are extracted from a word association graph built by using a large unlabeled social media text corpus. The graph encodes the relative positions of the words with respect to each other, as well as their part-of-speech tags. The lexical features are obtained by using the longest common subsequence ratio and edit distance measures to encode the surface similarity among words, and the double metaphone algorithm to represent the phonetic similarity. Unlike most of the recent approaches that are based on generating normalization dictionaries, the proposed approach performs normalization by considering the context of the noisy words in the input text. Our results show that it achieves state-of-the-art F-score performance on standard data sets. In addition, the system can be tuned to achieve very high precision without sacrificing much from recall.

[TACL] Entity Linking on Microblogs with Spatial and Temporal Signals
Yuan Fang and Ming-Wei Chang   13:55–14:20

Microblogs present an excellent opportunity for monitoring and analyzing world happenings. Given that words are often ambiguous, entity linking becomes a crucial step towards understanding microblogs. In this paper, we re-examine the problem of entity linking on microblogs. We first observe that spatiotemporal (i.e., spatial and temporal) signals play a key role, but they are not utilized in existing approaches. Thus, we propose a novel entity linking framework that incorporates spatiotemporal signals through a weakly supervised process. Using entity annotations1 on real-world data, our experiments show that the spatiotemporal model improves F1 by more than 10 points over existing systems. Finally, we present a qualitative study to visualize the effectiveness of our approach.

ReNoun: Fact Extraction for Nominal Attributes
Mohamed Yahya, Steven Whang, Rahul Gupta, and Alon Halevy   14:20–14:45

Search engines are increasingly relying on large knowledge bases of facts to provide direct answers to users' queries. However, the construction of these knowledge bases is largely manual and does not scale to the long and heavy tail of facts. Open information extraction tries to address this challenge, but typically assumes that facts are expressed with verb phrases, and therefore has had difficulty extracting facts for noun-based relations. We describe ReNoun, an open information extraction system that complements previous efforts by focusing on nominal attributes and on the long tail. ReNoun's approach is based on leveraging a large ontology of noun attributes mined from a text corpus and from user queries. ReNoun creates a seed set of training data by using specialized patterns and requiring that the facts mention an attribute in the ontology. ReNoun then generalizes from this seed set to produce a much larger set of extractions that are then scored. We describe experiments that show that we extract facts with high precision and for attributes that cannot be extracted with verb-based techniques.

Hierarchical Discriminative Classification for Text-Based Geolocation
Benjamin Wing and Jason Baldridge   14:45–15:10

Text-based document geolocation is commonly rooted in language-based information retrieval techniques over geodesic grids. These methods ignore the natural hierarchy of cells in such grids and fall afoul of independence assumptions. We demonstrate the effectiveness of using logistic regression models on a hierarchy of nodes in the grid, which improves upon the state of the art accuracy by several percent and reduces mean error distances by hundreds of kilometers on data from Twitter, Wikipedia, and Flickr. We also show that logistic regression performs feature selection effectively, assigning high weights to geocentric terms.

Session 2-posters: Semantics
Al Areen 5-6   13:30–15:10

Probabilistic Models of Cross-Lingual Semantic Similarity in Context Based on Latent Cross-Lingual Concepts Induced from Comparable Data
Ivan Vulić and Marie-Francine Moens

We propose the first probabilistic approach to modeling cross-lingual semantic similarity (CLSS) in context which requires only comparable data. The approach relies on an idea of projecting words and sets of words into a shared latent semantic space spanned by language-pair independent latent semantic concepts (e.g., cross-lingual topics obtained by a multilingual topic model). These latent cross-lingual concepts are induced from a comparable corpus without any additional lexical resources. Word meaning is represented as a probability distribution over the latent concepts, and a change in meaning is represented as a change in the distribution over these latent concepts. We present new models that modulate the isolated out-of-context word representations with contextual knowledge. Results on the task of suggesting word translations in context for 3 language pairs reveal the utility of the proposed contextualized models of cross-lingual semantic similarity.

Multi-Predicate Semantic Role Labeling
Haitong Yang and Chengqing Zong

The current approaches to Semantic Role Labeling (SRL) usually perform role classification for each predicate separately and the interaction among individual predicate's role labeling is ignored if there is more than one predicate in a sentence. In this paper, we prove that different predicates in a sentence could help each other during SRL. In multi-predicate role labeling, there are mainly two key points: argument identification and role labeling of the arguments shared by multiple predicates. To address these issues, in the stage of argument identification, we propose novel predicate-related features which help remove many argument identification errors; in the stage of argument classification, we adopt a discriminative reranking approach to perform role classification of the shared arguments, in which a large set of global features are proposed. We conducted experiments on two standard benchmarks: Chinese PropBank and English PropBank. The experimental results show that our approach can significantly improve SRL performance, especially in Chinese PropBank.

Werdy: Recognition and Disambiguation of Verbs and Verb Phrases with Syntactic and Semantic Pruning
Luciano Del Corro, Rainer Gemulla, and Gerhard Weikum

Word-sense recognition and disambiguation (WERD) is the task of identifying word phrases and their senses in natural language text. Though it is well understood how to disambiguate noun phrases, this task is much less studied for verbs and verbal phrases. We present Werdy, a framework for WERD with particular focus on verbs and verbal phrases. Our framework first identifies multi-word expressions based on the syntactic structure of the sentence; this allows us to recognize both contiguous and non-contiguous phrases. We then generate a list of candidate senses for each word or phrase, using novel syntactic and semantic pruning techniques. We also construct and leverage a new resource of pairs of senses for verbs and their object arguments. Finally, we feed the so-obtained candidate senses into standard word-sense disambiguation (WSD) methods, and boost their precision and recall. Our experiments indicate that Werdy significantly increases the performance of existing WSD methods.

Multi-Resolution Language Grounding with Weak Supervision
R. Koncel-Kedziorski, Hannaneh Hajishirzi, and Ali Farhadi

Language is given meaning through its correspondence with a world representa- tion. This correspondence can be at mul- tiple levels of granularity or resolutions. In this paper, we introduce an approach to multi-resolution language grounding in the extremely challenging domain of pro- fessional soccer commentaries. We define and optimize a factored objective function that allows us to leverage discourse struc- ture and the compositional nature of both language and game events. We show that finer resolution grounding helps coarser resolution grounding, and vice versa. Our method results in an F1 improvement of more than 48% versus the previous state of the art for fine-resolution grounding.

Incorporating Vector Space Similarity in Random Walk Inference over Knowledge Bases
Matt Gardner, Partha Talukdar, Jayant Krishnamurthy, and Tom Mitchell

Much work in recent years has gone into the construction of large knowledge bases (KBs), such as Freebase, DBPedia, NELL, and YAGO. While these KBs are very large, they are still very incomplete, necessitating the use of inference to fill in gaps. Prior work has shown how to make use of a large text corpus to augment random walk inference over

KBs. We present two improvements to the use of such large corpora to augment KB inference. First, we present a new technique for combining KB relations and surface text into a single graph representation that is much more compact than graphs used in prior work. Second, we describe how to incorporate vector space similarity into random walk inference over KBs, reducing the feature sparsity inherent in using surface text. This allows us to combine distributional similarity with symbolic logical inference in novel and effective ways. With experiments on many relations from two separate KBs, we show that our methods significantly outperform prior work on KB inference.

Composition of Word Representations Improves Semantic Role Labelling
Michael Roth and Kristian Woodsend
State-of-the-art semantic role labelling systems require large annotated corpora to achieve full performance. Unfortunately, such corpora are expensive to produce and often do not generalize well across domains. Even in domain, errors are often made where syntactic information does not provide sufficient cues. In this paper, we mitigate both of these problems by employing distributional word representations gathered from unlabelled data. While straight-forward word representations of predicates and arguments improve performance, we show that further gains are achieved by composing representations that model the interaction between predicate and argument, and capture full argument spans.

Automatic Domain Assignment for Word Sense Alignment
Tommaso Caselli and Carlo Strapparava
This paper reports on the development of a hybrid and simple method based on a machine learning classifier (Naive Bayes), Word Sense Disambiguation and rules, for the automatic assignment of WordNet Domains to nominal entries of a lexicographic dictionary, the Senso Comune De Mauro Lexicon. The system obtained an F1 score of 0.58, with a Precision of 0.70. We further used the automatically assigned domains to filter out word sense alignments between MultiWordNet and Senso Comune. This has led to an improvement in the quality of the sense alignments showing the validity of the approach for domain assignment and the importance of domain information for achieving good sense alignments.

Nothing like Good Old Frequency: Studying Context Filters for Distributional Thesauri
Muntsa Padró, Marco Idiart, Aline Villavicencio, and Carlos Ramisch
Much attention has been given to the impact of informativeness and similarity measures on distributional thesauri. We investigate the effects of context filters on thesaurus quality and propose the use of cooccurrence frequency as a simple and inexpensive criterion. For evaluation, we measure thesaurus agreement with WordNet and performance in answering TOEFL-like questions. Results illustrate the sensitivity of distributional thesauri to filters.

Aligning English Strings with Abstract Meaning Representation Graphs
Nima Pourdamghani, Yang Gao, Ulf Hermjakob, and Kevin Knight
We align pairs of English sentences and corresponding Abstract Meaning Representations (AMR), at the token level. Such alignments will be useful for downstream extraction of semantic interpretation and generation rules. Our method involves linearizing AMR structures and performing symmetrized EM training. We obtain 86.5% and 83.1% alignment F score on development and test sets.

A Shortest-path Method for Arc-factored Semantic Role Labeling
Xavier Lluís, Xavier Carreras, and Lluís Màrquez
We introduce a Semantic Role Labeling (SRL) parser that finds semantic roles for a predicate together with the syntactic paths linking predicates and arguments. Our main contribution is to formulate SRL in terms of shortest-path inference, on the assumption that the SRL model is restricted to arc-factored features of the syntactic paths behind semantic roles. Overall, our method for SRL is a novel way to exploit larger variability in the syntactic realizations of predicate-argument relations, moving away from pipeline architectures. Experiments show that our approach improves the robustness of the predictions, producing arc-factored models that perform closely to methods using unrestricted features from the syntax.

Semantic Kernels for Semantic Parsing
Iman Saleh, Alessandro Moschitti, Preslav Nakov, Lluís Màrquez, and Shafiq Joty
We present an empirical study on the use of semantic information for Concept Segmentation and Labeling (CSL), which is an important step for semantic parsing. We represent the hypotheses output by a state-of-the-art CSL parser with tree structures, which we rerank with a classifier trained on two types of semantic tree kernels: one processing structures built with words, concepts and Brown clusters, and another one using semantic similarity between the words

composing the structure. The results on a corpus related to the restaurant domain show that our semantic kernels exploiting similarity measure outperform state-of-the-art rerankers.

[TACL] Multi-Modal Models for Concrete and Abstract Concept Meaning
Felix Hill, Roi Reichart, and Anna Korhonen

Multi-modal models that learn semantic representations from both linguistic and perceptual input outperform language-only models on a range of evaluations, and better reflect human concept acquisition. Most perceptual input to such models corresponds to concrete noun concepts and the superiority of the multi-modal approach has only been established when evaluating on such concepts. We therefore investigate which concepts can be effectively learned by multi-modal models. We show that concreteness determines both which linguistic features are most informative and the impact of perceptual input in such models. We then introduce ridge regression as a means of propagating perceptual information from concrete nouns to more abstract concepts that is more robust than previous approaches. Finally, we present weighted gram matrix combination, a means of combining representations from distinct modalities that outperforms alternatives when both modalities are sufficiently rich.

An I-vector Based Approach to Compact Multi-Granularity Topic Spaces Representation of Textual Documents
Mohamed Morchid, Mohamed Bouallegue, Richard Dufour, Georges Linares, Driss Matrouf, and Renato de Mori

Various studies highlighted that topic-based approaches allow a powerful spoken content representation of documents. Nonetheless, these documents may content more that one main theme, and their automatic transcription inevitably contains errors. In this study, we propose an original and promising framework based on a compact representation of a textual document, to solve issues related to topic space granularity. Firstly, various topic spaces are estimated with different number of classes from a Latent Dirichlet Allocation. Then, this multiple topic space representation is compacted into an elementary segment, named c-vector, originally developed in the context of speaker recognition. Experiments are conducted on the DECODA corpus of conversations. Results show the effectiveness of the proposed multi-view compact representation paradigm, our identification system reaching an accuracy of 85%, with a significant gain of 9 points compared to the baseline (best single topic space configuration).

Parallel Session 3 Overview – Sunday, October 26, 2014

	Track A	Track B	Track C
	Sentiment Analysis and Opinion Mining	*Semantics*	*Machine Translation*
	Al Areen 1-2	Al Areen 3	Al Areen 4
15:40	Explaining the Stars: Weighted Multiple-Instance Learning for Aspect-Based Sentiment Analysis *Nikolaos Pappas and Andrei Popescu-Belis*	Resolving Shell Nouns *Varada Kolhatkar and Graeme Hirst*	Modeling Term Translation for Document-informed Machine Translation *Fandong Meng, Deyi Xiong, Wenbin Jiang, and Qun Liu*
16:05	Sentiment Analysis on the People's Daily *Jiwei Li and Eduard Hovy*	A Comparison of Selectional Preference Models for Automatic Verb Classification *Will Roberts and Markus Egg*	Beyond Parallel Data: Joint Word Alignment and Decipherment Improves Machine Translation *Qing Dou, Ashish Vaswani, and Kevin Knight*
16:30	A Joint Segmentation and Classification Framework for Sentiment Analysis *Duyu Tang, Furu Wei, Bing Qin, Li Dong, Ting Liu, and Ming Zhou*	Learning to Solve Arithmetic Word Problems with Verb Categorization *Mohammad Javad Hosseini, Hannaneh Hajishirzi, Oren Etzioni, and Nate Kushman*	Latent Domain Phrase-based Models for Adaptation *Cuong Hoang and Khalil Sima'an*
16:55	Positive Unlabeled Learning for Deceptive Reviews Detection	NaturalLI: Natural Logic Inference for Common Sense Reasoning *Gabor Angeli and Christopher Manning*	Translation Rules with Right-Hand Side Lattices *Fabien Cromieres and Sadao Kurohashi*

Track P

Posters: Information Retrieval, Summarization and Question Answering

Al Areen 5-6   15:40-17:20

- Learning to Translate: A Query-Specific Combination Approach for Cross-Lingual Information Retrieval
  Ferhan Ture and Elizabeth Boschee

- Semantic-Based Multilingual Document Clustering via Tensor Modeling
  Salvatore Romeo, Andrea Tagarelli, and Dino Ienco

- Lexical Substitution for the Medical Domain
  Martin Riedl, Michael Glass, and Alfio Gliozzo

- Question Answering with Subgraph Embeddings
  Antoine Bordes, Sumit Chopra, and Jason Weston

- Correcting Keyboard Layout Errors and Homoglyphs in Queries
  Derek Barnes, Mahesh Joshi, and Hassan Sawaf

- Non-linear Mapping for Improved Identification of 1300+ Languages
    Ralf Brown

- A Neural Network for Factoid Question Answering over Paragraphs
    Mohit Iyyer, Jordan Boyd-Graber, Leonardo Claudino, Richard Socher, and Hal Daumé III

- Joint Relational Embeddings for Knowledge-based Question Answering
    Min-Chul Yang, Nan Duan, Ming Zhou, and Hae-Chang Rim

- Adding High-Precision Links to Wikipedia
    Thanapon Noraset, Chandra Bhagavatula, and Doug Downey

- [TACL] Crosslingual and Multilingual Construction of Syntax-Based Vector Space Models
    Jason Utt and Sebastian Padó

- Finding Good Enough: A Task-Based Evaluation of Query Biased Summarization for Cross-Language Information Retrieval
    Jennifer Williams, Sharon Tam, and Wade Shen

- Chinese Poetry Generation with Recurrent Neural Networks
    Xingxing Zhang and Mirella Lapata

- Fear the REAPER: A System for Automatic Multi-Document Summarization with Reinforcement Learning
    Cody Rioux, Sadid A. Hasan, and Yllias Chali

- Improving Multi-documents Summarization by Sentence Compression based on Expanded Constituent Parse Trees
    Chen Li, Yang Liu, Fei Liu, Lin Zhao, and Fuliang Weng

- Analyzing Stemming Approaches for Turkish Multi-Document Summarization
    Muhammed Yavuz Nuzumlalı and Arzucan Özgür

Parallel Session 3

Session 3a: Sentiment Analysis and Opinion Mining
Al Areen 1-2 *Chair: Preslav Nakov*

Explaining the Stars: Weighted Multiple-Instance Learning for Aspect-Based Sentiment Analysis
Nikolaos Pappas and Andrei Popescu-Belis 15:40–16:05

This paper introduces a model of multiple-instance learning applied to the prediction of aspect ratings or judgments of specific properties of an item from user-contributed texts such as product reviews. Each variable-length text is represented by several independent feature vectors; one word vector per sentence or paragraph. For learning from texts with known aspect ratings, the model performs multiple-instance regression (MIR) and assigns importance weights to each of the sentences or paragraphs of a text, uncovering their contribution to the aspect ratings. Next, the model is used to predict aspect ratings in previously unseen texts, demonstrating interpretability and explanatory power for its predictions. We evaluate the model on seven multi-aspect sentiment analysis data sets, improving over four MIR baselines and two strong bag-of-words linear models, namely SVR and Lasso, by more than 10

Sentiment Analysis on the People's Daily
Jiwei Li and Eduard Hovy 16:05–16:30

We propose a semi-supervised bootstrapping algorithm for analyzing China's foreign relations from the People's Daily. Our approach addresses sentiment target clustering, subjective lexicons extraction and sentiment prediction in a unified framework. Different from existing algorithms in the literature, time information is considered in our algorithm through a hierarchical bayesian approach to guide the bootstrapping approach. We are hopeful that our approach can facilitate quantitative political analysis conducted by social scientists and politicians.

A Joint Segmentation and Classification Framework for Sentiment Analysis
Duyu Tang, Furu Wei, Bing Qin, Li Dong, Ting Liu, and Ming Zhou 16:30–16:55

In this paper, we propose a joint segmentation and classification framework for sentiment analysis. Existing sentiment classification algorithms typically split a sentence as a word sequence, which does not effectively handle the inconsistent sentiment polarity between a phrase and the words it contains, such as "not bad" and "a great deal of". We address this issue by developing a joint segmentation and classification framework (JSC), which simultaneously conducts sentence segmentation and sentence-level sentiment classification. Specifically, we use a log-linear model to score each segmentation candidate, and exploit the phrasal information of top-ranked segmentations as features to build the sentiment classifier. A marginal log-likelihood objective function is devised for the segmentation model, which is optimized for enhancing the sentiment classification performance. The joint model is trained only based on the annotated sentiment polarity of sentences, without any segmentation annotations. Experiments on a benchmark Twitter sentiment classification dataset in SemEval 2013 show that, our joint model performs comparably with the state-of-the-art methods.

Positive Unlabeled Learning for Deceptive Reviews Detection
16:55–17:20

Deceptive reviews detection has attracted significant attention from both business and research communities. However, due to the difficulty of human labeling needed for supervised learning, the problem remains to be highly challenging. This paper proposed a novel angle to the problem by modeling PU (positive unlabeled) learning. A semi-supervised model, called mixing population and individual property PU learning (MPIPUL), is proposed. Firstly, some reliable negative examples are identified from the unlabeled dataset. Secondly, some representative positive examples and negative examples are generated based on LDA (Latent Dirichlet Allocation). Thirdly, for the remaining unlabeled examples (we call them spy examples), which can not be explicitly identified as positive and negative, two similarity weights are assigned, by which the probability of a spy example belonging to the positive class and the negative class are displayed. Finally, spy examples and their similarity weights are incorporated into SVM (Support Vector Machine) to build an accurate classifier. Experiments on gold-standard dataset demonstrate the effectiveness of MPIPUL which outperforms the state-of-the-art baselines.

Session 3b: Semantics
Al Areen 3 *Chair: Kristina Toutanova*

Resolving Shell Nouns
Varada Kolhatkar and Graeme Hirst 15:40–16:05

'Shell nouns', such as 'fact' and 'problem', occur frequently in all kinds of texts. These nouns themselves are unspecific, and can only be interpreted together with the 'shell content'. We propose a general approach to automatically identify shell content of shell nouns. Our approach exploits lexico-syntactic knowledge derived from the linguistics literature. We evaluate the approach on a variety of shell nouns with a variety of syntactic expectations, achieving accuracies in the range of 62% (baseline = 33%) to 83% (baseline = 74%) on crowd-annotated data.

A Comparison of Selectional Preference Models for Automatic Verb Classification
Will Roberts and Markus Egg 16:05–16:30

We present a comparison of different selectional preference models and evaluate them on an automatic verb classification task in German. We find that all the models we compare are effective for verb clustering; the best-performing model uses syntactic information to induce nouns classes from unlabelled data in an unsupervised manner. A very simple model based on lexical preferences is also found to perform well.

Learning to Solve Arithmetic Word Problems with Verb Categorization
Mohammad Javad Hosseini, Hannaneh Hajishirzi, Oren Etzioni, and Nate Kushman 16:30–16:55

This paper presents a novel approach to learning to solve simple arithmetic word problems. Our system, ARIS, analyzes each of the sentences in the problem statement to identify the relevant variables and their values. ARIS then maps this information into an equation that represents the problem, and enables its (trivial) solution as shown in Figure 1. The paper analyzes the arithmetic-word problems "genre", identifying seven categories of verbs used in such problems. ARIS learns to categorize verbs with 81.2% accuracy, and is able to solve 77.7% of the problems in a corpus of standard primary school test questions. We report the first learning results on this task without reliance on pre-defined templates and make our data publicly available.

NaturalLI: Natural Logic Inference for Common Sense Reasoning
Gabor Angeli and Christopher Manning 16:55–17:20

Common-sense reasoning is important for AI applications, both in NLP and many vision and robotics tasks. We propose NaturalLI: a Natural Logic inference system for inferring common sense facts – for instance, that cats have tails or tomatoes are fragile – from a very large database of known facts. In addition being able to provide strictly valid derivations, the system is also able to produce derivations which are only likely valid, accompanied by an associated confidence. We show that our system is able to capture strict Natural Logic inferences on the FraCaS test suite, and demonstrate its ability to infer previously unseen facts with 48% recall and 93% precision. Common-sense reasoning is important for AI applications, both in NLP and many vision and robotics tasks. We propose NaturalLI: a Natural Logic inference system for inferring common sense facts – for instance, that cats have tails or tomatoes are round – from a very large database of known facts. In addition to being able to provide strictly valid derivations, the system is also able to produce derivations which are only likely valid, accompanied by an associated confidence. We both show that our system is able to capture strict Natural Logic inferences on the FraCaS test suite, and demonstrate its ability to predict common sense facts with 49% recall and 91% precision.

Session 3c: Machine Translation
Al Areen 4   *Chair: Taro Watanabe*

Modeling Term Translation for Document-informed Machine Translation
Fandong Meng, Deyi Xiong, Wenbin Jiang, and Qun Liu   15:40–16:05

Term translation is of great importance for statistical machine translation (SMT), especially document-informed SMT. In this paper, we investigate three issues of term translation in the context of document-informed SMT and propose three corresponding models: (a) a term translation disambiguation model which selects desirable translations for terms in the source language with domain information, (b) a term translation consistency model that encourages consistent translations for terms with a high strength of translation consistency throughout a document, and (c) a term bracketing model that rewards translation hypotheses where bracketable source terms are translated as a whole unit. We integrate the three models into hierarchical phrase-based SMT and evaluate their effectiveness on NIST Chinese-English translation tasks with large-scale training data. Experiment results show that all three models can achieve significant improvements over the baseline. Additionally, we can obtain a further improvement when combining the three models.

Beyond Parallel Data: Joint Word Alignment and Decipherment Improves Machine Translation
Qing Dou, Ashish Vaswani, and Kevin Knight   16:05–16:30

Inspired by previous work, where decipherment is used to improve machine translation, we propose a new idea to combine word alignment and decipherment into a single learning process. We use EM to estimate the model parameters, not only to maximize the probability of parallel corpus, but also the monolingual corpus. We apply our approach to improve Malagasy-English machine translation, where only a small amount of parallel data is available. In our experiments, we observe gains of 0.9 to 2.1 BLEU over a strong baseline.

Latent Domain Phrase-based Models for Adaptation
Cuong Hoang and Khalil Sima'an   16:30–16:55

Phrase-based models directly trained on mix-of-domain corpora can be sub-optimal. In this paper we equip phrase-based models with a latent domain variable and present a novel method for adapting them to an in-domain task represented by a seed corpus. We derive an EM algorithm which alternates between inducing domain-focused phrase pair estimates, and weights for mix-domain sentence pairs reflecting their relevance for the in-domain task. By embedding our latent domain phrase model in a sentence-level model and training the two in tandem we are able to adapt all core translation components together — phrase, lexical and reordering. We show experiments on weighing sentence pairs for relevance as well as adapting phrase-based models showing significant performance improvement in both tasks.

Translation Rules with Right-Hand Side Lattices
Fabien Cromieres and Sadao Kurohashi   16:55–17:20

In Corpus-Based Machine Translation, the search space of the translation candidates for a given input sentence is often defined by a set of (cycle-free) context-free grammar rules. This happen naturally in Syntax-Based Machine Translation and Hierarchical Phrase-Based Machine Translation. But it is also possible to describe Phrase-Based Machine Translation in this framework. We propose a natural extension to this representation by using lattice-rules that allow to easily encode an exponential number of variations of each rules. We also demonstrate how the representation of the search space has an impact on decoding efficiency, and how it is possible to optimize this representation.

Session 3-posters: Information Retrieval, Summarization and Question Answering

Al Areen 5-6 · 15:40–17:20

Learning to Translate: A Query-Specific Combination Approach for Cross-Lingual Information Retrieval

Ferhan Ture and Elizabeth Boschee

When documents and queries are presented in different languages, the common approach is to translate the query into the document language. While there are a variety of query translation approaches, recent research suggests that combining multiple methods into a single "structured query" is the most effective. In this paper, we introduce a novel approach for producing a unique combination recipe for each query, as it has also been shown that the optimal combination weights differ substantially across queries and other task specifics. Our query-specific combination method generates statistically significant improvements over other combination strategies presented in the literature, such as uniform and task-specific weighting. An in-depth empirical analysis presents insights about the effect of data size, domain differences, labeling and tuning strategies on the end performance of our approach.

Semantic-Based Multilingual Document Clustering via Tensor Modeling

Salvatore Romeo, Andrea Tagarelli, and Dino Ienco

A major challenge in document clustering research arises from the growing amount of text data written in different languages. Previous approaches depend on language-specific solutions (e.g., bilingual dictionaries, sequential machine translation) to evaluate document similarities, and the required transformations may alter the original document semantics. To cope with this issue we propose a new document clustering approach for multilingual corpora that (i) exploits a large-scale multilingual knowledge base, (ii) takes advantage of the multi-topic nature of the text documents, and (iii) employs a tensor-based model to deal with high dimensionality and sparseness. Results have shown the significance of our approach and its better performance w.r.t. classic document clustering approaches, in both a balanced and an unbalanced corpus evaluation.

Lexical Substitution for the Medical Domain

Martin Riedl, Michael Glass, and Alfio Gliozzo

In this paper we examine the lexical substitution task for the medical domain. We adapt the current best system from the open domain, which trains a single classifier for all instances using delexicalized features. We show significant improvements over a strong baseline coming from a distributional thesaurus (DT). Whereas in the open domain system, features derived from WordNet show only slight improvements, we show that its counterpart for the medical domain (UMLS) shows a significant additional benefit when used for feature generation.

Question Answering with Subgraph Embeddings

Antoine Bordes, Sumit Chopra, and Jason Weston

This paper presents a system which learns to answer questions on a broad range of topics from a knowledge base using few hand-crafted features. Our model learns low-dimensional embeddings of words and knowledge base constituents; these representations are used to score natural language questions against candidate answers. Training our system using pairs of questions and structured representations of their answers, and pairs of question paraphrases, yields competitive results on a recent benchmark of the literature.

Correcting Keyboard Layout Errors and Homoglyphs in Queries

Derek Barnes, Mahesh Joshi, and Hassan Sawaf

Keyboard layout errors and homoglyphs in cross-language queries impact our ability to correctly interpret user information needs and offer relevant results. We present a machine learning approach to correcting these errors, based largely on character-level n-gram features. We demonstrate superior performance over rule-based methods, as well as a significant reduction in the number of queries that yield null search results.

Non-linear Mapping for Improved Identification of 1300+ Languages

Ralf Brown

Non-linear mappings of the form $P(ngram)^\gamma$ and $\frac{log(1+\tau P(ngram))}{log(1+\tau)}$ are applied to the n-gram probabilities in five trainable open-source language identifiers. The first mapping reduces classification errors by 4.0% to 83.9% over a test set of more than one million 65-character strings in 1366 languages, and by 2.6% to 76.7% over a subset of 781 languages. The second mapping improves four of the five identifiers by 10.6% to 83.8% on the larger corpus and 14.4%

to 76.7% on the smaller corpus. The subset corpus and the modified programs are made freely available for download at http://www.cs.cmu.edu/~ralf/langid.html.

A Neural Network for Factoid Question Answering over Paragraphs
Mohit Iyyer, Jordan Boyd-Graber, Leonardo Claudino, Richard Socher, and Hal Daumé III

Text classification methods for tasks like factoid question answering typically use manually defined string matching rules or bag of words representations. These methods are ineffective when question text contains very few individual words (e.g., named entities) that are indicative of the answer. We introduce a recursive neural network (RNN) model that can reason over such input by modeling textual compositionality. We apply our model, QANTA, to a dataset of questions from a trivia competition called quiz bowl. Unlike previous RNN models, QANTA learns word and phrase-level representations that combine across sentences to reason about entities. The model outperforms multiple baselines and, when combined with information retrieval methods, rivals the best human players.

Joint Relational Embeddings for Knowledge-based Question Answering
Min-Chul Yang, Nan Duan, Ming Zhou, and Hae-Chang Rim

Transforming a natural language (NL) question into a corresponding logical form (LF) is central to the knowledge-based question answering (KB-QA) task. Unlike most previous methods that achieve this goal based on mappings between lexicalized phrases and logical predicates, this paper goes one step further and proposes a novel embedding-based approach that maps NL-questions into LFs for KB-QA by leveraging semantic associations between lexical representations and KB-properties in the latent space. Experimental results demonstrate that our proposed method outperforms three KB-QA baseline methods on two publicly released QA data sets.

Adding High-Precision Links to Wikipedia
Thanapon Noraset, Chandra Bhagavatula, and Doug Downey

Wikipedia's link structure is a valuable resource for natural language processing tasks, but only a fraction of the concepts mentioned in each article are annotated with hyperlinks. In this paper, we study how to augment Wikipedia with additional high-precision links. We present 3W , a system that identifies concept mentions in Wikipedia text, and links each mention to its referent page. 3W leverages rich semantic information present in Wikipedia to achieve high precision. Our experiments demonstrate that 3W can add an average of seven new links to each Wikipedia article, at a precision of 0.98.

[TACL] Crosslingual and Multilingual Construction of Syntax-Based Vector Space Models
Jason Utt and Sebastian Padó

Syntax-based distributional models of lexical semantics provide a flexible and linguistically adequate representation of co-occurrence information. However, their construction requires large, accurately parsed corpora, which are unavailable for most languages. In this paper, we develop a number of methods to overcome this obstacle. We describe (a) a crosslingual approach that constructs a syntax-based model for a new language requiring only an English resource and a translation lexicon; and (b) multilingual approaches that combine crosslingual with monolingual information, subject to availability. We evaluate on two lexical semantic benchmarks in German and Croatian. We find that the models exhibit complementary profiles: crosslingual models yield higher accuracies while monolingual models provide better coverage. In addition, we show that simple multilingual models can successfully combine their strengths.

Finding Good Enough: A Task-Based Evaluation of Query Biased Summarization for Cross-Language Information Retrieval
Jennifer Williams, Sharon Tam, and Wade Shen

In this paper we present our task-based evaluation of query biased summarization for cross-language information retrieval (CLIR) using relevance prediction. We describe our 13 summarization methods each from one of four summarization strategies. We show how well our methods perform using Farsi text from the CLEF 2008 shared-task, which we translated to English automtatically. We report precision/recall/F1, accuracy and time-on-task. We found that different summarization methods perform optimally for different evaluation metrics, but overall query biased word clouds are the best summarization strategy. In our analysis, we demonstrate that using the ROUGE metric on our sentence-based summaries cannot make the same kinds of distinctions as our evaluation framework does. Finally, we present our recommendations for creating much-needed evaluation standards and datasets.

Chinese Poetry Generation with Recurrent Neural Networks
Xingxing Zhang and Mirella Lapata

We propose a model for Chinese poem generation based on recurrent neural networks which we argue is ideally suited to capturing poetic content and form. Our generator jointly performs content selection ("what to say") and surface realization ("how to say") by learning representations of individual characters, and their combinations into one or more lines as well as how these mutually reinforce and constrain each other. Poem lines are generated incrementally by taking into account the entire history of what has been generated so far rather than the limited horizon imposed by the previous line or lexical n-grams. Experimental results show that our model outperforms competitive Chinese poetry generation systems using both automatic and manual evaluation methods.

Fear the REAPER: A System for Automatic Multi-Document Summarization with Reinforcement Learning

Cody Rioux, Sadid A. Hasan, and Yllias Chali

This paper explores alternate algorithms, reward functions and feature sets for per- forming multi-document summarization using reinforcement learning with a high focus on reproducibility. We show that ROUGE results can be improved using a unigram and bigram similarity metric when training a learner to select sentences for summarization. Learners are trained to summarize document clusters based on various algorithms and reward functions and then evaluated using ROUGE. Our ex- periments show a statistically significant improvement of 1.33%, 1.58%, and 2.25% for ROUGE-1, ROUGE-2 and ROUGE- L scores, respectively, when compared with the performance of the state of the art in automatic summarization with re- inforcement learning on the DUC2004 dataset. Furthermore query focused extensions of our approach show an improve- ment of 1.37% and 2.31% for ROUGE-2 and ROUGE-SU4 respectively over query focused extensions of the state of the art with reinforcement learning on the DUC2006 dataset.

Improving Multi-documents Summarization by Sentence Compression based on Expanded Constituent Parse Trees

Chen Li, Yang Liu, Fei Liu, Lin Zhao, and Fuliang Weng

In this paper, we focus on the problem of using sentence compression techniques to improve multi-document summarization. We propose an innovative sentence compression method by considering every node in the constituent parse tree and deciding its status C remove or retain. Integer liner programming (ILP) with discriminative training is used to solve the problem. Under this model, we incorporate various constraints to improve the linguistic quality of the compressed sentences. For summarization, we utilize a pipeline framework where sentences are compressed first by our proposed compression model to obtain top-n candidates and then a sentence selection module is used to generate the final summary. Comparing with the state-of-the-art algorithms, our model provides better linguistic quality and similar ROUGE-2 scores on TAC data.

Analyzing Stemming Approaches for Turkish Multi-Document Summarization

Muhammed Yavuz Nuzumlalı and Arzucan Özgür

In this study, we analyzed the effects of applying different levels of stemming approaches such as fixed-length word truncation and morphological analysis for multi-document summarization (MDS) on Turkish, which is an agglutinative and morphologically rich language. We constructed a manually annotated MDS data set, and to our best knowledge, reported the first results on Turkish MDS. Our results show that a simple fixed-length word truncation approach performs slightly better than no stemming, whereas applying complex morphological analysis does not improve Turkish MDS.

5

Overview

8:00 – 17:00	**Registration**			(Area 3)
8:00 – 9:00	**Refreshments**			(Area 3-4)
9:00 – 10:00	**Invited Talk: Thorsten Joachims**			(Al Areen 1-4)
10:00 – 10:30	**Coffee Break**			(Area 3-4)

Session 4

10:30 – 12:10	Neural Net Mixer	Discourse, Dialogue and Pragmatics / Summarization and Generation	Information Extraction	Posters: Segmentation, Tagging and Parsing

12:10 – 13:30	**Lunch Break**			(Area 3-4)
12:50 – 13:30	**SIGDAT Business Meeting** (bring your lunch)			(Al Areen 3)

Session 5

13:30 – 15:10	Tagging, Chunking, Parsing and Syntax	Semantics	Information Retrieval and Question Answering	Posters: NLP for theWeb, Social Media and Sentiment Analysis

15:10 – 15:40	**Coffee Break**			(Area 3-4)

Session 6

15:40 – 17:20	Machine Translation	Semantic Parsing	NLP-Related Machine Learning	Posters: Computational Psycholinguistics, Text Mining and NLP Applications

19:00 – 21:30	**Social Dinner**			(Marriott Marquis: pool side – 7th floor)

Come and enjoy an evening of Arabic culture and cuisine while networking under the stars.

Invited Talk: Thorsten Joachims

"Learning from Rational* Behavior"

Monday, October 27, 2014, 9:00–10:00

Al Areen 1-4

Abstract: The ability to learn from user interactions can give systems access to unprecedented amounts of knowledge. This is evident in search engines, recommender systems, and electronic commerce, and it can be the key to solving other knowledge intensive tasks. However, extracting the knowledge conveyed by user interactions is less straightforward than standard machine learning, since it requires learning systems that explicitly account for human decision making, human motivation, and human abilities.

In this talk, I argue that the design space of such interactive learning systems encompasses not only the machine learning algorithm itself, but also the design of the interaction under an appropriate model of user behavior. To this effect, the talk explores how integrating microeconomic models of human behavior into the learning process leads to new interaction models and their associated learning algorithms, leading to systems that have provable guarantees and that perform robustly in practice.

* Restrictions apply. Some modeling required.

Biography: Thorsten Joachims is a Professor in the Department of Computer Science and the Department of Information Science at Cornell University. His research interests center on a synthesis of theory and system building in machine learning, with applications in information access, language technology, and recommendation. His past research focused on support vector machines, text classification, structured output prediction, convex optimization, learning to rank, learning with preferences, and learning from implicit feedback. In 2001, he finished his dissertation advised by Prof. Katharina Morik at the University of Dortmund. From there he also received his Diploma in Computer Science in 1997. Between 2000 and 2001 he worked as a PostDoc at the GMD Institute for Autonomous Intelligent Systems. From 1994 to 1996 he was a visiting scholar with Prof. Tom Mitchell at Carnegie Mellon University.

Parallel Session 4 Overview – Monday, October 27, 2014

Track A	Track B	Track C	
Neural Net Mixer	*Discourse, Dialogue and Pragmatics / Summarization and Generation*	*Information Extraction*	
Al Areen 1-2	Al Areen 3	Al Areen 4	
Evaluating Neural Word Representations in Tensor-Based Compositional Settings *Dmitrijs Milajevs, Dimitri Kartsaklis, Mehrnoosh Sadrzadeh, and Matthew Purver*	Why are You Taking this Stance? Identifying and Classifying Reasons in Ideological Debates *Kazi Saidul Hasan and Vincent Ng*	Unsupervised Template Mining for Semantic Category Understanding *Lei Shi, Shuming Shi, Chin-Yew Lin, Yi-Dong Shen, and Yong Rui*	10:30
Opinion Mining with Deep Recurrent Neural Networks *Ozan Irsoy and Claire Cardie*	Chinese Zero Pronoun Resolution: An Unsupervised Probabilistic Model Rivaling Supervised Resolvers *Chen Chen and Vincent Ng*	Taxonomy Construction Using Syntactic Contextual Evidence *Tuan Luu Anh, Jung-jae Kim, and See Kiong Ng*	10:55
The Inside-Outside Recursive Neural Network model for Dependency Parsing *Phong Le and Willem Zuidema*	Unsupervised Sentence Enhancement for Automatic Summarization *Jackie Chi Kit Cheung and Gerald Penn*	Analysing recall loss in named entity slot filling *Glen Pink, Joel Nothman, and James R. Curran*	11:20
A Fast and Accurate Dependency Parser using Neural Networks *Danqi Chen and Christopher Manning*	ReferItGame: Referring to Objects in Photographs of Natural Scenes *Sahar Kazemzadeh, Vicente Ordonez, Mark Matten, and Tamara Berg*	Relieving the Computational Bottleneck: Joint Inference for Event Extraction with High-Dimensional Features *Deepak Venugopal, Chen Chen, Vibhav Gogate, and Vincent Ng*	11:45

Track P

Posters: Segmentation, Tagging and Parsing

Al Areen 5-6   10:30-12:10

- Syllable weight encodes mostly the same information for English word segmentation as dictionary stress
  John K Pate and Mark Johnson

- A Joint Model for Unsupervised Chinese Word Segmentation
  Miaohong Chen, Baobao Chang, and Wenzhe Pei

- Domain Adaptation for CRF-based Chinese Word Segmentation using Free Annotations
  Yijia Liu, Yue Zhang, Wanxiang Che, Ting Liu, and Fan Wu

- Balanced Korean Word Spacing with Structural SVM
  Changki Lee, Edward Choi, and Hyunki Kim

- Morphological Segmentation for Keyword Spotting
  Karthik Narasimhan, Damianos Karakos, Richard Schwartz, Stavros Tsakalidis, and Regina Barzilay

- What Can We Get From 1000 Tokens? A Case Study of Multilingual POS Tagging For Resource-Poor Languages
  Long Duong, Trevor Cohn, Karin Verspoor, Steven Bird, and Paul Cook

- An Experimental Comparison of Active Learning Strategies for Partially Labeled Sequences
  Diego Marcheggiani and Thierry Artières

- Language Modeling with Functional Head Constraint for Code Switching Speech Recognition
  Ying Li and Pascale Fung

- A Polynomial-Time Dynamic Oracle for Non-Projective Dependency Parsing
  Carlos Gómez-Rodríguez, Francesco Sartorio, and Giorgio Satta

- Ambiguity Resolution for Vt-N Structures in Chinese
  Yu-Ming Hsieh, Jason S. Chang, and Keh-Jiann Chen

- Neural Networks Leverage Corpus-wide Information for Part-of-speech Tagging
  Yuta Tsuboi

- System Combination for Grammatical Error Correction
  Raymond Hendy Susanto, Peter Phandi, and Hwee Tou Ng

- Dependency parsing with latent refinements of part-of-speech tags
  Thomas Mueller, Richárd Farkas, Alex Judea, and Helmut Schmid

- Importance weighting and unsupervised domain adaptation of POS taggers: a negative result
  Barbara Plank, Anders Johannsen, and Anders Søgaard

- POS Tagging of English-Hindi Code-Mixed Social Media Content
  Yogarshi Vyas, Spandana Gella, Jatin Sharma, Kalika Bali, and Monojit Choudhury

- Data Driven Grammatical Error Detection in Transcripts of Children's Speech
  Eric Morley, Anna Eva Hallin, and Brian Roark

Parallel Session 4

Session 4a: Neural Net Mixer
Al Areen 1-2   *Chair: Wen-tau Yih*

Evaluating Neural Word Representations in Tensor-Based Compositional Settings
Dmitrijs Milajevs, Dimitri Kartsaklis, Mehrnoosh Sadrzadeh, and Matthew Purver   10:30–10:55

We provide a comparative study between neural word representations and traditional vector spaces based on co-occurrence counts, in a number of compositional tasks. We use three different semantic spaces and implement seven tensor-based compositional models, which we then test (together with simpler additive and multiplicative approaches) in tasks involving verb disambiguation and sentence similarity. To check their scalability, we additionally evaluate the spaces using simple compositional methods on larger-scale tasks with less constrained language: paraphrase detection and dialogue act tagging. In the more constrained tasks, co-occurrence vectors are competitive, although choice of compositional method is important; on the larger-scale tasks, they are outperformed by neural word embeddings, which show robust, stable performance across the tasks.

Opinion Mining with Deep Recurrent Neural Networks
Ozan Irsoy and Claire Cardie   10:55–11:20

Recurrent neural networks (RNNs) are connectionist models of sequential data that are naturally applicable to the analysis of natural language. Recently, "depth in space" as an orthogonal notion to "depth in time" in RNNs has been investigated by stacking multiple layers of RNNs and shown empirically to bring temporal hierarchy to the architecture. In this work we apply these deep RNNs to the task of opinion expression extraction formulated as a token-level sequence-labeling task. Experimental results show that deep, narrow RNNs outperform traditional shallow, wide RNNs with the same number of parameters. Furthermore, our approach outperforms previous CRF-based baselines, including the state-of-the-art semi-Markov CRF model, and does so without access to the powerful opinion lexicons and syntactic features relied upon by the semi-CRF, as well as without the standard layer-by-layer pre-training typically required of RNN architectures.

The Inside-Outside Recursive Neural Network model for Dependency Parsing
Phong Le and Willem Zuidema   11:20–11:45

We propose the first implementation of an infinite-order generative dependency model. The model is based on a new recursive neural network architecture, the Inside-Outside Recursive Neural Network. This architecture allows information to flow not only bottom-up, as in traditional recursive neural networks, but also top-down. This is achieved by computing content as well as context representations for any constituent, and letting these representations interact. Experimental results on the English section of the Universal Dependency Treebank show that the infinite-order model achieves a perplexity seven times lower than the traditional third-order model using counting, and tends to choose more accurate parses in k-best lists. In addition, reranking with this model achieves state-of-the-art unlabelled attachment scores and unlabelled exact match scores.

A Fast and Accurate Dependency Parser using Neural Networks
Danqi Chen and Christopher Manning   11:45–12:10

Almost all current dependency parsers classify based on millions of sparse indicator features. Not only do these features generalize poorly, but the cost of feature computation restricts parsing speed significantly. In this work, we propose a novel way of learning a neural network classifier for use in a greedy, transition-based dependency parser. Because this classifier learns and uses just a small number of dense features, it can work very fast, while achieving an about 2% improvement in unlabeled and labeled attachment scores on both English and Chinese datasets. Concretely, our parser is able to parse more than 1000 sentences per second at 92.2% unlabeled attachment score on the English Penn Treebank.

Session 4b: Discourse, Dialogue and Pragmatics / Summarization and Generation
Al Areen 3 *Chair: Yejin Choi*

Why are You Taking this Stance? Identifying and Classifying Reasons in Ideological Debates
Kazi Saidul Hasan and Vincent Ng 10:30–10:55

Recent years have seen a surge of interest in stance classification in online debates. Oftentimes, however, it is important to determine not only the stance expressed by an author in her debate posts, but also the reasons behind her supporting or opposing the issue under debate. We therefore examine the new task of reason classification in this paper. Given the close interplay between stance classification and reason classification, we design computational models for examining how automatically computed stance information can be profitably exploited for reason classification. Experiments on our reason-annotated corpus of ideological debate posts from four domains demonstrate that sophisticated models of stances and reasons can indeed yield more accurate reason and stance classification results than their simpler counterparts.

Chinese Zero Pronoun Resolution: An Unsupervised Probabilistic Model Rivaling Supervised Resolvers
Chen Chen and Vincent Ng 10:55–11:20

State-of-the-art Chinese zero pronoun resolution systems are supervised, thus relying on training data containing manually resolved zero pronouns. To eliminate the reliance on annotated data, we present a generative model for unsupervised Chinese zero pronoun resolution. At the core of our model is a novel hypothesis: a probabilistic pronoun resolver trained on overt pronouns in an unsupervised manner can be used to resolve zero pronouns. Experiments demonstrate that our unsupervised model rivals its state-of-the-art supervised counterparts in performance when resolving the Chinese zero pronouns in the OntoNotes corpus.

Unsupervised Sentence Enhancement for Automatic Summarization
Jackie Chi Kit Cheung and Gerald Penn 11:20–11:45

We present sentence enhancement as a novel technique for text-to-text generation in abstractive summarization. Compared to extraction or previous approaches to sentence fusion, sentence enhancement increases the range of possible summary sentences by allowing a combination of dependency subtrees from any sentence from the source text. Our experiments indicate that our approach yields summary sentences that are competitive with a sentence fusion baseline in terms of content quality, but better in terms of grammaticality, and that the benefit of sentence enhancement relies crucially on an event coreference resolution algorithm that uses distributional semantics. We also consider how text-to-text generation approaches to summarization can be extended beyond the source text by examining how human summary writers incorporate source-text-external elements into their summary sentences.

ReferItGame: Referring to Objects in Photographs of Natural Scenes
Sahar Kazemzadeh, Vicente Ordonez, Mark Matten, and Tamara Berg 11:45–12:10

In this paper we introduce a new game to crowd-source natural language referring expressions. By designing a two player game, we can both collect and verify referring expressions directly within the game. To date, the game has produced a dataset containing 130,525 expressions, referring to 96,654 distinct objects, in 19,894 photographs of natural scenes. This dataset is larger and more varied than previous REG datasets and allows us to study referring expressions in real-world scenes. We provide an in depth analysis of the resulting dataset. Based on our findings, we design a new optimization based model for generating referring expressions and perform experimental evaluations on 3 test sets.

Session 4c: Information Extraction
Al Areen 4   *Chair: Heng Ji*

Unsupervised Template Mining for Semantic Category Understanding
Lei Shi, Shuming Shi, Chin-Yew Lin, Yi-Dong Shen, and Yong Rui   10:30–10:55

We propose an unsupervised approach to constructing templates from a large collection of semantic category names, and use the templates as the semantic representation of categories. The main challenge is that many terms have multiple meanings, resulting in a lot of wrong templates. Statistical data and semantic knowledge are extracted from a web corpus to improve template generation. A nonlinear scoring function is proposed and demonstrated to be effective. Experiments show that our approach achieves significantly better results than baseline methods. As an immediate application, we apply the extracted templates to the cleaning of a category collection and see promising results (precision improved from 81% to 89%).

Taxonomy Construction Using Syntactic Contextual Evidence
Tuan Luu Anh, Jung-jae Kim, and See Kiong Ng   10:55–11:20

Taxonomies are the backbone of many structured, semantic knowledge resources. Recent works for extracting taxonomic relations from text focused on collecting lexical-syntactic patterns to extract the taxonomic relations by matching the patterns to text. These approaches, however, often show low coverage due to the lack of contextual analysis across sentences. To address this issue, we propose a novel approach that collectively utilizes contextual information of terms in syntactic structures such that if the set of contexts of a term includes most of contexts of another term, a subsumption relation between the two terms is inferred. We apply this method to the task of taxonomy construction from scratch, where we introduce another novel graph-based algorithm for taxonomic structure induction. Our experiment results show that the proposed method is well complementary with previous methods of linguistic pattern matching and significantly improves recall and thus F-measure.

Analysing recall loss in named entity slot filling
Glen Pink, Joel Nothman, and James R. Curran   11:20–11:45

State-of-the-art fact extraction is heavily constrained by recall, as demonstrated by recent performance in TAC Slot Filling. We isolate this recall loss for NE slots by systematically analysing each stage of the slot filling pipeline as a filter over correct answers. Recall is critical as candidates never generated can never be recovered, whereas precision can always be increased in downstream processing. We provide precise, empirical confirmation of previously hypothesised sources of recall loss in slot filling. While NE type constraints substantially reduce the search space with only a minor recall penalty, we find that 10% to 39% of slot fills will be entirely ignored by most systems. One in six correct answers are lost if coreference is not used, but this can be mostly retained by simple name matching rules.

Relieving the Computational Bottleneck: Joint Inference for Event Extraction with High-Dimensional Features
Deepak Venugopal, Chen Chen, Vibhav Gogate, and Vincent Ng   11:45–12:10

Several state-of-the-art event extraction systems employ models based on Support Vector Machines (SVMs) in a pipeline architecture, which fails to exploit the joint dependencies that typically exist among events and arguments. While there have been attempts to overcome this limitation using Markov Logic Networks (MLNs), it remains challenging to perform joint inference in MLNs when the model encodes many high-dimensional sophisticated features such as those essential for event extraction. In this paper, we propose a new model for event extraction that combines the power of MLNs and SVMs, dwarfing their limitations. The key idea is to reliably learn and process high-dimensional features using SVMs; encode the output of SVMs as low-dimensional, soft formulas in MLNs; and use the superior joint inferencing power of MLNs to enforce joint consistency constraints over the soft formulas. We evaluate our approach for the task of extracting biomedical events on the BioNLP 2013, 2011 and 2009 Genia shared task datasets. Our approach yields the best F1 score to date on the BioNLP'13 (53.61) and BioNLP'11 (58.07) datasets and the second-best F1 score to date on the BioNLP'09 dataset (58.16).

Session 4-posters: Segmentation, Tagging and Parsing

Al Areen 5-6   10:30–12:10

Syllable weight encodes mostly the same information for English word segmentation as dictionary stress
John K Pate and Mark Johnson

Stress is a useful cue for English word segmentation. A wide range of computational models have found that stress cues enable a 2-10% improvement in segmentation accuracy, depending on the kind of model, by using input that has been annotated with stress using a pronouncing dictionary. However, stress is neither invariably produced nor unambiguously identifiable in real speech. Heavy syllables, i.e. those with long vowels or syllable codas, attract stress in English. We devise Adaptor Grammar word segmentation models that exploit either stress, or syllable weight, or both, and evaluate the utility of syllable weight as a cue to word boundaries. Our results suggest that syllable weight encodes the same information for word segmentation in English that annotated dictionary stress does.

A Joint Model for Unsupervised Chinese Word Segmentation
Miaohong Chen, Baobao Chang, and Wenzhe Pei

In this paper, we propose a joint model for unsupervised Chinese word segmentation. Inspired by the "products of experts" idea, our joint model firstly combines two generative models, which are word-based hierarchical Dirichlet process model and character-based hidden Markov model, by simply multiplying their probabilities together. Gibbs sampling is used for model inference. In order to further combine the strength of goodness-based model, we then integrated nVBE into our joint model by using it to initializing the Gibbs sampler. We conduct our experiments on PKU and MSRA datasets provided by the second SIGHAN bakeoff. Test results on these two datasets show that the joint model achieves much better results than all of its component models. Statistical significance tests also show that it is significantly better than state-of-the-art systems, achieving the highest F-scores. Finally, analysis indicates that compared with nVBE and HDP, the joint model has a stronger ability to solve both combinational and overlapping ambiguities in Chinese word segmentation.

Domain Adaptation for CRF-based Chinese Word Segmentation using Free Annotations
Yijia Liu, Yue Zhang, Wanxiang Che, Ting Liu, and Fan Wu

Supervised methods have been the dominant approach for Chinese word segmentation. The performance can drop significantly when the test domain is different from the training domain. In this paper, we study the problem of obtaining partial annotation from freely available data to help Chinese word segmentation on different domains. Different sources of free annotations are transformed into a unified form of partial annotation and a variant CRF model is used to leverage both fully and partially annotated data consistently. Experimental results show that the Chinese word segmentation model benefits from free partially annotated data. On the SIGHAN Bakeoff 2010 data, we achieve results that are competitive to the best reported in the literature.

Balanced Korean Word Spacing with Structural SVM
Changki Lee, Edward Choi, and Hyunki Kim

Most studies on statistical Korean word spacing do not utilize the information provided by the input sentence and assume that it was completely concatenated. This makes the word spacer ignore the correct spaced parts of the input sentence and erroneously alter them. To overcome such limit, this paper proposes a structural SVM-based Korean word spacing method that can utilize the space information of the input sentence. The experiment on sentences with 10% spacing errors showed that our method achieved 96.81% F-score, while the basic structural SVM method only achieved 92.53% F-score. The more the input sentence was correctly spaced, the more accurately our method performed.

Morphological Segmentation for Keyword Spotting
Karthik Narasimhan, Damianos Karakos, Richard Schwartz, Stavros Tsakalidis, and Regina Barzilay

We explore the impact of morphological segmentation on keyword spotting (KWS). Despite potential benefits, state-of-the-art KWS systems do not use morphological information. In this paper, we augment a state-of-the-art KWS system with sub-word units derived from supervised and unsupervised morphological segmentations, and compare with phonetic and syllabic segmentations. Our experiments demonstrate that morphemes improve overall performance of KWS systems. Syllabic units, however, rival the performance of morphological units when used in KWS. By combining morphological and syllabic segmentations, we demonstrate substantial performance gains.

What Can We Get From 1000 Tokens? A Case Study of Multilingual POS Tagging For Resource-Poor Languages
Long Duong, Trevor Cohn, Karin Verspoor, Steven Bird, and Paul Cook

In this paper we address the problem of multilingual part-of-speech tagging for resource-poor languages. We use parallel data to transfer part-of-speech information from resource-rich to resource-poor languages. Additionally, we use a small amount of annotated data to learn to "correct" errors from projected approach such as tagset mismatch between languages, achieving state-of-the-art performance (91.3%) across 8 languages. Our approach is based on modest data requirements, and uses minimum divergence classification. For situations where no universal tagset mapping is available, we propose an alternate method, resulting in state-of-the-art 85.6% accuracy on the resource-poor language Malagasy.

An Experimental Comparison of Active Learning Strategies for Partially Labeled Sequences
Diego Marcheggiani and Thierry Artières

Active learning (AL) consists of asking human annotators to annotate automatically selected data that are assumed to bring the most benefit in the creation of a classifier. AL allows to learn accurate systems with much less annotated data than what is required by pure supervised learning algorithms, hence limiting the tedious effort of annotating a large collection of data. We experimentally investigate the behavior of several AL strategies for sequence labeling tasks (in a partially-labeled scenario) tailored on Partially-Labeled Conditional Random Fields, on four sequence labeling tasks: phrase chunking, part-of-speech tagging, named-entity recognition, and bio-entity recognition.

Language Modeling with Functional Head Constraint for Code Switching Speech Recognition
Ying Li and Pascale Fung

In this paper, we propose novel structured language modeling methods for code mixing speech recognition by incorporating a well-known syntactic constraint for switching code, namely the Functional Head Constraint (FHC). Code mixing data is not abundantly available for training language models. Our proposed methods successfully alleviate this core problem for code mixing speech recognition by using bilingual data to train a structured language model with syntactic constraint. Linguists and bilingual speakers found that code switch do not happen between the functional head and its complements. We propose to learn the code mixing language model from bilingual data with this constraint in a weighted finite state transduction (WFST) framework. The constrained code switch language model is obtained by first expanding the search network with a translation model, and then using parsing to restrict paths to those permissible under the constraint. We implement and compare two approaches - lattice parsing enables a sequential coupling whereas partial parsing enables a tight coupling between parsing and filtering. We tested our system on a lecture speech dataset with 16% embedded second language, and on a lunch conversation dataset with 20% embedded language. Our language models with lattice parsing and partial parsing reduce word error rates from a baseline mixed language model by 3.8% and 3.9% in terms of word error rate relatively on the average on the first and second tasks respectively. It outperforms the interpolated language model by 3.72% and 5.6% in terms of word error rate relatively, and outperforms the adapted language model by 2.57% and 4.6% relatively. Our proposed approach avoids making early decisions on code switch boundaries and is therefore more robust. We address the code switch data scarcity challenge by using bilingual data with syntactic structure.

A Polynomial-Time Dynamic Oracle for Non-Projective Dependency Parsing
Carlos Gómez-Rodríguez, Francesco Sartorio, and Giorgio Satta

The introduction of dynamic oracles has considerably improved the accuracy of greedy transition-based dependency parsers, without sacrificing parsing efficiency. However, this enhancement is limited to projective parsing, and dynamic oracles have not yet been implemented for parsers supporting non-projectivity. In this paper we introduce the first such oracle, for a non-projective parser based on Attardi's parser. We show that training with this oracle improves parsing accuracy over a conventional (static) oracle on a wide range of datasets.

Ambiguity Resolution for Vt-N Structures in Chinese
Yu-Ming Hsieh, Jason S. Chang, and Keh-Jiann Chen

The syntactic ambiguity of a transitive verb (Vt) followed by a noun (N) has long been a problem in Chinese parsing. In this paper, we propose a classifier to resolve the ambiguity of Vt-N structures. The design of the classifier is based on three important guidelines, namely, adopting linguistically motivated features, using all available resources, and easy integration into a parsing model. The linguistically motivated features include semantic relations, context, and morphological structures; and the available resources are treebank, thesaurus, affix database, and large corpora. We also propose two learning approaches that resolve the problem of data sparseness by auto-parsing and extracting relative

knowledge from large-scale unlabeled data. Our experiment results show that the Vt-N classifier outperforms the current PCFG parser. Furthermore, it can be easily and effectively integrated into the PCFG parser and general statistical parsing models. Evaluation of the learning approaches indicates that world knowledge facilitates Vt-N disambiguation through data selection and error correction.

Neural Networks Leverage Corpus-wide Information for Part-of-speech Tagging
Yuta Tsuboi

We propose a neural network approach to benefit from the non-linearity of corpus-wide statistics for part-of-speech (POS) tagging. We investigated several types of corpus-wide information for the words, such as word embeddings and POS tag distributions. Since these statistics are encoded as dense continuous features, it is not trivial to combine these features comparing with sparse discrete features. Our tagger is designed as a combination of a linear model for discrete features and a feed-forward neural network that captures the non-linear interactions among the continuous features. By using several recent advances in the activation functions for neural networks, the proposed method marks new state-of-the-art accuracies for English POS tagging tasks.

System Combination for Grammatical Error Correction
Raymond Hendy Susanto, Peter Phandi, and Hwee Tou Ng

Different approaches to high-quality grammatical error correction have been proposed recently, many of which have their own strengths and weaknesses. Most of these approaches are based on classification or statistical machine translation (SMT). In this paper, we propose to combine the output from a classification-based system and an SMT-based system to improve the correction quality. We adopt the system combination technique of Heafield and Lavie (2010). We achieve an $F_{0.5}$ score of 39.39

Dependency parsing with latent refinements of part-of-speech tags
Thomas Mueller, Richárd Farkas, Alex Judea, and Helmut Schmid

In this paper we propose a method to increase dependency parser performance without using additional labeled or unlabeled data by refining the layer of predicted part-of-speech (POS) tags. We perform experiments on English and German and show significant improvements for both languages. The refinement is based on generative split-merge training for Hidden Markov models (HMMs).

Importance weighting and unsupervised domain adaptation of POS taggers: a negative result
Barbara Plank, Anders Johannsen, and Anders Søgaard

Importance weighting is a generalization of various statistical bias correction techniques. While our labeled data in NLP is heavily biased, importance weighting has seen only few applications in NLP, most of them relying on a small amount of labeled target data. The publication bias toward reporting positive results makes it hard to say whether researchers have tried. This paper presents a negative result on unsupervised domain adaptation for POS tagging. In this setup, we only have unlabeled data and thus only indirect access to the bias in emission and transition probabilities. Moreover, most errors in POS tagging are due to unseen words, and there, importance weighting cannot help. We present experiments with a wide variety of weight functions, quantilizations, as well as with randomly generated weights, to support these claims.

POS Tagging of English-Hindi Code-Mixed Social Media Content
Yogarshi Vyas, Spandana Gella, Jatin Sharma, Kalika Bali, and Monojit Choudhury

Code-mixing is frequently observed in user generated content on social media, especially from multilingual users. The linguistic complexity of such content is compounded by presence of spelling variations, transliteration and non-adherence to formal grammar. We describe our initial efforts to create a multi-level annotated corpus of Hindi-English code-mixed text collated from Facebook forums, and explore language identification, back-transliteration, normalization and POS tagging of this data. Our results show that language identification and transliteration for Hindi are two major challenges that impact POS tagging accuracy.

Data Driven Grammatical Error Detection in Transcripts of Children's Speech
Eric Morley, Anna Eva Hallin, and Brian Roark

We investigate grammatical error detection in spoken language, and present a data-driven method to train a dependency parser to automatically identify and label grammatical errors. This method is agnostic to the label set used, and the only manual annotations needed for training are grammatical error labels. We find that the proposed system is robust

to disfluencies, so that a separate stage to elide disfluencies is not required. The proposed system outperforms two baseline systems on two different corpora that use different sets of error tags. It is able to identify utterances with grammatical errors with an F1-score as high as 0.623, as compared to a baseline F1 of 0.350 on the same data.

Parallel Session 5 Overview – Monday, October 27, 2014

	Track A	Track B	Track C
	Tagging, Chunking, Parsing and Syntax	*Semantics*	*Information Retrieval and Question Answering*
	Al Areen 1-2	Al Areen 3	Al Areen 4
13:30	[TACL] Improved CCG Parsing with Semi-supervised Supertagging *Mike Lewis and Mark Steedman*	A Unified Model for Word Sense Representation and Disambiguation *Xinxiong Chen, Zhiyuan Liu, and Maosong Sun*	Tailor knowledge graph for query understanding: linking intent topics by propagation *Shi Zhao and Yan Zhang*
13:55	A* CCG Parsing with a Supertag-factored Model *Mike Lewis and Mark Steedman*	Reducing Dimensions of Tensors in Type-Driven Distributional Semantics *Tamara Polajnar, Luana Fagarasan, and Stephen Clark*	Queries as a Source of Lexicalized Commonsense Knowledge *Marius Pasca*
14:20	A Dependency Parser for Tweets *Lingpeng Kong, Nathan Schneider, Swabha Swayamdipta, Archna Bhatia, Chris Dyer, and Noah A. Smith*	An Etymological Approach to Cross-Language Orthographic Similarity. Application on Romanian *Alina Maria Ciobanu and Liviu P. Dinu*	Question Answering over Linked Data Using First-order Logic *Shizhu He, Kang Liu, Yuanzhe Zhang, Liheng Xu, and Jun Zhao*
14:45	Greed is Good if Randomized: New Inference for Dependency Parsing *Yuan Zhang, Tao Lei, Regina Barzilay, and Tommi Jaakkola*	Efficient Non-parametric Estimation of Multiple Embeddings per Word in Vector Space *Arvind Neelakantan, Jeevan Shankar, Alexandre Passos, and Andrew McCallum*	Knowledge Graph and Corpus Driven Segmentation and Answer Inference for Telegraphic Entity-seeking Queries *Mandar Joshi, Uma Sawant, and Soumen Chakrabarti*

Track P

Posters: NLP for the Web, Social Media and Sentiment Analysis

Al Areen 5-6   13:30-15:10

- A Regularized Competition Model for Question Difficulty Estimation in Community Question Answering Services
  Quan Wang, Jing Liu, Bin Wang, and Li Guo

- Vote Prediction on Comments in Social Polls
  Isaac Persing and Vincent Ng

- Exploiting Social Relations and Sentiment for Stock Prediction
  Jianfeng Si, Arjun Mukherjee, Bing Liu, Sinno Jialin Pan, Qing Li, and Huayi Li

- Developing Age and Gender Predictive Lexica over Social Media
  Maarten Sap, Gregory Park, Johannes Eichstaedt, Margaret Kern, David Stillwell, Michal Kosinski, Lyle Ungar, and Hansen Andrew Schwartz

- Dependency Parsing for Weibo: An Efficient Probabilistic Logic Programming Approach
  William Yang Wang, Lingpeng Kong, Kathryn Mazaitis, and William W Cohen

- Exploiting Community Emotion for Microblog Event Detection
  Gaoyan Ou, Wei Chen, Tengjiao Wang, Zhongyu Wei, Binyang Li, Dongqing Yang, and Kam-Fai Wong

- Detecting Disagreement in Conversations using Pseudo-Monologic Rhetorical Structure
  Kelsey Allen, Giuseppe Carenini, and Raymond Ng

- +/-EffectWordNet: Sense-level Lexicon Acquisition for Opinion Inference
  Yoonjung Choi and Janyce Wiebe

- A Sentiment-aligned Topic Model for Product Aspect Rating Prediction
  Hao Wang and Martin Ester

- Learning Emotion Indicators from Tweets: Hashtags, Hashtag Patterns, and Phrases
  Ashequl Qadir and Ellen Riloff

- Fine-Grained Contextual Predictions for Hard Sentiment Words
  Sebastian Ebert and Hinrich Schütze

- An Iterative Link-based Method for Parallel Web Page Mining
  Le Liu, Yu Hong, Jun Lu, Jun Lang, Heng Ji, and Jianmin Yao

- [TACL] Exploiting Social Network Structure for Person-to-Person Sentiment Analysis
  Robert West, Hristo Paskov, Jure Leskovec, and Christopher Potts

Parallel Session 5

Session 5a: Tagging, Chunking, Parsing and Syntax
Al Areen 1-2 *Chair: Trevor Cohn*

[TACL] Improved CCG Parsing with Semi-supervised Supertagging
Mike Lewis and Mark Steedman 13:30–13:55
Current supervised parsers are limited by the size of their labelled training data, making improving them with unlabelled data an important goal. We show how a state-of-the-art CCG parser can be enhanced, by predicting lexical categories using unsupervised vector-space embeddings of words. The use of word embeddings enables our model to better generalize from the labelled data, and allows us to accurately assign lexical categories without depending on a POS-tagger. Our approach leads to substantial improvements in dependency parsing results over the standard supervised CCG parser when evaluated on Wall Street Journal (0.8%), Wikipedia (1.8%) and biomedical (3.4%) text. We compare the performance of two recently proposed approaches for classification using a wide variety of word embeddings. We also give a detailed error analysis demonstrating where using embeddings outperforms traditional feature sets, and showing how including POS features can decrease accuracy.

A* CCG Parsing with a Supertag-factored Model
Mike Lewis and Mark Steedman 13:55–14:20
We introduce a new CCG parsing model which is factored on lexical category assignments. Parsing is then simply a deterministic search for the most probable category sequence that supports a CCG derivation. The parser is extremely simple, with a tiny feature set, no POS tagger, and no statistical model of the derivation or dependencies. Formulating the model in this way allows a highly effective heuristic for A* parsing, which makes parsing extremely fast. Compared to the standard C&C CCG parser, our model is more accurate out-of-domain, is four times faster, has higher coverage, and is greatly simplified. We also show that using our parser improves the performance of a state-of-the-art question answering system.

A Dependency Parser for Tweets
Lingpeng Kong, Nathan Schneider, Swabha Swayamdipta, Archna Bhatia, Chris Dyer, and Noah A. Smith 14:20–14:45
We describe a new dependency parser for English tweets. The parser builds on several contributions: new syntactic annotations for a corpus of tweets, with conventions heavily informed by the domain; adaptations to a statistical parsing algorithm; and a new approach to exploiting out-of-domain Penn Treebank data. Our experiments show that the parser achieves over 80% unlabeled attachment accuracy on our new, high-quality test set and measure the benefit of our contributions. The dataset and parser will be made publicly available on publication.

Greed is Good if Randomized: New Inference for Dependency Parsing
Yuan Zhang, Tao Lei, Regina Barzilay, and Tommi Jaakkola 14:45–15:10
Dependency parsing with high-order features results in a provably hard decoding problem. A lot of work has gone into developing powerful optimization methods for solving these combinatorial problems. In contrast, we explore, analyze, and demonstrate that a substantially simpler randomized greedy inference algorithm already suffices for near optimal parsing: a) we analytically quantify the number of local optima that the greedy method has to overcome in the context of first-order parsing; b) we show that, as a decoding algorithm, the greedy method surpasses dual decomposition in second-order parsing; c) we empirically demonstrate that our approach with up to third-order and global features outperforms the state-of-the-art dual decomposition and MCMC sampling methods when evaluated on 14 languages of non-projective CoNLL datasets.

Session 5b: Semantics
Al Areen 3   *Chair: Dipanjan Das*

A Unified Model for Word Sense Representation and Disambiguation
Xinxiong Chen, Zhiyuan Liu, and Maosong Sun   13:30–13:55

Most word representation methods assume that each word owns a single semantic vector. This is usually problematic because lexical ambiguity is ubiquitous, which is also the problem to be resolved by word sense disambiguation. In this paper, we present a unified model for joint word sense representation and disambiguation, which will assign distinct representations for each word sense. The basic idea is that both word sense representation (WSR) and word sense disambiguation (WSD) will benefit from each other: (1) high-quality WSR will capture rich information about words and senses, which should be helpful for WSD, and (2) high-quality WSD will provide reliable disambiguated corpora for learning better sense representations. Experimental results show that, our model improves the performance of contextual word similarity compared to existing WSR methods, outperforms state-of-the-art supervised methods on domain-specific WSD, and achieves competitive performance on coarse-grained all-words WSD.

Reducing Dimensions of Tensors in Type-Driven Distributional Semantics
Tamara Polajnar, Luana Fagarasan, and Stephen Clark   13:55–14:20

Compositional distributional semantics is a subfield of Computational Linguistics which investigates methods for representing the meanings of phrases and sentences. In this paper, we explore implementations of a framework based on Combinatory Categorial Grammar (CCG), in which words with certain grammatical types have meanings represented by multi-linear maps (i.e. multi-dimensional arrays, or tensors). An obstacle to full implementation of the framework is the size of these tensors. We examine the performance of lower dimensional approximations of transitive verb tensors on a sentence plausibility/selectional preference task. We find that the matrices perform as well as, and sometimes even better than, full tensors, allowing a reduction in the number of parameters needed to model the framework.

An Etymological Approach to Cross-Language Orthographic Similarity. Application on Romanian
Alina Maria Ciobanu and Liviu P. Dinu   14:20–14:45

In this paper we propose a computational method for determining the orthographic similarity between Romanian and related languages. We account for etymons and cognates and we investigate not only the number of related words, but also their forms, quantifying orthographic similarities. The method we propose is adaptable to any language, as far as resources are available.

Efficient Non-parametric Estimation of Multiple Embeddings per Word in Vector Space
Arvind Neelakantan, Jeevan Shankar, Alexandre Passos, and Andrew McCallum   14:45–15:10

There is rising interest in vector-space word embeddings and their use in NLP, especially given recent methods for their fast estimation at very large scale. Nearly all this work, however, assumes a single vector per word type—ignoring polysemy and thus jeopardizing their usefulness for downstream tasks. We present an extension to the Skip-gram model that efficiently learns multiple embeddings per word type. It differs from recent related work by jointly performing word sense discrimination and embedding learning, by non-parametrically estimating the number of senses per word type, and by its efficiency and scalability. We present new state-of-the-art results in the word similarity in context task and demonstrate its scalability by training with one machine on a corpus of nearly 1 billion tokens in less than 6 hours.

Session 5c: Information Retrieval and Question Answering
Al Areen 4   *Chair: Gregory Greffenstette*

Tailor knowledge graph for query understanding: linking intent topics by propagation
Shi Zhao and Yan Zhang   13:30–13:55
Knowledge graph are recently used for enriching query representations in an entity-aware way for the rich facts organized around entities in it. However, few of the methods pay attention to non-entity words and clicked websites in queries, which also help conveying user intent. In this paper, we tackle the problem of intent understanding with innovatively representing entity words, refiners and clicked urls as intent topics in a unified knowledge graph based framework, in a way to exploit and expand knowledge graph which we call 'tailor'. We collaboratively exploit global knowledge in knowledge graph and local context in query session to initialize intent representation, then propagate the enriched features in a graph consisting of intent topics using an unsupervised algorithm. The experiments prove intent topics with knowledge graph enriched features significantly enhance intent understanding.

Queries as a Source of Lexicalized Commonsense Knowledge
Marius Pasca   13:55–14:20
The role of Web search queries has been demonstrated in the extraction of attributes of instances and classes, or of sets of related instances and their class labels. This paper explores the acquisition of open-domain commonsense knowledge, usually available as factual knowledge, from Web search queries. Similarly to previous work in open-domain information extraction, knowledge extracted from text - in this case, from queries - takes the form of lexicalized assertions associated with open-domain classes. Experimental results indicate that facts extracted from queries complement, and have competitive accuracy levels relative to, facts extracted from Web documents by previous methods.

Question Answering over Linked Data Using First-order Logic
Shizhu He, Kang Liu, Yuanzhe Zhang, Liheng Xu, and Jun Zhao   14:20–14:45
Question Answering over Linked Data (QALD) aims to evaluate a question answering system over structured data, the key objective of which is to translate questions posed using natural language into structured queries. This technique can help common users to directly access open-structured knowledge on the Web and, accordingly, has attracted much attention. To this end, we propose a novel method using first-order logic. We formulate the knowledge for resolving the ambiguities in the main three steps of QALD (phrase detection, phrase-to-semantic-item mapping and semantic item grouping) as first-order logic clauses in a Markov Logic Network. All clauses can then produce interacted effects in a unified framework and can jointly resolve all ambiguities. Moreover, our method adopts a pattern-learning strategy for semantic item grouping. In this way, our method can cover more text expressions and answer more questions than previous methods using manually designed patterns. The experimental results using open benchmarks demonstrate the effectiveness of the proposed method.

Knowledge Graph and Corpus Driven Segmentation and Answer Inference for Telegraphic Entity-seeking Queries
Mandar Joshi, Uma Sawant, and Soumen Chakrabarti   14:45–15:10
Much recent work focuses on formal interpretation of natural question utterances, with the goal of executing the resulting structured queries on knowledge graphs (KGs) such as Freebase. Here we address two limitations of this approach when applied to open-domain, entity-oriented Web queries. First, Web queries are rarely well-formed questions. They are "telegraphic", with missing verbs, prepositions, clauses, case and phrase clues. Second, the KG is always incomplete, unable to directly answer many queries. We propose a novel technique to segment a telegraphic query and assign a coarse-grained purpose to each segment: a base entity e1 , a relation type r, a target entity type t2 , and contextual words s. The query seeks entity e2 of type t2 where r(e1 , e2) holds, further evidenced by schema-agnostic words s. Query segmentation is integrated with the KG and an unstructured corpus where mentions of entities have been linked to the KG. We do not trust the best or any specific query segmentation. Instead, evidence in favor of candidate e2s are aggregated across several segmentations. Extensive experiments on the ClueWeb corpus and parts of Freebase as our KG, using over a thousand telegraphic queries adapted from TREC, INEX, and WebQuestions, show the efficacy of our approach. For one benchmark, MAP improves from 0.2—0.29 (competitive baselines) to 0.42 (our system). NDCG10 improves from 0.29—0.36 to 0.54.

Session 5-posters: NLP for the Web, Social Media and Sentiment Analysis
Al Areen 5-6   13:30–15:10

A Regularized Competition Model for Question Difficulty Estimation in Community Question Answering Services
Quan Wang, Jing Liu, Bin Wang, and Li Guo

Estimating questions' difficulty levels is an important task in community question answering (CQA) services. Previous studies propose to solve this problem based on the question-user comparisons extracted from the question answering threads. However, they suffer from data sparseness problem as each question only gets a limited number of comparisons. Moreover, they cannot handle newly posted questions which get no comparisons. In this paper, we propose a novel question difficulty estimation approach called Regularized Competition Model (RCM), which naturally combines question-user comparisons and questions' textual descriptions into a unified framework. By incorporating textual information, RCM can effectively deal with data sparseness problem. We further employ a K-Nearest Neighbor approach to estimate difficulty levels of newly posted questions, again by leveraging textual similarities. Experiments on two publicly available data sets show that for both well-resolved and newly-posted questions, RCM performs the estimation task significantly better than existing methods, demonstrating the advantage of incorporating textual information. More interestingly, we observe that RCM might provide an automatic way to quantitatively measure the knowledge levels of words.

Vote Prediction on Comments in Social Polls
Isaac Persing and Vincent Ng

A poll consists of a question and a set of predefined answers from which voters can select. We present the new problem of vote prediction on comments, which involves determining which of these answers a voter selected given a comment she wrote after voting. To address this task, we exploit not only the information extracted from the comments but also extra-textual information such as user demographic information and inter-comment constraints. In an evaluation involving nearly one million comments collected from the popular SodaHead social polling website, we show that a vote prediction system that exploits only textual information can be improved significantly when extended with extra-textual information.

Exploiting Social Relations and Sentiment for Stock Prediction
Jianfeng Si, Arjun Mukherjee, Bing Liu, Sinno Jialin Pan, Qing Li, and Huayi Li

In this paper we first exploit cash-tags ("$" followed by stocks' ticker symbols) in Twitter to build a stock network, where nodes are stocks connected by edges when two stocks co-occur frequently in tweets. We then employ a labeled topic model to jointly model both the tweets and the network structure to assign each node and each edge a topic respectively. This Semantic Stock Network (SSN) summarizes discussion topics about stocks and stock relations. We further show that social sentiment about stock (node) topics and stock relationship (edge) topics are predictive of each stock's market. For prediction, we propose to regress the topic-sentiment time-series and the stock's price time series. Experimental results demonstrate that topic sentiments from close neighbors are able to help improve the prediction of a stock markedly.

Developing Age and Gender Predictive Lexica over Social Media
Maarten Sap, Gregory Park, Johannes Eichstaedt, Margaret Kern, David Stillwell, Michal Kosinski, Lyle Ungar, and Hansen Andrew Schwartz

Demographic lexica have potential for widespread use in social science, economic, and business applications. We derive predictive lexica (words and weights) for age and gender using regression and classification models from word usage in Facebook, blog, and Twitter data with associated demographic labels. The lexica, made publicly available, achieved state-of-the-art accuracy in language based age and gender prediction over Facebook and Twitter, and were evaluated for generalization across social media genres as well as in limited message situations.

Dependency Parsing for Weibo: An Efficient Probabilistic Logic Programming Approach
William Yang Wang, Lingpeng Kong, Kathryn Mazaitis, and William W Cohen

Dependency parsing is a core task in NLP, and it is widely used by many applications such as information extraction, question answering, and machine translation. In the era of social media, a big challenge is that parsers trained on traditional newswire corpora typically suffer from the domain mismatch issue, and thus perform poorly on social media data. We present a new GFL/FUDG-annotated Chinese treebank with more than 18K tokens from Sina Weibo (the Chinese equivalent of Twitter). We formulate the dependency parsing problem as many small and parallelizable

arc prediction tasks: for each task, we use a programmable probabilistic first-order logic to infer the dependency arc of a token in the sentence. In experiments, we show that the proposed model outperforms an off-the-shelf Stanford Chinese parser, as well as a strong MaltParser baseline that is trained on the same in-domain data.

Exploiting Community Emotion for Microblog Event Detection
Gaoyan Ou, Wei Chen, Tengjiao Wang, Zhongyu Wei, Binyang Li, Dongqing Yang, and Kam-Fai Wong

Microblog has become a major platform for information about real-world events. Automatically discovering real-world events from microblog has attracted the attention of many researchers. However, most of existing work ignore the importance of emotion information for event detection. We argue that people's emotional reactions immediately reflect the occurring of real-world events and should be important for event detection. In this study, we focus on the problem of community-related event detection by community emotions. To address the problem, we propose a novel framework which include the following three key components: microblog emotion classification, community emotion aggregation and community emotion burst detection. We evaluate our approach on real microblog data sets. Experimental results demonstrate the effectiveness of the proposed framework.

Detecting Disagreement in Conversations using Pseudo-Monologic Rhetorical Structure
Kelsey Allen, Giuseppe Carenini, and Raymond Ng

Casual online forums such as Reddit, Slashdot and Digg, are continuing to increase in popularity as a means of communication. Detecting disagreement in this domain is a considerable challenge. Many topics are unique to the conversation on the forum, and the appearance of disagreement may be much more subtle than on political blogs or social media sites such as twitter. In this analysis we present a crowd-sourced annotated corpus for topic level disagreement detection in Slashdot, showing that disagreement detection in this domain is difficult even for humans. We then proceed to show that a new set of features determined from the rhetorical structure of the conversation significantly improves the performance on disagreement detection over a baseline consisting of unigram/bigram features, discourse markers, structural features and meta-post features.

+/-EffectWordNet: Sense-level Lexicon Acquisition for Opinion Inference
Yoonjung Choi and Janyce Wiebe

Recently, work in NLP was initiated on a type of opinion inference that arises when opinions are expressed toward events which have positive or negative effects on entities +/-effect events. This paper addresses methods for creating a lexicon of such events, to support such work on opinion inference. Due to significant sense ambiguity, our goal is to develop a sense-level rather than word-level lexicon. To maximize the effectiveness of different types of information, we combine a graph-based method using WordNet relations and a standard classifier using gloss information. A hybrid between the two gives the best results. Further, we provide evidence that the model is an effective way to guide manual annotation to find +/-effect senses that are not in the seed set.

A Sentiment-aligned Topic Model for Product Aspect Rating Prediction
Hao Wang and Martin Ester

Aspect-based opinion mining has attracted lots of attention today. In this paper, we address the problem of product aspect rating prediction, where we would like to extract the product aspects, and predict aspect ratings simultaneously. Topic models have been widely adapted to jointly model aspects and sentiments, but existing models may not do the prediction task well due to their weakness in sentiment extraction. The sentiment topics usually do not have clear correspondence to commonly used ratings, and the model may fail to extract certain kinds of sentiments due to skewed data. To tackle this problem, we propose a sentiment-aligned topic model(SATM), where we incorporate two types of external knowledge: product-level overall rating distribution and word-level sentiment lexicon. Experiments on real dataset demonstrate that SATM is effective on product aspect rating prediction, and it achieves better performance compared to the existing approaches.

Learning Emotion Indicators from Tweets: Hashtags, Hashtag Patterns, and Phrases
Ashequl Qadir and Ellen Riloff

We present a weakly supervised approach for learning hashtags, hashtag patterns, and phrases associated with five emotions: AFFECTION, ANGER/RAGE, FEAR/ANXIETY, JOY, and SADNESS/DISAPPOINTMENT. Starting with seed hashtags to label an initial set of tweets, we train emotion classifiers and use them to learn new emotion hashtags and hashtag patterns. This process then repeats in a bootstrapping framework. Emotion phrases are also extracted from the learned hashtags and used to create phrase-based emotion classifiers. We show that the learned set of emotion

indicators yields a substantial improvement in F-scores, ranging from +%5 to +%18 over baseline classifiers.

Fine-Grained Contextual Predictions for Hard Sentiment Words
Sebastian Ebert and Hinrich Schütze

We put forward the hypothesis that high-accuracy sentiment analysis is only possible if word senses with different polarity are accurately recognized. We provide evidence for this hypothesis in a case study for the adjective "hard" and propose contextually enhanced sentiment lexicons that contain the information necessary for sentiment-relevant sense disambiguation. An experimental evaluation demonstrates that senses with different polarity can be distinguished well using a combination of standard and novel features.

An Iterative Link-based Method for Parallel Web Page Mining
Le Liu, Yu Hong, Jun Lu, Jun Lang, Heng Ji, and Jianmin Yao

Identifying parallel web pages from bilingual web sites is a crucial step of bilingual resource construction for cross-lingual information processing. In this paper, we propose a link-based approach to distinguish parallel web pages from bilingual web sites. Compared with the existing methods, which only employ the internal translation similarity (such as content-based similarity and page structural similarity), we hypothesize that the external translation similarity is an effective feature to identify parallel web pages. Within a bilingual web site, web pages are interconnected by hyperlinks. The basic idea of our method is that the translation similarity of two pages can be inferred from their neighbor pages, which can be adopted as an important source of external similarity. Thus, the translation similarity of page pairs will influence each other. An iterative algorithm is developed to estimate the external translation similarity and the final translation similarity. Both internal and external similarity measures are combined in the iterative algorithm. Experiments on six bilingual websites demonstrate that our method is effective and obtains significant improvement (6.2% F-Score) over the baseline which only utilizes internal translation similarity

[TACL] Exploiting Social Network Structure for Person-to-Person Sentiment Analysis
Robert West, Hristo Paskov, Jure Leskovec, and Christopher Potts

Person-to-person evaluations are prevalent in all kinds of discourse and important for establishing reputations, building social bonds, and shaping public opinion. Such evaluations can be analyzed separately using signed social networks and textual sentiment analysis, but this misses the rich interactions between language and social context. To capture such interactions, we develop a model that predicts individual A's opinion of individual B by synthesizing information from the signed social network in which A and B are embedded with sentiment analysis of the evaluative texts relating A to B. We prove that this problem is NP-hard but can be relaxed to an efficiently solvable hinge-loss Markov random field, and we show that this implementation outperforms text-only and network-only versions in two very different datasets involving community-level decision-making: the Convote U.S. Congressional speech corpus and the Wikipedia Requests for Adminship corpus.

Parallel Session 6 Overview – Monday, October 27, 2014

	Track A *Machine Translation* Al Areen 1-2	**Track B** *Semantic Parsing* Al Areen 3	**Track C** *NLP-Related Machine Learning* Al Areen 4
15:40	Human Effort and Machine Learnability in Computer Aided Translation *Spence Green, Sida I. Wang, Jason Chuang, Jeffrey Heer, Sebastian Schuster, and Christopher Manning*	Learning Compact Lexicons for CCG Semantic Parsing *Yoav Artzi, Dipanjan Das, and Slav Petrov*	Low-dimensional Embeddings for Interpretable Anchor-based Topic Inference *David Mimno and Moontae Lee*
16:05	Exact Decoding for Phrase-Based Statistical Machine Translation *Wilker Aziz, Marc Dymetman, and Lucia Specia*	Morpho-syntactic Lexical Generalization for CCG Semantic Parsing *Adrienne Wang, Tom Kwiatkowski, and Luke Zettlemoyer*	Weakly-Supervised Learning with Cost-Augmented Contrastive Estimation *Kevin Gimpel and Mohit Bansal*
16:30	Large-scale Expected BLEU Training of Phrase-based Reordering Models *Michael Auli, Michel Galley, and Jianfeng Gao*	Semantic Parsing Using Content and Context: A Case Study from Requirements Elicitation *Reut Tsarfaty, Ilia Pogrebezky, Guy Weiss, Yaarit Natan, Smadar Szekely, and David Harel*	Don't Until the Final Verb Wait: Reinforcement Learning for Simultaneous Machine Translation *Alvin Grissom II, He He, Jordan Boyd-Graber, John Morgan, and Hal Daumé III*
16:55	Confidence-based Rewriting of Machine Translation Output *Benjamin Marie and Aurélien Max*	Semantic Parsing with Relaxed Hybrid Trees *Wei Lu*	PCFG Induction for Unsupervised Parsing and Language Modelling *James Scicluna and Colin de la Higuera*

Track P

Posters: Computational Psycholinguistics, Text Mining and NLP Applications

Al Areen 5-6   15:40-17:20

- Can characters reveal your native language? A language-independent approach to native language identification
  Radu Tudor Ionescu, Marius Popescu, and Aoife Cahill

- Formalizing Word Sampling for Vocabulary Prediction as Graph-based Active Learning
  Yo Ehara, Yusuke Miyao, Hidekazu Oiwa, Issei Sato, and Hiroshi Nakagawa

- Language Transfer Hypotheses with Linear SVM Weights
  Shervin Malmasi and Mark Dras

- Predicting Dialect Variation in Immigrant Contexts Using Light Verb Constructions
  A. Seza Doğruöz and Preslav Nakov

- Device-Dependent Readability for Improved Text Understanding
  A-Yeong Kim, Hyun-Je Song, Seong-Bae Park, and Sang-Jo Lee

- Predicting Chinese Abbreviations with Minimum Semantic Unit and Global Constraints
  Longkai Zhang, Houfeng Wang, and Xu Sun

- Using Structured Events to Predict Stock Price Movement: An Empirical Investigation
  Xiao Ding, Yue Zhang, Ting Liu, and Junwen Duan

- Extracting Clusters of Specialist Terms from Unstructured Text
  Aaron Gerow

- Citation-Enhanced Keyphrase Extraction from Research Papers: A Supervised Approach
  Cornelia Caragea, Florin Adrian Bulgarov, Andreea Godea, and Sujatha Das Gollapalli

- Using Mined Coreference Chains as a Resource for a Semantic Task
  Heike Adel and Hinrich Schütze

- Financial Keyword Expansion via Continuous Word Vector Representations
  Ming-Feng Tsai and Chuan-Ju Wang

- Intrinsic Plagiarism Detection using N-gram Classes
  Imene Bensalem, Paolo Rosso, and Salim Chikhi

- Verifiably Effective Arabic Dialect Identification
  Kareem Darwish, Hassan Sajjad, and Hamdy Mubarak

- Keystroke Patterns as Prosody in Digital Writings: A Case Study with Deceptive Reviews and Essays
  Ritwik Banerjee, Song Feng, Jun Seok Kang, and Yejin Choi

- Leveraging Effective Query Modeling Techniques for Speech Recognition and Summarization
  Kuan-Yu Chen, Shih-Hung Liu, Berlin Chen, Ea-Ee Jan, Hsin-Min Wang, Wen-Lian Hsu, and Hsin-Hsi Chen

- Staying on Topic: An Indicator of Power in Political Debates
  Vinodkumar Prabhakaran, Ashima Arora, and Owen Rambow

Parallel Session 6

Session 6a: Machine Translation
Al Areen 1-2   *Chair: Andrei Popescu-Belis*

Human Effort and Machine Learnability in Computer Aided Translation
Spence Green, Sida I. Wang, Jason Chuang, Jeffrey Heer, Sebastian Schuster, and Christopher Manning
13:40–16:05

Analyses of computer aided translation typically focus on either frontend interfaces and human effort, or backend translation and machine learnability of corrections. However, these distinctions are artificial in practice since the frontend and backend must work in concert. We present the first holistic, quantitative evaluation of these issues by contrasting two assistive modes: post-editing and interactive machine translation (MT). We describe a new translator interface, extensive modifications to a phrase-based MT system, and a novel objective function for re-tuning to human corrections. Evaluation with professional bilingual translators shows that post-edit is faster than interactive at the cost of translation quality for French-English and English-German. However, re-tuning the MT system to interactive output leads to larger, statistically significant reductions in HTER versus re-tuning to post-edit. Analysis shows that tuning directly to HTER results in fine-grained corrections to subsequent machine output.

Exact Decoding for Phrase-Based Statistical Machine Translation
Wilker Aziz, Marc Dymetman, and Lucia Specia   16:05–16:30

The combinatorial space of translation derivations in phrase-based statistical machine translation is given by the intersection between a translation lattice and a target language model. We replace this intractable intersection by a tractable relaxation which incorporates a low-order upperbound on the language model. Exact optimisation is achieved through a coarse-to-fine strategy with connections to adaptive rejection sampling. We perform exact optimisation with unpruned language models of order 3 to 5 and show search-error curves for beam search and cube pruning on standard test sets. This is the first work to tractably tackle exact optimisation with language models of orders higher than 3.

Large-scale Expected BLEU Training of Phrase-based Reordering Models
Michael Auli, Michel Galley, and Jianfeng Gao   16:30–16:55

Recent work by Cherry (2013) has shown that directly optimizing phrase-based reordering models towards BLEU can lead to significant gains. Their approach is limited to small training sets of a few thousand sentences and a similar number of sparse features. We show how the expected BLEU objective allows us to train a simple linear discriminative reordering model with millions of sparse features on hundreds of thousands of sentences resulting in significant improvements. A comparison to likelihood training demonstrates that expected BLEU is vastly more effective. Our best results improve a hierarchical lexicalized reordering baseline by up to 2.0 BLEU in a single-reference setting on a French-English WMT 2012 setup.

Confidence-based Rewriting of Machine Translation Output
Benjamin Marie and Aurélien Max   16:55–17:20

Numerous works in Statistical Machine Translation (SMT) have attempted to identify better translation hypotheses obtained by an initial decoding using an improved, but more costly scoring function. In this work, we introduce an approach that takes the hypotheses produced by a state-of-the-art, reranked phrase-based SMT system, and explores new parts of the search space by applying rewriting rules selected on the basis of posterior phrase- level confidence. In the medical domain, we obtain a 1.9 BLEU improvement over a reranked baseline exploiting the same scoring function, corresponding to a 5.4 BLEU improvement over the original Moses baseline. We show that if an indication of which phrases require rewriting is provided, our automatic rewriting procedure yields an additional improvement of 1.5 BLEU. Various analyses, including a manual error analysis, further illustrate the good performance and potential for improvement of our approach in spite of its simplicity.

Session 6b: Semantic Parsing
Al Areen 3 *Chair: Gerald Penn*

Learning Compact Lexicons for CCG Semantic Parsing
Yoav Artzi, Dipanjan Das, and Slav Petrov 13:40–16:05

We present methods to control the lexicon size when learning a Combinatory Categorial Grammar semantic parser. Existing methods incrementally expand the lexicon by greedily adding entries, considering a single training datapoint at a time. We propose using corpus-level statistics for lexicon learning decisions. We introduce voting to globally consider adding entries to the lexicon, and pruning to remove entries no longer required to explain the training data. Our methods result in state-of-the-art performance on the task of executing sequences of natural language instructions, achieving up to 25% error reduction, with lexicons that are up to 70% smaller and are qualitatively less noisy.

Morpho-syntactic Lexical Generalization for CCG Semantic Parsing
Adrienne Wang, Tom Kwiatkowski, and Luke Zettlemoyer 16:05–16:30

In this paper, we demonstrate that significant performance gains can be achieved in CCG semantic parsing by introducing a linguistically motivated grammar induction scheme. We present a new morpho-syntactic factored lexicon that models systematic variations in morphology, syntax, and semantics across word classes. The grammar uses domain-independent facts about the English language to restrict the number of incorrect parses that must be considered, thereby enabling effective learning from less data. Experiments in benchmark domains match previous models with one quarter of the data and provide new state-of-the-art results with all available data, including up to 45% relative test-error reduction.

Semantic Parsing Using Content and Context: A Case Study from Requirements Elicitation
Reut Tsarfaty, Ilia Pogrebezky, Guy Weiss, Yaarit Natan, Smadar Szekely, and David Harel 16:30–16:55

We present a model for automatic semantic analysis of requirements elicitation documents. Our target semantic representation employs live sequence charts (LSC), a multi-modal visual language for scenario-based programming, which has a direct translation into executable code. The architecture we propose integrates sentence-level and discourse-level processing in a generative probabilistic model for the analysis and disambiguation of individual sentences in context. We empirically show that the joint model consistently outperforms a sentence-based model in terms of constructing a system that reflects all the static (entities, properties) and dynamic (behavioral scenarios) requirements.

Semantic Parsing with Relaxed Hybrid Trees
Wei Lu 16:55–17:20

We propose a novel model for parsing natural language sentences into their formal semantic representations. The model is able to perform integrated lexicon acquisition and semantic parsing, mapping each atomic element in a complete semantic representation to a contiguous word sequence in the input sentence in a recursive manner, where certain overlappings amongst such word sequences are allowed. It defines distributions over the novel relaxed hybrid tree structures which jointly represent both sentences and semantics. Such structures allow tractable dynamic programming algorithms to be developed for efficient learning and decoding. Trained under a discriminative setting, our model is able to incorporate a rich set of features where certain long-distance dependencies can be captured in a principled manner. We demonstrate through experiments that by exploiting a large collection of simple features, our model is shown to be competitive to previous works and achieves state-of-the-art performance on standard benchmark data across four different languages. The system and code can be downloaded from http://statnlp.org/research/sp/.

Session 6c: NLP-Related Machine Learning

Al Areen 4   *Chair: Xavier Carreras*

Low-dimensional Embeddings for Interpretable Anchor-based Topic Inference

David Mimno and Moontae Lee   13:40–16:05

The anchor words algorithm performs provably efficient topic model inference by finding an approximate convex hull in a high-dimensional word co-occurrence space. However, the existing greedy algorithm often selects poor anchor words, reducing topic quality and interpretability. Rather than finding an approximate convex hull in a high-dimensional space, we propose to find an exact convex hull in a visualizable 2- or 3-dimensional space. Such low-dimensional embeddings both improve topics and clearly show users why the algorithm selects certain words.

Weakly-Supervised Learning with Cost-Augmented Contrastive Estimation

Kevin Gimpel and Mohit Bansal   16:05–16:30

We generalize contrastive estimation in two ways that permit adding more knowledge to unsupervised learning. The first allows the modeler to specify not only the set of corrupted inputs for each observation, but also how bad each one is. The second allows specifying structural preferences on the latent variable used to explain the observations. They require setting additional hyperparameters, which can be problematic in unsupervised learning, so we investigate new methods for unsupervised model selection and system combination. We instantiate these ideas for part-of-speech induction without tag dictionaries, improving over contrastive estimation as well as strong benchmarks from the PASCAL 2012 shared task.

Don't Until the Final Verb Wait: Reinforcement Learning for Simultaneous Machine Translation

Alvin Grissom II, He He, Jordan Boyd-Graber, John Morgan, and Hal Daumé III   16:30–16:55

We introduce a reinforcement learning-based approach to simultaneous machine translation–producing a translation while receiving input words–between languages with drastically different word orders: from verb-final languages (e.g., German) to verb-medial languages (English). In traditional machine translation, a translator must "wait" for source material to appear before translation begins. We remove this bottleneck by predicting the final verb in advance. We use reinforcement learning to learn when to trust predictions about unseen, future portions of the sentence. We also introduce an evaluation metric to measure expeditiousness and quality. We show that our new translation model outperforms batch and monotone translation strategies.

PCFG Induction for Unsupervised Parsing and Language Modelling

James Scicluna and Colin de la Higuera   16:55–17:20

The task of unsupervised induction of probabilistic context-free grammars (PCFGs) has attracted a lot of attention in the field of computational linguistics. Although it is a difficult task, work in this area is still very much in demand since it can contribute to the advancement of language parsing and modelling. In this work, we describe a new algorithm for PCFG induction based on a principled approach and capable of inducing accurate yet compact artificial natural language grammars and typical context-free grammars. Moreover, this algorithm can work on large grammars and datasets and infers correctly even from small samples. Our analysis shows that the type of grammars induced by our algorithm are, in theory, capable of modelling natural language. One of our experiments shows that our algorithm can potentially outperform the state-of-the-art in unsupervised parsing on the WSJ10 corpus.

Session 6-posters: Computational Psycholinguistics, Text Mining and NLP Applications

Al Areen 5-6   15:40–17:20

Can characters reveal your native language? A language-independent approach to native language identification
Radu Tudor Ionescu, Marius Popescu, and Aoife Cahill

A common approach in text mining tasks such as text categorization, authorship identification or plagiarism detection is to rely on features like words, part-of-speech tags, stems, or some other high-level linguistic features. In this work, an approach that uses character n-grams as features is proposed for the task of native language identification. Instead of doing standard feature selection, the proposed approach combines several string kernels using multiple kernel learning. Kernel Ridge Regression and Kernel Discriminant Analysis are independently used in the learning stage. The empirical results obtained in all the experiments conducted in this work indicate that the proposed approach achieves state of the art performance in native language identification, reaching an accuracy that is 1.7% above the top scoring system of the 2013 NLI Shared Task. Furthermore, the proposed approach has an important advantage in that it is language independent and linguistic theory neutral. In the cross-corpus experiment, the proposed approach shows that it can also be topic independent, improving the state of the art system by 32.3%.

Formalizing Word Sampling for Vocabulary Prediction as Graph-based Active Learning
Yo Ehara, Yusuke Miyao, Hidekazu Oiwa, Issei Sato, and Hiroshi Nakagawa

Predicting vocabulary of second language learners is essential to support their language learning; however, because of the large size of language vocabularies, we cannot collect information on the entire vocabulary. For practical measurements, we need to sample a small portion of words from the entire vocabulary and predict the rest of the words. In this study, we propose a novel framework for this sampling method. Current methods rely on simple heuristic techniques involving inflexible manual tuning by educational experts. We formalize these heuristic techniques as a graph-based non-interactive active learning method as applied to a special graph. We show that by extending the graph, we can support additional functionality such as incorporating domain specificity and sampling from multiple corpora. In our experiments, we show that our extended methods outperform other methods in terms of vocabulary prediction accuracy when the number of samples is small.

Language Transfer Hypotheses with Linear SVM Weights
Shervin Malmasi and Mark Dras

Language transfer, the characteristic second language usage patterns caused by native language interference, is investigated by Second Language Acquisition (SLA) researchers seeking to find overused and underused linguistic features. In this paper we develop and present a methodology for deriving ranked lists of such features. Using very large learner data, we show our method's ability to find relevant candidates using sophisticated linguistic features. To illustrate its applicability to SLA research, we formulate plausible language transfer hypotheses supported by current evidence. This is the first work to extend Native Language Identification to a broader linguistic interpretation of learner data and address the automatic extraction of underused features on a per-native language basis.

Predicting Dialect Variation in Immigrant Contexts Using Light Verb Constructions
A. Seza Doğruöz and Preslav Nakov

Languages spoken by immigrants change due to contact with the local languages. Capturing these changes is problematic for current language technologies, which are typically developed for speakers of the standard dialect only. Even when dialectal variants are available for such technologies, we still need to predict which dialect is being used. In this study, we distinguish between the immigrant and the standard dialect of Turkish by focusing on Light Verb Constructions. We experiment with a number of grammatical and contextual features, achieving over 84% accuracy (56% baseline).

Device-Dependent Readability for Improved Text Understanding
A-Yeong Kim, Hyun-Je Song, Seong-Bae Park, and Sang-Jo Lee

Readability is used to provide users with high-quality service in text recommendation or text visualization. With the increasing use of hand-held devices, reading device is regarded as an important factor for readability. Therefore, this paper investigates the relationship between readability and reading devices such as a smart phone, a tablet, and paper. We suggest readability factors that are strongly related with the readability of a specific device by showing the correlations between various factors in each device and human-rated readability. Our experimental results show that

each device has its own readability characteristics, and thus different weights should be imposed on readability factors according to the device type. In order to prove the usefulness of the results, we apply the device-dependent readability to news article recommendation.

Predicting Chinese Abbreviations with Minimum Semantic Unit and Global Constraints
Longkai Zhang, Houfeng Wang, and Xu Sun

We propose a new Chinese abbreviation prediction method which can incorporate rich local information while generating the abbreviation globally. Different to previous character tagging methods, we introduce the minimum semantic unit, which is more fine-grained than character but more coarse-grained than word, to capture word level information in the sequence labeling framework. To solve the "character duplication" problem in Chinese abbreviation prediction, we also use a substring tagging strategy to generate local substring tagging candidates. We use an integer linear programming (ILP) formulation with various constraints to globally decode the final abbreviation from the generated candidates. Experiments show that our method outperforms the state-of-art systems, without using any extra resource.

Using Structured Events to Predict Stock Price Movement: An Empirical Investigation
Xiao Ding, Yue Zhang, Ting Liu, and Junwen Duan

It has been shown that news events influence the trends of stock price movements. However, previous work on news-driven stock market prediction rely on shallow features (such as bags-of-words, named entities and noun phrases), which do not capture structured entity-relation information, and hence cannot represent complete and exact events. Recent advances in Open Information Extraction (Open IE) techniques enable the extraction of structured events from web-scale data. We propose to adapt Open IE technology for event-based stock price movement prediction, extracting structured events from large-scale public news without manual efforts. Both linear and nonlinear models are employed to empirically investigate the hidden and complex relationships between events and the stock market. Large-scale experiments show that the accuracy of S&P 500 index prediction is 60%, and that of individual stock prediction can be over 70%. Our event-based system outperforms bags-of-words-based baselines, and previously reported systems trained on S&P 500 stock historical data.

Extracting Clusters of Specialist Terms from Unstructured Text
Aaron Gerow

Automatically identifying related specialist terms is a difficult and important task required to understand the lexical structure of language. This paper develops a corpus-based method of extracting coherent clusters of satellite terminology – terms on the edge of the lexicon – using co-occurrence networks of unstructured text. Term clusters are identified by extracting communities in the co-occurrence graph, after which the largest is discarded and the remaining words are ranked by centrality within a community. The method is tractable on large corpora, requires no document structure and minimal normalization. The results suggest that the model is able to extract coherent groups of satellite terms in corpora with varying size, content and structure. The findings also confirm that language consists of a densely connected core (observed in dictionaries) and systematic, semantically coherent groups of terms at the edges of the lexicon.

Citation-Enhanced Keyphrase Extraction from Research Papers: A Supervised Approach
Cornelia Caragea, Florin Adrian Bulgarov, Andreea Godea, and Sujatha Das Gollapalli

Given the large amounts of online textual documents available these days, e.g., news articles, weblogs, and scientific papers, effective methods for extracting keyphrases, which provide a high-level topic description of a document, are greatly needed. In this paper, we propose a supervised model for keyphrase extraction from research papers that are embedded in citation networks. To this end, we design novel features based on citation network information and use them in conjunction with traditional features for keyphrase extraction to obtain remarkable improvements in performance over strong baselines.

Using Mined Coreference Chains as a Resource for a Semantic Task
Heike Adel and Hinrich Schütze

We propose to use coreference chains extracted from a large corpus as a resource for semantic tasks. We extract three million coreference chains and train word embeddings on them. Then, we compare these embeddings to word vectors derived from raw text data and show that coreference-based word embeddings improve F1 on the task of antonym classification by up to .09.

Financial Keyword Expansion via Continuous Word Vector Representations
Ming-Feng Tsai and Chuan-Ju Wang

This paper proposes to apply the continuous vector representations of words for discovering keywords from a financial sentiment lexicon. In order to capture more keywords, we also incorporate syntactic information into the Continuous Bag-of-Words (CBOW) model. Experimental results on a task of financial risk prediction using the discovered keywords demonstrate that the proposed approach is good at predicting financial risk.

Intrinsic Plagiarism Detection using N-gram Classes
Imene Bensalem, Paolo Rosso, and Salim Chikhi

When it is not possible to compare the suspicious document to the source document(s) plagiarism has been committed from, the evidence of plagiarism has to be looked for intrinsically in the document itself. In this pa-per, we introduce a novel language-independent intrinsic plagiarism detection method which is based on a new text representation that we called n-gram classes. The proposed method was evaluated on three publicly available standard corpora. The obtained results are comparable to the ones obtained by the best state-of-the-art methods.

Verifiably Effective Arabic Dialect Identification
Kareem Darwish, Hassan Sajjad, and Hamdy Mubarak

Several recent papers on Arabic dialect identification have hinted that using a word unigram model is sufficient and effective for the task. However, most previous work was done on a standard fairly homogeneous dataset of dialectal user comments. In this paper, we show that training on the standard dataset does not generalize, because a unigram model may be tuned to topics in the comments and does not capture the distinguishing features of dialects. We show that effective dialect identification requires that we account for the distinguishing lexical, morphological, and phonological phenomena of dialects. We show that accounting for such can improve dialect detection accuracy by nearly 10% absolute.

Keystroke Patterns as Prosody in Digital Writings: A Case Study with Deceptive Reviews and Essays
Ritwik Banerjee, Song Feng, Jun Seok Kang, and Yejin Choi

In this paper, we explore the use of keyboard strokes as a means to access the real-time writing process of online authors, analogously to prosody in speech analysis, in the context of deception detection. We show that differences in keystroke patterns like editing maneuvers and duration of pauses can help distinguish between truthful and deceptive writing. Empirical results show that incorporating keystroke-based features lead to improved performance in deception detection in two different domains: online reviews and essays.

Leveraging Effective Query Modeling Techniques for Speech Recognition and Summarization
Kuan-Yu Chen, Shih-Hung Liu, Berlin Chen, Ea-Ee Jan, Hsin-Min Wang, Wen-Lian Hsu, and Hsin-Hsi Chen

Statistical language modeling (LM) that purports to quantify the acceptability of a given piece of text has long been an interesting yet challenging research area. In particular, language modeling for information retrieval (IR) has enjoyed remarkable empirical success; one emerging stream of the LM approach for IR is to employ the pseudo-relevance feedback process to enhance the representation of an input query so as to improve retrieval effectiveness. This paper presents a continuation of such a general line of research and the main contribution is three-fold. First, we propose a principled framework which can unify the relationships among several widely-used query modeling formulations. Second, on top of the successfully developed framework, we propose an extended query modeling formulation by incorporating critical query-specific information cues to guide the model estimation. Third, we further adopt and formalize such a framework to the speech recognition and summarization tasks. A series of empirical experiments reveal the feasibility of such an LM framework and the performance merits of the deduced models on these two tasks.

Staying on Topic: An Indicator of Power in Political Debates
Vinodkumar Prabhakaran, Ashima Arora, and Owen Rambow

We study the topic dynamics of interactions in political debates using the 2012 Republican presidential primary debates as data. We show that the tendency of candidates to shift topics changes over the course of the election campaign, and that it is correlated with their relative power. We also show that our topic shift features help predict candidates' relative rankings.

6

Main Conference: Tuesday, October 28

Overview

8:00 – 17:00	**Registration**			(Area 3)
8:00 – 9:00	**Refreshments**			(Area 3-4)
9:00 – 10:00	**Best Paper Awards and Presentations**			(Area 3-4)
10:00 – 10:30	**Coffee Break**			(Area 3-4)

Session 7

10:30 – 12:10	Semantics	Information Extraction	Sentiment Analysis and NLP Applications	Posters: Machine Translation and Machine Learning

12:10 – 13:30	**Lunch Break**

Session 8

13:30 – 15:10	Segmentation and Tagging / Spoken Language / Semantics	Sentiment Analysis / Social / Computational Psycholinguistics / Text Classification	Summarization / Machine Translation / Information Extraction	Posters: Information Extraction

15:10 – 15:40	**Coffee Break**			(Area 3-4)

Session 9

15:40 – 17:20	Machine Learning and Machine Translation	NLP for theWeb and Social Media	Semantics	Posters: Discourse, Dialogue and Pragmatics

17:25 – 17:45	**Closing session**			(Al Areen 3)

Parallel Session 7 Overview – Tuesday, October 28, 2014

	Track A *Semantics* Al Areen 1-2	Track B *Information Extraction* Al Areen 3	Track C *Sentiment Analysis and NLP Applications* Al Areen 4
10:30	Sensicon: An Automatically Constructed Sensorial Lexicon *Serra Sinem Tekiroglu, Gözde Özbal, and Carlo Strapparava*	Combining Distant and Partial Supervision for Relation Extraction *Gabor Angeli, Julie Tibshirani, Jean Wu, and Christopher Manning*	Abstractive Summarization of Product Reviews Using Discourse Structure *Shima Gerani, Yashar Mehdad, Giuseppe Carenini, Raymond Ng, and Bita Nejat*
10:55	Word Semantic Representations using Bayesian Probabilistic Tensor Factorization *Jingwei Zhang, Jeremy Salwen, Michael Glass, and Alfio Gliozzo*	Typed Tensor Decomposition of Knowledge Bases for Relation Extraction *Kai-Wei Chang, Wen-tau Yih, Bishan Yang, and Christopher Meek*	Clustering Aspect-related Phrases by Leveraging Sentiment Distribution Consistency *Li Zhao, Minlie Huang, Haiqiang Chen, Junjun Cheng, and Xiaoyan Zhu*
11:20	Glove: Global Vectors for Word Representation *Jeffrey Pennington, Richard Socher, and Christopher Manning*	A convex relaxation for weakly supervised relation extraction *Edouard Grave*	Automatic Generation of Related Work Sections in Scientific Papers: An Optimization Approach *Yue Hu and Xiaojun Wan*
11:45	Jointly Learning Word Representations and Composition Functions Using Predicate-Argument Structures *Kazuma Hashimoto, Pontus Stenetorp, Makoto Miwa, and Yoshimasa Tsuruoka*	Knowledge Graph and Text Jointly Embedding *Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen*	Fast and Accurate Misspelling Correction in Large Corpora *Octavian Popescu and Ngoc Phuoc An Vo*

Track P

Posters: Machine Translation and Machine Learning

Al Areen 5-6   10:30-12:10

- Assessing the Impact of Translation Errors on Machine Translation Quality with Mixed-effects Models
  Marcello Federico, Matteo Negri, Luisa Bentivogli, and Marco Turchi

- Refining Word Segmentation Using a Manually Aligned Corpus for Statistical Machine Translation
  Xiaolin Wang, Masao Utiyama, Andrew Finch, and Eiichiro Sumita

- Improving Pivot-Based Statistical Machine Translation by Pivoting the Co-occurrence Count of Phrase Pairs
  Xiaoning Zhu, Zhongjun He, Hua Wu, Conghui Zhu, Haifeng Wang, and Tiejun Zhao

- Word Translation Prediction for Morphologically Rich Languages with Bilingual Neural Networks
  Ke M. Tran, Arianna Bisazza, and Christof Monz

- Dependency-Based Bilingual Language Models for Reordering in Statistical Machine Translation
  Ekaterina Garmash and Christof Monz

- Combining String and Context Similarity for Bilingual Term Alignment from Comparable Corpora
  Georgios Kontonatsios, Ioannis Korkontzelos, Jun'ichi Tsujii, and Sophia Ananiadou

- Random Manhattan Integer Indexing: Incremental L1 Normed Vector Space Construction
  Behrang Zadeh and Siegfried Handschuh

- Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation
  Kyunghyun Cho, Bart van Merrienboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio

- Type-based MCMC for Sampling Tree Fragments from Forests
  Xiaochang Peng and Daniel Gildea

- Convolutional Neural Networks for Sentence Classification
  Yoon Kim

- Sometimes Average is Best: The Importance of Averaging for Prediction using MCMC Inference in Topic Modeling
  Viet-An Nguyen, Jordan Boyd-Graber, and Philip Resnik

- Large-scale Reordering Model for Statistical Machine Translation using Dual Multinomial Logistic Regression
  Abdullah Alrajeh and Mahesan Niranjan

- [TACL] Dynamic Language Models for Streaming Text
  Dani Yogatama, Chong Wang, Bryan Routledge, Noah A. Smith, and Eric P. Xing

- Improved Decipherment of Homophonic Ciphers
  Malte Nuhn, Julian Schamper, and Hermann Ney

- Cipher Type Detection
  Malte Nuhn and Kevin Knight

Parallel Session 7

Session 7a: Semantics
Al Areen 1-2   *Chair: Rebecca Passonneau*

Sensicon: An Automatically Constructed Sensorial Lexicon
Serra Sinem Tekiroglu, Gözde Özbal, and Carlo Strapparava   10:30–10:55

Connecting words with senses, namely, sight, hearing, taste, smell and touch, to comprehend the sensorial information in language is a straightforward task for humans by using commonsense knowledge. With this in mind, a lexicon associating words with senses would be crucial for the computational tasks aiming at interpretation of language. However, to the best of our knowledge, there is no systematic attempt in the literature to build such a resource. In this paper, we present a sensorial lexicon that associates English words with senses. To obtain this resource, we apply a computational method based on bootstrapping and corpus statistics. The quality of the resulting lexicon is evaluated with a gold standard created via crowdsourcing. The results show that a simple classifier relying on the lexicon outperforms two baselines on a sensory classification task, both at word and sentence level, and confirm the soundness of the proposed approach for the construction of the lexicon and the usefulness of the resource for computational applications.

Word Semantic Representations using Bayesian Probabilistic Tensor Factorization
Jingwei Zhang, Jeremy Salwen, Michael Glass, and Alfio Gliozzo   10:55–11:20

Many forms of word relatedness have been developed, providing different perspectives on word similarity. We introduce a Bayesian probabilistic tensor factorization model for synthesizing a single word vector representation and per-perspective linear transformations from any number of word similarity matrices. The result- ing word vectors, when combined with the per-perspective linear transformation, approximately recreate while also regularizing and generalizing, each word similarity perspective. Our method can combine manually created semantic resources with neural word embeddings to separate synonyms and antonyms, and is capable of generalizing to words outside the vocabulary of any particular perspective. We evaluated the word embeddings with GRE antonym questions, the result achieves the state-of- the-art performance.

Glove: Global Vectors for Word Representation
Jeffrey Pennington, Richard Socher, and Christopher Manning   11:20–11:45

Recent methods for learning vector space representations of words have succeeded in capturing fine-grained semantic and syntactic regularities using vector arithmetic, but the origin of these regularities has remained opaque. We analyze and make explicit the model properties needed for such regularities to emerge in word vectors. The result is a new global log-bilinear regression model that combines the advantages of the two major model families in the literature: global matrix factorization and local context window methods. Our model efficiently leverages statistical information by training only on the nonzero elements in a word-word co-occurrence matrix, rather than on the entire sparse matrix or on individual context windows in a large corpus. The model produces a vector space with meaningful substructure, as evidenced by its performance of 75% on a recent word analogy task. It also outperforms related models on similarity tasks and named entity recognition.

Jointly Learning Word Representations and Composition Functions Using Predicate-Argument Structures
Kazuma Hashimoto, Pontus Stenetorp, Makoto Miwa, and Yoshimasa Tsuruoka   11:45–12:10

We introduce a novel compositional language model that works on Predicate-Argument Structures (PASs). Our model jointly learns word representations and their composition functions using bag-of-words and dependency-based contexts. Unlike previous word-sequence-based models, our PAS-based model composes arguments into predicates by using the category information from the PAS. This enables our model to capture long-range dependencies between words and to better handle constructs such as verb-object and subject-verb-object relations. We verify this experimentally using two phrase similarity datasets and achieve results comparable to or higher than the previous best results. Our system achieves these results without the need for pre-trained word vectors and using a much smaller training corpus; despite this, for the subject-verb-object dataset our model improves upon the state of the art by as much as ~10% in relative performance.

Session 7b: Information Extraction
Al Areen 3 *Chair: Ruhi Sarikaya*

Combining Distant and Partial Supervision for Relation Extraction
Gabor Angeli, Julie Tibshirani, Jean Wu, and Christopher Manning 10:30–10:55

Broad-coverage relation extraction either requires expensive supervised training data, or suffers from drawbacks inherent to distant supervision. We present an approach for providing partial supervision to a distantly supervised relation extractor using a small number of carefully selected examples. We compare against established active learning criteria and propose a novel criterion to sample examples which are both uncertain and representative. In this way, we combine the benefits of fine-grained supervision for difficult examples with the coverage of a large distantly supervised corpus. Our approach gives a substantial increase of 3.9% end-to-end F1 on the 2013 KBP Slot Filling evaluation, yielding a net F1 of 37.7%.

Typed Tensor Decomposition of Knowledge Bases for Relation Extraction
Kai-Wei Chang, Wen-tau Yih, Bishan Yang, and Christopher Meek 10:55–11:20

While relation extraction has traditionally been viewed as a task relying solely on textual data, recent work has shown that by taking as input existing facts in the form of entity-relation triples from both knowledge bases and textual data, the performance of relation extraction can be improved significantly. Following this new paradigm, we propose a tensor decomposition approach for knowledge base embedding that is highly scalable, and is especially suitable for relation extraction. By leveraging relational domain knowledge about entity type information, our learning algorithm is significantly faster than previous approaches and is better able to discover new relations missing from the database. In addition, when applied to a relation extraction task, our approach alone is comparable to several existing systems, and improves the weighted mean average precision of a state-of-the-art method by 10 points when used as a subcomponent.

A convex relaxation for weakly supervised relation extraction
Edouard Grave 11:20–11:45

A promising approach to relation extraction, called weak or distant supervision, exploits an existing database of facts as training data, by aligning it to an unlabeled collection of text documents. Using this approach, the task of relation extraction can easily be scaled to hundreds of different relationships. However, distant supervision leads to a challenging multiple instance, multiple label learning problem. Most of the proposed solutions to this problem are based on non-convex formulations, and are thus prone to local minima. In this article, we propose a new approach to the problem of weakly supervised relation extraction, based on discriminative clustering and leading to a convex formulation. We demonstrate that our approach outperforms state-of-the-art methods on the challenging dataset introduced by Riedel et al. (2010).

Knowledge Graph and Text Jointly Embedding
Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen 11:45–12:10

We examine the embedding approach to reason new relational facts from a large-scale knowledge graph and a text corpus. We propose a novel method of jointly embedding entities and words into the same continuous vector space. The embedding process attempts to preserve the relations between entities in the knowledge graph and the concurrences of words in the text corpus. Entity names and Wikipedia anchors are utilized to align the embeddings of entities and words in the same space. Large scale experiments on Freebase and a Wikipedia/NY Times corpus show that jointly embedding brings promising improvement in the accuracy of predicting facts, compared to separately embedding knowledge graphs and text. Particularly, jointly embedding enables the prediction of facts containing entities out of the knowledge graph, which cannot be handled by previous embedding methods. At the same time, concerning the quality of the word embeddings, experiments on the analogical reasoning task show that jointly embedding is comparable to or slightly better than word2vec(Skip-Gram).

Session 7c: Sentiment Analysis and NLP Applications
Al Areen 4   *Chair: Vincent Ng*

Abstractive Summarization of Product Reviews Using Discourse Structure
Shima Gerani, Yashar Mehdad, Giuseppe Carenini, Raymond Ng, and Bita Nejat   10:30–10:55

We propose a novel abstractive summarization system for product reviews by taking advantage of their discourse structure. First, we apply a discourse parser to each review and obtain a discourse tree representation for every review. We then modify the discourse trees such that every leaf node only contains the aspect words. Second, we aggregate the aspect discourse trees and generate a graph. We then select a subgraph representing the most important aspects and the rhetorical relations between them using a PageRank algorithm, and transform the selected subgraph into an aspect tree. Finally, we generate a natural language summary by applying a template-based NLG framework which takes advantage of rhetorical relations. Quantitative and qualitative analysis of the results, based on two user studies, show that our approach significantly outperforms extractive and abstractive baselines. .

Clustering Aspect-related Phrases by Leveraging Sentiment Distribution Consistency
Li Zhao, Minlie Huang, Haiqiang Chen, Junjun Cheng, and Xiaoyan Zhu   10:55–11:20

Clustering aspect-related phrases in terms of product's property is a precursor process to aspect-level sentiment analysis which is a central task in sentiment analysis. Most of existing methods for addressing this problem are context-based models which assume that domain synonymous phrases share similar co-occurrence contexts. In this paper, we explore a novel idea, sentiment distribution consistency, which states that different phrases (e.g."price", "money", "worth", and "cost") of the same aspect tend to have consistent sentiment distribution. Through formalizing sentiment distribution consistency as soft constraint, we propose a novel unsupervised model in the framework of Posterior Regularization (PR) to cluster aspect-related phrases. Experiments demonstrate that our approach outperforms baselines remarkably. Keywords: Aspect-related Phrases Clustering, Text Mining, Natural Language Processing Applications

Automatic Generation of Related Work Sections in Scientific Papers: An Optimization Approach
Yue Hu and Xiaojun Wan   11:20–11:45

In this paper, we investigate a challenging task of automatic related work generation. Given multiple reference papers as input, the task aims to generate a related work section for a target paper. The generated related work section can be used as a draft for the author to complete his or her final related work section. We propose our Automatic Related Work Generation system called ARWG to address this task. It first exploits a PLSA model to split the sentence set of the given papers into different topic-biased parts, and then applies regression models to learn the importance of the sentences. At last it employs an optimization framework to generate the related work section. Our evaluation results on a test set of 150 target papers along with their reference papers show that our proposed ARWG system can generate related work sections with better quality. A user study is also performed to show ARWG can achieve an improvement over generic multi-document summarization baselines.

Fast and Accurate Misspelling Correction in Large Corpora
Octavian Popescu and Ngoc Phuoc An Vo   11:45–12:10

There are several NLP systems whose ac- curacy depends crucially on finding misspellings fast. However, the classical approach is based on a quadratic time algorithm with 80% coverage. We present a novel algorithm for misspelling detection, which runs in constant time and improves the coverage to more than 96%. We use this algorithm together with a cross document coreference system in order to find proper name misspellings. The experiments confirmed significant improvement over the state of the art.

Session 7-posters: Machine Translation and Machine Learning
Al Areen 5-6   10:30–12:10

Assessing the Impact of Translation Errors on Machine Translation Quality with Mixed-effects Models
Marcello Federico, Matteo Negri, Luisa Bentivogli, and Marco Turchi

Learning from errors is a crucial aspect of improving expertise. Based on this notion, we discuss a robust statistical framework for analysing the impact of different error types on machine translation (MT) output quality. Our approach is based on linear mixed-effects models, which allow the analysis of error-annotated MT output taking into account the variability inherent to the specific experimental setting from which the empirical observations are drawn. Our experiments are carried out on different language pairs involving Chinese, Arabic and Russian as target languages. Interesting findings are reported, concerning the impact of different error types both at the level of human perception of quality and with respect to performance results measured with automatic metrics.

Refining Word Segmentation Using a Manually Aligned Corpus for Statistical Machine Translation
Xiaolin Wang, Masao Utiyama, Andrew Finch, and Eiichiro Sumita

Languages that have no explicit word delimiters often have to be segmented for statistical machine translation (SMT). This is commonly performed by automated segmenters trained on manually annotated corpora. However, the word segmentation (WS) schemes of these annotated corpora are handcrafted for general usage, and may not be suitable for SMT. An analysis was performed to test this hypothesis using a manually annotated word alignment (WA) corpus for Chinese-English SMT. An analysis revealed that 74.60% of the sentences in the WA corpus if segmented using an automated segmenter trained on the Penn Chinese Treebank (CTB) will contain conflicts with the gold WA annotations. We formulated an approach based on word splitting with reference to the annotated WA to alleviate these conflicts. Experimental results show that the refined WS reduced word alignment error rate by 6.82

Improving Pivot-Based Statistical Machine Translation by Pivoting the Co-occurrence Count of Phrase Pairs
Xiaoning Zhu, Zhongjun He, Hua Wu, Conghui Zhu, Haifeng Wang, and Tiejun Zhao

To overcome the scarceness of bilingual corpora for some language pairs in machine translation, pivot-based SMT uses pivot language as a "bridge" to generate source-target translation from source-pivot and pivot-target translation. One of the key issues is to estimate the probabilities for the generated phrase pairs. In this paper, we present a novel approach to calculate the translation probability by pivoting the co-occurrence count of source-pivot and pivot-target phrase pairs. Experimental results on Europarl data and web data show that our method leads to significant improvements over the baseline systems.

Word Translation Prediction for Morphologically Rich Languages with Bilingual Neural Networks
Ke M. Tran, Arianna Bisazza, and Christof Monz

Translating into morphologically rich languages is a particularly difficult problem in machine translation due to the high degree of inflectional ambiguity in the target language, often only poorly captured by existing word translation models. We present a general approach that exploits source-side contexts of foreign words to improve translation prediction accuracy. Our approach is based on a probabilistic neural network which does not require linguistic annotation nor manual feature engineering. We report significant improvements in word translation prediction accuracy for three morphologically rich target languages. In addition, preliminary results for integrating our approach into our existing English-Russian statistical machine translation system show small but statistically significant improvements in translation quality.

Dependency-Based Bilingual Language Models for Reordering in Statistical Machine Translation
Ekaterina Garmash and Christof Monz

This paper presents a novel approach to improve reordering in phrase-based machine translation by using richer, syntactic representations of units of bilingual language models (BiLMs). Our method to include syntactic information is simple in implementation and requires minimal changes in the decoding algorithm. The approach is evaluated in a series of Arabic-English and Chinese-English translation experiments. The best models demonstrate significant improvements in BLEU and TER over the phrase-based baseline, as well as over the lexicalized BiLM by Niehues et al.

(2011). Further improvements of up to 0.45 BLEU for Arabic-English and up to 0.59 BLEU for Chinese English are obtained by combining our dependency BiLM with a lexicalized BiLM. An improvement of 0.98 BLEU is obtained for Chinese-English in the setting of an increased distortion limit.

Combining String and Context Similarity for Bilingual Term Alignment from Comparable Corpora
Georgios Kontonatsios, Ioannis Korkontzelos, Jun'ichi Tsujii, and Sophia Ananiadou

Automatically compiling bilingual dictionaries of technical terms from comparable corpora is a challenging problem, yet with many potential applications. In this paper, we exploit two independent observations about term translations: (a) terms are often formed by corresponding sub-lexical units across languages and (b) a term and its translation tend to appear in similar lexical context. Based on the first observation, we develop a new character n-gram compositional method, a logistic regression classifier, for learning a string similarity measure of term translations. According to the second observation, we use an existing context-based approach. For evaluation, we investigate the performance of compositional and context-based methods on: (a) similar and unrelated languages, (b) corpora of different degree of comparability and (c) the translation of frequent and rare terms. Finally, we combine the two translation clues, namely string and contextual similarity, in a linear model and we show substantial improvements over the two translation signals.

Random Manhattan Integer Indexing: Incremental L1 Normed Vector Space Construction
Behrang Zadeh and Siegfried Handschuh

Vector space models (VSMs) are mathematically well-defined frameworks that have been widely used in the distributional approaches to semantics. In VSMs, high-dimensional vectors represent linguistic entities. In an application, the similarity of vectors—and thus the entities that they represent—is computed by a distance formula. The high dimensionality of vectors, however, is a barrier to the performance of methods that employ VSMs. Consequently, a dimensionality reduction technique is employed to alleviate this problem. This paper introduces a novel technique called Random Manhattan Indexing (RMI) for the construction of l1 normed VSMs at reduced dimensionality. RMI combines the construction of a VSM and dimension reduction into an incremental and thus scalable two-step procedure. In order to attain its goal, RMI employs the sparse Cauchy random projections. We further introduce Random Manhattan Integer Indexing (RMII): a computationally enhanced version of RMI. As shown in the reported experiments, RMI and RMII can be used reliably to estimate the l1 distances between vectors in a vector space of low dimensionality.

Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation
Kyunghyun Cho, Bart van Merrienboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio

In this paper, we propose a novel neural network model called RNN Encoder–Decoder that consists of two recurrent neural networks (RNN). One RNN encodes a sequence of symbols into a fixed-length vector representation, and the other decodes the representation into another sequence of symbols. The encoder and decoder of the proposed model are jointly trained to maximize the conditional probability of a target sequence given a source sequence. The performance of a statistical machine translation system is empirically found to improve by using the conditional probabilities of phrase pairs computed by the RNN Encoder–Decoder as an additional feature in the existing log-linear model. Qualitatively, we show that the proposed model learns a semantically and syntactically meaningful representation of linguistic phrases.

Type-based MCMC for Sampling Tree Fragments from Forests
Xiaochang Peng and Daniel Gildea

This paper applies type-based Markov Chain Monte Carlo (MCMC) algorithms to the problem of learning Synchronous Context-Free Grammar (SCFG) rules from a forest that represents all possible rules consistent with a fixed word alignment. While type-based MCMC has been shown to be effective in a number of NLP applications, our setting, where the tree structure of the sentence is itself a hidden variable, presents a number of challenges to type-based inference. We describe methods for defining variable types and efficiently indexing variables in order to overcome these challenges. These methods lead to improvements in both log likelihood and BLEU score in our experiments.

Convolutional Neural Networks for Sentence Classification
Yoon Kim

We report on a series of experiments with convolutional neural networks (CNN) trained on top of pre-trained word vectors for sentence-level classification tasks. We show that a simple CNN with little hyperparameter tuning and static vectors achieves excellent results on multiple benchmarks. Learning task-specific vectors through fine-tuning offers further gains in performance. We additionally propose a simple modification to the architecture to allow for the use of both task-specific and static word vectors. The CNN models discussed herein improve upon the state of the art on 4 out of 7 tasks, which include sentiment analysis and question classification.

Sometimes Average is Best: The Importance of Averaging for Prediction using MCMC Inference in Topic Modeling
Viet-An Nguyen, Jordan Boyd-Graber, and Philip Resnik

Markov chain Monte Carlo (MCMC) approximates the posterior distribution of latent variable models by generating many samples and averaging over them. In practice, however, it is often more convenient to cut corners, using only a single sample or following a suboptimal averaging strategy. We systematically study different strategies for averaging MCMC samples and show empirically that averaging properly leads to significant improvements in prediction.

Large-scale Reordering Model for Statistical Machine Translation using Dual Multinomial Logistic Regression
Abdullah Alrajeh and Mahesan Niranjan

Phrase reordering is a challenge for statistical machine translation systems. Posing phrase movements as a prediction problem using contextual features modeled by maximum entropy based classifier is superior to the commonly used lexicalized reordering model. However, Training this discriminative model using large-scale parallel corpus might be computationally expensive. In this paper, we explore recent advancements in solving large scale classification problems. Using the dual problem to multinomial logistic regression, we managed to shrink the training data while iterating and produce significant saving in computation and memory while preserving the accuracy.

[TACL] Dynamic Language Models for Streaming Text
Dani Yogatama, Chong Wang, Bryan Routledge, Noah A. Smith, and Eric P. Xing

We present a probabilistic language model that captures temporal dynamics and conditions on arbitrary non-linguistic context features. These context features serve as important indicators of language changes that are otherwise difficult to capture using text data by itself. We learn our model in an efficient online fashion that is scalable for large, streaming data. With five streaming datasets from two different genres— economics news articles and social media—we evaluate our model on the task of sequential language modeling. Our model consistently outperforms competing models.

Improved Decipherment of Homophonic Ciphers
Malte Nuhn, Julian Schamper, and Hermann Ney

In this paper, we present two improvements to the beam search approach for solving homophonic substitution ciphers presented in Nuhn et al. (2013): An improved rest cost estimation together with an optimized strategy for obtaining the order in which the symbols of the cipher are deciphered reduces the beam size needed to successfully decipher the Zodiac-408 cipher from several million down to less than one hundred: The search effort is reduced from several hours of computation time to just a few seconds on a single CPU. These improvements allow us to successfully decipher the second part of the famous Beale cipher (see (Ward et al., 1885) and e.g. (King, 1993)): Having 182 different cipher symbols while having a length of just 762 symbols, the decipherment is way more challenging than the decipherment of the previously deciphered Zodiac-408 cipher (length 408, 54 different symbols). To the best of our knowledge, this cipher has not been deciphered automatically before.

Cipher Type Detection
Malte Nuhn and Kevin Knight

Manual analysis and decryption of enciphered documents is a tedious and error prone work. Often—even after spending large amounts of time on a particular cipher—no decipherment can be found. Automating the decryption of various types of ciphers makes it possible to sift through the large number of encrypted messages found in libraries and archives, and to focus human effort only on a small but potentially interesting subset of them. In this work, we train a classifier that is able to predict which encipherment method has been used to generate a given ciphertext. We are able to distinguish 50 different cipher types (specified by the American Cryptogram Association) with an accuracy of 58.5%. This is a 11.2% absolute improvement over the best previously published classifier.

Parallel Session 8 (Short Papers) Overview – Tuesday, October 28, 2014

	Track A	Track B	Track C
	Segmentation and Tagging / Spoken Language / Semantics	*Sentiment Analysis / Social / Computational Psycholinguistics / Text Classification*	*Summarization / Machine Translation / Information*
	Al Areen 1-2	Al Areen 3	Al Areen 4
13:30	Joint Learning of Chinese Words, Terms and Keywords *Ziqiang Cao, Sujian Li, and Heng Ji*	Joint Emotion Analysis via Multi-task Gaussian Processes *Daniel Beck, Trevor Cohn, and Lucia Specia*	Joint Decoding of Tree Transduction Models for Sentence Compression *Jin-ge Yao, Xiaojun Wan, and Jianguo Xiao*
13:50	Cross-Lingual Part-of-Speech Tagging through Ambiguous Learning *Guillaume Wisniewski, Nicolas Pécheux, Souhir Gahbiche-Braham, and François Yvon*	Detecting Latent Ideology in Expert Text: Evidence From Academic Papers in Economics *Zubin Jelveh, Bruce Kogut, and Suresh Naidu*	Dependency-based Discourse Parser for Single-Document Summarization *Yasuhisa Yoshida, Jun Suzuki, Tsutomu Hirao, and Masaaki Nagata*
14:10	Comparing Representations of Semantic Roles for String-To-Tree Decoding *Marzieh Bazrafshan and Daniel Gildea*	A Model of Individual Differences in Gaze Control During Reading *Niels Landwehr, Sebastian Arzt, Tobias Scheffer, and Reinhold Kliegl*	Improving Word Alignment using Word Similarity *Theerawat Songyot and David Chiang*
14:30	Detecting Non-compositional MWE Components using Wiktionary *Bahar Salehi, Paul Cook, and Timothy Baldwin*	Muli-label Text Categorization with Hidden Components *Longkai Zhang and Houfeng Wang*	Constructing Information Networks Using One Single Model *Qi Li, Heng Ji, Yu Hong, and Sujian Li*
14:50		#TagSpace: Semantic Embeddings from Hashtags *Jason Weston, Sumit Chopra, and Keith Adams*	Event Role Extraction using Domain-Relevant Word Representations *Emanuela Boros, Romaric Besançon, Olivier Ferret, and Brigitte Grau*

Track P

Posters: Information Extraction

Al Areen 5-6   13:30-15:10

- Modeling Joint Entity and Relation Extraction with Table Representation
  Makoto Miwa and Yutaka Sasaki

- ZORE: A Syntax-based System for Chinese Open Relation Extraction
  Likun Qiu and Yue Zhang

- Coarse-grained Candidate Generation and Fine-grained Re-ranking for Chinese Abbreviation Prediction

Parallel Session 8 (Short Papers)

Session 8sa: Segmentation and Tagging / Spoken Language / Semantics
Al Areen 1-2   *Chair: Lluís Màrquez*

Joint Learning of Chinese Words, Terms and Keywords
Ziqiang Cao, Sujian Li, and Heng Ji   13:30–13:50

Previous work often used a pipelined framework where Chinese word segmentation is followed by term extraction and keyword extraction. Such framework suffers from error propagation and is unable to leverage information in later modules for prior components. In this paper, we propose a four-level Dirichlet Process based model (named DP-4) to jointly learn the word distributions occurring in the corpus, domain and document levels simultaneously. Based on the DP-4 model, a sentence-wise Gibbs sampler is adopted to obtain the proper segmentation results. Meanwhile, terms and keywords are acquired in the sampling process. The experimental results have shown the effectiveness of our method.

Cross-Lingual Part-of-Speech Tagging through Ambiguous Learning
Guillaume Wisniewski, Nicolas Pécheux, Souhir Gahbiche-Braham, and François Yvon   13:50–14:10

When Part-of-Speech annotated data is scarce, e.g. for under-resourced languages, one can turn to cross-lingual transfer and crawled dictionaries to collect partially supervised data. We cast this problem in the framework of ambiguous learning and show how to learn an accurate history-based model. Experiments on ten languages show significant improvements over prior state of the art performance.

Comparing Representations of Semantic Roles for String-To-Tree Decoding
Marzieh Bazrafshan and Daniel Gildea   14:10–14:30

We introduce new features for incorporating semantic predicate-argument structures in machine translation (MT). The methods focus on the completeness of the semantic structures of the translations, as well as the order of the translated semantic roles. We experiment with translation rules which contain the core arguments for the predicates in the source side of a MT system, and observe that using these rules significantly improves the translation quality. We also present a new semantic feature that resembles a language model. Our results show that the language model feature can also significantly improve MT results.

Detecting Non-compositional MWE Components using Wiktionary
Bahar Salehi, Paul Cook, and Timothy Baldwin   14:30–14:50

We propose a simple unsupervised approach to detecting non-compositional components in multiword expressions based on Wiktionary. The approach makes use of the definitions, synonyms and translations in Wiktionary, and is applicable to any type of MWE in any language, assuming the MWE is contained in Wiktionary. Our experiments show that the proposed approach achieves higher F-score than state-of-the-art methods.

Session 8sb: Sentiment Analysis / Social / Computational Psycholinguistics / Text Classification

Al Areen 3   *Chair: Paolo Rosso*

Joint Emotion Analysis via Multi-task Gaussian Processes
Daniel Beck, Trevor Cohn, and Lucia Specia   13:30–13:50

In this paper we propose a model for jointly predict the emotions of natural language sentences. Our model is based on a low-rank coregionalization approach, which combines a vector-valued Gaussian Process with a rich parameterization scheme. We show that our approach is able to learn correlations and anti-correlations between emotions on a news headlines dataset. The proposed model also outperforms both single-task baselines and other multi-task approaches.

Detecting Latent Ideology in Expert Text: Evidence From Academic Papers in Economics
Zubin Jelveh, Bruce Kogut, and Suresh Naidu   13:50–14:10

Previous work on extracting ideology from text has focused on domains where expression of political views is expected, but it's unclear if current technology can work in domains where displays of ideology are considered inappropriate. We present a supervised ensemble n-gram model for ideology extraction with topic adjustments and apply it to one such domain: research papers written by academic economists. We show economists' political leanings can be correctly predicted, that our predictions generalize to new domains, and that they correlate with public policy-relevant research findings. We also present evidence that unsupervised models can under-perform in domains where ideological expression is discouraged.

A Model of Individual Differences in Gaze Control During Reading
Niels Landwehr, Sebastian Arzt, Tobias Scheffer, and Reinhold Kliegl   14:10–14:30

We develop a statistical model of saccadic eye movements during reading of isolated sentences. The model is focused on representing individual differences between readers and supports the inference of the most likely reader for a novel set of eye movement patterns. We empirically study the model for biometric reader identification using eye-tracking data collected from 20 individuals and observe that the model distinguishes between 20 readers with an accuracy of up to 98%.

Muli-label Text Categorization with Hidden Components
Longkai Zhang and Houfeng Wang   14:30–14:50

Multi-label text categorization (MTC) is supervised learning, where documents may be assigned with multiple categories (labels) simultaneously. The labels in the MTC are correlated. The labels with correlation result in some hidden components, which represent the "share" variance of correlated labels. In this paper, we propose multi-label text categorization method with hidden components. The proposed method employs PCA to obtain the hidden components, and incorporates them into a joint learning framework to improve the performance. Experiments with various data sets and evaluation metrics validate the effectiveness of the proposed method.

#TagSpace: Semantic Embeddings from Hashtags
Jason Weston, Sumit Chopra, and Keith Adams   14:50–15:10

We describe a convolutional neural network that learns feature representations for short textual posts using hashtags as a supervised signal. The proposed approach is trained on up to 5.5 billion words predicting 100,000 possible hashtags. As well as strong performance on the hashtag prediction task itself, we show that its learned representation of text (ignoring the hashtag labels) is useful for other tasks as well. To that end, we present results on a document recommendation task, where it also outperforms a number of baselines.

Session 8sc: Summarization / Machine Translation / Information Extraction
Al Areen 4   *Chair: Colin de la Higuera*

Joint Decoding of Tree Transduction Models for Sentence Compression
Jin-ge Yao, Xiaojun Wan, and Jianguo Xiao   13:30–13:50

In this paper, we provide a new method for decoding tree transduction based sentence compression models augmented with language model scores, by jointly decoding two components. In our proposed solution, rich local discriminative features can be easily integrated without increasing computational complexity. Utilizing an unobvious fact that the resulted two components can be independently decoded, we conduct efficient joint decoding based on dual decomposition. Experimental results show that our method outperforms traditional beam search decoding and achieves the state-of-the-art performance.

Dependency-based Discourse Parser for Single-Document Summarization
Yasuhisa Yoshida, Jun Suzuki, Tsutomu Hirao, and Masaaki Nagata   13:50–14:10

The current state-of-the-art single-document summarization method generates a summary by solving a Tree Knapsack Problem (TKP), which is the problem of finding the optimal rooted subtree of the dependency-based discourse tree (DEP-DT) of a document. We can obtain a gold DEP-DT by transforming a gold Rhetorical Structure Theory-based discourse tree (RST-DT). However, there is still a large difference between the ROUGE scores of a system with a gold DEP-DT and a system with a DEP-DT obtained from an automatically parsed RST-DT. To improve the ROUGE score, we propose a novel discourse parser that directly generates the DEP-DT. The evaluation results showed that the TKP with our parser outperformed that with the state-of-the-art RST-DT parser, and achieved almost equivalent ROUGE scores to the TKP with the gold DEP-DT.

Improving Word Alignment using Word Similarity
Theerawat Songyot and David Chiang   14:10–14:30

We show that semantic relationships can be used to improve word alignment, in addition to the lexical and syntactic features that are typically used. In this paper, we present a method based on a neural network to automatically derive word similarity from monolingual data. We present an extension to word alignment models that exploits word similarity. Our experiments, in both large-scale and, particularly, resource-limited settings, have shown improvement in word alignment tasks as well as translation tasks.

Constructing Information Networks Using One Single Model
Qi Li, Heng Ji, Yu Hong, and Sujian Li   14:30–14:50

In this paper, we propose a new framework that unifies the output of three information extraction (IE) tasks - entity mentions, relations and events as an information network representation, and extracts all of them using one single joint model based on structured prediction. This novel formulation allows different parts of the information network fully interact with each other. For example, many relations can now be considered as the resultant states of events. Our approach achieves substantial improvements over traditional pipelined approaches, and significantly advances state-of-the-art end-to-end event argument extraction.

Event Role Extraction using Domain-Relevant Word Representations
Emanuela Boros, Romaric Besançon, Olivier Ferret, and Brigitte Grau   14:50–15:10

The efficiency of Information Extraction systems is known to be heavily influenced by domain-specific knowledge but the cost of developing such systems is considerably high. In this paper, we consider the problem of event extraction and show that learning word representations from unlabeled domain-specific data and using them for representing event roles enable to outperform previous state-of-the-art event extraction models on the MUC-4 data set.

Session 8-posters: Information Extraction
Al Areen 5-6   13:30–15:10

Modeling Joint Entity and Relation Extraction with Table Representation
Makoto Miwa and Yutaka Sasaki

This paper proposes a history-based structured learning approach that jointly extracts entities and relations in a sentence. We introduce a novel simple and flexible table representation of entities and relations. We investigate several feature settings, search orders, and learning methods with inexact search on the table. The experimental results demonstrate that a joint learning approach significantly outperforms a pipeline approach by incorporating global features and by selecting appropriate learning methods and search orders.

ZORE: A Syntax-based System for Chinese Open Relation Extraction
Likun Qiu and Yue Zhang

Open Relation Extraction (ORE) overcomes the limitations of traditional IE techniques, which train individual extractors for every single relation type. Systems such as ReVerb, PATTY, OLLIE, and Exemplar have attracted much attention on English ORE. However, few studies have been reported on ORE for languages beyond English. This paper presents a syntax-based Chinese (Zh) ORE system, ZORE, for extracting relations and semantic patterns from Chinese text. ZORE identifies relation candidates from automatically parsed dependency trees, and then extracts relations with their semantic patterns iteratively through a novel double propagation algorithm. Empirical results on two data sets show the effectiveness of the proposed system.

Coarse-grained Candidate Generation and Fine-grained Re-ranking for Chinese Abbreviation Prediction
Longkai Zhang, Houfeng Wang, and Xu Sun

Correctly predicting abbreviations given the full forms is important in many natural language processing systems. In this paper we propose a two-stage method to find the corresponding abbreviation given its full form. We first use the contextual information given a large corpus to get abbreviation candidates for each full form and get a coarse-grained ranking through graph random walk. This coarse-grained rank list fixes the search space inside the top-ranked candidates. Then we use a similarity sensitive re-ranking strategy which can utilize the features of the candidates to give a fine-grained re-rank and select the final result. Our method achieves good results and outperforms the state-of-art systems. One advantage of our method is that it only needs weak supervision and can get competitive results with fewer training data. The candidate generation and coarse-grained ranking is totally unsupervised. The re-ranking phase can use a very small amount of training data to get a reasonably good result.

Type-Aware Distantly Supervised Relation Extraction with Linked Arguments
Mitchell Koch, John Gilmer, Stephen Soderland, and Daniel S. Weld

Distant supervision has become the leading method for training large-scale relation extractors, with nearly universal adoption in recent TAC knowledge-base population competitions. However, there are still many questions about the best way to learn such extractors. In this paper we investigate four orthogonal improvements: integrating named entity linking (NEL) and coreference resolution into argument identification for training and extraction, enforcing type constraints of linked arguments, and partitioning the model by relation type signature. We evaluate sentential extraction performance on two datasets: the popular set of NY Times articles partially annotated by Hoffmann et al. (2011) and a new dataset, called GoReCo, that is comprehensively annotated for 48 common relations. We find that using NEL for argument identification boosts performance over the traditional approach (named entity recognition with string match), and there is further improvement from using argument types. Our best system boosts precision by 44% and recall by 70%.

Automatic Inference of the Tense of Chinese Events Using Implicit Linguistic Information
Yuchen Zhang and Nianwen Xue

We address the problem of automatically inferring the tense of events in Chinese text. We use a new corpus annotated with Chinese semantic tense information and other implicit Chinese linguistic information using a "distant annotation" method. We propose three improvements over a relatively strong baseline method – a statistical learning method with extensive feature engineering. First, we add two sources of implicit linguistic information as features – eventuality type and modality of an event, which are also inferred automatically. Second, we perform joint learning on semantic tense, eventuality type, and modality of an event. Third, we train artificial neural network models for this problem and compare its performance with feature-based approaches. Experimental results show considerable improvements

on Chinese tense inference. Our best performance reaches 68.6

Joint Inference for Knowledge Base Population
Liwei Chen, Yansong Feng, Jinghui Mo, Songfang Huang, and Dongyan Zhao

Populating Knowledge Base (KB) with new knowledge facts from reliable text resources usually consists of linking name mentions to KB entities and identifying relationship between entity pairs. However, the task often suffers from errors propagating from upstream entity linkers to downstream relation extractors. In this paper, we propose a novel joint inference framework to allow interactions between the two subtasks and find an optimal assignment by addressing the coherence among preliminary local predictions: whether the types of entities meet the expectations of relations explicitly or implicitly, and whether the local predictions are globally compatible. We further measure the confidence of the extracted triples by looking at the details of the complete extraction process. Experiments show that the proposed framework can significantly reduce the error propagations thus obtain more reliable facts, and outperforms competitive baselines with state-of-the-art relation extraction models.

Combining Visual and Textual Features for Information Extraction from Online Flyers
Emilia Apostolova and Noriko Tomuro

Information in visually rich formats such as PDF and HTML is often conveyed by a combination of textual and visual features. In particular, genres such as marketing flyers and info-graphics often augment textual information by its color, size, positioning, etc. As a result, traditional text-based approaches to information extraction (IE) could underperform. In this study, we present a supervised machine learning approach to IE from online commercial real estate flyers. We evaluated the performance of SVM classifiers on the task of identifying 12 types of named entities using a combination of textual and visual features. Results show that the addition of visual features such as color, size, and positioning significantly increased classifier performance.

CTPs: Contextual Temporal Profiles for Time Scoping Facts using State Change Detection
Derry Tanti Wijaya, Ndapandula Nakashole, and Tom Mitchell

Temporal scope adds a time dimension to facts in Knowledge Bases (KBs). These time scopes specify the time periods when a given fact was valid in real life. Without temporal scope, many facts are underspecified, reducing the usefulness of the data for upper level applications such as Question Answering. Existing methods for temporal scope inference and extraction still suffer from low accuracy. In this paper, we present a new method that leverages temporal profiles augmented with context— Contextual Temporal Profiles (CTPs) of entities. Through change patterns in an entity's CTP, we model the entity's state change brought about by real world events that happen to the entity (e.g, hired, fired, divorced, etc.). This leads to a new formulation of the temporal scoping problem as a state change detection problem. Our experiments show that this formulation of the problem, and the resulting solution are highly effective for inferring temporal scope of facts.

Noisy Or-based model for Relation Extraction using Distant Supervision
Ajay Nagesh, Gholamreza Haffari, and Ganesh Ramakrishnan

Distant supervision, a paradigm of rela- tion extraction where training data is cre- ated by aligning facts in the database with a large unannotated corpus, is currently an attractive approach for training relation extractors. Various models are proposed in recent literature to align the facts in the database to their mentions in the cor- pus. In this paper, we discuss and crit- ically analyse a popular alignment strat- egy called "at least one" heuristic. We provide a simple, yet effective relaxation to this strategy. We formulate the in- ference procedures in training as integer linear programming (ILP) problems and implement this relaxation via a soft con- straint in this formulation. Empirically, we demonstrate that this simple strategy leads to a better performance under certain set- tings over the existing approaches.

Parallel Session 9 Overview – Tuesday, October 28, 2014

Track A	Track B	Track C	
Machine Learning and Machine Translation	*NLP for the Web and Social Media*	*Semantics*	
Al Areen 1-2	Al Areen 3	Al Areen 4	
Search-Aware Tuning for Machine Translation *Lemao Liu and Liang Huang*	Gender and Power: How Gender and Gender Environment Affect Manifestations of Power *Vinodkumar Prabhakaran, Emily E. Reid, and Owen Rambow*	Brighter than Gold: Figurative Language in User Generated Comparisons *Vlad Niculae and Cristian Danescu-Niculescu-Mizil*	15:40
Latent-Variable Synchronous CFGs for Hierarchical Translation *Avneesh Saluja, Chris Dyer, and Shay B. Cohen*	Online topic model for Twitter considering dynamics of user interests and topic trends *Kentaro Sasaki, Tomohiro Yoshikawa, and Takeshi Furuhashi*	Classifying Idiomatic and Literal Expressions Using Topic Models and Intensity of Emotions *Jing Peng, Anna Feldman, and Ekaterina Vylomova*	16:05
[TACL] Dynamically Shaping the Reordering Search Space of Phrase-Based Statistical Machine Translation *Arianna Bisazza and Marcello Federico*	Self-disclosure topic model for classifying and analyzing Twitter conversations *JinYeong Bak, Chin-Yew Lin, and Alice Oh*	[TACL] TREETALK: Composition and Compression of Trees for Image Descriptions *Polina Kuznetsova, Vicente Ordonez, Tamara Berg, and Yejin Choi*	16:30
	Major Life Event Extraction from Twitter based on Congratulations/Condolences Speech Acts *Jiwei Li, Alan Ritter, Claire Cardie, and Eduard Hovy*	Learning Spatial Knowledge for Text to 3D Scene Generation *Angel Chang, Manolis Savva, and Christopher Manning*	16:55

Track P

Posters: Discourse, Dialogue and Pragmatics

Al Areen 5-6   15:40-17:20

- A Model of Coherence Based on Distributed Sentence Representation
  Jiwei Li and Eduard Hovy

- Discriminative Reranking of Discourse Parses Using Tree Kernels
  Shafiq Joty and Alessandro Moschitti

- Recursive Deep Models for Discourse Parsing
  Jiwei Li, Rumeng Li, and Eduard Hovy

- Recall Error Analysis for Coreference Resolution
  Sebastian Martschat and Michael Strube

- A Rule-Based System for Unrestricted Bridging Resolution: Recognizing Bridging Anaphora and Finding Links to Antecedents
  Yufang Hou, Katja Markert, and Michael Strube

- Resolving Referring Expressions in Conversational Dialogs for Natural User Interfaces
  Asli Celikyilmaz, Zhaleh Feizollahi, Dilek Hakkani-Tür, and Ruhi Sarikaya

- Building Chinese Discourse Corpus with Connective-driven Dependency Tree Structure
  Yancui Li, Fang Kong, and Guodong Zhou

- Prune-and-Score: Learning for Greedy Coreference Resolution
  Chao Ma, Janardhan Rao Doppa, J. Walker Orr, Prashanth Mannem, Xiaoli Fern, Tom Dietterich, and Prasad Tadepalli

- Summarizing Online Forum Discussions – Can Dialog Acts of Individual Messages Help?
  Sumit Bhatia, Prakhar Biyani, and Prasenjit Mitra

Parallel Session 9

Session 9a: Machine Learning and Machine Translation
Al Areen 1-2   *Chair: Khalil Sima'an*

Search-Aware Tuning for Machine Translation
Lemao Liu and Liang Huang   15:40–16:05

Most current parameter tuning methods for machine translation (such as MERT and PRO) are agnostic about search, while search errors are well-known to adversely affect translation quality. We propose to promote potentially accurate partial translations and prevent them from being pruned, and develop two metrics to evaluate partial derivations. Our method can be applied to all of the three most popular tuning algorithms: MERT, PRO, and MIRA, where extensive experiments on Chinese-to- English and English-to-Chinese translation show up to +2.6 BLEU gains with each of the three algorithms.

Latent-Variable Synchronous CFGs for Hierarchical Translation
Avneesh Saluja, Chris Dyer, and Shay B. Cohen   16:05–16:30

Data-driven refinement of non-terminal categories has been demonstrated to be a reliable technique for improving monolingual parsing with PCFGs. In this paper, we extend these techniques to learn latent refinements of single-category synchronous grammars, so as to improve translation performance. We compare two estimators for this latent-variable model: one based on EM and the other is a spectral algorithm based on the method of moments. We evaluate their performance on a Chinese–English translation task. The results indicate that we can achieve significant gains over the baseline with both approaches, but in particular the moments-based estimator is both faster and performs better than EM.

[TACL] Dynamically Shaping the Reordering Search Space of Phrase-Based Statistical Machine Translation
Arianna Bisazza and Marcello Federico   16:30–16:55

Defining the reordering search space is a crucial issue in phrase-based SMT between distant languages. In fact, the optimal trade-off between accuracy and complexity of decoding is nowadays reached by harshly limiting the input permutation space. We propose a method to dynamically shape such space and, thus, capture long-range word movements without hurting translation quality nor decoding time. The space defined by loose reordering constraints is dynamically pruned through a binary classifier that predicts whether a given input word should be translated right after another. The integration of this model into a phrase-based decoder improves a strong Arabic-English baseline that already includes state-of-the-art early distortion cost (Moore and Quirk, 2007) and hierarchical phrase orientation models (Galley and Manning, 2008). Significant improvements in the reordering of verbs are achieved by a system that is notably faster than the baseline, while BLEU and METEOR remain stable, or even increase, at a very high distortion limit.

Session 9b: NLP for the Web and Social Media
Al Areen 3 *Chair: Ingmar Weber*

Gender and Power: How Gender and Gender Environment Affect Manifestations of Power
Vinodkumar Prabhakaran, Emily E. Reid, and Owen Rambow 15:40–16:05

We investigate the interaction of power, gender, and language use in the Enron email corpus. We present a freely available extension to the Enron corpus, with the gender of senders of 87% messages reliably identified. Using this data, we test two specific hypotheses drawn from the sociolinguistic literature pertaining to gender and power: women managers use face-saving communicative strategies, and women use language more explicitly than men to create and maintain social relations. We introduce the notion of "gender environment" to the computational study of written conversations; we interpret this notion as the gender makeup of an email thread, and show that some manifestations of power differ significantly between gender environments. Finally, we show the utility of gender information in the problem of automatically predicting the direction of power between pairs of participants in email interactions.

Online topic model for Twitter considering dynamics of user interests and topic trends
Kentaro Sasaki, Tomohiro Yoshikawa, and Takeshi Furuhashi 16:05–16:30

Latent Dirichlet Allocation (LDA) is a topic model which has been applied to various fields. It has been also applied to user profiling or event summarization on Twitter. In the application of LDA to tweet collection, it generally treats aggregated all tweets of a user as a single document. On the other hand, Twitter-LDA which assumes a single tweet consists of a single topic has been proposed and showed that it is superior to the former way in topic semantic coherence. However, Twitter-LDA has a problem that it is not capable of online inference. In this paper, we extend Twitter-LDA in the following two points. First, we model the generation process of tweets more accurately by estimating the ratio between topic words and general words for each user. Second, we enable it to estimate dynamics of user interests and topic trends in online based on Topic Tracking Model (TTM) which models consumer purchase behaviors.

Self-disclosure topic model for classifying and analyzing Twitter conversations
JinYeong Bak, Chin-Yew Lin, and Alice Oh 16:30–16:55

Self-disclosure, the act of revealing oneself to others, is an important social behavior that strengthens interpersonal relationships and increases social support. Although there are many social science studies of self-disclosure, they are based on manual coding of small datasets and questionnaires. We conduct a computational analysis of self-disclosure with a large dataset of naturally-occurring conversations, a semi-supervised machine learning algorithm, and a computational analysis of the effects of self-disclosure on subsequent conversations. We use a longitudinal dataset of 17 million tweets, all of which occurred in conversations that consist of five or more tweets directly replying to the previous tweet, and from dyads with twenty of more conversations each. We develop self-disclosure topic model (SDTM), a variant of latent Dirichlet allocation (LDA) for automatically classifying the level of self-disclosure for each tweet. We take the results of SDTM and analyze the effects of self-disclosure on subsequent conversations. Our model significantly outperforms several comparable methods on classifying the level of self-disclosure, and the analysis of the longitudinal data using SDTM uncovers significant and positive correlation between self-disclosure and conversation frequency and length.

Major Life Event Extraction from Twitter based on Congratulations/Condolences Speech Acts
Jiwei Li, Alan Ritter, Claire Cardie, and Eduard Hovy 16:55–17:20

Social media websites provide a platform for anyone to describe significant events taking place in their lives in realtime. Currently, the majority of personal news and life events are published in a textual format, motivating information extraction systems that can provide a structured representations of major life events (weddings, graduation, etc...). This paper demonstrates the feasibility of accurately extracting major life events. Our system extracts a fine-grained description of users' life events based on their published tweets. We are optimistic that our system can help Twitter users more easily grasp information from users they take interest in following and also facilitate many downstream applications, for example realtime friend recommendation.

Session 9c: Semantics
Al Areen 4 *Chair: Mehrnoosh Sadrzadeh*

Brighter than Gold: Figurative Language in User Generated Comparisons
Vlad Niculae and Cristian Danescu-Niculescu-Mizil 15:40–16:05

Comparisons are common linguistic devices used to indicate the likeness of two things. Often, this likeness is not meant in the literal sense—for example, "I slept as a log" does not imply that logs actually sleep. In this paper we propose a computational study of figurative comparisons, or similes. Our starting point is a new large dataset of comparisons extracted from product reviews and annotated for figurativeness. We use this dataset to characterize figurative language in naturally occurring comparisons and reveal linguistic patterns indicative of this phenomenon. We operationalize these insights and apply them to a new task with high relevance to text understanding: distinguishing between figurative and literal comparisons. Finally, we apply this framework to explore the social context in which figurative language is produced, showing that similes are more likely to accompany opinions showing extreme sentiment, and that they are uncommon in reviews deemed helpful.

Classifying Idiomatic and Literal Expressions Using Topic Models and Intensity of Emotions
Jing Peng, Anna Feldman, and Ekaterina Vylomova 16:05–16:30

We describe an algorithm for automatic classification of idiomatic and literal expressions. Our starting point is that words in a given text segment, such as a paragraph, that are high-ranking representatives of a common topic of discussion are less likely to be a part of an idiomatic expression. Our additional hypothesis is that contexts in which idioms occur, typically, are more affective and therefore, we incorporate a simple analysis of the intensity of the emotions expressed by the contexts. We investigate the bag of words topic representation of one to three paragraphs containing an expression that should be classified as idiomatic or literal (a target phrase). We extract topics from paragraphs containing idioms and from paragraphs containing literals using an unsupervised clustering method, Latent Dirichlet Allocation (LDA). Since idiomatic expressions exhibit the property of non-compositionality, we assume that they usually present different semantics than the words used in the local topic. We treat idioms as semantic outliers, and the identification of a semantic shift as outlier detection. Thus, this topic representation allows us to differentiate idioms from literals using local semantic contexts. Our results are encouraging.

[TACL] TREETALK: Composition and Compression of Trees for Image Descriptions
Polina Kuznetsova, Vicente Ordonez, Tamara Berg, and Yejin Choi 16:30–16:55

We present a new tree based approach to composing expressive image descriptions that makes use of naturally occurring web images with captions. We investigate two related tasks: image caption generalization and generation, where the former is an optional sub- task of the latter. The high-level idea of our approach is to harvest expressive phrases (as tree fragments) from existing image descriptions, then to compose a new description by selectively combining the extracted (and optionally pruned) tree fragments. Key algorithmic components are tree composition and compression, both integrating tree structure with sequence structure. Our proposed system attains significantly better performance than previous approaches for both image caption generalization and generation. In addition, our work is the first to show the empirical benefit of automatically generalized captions for composing natural image descriptions.

Learning Spatial Knowledge for Text to 3D Scene Generation
Angel Chang, Manolis Savva, and Christopher Manning 16:55–17:20

We address the grounding of natural language to concrete spatial constraints, and inference of implicit pragmatics in 3D environments. We apply our approach to the task of text-to-3D scene generation. We present a representation for common sense spatial knowledge and an approach to extract it from 3D scene data. In text-to-3D scene generation, a user provides as input natural language text from which we extract explicit constraints on the objects that should appear in the scene. The main innovation of this work is to show how to augment these explicit constraints with learned spatial knowledge to infer missing objects and likely layouts for the objects in the scene. We demonstrate that spatial knowledge is useful for interpreting natural language and show examples of learned knowledge and generated 3D scenes.

Session 9-posters: Discourse, Dialogue and Pragmatics
Al Areen 5-6   15:40–17:20

A Model of Coherence Based on Distributed Sentence Representation
Jiwei Li and Eduard Hovy

Coherence is what makes a multi-sentence text meaningful, both logically and syntactically. To solve the challenge of ordering a set of sentences into coherent order, existing approaches focus mostly on defining and using sophisticated features to capture the cross-sentence argumentation logic and syntactic relationships. But both argumentation semantics and cross-sentence syntax (such as coreference and tense rules) are very hard to formalize. In this paper, we introduce a neural network model for the coherence task based on distributed sentence representation. Our approach learns a syntactico-semantic representation for sentences automatically, using either recurrent or recursive neural networks. The deep learning architecture not only obviated the need for feature engineering, but learns which sentence representation is able to capture the 'rules' governing coherent sentence structure. The approach outperforms existing baselines by a large margin and generates the state-of-art performance in standard coherence evaluation tasks[1].

Discriminative Reranking of Discourse Parses Using Tree Kernels
Shafiq Joty and Alessandro Moschitti

In this paper, we present a discriminative approach for reranking discourse trees generated by an existing probabilistic discourse parser. The reranker relies on tree kernels (TKs) to capture the global dependencies between discourse units in a tree. In particular, we design new computational structures of discourse trees, which combined with standard TKs, originate novel discourse TKs. The empirical evaluation shows that our reranker can improve the state-of-the-art sentence-level parsing accuracy from 79.77% to 82.15%, a relative error reduction of 11.8%, which in turn pushes the state-of-the-art document-level accuracy from 55.8% to 57.3%.

Recursive Deep Models for Discourse Parsing
Jiwei Li, Rumeng Li, and Eduard Hovy

Text-level discourse parsing remains a challenge: most approaches employ features that fail to capture the intentional, semantic, and syntactic aspects that govern discourse coherence. In this paper, we propose a recursive deep learning model for discourse parsing that jointly models distributed representations for clauses, sentences, and entire discourses. The deep learning approach is capable of capturing the semantic and intentional import of words and larger discourse units automatically, freeing the researcher from the effort required for feature engineering. The proposed framework obtains comparable performance regarding standard discoursing parsing evaluations when compared against current state-of-art systems.

Recall Error Analysis for Coreference Resolution
Sebastian Martschat and Michael Strube

We present a novel method for coreference resolution error analysis which we apply to perform a recall error analysis of four state-of-the-art English coreference resolution systems. Our analysis highlights differences between the systems and identifies that the majority of recall errors for nouns and names are shared by all systems. We characterize this set of common challenging errors in terms of a broad range of lexical and semantic properties.

A Rule-Based System for Unrestricted Bridging Resolution: Recognizing Bridging Anaphora and Finding Links to Antecedents
Yufang Hou, Katja Markert, and Michael Strube

Bridging resolution plays an important role in establishing (local) entity coherence. This paper proposes a rule-based approach for the challenging task of unrestricted bridging resolution, where bridging anaphors are not limited to definite NPs and semantic relations between anaphors and their antecedents are not restricted to meronymic relations. The system consists of eight rules which target different relations based on linguistic insights. Our rule-based system significantly outperforms a reimplementation of a previous rule-based system (Vieira and Poe- sio, 2000). Furthermore, it performs better than a learning-based approach which has access to the same knowledge resources as the rule-based system. Additionally, incorporating the rules and more features into the learning-based system yields a minor improvement over the rule-based system.

Resolving Referring Expressions in Conversational Dialogs for Natural User Interfaces
Asli Celikyilmaz, Zhaleh Feizollahi, Dilek Hakkani-Tür, and Ruhi Sarikaya

---

[1]Code available at `stanford.edu/{~}jiweil/` or by request from the first author.

Unlike traditional over-the-phone spoken dialog systems (SDSs), modern dialog systems tend to have visual rendering on the device screen as an additional modality to communicate the system's response to the user. Visual display of the system's response not only changes human behavior when interacting with devices, but also creates new research areas in SDSs. On-screen item identification and resolution in utterances is one critical problem to achieve a natural and accurate human-machine communication. We pose the problem as a classification task to correctly identify intended on-screen item(s) from user utterances. Using syntactic, semantic as well as context features from the display screen, our model can resolve different types of referring expressions with up to 90% accuracy. In the experiments we also show that the proposed model is robust to domain and screen layout changes.

Building Chinese Discourse Corpus with Connective-driven Dependency Tree Structure
Yancui Li, Fang Kong, and Guodong Zhou

In this paper, we propose a Connective-driven Dependency Tree (CDT) scheme to represent the discourse rhetorical structure in Chinese language, with ele-mentary discourse units as leaf nodes and connectives as non-leaf nodes, largely motivated by the Penn Discourse Treebank and the Rhetorical Structure Theory. In particular, connectives are employed to directly represent the hier-archy of the tree structure and the rhe-torical relation of a discourse, while the nuclei of discourse units are globally de-termined with reference to the depend-ency theory. Guided by the CDT scheme, we manually annotate a Chinese Dis-course Treebank (CDTB) of 500 docu-ments. Preliminary evaluation justifies the appropriateness of the CDT scheme to Chinese discourse analysis and the usefulness of our manually annotated CDTB corpus.

Prune-and-Score: Learning for Greedy Coreference Resolution
Chao Ma, Janardhan Rao Doppa, J. Walker Orr, Prashanth Mannem, Xiaoli Fern, Tom Dietterich, and Prasad Tadepalli

We propose a novel search-based approach for greedy coreference resolution, where the mentions are processed in order and added to previous coreference clusters. Our method is distinguished by the use of two functions to make each coreference decision: a pruning function that prunes bad coreference decisions from further consideration, and a scoring function that then selects the best among the remaining decisions. Our framework reduces learning of these functions to rank learning, which helps leverage powerful off-the-shelf rank-learners. We show that our Prune-and-Score approach is superior to using a single scoring function to make both decisions and outperforms several state-of-the-art approaches on multiple benchmark corpora including OntoNotes.

Summarizing Online Forum Discussions – Can Dialog Acts of Individual Messages Help?
Sumit Bhatia, Prakhar Biyani, and Prasenjit Mitra

A typical discussion thread in an online forum spans multiple pages involving participation from multiple users and thus, may contain multiple view-points and solutions. A user interested in the topic of discussion or having a problem similar to being discussed in the thread may not want to read all the previous posts but only a few selected posts that provide her a concise summary of the ongoing discussion. This paper describes an extractive summarization technique that uses textual features and dialog act information of individual messages to select a subset of posts. Proposed approach is evaluated using two real life forum datasets.

Tutorials: Wednesday, October 29

Overview

Tutorial 5

Syntax-Based Statistical Machine Translation

Philip Williams (University of Edinburgh) and Philipp Koehn (Johns Hopkins University / University of Edinburgh)

Wednesday, October 29, 2014, 9:00 – 12:30pm

Al Areen 4

The tutorial explains in detail syntax-based statistical machine translation with synchronous context free grammars (SCFG). It is aimed at researchers who have little background in this area, and gives a comprehensive overview about the main models and methods.

While syntax-based models in statistical machine translation have a long history, spanning back almost 20 years, they have only recently shown superior translation quality over the more commonly used phrase-based models, and are now considered state of the art for some language pairs, such as Chinese-English (since ISI's submission to NIST 2006), and English-German (since Edinburgh's submission to WMT 2012).

While the field is very dynamic, there is a core set of methods that have become dominant. Such SCFG models are implemented in the open source machine translation toolkit Moses, and the tutors draw from the practical experience of its development.

The tutorial focuses on explaining core established concepts in SCFG-based approaches, which are the most popular in this area. The main goal of the tutorial is for the audience to understand how these systems work end-to-end. We review as much relevant literature as necessary, but the tutorial is not a primarily research survey.

The tutorial is rounded up with open problems and advanced topics, such as computational challenges, different formalisms for syntax-based models and inclusion of semantics.

Philip Williams completed his PhD at the University of Edinburgh in 2014, with a thesis on unification-based formalisms applied to syntax-based statistical machine translation. He has been recently the main contributor to the syntax-based MT model in the Moses decoder.

Philipp Koehn has worked on statistical machine translation since 1997, is leading the development of the open source machine translation toolkit Moses, and wrote the textbook for the field (published in 2010). He has worked on many aspects on machine translation, including syntax-based approaches.

Tutorial 6

Embedding Methods for Natural Language Processing

Antoine Bordes and Jason Weston (Facebook AI Research)

Wednesday, October 29, 2014, 9:00 – 12:30pm

Al Areen 5

Embedding-based models, that consists in learning latent representations of symbols (words, database entries), are popular tools in Natural Language Processing these days. In this tutorial, our goal is to provide an overview of the main advances in this domain. The vector latent representations learned by these methods have proved to be very efficient in many applications ranging from semantic search, automatic knowledge base construction, to natural language understanding or document retrieval.

Antoine Bordes is a research scientist at Facebook AI Research. Before joining Facebook, he was a CNRS researcher in the Heudiasyc laboratory of Universite de Technologie of Compiegne in France. Antoine's current research concerns large-scale machine learning applied to natural language understanding and knowledge base modeling, mostly with embedding models.

Jason Weston is a research scientist at Facebook AI Research. His previous affiliations include Google Research, NEC Research and the Max Planck Institute for Biological Cybernetics. Jason's current research concerns applying neural network and embedding models towards understanding text.

Tutorial 7

Natural Language Processing of Arabic and its Dialects

Mona Diab (George Washington University) and Nizar Habash (New York University Abu Dhabi)

Wednesday, October 29, 2014, 14:00 – 17:30pm

Al Areen 4

This tutorial introduces the different challenges and current solutions to the automatic processing of Arabic and its dialects. The tutorial is divided into sections on orthography, morphology, syntax, lexical variations, and translation. Each section will introduce basic concepts and ideas, detail dialectal variation issues, and discuss relevant enabling technologies and applications, e.g., spelling correction, morphological processing (analysis, disambiguation, tokenization, POS tagging), parsing, dialect identification, and machine translation from and into Arabic.

Mona Diab is an Associate Professor of Computer Science at the George Washington University.
Nizar Habash is an Associate Professor of Computer Science at New York University Abu Dhabi.

Tutorial 8

Text Quantification

Fabrizio Sebastiani (Qatar Computing Research Institute)

Wednesday, October 29, 2014, 14:00 – 17:30pm

Al Areen 5

In recent years it has been pointed out that, in a number of applications involving text classification, the final goal is not determining which class (or classes) individual unlabelled data items belong to, but determining the prevalence (or "relative frequency") of each class in the unlabelled data. The latter task is known as quantification.

Assume a market research agency runs a poll in which they ask the question "What do you think of the recent ad campaign for product X?" Once the poll is complete, they may want to classify the resulting textual answers according to whether they belong or not to the class LovedTheCampaign. The agency is likely not interested in whether a specific individual belongs to the class LovedTheCampaign, but in knowing how many respondents belong to it, i.e., in knowing the prevalence of the class. In other words, the agency is interested not in classification, but in quantification. Essentially, quantification is classification tackled at the aggregate (rather than at the individual) level.

The research community has recently shown a growing interest in tackling text quantification as a task in its own right. One of the reasons is that, since the goal of quantification is different than that of classification, quantification requires evaluation measures different than for classification. A second, related reason is that using a method optimized for classification accuracy is suboptimal when quantification accuracy is the real goal. A third reason is the growing awareness that quantification is going to be more and more important; with the advent of big data, more and more application contexts are going to spring up in which we will simply be happy with analyzing data at the aggregate (rather than at the individual) level.

The goal of this tutorial is to introduce the audience to the problem of quantification, to the techniques that have been proposed for solving it, to the metrics used to evaluate them, and to the problems that are still open in the area.

––––––––––––––––––––

Fabrizio Sebastiani is a Principal Scientist at QCRI; he was (until June 2014) a Senior Researcher at the Italian National Council of Research (from which he is currently on leave), and (until February 2006) an Associate Professor at the Department of Pure and Applied Mathematics of the University of Padova, Italy. His main current research interests are at the intersection of information retrieval, machine learning, and human language technologies, with particular emphasis on text classification, information extraction, opinion mining, and their applications.

8

Workshops: Wednesday, October 29

Overview

Workshop 5: LT4CloseLang: Language Technology for Closely-Related Languages and Language Variants

Organizers: *Preslav Nakov, Petya Osenova, and Cristina Vertan*

Venue: Al Areen 1

Wednesday, October 29, 2014

Opening Session and Invited Talk 1

8:50–9:00 Opening Remarks
The organizers

9:00–10:00 INVITED TALK 1: Computational Processing of Arabic Dialects
Nizar Habash

Session 1: Parsing

10:00–10:20 Learning from a Neighbor: Adapting a Japanese Parser for Korean Through Feature Transfer Learning
Hiroshi Kanayama, Youngja Park, Yuta Tsuboi, and Dongmook Yi

10:20–10:40 Cross-lingual Dependency Parsing of Related Languages with Rich Morphosyntactic Tagsets
Željko Agić, Jörg Tiedemann, Danijela Merkler, Simon Krek, Kaja Dobrovoljc, and Sara Moze

10:40–11:00 **Coffee Break**

Session 2: Variety and Adaptation

11:00–11:20 Language variety identification in Spanish tweets
Wolfgang Maier and Carlos Gómez-Rodríguez

11:20–11:40 Exploiting Language Variants Via Grammar Parsing Having Morphologically Rich Information
Qaiser Abbas

11:40–12:00 Adapting Predicate Frames for Urdu PropBanking
Riyaz Ahmad Bhat, Naman Jain, Ashwini Vaidya, Martha Palmer, Tafseer Ahmed Khan, Dipti Misra Sharma, and James Babani

12:00–12:20 Measuring Language Closeness by Modeling Regularity
Javad Nouri and Roman Yangarber

12:20–2:00 **Lunch**

Invited Talk 2

2:00–3:00 INVITED TALK 2: Towards Universal Syntactic Processing of Natural Language
Slav Petrov

Session 3: Machine Translation I

3:00–3:20 Proper Name Machine Translation from Japanese to Japanese Sign Language
Taro Miyazaki, Naoto Kato, Seiki Inoue, Shuichi Umeda, Makiko Azuma, Nobuyuki Hiruma, and Yuji Nagashima

3:20–3:40 Exploring cross-language statistical machine translation for closely related South Slavic languages
Maja Popović and Nikola Ljubešić

3:40–4:00 **Coffee Break**

Session 4: Machine Translation II

4:00–4:20 Exploring System Combination approaches for Indo-Aryan MT Systems
*Karan Singla, Anupam Singh, Nishkarsh Shastri, Megha Jhunjhunwala,
Srinivas Bangalore, and Dipti Misra Sharma*

4:20–4:40 A Comparison of MT Methods for Closely Related Languages: a Case Study on
Czech - Slovak Language Pair
Vladislav Kubon and Jernej Vicic

4:40–5:00 Handling OOV Words in Dialectal Arabic to English Machine Translation
Maryam Aminian, Mahmoud Ghoneim, and Mona Diab

Closing Session

5:00–6:00 Panel
*Houda Bouamor, Kareem Darwish, Vladislav Kubon, Wolfgang Maier, and
Kemal Oflazer*

6:00–6:10 Closing Remarks
The organizers

Workshop 6: TextGraphs-9: Graph-based Methods for Natural Language Processing

Organizers: *V. G. Vinod Vydiswaran, Amarnag Subramanya, Gabor Melli, and Irina Matveeva*

Venue: Al Areen 3

Wednesday, October 29, 2014

Session 1

9:00–9:10 Welcome and Introduction
The organizers

9:10–10:15 Keynote Talk
Mohammed J. Zaki

10:15–10:30 Normalized Entity Graph for Computing Local Coherence
Mohsen Mesgar and Michael Strube

10:30–11:00 **Coffee break**

Session 2

11:00–11:25 Exploiting Timegraphs in Temporal Relation Classification
Natsuda Laokulrat, Makoto Miwa, and Yoshimasa Tsuruoka

11:25–11:50 Multi-document Summarization Using Bipartite Graphs
Daraksha Parveen and Michael Strube

11:50–12:15 A Novel Two-stage Framework for Extracting Opinionated Sentences from News Articles
Pujari Rajkumar, Swara Desai, Niloy Ganguly, and Pawan Goyal

12:15–12:30 Constructing Coherent Event Hierarchies from News Stories
Goran Glavaš and Jan Snajder

12:30–2:00 **Lunch**

Session 3

2:00–3:05 Invited Talk
Partha Talukdar

3:05–3:30 Semi-supervised Graph-based Genre Classification for Web Pages
Noushin Rezapour Asheghi, Katja Markert, and Serge Sharoff

3:30–4:00 **Coffee break**

Session 4

4:00–4:25 The Modular Community Structure of Linguistic Predication Networks
Aaron Gerow and James Evans

4:25–4:50 From Visualisation to Hypothesis Construction for Second Language Acquisition
Shervin Malmasi and Mark Dras

4:50–5:00 Conclusion
The organizers

Index

115

9

Useful Local Information

EMNLP 2014 will take place in the Marriott Marquis City Center Hotel, in Doha's West Bay area, also known as Doha's Central Business District. The Marriott is adjoined to City Center Mall, one of Doha's largest and most vibrant shopping malls with a wide range of retail stores with international brands, cafes and restaurants, and entertainment options including a cinema, bowling alley and an ice skating rink. The conference it is also a short journey to a number of key local landmarks that are not to be missed including the Corniche, The Museum of Islamic Art, Souq Waqif, Katara Cultural Village and The Pearl.

Numbers, internet, and websites

Important phone numbers

- Police, Fire, Ambulance: 999

- Telephone Directory: 180

- EMNLP emergency contact:

    - Amanda Barnett: +974-33850652
    - Kareem Darwish: +974-33567421

- Traffic & Patrol Police Department: +974-44890666

- Hamad Hospital: +974-44392111

Internet at conference venue There are two internet options for conference participants:

- Exclusive internet connection for conference participants: Details are provided on name tags

- Free public internet in public areas in Marriott hotel

Important web links

- Conference website: `www.emnlp2014.org`

- Conference mobile application: **Conference4Me** – `conference4me.psnc.pl`

- IloveQatar.net local guide: `www.iloveqatar.net/guide`

- Qatar Tourism Authority: `www.qatartourism.gov.qa`

- Qatar International Adventures (for desert safaris): `www.qia-qatar.com`

Transportation

City Public Transportation One of the easier ways is by Taxi. There are a few local taxi operators. It is generally difficult to pick up a taxi off the street. You would need to call the operator before hand or go to their meeting points. The closest meeting point to the conference venue is in front the City Center Mall (less than 50 meters from the conference venue). A taxi ride is typically around 30 Qatari Riyals (about $8). Some of the operators are:

- Karwa Taxi Doha +974-4458-8888

- Super Limousine +974-5546-9299

- Super Transport Service +974-6612-2030

- Fox Transport +974-4462-2777

- Doha Limousine +974-4483-9999

- Rez Auto +974-3003-0149

- Cabz Transport +974-4441-0383

There are three different bus service options, namely:

- Public bus service. The service is sparse and not recommended.

- The West Bay Bus service. This is a free service with buses every 15 minutes between 6 am and 12 midnight. It only operates in the West Bay area and has a stop within 50 meters of the conference venue. The bus map is available here: `www.baladiya.gov.qa/pdocs/cview?siteID=1&docID=5135&year=2014`.

- The Doha Bus (aka the Hop-on Hop-off). This is a paid service with daily passes with buses every 30 minutes. It connects most of Doha's tourist attractions including the Museum of Islamic Art, the Pearl, and Souq Waqif. They are bit pricey. They make sense only when you are traveling alone and you want to visit several places on the same day. You can find more information by visiting: `www.dohabus.com/`.

To the airport The most reliable method to go the airport is to use a taxi. A taxi can be obtained from in front of City Center Mall or by booking directly or through the hotel. If you plan to book a taxi, make sure to book it book at least 24 hours before the pick up time. The cost of the trip to the airport should be 50-60 Qatari Riyals ($14-17).

10

Sponsors

List of EMNLP 2014 sponsors

Diamond

Qatar National Research Fund

Platinum

Qatar Computing Research Institute

Gold

Carnegie Mellon University Qatar

Silver

Facebook

Yandex

Bronze

iHorizons

Sponsor of student volunteers

Google

Supporter

IBM Research

Creating a
Knowledge-Based
Society

Creating a
Research
Culture

Supporting
Research
in the
National
Interest

Facilitating
Collaboration in
Research

Contributing
to the Arab
Scientific
Renaissance

Setting
Standards in
Research

الصندوق القطري لرعاية البحث العلمي
Qatar National Research Fund

عضو في مؤسسة قطر
Member of Qatar Foundation

www.qnrf.org

معهد قطر لبحوث الحوسبة
Qatar Computing Research Institute

عضو في مؤسسة قطر
Member of Qatar Foundation

Data Analytics
Community Welfare
Global Impact
Potential
Cyber Security
Knowledge
Unique
Qatar
Science
Education
Collaborative
Arabic Language
Technologies
Computing
Futuristic
Innovation
Research
Institute
Computational
Science &
Engineering
National
Social Computing
People
Distributed Systems

/QCRI.QA   @qatarcomputing   /qatar-computing-research-institute   /QatarComputing

Please visit
www. qcri.com

IDEAS THAT SHAPE YOUR WORLD START HERE

Carnegie Mellon University Qatar
www.qatar.cmu.edu

facebook
(https://www.facebook.com/academics/)

Helping a billion people share requires innovation. At Facebook, we solve technical problems no one else has seen because no one else has built a social network of this size. We work at the intersection of research and engineering to make the world more open and connected. With more than 20% of our engineers having PhDs, Facebook has a culture that encourages research activity and interactions with academia. We offer graduate fellowships to support the research of promising students, establish research collaborations to help solve problems, and publish at many top conferences.

iHorizons ™

FOCUS ON
R&D IN ARABIC

NLP &
SOCIAL
ANALYTICS

Arabic Search Engine

Arabic Social Media Analytics

sponsors

Qatar National Research Fund

Member of Qatar Foundation

معهد قطر لبحوث الحوسبة
Qatar Computing Research Institute

عضو في مؤسسة قطر
Member of Qatar Foundation

Carnegie Mellon University Qatar

facebook

Yandex

iHorizons™

Google™

IBM **Research**