Combining Supervised and Unsupervised Ensembles for Knowledge Base Population

Nazneen Fatema Rajani and Raymond J. Mooney

Department of Computer Science, The University of Texas at Austin

INTRODUCTION

- In our past work, we demonstrated **Stacking with auxiliary features (SWAF)** for improving performance of slot filler systems[1]
- Stacking uses **supervised** learning and thus relies on training data
- Sometimes, we might not have access to individual systems and only to the system outputs
- We first use **unsupervised** ensembling to combine such systems without training data and then use **stacking** to combine the **supervised** ensemble and the unsupervised ensemble
- We obtain state-of-the-art results on two KBP tasks for 2015:

1. Cold Start Slot Filling (CSSF)
2. Tri-lingual Entity Discovery and Linking (TEDL)

OVERVIEW OF THE KBP TASKS

1. Cold Start Slot Filling (CSSF)

- Extract **information** (fills) about specific **attributes** (slots) for a set of **entities** (queries) from a given corpus
- Query entities can be PER/ORG/GPE

org: Microsoft
city_of_headquarters Redmond
website microsoft.com
subsidiaries Skype Nokia

org: Microsoft
<eng-NG-31-1007>: Microsoft is a technology company, headquartered in Redmond, Washington that develops…
city_of_headquarters Redmond
Doc ID eng-NG-31-1007
Start Offset 48
End Offset 54

2. Tri-lingual Entity Discovery and Linking (TEDL)

- Find all entity mentions in a corpus of English, Chinese and Spanish documents and link to a **FreeBase KB**
- If there is no KB entry for the entity, systems are expected to cluster all the mentions for that entity using a NIL ID
- Entities can either be PER/ORG/GPE/LOC/FAC

per: Hillary Clinton
Corpus Document <doc id="ENG_NW_001429.nw.xml"> : Hillary Clinton Not Talking About '92 Clinton-Gore Confederate Campaign Button
FreeBase <ID= m.0d06m >: Hillary Diane Rodham Clinton is a US Secretary of State, U.S. Senator, and First Lady of the United States. From 2009 to 2013, she was the 67th Secretary of State, serving under President Barack Obama. She previously represented New York in the U.S. Senate. Before that, as the wife of President Bill Clinton, she was First Lady from 1993 to 2001.

ENSEMBLING ALGORITHM

- For systems common between 2014 and 2015, we use **stacking** to aggregate the outputs
- For systems without training data, we use **constrained optimization** approach described in Weng et al. to aggregate confidence scores
- Use the aggregated unsupervised ensemble as **one more** system to the stacker
- Unsupervised ensemble from 2014 is used as training data for this new aggregated system
- Instances classified as correct are kept and post-processed
- The final output is made to look like its produced by a single system by resolving any conflicts

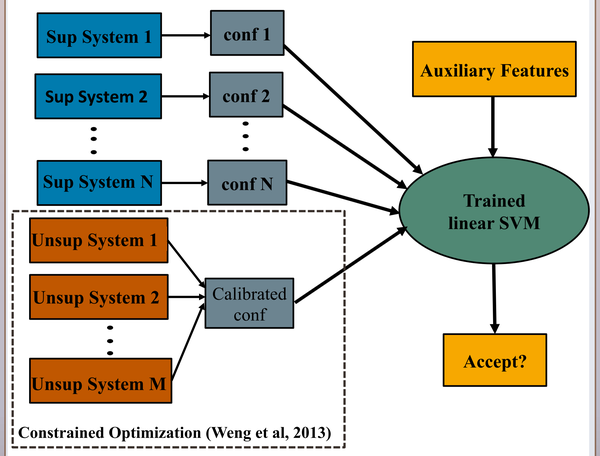


Figure 1: Illustration of our model to ensemble supervised and unsupervised systems using SWAF

	Number of supervised systems (N)			Number of unsupervised systems (M)		
	English	Chinese	Spanish	English	Chinese	Spanish
TEDL	6	4	5	4	3	2
CSSF	10	-	-	13	-	-

Table 1: Number of supervised and unsupervised systems for each of the tasks and languages

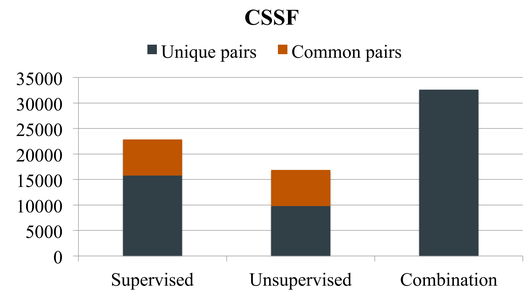
RESULTS

Approach	Precision	Recall	F1
Oracle Voting Baseline (>= 3)	0.4384	0.2720	0.3357
Top ranked CSSF system in 2015 (Angeli et al.)	0.3989	0.3058	0.3462
Stacking approach by Rajani et al.	0.4656	0.3312	0.3871
Combining supervised and unsupervised ensembles	**0.4679**	**0.4314**	**0.4489**

Table 2: Results on the 2015 CSSF task using the official NIST scorer

Approach	Precision	Recall	F1
Oracle Voting Baseline (>= 4)	0.514	0.601	0.554
Top ranked TEDL system in 2015 (Sil et al.)	0.693	0.547	0.611
Stacking approach by Rajani et al.	**0.813**	0.515	0.630
Combining supervised and unsupervised ensembles	0.686	**0.624**	**0.653**

Table 3: Results on the 2015 TEDL task using the official NIST scorer and CEAF metric[6]



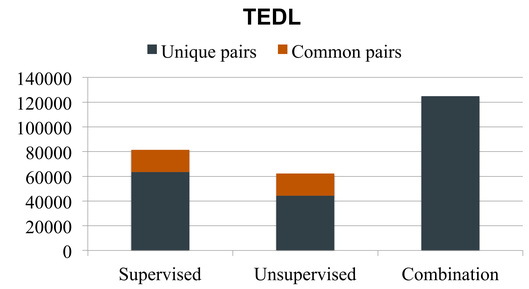


Figure 2: Proportion of unique and common input pairs in the ensemble for the CSSF and TEDL tasks

CONCLUSIONS

- Stacking-based approach to ensembling both **supervised** and **unsupervised** systems is very promising on CSSF and TEDL tasks
- Our model outperforms **top ranked systems** from the 2015 competition as well as several **other ensembling methods** on both tasks
- We obtained state-of-the-art results for **CSSF** and **TEDL**
- Adding the unsupervised ensemble along with the shared systems leads to a substantial **increase in recall**

REFERENCES

1. Nazneen Fatema Rajani and Raymond J. Mooney. 2016. Stacking With Auxiliary Features. ArXiv e-prints
2. Nazneen Fatema Rajani, Vidhoon Viswanathan, YinonBentor, and Raymond J. Mooney. 2015. Stacked Ensembles of Information Extractors for Knowledge-Base Population. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics (ACL2015) , pages 177–187, Beijing, China, July.
3. I-Jeng Wang, Edwina Liu, Cash Costello, and Christine Piatko. 2013. JHUAPL TAC-KBP2013 Slot Filler Validation System. In Proceedings of the Sixth Text Analysis Conference (TAC2013) .
4. Gabor Angeli, Victor Zhong, Danqi Chen, Arun Chaganty, Jason Bolton, Melvin Johnson Premkumar, Panupong Pasupat, Sonal Gupta, and Christopher D. Manning. 2015. Bootstrapped Self Training for Knowledge Base Population. In Proceedings of the Eighth Text Analysis Conference (TAC2015) .
5. Avirup Sil, Georgiana Dinu, and Radu Florian. 2015. The IBM systems for trilingual entity discovery and linking at TAC 2015. In Proceedings of the Eighth Text Analysis Conference (TAC2015).
6. Heng Ji, Joel Nothman, Ben Hachey, and Radu Florian. 2015. Overview of TAC-KBP2015 Tri-lingual Entity Discovery and Linking. In Proceedings of the Eighth Text Analysis Conference (TAC2015) .

ACKNOWLEDGEMENTS

This research was supported by the DARPA DEFT program under AFRL grant FA8750-13-2-0026.

CONTACT

Nazneen Rajani: nrajani@cs.utexas.edu cs.utexas.edu/~nrajani
Ray Mooney: mooney@cs.utexas.edu cs.utexas.edu/~mooney