Combining Distant and Partial Supervision for Relation Extraction

Gabor Angeli, Julie Tibshirani, Jean Y. Wu, Christopher D. Manning

Stanford University

October 28, 2014



Angeli, Tibshirani, Wu, Manning (Stanford) Combining Distant and Partial Supervision ...

Motivation: Knowledge Base Completion

Unstructured Text

1.47%	And when you are a long bet by the Applement Applement Applements and the		
12.20	Marting Rooms, Bud Vacanta C.	(So.) (Sect.)	
-	Thipste 3 house. Or start heatings (e descharges, his a print t	
WARRING	Visione to Waterda.	- ND - NDV - TUN	
	the local data in the same of the	· Repair · Materials · Telescop	
	1.0.7 standups	a longenty a horar or sharenes	
Fair and			
Tarburb.			
Careed wards	Tably C Carulana	a na nasa	
		Antiputer State	
the second second		the second se	
		Include Sectors	
Particul Differences	11 and along Talants	The Spins Island a family of	
Table & Tomatik	Actual Print of		
100	instantial to be inclusion of our light instant		
	which indicative Reportion/Instation. In 1988, Inc.	· Exterl politicae Earlienten Adrian in	
Tanka .	starband brocks a the book tager, which	which is to be tag	
	superterformed handwork enals fatignate	Remaining the Caracter and	
Particular.	Falory is its Typetch Ent Har, and and down 10	to hange Marc and Society/Fairly	
1000	and A. Ball the Line hand Whatled	· · · Print Dates Faire could be	
10000	research and thread all the heads. Once of	data Tarta of Theiraga and Indiana	

The New York Times





Structured Knowledge Base



Born	Barack Hussein Obama II August 4, 1961 (age 52) Honolulu, Hawaii, U.S.
Political party	Democratic
Spouse(s)	Michelle LaVaughn Robinson (m. 1992-present)
Children	Malia Ann Obama (b. 1998) Natasha Obama (b. 2001)



Angeli, Tibshirani, Wu, Manning (Stanford) Combining Distant and Partial Supervision ...

-

Where is Chris Manning from?



Angeli, Tibshirani, Wu, Manning (Stanford) Combining Distant and Partial Supervision ...

A DESCRIPTION OF THE OWNER OWNER OF THE OWNER OWNER OF THE OWNER	Christopher D. Manning ^{en}		
	People /people	≈ Freebase Commons	
	Person /people/person		х
	Date of birth /people/person/date_of_birth		
	9/18/1965		
	Place of birth /people/person/place_of_birth		
	Country of nationality /people/person/nationality		
	Gender /people/person/gender		
	Male		
	Profession /people/person/profession		
	-		



October 28, 2014 3 / 19

Christopher Manning

Professor of Linguistics and Computer Science

Natural Language Processing Group, Stanford University

Brief Bio

- I'm Australian ("I come from a land of wide open spaces ...")
- BA (Hons) Australian National University 1989 (majors in mathematics, computer science and linguistics)
- · PhD Stanford Linguistics 1995
- · Asst Professor Carnegie Mellon University Computational Linguistics Program 1994-96
- · Lecturer University of Sydney Dept of Linguistics 1996-99
- · Asst Professor Stanford University Depts of Computer Science and Linguistics 1999-2006
- Assoc Professor Stanford University Depts of Linguistics and Computer Science 2006-2012
- · Professor Stanford University Depts of Linguistics and Computer Science 2012-





(日)

Australia

Christopher D. Manning, origin

Feedback



Angeli, Tibshirani, Wu, Manning (Stanford) Combining Distant and Partial Supervision

Relation Extraction

Input: Sentences containing (entity, slot value). **Output**: Relation between entity and slot value.



Angeli, Tibshirani, Wu, Manning (Stanford) Combining Distant and Partial Supervision

Consider two approaches:

• **Supervised:** Trivial as a supervised classifier. Training data: {(sentence, relation)}. *But...*



Consider two approaches:

• **Supervised:** Trivial as a supervised classifier. Training data: {(sentence, relation)}. *But...* this training data is expensive to produce.



Consider two approaches:

- **Supervised:** Trivial as a supervised classifier. Training data: {(sentence, relation)}. *But...* this training data is expensive to produce.
- **Distantly Supervised:** Artificially produce "supervised" data. Training data: {(entity, relation, slot value)}. *But...*



Consider two approaches:

- **Supervised:** Trivial as a supervised classifier. Training data: {(sentence, relation)}. *But...* this training data is expensive to produce.
- **Distantly Supervised:** Artificially produce "supervised" data. Training data: {(entity, relation, slot value)}. *But...* this training data is much more noisy.



Adding carefully selected supervision improves distantly supervised relation extraction.



Angeli, Tibshirani, Wu, Manning (Stanford) Combining Distant and Partial Supervision

- What is "carefully selected": Propose new active learning criterion.
- Evaluate a number of questions:



- What is "carefully selected": Propose new active learning criterion.
- Evaluate a number of questions:
 - Is the proposed criterion better than other methods?



- What is "carefully selected": Propose new active learning criterion.
- Evaluate a number of questions:
 - Is the proposed criterion better than other methods?
 - Where is the supervision helping?



- What is "carefully selected": Propose new active learning criterion.
- Evaluate a number of questions:
 - Is the proposed criterion better than other methods?
 - Where is the supervision helping?
 - How far can we get with a supervised classifier?



Distant Supervision

(Barack Obama, EmployedBy, United States)



October 28, 2014 6 / 19

< □ > < 同 > < 回 > < 回 > < 回 >

Multiple-Instance Multiple-Label (MIML) Learning

(Barack Obama, EmployedBy, United States)



Distant Supervision





Angeli, Tibshirani, Wu, Manning (Stanford) Combining Distant and Partial Supervision

Distant Supervision



↑ Barack Obama is the 44th and current president of the United States



October 28, 2014 7 / 19

Multiple-Instance





Angeli, Tibshirani, Wu, Manning (Stanford) Combining Distant and Partial Supervision

イロト イポト イヨト イヨト

Multiple-Instance





Multiple-Instance Multiple-Label (MIML-RE)





Image: A math a math

Old problem: Supervision is expensive, but very useful.

Old solution: Active learning!



Angeli, Tibshirani, Wu, Manning (Stanford) Combining Distant and Partial Supervision

Old problem: Supervision is expensive, but very useful.

Old solution: Active learning!

- Select a subset of latent z to annotate.
- Fix these labels during training.



Old problem: Supervision is expensive, but very useful.

Old solution: Active learning!

- Select a subset of latent z to annotate.
- Fix these labels during training.
- Bonus: this creates a supervised training set.
 - We initialize from a supervised classifier on this training set.



Old problem: Supervision is expensive, but very useful.

Old solution: Active learning!

- Select a subset of latent z to annotate.
- Fix these labels during training.
- Bonus: this creates a supervised training set.
 - We initialize from a supervised classifier on this training set.

Some Statistics

- 1,208,524 latent z which we could annotate.
- \$0.13 per annotation.
- \$160,000 to annotate everything.



Old problem: Supervision is expensive, but very useful.

Old solution: Active learning!

- Select a subset of latent *z* to annotate.
- Fix these labels during training.
- Bonus: this creates a supervised training set.
 - We initialize from a supervised classifier on this training set.

Some Statistics

- 1,208,524 latent *z* which we could annotate.
- \$0.13 per annotation.
- \$160,000 to annotate everything.

New spin: Have to get it right the first time.



Train *k* MIML-RE models on *k* subsets of the data.





Angeli, Tibshirani, Wu, Manning (Stanford) Combining Distant and Partial Supervision ...

Train k MIML-RE models on k subsets of the data.

2 For each latent z, each trained model c predicts a multinomial $P_c(z)$.



Angeli, Tibshirani, Wu, Manning (Stanford) Combining Distant and Partial Supervision ...

Train k MIML-RE models on k subsets of the data.

- **2** For each latent z, each trained model c predicts a multinomial $P_c(z)$.
- **2** Calculate Jensen-Shannon divergence: $\frac{1}{k}\sum_{c=1}^{k} \text{KL}(p_{c}(z)||p_{mean}(z))$.



Train k MIML-RE models on k subsets of the data.

- **2** For each latent z, each trained model c predicts a multinomial $P_c(z)$.
- **3** Calculate Jensen-Shannon divergence: $\frac{1}{k}\sum_{c=1}^{k} \text{KL}(p_{c}(z)||p_{mean}(z))$.
- We have measure of disagreement for each z.

10 / 19

October 28, 2014

Train k MIML-RE models on k subsets of the data.

2 For each latent z, each trained model c predicts a multinomial $P_c(z)$.

- **3** Calculate Jensen-Shannon divergence: $\frac{1}{k}\sum_{c=1}^{k} \text{KL}(p_{c}(z)||p_{mean}(z))$.
- We have measure of disagreement for each z.

Three selection criteria

• Sample uniformly (uniform).



Train k MIML-RE models on k subsets of the data.

2 For each latent z, each trained model c predicts a multinomial $P_c(z)$.

- **③** Calculate Jensen-Shannon divergence: $\frac{1}{k}\sum_{c=1}^{k} \text{KL}(p_{c}(z)||p_{mean}(z))$.
- We have measure of disagreement for each z.

Three selection criteria

- Sample uniformly (uniform).
- Take z with highest disagreement (highJS).



Train k MIML-RE models on k subsets of the data.

2 For each latent z, each trained model c predicts a multinomial $P_c(z)$.

- **③** Calculate Jensen-Shannon divergence: $\frac{1}{k}\sum_{c=1}^{k} \text{KL}(p_{c}(z)||p_{mean}(z))$.
- We have measure of disagreement for each z.

Three selection criteria

- Sample uniformly (uniform).
- Take *z* with highest disagreement (*highJS*).
- Sample z with highest disagreement (sampleJS).





Angeli, Tibshirani, Wu, Manning (Stanford)








Committee Member Judgments





Sentence

Obama was born in Hawaii Obama grew up in Hawaii Obama Bear visits Hawaii President Obama ... Obama employed president ... Member A born born no relation title employee of

Member B born lived in born title title



Member C no relation born employee of title employee of



Committee Member Judgments







Sentence	Member A	Member B	Member C
Obama was born in Hawaii	born	born	no relation
Obama grew up in Hawaii	born	lived in	born
Obama Bear visits Hawaii	no relation	born	employee of
President Obama	title	title	title
Obama employed president	employee of	title	employee of

Uniform: Often annotates easy sentences.



Committee Member Judgments

	\sim		×.
ų.	Ŷ	¢	
((0)		





Sentence	Member A	Member B	Member C
Obama was born in Hawaii	born	born	no relation
Obama grew up in Hawaii	born	lived in	born
Obama Bear visits Hawaii	no relation	born	employee of
President Obama	title	title	title
Obama employed president	employee of	title	employee of

Uniform: Often annotates easy sentences.

High JS (disagreement): More likely to annotate "rare" sentences.



Committee Member Judgments







Sentence	Member A	Member B	Member C
Obama was born in Hawaii	born	born	no relation
Obama grew up in Hawaii	born	lived in	born
Obama Bear visits Hawaii	no relation	born	employee of
President Obama	title	title	title
Obama employed president	employee of	title	employee of

Uniform: Often annotates easy sentences.

High JS (disagreement): More likely to annotate "rare" sentences. Sample JS (disagreement): Mix of hard and representative sentences.



Experiments

Recall our questions:

- Is the proposed criterion better than other methods?
- Where is the supervision helping?
- How far can we get with a supervised classifier?



Experiments

Recall our questions:

- Is the proposed criterion better than other methods?
- Where is the supervision helping?
- How far can we get with a supervised classifier?

Two experimental setups:

- Slot filling evaluation of Surdeanu et al. (2012).
- Stanford's 2013 TAC-KBP slot filling system



Experiments

Recall our questions:

- Is the proposed criterion better than other methods?
- Where is the supervision helping?
- How far can we get with a supervised classifier?

Two experimental setups:

- Slot filling evaluation of Surdeanu et al. (2012).
- Stanford's 2013 TAC-KBP slot filling system

Bonus: 4.4 F₁ improvement on 2014 TAC-KBP competition



Old News: MIML-RE Works Well

Slot filling evaluation of Surdeanu et al. (2012).



Old News: MIML-RE Works Well

Slot filling evaluation of Surdeanu et al. (2012).



Old News: MIML-RE Works Well

Slot filling evaluation of Surdeanu et al. (2012).



Slot filling evaluation of Surdeanu et al. (2012).



Slot filling evaluation of Surdeanu et al. (2012).



Combining Distant and Partial Supervision ... October 28, 2014

15/19

Slot filling evaluation of Surdeanu et al. (2012).



Slot filling evaluation of Surdeanu et al. (2012).



Slot filling evaluation of Surdeanu et al. (2012).



Slot filling evaluation of Surdeanu et al. (2012).





TAC-KBP 2013 Slot Filling Challenge:

End-to-end task – includes IR + consistency.



- End-to-end task includes IR + consistency.
- **Precision:** facts LDC evaluators judged as correct. **Recall:** facts other teams (including LDC annotators) also found.



- End-to-end task includes IR + consistency.
- Precision: facts LDC evaluators judged as correct. Recall: facts other teams (including LDC annotators) also found.

System	Р	R	F ₁
No active learning	38.0	30.5	33.8



- End-to-end task includes IR + consistency.
- Precision: facts LDC evaluators judged as correct.
 Recall: facts other teams (including LDC annotators) also found.

System	Р	R	\mathbf{F}_1
No active learning	38.0	30.5	33.8
Sample uniformly	34.4	35.0	34.7



- End-to-end task includes IR + consistency.
- Precision: facts LDC evaluators judged as correct.
 Recall: facts other teams (including LDC annotators) also found.

System	Ρ	R	F ₁
No active learning	38.0	30.5	33.8
Sample uniformly	34.4	35.0	34.7
Highest JS disagreement	46.2	30.8	37.0



- End-to-end task includes IR + consistency.
- Precision: facts LDC evaluators judged as correct.
 Recall: facts other teams (including LDC annotators) also found.

System	Ρ	R	F ₁
No active learning	38.0	30.5	33.8
Sample uniformly	34.4	35.0	34.7
Highest JS disagreement	46.2	30.8	37.0
Sample JS disagreement	39.4	36.2	37.7



TAC-KBP 2013 Slot Filling Challenge:

- End-to-end task includes IR + consistency.
- Precision: facts LDC evaluators judged as correct.
 Recall: facts other teams (including LDC annotators) also found.

System	Ρ	R	F ₁
No active learning	38.0	30.5	33.8
Sample uniformly	34.4	35.0	34.7
Highest JS disagreement	46.2	30.8	37.0
Sample JS disagreement	39.4	36.2	37.7

• 2014 TAC-KBP Slot Filling Challenge: $27.6 \rightarrow 32.0 F_1$.



Is initialization or fixing latent zs during EM helping?

• What if we initialize with distant supervision?





Is initialization or fixing latent zs during EM helping?

• What if we initialize with distant supervision?





Is initialization or fixing latent zs during EM helping?

• What if we initialize with distant supervision?





Is initialization or fixing latent zs during EM helping?

• What if we initialize with distant supervision?





Is initialization or fixing latent zs during EM helping?

• What if we initialize with distant supervision?



Hypothesis: Supervision not only *smooths the objective* but provides *better initialization* for the non-convex objective.



- End-to-end task includes IR + consistency.
- Precision: facts LDC evaluators judged as correct. Recall: facts other teams (including LDC annotators) also found.

System	Ρ	R	F ₁
MIML-RE (baseline)	38.0	30.5	33.8



- End-to-end task includes IR + consistency.
- **Precision:** facts LDC evaluators judged as correct. **Recall:** facts other teams (including LDC annotators) also found.

System	Р	R	F ₁
MIML-RE (baseline)	38.0	30.5	33.8
Supervised from SampleJS	33.5	35.0	34.2



TAC-KBP 2013 Slot Filling Challenge:

- End-to-end task includes IR + consistency.
- **Precision:** facts LDC evaluators judged as correct. **Recall:** facts other teams (including LDC annotators) also found.

System	Ρ	R	F ₁
MIML-RE (baseline)	38.0	30.5	33.8
Supervised from SampleJS	33.5	35.0	34.2
MIML-RE Supervised init.	35.1	35.6	35.5
SampleJS	39.4	36.2	37.7



TAC-KBP 2013 Slot Filling Challenge:

- End-to-end task includes IR + consistency.
- Precision: facts LDC evaluators judged as correct. Recall: facts other teams (including LDC annotators) also found.

System	Ρ	R	F ₁
MIML-RE (baseline)	38.0	30.5	33.8
Supervised from SampleJS	33.5	35.0	34.2
MIML-RE Supervised init.	35.1	35.6	35.5
SampleJS	39.4	36.2	37.7

• A bit circular: Need MIML-RE to get supervised examples.



A Case for Supervised Classifiers





Stanford's KBP system (artist rendition)

Supervised Classifier (150 lines + featurizer)



A Case for Supervised Classifiers







Stanford's KBP system (artist rendition)

Supervised Classifier (150 lines + featurizer)



Angeli, Tibshirani, Wu, Manning (Stanford) Combining Distant and Partial Supervision ...

October 28, 2014 18 / 19
A Case for Supervised Classifiers







Stanford's KBP system (artist rendition)

Supervised Classifier (150 lines + featurizer)

Annotating examples: \$1330



Angeli, Tibshirani, Wu, Manning (Stanford) Combining Distant and Partial Supervision ...

A Case for Supervised Classifiers







Stanford's KBP system (artist rendition)

Supervised Classifier (150 lines + featurizer)

 Annotating examples:
 \$1330

 Flight to Qatar:
 \$1027

 Apple 27" Screen:
 \$999



Angeli, Tibshirani, Wu, Manning (Stanford) Combining Distant and Partial Supervision ...

Conclusions

Things you can use:

- New active learning criterion: *Sampling* disagreement between a committee of classifiers.
- Corpus of supervised examples for TAC-KBP relations.
- 4.4 F₁ improvement on 2014 KBP Slot Filling.



Conclusions

Things you can use:

- New active learning criterion: *Sampling* disagreement between a committee of classifiers.
- Corpus of supervised examples for TAC-KBP relations.
- 4.4 F₁ improvement on 2014 KBP Slot Filling.

Things we've learned:

- Example selection is *very* important for performance.
- MIML-RE is sensitive to initialization.
- Supervised classifiers can perform similarly to distantly supervised methods.



Conclusions

Things you can use:

- New active learning criterion: *Sampling* disagreement between a committee of classifiers.
- Corpus of supervised examples for TAC-KBP relations.
- 4.4 F₁ improvement on 2014 KBP Slot Filling.

Things we've learned:

- Example selection is *very* important for performance.
- MIML-RE is sensitive to initialization.
- Supervised classifiers can perform similarly to distantly supervised methods.

Thank You!

Comparison to Pershina et al. (ACL 2014)

Slot filling evaluation of Surdeanu et al. (2012).



Angeli, Tibshirani, Wu, Manning (Stanford) Combining Distant and Partial Supervision ...