Seoul, Korea
WWW 2014
The 23rd International World Wide Web Conference
April 7-11, 2014 coex

Linking Entities in #Microposts

ROMIL BANSAL, SANDEEP PANEM, PRIYA RADHAKRISHNAN, MANISH GUPTA, VASUDEVA VARMA

INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY, HYDERABAD

7TH APRIL 2014

Introduction

- Entity Linking is the task of **associating entity name mentions** in text to the correct **referent entities in the knowledge base**, with the goal of understanding and extracting useful information from the document.
- Entity Linking could be helpful for various IR tasks like document classification and **clustering, tags recommendation, relation extraction** etc.

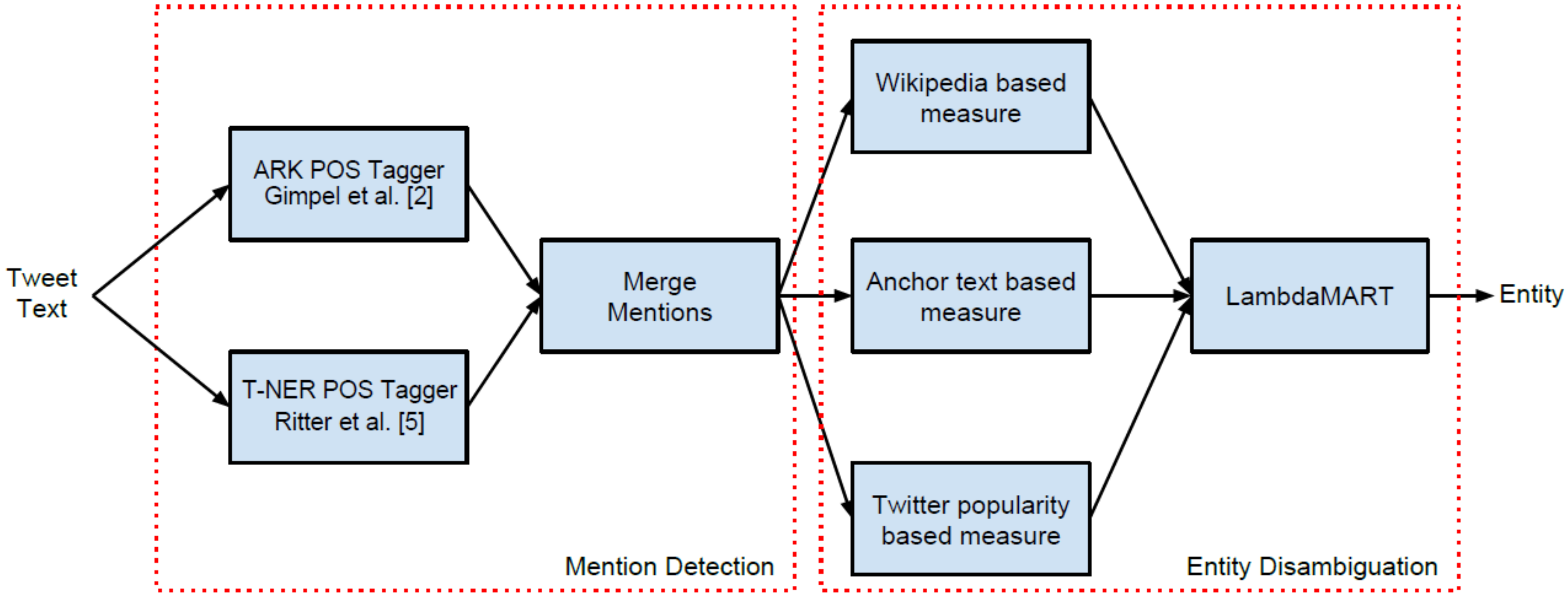
Motivation

- **Social Media** like Twitter is a source of a wide variety of information. Identifying entities in tweets can help in various tasks like tracking products, events etc.
- **Tweets** being short and noisy lack sufficient context for entity mention to be disambiguated completely.
- So we tried to **enhance the context** based on the **information shared by the other users** about the entity on social media like Twitter along with the local context of the entity.

Related Work

- Various approaches for tweet entity Linking have been proposed in the past.
- Leu *et. al* [ELFT13] use mention-entry similarity, entry-entry similarity, and mention-mention similarity and simultaneously resolve a set of mentions from tweets.
- Meij *et. al* [ASMP12] tried to link the entities in the tweets based on various n-grams, tweets and concept features.
- Guo *et. al* [TLNL13] tried to model entity linking as structured learning problem by simultaneously learning mention detection and entity linking.

Our Approach (System Architecture)



Mention Detection

- Mention Detection is the task of **detecting phrases in the text that could be linked to possible entities** in the knowledge base. We used POS patterns from ARK POS Tagger [POST11] coupled with the T-NER Tagger [NERT11] to find the mentions in the given text.
  1. **ARK POS Tagger**: Extract all sequences of proper nouns, and label longest continuous sequence as a mention.
  2. **T-NER POS Tagger**: Extract chunks with at least one proper noun, and label them as mention.
- **Merging Mentions**: Merge the entity lists from the two systems. In case of conflict, select the longest possible sequence as entity mention in the text.

Entity Disambiguation

- Entity Disambiguation is the task of **selecting the correct candidate from the possible list of candidates for the given Entity Mention**. We treated the problem of entity disambiguation as a ranking problem. We extracted the ranked entities using 3 different methods and later merged the ranked lists based on the machine learning model.

1. **Wikipedia Based Measure** (M1): Extract the entities that best matches the Wikipedia's pages title and body text and rank them according to the Wikipedia's page similarity with the mention.
2. **Google Cross-Wiki Based Measure** (M2): Extract and rank the entities based on the similarity between the anchor text [CLDE12] used across various web pages (for referring a Wikipedia Entity) and the mention.
3. **Twitter Popularity Based Measure** (M3): Extract the entities based on the similarity between the anchor text and the text used while referring the mention (in other tweets) on Twitter.

Entity Disambiguation (cont.)

- The ranked lists from three different models (Wikipedia based (M1), Google Cross-Wiki Based (M2) and Twitter Popularity Based (M3)) are **merged based on the LambdaMART model**.
- LambdaMART [ABIR10] combines MART and LambdaRank to generate an overall ranking model combining the ranks of three individual measures.
- The top ranked entity is taken as the disambiguated entity for the given entity mention.

Dataset

- #Microposts2014 NEEL Challenge Dataset is used for evaluating the system.
- 2.3K Tweets, manually annotated
- 70% Training – 30% Testing

Results

Entity Mention Detection and Entity Disambiguation

Method	Accuracy
ARK POS Tagger	77%
T-NER POS Tagger	92%
ARK + T-NER (Merged)	98%

Table 1: Performance for Mention Detection

Method	F1- measure
M1	0.335
M2	0.100
M3	0.194
M1+M2	0.335
M2+M3	0.244
M1+M3	0.405
M1+M2+M3	0.512

Table 2: Performance for Entity Disambiguation

Conclusion

- For effective entity linking, mention detection in tweets is important. We improve the accuracy of detecting mentions by combining two Twitter POS taggers.
- We resolve multiple mentions, abbreviations and spell variations of a named entity using the Wikipedia and Google Cross-Wiki Dictionary.
- We also use popularity of an entity on Twitter for improving the disambiguation. Our system performed well with a F1 score of 0.512 on the given dataset.

Thank you!

Questions?

References

[TNLN13] S. Guo, M.-W. Chang, and E. Kıcıman. To Link or Not to Link? A Study on End-to-End Tweet Entity Linking. In Proc. of the Human Language Technologies: The Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT)

[ASMP12] E. Meij, W. Weerkamp, and M. de Rijke. Adding semantics to microblog posts. In WSDM 2012. ACM, 2012

[ELFT13] X. Liu, Y. Li, H. Wu, M. Zhou, F. Wei, and Y. Lu. 2013. Entity linking for tweets. In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics

[NERT11] A. Ritter, S. Clark, Mausam, and O. Etzioni. Named Entity Recognition in Tweets: An Experimental study. In Proc. Of the 2011 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2011

[POST11] K. Gimpel, N. Schneider, B. O'Connor, D. Das, D. Mills, J. Eisenstein, M. Heilman, D. Yogatama, J. Flanigan, and N. A. Smith. Part-of-speech Tagging for Twitter: Annotation, Features, and Experiments. In Proc. of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers - Volume 2 (NAACL-HLT), pages 42–47, 2011

[CLDE12] V. I. Spitkovsky and A. X. Chang. A Cross-Lingual Dictionary for English Wikipedia Concepts. In Proc. of the 8th Intl. Conf. on Language Resources and Evaluation (LREC), 2012.

[ABIR10] Q. Wu, C. J. Burges, K. M. Svore, and J. Gao. Adapting Boosting for Information Retrieval Measures. Journal of Information Retrieval, 13(3):254–270, Jun 2010