# Context Representation for Named Entity Linking

**Xiao Ling**
University of Washington

**Sameer Singh**
University of Washington

**Daniel S. Weld**
University of Washington

{xiaoling,sameer,weld}@cs.washington.edu

## 1 Introduction

Named Entity Linking (NEL) is an Information Extraction task of linking a textual span (or an entity mention) to its corresponding entry in a given Knowledge Base (KB, e.g. Wikipedia). The links not only provide semantic annotations to the human readers but also a machine-consumable representation of the knowledge in the text. Other NLP applications could benefit from such links, e.g. for distantly supervised relation extraction (Mintz et al., 2009; Riedel et al., 2010; Hoffmann et al., 2011)) as well as bootstrapping from auxiliary knowledge resources (Hajishirzi et al., 2013).

In this work, we propose to learn a compact context representation based on dependency features to improve the quality of Named Entity Linking. Despite the massive attention on this task recently (Cucerzan, 2007; Milne and Witten, 2008; Ratinov et al., 2011; Hoffart et al., 2011; Hajishirzi et al., 2013), the majority of existing research has been using a rather shallow *bag-of-words* (BOW) representation for the context of the mentions and the entities. Unfortunately, the BOW representation has a number of drawbacks. The neighboring words are included indiscriminately, likely exposing noise to the context modeling. When modeling the context of a target mention, bag-of-words might take less tangential words to the mention and/or miss the important information for disambiguation. The same happens to modeling the entities. Further, sparsity is a critical problem for the BOW representation; the words found in the context of mentions may not have a high overlap with the words used to describe the entity, leading to a low cosine score. Less popular Wikipedia entities are especially prone to this issue because their Wikipedia pages do not have as many words.

To address the fore-mentioned challenges, we choose to create features for a mention based on the dependency graph of the sentence. Unlike the unigram bag-of-words model, the dependency-based features are defined over head words. Such features can capture lexically distant yet relevant words, and at the same time avoid words in the context that are not relevant. For instance, the BOW-style context for "Seattle" in "Seattle beat Portland yesterday" is {beat, portland, yesterday}. Here "Seattle" refers to a soccer team *Seattle Sounders*. Notice that the word "yesterday" has little to do with disambiguation of the mention and the verb "beat" that strongly indicates the type of the entity is weighted equally with "yesterday". Our proposed dependency-based features will 1) have "X (subj-of) beat" which captures the important word (the head of this sentence) as well as its relationship with the mention (i.e. "subject of"); 2) exclude a less relevant depedency such as the temporal modifier "yesterday".

By using the more informative features, lack of overlap between mention and entity context is not addressed, and, in fact, is exacerbated. For example, we might not see the exact feature "X (subj-of) beat" from the entity context of *Seattle Sounders*, leading to a lower matching score. To address the sparsity of the contexts, we propose a matrix completion model that combines a number of data sources (such as Wikipedia text, Freebase types, and high-precision links) to create a low-dimensional embedding of the entities and the context features. During linking, we can use this embedding to predict the probability that a mention context feature appears for an entity, even if the feature is absent from the entity context. This context completion allows the cosine distance to be defined on all the features of the mention context, instead of solely relying on the overlap between the mention and the entity contexts.

## 2 Model

Given the extracted mentions from a document, entity linking systems first identify a set of candidates from Wikipedia to link to. The set of mentions and entities are then linked using a combination of mention-entity string similarity, entity-entity coherence, and mention-entity context similarity, possibly with a step of NIL clustering for unlinkable mentions. Although we focus on improving mention-entity context similarity in this paper, our contribution can be combined easily with existing approaches for coherence, mention-entity string similarity, candidate generation, and NIL clustering. In particular, we define the mention-entity similarity as the cosine distance of features based on the mention context and the entity description.

In contrast with the traditional bag-of-words representation, we present a dependency-based feature representation. The features include the premodifier of the mention (e.g. the title of a person), the apposition of the mention (e.g. a textual description of the entity (Radford and Curran, 2013)) and the dependency paths from the head of the mention to a content word outside the mention (Ling and Weld, 2012; Riedel et al., 2013). The same feature generation procedure applies to all the coreferences of the target mention and then we use the union set of all the features (Cheng and Roth, 2013).

The representation of the entities also uses the same feature space. The first (summary) section of each Wikipedia article is used as the source text. According to the Wikipedia writing style, the first textual span in bold is extracted as the first mention of the entity. To resolve coreference, the Stanford Sieve-based system (Lee et al., 2013) is applied to the summary text. We also employ a heuristic for pronoun detection: the most frequent pronoun in the article are considered as coreferences of the main entity (Wu and Weld, 2007).

Sparsity is a common problem when using cosine distance based context similarity, further compounded in our approach as we filter some of the context features. The extracted entity features from Wikipedia may be viewed as a binary matrix in which the rows are the entities and the columns are the dependency-based features. Although this matrix is very sparse, we expect the matrix to contain redundancies and dependencies, i.e. many entities will have similar features (for example two people that are politicians) and many features are imply others (for example, "X, the CEO of Y" and "X be appointed as the CEO of Y"). The linear combinations that exist between entities and features suggest that a low-rank matrix completion may be used to predict the values of the missing features. To further encourage similarity between entities with similar types, we also include the entities' Freebase types as additional features (e.g. the dependency feature "X writes a book" and the Freebase type "/book/author" might have a strong correlation). We use a stochastic gradient descent with the maximum-likelihood objective to learn the latent $k$-dimensional vectors for each entity and feature. To compute the similarity between a mention and an entity, we use the latent vectors to predict the probability that the mention features appear as entity features, essentially completing the entity features for the given mention.

## 3 Preliminary Results

To evaluate the linking performance, we developed a prototype system that uses our mention-entity context similarity, a simple token overlap based string similarity, CrossWiki-based candidate generation (Spitkovsky and Chang, 2012), and ignores entity coherence[1] and NIL clustering. We evaluate our system on 5 TAC KBP Entity Linking data sets(tac, 2013). Each data set consists of a few thousand queries each of which specifies a mention in a document. The gold link could be either an entity in the KB or a NIL link. The performance is measured by both a micro-average accuracy and an entity-based B-cube+ F1 score. Unlike B-cube+, accuracy of the links is insensitive to the NIL clustering. As of the submission, the prototype's performance still trails the best system's by 5.5% in accuracy and 4.2% in F1.

## 4 Related Work and Conclusion

Other related work includes (He et al., 2013) that employs deep neural networks to learn entity representations and a generative topic model is developed by Han and Sun (2012) where each entity is generated based on the selection of an underlying topic. Despite the sophistication of the models, the context is still represented by bag-of-words. Among the few using a representation beyond bag-of-words, Hoffart et al. (2011) have shown inferior results using syntactic dependencies as features.

In conclusion, we propose a novel context representation beyond the previously dominant bag-of-words. We intend to investigate the effects of different designs of the context representation. We would also like to find out if a mixture of different representations can result in better performance.

[1]Note that our prototype independently predicts the link for each mention, which could lead to suboptimal decisions. For example, it could be hard to choose between two candidates have similar dependency features (*Seattle Seahawks* vs *Seattle Sounders*). Knowing that "Portland" refers to *Portland Timbers* could be really helpful. In the future version of the system, we plan to jointly link the mentions within the same document.

## References


Xiao Cheng and Dan Roth. 2013. Relational inference for wikification. In *EMNLP*.

S. Cucerzan. 2007. Large-scale named entity disambiguation based on wikipedia data. In *Proceedings of EMNLP-CoNLL*, volume 2007, pages 708–716.

Hannaneh Hajishirzi, Leila Zilles, Daniel S. Weld, and Luke Zettlemoyer. 2013. Joint Coreference Resolution and Named-Entity Linking with Multi-pass Sieves. In *EMNLP*.

Xianpei Han and Le Sun. 2012. An entity-topic model for entity linking. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 105–115. Association for Computational Linguistics.

Zhengyan He, Shujie Liu, Mu Li, Ming Zhou, Longkai Zhang, and Houfeng Wang. 2013. Learning entity representation for entity disambiguation. *Proc. ACL2013*.

Johannes Hoffart, Mohamed A. Yosef, Ilaria Bordino, Hagen Fürstenau, Manfred Pinkal, Marc Spaniol, Bilyana Taneva, Stefan Thater, and Gerhard Weikum. 2011. Robust disambiguation of named entities in text. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 782–792. Association for Computational Linguistics.

Raphael Hoffmann, Congle Zhang, Xiao Ling, Luke Zettlemoyer, and Daniel S Weld. 2011. Knowledge-based weak supervision for information extraction of overlapping relations. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, volume 1, pages 541–550.

Heeyoung Lee, Angel Chang, Yves Peirsman, Nathanael Chambers, Mihai Surdeanu, and Dan Jurafsky. 2013. Deterministic coreference resolution based on entity-centric, precision-ranked rules. *Computational Linguistics*, pages 1–54.

Xiao Ling and Daniel S Weld. 2012. Fine-grained entity recognition. In *AAAI*.

David Milne and Ian H. Witten. 2008. Learning to link with wikipedia. In *Proceedings of the 17th ACM conference on Information and knowledge management*, pages 509–518. ACM.

Mike Mintz, Steven Bills, Rion Snow, and Daniel Jurafsky. 2009. Distant supervision for relation extraction without labeled data. In *Proceedings of the 47th Annual Meeting of the Association for Computational Linguistics (ACL-2009)*, pages 1003–1011.

Will Radford and James R Curran. 2013. Joint apposition extraction with syntactic and semantic constraints.

Lev-Arie Ratinov, Dan Roth, Doug Downey, and Mike Anderson. 2011. Local and global algorithms for disambiguation to wikipedia. In *ACL*, volume 11, pages 1375–1384.

Sebastian Riedel, Limin Yao, and Andrew McCallum. 2010. Modeling relations and their mentions without labeled text. In *ECML/PKDD (3)*, pages 148–163.

Sebastian Riedel, Limin Yao, Andrew McCallum, and Benjamin M Marlin. 2013. Relation extraction with matrix factorization and universal schemas.

Valentin I Spitkovsky and Angel X Chang. 2012. A cross-lingual dictionary for english wikipedia concepts. In *LREC*, pages 3168–3175.

2013. Tac kbp 2013 entity linking track.

Fei Wu and Daniel S. Weld. 2007. Autonomously semantifying wikipedia. In *Proceedings of the International Conference on Information and Knowledge Management (CIKM-2007)*, pages 41–50.