



emnlp

2013

The 2013 Conference
on
Empirical Methods
in
Natural Language Processing

Conference Handbook

October 18-21, 2013
Grand Hyatt Hotel
Seattle, Washington

General Chair

David Yarowsky (John Hopkins University, USA)

Program Chairs

Timothy Baldwin (University of Melbourne, Australia)

Anna Korhonen (University of Cambridge, UK)

Workshops Chair

Karen Livescu (Toyota Technological Institute at Chicago, USA)

Publication Chair

Steven Bethard (University of Alabama at Birmingham, USA)

Local Arrangements Chair:

Priscilla Rasmussen (ACL Business Office)

Webmaster:

Francesco Figari (The University of Edinburgh)

Area Chairs

Phonology, Morphology, Tagging, Chunking and Segmentation:

Kemal Oflazer (Carnegie Mellon University Qatar)

Anna Feldman (Montclair State University, USA)

Syntax and Parsing:

Jennifer Foster (Dublin City University, Ireland)

Yoav Goldberg (Bar Ilan University, Israel)

Semantics:

Mark Stevenson (University of Sheffield, UK)

Luke Zettlemoyer (University of Washington, USA)

Discourse, Dialogue, and Pragmatics:

Carolyn Rose (Carnegie Mellon University, USA)

Matt Purver (Queen Mary, University of London, UK)

Language resources:

Emily M. Bender (University of Washington, USA)

Aline Villavicencio (Federal University of Rio Grande do Sul, Brazil)

Summarization and Generation:

Dragomir Radev (University of Michigan, USA)

Yang Liu (University of Texas at Dallas, USA)

NLP-related Machine Learning: theory, methods and algorithms:

Amir Globerson (The Hebrew University of Jerusalem, Israel)

Antal van den Bosch (Radboud University Nijmegen, Netherlands)

Machine Translation:

Taro Watanabe (NICT, Japan)

Kevin Knight (Information Sciences Institute, USA)

Information Retrieval and Question Answering:

Bernardo Magnini (Fondazione Bruno Kessler, Italy)

Soumen Chakrabarti (Indian Institute of Technology, India)

Information Extraction:

Mausam (Indian Institute of Technology, India)

Heng Ji (City University of New York, USA)

Spoken Language Processing:

Haizhou Li (Institute for Infocomm Research, Singapore)

Amanda Stent (AT&T Labs, USA)

Text Mining and Natural Language Processing Applications:

Hang Li (Huawei Noah's Ark Lab, Hong Kong)

Kevin Cohen (University of Colorado School of Medicine, USA)

Sentiment Analysis and Opinion Mining:

Janyce Weibe (University of Pittsburgh, USA)

Bing Liu (University of Illinois at Chicago, USA)

NLP for the Web and Social Media:

Miles Osborne (University of Edinburgh, UK)

Chin-Yew Lin (Microsoft Research Asia, China)

Computational Models of Human Language Acquisition and Processing:

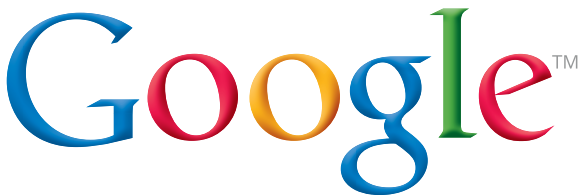
Alessandro Lenci (University of Pisa, Italy)

Afra Alishahi (Tilburg University, Netherlands)

Platinum Sponsor



Gold Sponsors



Bronze Sponsors



Supporters



Contents

1 Friday, October 19, 2013: Workshops	1
TextGraphs-8: Graph-based Methods for Natural Language Processing	2
Statistical Parsing of Morphologically-Rich Languages	5
Twenty Years of Bitext	10
2 Saturday, October 19, 2013: Main Conference	13
Information Extraction I	14
Language Acquisition and Processing	15
NLP for Social Media I	16
NLP Applications I	17
Semantics I	18
Language Resources	20
Invited talk: Andrew Ng	21
Machine Translation I	22
Dialogue and Discourse	24
Morphology and Phonology	26
Poster Session A	27
Poster Session B	41
3 Sunday, October 20, 2013: Main Conference	55
Invited talk: Fernando Pereira	56
Machine Learning for NLP	57
Summarization and Generation	59
Information Extraction and Social Media Analysis	61
Machine Translation II	63
Semantics II	64
Opinion Mining and Sentiment Analysis I	65
Machine Translation III	66
Information Extraction II	67
NLP Applications II	69

4 Monday, October 21, 2013: Main Conference	71
Information Extraction III	72
Opinion Mining and Sentiment Analysis II	73
NLP for Social Media II	74
Parsing	75
Semantics III	76
NLP Applications III	77
Plenary session I	78
Plenary Session II	79

1

Friday, October 19, 2013: Workshops

Overview

8:00–9:00	Breakfast	(Leonesa Foyer)
9:00–5:30	Workshop 1: TextGraphs-8	(Eliza Anderson Amphitheater)
9:00–5:30	Workshop 2: SPMRL-2013	(Blewett Suite)
9:00–5:30	Workshop 3: Twenty Years of Bitext	(Portland)

TextGraphs-8: Graph-based Methods for Natural Language Processing

Program Chairs: Zornitsa Kozareva, Irina Matveeva, Gabor Melli, Vivi Nastase

Invited talk: Opportunities for Graph-based Machine Reading Oren Etzioni

Friday 9:00–10:05 — Eliza Anderson Amphitheater

Event-Centered Information Retrieval Using Kernels on Event Graphs Goran Glavaš, Jan Šnajder

Friday 10:05–10:20 — Eliza Anderson Amphitheater

Traditional information retrieval models assume keyword-based queries and use unstructured document representations. There is an abundance of event-centered texts (e.g., breaking news) and event-oriented information needs that often involve structure that cannot be expressed using keywords. We present a novel retrieval model that uses a structured event-based representation. We structure queries and documents as graphs of event mentions and employ graph kernels to measure the query-document similarity. Experimental results on two event-oriented test collections show significant improvements over state-of-the-art keyword-based models.

JoBimText Visualizer: A Graph-based Approach to Contextualizing Distributional Similarity

Chris Biemann, Bonaventura Coppola, Michael R. Glass, Alfio Gliozzo, Matthew Hatem, Martin Riedl

Friday 10:20–10:30 — Eliza Anderson Amphitheater

We introduce an interactive visualization component for the JoBimText project. JoBimText is an open source platform for large-scale distributional semantics based on graph representations. First we describe the underlying technology for computing a distributional thesaurus on words using bipartite graphs of words and context features, and contextualizing the list of semantically similar words towards a given sentential context using graph-based ranking. Then we demonstrate the capabilities of this contextualized text expansion technology in an interactive visualization. The visualization can be used as a semantic parser providing contextualized expansions of words in text as well as disambiguation to word senses induced by graph clustering, and is provided as an open source tool.

Merging Word Senses

Sumit Bhagwani, Shrutiranjana Satapathy, Harish Karnick

Friday 11:00–11:25 — Eliza Anderson Amphitheater

WordNet, a widely used sense inventory for Word Sense Disambiguation, is often too fine-grained for many Natural Language applications because of its narrow sense distinctions. We present a semi-supervised approach to learn similarity between WordNet synsets using a graph based recursive similarity definition. We seed our framework with sense similarities of all the word-sense pairs, learnt using supervision on human-labelled sense clusterings. Finally we discuss our method to derive coarse sense inventories at arbitrary granularities and show that the coarse grained sense inventory obtained significantly boosts the disambiguation of nouns on standard test sets.

Reconstructing Big Semantic Similarity Networks

Ai He, Shefali Sharma, Chun-Nan Hsu

Friday 11:25–11:50 — Eliza Anderson Amphitheater

Distance metric learning from high (thousands or more) dimensional data with hundreds or thousands of classes is intractable but in NLP and IR, high dimensionality is usually required to represent data points, such as in modeling semantic similarity. This paper presents algorithms to scale up learning of a Mahalanobis distance metric from a large data graph in a high dimensional space. Our novel contributions include random projection that reduces dimensionality and a new objective function that regularizes intra-class and inter-class

distances to handle a large number of classes. We show that the new objective function is convex and can be efficiently optimized by a stochastic-batch subgradient descent method. We applied our algorithm to two different domains; semantic similarity of documents collected from the Web, and phenotype descriptions in genomic data. Experiments show that our algorithm can handle the high-dimensional big data and outperform competing approximations in both domains.

Graph-Based Unsupervised Learning of Word Similarities Using Heterogeneous Feature Types

Avneesh Saluja, Jiri Navratil

Friday 11:50–12:15 — Eliza Anderson Amphitheater

In this work, we propose a graph-based approach to computing similarities between words in an unsupervised manner, and take advantage of heterogeneous feature types in the process. The approach is based on the creation of two separate graphs, one for words and one for features of different types (alignment- based, orthographic, etc.). The graphs are connected through edges that link nodes in the feature graph to nodes in the word graph, the edge weights representing the importance of a particular feature for a particular word. High quality graphs are learned during training, and the proposed method outperforms experimental baselines.

From Global to Local Similarities: A Graph-Based Contextualization Method using Distributional Thesauri

Martin Riedl, Chris Biemann

Friday 12:15–12:30 — Eliza Anderson Amphitheater

After recasting the computation of a distributional thesaurus in a graph-based framework for term similarity, we introduce a new contextualization method that generates, for each term occurrence in a text, a ranked list of terms that are semantically similar and compatible with the given context. The framework is instantiated by the definition of term and context, which we derive from dependency parses in this work. Evaluating our approach on a standard data set for lexical substitution, we show substantial improvements over a strong non-contextualized baseline across all parts of speech. In contrast to comparable approaches, our framework defines an unsupervised generative method for similarity in context and does not rely on the existence of lexical resources as a source for candidate expansions.

Invited talk: Extracting Tractable Probabilistic Knowledge Graphs from Text

Pedro Domingos

Friday 2:00–3:05 — Eliza Anderson Amphitheater

Understanding seed selection in bootstrapping

Yo Ehara, Issei Sato, Hidekazu Oiwa, Hiroshi Nakagawa

Friday 3:05–3:30 — Eliza Anderson Amphitheater

Bootstrapping has recently become the focus of much attention in natural language processing to reduce labeling cost. In bootstrapping, unlabeled instances can be harvested from the initial labeled “seed” set. The selected seed set affects accuracy, but how to select a good seed set is not yet clear. Thus, an “iterative seeding” framework is proposed for bootstrapping to reduce its labeling cost. Our framework iteratively selects the unlabeled instance that has the best “goodness of seed” and labels the unlabeled instance in the seed set. Our framework deepens understanding of this seeding process in bootstrapping by deriving the dual problem. We propose a method called expected model rotation (EMR) that works well on not well-separated data which frequently occur as realistic data. Experimental results show that EMR can select seed sets that provide significantly higher mean reciprocal rank on realistic data than existing naive selection methods or random seed sets.

Graph-Structures Matching for Review Relevance Identification

Lakshmi Ramachandran, Edward Gehring

Friday 4:00–4:25 — Eliza Anderson Amphitheater

Review quality is determined by identifying the relevance of a review to a submission (the article or paper the review was written for). We identify relevance in terms of the semantic and syntactic similarities between

two texts. We use a word order graph, whose vertices, edges and double edges help determine structure-based match across texts. We use WordNet to determine semantic relatedness. Ours is a lexico-semantic approach, which predicts relevance with an accuracy of 66% and f-measure of 0.67.

Automatic Extraction of Reasoning Chains from Textual Reports

Gleb Sizov, Pinar Öztürk

Friday 4:25–4:50 — Eliza Anderson Amphitheater

Many organizations possess large collections of textual reports that document how a problem is solved or analysed, e.g. medical patient records, industrial accident reports, lawsuit records and investigation reports. Effective use of expert knowledge contained in these reports may greatly increase productivity of the organization. In this article, we propose a method for automatic extraction of reasoning chains that contain information used by the author of a report to analyse the problem at hand. For this purpose, we developed a graph-based text representation that makes the relations between textual units explicit. This representation is acquired automatically from a report using natural language processing tools including syntactic and discourse parsers. When applied to aviation investigation reports, our method generates reasoning chains that reveal the connection between initial information about the aircraft incident and its causes.

Graph-based Approaches for Organization Entity Resolution in MapReduce

Hakan Kardes, Deepak Konidena, Siddharth Agrawal, Micah Huff, Ang Sun

Friday 4:50–5:15 — Eliza Anderson Amphitheater

Entity Resolution is the task of identifying which records in a database refer to the same entity. A standard machine learning pipeline for the entity resolution problem consists of three major components: blocking, pairwise linkage, and clustering. The blocking step groups records by shared properties to determine which pairs of records should be examined by the pairwise linker as potential duplicates. Next, the linkage step assigns a probability score to pairs of records inside each block. If a pair scores above a user-defined threshold, the records are presumed to represent the same entity. Finally, the clustering step turns the input records into clusters of records (or profiles), where each cluster is uniquely associated with a single real-world entity. This paper describes the blocking and clustering strategies used to deploy a massive database of organization entities to power a major commercial People Search Engine. We demonstrate the viability of these algorithms for large data sets on a 50-node hadoop cluster.

A Graph-Based Approach to Skill Extraction from Text

Ilkka Kivimäki, Alexander Panchenko, Adrien Dessy, Dries Verdegem, Pascal Francq, Hugues Bersini, Marco Saerens

Friday 5:15–5:40 — Eliza Anderson Amphitheater

This paper presents a system, that performs skill extraction from text documents. It outputs a list of professional skills that are relevant to a given input text. We argue that the system can be practical for hiring and management of personnel in an organization. We make use of the texts and hyperlink graph of Wikipedia, as well as a list of professional skills obtained from the LinkedIn social network.

Statistical Parsing of Morphologically-Rich Languages

Program Chairs: Mohammed Attia, Bernd Bohnet, Marie Candito, Aoife Cahill, Özlem Cetinoglu, Jinho D. Choi, Grzegorz Chrupała, Benoit Crabbé, Gülşen Cebiroğlu Eryiğit, Michael Elhadad, Richard Farkas, Jennifer Foster, Josef van Genabith, Koldo Gojenola, Spence Green, Samar Husain, Sandra Kübler, Jonas Kuhn, Alberto Lavelli, Joseph Le Roux, Wolfgang Maier, Takuya Matsuzaki, Joakim Nivre, Kemal Oflazer, Adam Przepiórkowski, Owen Rambow, Kenji Sagae, Benoit Sagot, Djamé Seddah, Reut Tsarfaty, Lamia Tounsi, Daniel Zeman

Invited talk: An HDP Model for Inducing CCGs

Julia Hockenmaier

Friday 9:10–10:05 — Blewett Suite

Working with a small dataset - semi-supervised dependency parsing for Irish

Teresa Lynn, Jennifer Foster, Mark Dras

Friday 10:05–10:30 — Blewett Suite

We present a number of semi-supervised parsing experiments on the Irish language carried out using a small seed set of manually parsed trees and a larger, yet still relatively small, set of unlabelled sentences. We take two popular dependency parsers – one graph-based and one transition-based – and compare results for both. Results show that using semi-supervised learning in the form of self-training and co-training yields only very modest improvements in parsing accuracy. We also try to use morphological information in a targeted way and fail to see any improvements.

Lithuanian Dependency Parsing with Rich Morphological Features

Jurgita Kapociute-Dzikiene, Joakim Nivre, Algis Krupavicius

Friday 11:00–11:25 — Blewett Suite

We present the first statistical dependency parsing results for Lithuanian, a morphologically rich language in the Baltic branch of the Indo-European family. Using a greedy transition-based parser, we obtain a labeled attachment score of 74.7 with gold morphology and 68.1 with predicted morphology (77.8 and 72.8 unlabeled). We investigate the usefulness of different features and find that rich morphological features improve parsing accuracy significantly, by 7.5 percentage points with gold features and 5.6 points with predicted features. As expected, CASE is the single most important morphological feature, but virtually all available features bring some improvement, especially under the gold condition.

Parsing Croatian and Serbian by Using Croatian Dependency Treebanks

Željko Agić, Danijela Merkle, Daša Berović

Friday 11:25–11:50 — Blewett Suite

We investigate statistical dependency parsing of two closely related languages, Croatian and Serbian. As these two morphologically complex languages of relaxed word order are generally under-resourced – with the topic of dependency parsing still largely unaddressed, especially for Serbian – we make use of the two available dependency treebanks of Croatian to produce state-of-the-art parsing models for both languages. We observe parsing accuracy on four test sets from two domains. We give insight into overall parser performance for Croatian and Serbian, impact of preprocessing for lemmas and morphosyntactic tags and influence of selected morphosyntactic features on parsing accuracy.

A Cross-Task Flexible Transition Model for Arabic Tokenization, Affix Detection, Affix Labeling, POS Tagging, and Dependency Parsing

Stephen Tratz

Friday 11:50–12:15 — Blewett Suite

This paper describes cross-task flexible transition models (CTF-TMs) and demonstrates their effectiveness for Arabic natural language processing (NLP). NLP pipelines often suffer from error propagation, as errors committed in lower-level tasks cascade through the remainder of the processing pipeline. By allowing a

flexible order of operations across and within multiple NLP tasks, a CTF-TM can mitigate both cross-task and within-task error propagation. Our Arabic CTF-TM models tokenization, affix detection, affix labeling, part-of-speech tagging, and dependency parsing, achieving state-of-the-art results. We present the details of our general framework, our Arabic CTF-TM, and the setup and results of our experiments.

The LIGM-Alpage architecture for the SPMLR 2013 Shared Task: Multiword Expression Analysis and Dependency Parsing

Mathieu Constant, Marie Candito, Djamel Seddah

Friday 14:10–14:20 — Blewett Suite

This paper describes the LIGM-Alpage system for the SPMLR 2013 Shared Task. We only participated to the French part of the dependency parsing track, focusing on the realistic setting where the system is informed neither with gold tagging and morphology nor (more importantly) with gold grouping of tokens into multiword expressions (MWEs). While the realistic scenario of predicting both MWEs and syntax has already been investigated for constituency parsing, the SPMLR 2013 shared task datasets offer the possibility to investigate it in the dependency framework. We obtain the best results for French, both for overall parsing and for MWE recognition, using a reparsing architecture that combines several parsers, with both pipeline architecture (MWE recognition followed by parsing), and joint architecture (MWE recognition performed by the parser).

Exploring beam-based shift-reduce dependency parsing with DyALog: Results from the SPMLR 2013 shared task

Eric De La Clergerie

Friday 14:20–14:30 — Blewett Suite

The SPMLR 2013 shared task was the opportunity to develop and test, with promising results, a simple beam-based shift-reduce dependency parser on top of the tabular logic programming system DyALog. The parser was also extended to handle ambiguous word lattices, with almost no loss w.r.t. disambiguated input, thanks to specific training, use of oracle segmentation, and large beams. We believe that this result is an interesting new one for shift-reduce parsing.

Effective Morphological Feature Selection with MaltOptimizer at the SPMLR 2013 Shared Task

Miguel Ballesteros

Friday 14:30–14:55 — Blewett Suite

The inclusion of morphological features provides very useful information that helps to enhance the results when parsing morphologically rich languages. MaltOptimizer is a tool, that given a data set, searches for the optimal parameters, parsing algorithm and optimal feature set achieving the best results that it can find for parsers trained with MaltParser. In this paper, we present an extension of MaltOptimizer that explores, one by one and in combination, the features that are geared towards morphology. From our experiments in the context of the Shared Task on Parsing Morphologically Rich Languages, we extract an in-depth study that shows which features are actually useful for transition-based parsing and we provide competitive results, in a fast and simple way.

Exploiting the Contribution of Morphological Information to Parsing: the BASQUE TEAM system in the SPMLR'2013 Shared Task

Iakes Goenaga, Koldo Gojenola, Nerea Ezeiza

Friday 14:55–15:10 — Blewett Suite

This paper presents a dependency parsing system, presented as BASQUE_TEAM at the SPMLR'2013 Shared Task, based on the analysis of each morphological feature of the languages. Once the specific relevance of each morphological feature is calculated, this system uses the most significant of them to create a series of analyzers using two freely available and state of the art dependency parsers, MaltParser and Mate. Finally, the system will combine previously achieved parses using a voting approach.

The AI-KU System at the SPMRL 2013 Shared Task : Unsupervised Features for Dependency Parsing

Volkan Cirik, Hüsnü Şenşoş
Friday 15:10–15:20 — Blewett Suite

We propose the use of the word categories and embeddings induced from raw text as auxiliary features in dependency parsing. To induce word features, we make use of contextual, morphologic and orthographic properties of the words. To exploit the contextual information, we make use of substitute words, the most likely substitutes for target words, generated by using a statistical language model. We generate morphologic and orthographic properties of word types in an unsupervised manner. We use a co-occurrence model with these properties to embed words onto a 25-dimensional unit sphere. The AI-KU system shows improvements for some of the languages it is trained on for the first Shared Task of Statistical Parsing of Morphologically Rich Languages.

SPMRL'13 Shared Task System: The CADIM Arabic Dependency Parser

Yuval Marton, Nizar Habash, Owen Rambow, Sarah Alkhalani
Friday 15:20–15:30 — Blewett Suite

We describe the submission from the Columbia Arabic & Dialect Modeling group (CADIM) for the Shared Task at the Fourth Workshop on Statistical Parsing of Morphologically Rich Languages (SPMRL'2013). We participate in the Arabic Dependency parsing task for predicted POS tags and features. Our system is based on Marton et al. (2013).

A Statistical Approach to Prediction of Empty Categories in Hindi Dependency Treebank

Puneeth Kukkadapu, Prashanth Mannem
Friday 16:00–16:20 — Blewett Suite

In this paper we use statistical dependency parsing techniques to detect NULL or Empty categories in the Hindi sentences. We have currently worked on Hindi dependency treebank which is released as part of COLING-MTPIL 2012 Workshop. Earlier Rule based approaches are employed to detect Empty heads for Hindi language but statistical learning for automatic prediction is not explored. In this approach we used a technique of introducing complex labels into the data to predict Empty Categories in sentences. We have also discussed about shortcomings and difficulties in this approach and evaluated the performance of this approach on different Empty categories.

An Empirical Study on the Effect of Morphological and Lexical Features in Persian Dependency Parsing

Mojtaba Khallash, Ali Hadian, Behrouz Minaei-Bidgoli
Friday 16:00–16:20 — Blewett Suite

This paper investigates the impact of different morphological and lexical information on data-driven dependency parsing of Persian, a morphologically rich language. We explore two state-of-the-art parsers, namely MSTParser and MaltParser, on the recently released Persian dependency treebank and establish some baselines for dependency parsing performance. Three sets of issues are addressed in our experiments: effects of using gold and automatically derived features, finding the best features for the parser, and a suitable way to alleviate the data sparsity problem. The final accuracy is 87.91% and 88.37% labeled attachment scores for MaltParser and MSTParser, respectively.

Constructing a Practical Constituent Parser from a Japanese Treebank with Function Labels

Takaaki Tanaka, Masaaki NAGATA
Friday 16:00–16:20 — Blewett Suite

We present an empirical study on constructing a Japanese constituent parser, which can output function labels to deal with more detailed syntactic information. Japanese syntactic parse trees are usually represented as unlabeled dependency structure between bunsetsu chunks, however, such expression is insufficient to uncover the syntactic information about distinction between complements and adjuncts and coordination structure, which is required for practical applications such as syntactic reordering of machine translation. We describe a preliminary effort on constructing a Japanese constituent parser by a Penn Treebank style treebank semi-

automatically made from a dependency-based corpus. The evaluations show the parser trained on the treebank has comparable bracketing accuracy as conventional bunsetsu-based parsers, and can output such function labels as the grammatical role of the argument and the type of adnominal phrases.

Context Based Statistical Morphological Analyzer and its Effect on Hindi Dependency Parsing

Deepak Kumar Malladi, Prashanth Mannem
Friday 16:00–16:20 — Blewett Suite

This paper revisits the work of (Malladi and Mannem, 2013) which focused on building a Statistical Morphological Analyzer (SMA) for Hindi and compares the performance of SMA with other existing statistical analyzer, Morfette. We shall evaluate SMA in various experiment scenarios and look at how it performs for unseen words. The later part of the paper presents the effect of the predicted morph features on dependency parsing and extends the work to other morphologically rich languages: Hindi and Telugu, without any language-specific engineering.

Representation of Morphosyntactic Units and Coordination Structures in the Turkish Dependency Treebank

Umut Sulubacak, Gülşen Eryiğit
Friday 16:00–16:20 — Blewett Suite

This paper presents our preliminary conclusions as part of an ongoing effort to construct a new dependency representation framework for Turkish. We aim for this new framework to accommodate the highly agglutinative morphology of Turkish as well as to allow the annotation of unedited web data, and shape our decisions around these considerations. In this paper, we firstly describe a novel syntactic representation for morphosyntactic subword units (namely inflectional groups (IGs) in Turkish) which allows inter-IG relations to be discerned with perfect accuracy without having to hide lexical information. Secondly, we investigate alternative annotation schemes for coordination structures and present a better scheme (nearly 11% increase in recall scores) than the one in Turkish Treebank (Oflazler et al., 2003) for both parsing accuracies and compatibility for colloquial language.

(Re)ranking Meets Morphosyntax: State-of-the-art Results from the SPMRL 2013 Shared Task

Anders Björkelund, Ozlem Cetinoglu, Richárd Farkas, Thomas Mueller, Wolfgang Seeker
Friday 16:20–16:45 — Blewett Suite

This paper describes the IMS-SZEGED-CIS contribution to the SPMRL 2013 Shared Task. We participate in both the constituency and dependency tracks, and achieve state-of-the-art for all languages. For both tracks we make significant improvements through high quality preprocessing and (re)ranking on top of strong baselines. Our system came out first for both tracks.

Invited talk: Dependency Parsing with Selectional Branching

Jinho D. Choi
Friday 16:45–17:10 — Blewett Suite

Overview of the SPMRL 2013 Shared Task: A Cross-Framework Evaluation of Parsing Morphologically Rich Languages

Djamé Seddah, Reut Tsarfaty, Sandra Kübler, Marie Candito, Jinho D. Choi, Richárd Farkas, Jennifer Foster, Iakes Goenaga, Koldo Gojenola Gallettebeitia, Yoav Goldberg, Spence Green, Nizar Habash, Marco Kuhlmann, Wolfgang Maier, Yuval Marton, Joakim Nivre, Adam Przepiórkowski, Ryan Roth, Wolfgang Seeker, Yannick Versley, Veronika Vincze, Marcin Woliński, Alina Wróblewska
Friday 18:00–18:15 — Blewett Suite

This paper reports on the first shared task on statistical parsing of Morphologically Rich Languages (MRLs). The task features data sets from nine languages, each available both in constituency and dependency annotation. We report on the alignment of the data sets, on the proposed parsing scenarios, and on the evaluation

metrics for parsing MRLs given different representation types. We present and analyze parsing results obtained by the task participants, and then provide an analysis and cross- comparison of the parsers across languages and frameworks, reported for gold input as well as more realistic parsing scenarios.

Twenty Years of Bitext

Program Chairs: Chris Dyer, Noah A. Smith, Phil Blunsom

What a Translation Model Tastes Like

Kevin Knight, USC/ISI
Friday 9:30–10:00 — Portland

Between Machine Translation and Human Translation

Philipp Koehn, Edinburgh
Friday 10:00–10:30 — Portland

Oh, Yes, Everything's Right on Schedule, Fred

Peter F. Brown and Robert L. Mercer, Renaissance Technologies
Friday 11:00–12:00 — Portland

Beyond Bitext: Five Open Problems in Machine Translation

Adam Lopez and Matt Post
Friday 3:00–4:00 — Portland

Twenty Flavors of One Text

Daniel Zeman and Ondřej Bojar
Friday 3:00–4:00 — Portland

What SMT Learns

Dekai Wu
Friday 3:00–4:00 — Portland

Violation-Fixing Perceptron and Forced Decoding for Scalable MT Training

Heng Yu, Liang Huang, and Haito Mi
Friday 3:00–4:00 — Portland

Aligning Words in Bitexts using the Bilingual Web

Jim Chang, Joseph Chee Chang, Jian-cheng Wu, and Jason S. Chang
Friday 3:00–4:00 — Portland

The Pedantic Javadoc Corpus: Comments and Code as Bitext

Jim White
Friday 3:00–4:00 — Portland

Bitexts as Semantic Mirrors

Jörg Tiedemann, Lonneke van der Plas, and Begoña Villada Moirón
Friday 3:00–4:00 — Portland

Using parallel corpora to bootstrap multilingual semantic parsers

Kilian Evang and Johan Bos
Friday 3:00–4:00 — Portland

Bitext as Word Sense Annotation: Lessons from Evaluating Machine Translation on Lexical Semantics Tasks

Marine Carpuat
Friday 3:00–4:00 — Portland

Beyond Parallel Data – A Decipherment Approach to Machine Translation

Qing Dou and Kevin Knight
Friday 3:00–4:00 — Portland

On Applying and Extending Bitext for Entity Translation

Seung-won Hwang
Friday 3:00–4:00 — Portland

Lexicalized Reordering Model in Chart-based Machine Translation

ThuyLinh Nguyen
Friday 3:00–4:00 — Portland

Discrete Log-Linear Autoencoders for Unsupervised Learning of Linguistic Structure

Waleed Ammar, Chris Dyer, and Noah A. Smith
Friday 3:00–4:00 — Portland

Google Translate: Past, Present, Future

Franz Josef Och, Google
Friday 4:00–4:30 — Portland

Saturday, October 19, 2013: Main Conference

Overview

7:30 – 9:00	Breakfast	(Leonesa Foyer)
9:00 – 9:15	Opening Remarks	(Leonesa Ballroom)
9:20 – 10:30	Parallel Sessions <i>Information Extraction I</i> <i>Language Acquisition and Processing</i> <i>NLP for Social Media I</i>	(Leonesa I and II) (Leonesa III) (Eliza Anderson Amphitheater)
10:30–11:00	Break	(Leonesa Foyer)
11:00–12:35	Parallel Sessions <i>NLP Applications I</i> <i>Semantics I</i> <i>Language Resources</i>	(Leonesa I and II) (Leonesa III) (Eliza Anderson Amphitheater)
12:35– 2:00	Lunch	
2:00 – 3:15	Invited talk: Andrew Ng	(Leonesa Ballroom)
3:15 – 3:45	Break	(Leonesa Foyer)
3:45 – 5:50	Parallel Sessions <i>Machine Translation I</i> <i>Dialogue and Discourse</i> <i>Morphology and Phonology</i>	(Leonesa I and II) (Leonesa III) (Eliza Anderson Amphitheater)
6:00 – 7:30	Poster Session A	(Princessa Ballroom and Foyer)
7:30 – 9:00	Poster Session B	(Princessa Ballroom and Foyer)

Information Extraction I

Leonesa I and II

Chair: Steven Bethard

Event-Based Time Label Propagation for Automatic Dating of News Articles

Tao Ge, Baobao Chang, Sujian Li, Zhifang Su

Saturday 9:20–9:45 — Leonesa I and II

Since many applications such as timeline summaries and temporal IR involving temporal analysis rely on document timestamps, the task of automatic dating of documents has been increasingly important. Instead of using feature-based methods as conventional models, our method attempts to date documents in a year level by exploiting relative temporal relations between documents and events, which are very effective for dating documents. Based on this intuition, we proposed an event-based time label propagation model called confidence boosting in which time label information can be propagated between documents and events on a bipartite graph. The experiments show that our event-based propagation model can predict document timestamps in high accuracy and the model combined with a MaxEnt classifier outperforms the state-of-the-art method for this task especially when the size of the training set is small.

Exploiting Discourse Analysis for Article-Wide Temporal Classification

Jun-Ping Ng, Min-Yen Kan, Ziheng Lin, Wei Feng, Bin Chen, Jian Su, Chew Lim Tan

Saturday 9:45–10:10 — Leonesa I and II

In this paper we classify the temporal relations between pairs of events on an article-wide basis. This is in contrast to much of the existing literature which focuses on just event pairs which are found within the same or adjacent sentences. To achieve this, we leverage on discourse analysis as we believe that it provides more useful semantic information than typical lexico-syntactic features. We propose the use of several discourse analysis frameworks, including 1) Rhetorical Structure Theory (RST), 2) PDTB-styled discourse relations, and 3) topical text segmentation. We explain how features derived from these frameworks can be effectively used with support vector machines (SVM) paired with convolution kernels. Experiments show that our proposal is effective in improving on the state-of-the-art significantly by as much as 16% in terms of F1, even if we only adopt less-than-perfect automatic discourse analyzers and parsers. Making use of more accurate discourse analysis can further boost gains to 35%.

Combining Generative and Discriminative Model Scores for Distant Supervision

Benjamin Roth, Dietrich Klakow

Saturday 10:10–10:30 — Leonesa I and II

Distant supervision is a scheme to generate noisy training data for relation extraction by aligning entities of a knowledge base with text. In this work we combine the output of a discriminative at-least-one learner with that of a generative hierarchical topic model to reduce the noise in distant supervision data. The combination significantly increases the ranking quality of extracted facts and achieves state-of-the-art extraction performance in an end-to-end setting. A simple linear interpolation of the model scores performs better than a parameter-free scheme based on non-dominated sorting.

Language Acquisition and Processing

Leonesa III

Chair: Julia Hockenmaier

Exploring the Utility of Joint Morphological and Syntactic Learning from Child-directed Speech

Stella Frank, Frank Keller, Sharon Goldwater

Saturday 9:20–9:45 — Leonesa III

Children learn various levels of linguistic structure concurrently, yet most existing models of language acquisition deal with only a single level of structure, implicitly assuming a sequential learning process. Developing models that learn multiple levels simultaneously can provide important insights into how these levels might interact synergistically during learning. Here, we present a model that jointly induces syntactic categories and morphological segmentations by combining two well-known models for the individual tasks. We test on child-directed utterances in English and Spanish and compare to single-task baselines. In the morphologically poorer language (English), the model improves morphological segmentation, while in the morphologically richer language (Spanish), it leads to better syntactic categorization. These results provide further evidence that joint learning is useful, but also suggest that the benefits may be different for typologically different languages.

A Joint Learning Model of Word Segmentation, Lexical Acquisition, and Phonetic Variability

Micha Elsner, Sharon Goldwater, Naomi Feldman, Frank Wood

Saturday 9:45–10:10 — Leonesa III

We present a cognitive model of early lexical acquisition which jointly performs word segmentation and learns an explicit model of phonetic variation. We define the model as a Bayesian noisy channel; we sample segmentations and word forms simultaneously from the posterior, using beam sampling to control the size of the search space. Compared to a pipelined approach in which segmentation is performed first, our model is qualitatively more similar to human learners. On data with variable pronunciations, the pipelined approach learns to treat syllables or morphemes as words. In contrast, our joint model, like infant learners, tends to learn multiword collocations. We also conduct analyses of the phonetic variations that the model learns to accept and its patterns of word recognition errors, and relate these to developmental evidence.

Animacy Detection with Voting Models

Joshua Moore, Christopher J.C. Burges, Erin Renshaw, Wen-tau Yih

Saturday 10:10–10:30 — Leonesa III

Animacy detection is a problem whose solution has been shown to be beneficial for a number of syntactic and semantic tasks. We present a state-of-the-art system for this task which uses a number of simple classifiers with heterogeneous data sources in a voting scheme. We show how this framework can give us direct insight into the behavior of the system, allowing us to more easily diagnose sources of error.

NLP for Social Media I

Eliza Anderson Amphitheater

Chair: Noah Smith

A Log-Linear Model for Unsupervised Text Normalization

Yi Yang, Jacob Eisenstein

Saturday 9:20–9:45 — Eliza Anderson Amphitheater

We present a unified unsupervised statistical model for text normalization. The relationship between standard and non-standard tokens is characterized by a log-linear model, permitting arbitrary features. The weights of these features are trained in a maximum-likelihood framework, employing a novel sequential Monte Carlo training algorithm to overcome the large label space, which would be impractical for traditional dynamic programming solutions. This model is implemented in a normalization system called unLOL, which achieves the best known results on two normalization datasets, outperforming more complex systems. We use the output of unLOL to automatically normalize a large corpus of social media text, revealing a set of coherent orthographic styles that underlie online language variation.

Paraphrasing 4 Microblog Normalization

Wang Ling, Chris Dyer, Alan W Black, Isabel Trancoso

Saturday 9:45–10:10 — Eliza Anderson Amphitheater

Compared to the edited genres that have played a central role in NLP research, microblog texts use a more informal register with nonstandard lexical items, abbreviations, and free orthographic variation. When confronted with such input, conventional text analysis tools often perform poorly. Normalization — replacing orthographically or lexically idiosyncratic forms with more standard variants — can improve performance. We propose a method for learning normalization rules from machine translations of a parallel corpus of microblog messages. To validate the utility of our approach, we evaluate extrinsically, showing that normalizing English tweets and then translating improves translation quality (compared to translating unnormalized text) using three standard web translation services as well as a phrase-based translation system trained on parallel microblog data.

Question Difficulty Estimation in Community Question Answering Services

Jing Liu, Quan Wang, Chin-Yew Lin, Hsiao-Wuen Hon

Saturday 10:10–10:30 — Eliza Anderson Amphitheater

In this paper, we address the problem of estimating question difficulty in community question answering services. We propose a competition-based model for estimating question difficulty by leveraging pairwise comparisons between questions and users. Our experimental results show that our model significantly outperforms a PageRank-based approach. Most importantly, our analysis shows that the text of question descriptions reflects the question difficulty. This implies the possibility of predicting question difficulty from the text of question descriptions.

NLP Applications I

Leonesa I and II

Chair: Hang Li

Measuring Ideological Proportions in Political Speeches

Yanchuan Sim, Brice D. L. Acree, Justin H. Gross, Noah A. Smith

Saturday 11:00–11:25 — Leonesa I and II

We seek to measure political candidates' ideological positioning from their speeches. To accomplish this, we infer ideological cues from a corpus of political writings annotated with known ideologies. We then represent the speeches of U.S. Presidential candidates as sequences of cues and lags (filler distinguished only by its length in words). We apply a domain-informed Bayesian HMM to infer the proportions of ideologies each candidate uses in each campaign. The results are validated against a set of preregistered, domain expert-authored hypotheses.

Learning to Freestyle: Hip Hop Challenge-Response Induction via Transduction Rule Segmentation

Dekai Wu, Karteek Addanki, Markus Saers, Meriem Beloucif

Saturday 11:25–11:50 — Leonesa I and II

We present a novel model that learns to improvise rhyming and fluent responses upon being challenged with a line of hip hop lyrics, by combining both bottom-up token based rule induction and top-down rule segmentation strategies to learn a stochastic transduction grammar that simultaneously learns both phrasing and rhyming associations. In this attack on the woefully under-explored natural language genre of music lyrics, we exploit a strictly unsupervised transduction grammar induction approach. Our task is particularly ambitious in that no use of any a priori linguistic or phonetic information is allowed, even though the domain of hip hop lyrics is particularly noisy and unstructured. We evaluate the performance of the learned model against a model learned only using the more conventional bottom-up token based rule induction, and demonstrate the superiority of our combined token based and rule segmentation induction method toward generating higher quality improvised responses, measured on fluency and rhyming criteria as judged by human evaluators. To highlight some of the inherent challenges in adapting other algorithms to this novel task, we also compare the quality of the responses generated by our model to those generated by an out-of-the-box phrase based SMT system. We tackle the challenge of selecting appropriate training data for our task via a dedicated rhyme scheme detection module, which is also acquired via unsupervised learning and report improved quality of the generated responses. Finally, we report results with French hip hop lyrics indicating that our model performs surprisingly well with no special adaptation to other languages.

Modeling Scientific Impact with Topical Influence Regression

James Foulds, Padhraic Smyth

Saturday 11:50–12:15 — Leonesa I and II

When reviewing scientific literature, it would be useful to have automatic tools that identify the most influential scientific articles as well as how ideas propagate between articles. In this context, this paper introduces topical influence, a quantitative measure of the extent to which an article tends to spread its topics to the articles that cite it. Given the text of the articles and their citation graph, we show how to learn a probabilistic model to recover both the degree of topical influence of each article and the influence relationships between articles. Experimental results on corpora from two well-known computer science conferences are used to illustrate and validate the proposed approach.

Joint Parsing and Disfluency Detection in Linear Time

Mohammad Sadeq Rasooli, Joel Tetreault

Saturday 12:15–12:35 — Leonesa I and II

We introduce a novel method to jointly parse and detect disfluencies in spoken utterances. Our model can use arbitrary features for parsing sentences and adapt itself with out-of-domain data. We show that our method, based on transition-based parsing, performs at a high level of accuracy for both the parsing and disfluency detection tasks. Additionally, our method is the fastest for the joint task, running in linear time.

Semantics I

Leonesa III

Chair: Luke Zettlemoyer

Modeling and Learning Semantic Co-Compositionality through Prototype Projections and Neural Networks

Masashi Tsubaki, Kevin Duh, Masashi Shimbo, Yuji Matsumoto

Saturday 11:00–11:25 — Leonesa III

We present a novel vector space model for semantic co-compositionality. Inspired by Generative Lexicon Theory (Pustejovsky, 1995), our goal is a compositional model where both predicate and argument are allowed to modify each others' meaning representations while generating the overall semantics. This readily addresses some major challenges with current vector space models, notably the polysemy issue and the use of one representation per word type. We implement co-compositionality using prototype projections on predicates/arguments and show that this is effective in adapting their word representations. We further cast the model as a neural network and propose an unsupervised algorithm to jointly train word representations with co-compositionality. The model achieves the best result to date ($\rho = 0.47$) on the semantic similarity task of transitive verbs (Grefenstette and Sadrzadeh, 2011).

Studying the Recursive Behaviour of Adjectival Modification with Compositional Distributional Semantics

Eva Maria Vecchi, Roberto Zamparelli, Marco Baroni

Saturday 11:25–11:50 — Leonesa III

In this study, we use compositional distributional semantic methods to investigate restrictions in adjective ordering. Specifically, we focus on properties distinguishing Adjective-Adjective-Noun phrases in which there is flexibility in the adjective ordering from those bound to a rigid order. We explore a number of measures extracted from the distributional representation of AAN phrases which may indicate a word order restriction. We find that we are able to distinguish the relevant classes and the correct order based primarily on the degree of modification of the adjectives. Our results offer fresh insight into the semantic properties that determine adjective ordering, building a bridge between syntax and distributional semantics.

Learning Latent Word Representations for Domain Adaptation using Supervised Word Clustering

Min Xiao, Feipeng Zhao, Yuhong Guo

Saturday 11:50–12:15 — Leonesa III

Domain adaptation has been popularly studied on exploiting labeled information from a source domain to learn a prediction model in a target domain. In this paper, we develop a novel representation learning approach to address domain adaptation for text classification with automatically induced discriminative latent features, which are generalizable across domains while informative to the prediction task. Specifically, we propose a hierarchical multinomial Naive Bayes model with latent variables to conduct supervised word clustering on labeled documents from both source and target domains, and then use the produced cluster distribution of each word as its latent feature representation for domain adaptation. We train this latent graphical model using a simple expectation-maximization (EM) algorithm. We empirically evaluate the proposed method with both cross-domain document categorization tasks on Reuters-21578 dataset and cross-domain sentiment classification tasks on Amazon product review dataset. The experimental results demonstrate that our proposed approach achieves superior performance compared with alternative methods.

Appropriately Incorporating Statistical Significance in PMI

Om P. Damani, Shweta Ghonge

Saturday 12:15–12:35 — Leonesa III

Two recent measures incorporate the notion of statistical significance in basic PMI formulation. In some tasks, we find that the new measures perform worse than the PMI. Our analysis shows that while the basic ideas in incorporating statistical significance in PMI are reasonable, they have been applied slightly inappropriately. By fixing this, we get new measures that improve performance over not just PMI but on other popular co-

occurrence measures as well. In fact, the revised measures perform reasonably well compared with more resource intensive non co-occurrence based methods also.

Language Resources

Eliza Anderson Amphitheater

Chair: Emily Bender

Growing Multi-Domain Glossaries from a Few Seeds using Probabilistic Topic Models

Stefano Faralli, Roberto Navigli

Saturday 11:00–11:25 — Eliza Anderson Amphitheater

In this paper we present a minimally-supervised approach to the multi-domain acquisition of wide-coverage glossaries. We start from a small number of hypernymy relation seeds and bootstrap glossaries from the Web for dozens of domains using Probabilistic Topic Models. Our experiments show that we are able to extract high-precision glossaries comprising thousands of terms and definitions.

Joint Learning of Phonetic Units and Word Pronunciations for ASR

Chia-ying Lee, Yu Zhang, James Glass

Saturday 11:25–11:50 — Eliza Anderson Amphitheater

The creation of a pronunciation lexicon remains the most inefficient process in developing an Automatic Speech Recognizer (ASR). In this paper, we propose an unsupervised alternative - requiring no language-specific knowledge - to the conventional manual approach for creating pronunciation dictionaries. We present a hierarchical Bayesian model, which jointly discovers the phonetic inventory and the Letter-to-Sound (L2S) mapping rules in a language using only transcribed data. When tested on a corpus of spontaneous queries, the results demonstrate the superiority of the proposed joint learning scheme over its sequential counterpart, in which the latent phonetic inventory and L2S mappings are learned separately. Furthermore, the recognizers built with the automatically induced lexicon consistently outperform grapheme-based recognizers and even approach the performance of recognition systems trained using conventional supervised procedures.

MCTest: A Challenge Dataset for the Open-Domain Machine Comprehension of Text

Matthew Richardson, Christopher J.C. Burges, Erin Renshaw

Saturday 11:50–12:15 — Eliza Anderson Amphitheater

We present MCTest, a freely available set of stories and associated questions intended for research on the machine comprehension of text. Previous work on machine comprehension (e.g., semantic modeling) has made great strides, but primarily focuses either on limited-domain datasets, or on solving a more restricted goal (e.g., open-domain relation extraction). In contrast, MCTest requires machines to answer multiple-choice reading comprehension questions about fictional stories, directly tackling the high-level goal of open-domain machine comprehension. Reading comprehension can test advanced abilities such as causal reasoning and understanding the world, yet, by being multiple-choice, still provide a clear metric. By being fictional, the answer typically can be found only in the story itself. The stories and questions are also carefully limited to those a young child would understand, reducing the world knowledge that is required for the task. We present the scalable crowd-sourcing methods that allow us to cheaply construct a dataset of 500 stories and 2000 questions. By screening workers (with grammar tests) and stories (with grading), we have ensured that the data is the same quality as another that we manually edited, but at one tenth the editing cost. By being open-domain, yet carefully restricted, we hope MCTest will serve to encourage research and provide a clear metric for advancement on the machine comprehension of text.

Noise-Aware Character Alignment for Bootstrapping Statistical Machine Transliteration from Bilingual Corpora

Katsuhito Sudoh, Shinsuke Mori, Masaaki Nagata

Saturday 12:15–12:35 — Eliza Anderson Amphitheater

This paper proposes a novel noise-aware character alignment method for bootstrapping statistical machine transliteration from automatically extracted phrase pairs. The model is an extension of a Bayesian many-to-many alignment method for distinguishing non-transliteration (noise) parts in phrase pairs. It worked effectively in the experiments of bootstrapping Japanese-to-English statistical machine transliteration in patent domain using patent bilingual corpora.

Invited talk: Andrew Ng

Leonessa Ballroom

The Online Revolution: Education for Everyone
Dr Andrew Ng, Co-CEO and Co-founder of Coursera
Saturday 2:00–3:15 — Leonessa Ballroom

In 2011, Stanford University offered three online courses, which anyone in the world could enroll in and take for free. Together, these three courses had enrollments of around 350,000 students, making this one of the largest experiments in online education ever performed.

Since the beginning of 2012, we have transitioned this effort into a new venture, Coursera, a social entrepreneurship company whose mission is to make high-quality education accessible to everyone by allowing the best universities to offer courses to everyone around the world, for free. Coursera classes provide a real course experience to students, including video content, interactive exercises with meaningful feedback, using both auto-grading and peer-grading, and a rich peer-to-peer interaction around the course materials.

Currently, Coursera has 80 university and other partners, and 3.6 million students enrolled in its nearly 400 courses. These courses span a range of topics including computer science, business, medicine, science, humanities, social sciences, and more.

In this talk, I'll report on this far-reaching experiment in education, and why we believe this model can provide both an improved classroom experience for our on-campus students, via a flipped classroom model, as well as a meaningful learning experience for the millions of students around the world who would otherwise never have access to education of this quality.

Machine Translation I

Leonesa I and II

Chair: Kevin Knight

Optimal Beam Search for Machine Translation

Alexander Rush, Yin-Wen Chang, Michael Collins

Saturday 3:45–4:10 — Leonesa I and II

Beam search is a fast and empirically effective method for translation decoding, but it lacks any formal guarantees about search error. We develop a new decoding algorithm that combines the speed of beam search with the optimal certificate property of Lagrangian relaxation, and apply it to phrase- and syntax-based translation decoding. The new method is efficient, utilizes standard MT algorithms, and returns an exact solution on the majority of translation examples in our test data. The algorithm is 3.5 times faster than a Lagrangian relaxation-based method for phrase-based translation and 4 times faster for syntax-based translation.

An Efficient Language Model Using Double-Array Structures

Makoto Yasuhara, Toru Tanaka, Jun-ya Norimatsu, Mikio Yamamoto

Saturday 4:10–4:35 — Leonesa I and II

Ngram language models tend to increase in size with inflating the corpus size, and consume considerable resources. In this paper, we propose an efficient method for implementing ngram models based on double-array structures. First, we propose a method for representing backwards suffix trees using double-array structures and demonstrate its efficiency. Next, we propose two optimization methods for improving the efficiency of data representation in the double-array structures. Embedding probabilities into unused spaces in double-array structures reduces the model size. Moreover, tuning the word IDs in the language model makes the model smaller and faster. We also show that our method can be used for building large language models using the division method. Lastly, we show that our method outperforms methods based on recent related works from the viewpoints of model size and query speed when both optimization methods are used.

Structured Penalties for Log-Linear Language Models

Anil Kumar Nelakanti, Cedric Archambeau, Julien Mairal, Francis Bach, Guillaume Bouchard

Saturday 4:35–5:00 — Leonesa I and II

Language models can be formalized as log-linear regression models where the input features represent previously observed contexts up to a certain length m . The complexity of existing algorithms to learn the parameters by maximum likelihood scale linearly in nd , where n is the length of the training corpus and d is the number of observed features. We present a model that grows logarithmically in d , making it possible to efficiently leverage longer contexts. We account for the sequential structure of natural language using tree-structured penalized objectives to avoid overfitting and achieve better generalization.

Interactive Machine Translation using Hierarchical Translation Models

Jesús González-Rubio, Daniel Ortíz-Martínez, José-Miguel Benedí, Francisco Casacuberta

Saturday 5:00–5:25 — Leonesa I and II

Current automatic machine translation systems are not able to generate error-free translations and human intervention is often required to correct their output. Alternatively, an interactive framework that integrates the human knowledge into the translation process has been presented in previous works. Here, we describe a new interactive machine translation approach that is able to work with phrase-based and hierarchical translation models, and integrates error-correction all in a unified statistical framework. In our experiments, our approach outperforms previous interactive translation systems, and achieves relative effort reductions of as much as 48% over a traditional post-edition system.

Max-Margin Synchronous Grammar Induction for Machine Translation

Xinyan Xiao, Deyi Xiong

Saturday 5:25–5:50 — Leonesa I and II

Traditional synchronous grammar induction estimates parameters by maximizing likelihood, which only has a loose relation to translation quality. Alternatively, we propose a max-margin estimation approach to discrimi-

natively inducing synchronous grammars for machine translation, which directly optimizes translation quality measured by BLEU. In the max-margin estimation of parameters, we only need to calculate Viterbi translations. This further facilitates the incorporation of various non-local features that are defined on the target side. We test the effectiveness of our max-margin estimation framework on a competitive hierarchical phrase-based system. Experiments show that our max-margin method significantly outperforms the traditional two-step pipeline for synchronous rule extraction by 1.3 BLEU points and is also better than previous max-likelihood estimation method.

Dialogue and Discourse

Leonesa III

Chair: Eugene Charniak

Error-Driven Analysis of Challenges in Coreference Resolution

Jonathan K. Kummerfeld, Dan Klein

Saturday 3:45–4:10 — Leonesa III

Coreference resolution metrics quantify errors but do not analyze them. Here, we consider an automated method of categorizing errors in the output of a coreference system into intuitive underlying error types. Using this tool, we first compare the error distributions across a large set of systems, then analyze common errors across the top ten systems, empirically characterizing the major unsolved challenges of the coreference resolution task.

Exploiting Zero Pronouns to Improve Chinese Coreference Resolution

Fang Kong, Hwee Tou Ng

Saturday 4:10–4:35 — Leonesa III

Coreference resolution plays a critical role in discourse analysis. This paper focuses on exploiting zero pronouns to improve Chinese coreference resolution. In particular, a simplified semantic role labeling framework is proposed to identify clauses and to detect zero pronouns effectively, and two effective methods (refining syntactic parser and refining learning example generation) are employed to exploit zero pronouns for Chinese coreference resolution. Evaluation on the CoNLL-2012 shared task data set shows that zero pronouns can significantly improve Chinese coreference resolution.

Joint Coreference Resolution and Named-Entity Linking with Multi-Pass Sieves

Hannaneh Hajishirzi, Leila Zilles, Daniel S. Weld, Luke Zettlemoyer

Saturday 4:35–5:00 — Leonesa III

Many errors in coreference resolution come from semantic mismatches due to inadequate world knowledge. Errors in named-entity linking (NEL), on the other hand, are often caused by superficial modeling of entity context. This paper demonstrates that these two tasks are complementary. We introduce NECo, a new model for named entity linking and coreference resolution, which solves both problems jointly, reducing the errors made on each. NECo extends the Stanford deterministic coreference system by automatically linking mentions to Wikipedia and introducing new NEL-informed mention-merging sieves. Linking improves mention-detection and enables new semantic attributes to be incorporated from Freebase, while coreference provides better context modeling by propagating named-entity links within mention clusters. Experiments show consistent improvements across a number of datasets and experimental conditions, including over 11% reduction in MUC coreference error and nearly 21% reduction in F1 NEL error on ACE 2004 newswire data.

Interpreting Anaphoric Shell Nouns using Antecedents of Cataphoric Shell Nouns as Training Data

Varada Kolhatkar, Heike Zinsmeister, Graeme Hirst

Saturday 5:00–5:25 — Leonesa III

Interpreting anaphoric shell nouns (ASNs) such as "this issue" and "this fact" is essential to understanding virtually any substantial natural language text. One obstacle in developing methods for automatically interpreting ASNs is the lack of annotated data. We tackle this challenge by exploiting cataphoric shell nouns (CSNs) whose construction makes them particularly easy to interpret (e.g., "the fact that X"). We propose an approach that uses automatically extracted antecedents of CSNs as training data to interpret ASNs. We achieve precisions in the range of 0.35 (baseline = 0.21) to 0.72 (baseline = 0.44), depending upon the shell noun.

Parsing Entire Discourses as Very Long Strings: Capturing Topic Continuity in Grounded Language Learning [TACL]

Minh-Thang Luong, Michael C. Frank, Mark Johnson

Saturday 5:25–5:50 — Leonesa III

Grounded language learning, the task of mapping from natural language to a representation of meaning, has attracted more and more interest in recent years. In most work on this topic, however, utterances in a conversation are treated independently and discourse structure information is largely ignored. In the context of language acquisition, this independence assumption discards cues that are important to the learner, e.g., the fact that consecutive utterances are likely to share the same referent (Frank et al., 2013). The current paper describes an approach to the problem of simultaneously modeling grounded language at the sentence and discourse levels. We combine ideas from parsing and grammar induction to produce a parser that can handle long input strings with thousands of tokens, creating parse trees that represent full discourses. By casting grounded language learning as a grammatical inference task, we use our parser to extend the work of Johnson et al. (2012), investigating the importance of discourse continuity in children’s language acquisition and its interaction with social cues. Our model boosts performance in a language acquisition task and yields good discourse segmentations compared with human annotators.

Morphology and Phonology

Eliza Anderson Amphitheater

Chair: Anders Sogaard

Exploring Representations from Unlabeled Data with Co-training for Chinese Word Segmentation

Longkai Zhang, Houfeng Wang, Xu Sun, Mairgup Mansur

Saturday 3:45–4:10 — Eliza Anderson Amphitheater

Nowadays supervised sequence labeling models can reach competitive performance on the task of Chinese word segmentation. However, the ability of these models is restricted by the availability of annotated data and the design of features. We propose a scalable semi-supervised feature engineering approach. In contrast to previous works using pre-defined task-specific features with fixed values, we dynamically extract representations of label distributions from both an in-domain corpus and an out-of-domain corpus. We update the representation values with a semi-supervised approach. Experiments on the benchmark datasets show that our approach achieves good results and reaches an f-score of 0.961. The feature engineering approach proposed here is a general iterative semi-supervised method and not limited to the word segmentation task.

Efficient Higher-Order CRFs for Morphological Tagging

Thomas Mueller, Helmut Schmid, Hinrich Schütze

Saturday 4:10–4:35 — Eliza Anderson Amphitheater

Training higher-order conditional random fields is prohibitive for huge tag sets. We present an approximated conditional random field using coarse-to-fine decoding and early updating. We show that our implementation yields fast and accurate morphological taggers across six languages with different morphological properties and that across languages higher-order models give significant improvements over first-order models.

The Effects of Syntactic Features in Automatic Prediction of Morphology

Wolfgang Seeker, Jonas Kuhn

Saturday 4:35–5:00 — Eliza Anderson Amphitheater

Morphology and syntax interact considerably in many languages and language processing should pay attention to these interdependencies. We analyze the effect of syntactic features when used in automatic morphology prediction on four typologically different languages. We show that predicting morphology for languages with highly ambiguous word forms profits from taking the syntactic context of words into account and results in state-of-the-art models.

Adaptor Grammars for Learning Non-Concatenative Morphology

Jan A. Botha, Phil Blunsom

Saturday 5:00–5:25 — Eliza Anderson Amphitheater

This paper contributes an approach for expressing non-concatenative morphological phenomena, such as stem derivation in Semitic languages, in terms of a mildly context-sensitive grammar formalism. This offers a convenient level of modelling abstraction while remaining computationally tractable. The nonparametric Bayesian framework of adaptor grammars is extended to this richer grammar formalism to propose a probabilistic model that can learn word segmentation and morpheme lexicons, including ones with discontinuous strings as elements, from unannotated data. Our experiments on Hebrew and three variants of Arabic data find that the additional expressiveness to capture roots and templates as atomic units improves the quality of concatenative segmentation and stem identification. We obtain 74% accuracy in identifying trilateral Hebrew roots, while performing morphological segmentation with an F1-score of 78.1.

Poster Session A

Princessa Ballroom and Foyer

Grounding Strategic Conversation: Using Negotiation Dialogues to Predict Trades in a Win-Lose Game

Anais Cadilhac, Nicholas Asher, Farah Benamara, Alex Lascarides

Saturday 6:00–7:30 — Princessa Ballroom and Foyer

This paper describes a method that predicts which trades players execute during a win-lose game. Our method uses data collected from chat negotiations of the game "The Settlers of Catan". We use this corpus to learn classifiers that map each utterance to its speech act and to other acts that are pertinent to bargaining. And we develop a symbolic algorithm that, from the classifiers' output, dynamically constructs a model of each player's preferences as the conversation proceeds (for instance, the preference to receive a certain resource, or to accept a certain trade). This preference model is based on CP-nets (Boutilier et al., 2004), a logic of preferences for which algorithms for computing equilibrium strategies exist. We adapt those algorithms to predict whether a trade is executed as a result of the players' negotiations. A comparison of our method against 4 baselines shows that tracking how preferences evolve through the dialogue and reasoning about equilibrium moves are both crucial to success.

Unsupervised Induction of Contingent Event Pairs from Film Scenes

Zhichao Hu, Elahe Rahimtoroghi, Larissa Munishkina, Reid Swanson, Marilyn A. Walker

Saturday 6:00–7:30 — Princessa Ballroom and Foyer

Human engagement in narrative is partially driven by reasoning about discourse relations between narrative events, and the expectations about what is likely to happen next that results from such reasoning. Researchers in NLP have tackled modeling such expectations from a range of perspectives, including treating it as the inference of the CONTINGENT discourse relation, or as a type of common-sense reasoning. Our approach is to model likelihood between events by drawing on several of these lines of previous work. We implement and evaluate different unsupervised methods for learning event pairs that are likely to be CONTINGENT on one another. We refine event pairs that we learn from a corpus of film scene descriptions utilizing Web search counts, and evaluate our results by collecting human judgments of contingency. Our results indicate that the use of Web search counts increases the average accuracy of our best method to 85.64% over a baseline of 50%, as compared to an average accuracy of 75.15% without web search.

Latent Anaphora Resolution for Cross-Lingual Pronoun Prediction

Christian Hardmeier, Jörg Tiedemann, Joakim Nivre

Saturday 6:00–7:30 — Princessa Ballroom and Foyer

This paper addresses the task of predicting the correct French translations of third-person subject pronouns in English discourse, a problem that is relevant as a prerequisite for machine translation and that requires anaphora resolution. We present an approach based on neural networks that models anaphoric links as latent variables and show that its performance is competitive with that of a system with separate anaphora resolution while not requiring any coreference-annotated training data. This demonstrates that the information contained in parallel bitexts can successfully be used to acquire knowledge about pronominal anaphora in an unsupervised way.

Towards Situated Dialogue: Revisiting Referring Expression Generation

Rui Fang, Changsong Liu, Lanbo She, Joyce Y. Chai

Saturday 6:00–7:30 — Princessa Ballroom and Foyer

In situated dialogue, humans and agents have mismatched capabilities of perceiving the shared environment. Their representations of the shared world are misaligned. Thus referring expression generation (REG) will need to take this discrepancy into consideration. To address this issue, we developed a hypergraph-based approach to account for group-based spatial relations and uncertainties in perceiving the environment. Our empirical results have shown that this approach outperforms a previous graph-based approach with an absolute gain of 9%. However, while these graph-based approaches perform effectively when the agent has perfect knowledge or perception of the environment (e.g., 84%), they perform rather poorly when the agent has imperfect perception of the environment (e.g., 45%). This big performance gap calls for new solutions to REG

that can mediate a shared perceptual basis in situated dialogue.

Open-Domain Fine-Grained Class Extraction from Web Search Queries

Marius Pasca

Saturday 6:00–7:30 — Princessa Ballroom and Foyer

This paper introduces a method for extracting fine-grained class labels (countries with double taxation agreements with india) from Web search queries. The class labels are more numerous and more diverse than those produced by previous extraction methods. Also extracted are representative sets of instances (singapore, united kingdom) for the class labels.

Unsupervised Relation Extraction with General Domain Knowledge

Oier Lopez de Lacalle, Mirella Lapata

Saturday 6:00–7:30 — Princessa Ballroom and Foyer

In this paper we present an unsupervised approach to relational information extraction. Our model partitions tuples representing an observed syntactic relationship between two named entities (e.g., “X was born in Y” and “X is from Y”) into clusters corresponding to underlying semantic relation types (e.g., BornIn, Located). Our approach incorporates general domain knowledge which we encode as First Order Logic rules and automatically combine with a topic model developed specifically for the relation extraction task. Evaluation results on the ACE 2007 English Relation Detection and Categorization (RDC) task show that our model outperforms competitive unsupervised approaches by a wide margin and is able to produce clusters shaped by both the data and the rules.

Efficient Collective Entity Linking with Stacking

Zhengyan He, Shujie Liu, Yang Song, Mu Li, Ming Zhou, Hongfeng Wang

Saturday 6:00–7:30 — Princessa Ballroom and Foyer

Entity disambiguation works by linking ambiguous mentions in text to their corresponding real-world entities in knowledge base. Recent collective disambiguation methods enforce coherence among contextual decisions at the cost of non-trivial inference processes. We propose a fast collective disambiguation approach based on stacking. First, we train local predictor $g_{0\$}$ with learning to rank as base learner, to generate initial ranking list of candidates. Second, top k candidates of related instances are searched for constructing expressive global coherence features. A global predictor $g_{1\$}$ is trained in the augmented feature space and stacking is employed to tackle the train/test mismatch problem. The proposed method is fast and easy to implement. Experiments show its effectiveness over various algorithms on several public datasets. By learning a rich semantic relatedness measure between entity categories and context document, performance is further improved.

Joint Bootstrapping of Corpus Annotations and Entity Types

Hrushikesh Mohapatra, Siddhanth Jain, Soumen Chakrabarti

Saturday 6:00–7:30 — Princessa Ballroom and Foyer

Web search can be enhanced in powerful ways if token spans in Web text are annotated with disambiguated entities from large catalogs like Freebase. Entity annotators need to be trained on sample mention snippets. Wikipedia entities and annotated pages offer high-quality labeled data for training and evaluation. Unfortunately, Wikipedia features only one-ninth the number of entities as Freebase, and these are a highly biased sample of well-connected, frequently mentioned “head” entities. To bring hope to “tail” entities, we broaden our goal to a second task: assigning types to entities in Freebase but not Wikipedia. The two tasks are synergistic: knowing the types of unfamiliar entities helps disambiguate mentions, and words in mention contexts help assign types to entities. We present TMI, a bipartite graphical model for joint type-mention inference. TMI attempts no schema integration or entity resolution, but exploits the above-mentioned synergy. In experiments involving 780,000 people in Wikipedia, 2.3 million people in Freebase, 700 million Web pages, and over 20 professional editors, TMI shows considerable annotation accuracy improvement (e.g., 70%) compared to baselines (e.g., 46%), especially for “tail” and emerging entities. We also compare with Google’s recent annotations of the same corpus with Freebase entities, and report considerable improvements within the people domain.

Effectiveness and Efficiency of Open Relation Extraction

Filipe Mesquita, Jordan Schmedek, Denilson Barbosa

Saturday 6:00–7:30 — Princessa Ballroom and Foyer

A large number of Open Relation Extraction approaches have been proposed recently, covering a wide range of NLP machinery, from "shallow" (e.g., part-of-speech tagging) to "deep" (e.g., semantic role labeling-SRL). A natural question then is what is the trade-off between NLP depth (and associated computational cost) versus effectiveness. This paper presents a fair and objective experimental comparison of 8 state-of-the-art approaches over 5 different datasets, and sheds some light on the issue. The paper also describes a novel method, EX-EMPLAR, which adapts ideas from SRL to less costly NLP machinery, resulting in substantial gains both in efficiency and effectiveness, over binary and n-ary relation extraction tasks.

Automatic Feature Engineering for Answer Selection and Extraction

Aliaksei Severyn, Alessandro Moschitti

Saturday 6:00–7:30 — Princessa Ballroom and Foyer

This paper proposes a framework for automatically engineering features for two important tasks of question answering: answer sentence selection and answer extraction. We represent question and answer sentence pairs with linguistic structures enriched by semantic information, where the latter is produced by automatic classifiers, e.g., question classifier and Named Entity Recognizer. Tree kernels applied to such structures enable a simple way to generate highly discriminative structural features that combine syntactic and semantic information encoded in the input trees. We conduct experiments on a public benchmark from TREC to compare with previous systems for answer sentence selection and answer extraction. The results show that our models greatly improve on the state of the art, e.g., up to 22% on F1 (relative improvement) for answer extraction, while using no additional resources and no manual feature engineering.

Improving Web Search Ranking by Incorporating Structured Annotation of Queries

Xiao Ding, Zhicheng Dou, Bing Qin, Ting Liu, Ji-rong Wen

Saturday 6:00–7:30 — Princessa Ballroom and Foyer

Web users are increasingly looking for structured data, such as lyrics, job, or recipes, using unstructured queries on the web. However, retrieving relevant results from such data is a challenging problem due to the unstructured language of the web queries. In this paper, we propose a method to improve web search rankings by detecting Structured Annotation of queries based on top search results. In a structured annotation, the original query is split into different units that are associated with semantic attributes in the corresponding domain. We evaluate our techniques using real world queries and achieve significant improvement.

Building Specialized Bilingual Lexicons Using Large Scale Background Knowledge

Dhouha Bouamor, Adrian Popescu, Nasredine Semmar, Pierre Zweigenbaum

Saturday 6:00–7:30 — Princessa Ballroom and Foyer

Bilingual lexicons are central components of machine translation and cross-lingual information retrieval systems. Their manual construction requires strong expertise in both languages involved and is a costly process. Several automatic methods were proposed as an alternative but they often rely of resources available in a limited number of languages and their performances are still far behind the quality of manual translations. We introduce a novel approach to the creation of specific domain bilingual lexicon that relies on Wikipedia. This massively multilingual encyclopedia makes it possible to create lexicons for a large number of language pairs. Wikipedia is used to extract domains in each language, to link domains between languages and to create generic translation dictionaries. The approach is tested on four specialized domains and is compared to three state of the art approaches using two language pairs: French-English and Romanian-English. The newly introduced method compares favorably to existing methods in all configurations tested.

Document Summarization via Guided Sentence Compression

Chen Li, Fei Liu, Fuliang Weng, Yang Liu

Saturday 6:00–7:30 — Princessa Ballroom and Foyer

Joint compression and summarization has been used recently to generate high quality summaries. However, such word-based joint optimization is computationally expensive. In this paper we adopt the 'sentence compression + sentence selection' pipeline approach for compressive summarization, but propose to perform sum-

mary guided compression, rather than generic sentence-based compression. To create an annotated corpus, the human annotators were asked to compress sentences while explicitly given the important summary words in the sentences. Using this corpus, we train a supervised sentence compression model using a set of token-, syntax-, and document level features. During summarization, we use multiple compressed sentences in the integer linear programming framework to select salient summary sentences. Our results on the TAC 2008 and 2011 summarization data sets show that by incorporating the guided sentence compression model, our summarization system can yield significant performance gain as compared to the state-of-the-art.

Anchor Graph: Global Reordering Contexts for Statistical Machine Translation

Hendra Setiawan, Bowen Zhou, Bing Xiang
Saturday 6:00–7:30 — Princessa Ballroom and Foyer

Reordering poses one of the greatest challenges in Statistical Machine Translation research as the key contextual information may well be beyond the confine of translation units. We present the “Anchor Graph” (AG) model where we use a graph structure to model global contextual information that is crucial in reordering. The key ingredient of our AG model is the edges that model the relationship between the reordering around a set of selected translation units, which we refer to as emphanchors. As the edges link anchors that may span multiple translation units at decoding time, our AG model effectively encodes global contextual information that is previously absent. We integrate our proposed model into a state-of-the-art translation system and demonstrate the efficacy of our proposal in a large-scale Chinese-to-English translation task.

Source-Side Classifier Preordering for Machine Translation

Uri Lerner, Slav Petrov
Saturday 6:00–7:30 — Princessa Ballroom and Foyer

We present a simple and novel classifier-based preordering approach. Unlike existing preordering models, we train feature-rich discriminative classifiers that directly predict the target-side word order. Our approach combines the strengths of lexical reordering and syntactic preordering models by performing long-distance reorderings using the structure of the parse tree, while utilizing a discriminative model with a rich set of features, including lexical features. We present extensive experiments on 22 language pairs, including preordering into English from 7 other languages. We obtain improvements of up to 1.4 BLEU on language pairs in the WMT 2010 shared task. For languages from different families the improvements often exceed 2 BLEU. Many of these gains are also significant in human evaluations.

Improving Pivot-Based Statistical Machine Translation Using Random Walk

Xiaoning Zhu, Zhongjun He, Hua Wu, Haifeng Wang, Conghui Zhu, Tiejun Zhao
Saturday 6:00–7:30 — Princessa Ballroom and Foyer

This paper proposes a novel approach that utilizes a machine learning method to improve pivot-based statistical machine translation (SMT). For language pairs with few bilingual data, a possible solution in pivot-based SMT using another language as a “bridge” to generate source-target translation. However, one of the weaknesses is that some useful source-target translations cannot be generated if the corresponding source phrase and target phrase connect to different pivot phrases. To alleviate the problem, we utilize Markov random walks to connect possible translation phrases between source and target language. Experimental results on European Parliament data, spoken language data and web data show that our method leads to significant improvements on all the tasks over the baseline system.

Improving Alignment of System Combination by Using Multi-objective Optimization

Tian Xia, Zongcheng Ji, Shaodan Zhai, Yidong Chen, Qun Liu, Shaojun Wang
Saturday 6:00–7:30 — Princessa Ballroom and Foyer

This paper proposes a multi-objective optimization framework which supports heterogeneous information sources to improve alignment in machine translation system combination techniques. In this area, most of techniques usually utilize confusion networks (CN) as their central data structure to compact an exponential number of a potential hypotheses, and because better hypothesis alignment may benefit constructing better quality confusion networks, it is natural to add more useful information to improve alignment results. However, these information may be heterogeneous, so the widely-used Viterbi algorithm for searching the best alignment may not apply here. In the multi-objective optimization framework, each information source is viewed as an

independent objective, and a new goal of improving all objectives can be searched by mature algorithms. The solutions from this framework, termed Pareto optimal solutions, are then combined to construct confusion networks. Experiments on two Chinese-to-English translation datasets show significant improvements, 0.97 and 1.06 BLEU points over a strong Indirected Hidden Markov Model-based (IHMM) system, and 4.75 and 3.53 points over the best single machine translation systems.

Flexible and Efficient Hypergraph Interactions for Joint Hierarchical and Forest-to-String Decoding

Martin Cmejrek, Haitao Mi, Bowen Zhou

Saturday 6:00–7:30 — Princessa Ballroom and Foyer

Machine translation benefits from system combination. We propose flexible interaction of hypergraphs as a novel technique combining different translation models within one decoder. We introduce features controlling the interactions between the two systems and explore three interaction schemes of hierarchical and forest-to-string models—specification, generalization, and interchange. The experiments are carried out on large training data with strong baselines utilizing rich sets of dense and sparse features. All three schemes significantly improve results of any single system on four test sets. We find that specification—more constrained scheme that almost entirely uses forest-to-string rules, but optionally uses hierarchical rules for shorter spans—comes out as the strongest, yielding improvement up to 0.9 (Ter-Bleu)/2 points. We also provide a detailed experimental and qualitative analysis of the results.

Factored Soft Source Syntactic Constraints for Hierarchical Machine Translation

Zhongqiang Huang, Jacob Devlin, Rabih Zbib

Saturday 6:00–7:30 — Princessa Ballroom and Foyer

This paper describes a factored approach to incorporating soft source syntactic constraints into a hierarchical phrase-based translation system. In contrast to traditional approaches that directly introduce syntactic constraints to translation rules by explicitly decorating them with syntactic annotations, which often exacerbate the data sparsity problem and cause other problems, our approach keeps translation rules intact and factorizes the use of syntactic constraints through two separate models: 1) a syntax mismatch model that associates each nonterminal of a translation rule with a distribution of tags that is used to measure the degree of syntactic compatibility of the translation rule on source spans; 2) a syntax-based reordering model that predicts whether a pair of sibling constituents in the constituent parse tree of the source sentence should be reordered or not when translated to the target language. The features produced by both models are used as soft constraints to guide the translation process. Experiments on Chinese-English translation show that the proposed approach significantly improves a strong string-to-dependency translation system on multiple evaluation sets.

Recursive Autoencoders for ITG-Based Translation

Peng Li, Yang Liu, Maosong Sun

Saturday 6:00–7:30 — Princessa Ballroom and Foyer

While inversion transduction grammar (ITG) is well suited for modeling ordering shifts between languages, how to make applying the two reordering rules (i.e., straight and inverted) dependent on actual blocks being merged remains a challenge. Unlike previous work that only uses boundary words, we propose to use recursive autoencoders to make full use of the entire merging blocks alternatively. The recursive autoencoders are capable of generating vector space representations for variable-sized phrases, which enable predicting orders to exploit syntactic and semantic information from a neural language modeling's perspective. Experiments on the NIST 2008 dataset show that our system significantly improves over the MaxEnt classifier by 1.07 BLEU points.

Automatically Classifying Edit Categories in Wikipedia Revisions

Johannes Daxenberger, Iryna Gurevych

Saturday 6:00–7:30 — Princessa Ballroom and Foyer

In this paper, we analyze a novel set of features for the task of automatic edit category classification. Edit category classification assigns categories such as spelling error correction, paraphrase or vandalism to edits in a document. Our features are based on differences between two versions of a document including meta data, textual and language properties and markup. In a supervised machine learning experiment, we achieve a micro-

averaged F1 score of .62 on a corpus of edits from the English Wikipedia. In this corpus, each edit has been multi-labeled according to a 21-category taxonomy. A model trained on the same data achieves state-of-the-art performance on the related task of fluency edit classification. We apply pattern mining to automatically labeled edits in the revision histories of different Wikipedia articles. Our results suggest that high-quality articles show a higher degree of homogeneity with respect to their collaboration patterns as compared to random articles.

Semi-Markov Phrase-Based Monolingual Alignment

Xuchen Yao, Benjamin Van Durme, Chris Callison-Burch, Peter Clark

Saturday 6:00–7:30 — Princessa Ballroom and Foyer

We introduce a novel discriminative model for phrase-based monolingual alignment using a semi-Markov CRF. Our model achieves state-of-the-art alignment accuracy on two phrase-based alignment datasets (RTE and paraphrase), while doing significantly better than other strong baselines in both non-identical alignment and phrase-only alignment. Additional experiments highlight the potential benefit of our alignment model to RTE, paraphrase identification and question answering, where even a naive application of our model's alignment score approaches the state of the art.

A Constrained Latent Variable Model for Coreference Resolution

Kai-Wei Chang, Rajhans Samdani, Dan Roth

Saturday 6:00–7:30 — Princessa Ballroom and Foyer

Coreference resolution is a well known clustering task in Natural Language Processing. In this paper, we describe the Latent Left Linking model (L3 M), a novel, principled, and linguistically motivated latent structured prediction approach to coreference resolution. We show that L3 M admits efficient inference and can be augmented with knowledge-based constraints; we also present a fast stochastic gradient based learning. Experiments on ACE and Ontonotes data show that L3 M and its constrained version, CL3 M, are more accurate than several state-of-the-art approaches as well as some structured prediction models proposed in the literature.

Centering Similarity Measures to Reduce Hubs

Ikumi Suzuki, Kazuo Hara, Masashi Shimbo, Marco Saerens, Kenji Fukumizu

Saturday 6:00–7:30 — Princessa Ballroom and Foyer

The performance of nearest neighbor methods is degraded by the presence of hubs, i.e., objects in the dataset that are similar to many other objects. In this paper, we show that the classical method of centering, the transformation that shifts the origin of the space to the data centroid, provides an effective way to reduce hubs. We show analytically why hubs emerge and why they are suppressed by centering, under a simple probabilistic model of data. To further reduce hubs, we also move the origin more aggressively towards hubs, through weighted centering. Our experimental results show that (weighted) centering is effective for natural language data; it improves the performance of the k -nearest neighbor classifiers considerably in word sense disambiguation and document classification tasks.

Unsupervised Spectral Learning of WCFG as Low-rank Matrix Completion

Raphaël Bailly, Xavier Carreras, Franco M. Luque, Ariadna Quattoni

Saturday 6:00–7:30 — Princessa Ballroom and Foyer

We derive a spectral method for unsupervised learning of Weighted Context Free Grammars. We frame WCFG induction as finding a Hankel matrix that has low rank and is linearly constrained to represent a function computed by an inside-outside recursions. The proposed algorithm picks the grammar that agrees with a sample and is the simplest with respect to the nuclear norm of the Hankel matrix.

Identifying Phrasal Verbs Using Many Bilingual Corpora

Karl Pichotta, John DeNero

Saturday 6:00–7:30 — Princessa Ballroom and Foyer

We address the problem of identifying multiword expressions in a language, focusing on English phrasal verbs. Our polyglot ranking approach integrates frequency statistics from translated corpora in 50 different languages. Our experimental evaluation demonstrates that combining statistical evidence from many parallel

corpora using a novel ranking-oriented boosting algorithm produces a comprehensive set of English phrasal verbs, achieving performance comparable to a human-curated set.

Deep Learning for Chinese Word Segmentation and POS Tagging

Xiaoqing Zheng, Hanyang Chen, Tianyu Xu

Saturday 6:00–7:30 — Princessa Ballroom and Foyer

This study explores the feasibility of performing Chinese word segmentation (CWS) and POS tagging by deep learning. We try to avoid task-specific feature engineering, and use deep layers of neural networks to discover relevant features to the tasks. We leverage large-scale unlabeled data to improve internal representation of Chinese characters, and use these improved representations to enhance supervised word segmentation and POS tagging models. Our networks achieved close to state-of-the-art performance with minimal computational cost. We also describe a perceptron-style algorithm for training the neural networks, as an alternative to maximum-likelihood method, to speed up the training process and make the learning algorithm easier to be implemented.

Joint Chinese Word Segmentation and POS Tagging on Heterogeneous Annotated Corpora with Multiple Task Learning

Xipeng Qiu, Jiayi Zhao, Xuanjing Huang

Saturday 6:00–7:30 — Princessa Ballroom and Foyer

Chinese word segmentation and part-of-speech tagging (S&T) are fundamental steps for more advanced Chinese language processing tasks. Recently, it has attracted more and more research interests to exploit heterogeneous annotation corpora for Chinese S&T. In this paper, we propose a unified model for Chinese S&T with heterogeneous annotation corpora. We first automatically construct a loose and uncertain mapping between two representative heterogeneous corpora, Penn Chinese Treebank (CTB) and PKU’s People’s Daily (PPD). Then we regard the Chinese S&T with heterogeneous corpora as two “related” tasks and train our model on two heterogeneous corpora simultaneously. Experiments show that our method can boost the performances of both of the heterogeneous corpora by using the shared information, and achieves significant improvements over the state-of-the-art methods.

The Topology of Semantic Knowledge

Jimmy Dubuisson, Jean-Pierre Eckmann, Christian Scheible, Hinrich Schütze

Saturday 6:00–7:30 — Princessa Ballroom and Foyer

Studies of the graph of dictionary definitions (DD) (Picard et al., 2009; Levary et al., 2012) have revealed strong semantic coherence of local topological structures. The techniques used in these papers are simple and the main results are found by understanding the structure of cycles in the directed graph (where words point to definitions). Based on our earlier work (Levary et al., 2012), we study a different class of word definitions, namely those of the Free Association (FA) dataset (Nelson et al., 2004). These are responses by subjects to a cue word, which are then summarized by a directed, free association graph. We find that the structure of this network is quite different from both the Wordnet and the dictionary networks. This difference can be explained by the very nature of free association as compared to the more “logical” construction of dictionaries. It thus sheds some (quantitative) light on the psychology of free association. In NLP, semantic groups or clusters are interesting for various applications such as word sense disambiguation. The FA graph is tighter than the DD graph, because of the large number of triangles. This also makes drift of meaning quite measurable so that FA graphs provide a quantitative measure of the semantic coherence of small groups of words.

Unsupervised Induction of Cross-Lingual Semantic Relations

Mike Lewis, Mark Steedman

Saturday 6:00–7:30 — Princessa Ballroom and Foyer

Creating a language-independent meaning representation would benefit many cross-lingual NLP tasks. We introduce the first unsupervised approach to this problem, learning clusters of semantically equivalent English and French relations between referring expressions, based on their named-entity arguments in large monolingual corpora. The clusters can be used as language-independent semantic relations, by mapping clustered expressions in different languages onto the same relation. Our approach needs no parallel text for training, but outperforms a baseline that uses machine translation on a cross-lingual question answering task. We also show

how to use the semantics to improve the accuracy of machine translation, by using it in a simple reranker.

Two-Stage Method for Large-Scale Acquisition of Contradiction Pattern Pairs using Entailment

Julien Kloetzer, Stijn De Saeger, Kentaro Torisawa, Chikara Hashimoto, Jong-Hoon Oh, Motoki Sano, Kiyonori Ohtake

Saturday 6:00–7:30 — Princessa Ballroom and Foyer

In this paper we propose a two-stage method to acquire contradiction relations between typed lexico-syntactic patterns such as X_{drug} prevents $Y_{disease}$ and $Y_{disease}$ caused by X_{drug} . In the first stage, we train an SVM classifier to detect contradiction pattern pairs in a large web archive by exploiting the excitation polarity (Hashimoto et al., 2012) of the patterns. In the second stage, we enlarge the first stage classifier’s training data with new contradiction pairs obtained by combining the output of the first stage’s classifier and that of an entailment classifier. We acquired this way 750,000 typed Japanese contradiction pattern pairs with an estimated precision of 80%. We plan to release this resource to the NLP community.

Sarcasm as Contrast between a Positive Sentiment and Negative Situation

Ellen Riloff, Ashequl Qadir, Prafulla Surve, Lalindra De Silva, Nathan Gilbert, Ruihong Huang

Saturday 6:00–7:30 — Princessa Ballroom and Foyer

A common form of sarcasm on Twitter consists of a positive sentiment contrasted with a negative situation. For example, many sarcastic tweets include a positive sentiment, such as "love" or "enjoy", followed by an expression that describes an undesirable activity or state (e.g., "taking exams" or "being ignored"). We have developed a sarcasm recognizer to identify this type of sarcasm in tweets. We present a novel bootstrapping algorithm that automatically learns lists of positive sentiment phrases and negative situation phrases from sarcastic tweets. We show that identifying contrasting contexts using the phrases learned through bootstrapping yields improved recall for sarcasm recognition.

Collective Personal Profile Summarization with Social Networks

Zhongqing Wang, Shoushan LI, Fang Kong, Guodong Zhou

Saturday 6:00–7:30 — Princessa Ballroom and Foyer

Personal profile information on social media like LinkedIn.com and Facebook.com is at the core of many interesting applications, such as talent recommendation and contextual advertising. However, personal profiles usually lack organization confronted with the large amount of available information. Therefore, it is always a challenge for people to find desired information from them. In this paper, we address the task of personal profile summarization by leveraging both personal profile textual information and social networks. Here, using social networks is motivated by the intuition that, people with similar academic, business or social connections (e.g. co-major, co-university, and co-corporation) tend to have similar experience and summaries. To achieve the learning process, we propose a collective factor graph (CoFG) model to incorporate all these re-sources of knowledge to summarize personal profiles with local textual attribute functions and social connection factors. Extensive evaluation on a large-scale dataset from LinkedIn.com demonstrates the effectiveness of the proposed approach.

Optimized Event Storyline Generation based on Mixture-Event-Aspect Model

Lifu Huang, Lian'en Huang

Saturday 6:00–7:30 — Princessa Ballroom and Foyer

Recently, much research focuses on event storyline generation, which aims to produce a concise, global and temporal event summary from a collection of articles. Generally, each event contains multiple sub-events and the storyline should be composed by the component summaries of all the sub-events. However, different sub-events have different part-whole relationship with the major event, which is important to correspond to users’ interests but seldom considered in previous work. To distinguish different types of sub-events, we propose a mixture-event-aspect model which models different sub-events into local and global aspects. Combining these local/global aspects with summarization requirements together, we utilize an optimization method to generate the component summaries along the timeline. We develop experimental systems on 6 distinctively different datasets. Evaluation and comparison results indicate the effectiveness of our proposed method.

Automatically Determining a Proper Length for Multi-Document Summarization: A Bayesian Nonparametric Approach

Tengfei Ma, Hiroshi Nakagawa

Saturday 6:00–7:30 — Princessa Ballroom and Foyer

Document summarization is an important task in the area of natural language processing, which aims to extract the most important information from a single document or a cluster of documents. In various summarization tasks, the summary length is manually defined. However, how to find the proper summary length is quite a problem; and keeping all summaries restricted to the same length is not always a good choice. It is obviously improper to generate summaries with the same length for two clusters of documents which contain quite different quantity of information. In this paper, we propose a Bayesian nonparametric model for multi-document summarization in order to automatically determine the proper lengths of summaries. Assuming that an original document can be reconstructed from its summary, we describe the "reconstruction" by a Bayesian framework which selects sentences to form a good summary. Experimental results on DUC2004 data sets and some expanded data demonstrate the good quality of our summaries and the rationality of the length determination.

A Discourse-Driven Content Model for Summarising Scientific Articles Evaluated in a Complex Question Answering Task

Maria Liakata, Simon Dobnik, Shyamasree Saha, Colin Batchelor, Dietrich Rebbholz-Schuhmann

Saturday 6:00–7:30 — Princessa Ballroom and Foyer

We present a method which exploits automatically generated scientific discourse annotations to create a content model for the summarisation of scientific articles. Full papers are first automatically annotated using the CoreSC scheme, which captures 11 content-based concepts such as Hypothesis, Result, Conclusion etc at the sentence level. A content model which follows the sequence of CoreSC categories observed in abstracts is used to provide the skeleton of the summary, making a distinction between dependent and independent categories. Summary creation is also guided by the distribution of CoreSC categories found in the full articles, in order to adequately represent the article content. Finally, we demonstrate the usefulness of the summaries by evaluating them in a complex question answering task. Results are very encouraging as summaries of papers from automatically obtained CoreSCs enable experts to answer 66 % of complex content-related questions designed on the basis of paper abstracts. The questions were answered with a precision of 75 %, where the upper bound for human summaries (abstracts) was 95 %.

Optimal Incremental Parsing via Best-First Dynamic Programming

Kai Zhao, James Cross, Liang Huang

Saturday 6:00–7:30 — Princessa Ballroom and Foyer

We present the first provably optimal polynomial-time dynamic programming (DP) algorithm for best-first shift-reduce parsing, which applies the DP idea of Huang and Sagae (2010) to the best-first parser of Sagae and Lavie (2006) in a non-trivial way, reducing the complexity of the latter from exponential to polynomial. We prove the correctness of our algorithm rigorously. Experiments confirm that DP leads to a significant speedup on a probabilistic best-first shift-reduce parser, and makes exact search under such a model tractable for the first time.

Exploiting Language Models for Visual Recognition

Dieu-Thu Le, Jasper Uijlings, Raffaella Bernardi

Saturday 6:00–7:30 — Princessa Ballroom and Foyer

The problem of learning language models from large text corpora has been widely studied within the computational linguistic community. However, little is known about the performance of these language models when applied to the computer vision domain. In this work, we compare representative models: a window-based model, a topic model, a distributional memory and a commonsense knowledge database, ConceptNet, in two visual recognition scenarios: human action recognition and object prediction. We examine whether the knowledge extracted from texts through these models are compatible to the knowledge represented in images. We determine the usefulness of different language models in aiding the two visual recognition tasks. The study shows that the language models built from general text corpora can be used instead of expensive annotated images and even outperform the image model when testing on a big general dataset.

Mining Scientific Terms and their Definitions: A Study of the ACL Anthology

Yiping Jin, Min-Yen Kan, Jun-Ping Ng, Xiangnan He

Saturday 6:00–7:30 — Princessa Ballroom and Foyer

This paper presents DefMiner, a supervised sequence labeling system that identifies scientific terms and their accompanying definitions. DefMiner achieves 85% F1 on a Wikipedia benchmark corpus, significantly improving the previous state-of-the-art by 8%. We exploit DefMiner to process the ACL Anthology Reference Corpus (ARC) – a large, real-world digital library of scientific articles in computational linguistics. The resulting automatically-acquired glossary represents the terminology defined over several thousand individual research articles. We highlight several interesting observations: more definitions are introduced for conference and workshop papers over the years and that multiword terms account for slightly less than half of all terms. Obtaining a list of popular defined terms in a corpus of computational linguistics papers, we find that concepts can often be categorized into one of three categories: resources, methodologies and evaluation metrics.

Joint Learning and Inference for Grammatical Error Correction

Alla Rozovskaya, Dan Roth

Saturday 6:00–7:30 — Princessa Ballroom and Foyer

State-of-the-art systems for grammatical error correction are based on a collection of independently-trained models for specific errors. Such models ignore linguistic interactions at the sentence level and thus do poorly on mistakes that involve grammatical dependencies among several words. In this paper, we identify linguistic structures with interacting grammatical properties and propose to address such dependencies via joint inference and joint learning. We show that it is possible to identify interactions well enough to facilitate a joint approach and, consequently, that joint methods correct incoherent predictions that independently-trained classifiers tend to produce. Furthermore, because the joint learning model considers interacting phenomena during training, it is able to identify mistakes that require making multiple changes simultaneously and that standard approaches miss. Overall, our model significantly outperforms the Illinois system that placed first in the CoNLL-2013 shared task on grammatical error correction.

With Blinkers on: Robust Prediction of Eye Movements across Readers

Franz Matthias, Anders Søgaard

Saturday 6:00–7:30 — Princessa Ballroom and Foyer

Nilsson and Nivre (2009) introduced a tree-based model of persons' eye movements in reading. The individual variation between readers reportedly made application across readers impossible. While a tree-based model seems plausible for eye movements, we show that competitive results can be obtained with a linear CRF model. Increasing the inductive bias also makes learning across readers possible. In fact we observe next-to-no performance drop when evaluating models trained on gaze records of multiple readers on new readers.

Using Paraphrases and Lexical Semantics to Improve the Accuracy and the Robustness of Supervised Models in Situated Dialogue Systems

Claire Gardent, Lina M. Rojas Barahona

Saturday 6:00–7:30 — Princessa Ballroom and Foyer

This paper explores to what extent lemmatisation, lexical resources, distributional semantics and paraphrases can increase the accuracy of supervised models for dialogue management. The results suggest that each of these factors can help improve performance but that the impact will vary depending on their combination and on the evaluation mode.

Cascading Collective Classification for Bridging Anaphora Recognition using a Rich Linguistic Feature Set

Yufang Hou, Katja Markert, Michael Strube

Saturday 6:00–7:30 — Princessa Ballroom and Foyer

Recognizing bridging anaphora is difficult due to the wide variation within the phenomenon, the resulting lack of easily identifiable surface markers and their relative rarity. We develop linguistically motivated discourse structure, lexico-semantic and genericity detection features and integrate these into a cascaded minority preference algorithm that models bridging recognition as a subtask of learning fine-grained information status (IS).

We substantially improve bridging recognition without impairing performance on other IS classes.

A Synchronous Context Free Grammar for Time Normalization

Steven Bethard

Saturday 6:00–7:30 — Princessa Ballroom and Foyer

We present an approach to time normalization (e.g. "the day before yesterday" => 2013-04-12) based on a synchronous context free grammar. Synchronous rules map the source language to formally defined operators for manipulating times (FindEnclosed, StartAtEndOf, etc.). Time expressions are then parsed using an extended CYK+ algorithm, and converted to a normalized form by applying the operators recursively. For evaluation, a small set of synchronous rules for English time expressions were developed. Our model outperforms Hei-delTime, the best time normalization system in TempEval 2013, on four different time normalization corpora.

Rule-Based Information Extraction is Dead! Long Live Rule-Based Information Extraction Systems!

Laura Chiticariu, Yunyao Li, Frederick R. Reiss

Saturday 6:00–7:30 — Princessa Ballroom and Foyer

The rise of "Big Data" analytics over unstructured text has led to renewed interest in information extraction (IE). We surveyed the landscape of IE technologies and identified a major disconnect between industry and academia: while rule-based IE dominates the commercial world, it is widely regarded as dead-end technology by the academia. We believe the disconnect stems from the way in which the two communities measure the benefits and costs of IE, as well as academia's perception that rule-based IE is devoid of research challenges. We make a case for the importance of rule-based IE to industry practitioners. We then lay out a research agenda in advancing the state-of-the-art in rule-based IE systems which we believe has the potential to bridge the gap between academic research and industry practice.

Improving Learning and Inference in a Large Knowledge-Base using Latent Syntactic Cues

Matt Gardner, Partha Pratim Talukdar, Bryan Kisiel, Tom Mitchell

Saturday 6:00–7:30 — Princessa Ballroom and Foyer

Automatically constructed knowledge bases (KBs) are often incomplete and there is a genuine need to improve their coverage. Path Ranking Algorithm (PRA) is a recently proposed method which aims to improve KB coverage by performing inference directly over the KB graph. For the first time, we demonstrate that addition of edges labeled with latent features mined from a large dependency parsed corpus of 500 million Web documents can significantly outperform previous PRA-based approaches on the KB inference task. We present extensive experimental results validating this finding. The resources presented in this paper are publicly available.

What is Hidden among Translation Rules

Libin Shen, Bowen Zhou

Saturday 6:00–7:30 — Princessa Ballroom and Foyer

Most of the machine translation systems rely on a large set of translation rules. These rules are treated as discrete and independent events. In this work-in-progress short paper, we propose a novel method to model rules as observed generation output of a compact hidden model, which leads to better generalization capability. We present a preliminary generative model to test this idea. Experimental results show about one point improvement on TER-BLEU over a strong baseline in Chinese-to-English translation.

Converting Continuous-Space Language Models into N-Gram Language Models for Statistical Machine Translation

Rui Wang, Masao Utiyama, Isao Goto, Eiichiro Sumita, Hai Zhao, Bao-Liang Lu

Saturday 6:00–7:30 — Princessa Ballroom and Foyer

Neural network language models, or continuous-space language models (CSLMs), have been shown to improve the performance of statistical machine translation (SMT) when they are used for reranking n-best translations. However, CSLMs have not been used in the first pass decoding of SMT, because using CSLMs in

decoding takes a lot of time. In contrast, we propose a method for converting CSLMs into back-off n-gram language models (BNLMs) so that we can use converted CSLMs in decoding. We show that they outperform the original BNLMs and are comparable with the traditional use of CSLMs in reranking.

A Corpus Level MIRA Tuning Strategy for Machine Translation

Ming Tan, Tian Xia, Shaojun Wang, Bowen Zhou

Saturday 6:00–7:30 — Princessa Ballroom and Foyer

MIRA based tuning methods have been widely used in statistical machine translation (SMT) system with a large number of features. Since the corpus-level BLEU is not decomposable, these MIRA approaches usually define a variety of heuristic-driven sentence-level BLEUs in their model losses. Instead, we present a new MIRA method, which employs an exact corpus-level BLEU to compute the model loss. Our method is simpler in implementation. Experiments on Chinese-to-English translation show its effectiveness over two state-of-the-art MIRA implementations.

Word Level Language Identification in Online Multilingual Communication

Dong Nguyen, A. Seza Dogruoz

Saturday 6:00–7:30 — Princessa Ballroom and Foyer

Multilingual speakers switch between languages in online and spoken communication. Analyses of large scale multilingual data require automatic language identification at the word level. For our experiments with multilingual online discussions, we first tag the language of individual words using language models and dictionaries. Secondly, we incorporate context to improve the performance. We achieve an accuracy of 98%. Besides word level accuracy, we use two new metrics to evaluate this task.

Microblog Entity Linking by Leveraging Extra Posts

Yuhang Guo, Bing Qin, Ting Liu, Sheng Li

Saturday 6:00–7:30 — Princessa Ballroom and Foyer

Linking name mention in microblog posts to a knowledge base, namely microblog entity linking, is useful for text mining tasks on microblog. Entity linking in long text has been well studied in previous works. However few work has focused on short text like microblog post. Microblog posts are short and noisy. Previous method can extract few features from the post context. In this paper we propose to use multiple posts for the microblog entity linking task. Experimental results show that our methods significantly improve the linking accuracy over traditional methods.

Automatic Domain Partitioning for Multi-Domain Learning

Di Wang, Chenyan Xiong, William Yang Wang

Saturday 6:00–7:30 — Princessa Ballroom and Foyer

Multi-Domain learning (MDL) assumes that the domain labels in the dataset are known. However, when there are multiple metadata attributes available, it is not always straightforward to select a single best attribute for domain partition, and it is possible that combining more than one metadata attributes (including continuous attributes) can lead to better MDL performance. In this work, we propose an automatic domain partitioning approach that aims at providing better domain identities for MDL. We use a supervised clustering approach that learns the domain distance between data instances, and then cluster the data into better domains for MDL. Our experiment on real multi-domain datasets shows that using our automatically generated domain partition improves over popular MDL methods.

Decipherment with a Million Random Restarts

Taylor Berg-Kirkpatrick, Dan Klein

Saturday 6:00–7:30 — Princessa Ballroom and Foyer

This paper investigates the utility and effect of running numerous random restarts when using EM to attack decipherment problems. We find that simple decipherment models are able to crack homophonic substitution ciphers with high accuracy if a large number of random restarts are used but almost completely fail with only a few random restarts. For particularly difficult homophonic ciphers, we find that big gains in accuracy are to be had by running upwards of 100K random restarts, which we accomplish efficiently using a GPU-based parallel

implementation. We run a series of experiments using millions of random restarts in order to investigate other empirical properties of decipherment problems, including the famously uncracked Zodiac 340.

Russian Stress Prediction using Maximum Entropy Ranking

Keith Hall, Richard Sproat

Saturday 6:00–7:30 — Princessa Ballroom and Foyer

We explore a model of stress prediction in Russian using a combination of local contextual features and linguistically-motivated features associated with the word’s stem and suffix. We frame this as a ranking problem, where the objective is to rank the pronunciation with the correct stress above those with incorrect stress. We train our models using a simple Maximum Entropy ranking framework allowing for efficient prediction. An empirical evaluation shows that a model combining the local contextual features and the linguistically-motivated non-local features performs best in predicting both primary and secondary stress.

Scaling to Large³ Data: An Efficient and Effective Method to Compute Distributional Thesauri

Martin Riedl, Chris Biemann

Saturday 6:00–7:30 — Princessa Ballroom and Foyer

We introduce a new highly scalable approach to computing Distributional Thesauri (DTs). By employing pruning techniques and a distributed framework, we make the computation for very large corpora feasible on comparably small computational resources. We demonstrate this by releasing a DT for the whole vocabulary of Google Books syntactic n-grams. Evaluating against lexical resources using two measures, we show that our approach produces higher quality DTs than previous approaches, and is thus preferable in terms of speed and quality for large corpora.

Discriminative Improvements to Distributional Sentence Similarity

Yangfeng Ji, Jacob Eisenstein

Saturday 6:00–7:30 — Princessa Ballroom and Foyer

Matrix and tensor factorization have been applied to a number of semantic relatedness tasks, including paraphrase identification. The key idea is that similarity in the latent space implies semantic relatedness. We describe three ways in which labeled data can improve the accuracy of these approaches on paraphrase classification. First, we design a new discriminative term-weighting metric called TF-KLD, which outperforms TF-IDF. Next, we show that using the latent representation from matrix factorization as features in a classification algorithm substantially improves accuracy. Finally, we combine latent features with fine-grained n-gram overlap features, yielding performance that is 3% more accurate than the prior state-of-the-art.

Is Twitter A Better Corpus for Measuring Sentiment Similarity?

Shi Feng, Le Zhang, Binyang Li, Daling Wang, Ge Yu, Kam-Fai Wong

Saturday 6:00–7:30 — Princessa Ballroom and Foyer

Extensive experiments have validated the effectiveness of the corpus-based method for classifying the word’s sentiment polarity. However, no work is done for comparing different corpora in the polarity classification task. Nowadays, Twitter has aggregated huge amount of data that are full of people’s sentiments. In this paper, we empirically evaluate the performance of different corpora in sentiment similarity measurement, which is the fundamental task for word polarity classification. Experiment results show that the Twitter data can achieve a much better performance than the Google, Web1T and Wikipedia based methods.

Implicit Feature Detection via a Constrained Topic Model and SVM

Wei Wang, Hua Xu, Xiaoqiu Huang

Saturday 6:00–7:30 — Princessa Ballroom and Foyer

Implicit feature detection, also known as implicit feature identification, is an essential aspect of feature-specific opinion mining but previous works have often ignored it. We think, based on the explicit sentences, several Support Vector Machine (SVM) classifiers can be established to do this task. Nevertheless, we believe it is possible to do better by using a constrained topic model instead of traditional attribute selection methods. Experiments show that this method outperforms the traditional attribute selection methods by a large margin

and the detection task can be completed better.

Online Learning for Inexact Hypergraph Search

Hao Zhang, Liang Huang, Kai Zhao, Ryan McDonald

Saturday 6:00–7:30 — Princessa Ballroom and Foyer

Online learning algorithms like the perceptron are widely used for structured prediction tasks. For sequential search problems, like left-to-right tagging and parsing, beam search has been successfully combined with perceptron variants that accommodate search errors (Collins and Roark, 2004; Huang et al., 2012). However, perceptron training with inexact search is less studied for bottom-up parsing and, more generally, inference over hypergraphs. In this paper, we generalize the violation-fixing perceptron of Huang et al. (2012) to hypergraphs and apply it to the cube-pruning parser of Zhang and McDonald (2012). This results in the highest reported scores on WSJ evaluation set (UAS 93.50% and LAS 92.41% respectively) without the aid of additional resources.

Poster Session B

Princessa Ballroom and Foyer

Predicting the Presence of Discourse Connectives

Gary Patterson, Andrew Kehler

Saturday 7:30–9:00 — Princessa Ballroom and Foyer

We present a classification model that predicts the presence or omission of a lexical connective between two clauses, based upon linguistic features of the clauses and the type of discourse relation holding between them. The model is trained on a set of high frequency relations extracted from the Penn Discourse Treebank and achieves an accuracy of 86.6%. Analysis of the results reveals that the most informative features relate to the discourse dependencies between sequences of coherence relations in the text. We also present results of an experiment that provides insight into the nature and difficulty of the task.

Japanese Zero Reference Resolution Considering Exophora and Author/Reader Mentions

Masatsugu Hangyo, Daisuke Kawahara, Sadao Kurohashi

Saturday 7:30–9:00 — Princessa Ballroom and Foyer

In Japanese, zero references often occur and many of them are categorized into zero exophora, in which a referent is not mentioned in a document. However, previous studies have focused on only zero endophora, in which a referent explicitly appears. We present a zero reference resolution model considering zero exophora and author/reader of a document. To deal with zero exophora, our model adds pseudo entities corresponding to zero exophora to candidate referents of zero pronouns. In addition, we automatically detect mentions that refer to author and reader of a document by using lexico-syntactic patterns. We represent their particular behavior in a discourse as a feature vector of a machine learning model. We report the experimental results on Japanese zero reference resolution and demonstrate the effectiveness our model for not only zero exophora but also zero endophora.

A Dataset for Research on Short-Text Conversations

Hao Wang, Zhengdong Lu, Hang Li, Enhong Chen

Saturday 7:30–9:00 — Princessa Ballroom and Foyer

Natural language conversation is widely regarded as a highly difficult problem, which is usually attacked with either rule-based or learning-based models. In this paper we propose a retrieval-based automatic response model for short-text conversation, to exploit the vast amount of short conversation instances available on social media. For this purpose we introduce a dataset of short-text conversation based on the real-world instances from Sina Weibo (a popular Chinese microblog service), which will be soon released to public. This data set provides rich collection of instances for the research on finding natural and relevant short responses to a given short text, and useful for both training and testing of conversation models. This data set consists of both naturally formed conversations, manually labeled data, and a large repository of candidate responses. Our preliminary experiments demonstrate that the simple retrieval-based conversation model performs reasonably well when combined with the rich instances in our data set

Discourse Level Explanatory Relation Extraction from Product Reviews Using First-Order Logic

Qi Zhang, Jin Qian, Huan Chen, Jihua Kang, Xuanjing Huang

Saturday 7:30–9:00 — Princessa Ballroom and Foyer

Explanatory sentences are employed to clarify reasons, details, facts, and so on. High quality online product reviews usually include not only positive or negative opinions, but also a variety of explanations of why these opinions were given. These explanations can help readers get easily comprehensible information of the discussed products and aspects. Moreover, explanatory relations can also benefit sentiment analysis applications. In this work, we focus on the task of identifying subjective text segments and extracting their corresponding explanations from product reviews in discourse level. We propose a novel joint extraction method using first-order logic to model rich linguistic features and long distance constraints. Experimental results demonstrate the effectiveness of the proposed method.

Building Event Threads out of Multiple News Articles

Xavier Tannier, Véronique Moriceau

Saturday 7:30–9:00 — Princessa Ballroom and Foyer

We present an approach for building multidocument event threads from a large corpus of newswire articles. An event thread is basically a succession of events belonging to the same story. It helps the reader to contextualize the information contained in a single article, by navigating backward or forward in the thread from this article. A specific effort is also made on the detection of reactions to a particular event. In order to build these event threads, we use a cascade of classifiers and other modules, taking advantage of the redundancy of information in the newswire corpus. We also share interesting comments concerning our manual annotation procedure for building a training and testing set.

Tree Kernel-based Negation and Speculation Scope Detection with Structured Syntactic Parse Features

Bowei Zou, Guodong Zhou, Qiaoming Zhu

Saturday 7:30–9:00 — Princessa Ballroom and Foyer

Scope detection is a key task in information extraction. This paper proposes a new approach for tree kernel-based scope detection by using the structured syntactic parse information. In addition, we have explored the way of selecting compatible features for different part-of-speech cues. Experiments on the BioScope corpus show that both constituent and dependency structured syntactic parse features have the advantage in capturing the potential relationships between cues and their scopes. Compared with the state of the art scope detection systems, our system achieves substantial improvement.

A temporal model of text periodicities using Gaussian Processes

Daniel Preotjuc-Pietro, Trevor Cohn

Saturday 7:30–9:00 — Princessa Ballroom and Foyer

Temporal variations of text are usually ignored in NLP applications. However, text use changes with time, which can affect many applications. In this paper we model periodic distributions of words over time. Focusing on hashtag frequency in Twitter, we first automatically identify the periodic patterns. We use this for regression in order to forecast the volume of a hashtag based on past data. We use Gaussian Processes, a state-of-the-art bayesian non-parametric model, with a novel periodic kernel. We demonstrate this in a text classification setting, assigning the tweet hashtag based on the rest of its text. This method shows significant improvements over competitive baselines.

Automatically Detecting and Attributing Indirect Quotations

Silvia Pareti, Tim O’Keefe, Ioannis Konstas, James R. Curran, Irena Koprinska

Saturday 7:30–9:00 — Princessa Ballroom and Foyer

Direct quotations are used for opinion mining and information extraction as they have an easy to extract span and they can be attributed to a speaker with high accuracy. However, simply focusing on direct quotations ignores around half of all reported speech, which is in the form of indirect or mixed speech. This work presents the first large-scale experiments in indirect and mixed quotation extraction and attribution. We propose two methods of extracting all quote types from news articles and evaluate them on two large annotated corpora, one of which is a contribution of this work. We further show that direct quotation attribution methods can be successfully applied to indirect and mixed quotation attribution.

Identifying Web Search Query Reformulation using Concept based Matching

Ahmed Hassan

Saturday 7:30–9:00 — Princessa Ballroom and Foyer

Web search users frequently modify their queries in hope of receiving better results. This process is referred to as "Query Reformulation". Previous research has mainly focused on proposing query reformulations in the form of suggested queries for users. Some research has studied the problem of predicting whether the current query is a reformulation of the previous query or not. However, this work has been limited to bag-of-words models where the main signals being used are word overlap, character level edit distance and word level edit distance. In this work, we show that relying solely on surface level text similarity results in many false positives where queries with different intents yet similar topics are mistakenly predicted as query reformulations. We

propose a new representation for Web search queries based on identifying the concepts in queries and show that we can significantly improve query reformulation performance using features of query concepts.

The Answer is at your Fingertips: Improving Passage Retrieval for Web Question Answering with Search Behavior Data

Mikhail Ageev, Dmitry Lagun, Eugene Agichtein

Saturday 7:30–9:00 — Princessa Ballroom and Foyer

Passage retrieval is a crucial first step of automatic Question Answering (QA). While existing passage retrieval algorithms are effective at selecting document passages most similar to the question, or those that contain the expected answer types, they do not take into account which parts of the document the searchers actually found useful. We propose, to the best of our knowledge, the first successful attempt to incorporate searcher examination data into passage retrieval for question answering. Specifically, we exploit detailed examination data, such as mouse cursor movements and scrolling, to infer the parts of the document the searcher found interesting, and then incorporate this signal into passage retrieval for QA. Our extensive experiments and analysis demonstrate that our method significantly improves passage retrieval, compared to using textual features alone. As an additional contribution, we make available to the research community the code and the search behavior data used in this study, with the hope of encouraging further research in this area.

Assembling the Kazakh Language Corpus

Olzhas Makhambetov, Aibek Makazhanov, Zhandos Yessenbayev, Bakhyt Matkarimov, Islam Sabyrgaliyev, Anuar Sharafudinov

Saturday 7:30–9:00 — Princessa Ballroom and Foyer

This paper presents the Kazakh Language Corpus (KLC), which is one of the first attempts made within a local research community to assemble a Kazakh corpus. KLC is designed to be a large scale corpus containing over 135 million words and conveying five stylistic genres: literary, publicistic, official, scientific and informal. Along with its primary part KLC comprises such parts as: (i) annotated sub-corpus, containing segmented documents encoded in the eXtensible Markup Language (XML) that marks complete morphological, syntactic, and structural characteristics of texts; (ii) as well as a sub-corpus with the annotated speech data. KLC has a web-based corpus management system that helps to navigate the data and retrieve necessary information. KLC is also open for contributors, who are willing to make suggestions, donate texts and help with annotation of existing materials.

Automatic Extraction of Morphological Lexicons from Morphologically Annotated Corpora

Ramy Eskander, Nizar Habash, Owen Rambow

Saturday 7:30–9:00 — Princessa Ballroom and Foyer

We present a method for automatically learning inflectional classes and associated lemmas from morphologically annotated corpora. The method consists of a core language independent algorithm, which can be optimized for specific languages. The method is demonstrated on Egyptian Arabic and German, two morphologically rich languages. Our best method for Egyptian Arabic provides an error reduction of 55.6% over a simple baseline; our best method for German achieves a 66.7% error reduction.

Joint Language and Translation Modeling with Recurrent Neural Networks

Michael Auli, Michel Galley, Chris Quirk, Geoffrey Zweig

Saturday 7:30–9:00 — Princessa Ballroom and Foyer

We present a joint language and translation model based on a recurrent neural network which predicts target words based on an unbounded history of both source and target words. The weaker independence assumptions of this model result in a vastly larger search space compared to related feed-forward-based language or translation models. We tackle this issue with a new lattice rescoring algorithm and demonstrate its effectiveness empirically. Our joint model builds on a well known recurrent neural network language model (Mikolov, 2012) augmented by a layer of additional inputs from the source language. We show competitive accuracy compared to the traditional channel model features. Our best results improve the output of a system trained on WMT 2012 French-English data by up to 1.5 BLEU, and by 1.1 BLEU on average across several test sets.

Multi-Domain Adaptation for SMT Using Multi-Task Learning

Lei Cui, Xilun Chen, Dongdong Zhang, Shujie Liu, Mu Li, Ming Zhou

Saturday 7:30–9:00 — Princessa Ballroom and Foyer

Domain adaptation for SMT usually adapts models to an individual specific domain. However, it often lacks some correlation among different domains where common knowledge could be shared to improve the overall translation quality. In this paper, we propose a novel multi-domain adaptation approach for SMT using Multi-Task Learning (MTL), with in-domain models tailored for each specific domain and a general-domain model shared by different domains. The parameters of these models are tuned jointly via MTL so that they can learn general knowledge more accurately and exploit domain knowledge better. Our experiments on a large-scale English-to-Chinese translation task validate that the MTL-based adaptation approach significantly and consistently improves the translation quality compared to a non-adapted baseline. Furthermore, it also outperforms the individual adaptation of each specific domain.

Translation with Source Constituency and Dependency Trees

Fandong Meng, Jun Xie, Linfeng Song, Yajuan Lü, Qun Liu

Saturday 7:30–9:00 — Princessa Ballroom and Foyer

We present a novel translation model, which simultaneously exploits the constituency and dependency trees on the source side, to combine the advantages of two types of trees. We take head-dependents relations of dependency trees as backbone and incorporate phrasal n-odes of constituency trees as the source side of our translation rules, and the target side as strings. Our rules hold the property of long distance reorderings and the compatibility with phrases. Large-scale experimental results show that our model achieves significant improvements over the constituency-to-string (+2.45 BLEU on average) and dependency-to-string (+0.91 BLEU on average) models, which only employ single type of trees, and significantly outperforms the state-of-the-art hierarchical phrase-based model (+1.12 BLEU on average), on three Chinese-English NIST test sets.

Monolingual Marginal Matching for Translation Model Adaptation

Ann Irvine, Chris Quirk, Hal Daumé III

Saturday 7:30–9:00 — Princessa Ballroom and Foyer

When using a machine translation (MT) model trained on OLD-domain parallel data to translate NEW-domain text, one major challenge is the large number of out-of-vocabulary (OOV) and new-translation-sense words. We present a method to identify new translations of both known and unknown source language words that uses NEW-domain comparable document pairs. Starting with a joint distribution of source-target word pairs derived from the OLD-domain parallel corpus, our method recovers a new joint distribution that matches the marginal distributions of the NEW-domain comparable document pairs, while minimizing the divergence from the OLD-domain distribution. Adding learned translations to our French-English MT model results in gains of about 2 BLEU points over strong baselines.

Efficient Left-to-Right Hierarchical Phrase-Based Translation with Improved Reordering

Maryam Siahbani, Baskaran Sankaran, Anoop Sarkar

Saturday 7:30–9:00 — Princessa Ballroom and Foyer

Left-to-right (LR) decoding (Watanabe et al., 2006b) is a promising decoding algorithm for hierarchical phrase-based translation (Hiero). It generates the target sentence by extending the hypotheses only on the right edge. LR decoding has complexity $O(n^2b)$ for input of n words and beam size b , compared to $O(n^3)$ for the CKY algorithm. It requires a single language model (LM) history for each target hypothesis rather than two LM histories per hypothesis as in CKY. In this paper we present an augmented LR decoding algorithm that builds on the original algorithm in (Watanabe et al., 2006b). Unlike that algorithm, using experiments over multiple language pairs we show two new results: our LR decoding algorithm provides demonstrably more efficient decoding than CKY Hiero, four times faster; and by introducing new distortion and reordering features for LR decoding, it maintains the same translation quality (as in BLEU scores) obtained phrase-based and CKY Hiero with the same translation model.

A Systematic Exploration of Diversity in Machine Translation

Kevin Gimpel, Dhruv Batra, Chris Dyer, Gregory Shakhnarovich
Saturday 7:30-9:00 — Princessa Ballroom and Foyer

This paper addresses the problem of producing a diverse set of plausible translations. We present a simple procedure that can be used with any statistical machine translation (MT) system. We explore three ways of using diverse translations: (1) system combination, (2) discriminative reranking with rich features, and (3) a novel post-editing scenario in which multiple translations are presented to users. We find that diversity can improve performance on these tasks, especially for sentences that are difficult for MT.

Max-Violation Perceptron and Forced Decoding for Scalable MT Training

Heng Yu, Liang Huang, Haitao Mi, Kai Zhao
Saturday 7:30-9:00 — Princessa Ballroom and Foyer

While large-scale discriminative training has triumphed in many NLP problems, its definite success on machine translation has been largely elusive. Most recent efforts along this line are not scalable (training on the small dev set with features from top 100 most frequent words) and overly complicated. We instead present a very simple yet theoretically motivated approach by extending the recent framework of “violation-fixing perceptron”, using forced decoding to compute the target derivations. Extensive phrase-based translation experiments on both Chinese-to-English and Spanish-to-English tasks show substantial gains in BLEU by up to +2.3/+2.0 on dev/test over MERT, thanks to 20M+ sparse features. This is the first successful effort of large-scale online discriminative training for MT.

Identifying Multiple Userids of the Same Author

Tieyun Qian, Bing Liu
Saturday 7:30-9:00 — Princessa Ballroom and Foyer

This paper studies the problem of identifying users who use multiple userids to post in social media. Since multiple userids may belong to the same author, it is hard to directly apply supervised learning to solve the problem. This paper proposes a new method, which still uses supervised learning but does not require training documents from the involved userids. Instead, it uses documents from other userids for classifier building. The classifier can be applied to documents of the involved userids. This is possible because we transform the document space to a similarity space and learning is performed in this new space. Our evaluation is done in the online review domain. The experimental results using a large number of userids and their reviews show that the proposed method is very effective.

Gender Inference of Twitter Users in Non-English Contexts

Morgane Ciot, Morgan Sonderegger, Derek Ruths
Saturday 7:30-9:00 — Princessa Ballroom and Foyer

While much work has considered the problem of latent attribute inference for users of social media such as Twitter, little has been done on non-English-based content and users. Here, we conduct the first assessment of latent attribute inference in languages beyond English, focusing on gender inference. We find that the gender inference problem in quite diverse languages can be addressed using existing machinery. Further, accuracy gains can be made by taking language-specific features into account. We identify languages with complex orthography, such as Japanese, as difficult for existing methods, suggesting a valuable direction for future research.

A Multimodal LDA Model Integrating Textual, Cognitive and Visual Modalities

Stephen Roller, Sabine Schulte im Walde
Saturday 7:30-9:00 — Princessa Ballroom and Foyer

Recent investigations into grounded models of language have shown that holistic views of language and perception can provide higher performance than independent views. In this work, we improve a two-dimensional multimodal version of Latent Dirichlet Allocation (Andrews et al., 2009) in various ways. (1) We outperform text-only models in two different evaluations, and demonstrate that low-level visual features are directly compatible with the existing model. (2) We present a novel way to integrate visual features into the LDA model using unsupervised clusters of images. The clusters are directly interpretable and improve on our evaluation tasks. (3) We provide two novel ways to extend the bimodal models to support three or more modalities. We

find that the three-, four-, and five-dimensional models significantly outperform models using only one or two modalities, and that nontextual modalities each provide separate, disjoint knowledge that cannot be forced into a shared, latent structure.

Combining PCFG-LA Models with Dual Decomposition: A Case Study with Function Labels and Binarization

Joseph Le Roux, Antoine Rozenknop, Jennifer Foster
Saturday 7:30–9:00 — Princessa Ballroom and Foyer

It has recently been shown that different NLP models can be effectively combined using dual decomposition. In this paper we demonstrate that PCFG-LA parsing models are suitable for combination in this way. We experiment with the different models which result from alternative methods of extracting a grammar from a treebank (retaining or not function labels, left binarization versus right binarization) and achieve a labeled Parseval f- score of 92.4 on Wall Street Journal Section 23 - this represents an absolute improvement of 0.7 and an error reduction rate of 7% over a strong PCFG-LA product-model baseline. Although we experiment only with binarization and function labels in this study, there is much scope for applying this approach to other grammar extraction strategies.

Feature Noising for Log-Linear Structured Prediction

Sida Wang, Mengqiu Wang, Stefan Wager, Percy Liang, Christopher D. Manning
Saturday 7:30–9:00 — Princessa Ballroom and Foyer

NLP models have many and sparse features, and regularization is key for balancing model overfitting versus underfitting. A recently repopularized form of regularization is to generate fake training data by repeatedly adding noise to real data. We reinterpret this noising as an explicit regularizer, and approximate it with a second-order formula that can be used during training without actually generating fake data. We show how to apply this method to structured prediction using multinomial logistic regression and linear-chain CRFs. We tackle the key challenge of developing a dynamic program to compute the gradient of the regularizer efficiently. The regularizer is a sum over inputs, so we can estimate it more accurately via a semi-supervised or transductive extension. Applied to text classification and NER, our method provides a $\$>\1 % absolute performance gain over use of standard $\$$ LIIS regularization.

Improvements to the Bayesian Topic N-Gram Models

Hiroshi Noji, Daichi Mochihashi, Yusuke Miyao
Saturday 7:30–9:00 — Princessa Ballroom and Foyer

One of the language phenomena that n-gram language model fails to capture is the topic information of a given situation. We advance the previous study of the Bayesian topic language model by Wallach (2006) in two directions: one, investigating new priors to alleviate the sparseness problem caused by dividing all n-grams into exclusive topics, and two, developing a novel Gibbs sampler that enables moving multiple n-grams across different documents to other topics. Our blocked sampler can efficiently search for higher probability space even with higher order n-grams. In terms of modeling assumption, we found it is effective to assign a topic to only some parts of a document.

An Empirical Study Of Semi-Supervised Chinese Word Segmentation Using Co-Training

Fan Yang, Paul Vozila
Saturday 7:30–9:00 — Princessa Ballroom and Foyer

In this paper we report an empirical study on semi-supervised Chinese word segmentation using co-training. We utilize two segmenters: 1) a word-based segmenter leveraging a word-level language model, and 2) a character-based segmenter using character-level features within a CRF-based sequence labeler. These two segmenters are initially trained with a small amount of segmented data, and then iteratively improve each other using the large amount of unlabelled data. Our experimental results show that co-training captures 20% and 31% of the performance improvement achieved by supervised training with an order of magnitude more data for the SIGHAN Bakeoff 2005 PKU and CU corpora respectively.

Ubertagging: Joint Segmentation and Supertagging for English

Rebecca Dridan

Saturday 7:30–9:00 — Princessa Ballroom and Foyer

A precise syntacto-semantic analysis of English requires a large detailed lexicon with the possibility of treating multiple tokens as a single meaning-bearing unit, a word-with-spaces. However parsing with such a lexicon, as included in the English Resource Grammar, can be very slow. We show that we can apply supertagging techniques over an ambiguous token lattice without resorting to previously used heuristics, a process we call ubertagging. Our model achieves an ubertagging accuracy that can lead to a four to eight fold speed up while improving parser accuracy.

Automatic Knowledge Acquisition for Case Alternation between the Passive and Active Voices in Japanese

Ryohei Sasano, Daisuke Kawahara, Sadao Kurohashi, Manabu Okumura

Saturday 7:30–9:00 — Princessa Ballroom and Foyer

We present a method for automatically acquiring knowledge for case alternation between the passive and active voices in Japanese. By leveraging several linguistic constraints on alternation patterns and lexical case frames obtained from a large Web corpus, our method aligns a case frame in the passive voice to a corresponding case frame in the active voice and finds an alignment between their cases. We then apply the acquired knowledge to a case alternation task and prove its usefulness.

Exploiting Multiple Sources for Open-Domain Hypernym Discovery

Ruiji Fu, Bing Qin, Ting Liu

Saturday 7:30–9:00 — Princessa Ballroom and Foyer

Hypernym discovery aims to extract such noun pairs that one noun is a hypernym of the other. Most previous methods are based on lexical patterns but perform badly on open-domain data. Other work extracts hypernym relations from encyclopedias but has limited coverage. This paper proposes a simple yet effective distant supervision framework for Chinese open-domain hypernym discovery. Given an entity name, we try to discover its hypernyms by leveraging knowledge from multiple sources, i.e., search engine results, encyclopedias, and morphology of the entity name. First, we extract candidate hypernyms from the above sources. Then, we apply a statistical ranking model to select correct hypernyms. A set of novel features is proposed for the ranking model. We also present a heuristic strategy to build a large-scale noisy training data for the model without human annotation. Experimental results demonstrate that our approach outperforms the state-of-the-art methods on a manually labeled test dataset.

A Semantically Enhanced Approach to Determine Textual Similarity

Eduardo Blanco, Dan Moldovan

Saturday 7:30–9:00 — Princessa Ballroom and Foyer

This paper presents a novel approach to determine textual similarity. A layered methodology to transform text into logic forms is proposed, and semantic features are derived from a logic prover. Experimental results show that incorporating the semantic structure of sentences is beneficial. When training data is unavailable, scores obtained from the logic prover in an unsupervised manner outperform supervised methods.

Understanding and Quantifying Creativity in Lexical Composition

Polina Kuznetsova, Jianfu Chen, Yejin Choi

Saturday 7:30–9:00 — Princessa Ballroom and Foyer

Why do certain combination of words such as “disadvantageous peace” or “metal to the petal” appeal to our minds as interesting expressions with a sense of creativity, while other phrases such as “quiet teenager”, or “geometrical base” not as much? We present statistical explorations to understand the characteristics of lexical compositions that give rise to the perception of being original, interesting, and at times even artistic. We first examine various correlates of perceived creativity based on information theoretic measures and the connotation of words, then present experiments based on supervised learning that give us further insights on how different aspects of lexical composition collectively contribute to the perceived creativity.

Sentiment Analysis: How to Derive Prior Polarities from SentiWordNet

Marco Guerini, Lorenzo Gatti, Marco Turchi
Saturday 7:30–9:00 — Princessa Ballroom and Foyer

Assigning a positive or negative score to a word out of context (i.e. a word’s prior polarity) is a challenging task for sentiment analysis. In the literature, various approaches based on SentiWordNet have been proposed. In this paper, we compare the most often used techniques together with newly proposed ones and incorporate all of them in a learning framework to see whether blending them can further improve the estimation of prior polarity scores. Using two different versions of SentiWordNet and testing regression and classification models across tasks and datasets, our learning approach consistently outperforms the single metrics, providing a new state-of-the-art approach in computing words’ prior polarity for sentiment analysis. We conclude our investigation showing interesting biases in calculated prior polarity scores when word Part of Speech and annotator gender are considered.

Simulating Early-Termination Search for Verbose Spoken Queries

Jerome White, Douglas W. Oard, Nitendra Rajput, Marion Zalk
Saturday 7:30–9:00 — Princessa Ballroom and Foyer

Building search engines that can respond to spoken queries with spoken content requires that the system not just be able to find useful responses, but also that it know when it has heard enough about what the user wants to be able to do so. This paper describes a simulation study with queries spoken by non-native speakers that suggests that indicates that finding relevant content is often possible within a half minute, and that combining features based on automatically recognized words with features designed for automated prediction of query difficulty can serve as a useful basis for predicting when that useful content has been found.

Summarizing Complex Events: a Cross-Modal Solution of Storylines Extraction and Reconstruction

Shize Xu, Shanshan Wang, Yan Zhang
Saturday 7:30–9:00 — Princessa Ballroom and Foyer

The rapid development of Web2.0 leads to significant information redundancy. Especially for a complex news event, it is difficult to understand its general idea within a single coherent picture. A complex event often contains branches, intertwining narratives and side news which are all called storylines. In this paper, we propose a novel solution to tackle the challenging problem of storylines extraction and reconstruction. Specifically, we first investigate two requisite properties of an ideal storyline. Then a unified algorithm is devised to extract all effective storylines by optimizing these properties at the same time. Finally, we reconstruct all extracted lines and generate the high-quality story map. Experiments on real-world datasets show that our method is quite efficient and highly competitive, which can bring about quicker, clearer and deeper comprehension to readers.

Image Description using Visual Dependency Representations

Desmond Elliott, Frank Keller
Saturday 7:30–9:00 — Princessa Ballroom and Foyer

Describing the main event of an image involves identifying the objects depicted and predicting the relationships between them. Previous approaches have represented images as unstructured bags of regions, which makes it difficult to accurately predict meaningful relationships between regions. In this paper, we introduce visual dependency representations to capture the relationships between the objects in an image, and hypothesize that this representation can improve image description. We test this hypothesis using a new data set of region-annotated images, associated with visual dependency representations and gold-standard descriptions. We describe two template-based description generation models that operate over visual dependency representations. In an image description task, we find that these models outperform approaches that rely on object proximity or corpus information to generate descriptions on both automatic measures and on human judgements.

Semi-Supervised Feature Transformation for Dependency Parsing

Wenliang Chen, Min Zhang, Yue Zhang
Saturday 7:30–9:00 — Princessa Ballroom and Foyer

In current dependency parsing models, conventional features (i.e. base features) defined over surface words and part-of-speech tags in a relatively high-dimensional feature space may suffer from the data sparseness problem and thus exhibit less discriminative power on unseen data. In this paper, we propose a novel semi-supervised approach to addressing the problem by transforming the base features into high-level features (i.e. meta features) with the help of a large amount of automatically parsed data. The meta features are used together with base features in our final parser. Our studies indicate that our proposed approach is very effective in processing unseen data and features. Experiments on Chinese and English data sets show that the final parser achieves the best-reported accuracy on the Chinese data and comparable accuracy with the best known parsers on the English data.

Leveraging Lexical Cohesion and Disruption for Topic Segmentation

Anca-Roxana Simon, Guillaume Gravier, Pascale Sébillot

Saturday 7:30–9:00 — Princessa Ballroom and Foyer

Topic segmentation classically relies on one of two criteria, either finding areas with coherent vocabulary use or detecting discontinuities. In this paper, we propose a segmentation criterion combining both lexical cohesion and disruption, enabling a trade-off between the two. We provide the mathematical formulation of the criterion and an efficient graph based decoding algorithm for topic segmentation. Experimental results on standard textual data sets and on a more challenging corpus of automatically transcribed broadcast news shows demonstrate the benefit of such a combination. Gains were observed in all conditions, with segments of either regular or varying length and abrupt or smooth topic shifts. Long segments benefit more than short segments. However the algorithm has proven robust on automatic transcripts with short segments and limited vocabulary recurrences.

This Text Has the Scent of Starbucks: A Laplacian Structured Sparsity Model for Computational Branding Analytics

William Yang Wang, Edward Lin, John Kominek

Saturday 7:30–9:00 — Princessa Ballroom and Foyer

We propose a Laplacian structured sparsity model to study computational branding analytics. To do this, we collected customer reviews from Starbucks, Dunkin' Donuts, and other coffee shops across 38 major cities in the Midwest and Northeastern regions of USA. We study the brand related language use through these reviews, with focuses on the brand satisfaction and gender factors. In particular, we perform three tasks: automatic brand identification from raw text, joint brand-satisfaction prediction, and joint brand-gender-satisfaction prediction. This work extends previous studies in text classification by incorporating the dependency and interaction among local features in the form of structured sparsity in a log-linear model. Our quantitative evaluation shows that our approach which combines the advantages of graphical modeling and sparsity modeling techniques significantly outperforms various standard and state-of-the-art text classification algorithms. In addition, qualitative analysis of our model reveals important features of the language uses associated with the specific brands.

Mining New Business Opportunities: Identifying Trend related Products by Leveraging Commercial Intents from Microblogs

Jinpeng Wang, Wayne Xin Zhao, Haitian Wei, Hongfei Yan, Xiaoming Li

Saturday 7:30–9:00 — Princessa Ballroom and Foyer

Hot trends are likely to bring new business opportunities. For example, "Air Pollution" might lead to a significant increase of the sales of related products, e.g., mouth mask. For ecommerce companies, it is very important to make rapid and correct response to these hot trends in order to improve product sales. In this paper, we take the initiative to study the task of how to identify trend related products. The major novelty of our work is that we automatically learn commercial intents revealed from microblogs. We carefully construct a data collection for this task and present quite a few insightful findings. In order to solve this problem, we further propose a graph based method, which jointly models relevance and associativity. We perform extensive experiments and the results showed that our methods are very effective.

Using Topic Modeling to Improve Prediction of Neuroticism and Depression in College Students

Philip Resnik, Anderson Garron, Rebecca Resnik
Saturday 7:30–9:00 — Princessa Ballroom and Foyer

We investigate the value-add of topic modeling in text analysis for depression, and for neuroticism as a strongly associated personality measure. Using Pennebaker’s Linguistic Inquiry and Word Count (LIWC) lexicon to provide baseline features, we show that straightforward topic modeling using Latent Dirichlet Allocation (LDA) yields interpretable, psychologically relevant “themes” that add value in prediction of clinical assessments.

Predicting the Resolution of Referring Expressions from User Behavior

Nikos Engonopoulos, Martin Villalba, Ivan Titov, Alexander Koller
Saturday 7:30–9:00 — Princessa Ballroom and Foyer

We present a statistical model for predicting how the user of an interactive, situated NLP system resolved a referring expression. The model makes an initial prediction based on the meaning of the utterance, and revises it continuously based on the user’s behavior. The combined model outperforms its components in predicting reference resolution and when to give feedback.

Chinese Zero Pronoun Resolution: Some Recent Advances

Chen Chen, Vincent Ng
Saturday 7:30–9:00 — Princessa Ballroom and Foyer

We extend Zhao and Ng’s (2007) Chinese anaphoric zero pronoun resolver by (1) using a richer set of features and (2) exploiting the coreference links between zero pronouns during resolution. Results on OntoNotes show that our approach significantly outperforms two state-of-the-art anaphoric zero pronoun resolvers. To our knowledge, this is the first work to report results obtained by an end-to-end Chinese zero pronoun resolver.

Connecting Language and Knowledge Bases with Embedding Models for Relation Extraction

Jason Weston, Antoine Bordes, Oksana Yakhnenko, Nicolas Usunier
Saturday 7:30–9:00 — Princessa Ballroom and Foyer

This paper proposes a novel approach for relation extraction from free text which is trained to jointly use information from the text and from existing knowledge. Our model is based on scoring functions that operate by learning low-dimensional embeddings of words, entities and relationships from a knowledge base. We empirically show on New York Times articles aligned with Freebase relations that our approach is able to efficiently use the extra information provided by a large subset of Freebase data (4M entities, 23k relationships) to improve over methods that rely on text features alone.

Simple Customization of Recursive Neural Networks for Semantic Relation Classification

Kazuma Hashimoto, Makoto Miwa, Yoshimasa Tsuruoka, Takashi Chikayama
Saturday 7:30–9:00 — Princessa Ballroom and Foyer

In this paper, we present a recursive neural network (RNN) model that works on a syntactic tree. Our model differs from previous RNN models in that the model allows for an explicit weighting of important phrases for the target task. We also propose to average parameters in training. Our experimental results on semantic relation classification show that both phrase categories and task-specific weighting significantly improve the prediction accuracy of the model. We also show that averaging the model parameters is effective in stabilizing the learning and improves generalization capacity. The proposed model marks scores competitive with state-of-the-art RNN-based models.

Improving Statistical Machine Translation with Word Class Models

Joern Wuebker, Stephan Peitz, Felix Rietig, Hermann Ney
Saturday 7:30–9:00 — Princessa Ballroom and Foyer

Automatically clustering words from a monolingual or bilingual training corpus into classes is a widely used technique in statistical natural language processing. We present a very simple and easy to implement method for using these word classes to improve translation quality. It can be applied across different machine transla-

tion paradigms and with arbitrary types of models. We show its efficacy on a small German-to-English and a larger French-to-German translation task with both standard phrase-based and hierarchical phrase-based translation systems for a common set of models. Our results show that with word class models, the baseline can be improved by up to 1.4% BLEU and 1.0% TER on the French-to-German task and 0.3% BLEU and 1.1% TER on the German-to-English task.

Shift-Reduce Word Reordering for Machine Translation

Katsuhiko Hayashi, Katsuhito Sudoh, Hajime Tsukada, Jun Suzuki, Masaaki Nagata

Saturday 7:30–9:00 — Princessa Ballroom and Foyer

This paper presents a novel word reordering model that employs a shift-reduce parser for inversion transduction grammars. Our model uses rich syntax parsing features for word reordering and runs in linear time. We apply it to postordering of phrase-based machine translation (PBMT) for Japanese-to-English patent tasks. Our experimental results show that our method achieves a significant improvement of +3.1 BLEU scores against 30.15 BLEU scores of the baseline PBMT system.

Decoding with Large-Scale Neural Language Models Improves Translation

Ashish Vaswani, Yinggong Zhao, Victoria Fossum, David Chiang

Saturday 7:30–9:00 — Princessa Ballroom and Foyer

We explore the application of neural language models to machine translation. We develop a new model that combines the neural probabilistic language model of Bengio et al., rectified linear units, and noise-contrastive estimation, and we incorporate it into a machine translation system both by reranking k -best lists and by direct integration into the decoder. Our large-scale, large-vocabulary experiments across four language pairs show that our neural language model improves translation quality by up to 1.1 BLEU

Bilingual Word Embeddings for Phrase-Based Machine Translation

Will Y. Zou, Richard Socher, Daniel Cer, Christopher D. Manning

Saturday 7:30–9:00 — Princessa Ballroom and Foyer

We introduce bilingual word embeddings: semantic embeddings associated across two languages in the context of neural language models. We propose a method to learn bilingual embeddings from a large unlabeled corpus, while utilizing MT word alignments to naturally constrain translational equivalence. The new embeddings significantly out-perform baselines in word semantic similarity. A single semantic similarity feature induced with bilingual embeddings adds near half a BLEU point to the results of NIST08 Chinese-English machine translation task.

Application of Localized Similarity for Web Documents

Peter Reberšek, Mateja Verlic

Saturday 7:30–9:00 — Princessa Ballroom and Foyer

In this paper we present a novel approach to automatic creation of anchor texts for hyperlinks in a document pointing to similar documents. Methods used in this approach rank parts of a document based on the similarity to a presumably related document. Ranks are then used to automatically construct the best anchor text for a link inside original document to the compared document. A number of different methods from information retrieval and natural language processing are adapted for this task. Automatically constructed anchor texts are manually evaluated in terms of relatedness to linked documents and compared to baseline consisting of originally inserted anchor texts. Additionally we use crowdsourcing for evaluation of original anchors and automatically constructed anchors. Results show that our best adapted methods rival the precision of the baseline method.

Dependency Language Models for Sentence Completion

Joseph Gubbins, Andreas Vlachos

Saturday 7:30–9:00 — Princessa Ballroom and Foyer

Sentence completion is a challenging semantic modeling task in which models must choose the most appropriate word from a given set to complete a sentence. While a variety of language models have been applied to this task in previous work, none of the existing approaches has attempted to incorporate syntactic information. In this paper we propose to tackle this task using a pair of simple language models in which the probability of

a sentence is estimated as the probability of the lexicalisation of a given syntactic dependency tree. We apply our approach to the Microsoft Research Sentence Completion Challenge and show that it improves on N-gram language models by 8.7 percentage points in accuracy, achieving the highest accuracy reported to date apart from the more complex and expensive to train neural language models.

A Walk-Based Semantically Enriched Tree Kernel Over Distributed Word Representations

Shashank Srivastava, Dirk Hovy, Eduard Hovy
Saturday 7:30–9:00 — Princessa Ballroom and Foyer

In this paper, we propose a walk-based graph kernel that generalizes the notion of tree-kernels to continuous spaces. Our proposed approach subsumes a general framework for word-similarity, and in particular, provides a flexible way to incorporate distributed representations. Using vector representations, such an approach captures both distributional semantic similarities among words as well as the structural relations between them (encoded as the structure of the parse tree). We show an efficient formulation to compute this kernel using simple matrix operations. We present our results on three diverse NLP tasks, showing state-of-the-art results.

Automatic Idiom Identification in Wiktionary

Grace Muzny, Luke Zettlemoyer
Saturday 7:30–9:00 — Princessa Ballroom and Foyer

Online resources, such as Wiktionary, provide an accurate but incomplete source of idiomatic phrases. In this paper, we study the problem of automatically identifying idiomatic dictionary entries with such resources. We train an idiom classifier on a newly gathered corpus of over 60,000 Wiktionary multi-word definitions, incorporating features that model whether phrase meanings are constructed compositionally. Experiments demonstrate that the learned classifier can provide high quality idiom labels, more than doubling the number of idiomatic entries from 7,764 to 18,155 at precision levels of over 65%. These gains also translate to idiom detection in sentences, by simply using known word sense disambiguation algorithms to match phrases to their definitions. In a set of Wiktionary definition example sentences, the more complete set of idioms boosts detection recall by over 28 percentage points.

Elephant: Sequence Labeling for Word and Sentence Segmentation

Kilian Evang, Valerio Basile, Grzegorz Chrupała, Johan Bos
Saturday 7:30–9:00 — Princessa Ballroom and Foyer

Tokenization is widely regarded as a solved problem due to the high accuracy that rule-based tokenizers achieve. But rule-based tokenizers are hard to maintain and their rules language specific. We show that high-accuracy word and sentence segmentation can be achieved by using supervised sequence labeling on the character level combined with unsupervised feature learning. We evaluated our method on three languages and obtained error rates of 0.27 %*cc*(English), 0.35 %*cc*(Dutch) and 0.76 %*cc*(Italian) for our best models.

Detecting Compositionality of Multi-Word Expressions using Nearest Neighbours in Vector Space Models

Douwe Kiela, Stephen Clark
Saturday 7:30–9:00 — Princessa Ballroom and Foyer

We present a novel unsupervised approach to detecting the compositionality of multiword expressions. We compute the compositionality of a phrase through substituting the constituent words with their "neighbours" in a semantic vector space and averaging over the distance between the original phrase and the substituted neighbour phrases. Several methods of obtaining neighbours are presented. The results are compared to existing supervised results and achieve state-of-the-art performance on a verb-object dataset of human compositionality ratings.

Naive Bayes Word Sense Induction

Do Kook Choe, Eugene Charniak

Saturday 7:30–9:00 — Princessa Ballroom and Foyer

We introduce an extended naive Bayes model for word sense induction (WSI) and apply it to a WSI task. The extended model incorporates the idea the words closer to the target word are more relevant in predicting its sense. The proposed model is very simple yet effective when evaluated on SemEval-2010 WSI data.

The VerbCorner Project: Toward an Empirically-Based Semantic Decomposition of Verbs

Joshua K. Hartshorne, Claire Bonial, Martha Palmer

Saturday 7:30–9:00 — Princessa Ballroom and Foyer

This research describes efforts to use crowd-sourcing to improve the validity of the semantic predicates in VerbNet, a lexicon of about 6300 English verbs. The current semantic predicates can be thought of semantic primitives, into which the concepts denoted by a verb can be decomposed. For example, the verb *spray* (of the *Spray* class), involves the predicates *Motion*, *Not*, and *Location*, where the event can be decomposed into an *Agent* causing a *Theme* that was originally not in a particular location to now be in that location. Although VerbNet’s predicates are theoretically well-motivated, systematic empirical data is scarce. This paper describes a recently-launched attempt to address this issue with a series of human judgment tasks, posed to subjects in the form of games.

Where Not to Eat? Improving Public Policy by Predicting Hygiene Inspections Using Online Reviews

Jun Seok Kang, Polina Kuznetsova, Michael Luca, Yejin Choi

Saturday 7:30–9:00 — Princessa Ballroom and Foyer

This paper offers an approach for governments to harness the information contained in social media in order to make public inspections and disclosure more efficient. As a case study, we turn to restaurant hygiene inspections - which are done for restaurants throughout the United States and in most of the world and are a frequently cited example of public inspections and disclosure. We present the first empirical study that shows the viability of statistical models that learn the mapping between textual signals in restaurant reviews and the hygiene inspection records from the Department of Public Health. The learned model achieves over 82% accuracy in discriminating severe offenders from places with no violation, and provides insights into salient cues in reviews that are indicative of the restaurant’s sanitary conditions. Our study suggests that public disclosure policy can be improved by mining public opinions from social media to target inspections and to provide alternative forms of disclosure to customers.

Automatically Identifying Pseudepigraphic Texts

Moshe Koppel, Shachar Seidman

Saturday 7:30–9:00 — Princessa Ballroom and Foyer

The identification of pseudepigraphic texts - texts not written by the authors to which they are attributed - has important historical, forensic and commercial applications. We introduce an unsupervised technique for identifying pseudepigrapha. The idea is to identify textual outliers in a corpus based on the pair-wise similarities of all documents in the corpus. The crucial point is that document similarity not be measured in any of the standard ways but rather be based on the output of a recently introduced algorithm for authorship verification. The proposed method strongly outperforms existing techniques in systematic experiments on a blog corpus.

Sunday, October 20, 2013: Main Conference

Overview

8:00 – 9:00	Breakfast	(Leonesa Foyer)
9:15 –10:30	Invited talk: Fernando Pereira	(Leonesa Ballroom)
10:30–11:00	Break	(Leonesa Foyer)
11:00–12:35	Parallel Sessions	
	<i>Machine Learning for NLP</i>	(Leonesa I and II)
	<i>Summarization and Generation</i>	(Leonesa III)
	<i>Information Extraction and Social Media Analysis</i>	(Eliza Anderson Amphitheater)
12:35– 2:00	Lunch	
2:00 – 3:35	Parallel Sessions	
	<i>Machine Translation II</i>	(Leonesa I and II)
	<i>Semantics II</i>	(Leonesa III)
	<i>Opinion Mining and Sentiment Analysis I</i>	(Eliza Anderson Amphitheater)
3:35 – 4:05	Break	(Leonesa Foyer)
4:05 – 5:55	Parallel Sessions	
	<i>Machine Translation III</i>	(Leonesa I and II)
	<i>Information Extraction II</i>	(Leonesa III)
	<i>NLP Applications II</i>	(Eliza Anderson Amphitheater)

Invited talk: Fernando Pereira

Leonesa Ballroom

Meaning in the Wild

Dr Fernando Pereira, Research Director at Google

Sunday 9:15–10:30 — Leonesa Ballroom

This meeting was founded on the premise that analytical approaches to computational linguistics could be beneficially replaced by machine learning from large corpora exhibiting the linguistic behaviors of interest. The successes of that program have been most notable in speech recognition and machine translation, where the behavior of interest is plentiful in the wild: people transcribe speech and translate texts for practical reasons, creating a voluminous record from which our algorithms can learn.

However, when people understand what they are told or what they read, the output of the process is a hidden mental state change, only accessible partially from whatever observable actions are triggered by the change: the desired input-output behavior is not available in the wild, even accepting the dubious assumption that the output can be abstracted away from the mental state where it came about. The standard escape route of enlisting linguists to create annotated training data is hard enough for parsing, and it quickly falls apart for semantics, even for seemingly ?constrained? tasks like coreference.

Nevertheless, meaning is all around us, in how people ask and respond to search queries, in how they write text so that it can be understood by others, in how they annotate their text with hyperlinks, and in many other common behaviors that are to some extent observable in the Web. We also have a growing body of structured text organized so that computers can use it meaningfully, such as WordNet and Freebase. I will take you on a tour of examples from search, coreference, and information extraction that show small successes and big failures in understanding, asking questions about the potential and limitations of our current approaches along the way. I'll not give you a complete recipe for machine understanding, but I hope you'll find the examples and research questions fun and useful.

Machine Learning for NLP

Leonesa I and II

Chair: Chris Dyer

Dual Coordinate Descent Algorithms for Efficient Large Margin Structured Prediction [TACL]

Ming-Wei Chang, Wen-tau Yih
Sunday 11:00–11:25 — Leonesa I and II

Due to the nature of complex NLP problems, structured prediction algorithms have been important modeling tools for a wide range of tasks. While there exists evidence showing that linear Structural Support Vector Machine (SSVM) algorithm performs better than structured Perceptron, the SSVM algorithm is still less frequently chosen in the NLP community because of its relatively slow training speed. In this paper, we propose a fast and easy-to-implement dual coordinate descent algorithm for SSVMs. Unlike algorithms such as Perceptron and stochastic gradient descent, our method keeps track of dual variables and updates the weight vector more aggressively. As a result, this training process is as efficient as existing online learning methods, and yet derives consistently better models, as evaluated on four benchmark NLP datasets for part-of-speech tagging, named-entity recognition and dependency parsing.

Dynamic Feature Selection for Dependency Parsing

He He, Hal Daumé III, Jason Eisner
Sunday 11:25–11:50 — Leonesa I and II

Feature computation and exhaustive search have significantly restricted the speed of graph-based dependency parsing. We propose a faster framework of dynamic feature selection, where features are added sequentially as needed, edges are pruned early, and decisions are made online for each sentence. We model this as a sequential decision-making problem and solve it by imitation learning techniques. We test our method on 7 languages. Our dynamic parser can achieve accuracies comparable or even superior to parsers using a full set of features, while computing fewer than 30% of the feature templates.

Semi-Supervised Representation Learning for Cross-Lingual Text Classification

Min Xiao, Yuhong Guo
Sunday 11:50–12:15 — Leonesa I and II

Cross-lingual adaptation aims to learn a prediction model in a label-scarce target language by exploiting labeled data from a label-rich source language. An effective cross-lingual adaptation system can substantially reduce the manual annotation effort required in many natural language processing tasks. In this paper, we propose a new cross-lingual adaptation approach for document classification based on learning cross-lingual discriminative distributed representations of words. Specifically, we propose to maximize the log-likelihood of the documents from both language domains under a cross-lingual log-bilinear document model, while minimizing the prediction log-losses of labeled documents. We conduct extensive experiments on cross-lingual sentiment classification tasks of Amazon product reviews. Our experimental results demonstrate the efficacy of the proposed cross-lingual adaptation approach.

Using Crowdsourcing to get Representations based on Regular Expressions

Anders Søgaard, Hector Martinez, Jakob Elming, Anders Johannsen
Sunday 12:15–12:35 — Leonesa I and II

Often the bottleneck in document classification is finding good representations that zoom in on the most important aspects of the documents. Most research uses n-gram representations, but relevant features often occur discontinuously, e.g., 'not ... good' in sentiment analysis. Discontinuous features can be expressed as regular expressions, but even if we limit the regular expressions that we derive from a set of documents to some fixed length, the number becomes astronomical. Some feature combination methods can be seen as digging into the space of regular expressions. In this paper we present experiments getting experts to provide regular expressions, as well as crowdsourced annotation tasks from which regular expressions can be derived. Somewhat surprisingly, it turns out that these crowdsourced feature combinations outperform automatic feature combination methods, as well as expert features, by a very large margin and reduce error by 24-41% over n -gram

representations.

Summarization and Generation

Leonesa III

Chair: Lucy Vanderwende

Overcoming the Lack of Parallel Data in Sentence Compression

Katja Filippova, Yasemin Altun

Sunday 11:00–11:25 — Leonesa III

A major challenge in supervised sentence compression is making use of rich feature representations because of very scarce parallel data. We address this problem and present a method to automatically build a compression corpus with hundreds of thousands of instances on which deletion-based algorithms can be trained. In our corpus, the syntactic trees of the compressions are subtrees of their uncompressed counterparts, and hence supervised systems which require a structural alignment between the input and output can be successfully trained. We also extend an existing unsupervised compression method with a learning module. The new system uses structured prediction to learn from lexical, syntactic and other features. An evaluation with human raters shows that the presented data harvesting method indeed produces a parallel corpus of high quality. Also, the supervised system trained on this corpus gets high scores both from human raters and in an automatic evaluation setting, significantly outperforming a strong baseline.

Fast Joint Compression and Summarization via Graph Cuts

Xian Qian, Yang Liu

Sunday 11:25–11:50 — Leonesa III

Extractive summarization typically uses sentences as summarization units. In contrast, joint compression and summarization can use smaller units such as words and phrases, resulting in summaries containing more information. The goal of compressive summarization is to find a subset of words that maximize the total score of concepts and cutting dependency arcs under the grammar constraints and summary length constraint. We propose an efficient decoding algorithm for fast compressive summarization using graph cuts. Our approach first relaxes the length constraint using Lagrangian relaxation. Then we propose to bound the relaxed objective function by the supermodular binary quadratic programming problem, which can be solved efficiently using graph max-flow/min-cut. Since finding the tightest lower bound suffers from local optimality, we use convex relaxation for initialization. Experimental results on TAC2008 dataset demonstrates our method achieves competitive ROUGE score and has good readability, while is much faster than the integer linear programming (ILP) method.

Inducing Document Plans for Concept-to-Text Generation

Ioannis Konstas, Mirella Lapata

Sunday 11:50–12:15 — Leonesa III

In a language generation system, a content planner selects which elements must be included in the output text and the ordering between them. Recent empirical approaches perform content selection without any ordering and have thus no means to ensure that the output is coherent. In this paper we focus on the problem of generating text from a database and present a trainable end-to-end generation system that includes both content selection and ordering. Content plans are represented intuitively by a set of grammar rules that operate on the document level and are acquired automatically from training data. We develop two approaches: the first one is inspired from Rhetorical Structure Theory and represents the document as a tree of discourse relations between database records; the second one requires little linguistic sophistication and uses tree structures to represent global patterns of database record sequences within a document.

Single-Document Summarization as a Tree Knapsack Problem

Tsutomu Hirao, Yasuhisa Yoshida, Masaaki Nishino, Norihito Yasuda, Masaaki Nagata

Sunday 12:15–12:35 — Leonesa III

Recent studies on extractive text summarization formulate it as a combinatorial optimization problem such as a Knapsack Problem, a Maximum Coverage Problem or a Budgeted Median Problem. These methods successfully improved summarization quality, but they did not consider the rhetorical relations between the textual units of a source document. Thus, summaries generated by these methods may lack logical coherence. This

paper proposes a single document summarization method based on the trimming of a discourse tree. This is a two-fold process. First, we propose rules for transforming a rhetorical structure theory-based discourse tree into a dependency-based discourse tree, which allows us to take a tree-trimming approach to summarization. Second, we formulate the problem of trimming a dependency-based discourse tree as a Tree Knapsack Problem, then solve it with integer linear programming (ILP). Evaluation results showed that our method improved Rouge scores.

Information Extraction and Social Media Analysis

Eliza Anderson Amphitheater

Chair: Soumen Chakrabarti

A Hierarchical Entity-Based Approach to Structuralize User Generated Content in Social Media: A Case of Yahoo! Answers

Baichuan Li, Jing Liu, Chin-Yew Lin, Irwin King, Michael R. Lyu
Sunday 11:00–11:25 — Eliza Anderson Amphitheater

Social media like forums and microblogs have accumulated a huge amount of user generated content (UGC) containing human knowledge. Currently, most of UGC is listed as a whole or in pre-defined categories. This "list-based" approach is simple, but hinders users from browsing and learning knowledge of certain topics effectively. To address this problem, we propose a hierarchical entity-based approach for structuralizing UGC in social media. By using a large-scale entity repository, we design a three-step framework to organize UGC in a novel hierarchical structure called "cluster entity tree (CET)". With Yahoo! Answers as a test case, we conduct experiments and the results show the effectiveness of our framework in constructing CET. We further evaluate the performance of CET on UGC organization in both user and system aspects. From a user aspect, our user study demonstrates that, with CET-based structure, users perform significantly better in knowledge learning than using traditional list-based approach. From a system aspect, CET substantially boosts the performance of two information retrieval models (i.e., vector space model and query likelihood language model).

Semantic Parsing on Freebase from Question-Answer Pairs

Jonathan Berant, Andrew Chou, Roy Frostig, Percy Liang
Sunday 11:25–11:50 — Eliza Anderson Amphitheater

In this paper, we train a semantic parser that scales up to Freebase. Instead of relying on annotated logical forms, which is especially expensive to obtain at large scale, we learn from question-answer pairs. The main challenge in this setting is narrowing down the huge number of possible logical predicates for a given question. We tackle this problem in two ways: First, we build a coarse mapping from phrases to predicates using a knowledge base and a large text corpus. Second, we use a "bridge" operation to generate additional predicates based on neighboring predicates. On the dataset of Cai and Yates (2013), despite not having annotated logical forms, our system outperforms their state-of-the-art parser. Additionally, we collected a more realistic and challenging dataset of question-answer pairs and improves over a natural baseline.

Scaling Semantic Parsers with On-the-Fly Ontology Matching

Tom Kwiatkowski, Eunsol Choi, Yoav Artzi, Luke Zettlemoyer
Sunday 11:50–12:15 — Eliza Anderson Amphitheater

We consider the challenge of learning semantic parsers that scale to large, open-domain problems, such as question answering with Freebase. In such settings, the sentences cover a wide variety of topics and include many phrases whose meaning is difficult to represent in a fixed target ontology. For example, even simple phrases such as 'daughter' and 'number of people living in' cannot be directly represented in Freebase, whose ontology instead encodes facts about gender, parenthood, and population. In this paper, we introduce a new semantic parsing approach that learns to resolve such ontological mismatches. The parser is learned from question-answer pairs, uses a probabilistic CCG to build linguistically motivated logical-form meaning representations, and includes an ontology matching model that adapts the output logical forms for each target ontology. Experiments demonstrate state-of-the-art performance on two benchmark semantic parsing datasets, including a nine point accuracy improvement on a recent Freebase QA corpus.

Classifying Message Board Posts with an Extracted Lexicon of Patient Attributes

Ruihong Huang, Ellen Riloff
Sunday 12:15–12:35 — Eliza Anderson Amphitheater

The goal of our research is to distinguish veterinary message board posts that describe a case involving a specific patient from posts that ask a general question. We create a text classifier that incorporates automatically generated attribute lists for veterinary patients to tackle this problem. Using a small amount of annotated data, we train an information extraction (IE) system to identify veterinary patient attributes. We then apply the IE

system to a large collection of unannotated texts to produce a lexicon of veterinary patient attribute terms. Our experimental results show that using the learned attribute lists to encode patient information in the text classifier yields improved performance on this task.

Machine Translation II

Leonesa I and II

Chair: Kristina Toutanova

Lexical Chain Based Cohesion Models for Document-Level Statistical Machine Translation

Deyi Xiong, Yang Ding, Min Zhang, Chew Lim Tan
Sunday 2:00–2:25 — Leonesa I and II

Lexical chains provide a representation of the lexical cohesion structure of a text. In this paper, we propose two lexical chain based cohesion models to incorporate lexical cohesion into document-level statistical machine translation: 1) a count cohesion model that rewards a hypothesis whenever a chain word occurs in the hypothesis, 2) and a probability cohesion model that further takes chain word translation probabilities into account. We compute lexical chains for each source document to be translated and generate target lexical chains based on the computed source chains via maximum entropy classifiers. We then use the generated target chains to provide constraints for word selection in document-level machine translation through the two proposed lexical chain based cohesion models. We verify the effectiveness of the two models using a hierarchical phrase-based translation system. Experiments on large-scale training data show that they can substantially improve translation quality in terms of BLEU and that the probability cohesion model outperforms previous models based on lexical cohesion devices.

Measuring machine translation errors in new domains [TACL]

Ann Irvine, John Morgan, Marine Carpuat, Hal Daumé III, Dragos Munteanu
Sunday 2:25–2:50 — Leonesa I and II

We develop two techniques for analyzing the effect of porting a machine translation system to a new domain. One is a macro-level analysis that measures how domain shift affects corpus-level evaluation; the second is a micro-level analysis for word-level errors. We apply these methods to understand what happens when a Parliament-trained phrase-based machine translation system is applied in four very different domains: news, medical texts, scientific articles and movie subtitles. We present quantitative and qualitative experiments that highlight opportunities for future research in domain adaptation for machine translation.

A Convex Alternative to IBM Model 2

Andrei Simion, Michael Collins, Cliff Stein
Sunday 2:50–3:15 — Leonesa I and II

The IBM translation models have been hugely influential in statistical machine translation; they are the basis of the alignment models used in modern translation systems. Excluding IBM Model 1, the IBM translation models, and practically all variants proposed in the literature, have relied on the optimization of likelihood functions or similar functions that are non-convex, having multiple local optima. In this paper we introduce a convex relaxation of IBM Model 2, and describe an optimization algorithm for the relaxation based on a subgradient method combined with exponentiated-gradient updates. The approach gives the same level of alignment accuracy as IBM Model 2.

Pair Language Models for Deriving Alternative Pronunciations and Spellings from Pronunciation Dictionaries

Russell Beckley, Brian Roark
Sunday 3:15–3:35 — Leonesa I and II

Pronunciation dictionaries provide a readily available parallel corpus for learning to transduce between character strings and phoneme strings or vice versa. Translation models can be used to derive character-level paraphrases on either side of this transduction, allowing for the automatic derivation of alternative pronunciations or spellings. We examine finite-state and SMT-based methods for these related tasks, and demonstrate that the tasks have different characteristics – finding alternative spellings is harder than alternative pronunciations and benefits from round-trip algorithms when the other does not. We also show that we can increase accuracy by modeling syllable stress.

Semantics II

Leonesa III

Chair: Dipanjan Das

Prior Disambiguation of Word Tensors for Constructing Sentence Vectors

Dimitri Kartsaklis, Mehrnoosh Sadrzadeh

Sunday 2:00–2:25 — Leonesa III

Recent work has shown that compositional-distributional models using element-wise operations on contextual word vectors benefit from the introduction of a prior disambiguation step. The purpose of this paper is to generalise these ideas to tensor-based models, where relational words such as verbs and adjectives are represented by linear maps (higher order tensors) acting on a number of arguments (vectors). We propose disambiguation algorithms for a number of tensor-based models, which we then test on a variety of tasks. The results show that disambiguation can provide better compositional representation even for the case of tensor-based models. Furthermore, we confirm previous findings regarding the positive effect of disambiguation on vector mixture models, and we compare the effectiveness of the two approaches.

Multi-Relational Latent Semantic Analysis

Kai-Wei Chang, Wen-tau Yih, Christopher Meek

Sunday 2:25–2:50 — Leonesa III

We present Multi-Relational Latent Semantic Analysis (MRLSA) which generalizes Latent Semantic Analysis (LSA). MRLSA provides an elegant approach to combining multiple relations between words by constructing a 3-way tensor. Similar to LSA, a low-rank approximation of the tensor is derived using a tensor decomposition. Each word in the vocabulary is thus represented by a vector in the latent semantic space and each relation is captured by a latent square matrix. The degree of two words having a specific relation can then be measured through simple linear algebraic operations. We demonstrate that by integrating multiple relations from both homogeneous and heterogeneous information sources, MRLSA achieves state-of-the-art performance on existing benchmark datasets for two relations, antonymy and is-a.

A Study on Bootstrapping Bilingual Vector Spaces from Non-Parallel Data (and Nothing Else)

Ivan Vulić, Marie-Francine Moens

Sunday 2:50–3:15 — Leonesa III

We present a new language pair agnostic approach to inducing bilingual vector spaces from non-parallel data without any other resource in a bootstrapping fashion. The paper systematically introduces and describes all key elements of the bootstrapping procedure: (1) starting point or seed lexicon, (2) the confidence estimation and selection of new dimensions of the space, and (3) convergence. We test the quality of the induced bilingual vector spaces, and analyze the influence of the different components of the bootstrapping approach in the task of bilingual lexicon extraction (BLE) for two language pairs. Results reveal that, contrary to conclusions from prior work, the seeding of the bootstrapping process has a heavy impact on the quality of the learned lexicons. We also show that our approach outperforms the best performing fully corpus-based BLE methods on these test sets.

Deriving Adjectival Scales from Continuous Space Word Representations

Joo-Kyung Kim, Marie-Catherine de Marneffe

Sunday 3:15–3:35 — Leonesa III

Continuous space word representations extracted from neural network language models have been used effectively for natural language processing, but until recently it was not clear whether the spatial relationships of such representations were interpretable. Mikolov et al. (2013) show that these representations do capture syntactic and semantic regularities. Here, we push the interpretation of continuous space word representations further by demonstrating that vector offsets can be used to derive adjectival scales (e.g., okay < good < excellent). We evaluate the scales on the indirect answers to yes/no questions corpus (de Marneffe et al., 2010). We obtain 72.8% accuracy, which outperforms previous results (60%) on this corpus and highlights the quality of the scales extracted, providing further support that the continuous space word representations are meaningful.

Opinion Mining and Sentiment Analysis I

Eliza Anderson Amphitheater

Chair: Oren Tsur

Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank

Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D. Manning, Andrew Ng, Christopher Potts

Sunday 2:00–2:25 — Eliza Anderson Amphitheater

Semantic word spaces have been very useful but cannot express the meaning of longer phrases in a principled way. Further progress towards understanding compositionality in tasks such as sentiment detection requires richer supervised training and evaluation resources and more powerful models of composition. To remedy this, we introduce a Sentiment Treebank. It includes fine grained sentiment labels for 215,154 phrases in the parse trees of 11,855 sentences and presents new challenges for sentiment compositionality. To address them, we introduce the Recursive Neural Tensor Network. When trained on the new treebank, this model outperforms all previous methods on several metrics. It pushes the state of the art in single sentence positive/negative classification from 80% up to 85.4%. The accuracy of predicting fine-grained sentiment labels for all phrases reaches 80.7%, an improvement of 9.7% over bag of features baselines. Lastly, it is the only model that can accurately capture the effect of contrastive conjunctions as well as negation and its scope at various tree levels for both positive and negative phrases.

Open Domain Targeted Sentiment

Margaret Mitchell, Jacqui Aguilar, Theresa Wilson, Benjamin Van Durme

Sunday 2:25–2:50 — Eliza Anderson Amphitheater

We propose a novel approach to sentiment analysis for a low resource setting. The intuition behind this work is that sentiment expressed towards an entity, targeted sentiment, may be viewed as a span of sentiment expressed across the entity. This representation allows us to model sentiment detection as a sequence tagging problem, jointly discovering people and organizations along with whether there is sentiment directed towards them. We compare performance in both Spanish and English on microblog data, using only a sentiment lexicon as an external resource. By leveraging linguistically informed features within conditional random fields (CRFs) trained to minimize empirical risk, our best models in Spanish significantly outperform a strong baseline, and reach around 90% accuracy on the combined task of named entity recognition and sentiment prediction. Our models in English, trained on a much smaller dataset, are not yet statistically significant against their baselines.

Exploiting Domain Knowledge in Aspect Extraction

Zhiyuan Chen, Arjun Mukherjee, Bing Liu, Meichun Hsu, Malu Castellanos, Riddhiman Ghosh

Sunday 2:50–3:15 — Eliza Anderson Amphitheater

Aspect extraction is one of the key tasks in sentiment analysis. In recent years, statistical models have been used for the task. However, such models without any domain knowledge often produce aspects that are not interpretable in applications. To tackle the issue, some knowledge-based topic models have been proposed, which allow the user to input some prior domain knowledge to generate coherent aspects. However, existing knowledge-based topic models have several major shortcomings, e.g., little work has been done to incorporate the cannot-link type of knowledge or to automatically adjust the number of topics based on domain knowledge. This paper proposes a more advanced topic model, called MC-LDA (LDA with m-set and c-set), to address these problems, which is based on an Extended generalized Pólya urn (E-GPU) model (which is also proposed in this paper). Experiments on real-life product reviews from a variety of domains show that MC-LDA outperforms the existing state-of-the-art models markedly.

Machine Translation III

Leonesa I and II

Chair: Taro Watanabe

Dependency-Based Decipherment for Resource-Limited Machine Translation

Qing Dou, Kevin Knight

Sunday 4:05–4:30 — Leonesa I and II

We introduce dependency relations into deciphering foreign languages and show that dependency relations help improve the state-of-the-art deciphering accuracy by over 500%. We learn a translation lexicon from large amounts of genuinely non parallel data with decipherment to improve a phrase-based machine translation system trained with limited parallel data. In experiments, we observe BLEU gains of 1.2 to 1.8 across three different test sets.

Translating into Morphologically Rich Languages with Synthetic Phrases

Victor Chahuneau, Eva Schlinger, Noah A. Smith, Chris Dyer

Sunday 4:30–4:55 — Leonesa I and II

Translation into morphologically rich languages is an important but recalcitrant problem in MT. We present a simple and effective approach that deals with the problem in two phases. First, a discriminative model is learned to predict inflections of target words from rich source-side annotations. Then, this model is used to create additional sentence-specific word- and phrase-level translations that are added to a standard translation model as "synthetic" phrases. Our approach relies on morphological analysis of the target language, but we show that an unsupervised Bayesian model of morphology can successfully be used in place of a supervised analyzer. We report significant improvements in translation quality when translating from English to Russian, Hebrew and Swahili.

Boosting Cross-Language Retrieval by Learning Bilingual Phrase Associations from Relevance Rankings

Artem Sokokov, Laura Jehl, Felix Hieber, Stefan Riezler

Sunday 4:55–5:20 — Leonesa I and II

We present an approach to learning bilingual n-gram correspondences from relevance rankings of English documents for Japanese queries. We show that directly optimizing cross-lingual rankings rivals and complements machine translation-based cross-language information retrieval (CLIR). We propose an efficient boosting algorithm that deals with very large cross-product spaces of word correspondences. We show in an experimental evaluation on patent prior art search that our approach, and in particular a consensus-based combination of boosting and translation-based approaches, yields substantial improvements in CLIR performance. Our training and test data are made publicly available.

Recurrent Continuous Translation Models

Nal Kalchbrenner, Phil Blunsom

Sunday 5:20–5:55 — Leonesa I and II

We introduce a class of probabilistic continuous translation models called Recurrent Continuous Translation Models that are purely based on continuous representations for words, phrases and sentences and do not rely on alignments or phrasal translation units. The models have a generation and a conditioning aspect. The generation of the translation is modelled with a target Recurrent Language Model, whereas the conditioning on the source sentence is modelled with a Convolutional Sentence Model. Through various experiments, we show first that our models obtain a perplexity with respect to gold translations that is >43% lower than that of state-of-the-art alignment-based translation models. Secondly, we show that they are remarkably sensitive to the word order, syntax, and meaning of the source sentence despite lacking alignments. Finally we show that they match a state-of-the-art system when rescoring n-best lists of translations.

Information Extraction II

Leonesa III

Chair: Ellen Riloff

Learning Biological Processes with Global Constraints

Aju Thalappillil Scaria, Jonathan Berant, Mengqiu Wang, Peter Clark, Justin Lewis, Brittany Harding, Christopher D. Manning
Sunday 4:05–4:30 — Leonesa III

Biological processes are complex phenomena involving a series of events that are related to one another through various relationships. Systems that can understand and reason over biological processes would dramatically improve the performance of semantic applications involving inference such as question answering (QA) - specifically “How?” and “Why?” questions. In this paper, we present the task of process extraction, in which events within a process and the relations between the events are automatically extracted from text. We represent processes by graphs whose edges describe a set of temporal, causal and co-reference event-event relations, and characterize the structural properties of these graphs (e.g., the graphs are connected). Then, we present a method for extracting relations between the events, which exploits these structural properties by performing joint inference over the set of extracted relations. On a novel dataset containing 148 descriptions of biological processes (released with this paper), we show significant improvement comparing to baselines that disregard process structure.

Generating Coherent Event Schemas at Scale

Niranjan Balasubramanian, Stephen Soderland, Mausam, Oren Etzioni
Sunday 4:30–4:55 — Leonesa III

Chambers and Jurafsky (2009) demonstrated that event schemas can be automatically induced from text corpora. However, our analysis of their schemas identifies several weaknesses, e.g., some schemas lack a common topic and distinct roles are incorrectly mixed into a single actor. It is due in part to their pair-wise representation that treats subject-verb independently from verb-object. This often leads to subject-verb-object triples that are not meaningful in the real-world. We present a novel approach to inducing open-domain event schemas that overcomes these limitations. Our approach uses co-occurrence statistics of semantically typed relational triples, which we call Rel-grams (relational n-grams). In a human evaluation, our schemas outperform Chambers’s schemas by wide margins on several evaluation criteria. Both Rel-grams and event schemas are freely available to the research community.

Modeling Missing Data in Distant Supervision for Information Extraction [TACL]

Alan Ritter, Luke Zettlemoyer, Mausam, Oren Etzioni
Sunday 4:55–5:20 — Leonesa III

Distant supervision algorithms learn information extraction models given only large readily available databases and text collections. Most previous work has used heuristics for generating labeled data, for example assuming that facts not contained in the database are not mentioned in the text, and facts in the database must be mentioned at least once. In this paper, we propose a new latent-variable approach that models missing data. This provides a natural way to incorporate side information, for instance modeling the intuition that text will often mention rare entities which are likely to be missing in the database. Despite the added complexity introduced by reasoning about missing data, we demonstrate that a carefully designed local search approach to inference is very accurate and scales to large datasets. Experiments demonstrate improved performance for binary and unary relation extraction when compared to learning with heuristic labels, including on average a 27% increase in area under the precision recall curve in the binary case.

Orthonormal Explicit Topic Analysis for Cross-Lingual Document Matching

John Philip McCrae, Philipp Cimiano, Roman Klinger
Sunday 5:20–5:55 — Leonesa III

Cross-lingual topic modelling has applications in machine translation, word sense disambiguation and terminology alignment. Multilingual extensions of approaches based on latent (LSI), generative (LDA, PLSI) as well as explicit (ESA) topic modelling can induce an interlingual topic space allowing documents in different languages to be mapped into the same space and thus to be compared across languages. In this paper, we

present a novel approach that combines latent and explicit topic modelling approaches in the sense that it builds on a set of explicitly defined topics, but then computes latent relations between these. Thus, the method combines the benefits of both explicit and latent topic modelling approaches. We show that on a cross-lingual mate retrieval task, our model significantly outperforms LDA, LSI, and ESA, as well as a baseline that translates every word in a document into the target language.

NLP Applications II

Eliza Anderson Amphitheater

Chair: Min-Yen Kan

Automated Essay Scoring by Maximizing Human-Machine Agreement

Hongbo Chen, Ben He

Sunday 4:05–4:30 — Eliza Anderson Amphitheater

Previous approaches for automated essay scoring (AES) learn a rating model by minimizing either the classification, regression, or pairwise classification loss, depending on the learning algorithm used. In this paper, we argue that the current AES systems can be further improved by taking into account the agreement between human and machine raters. To this end, we propose a rank-based approach that utilizes listwise learning to rank algorithms for learning a rating model, where the agreement between the human and machine raters is directly incorporated into the loss function. Various linguistic and statistical features are utilized to facilitate the learning algorithms. Experiments on the publicly available English essay dataset, Automated Student Assessment Prize (ASAP), show that our proposed approach outperforms the state-of-the-art algorithms, and achieves performance comparable to professional human raters, which suggests the effectiveness of our proposed method for automated essay scoring.

Powergrading: a Clustering Approach to Amplify Human Effort for Short Answer Grading [TACL]

Sumit Basu, Charles Jacobs, Lucy Vanderwende

Sunday 4:30–4:55 — Eliza Anderson Amphitheater

Adjectives like good, great, and excellent are similar in meaning, but differ in intensity. Intensity order information is very useful for language learners as well as in several NLP tasks, but is missing in most lexical resources (dictionaries, WordNet, and thesauri). In this paper, we present a primarily unsupervised approach that uses semantics from Web-scale data (e.g., phrases like good but not excellent) to rank words by assigning them positions on a continuous scale. We rely on Mixed Integer Linear Programming to jointly determine the ranks, such that individual decisions benefit from global information. When ranking English adjectives, our global algorithm achieves substantial improvements over previous work on both pairwise and rank correlation metrics (specifically, 70% pairwise accuracy as compared to only 56% by previous work). Moreover, our approach can incorporate external synonymy information (increasing its pairwise accuracy to 78%) and extends easily to new languages. We also make our code and data freely available.

Success with Style: Using Writing Style to Predict the Success of Novels

Vikas Ganjigunte Ashok, Song Feng, Yejin Choi

Sunday 4:55–5:20 — Eliza Anderson Amphitheater

Predicting the success of literary works is a curious question among publishers and aspiring writers alike. We examine the quantitative connection, if any, between writing style and successful literature. Based on novels over several different genres, we probe the predictive power of statistical stylometry in discriminating successful literary works, and identify characteristic stylistic elements that are more prominent in successful writings. Our study reports for the first time that statistical stylometry can be surprisingly effective in discriminating highly successful literature from less successful counterpart, achieving accuracy up to 84%. Closer analyses lead to several new insights into characteristics of the writing style in successful literature, including findings that are contrary to the conventional wisdom with respect to good writing style and readability.

A Generative Joint, Additive, Sequential Model of Topics and Speech Acts in Patient-Doctor Communication

Byron C. Wallace, Thomas A Trikalinos, M. Barton Laws, Ira B. Wilson, Eugene Charniak

Sunday 5:20–5:55 — Eliza Anderson Amphitheater

We develop a novel generative model of conversation that jointly captures both the topical content and the speech act type associated with each utterance. Our model expresses both token emission and state transition probabilities as log-linear functions of separate components corresponding to topics and speech acts (and their interactions). We apply this model to a dataset comprising annotated patient-physician visits and show that the

proposed joint approach outperforms a baseline univariate model.

Monday, October 21, 2013: Main Conference

Overview

8:00 – 9:00	Breakfast	(Leonesa Foyer)
9:00 –10:35	Parallel Sessions <i>Information Extraction III</i> <i>Opinion Mining and Sentiment Analysis II</i> <i>NLP for Social Media II</i>	(Leonesa I and II) (Leonesa III) (Princessa)
10:35–11:00	Break	(Leonesa Foyer)
11:00–11:45	Parallel Sessions <i>Parsing</i> <i>Semantics III</i> <i>NLP Applications III</i>	(Leonesa I and II) (Leonesa III) (Princessa)
11:45–12:15	SIGDAT business meeting	(Princessa)
12:15– 1:45	Lunch	
1:45 – 3:15	Plenary session I	(Leonesa Ballroom)
3:15 – 3:45	Break	(Leonesa Foyer)
3:45 – 4:45	Plenary Session II	(Leonesa Ballroom)
4:45 – 5:15	Closing session	(Leonesa Ballroom)

Information Extraction III

Leonesa I and II

Chair: Andreas Vlachos

Harvesting Parallel News Streams to Generate Paraphrases of Event Relations

Congle Zhang, Daniel S. Weld
Monday 9:00–9:25 — Leonesa I and II

The distributional hypothesis, which states that words that occur in similar contexts tend to have similar meanings, has inspired several Web mining algorithms for paraphrasing semantically equivalent phrases. Unfortunately, these methods have several drawbacks, such as confusing synonyms with antonyms and causes with effects. This paper introduces three Temporal Correspondence Heuristics, that characterize regularities in parallel news streams, and shows how they may be used to generate high precision paraphrases for event relations. We encode the heuristics in a probabilistic graphical model to create the NEWSSPIKE algorithm for mining news streams. We present experiments demonstrating that NEWSSPIKE significantly outperforms several competitive baselines. In order to spur further research, we provide a large annotated corpus of timestamped news articles as well as the paraphrases produced by NEWSSPIKE.

Relational Inference for Wikification

Xiao Cheng, Dan Roth
Monday 9:25–9:50 — Leonesa I and II

Wikification, commonly referred to as Disambiguation to Wikipedia (D2W), is the task of identifying concepts and entities in text and disambiguating them into the most specific corresponding Wikipedia pages. Previous approaches to D2W focused on the use of local and global statistics over the given text, Wikipedia articles and its link structures, to evaluate context compatibility among a list of probable candidates. However, these methods fail (often, embarrassingly), when some level of text understanding is needed to support Wikification. In this paper we introduce a novel approach to Wikification by incorporating, along with statistical methods, richer relational analysis of the text. We provide an extensible, efficient and modular Integer Linear Programming (ILP) formulation of Wikification that incorporates the entity-relation inference problem, and show that the ability to identify relations in text helps both candidate generation and ranking Wikipedia titles considerably. Our results show significant improvements in both Wikification and the TAC Entity Linking task.

Event Schema Induction with a Probabilistic Entity-Driven Model

Nathanael Chambers
Monday 9:50–10:15 — Leonesa I and II

Event schema induction is the task of learning high-level representations of complex events (e.g., a bombing) and their entity roles (e.g., perpetrator and victim) from unlabeled text. Event schemas have important connections to early NLP research on frames and scripts, as well as modern applications like template extraction. Recent research suggests event schemas can be learned from raw text. Inspired by a pipelined learner based on named entity coreference, this paper presents the first generative model for schema induction that integrates coreference chains into learning. Our generative model is conceptually simpler than the pipelined approach and requires far less training data. It also provides an interesting contrast with a recent HMM-based model. We evaluate on a common dataset for template schema extraction. Our generative model matches the pipeline's performance, and outperforms the HMM by 7 F1 points (20%).

Using Soft Constraints in Joint Inference for Clinical Concept Recognition

Prateek Jindal, Dan Roth
Monday 10:15–10:35 — Leonesa I and II

This paper introduces IQPs (Integer Quadratic Programs) as a way to model joint inference for the task of concept recognition in clinical domain. IQPs make it possible to easily incorporate soft constraints in the optimization framework and still support exact global inference. We show that soft constraints give statistically significant performance improvements when compared to hard constraints.

Opinion Mining and Sentiment Analysis II

Leonesa III

Chair: Bing Liu

Exploring Demographic Language Variations to Improve Multilingual Sentiment Analysis in Social Media

Svitlana Volkova, Theresa Wilson, David Yarowsky

Monday 9:00–9:25 — Leonesa III

Different demographics, e.g., gender or age, can demonstrate substantial variation in their language use, particularly in informal contexts such as social media. In this paper we focus on learning gender differences in the use of subjective language in English, Spanish, and Russian Twitter data, and explore cross-cultural differences in emoticon and hashtag use for male and female users. We show that gender differences in subjective language can effectively be used to improve sentiment analysis, and in particular, polarity classification for Spanish and Russian. Our results show statistically significant relative F-measure improvement over the gender-independent baseline 1.5% and 1% for Russian, 2% and 0.5% for Spanish, and 2.5% and 5% for English for polarity and subjectivity classification.

Opinion Mining in Newspaper Articles by Entropy-Based Word Connections

Thomas Scholz, Stefan Conrad

Monday 9:25–9:50 — Leonesa III

A very valuable piece of information in newspaper articles is the tonality of extracted statements. For the analysis of tonality of newspaper articles either a big human effort is needed, when it is carried out by media analysts, or an automated approach which has to be as precise as possible for a Media Response Analysis (MRA). To this end, we will compare several state-of-the-art approaches for Opinion Mining in newspaper articles in this paper. Furthermore, we will introduce a new technique to extract entropy-based word connections which identifies the word combinations which create a tonality. In the evaluation, we use two different corpora consisting of news articles, by which we show that the new approach achieves better results than the four state-of-the-art methods.

Collective Opinion Target Extraction in Chinese Microblogs

Xinjie Zhou, Xiaojun Wan, Jianguo Xiao

Monday 9:50–10:15 — Leonesa III

Microblog messages pose severe challenges for current sentiment analysis techniques due to some inherent characteristics such as the length limit and informal writing style. In this paper, we study the problem of extracting opinion targets of Chinese microblog messages. Such fine-grained word-level task has not been well investigated in microblogs yet. We propose an unsupervised label propagation algorithm to address the problem. The opinion targets of all messages in a topic are collectively extracted based on the assumption that similar messages may focus on similar opinion targets. Topics in microblogs are identified by hashtags or using clustering algorithms. Experimental results on Chinese microblogs show the effectiveness of our framework and algorithms.

Detecting Promotional Content in Wikipedia

Shruti Bhosale, Heath Vinicombe, Raymond Mooney

Monday 10:15–10:35 — Leonesa III

This paper presents an approach for detecting promotional content in Wikipedia. By incorporating stylistic features, including features based on n-gram and PCFG language models, we demonstrate improved accuracy at identifying promotional articles, compared to using only lexical information and meta-features.

NLP for Social Media II

Princessa

Chair: Chin-Yew Lin

Learning Topics and Positions from Debatepedia

Swapna Gottipati, Minghui Qiu, Yanchuan Sim, Jing Jiang, Noah A. Smith

Monday 9:00–9:25 — Princessa

We explore Debatepedia, a community-authored encyclopedia of sociopolitical debates, as evidence for inferring a low-dimensional, human-interpretable representation in the domain of issues and positions. We introduce a generative model positing latent topics and cross-cutting positions that gives special treatment to person mentions and opinion words. We evaluate the resulting representation's usefulness in attaching opinionated documents to arguments and its consistency with human judgments about positions.

A Unified Model for Topics, Events and Users on Twitter

Qiming Diao, Jing Jiang

Monday 9:25–9:50 — Princessa

With the rapid growth of social media, Twitter has become one of the most widely adopted platforms for people to post short and instant message. On the one hand, people tweets about their daily lives, and on the other hand, when major events happen, people also follow and tweet about them. Moreover, people's posting behaviors on events are often closely tied to their personal interests. In this paper, we try to model topics, events and users on Twitter in a unified way. We propose a model which combines an LDA-like topic model and the Recurrent Chinese Restaurant Process to capture topics and events. We further propose a duration-based regularization component to find bursty events. We also propose to use event-topic affinity vectors to model the association between events and topics. Our experiments shows that our model can accurately identify meaningful events and the event-topic affinity vectors are effective for event recommendation and grouping events by topics.

Authorship Attribution of Micro-Messages

Roy Schwartz, Oren Tsur, Ari Rappoport, Moshe Koppel

Monday 9:50–10:15 — Princessa

Work on authorship attribution has traditionally focused on long texts. In this work, we tackle the question of whether the author of a very short text can be successfully identified. We use Twitter as an experimental testbed. We introduce the concept of an author's unique "signature", and show that such signatures are typical of many authors when writing very short texts. We also present a new authorship attribution feature ("flexible patterns") and demonstrate a significant improvement over our baselines. Our results show that the author of a single tweet can be identified with good accuracy in an array of flavors of the authorship attribution task.

Detection of Product Comparisons - How Far Does an Out-of-the-Box Semantic Role Labeling System Take You?

Wiltrud Kessler, Jonas Kuhn

Monday 10:15–10:35 — Princessa

This short paper presents a pilot study investigating the training of a standard Semantic Role Labeling (SRL) system on product reviews for the new task of detecting comparisons. An (opinionated) comparison consists of a comparative "predicate" and up to three "arguments": the entity evaluated positively, the entity evaluated negatively, and the aspect under which the comparison is made. In user-generated product reviews, the "predicate" and "arguments" are expressed in highly heterogeneous ways; but since the elements are textually annotated in existing datasets, SRL is technically applicable. We address the interesting question how well training an out-of-the-box SRL model works for English data. We observe that even without any feature engineering or other major adaptations to our task, the system outperforms a reasonable heuristic baseline in all steps (predicate identification, argument identification and argument classification) and in three different datasets.

Parsing

Leonesa I and II

Chair: Keith Hall

A Multi-Teraflop Constituency Parser using GPUs

John Canny, David Hall, Dan Klein

Monday 11:00–11:25 — Leonesa I and II

Constituency parsing with rich grammars remains a computational challenge. Graphics Processing Units (GPUs) have previously been used to accelerate CKY chart evaluation, but gains over CPU parsers have been modest. In this paper, we describe a collection of new techniques that enable chart evaluation at close to the GPU's practical maximum speed (a Teraflop for recent GPUs), or around a half-trillion rule evaluations per second. Net parser performance on a 4-GPU system is over 1 thousands sentences/second (1 trillion rules/sec), for unpruned CKY evaluation of the Berkeley Parser Grammar. The techniques we introduce include grammar compilation, recursive symbol clustering and blocking, and cache-sharing.

Fish Transporters and Miracle Homes: How Compositional Distributional Semantics can Help NP Parsing

Angeliki Lazaridou, Eva Maria Vecchi, Marco Baroni

Monday 11:25–11:45 — Leonesa I and II

In this work, we argue that measures that have been shown to quantify the degree of semantic plausibility of phrases, as obtained from their compositionally-derived distributional semantic representations, can resolve syntactic ambiguities. We exploit this idea to choose the correct parsing of NPs (e.g., (live fish) transporter rather than live (fish transporter)). We show that our plausibility cues outperform a strong baseline and significantly improve performance when used in combination with state-of-the-art features.

Semantics III

Leonesa III

Chair: Martha Palmer

Learning Distributions over Logical Forms for Referring Expression Generation

Nicholas FitzGerald, Yoav Artzi, Luke Zettlemoyer

Monday 11:00–11:25 — Leonesa III

We present a new approach to referring expression generation, casting it as a density estimation problem where the goal is to learn distributions over logical expressions identifying sets of objects in the world. Despite an extremely large space of possible expressions, we demonstrate effective learning of a globally normalized log-linear distribution. This learning is enabled by a new, multi-stage approximate inference technique that uses a pruning model to construct only the most likely logical forms. We train and evaluate the approach on a new corpus of references to sets of visual objects. Experiments show the approach is able to learn accurate models, which generate over 87% of the expressions people used. Additionally, on the previously studied special case of single object reference, we show a 35% relative error reduction over previous state of the art.

Learning to Rank Lexical Substitutions

György Szarvas, Róbert Busa-Fekete, Eyke Hüllermeier

Monday 11:25–11:45 — Leonesa III

The problem to replace a word with a synonym that fits well in its sentential context is known as the lexical substitution task. In this paper, we tackle this task as a supervised ranking problem. Given a dataset of target words, their sentential contexts and the potential substitutions for the target words, the goal is to train a model that accurately ranks the candidate substitutions based on their contextual fitness. As a key contribution, we customize and evaluate several learning-to-rank models to the lexical substitution task, including classification-based and regression-based approaches. On two datasets widely used for lexical substitution, our best models significantly advance the state-of-the-art.

NLP Applications III

Princessa

Chair: Richart Sproat

Identifying Manipulated Offerings on Review Portals

Jiwei Li, Myle Ott, Claire Cardie

Monday 11:00–11:25 — Princessa

Recent work has developed supervised methods for detecting deceptive opinion spam, fake reviews written to sound authentic and deliberately mislead readers. We are instead interested in identifying which offerings (e.g., hotels, restaurants) are posting fake reviews and develop a semi-supervised approach to the task that is based on a manifold ranking algorithm and requires only a small set of labeled truthful and deceptive reviews for training. With no gold standard information to indicate with certainty which offerings are posting fake reviews, we introduce a novel evaluation procedure that ranks artificial instances of real offerings for which the number of deceptive reviews is known. Experiments are conducted on a newly constructed dataset of hotel reviews and results show that the proposed method outperforms state-of-art learning baselines.

Well-Argued Recommendation: Adaptive Models Based on Words in Recommender Systems

Julien Gaillard, Marc El-Beze, Eitan Altman, Emmanuel Ethis

Monday 11:25–11:45 — Princessa

Recommendation systems take advantage of products and users information in order to propose items to consumers. Collaborative, content-based and a few hybrid RS have been developed in the past. In contrast, we propose a new domain-independent semantic RS. By providing textually well-argued recommendations, we aim to give more responsibility to the end user in his decision to purchase or not the recommended item. The system includes a new similarity measure keeping up both the accuracy of rating predictions and coverage. We propose an innovative way to apply a fast adaptation scheme at a semantic level, providing recommendations and arguments in phase with the very recent past. We have performed several experiments on films data, providing textually well-argued recommendations.

Plenary session I

Leonesa Ballroom

Chair: Anna Korhonen

Regularized Minimum Error Rate Training

Michel Galley, Chris Quirk, Colin Cherry, Kristina Toutanova

Monday 1:45–2:15 — Leonesa Ballroom

Minimum Error Rate Training (MERT) remains one of the preferred methods for tuning linear parameters in machine translation systems, yet, it faces significant issues. First, MERT is an unregularized learner and is therefore easily prone to overfitting. Second, it incorporates a difficult non-convex optimization problem that becomes intricate with many parameters to tune. To address these issues, we study the addition of a regularization term to the MERT objective function. Since standard regularizers such as L2 are inapplicable to MERT due to the scale invariance of its objective function, we turn to two regularizers—L0 and a modification of L2—and present methods for efficiently integrating them during search. To improve search in large parameter spaces, we also present a new direction finding algorithm that uses the gradient of expected BLEU to orient MERT's exact line searches. Experiments with up to 3600 features show that these extensions of MERT yield results comparable to PRO, a learner often used with large feature sets.

Of Words, Eyes and Brains: Correlating Image-Based Distributional Semantic Models with Neural Representations of Concepts

Andrew J. Anderson, Elia Bruni, Ulisse Bordignon, Massimo Poesio, Marco Baroni

Monday 2:15–2:45 — Leonesa Ballroom

Traditional distributional semantic models extract word meaning representations from co-occurrence patterns of words in text corpora. Recently, the distributional approach has been extended to models that record the co-occurrence of words with visual features in image collections. These image-based models should be complementary to text-based ones, providing a more cognitively plausible view of meaning grounded in visual perception. In this study, we test whether image-based models capture the semantic patterns that emerge from fMRI recordings of the neural signal. Our results indicate that, indeed, there is a significant correlation between image-based and brain-based semantic similarities, and that image-based models complement text-based ones, so that the best correlations are achieved when the two modalities are combined. Despite some unsatisfactory, but explained outcomes (in particular, failure to detect differential association of models with brain areas), the results show, on the one hand, that image-based distributional semantic models can be a precious new tool to explore semantic representation in the brain, and, on the other, that neural data can be used as the ultimate test set to validate artificial semantic models in terms of their cognitive plausibility.

Easy Victories and Uphill Battles in Coreference Resolution

Greg Durrett, Dan Klein

Monday 2:45–3:15 — Leonesa Ballroom

Classical coreference systems encode various syntactic, discourse, and semantic phenomena explicitly, using heterogeneous features computed from hand-crafted heuristics. In contrast, we present a state-of-the-art coreference system that captures such phenomena implicitly, with a small number of homogeneous feature templates examining shallow properties of mentions. Surprisingly, our features are actually more effective than the corresponding hand-engineered ones at modeling these key linguistic phenomena, allowing us to win "easy victories" without crafted heuristics. These features are successful on syntax and discourse; however, they do not model semantic compatibility well, nor do we see gains from experiments with shallow semantic features from the literature, suggesting that this approach to semantics is an "uphill battle." Nonetheless, our final system outperforms the Stanford system (Lee et al. (2011), the winner of the CoNLL 2011 shared task) by 3.5% absolute on the CoNLL metric and outperforms the IMS system (Bjorkelund and Farkas (2012), the best publicly available English coreference system) by 1.9% absolute.

Plenary Session II

Leonesa Ballroom

Chair: Tim Baldwin

Breaking Out of Local Optima with Count Transforms and Model Recombination: A Study in Grammar Induction

Valentin I. Spitzkovsky, Hiyan Alshawi, Daniel Jurafsky

Monday 3:45–4:15 — Leonesa Ballroom

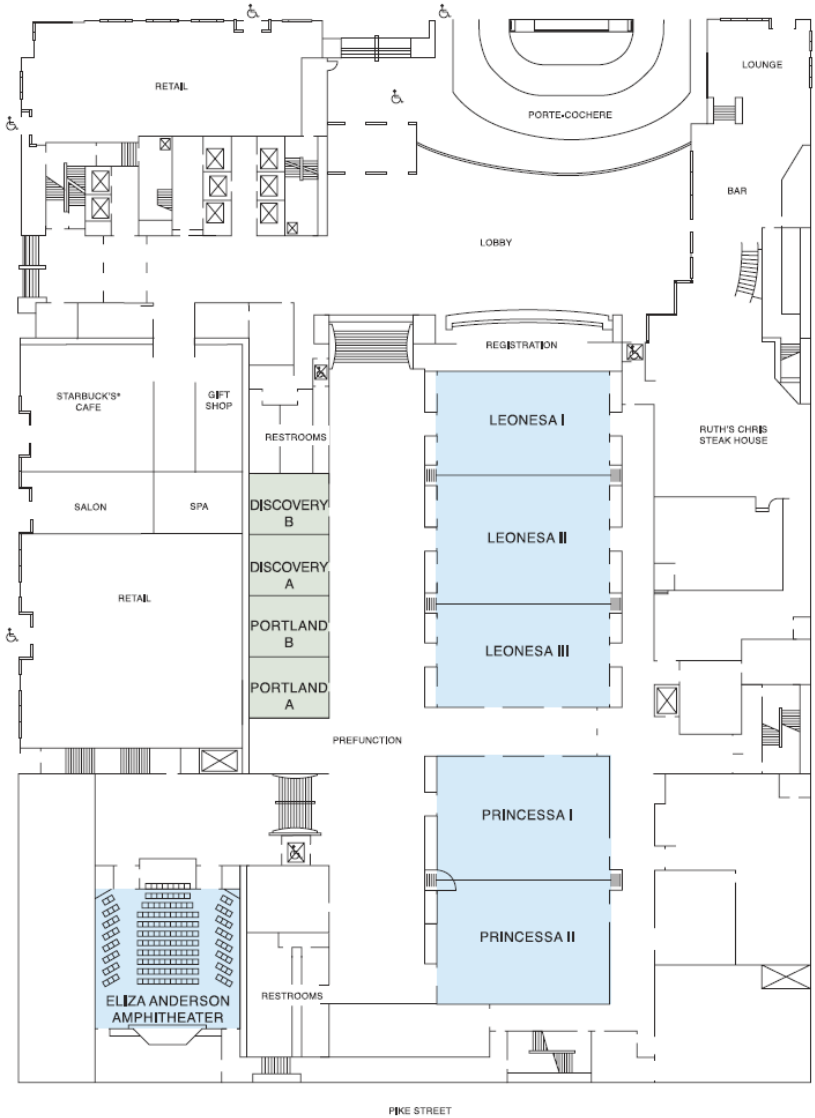
Many statistical learning problems in NLP call for local model search methods. But accuracy tends to suffer with current techniques, which often explore either too narrowly or too broadly: hill-climbers can get stuck in local optima, whereas samplers may be inefficient. We propose to arrange individual local optimizers into organized networks. Our building blocks are operators of two types: (i) transform, which suggests new places to search, via non-random restarts from already-found local optima; and (ii) join, which merges candidate solutions to find better optima. Experiments on grammar induction show that pursuing different transforms (e.g., discarding parts of a learned model or ignoring portions of training data) results in improvements. Groups of locally-optimal solutions can be further perturbed jointly, by constructing mixtures. We used these tools to design several modular dependency grammar induction networks of increasing complexity. Our complete system achieves 48.6% directed dependency accuracy (average over all 19 languages in the 2006/7 CoNLL data): more than 5% higher than the previous state-of-the-art.

Cross-Lingual Discriminative Learning of Sequence Models with Posterior Regularization

Kuzman Ganchev, Dipanjan Das

Monday 4:15–4:45 — Leonesa Ballroom

We present a framework for cross-lingual transfer of sequence information from a resource-rich source language to a resource-impoverished target language that incorporates soft constraints via posterior regularization. To this end, we use automatically word aligned bitext between the source and target language pair, and learn a discriminative conditional random field model on the target side. Our posterior regularization constraints are derived from simple intuitions about the task at hand and from cross-lingual alignment information. We show improvements over strong baselines for two tasks: part-of-speech tagging and named-entity segmentation.



The Blewett Suite is on the 7th floor