

Improving Entity Linking through Semantic Reinforced Entity Embeddings

Feng Hou^{#†}, Ruili Wang^{#†*}, Jun He[‡], Yi Zhou^{‡†}

[#]School of Computer and Information Engineering, Zhejiang Gongshang University, Hangzhou, China

[†]School of Natural and Computational Sciences, Massey University, New Zealand

[‡]School of Information Communication, National University of Defense Technology, China

[‡]Shanghai Research Center for Brain Science and Brain-Inspired Intelligence, Zhangjiang Lab, China

{f.hou, ruili.wang}@massey.ac.nz, prof.ruili.wang@gmail.com

hejun_nudt@nudt.edu.cn, yzhou@bsbii.cn

Abstract

Entity embeddings, which represent different aspects of each entity with a single vector like word embeddings, are a key component of neural entity linking models. Existing entity embeddings are learned from canonical Wikipedia articles and local contexts surrounding target entities. Such entity embeddings are effective, but too distinctive for linking models to learn contextual commonality. We propose a simple yet effective method, FGS2EE, to inject fine-grained semantic information into entity embeddings to reduce the distinctiveness and facilitate the learning of contextual commonality. FGS2EE first uses the embeddings of semantic type words to generate semantic embeddings, and then combines them with existing entity embeddings through linear aggregation. Extensive experiments show the effectiveness of such embeddings. Based on our entity embeddings, we achieved new state-of-the-art performance on entity linking.

1 Introduction

Entity Linking (EL) or Named Entity Disambiguation (NED) is to automatically resolve the ambiguity of entity mentions in natural language by linking them to concrete entities in a Knowledge Base (KB). For example, in Figure 1, mentions “Congress” and “Mr. Mueller” are linked to the corresponding Wikipedia entries, respectively.

Neural entity linking models use local and global scores to rank and select a set of entities for mentions in a document. Entity embeddings are critical for the local and global score functions. But the current entity embeddings (Ganea and Hofmann, 2017) encoded too many details of entities, thus are too distinctive for linking models to learn contextual commonality.

*Corresponding author



Figure 1: Entity linking with embedded fine-grained semantic types

We hypothesize that fine-grained semantic types of entities can let the linking models learn contextual commonality about semantic relatedness. For example, *rugby* related documents would have entities of *rugby player* and *rugby team*. If a linking model learns the contextual commonality of *rugby* related entities, it can correctly select entities of similar types using the similar contextual information.

In this paper, we propose a method FGS2EE to inject fine-grained semantic information into entity embeddings to reduce the distinctiveness and facilitate the learning of contextual commonality. FGS2EE uses the word embeddings of semantic words that represent the hallmarks of entities (e.g., *writer*, *carmaker*) to generate semantic embeddings. We find that the training converges faster when using semantic reinforced entity embeddings.

Our proposed FGS2EE consists of four steps:

(i) creating a dictionary of fine-grained semantic words; (ii) extracting semantic type words from each entity’s Wikipedia article; (iii) generating semantic embedding for each entity; (iv) combining semantic embeddings with existing embeddings through linear aggregation.

2 Background and Related Work

2.1 Local and Global Score for Entity Linking

The local score $\Psi(e_i, c_i)$ (Ganea and Hofmann, 2017) measures the relevance of entity candidates of each mention independently.

$$\Psi(e_i, c_i) = \mathbf{e}_i^\top \mathbf{B} f(c_i)$$

where $\mathbf{e}_i \in \mathbb{R}^d$ is the embedding of candidate entity e_i ; $\mathbf{B} \in \mathbb{R}^{d \times d}$ is a diagonal matrix; $f(c_i) \in \mathbb{R}^d$ is a feature representation of local context c_i surrounding mention m_i .

In addition to the local score, the global score adds a pairwise score $\Phi(e_i, e_j, D)$ to take the coherence of entities in document D into account.

$$\Phi(e_i, e_j, D) = \frac{1}{n-1} \mathbf{e}_i^\top \mathbf{C} \mathbf{e}_j$$

where \mathbf{e}_i and $\mathbf{e}_j \in \mathbb{R}^d$ are the embeddings of entities e_i, e_j , which are candidates for mention m_i and m_j respectively; $\mathbf{C} \in \mathbb{R}^{d \times d}$ is a diagonal matrix. The pairwise score of (Le and Titov, 2018) considers K relations between entities.

$$\Phi(e_i, e_j, D) = \sum_{k=1}^K \alpha_{ijk} \mathbf{e}_i^\top \mathbf{R}_k \mathbf{e}_j$$

where α_{ijk} is the weight for relation k , and \mathbf{R}_k is a diagonal matrix for measuring relations k between two entities.

2.2 Related Work

Our research focuses on improving the vector representations of entities through fine-grained semantic types. Related topics are as follows.

Entity Embeddings Similar to word embeddings, entity embeddings are the vector representations of entities. The methods of Yamada et al. (2016), Fang et al. (2016), Zwicklbauer et al. (2016), use data about entity-entity co-occurrences to learn entity embeddings and often suffer from sparsity of co-occurrence statistics. Ganea and Hofmann (2017) learned entity embeddings using

words from canonical Wikipedia articles and local context surrounding anchor links. They used Word2Vec vectors (Mikolov et al., 2013) of positive words and random negative words as input to the learning objective. Thus their entity embeddings are aligned with the Word2Vec word embeddings.

Fine-grained Entity Typing Fine-grained entity typing is a task of classifying entities into fine-grained types (Ling and Weld, 2012) or ultra fine-grained semantic labels (Choi et al., 2018). Bhowmik and de Melo (2018) used a memory-based network to generate a short description of an entity, e.g. “Roger Federer” is described as ‘Swiss tennis player’. In this paper, we heuristically extract fine-grained semantic types from the first sentence of Wikipedia articles.

Embeddings Aggregation Our research is closely related to the work on aggregation and evaluation of the information content of embeddings from different sources (e.g., polysemous words have multiple sense embeddings), and fusion of multiple data sources (Wang et al., 2018). Arora et al. (2018) hypothesizes that the global word embedding is a linear combination of its sense embeddings. They showed that senses can be recovered through sparse coding. Mu et al. (2017) showed that senses and word embeddings are linearly related and sense sub-spaces tend to intersect over a line. Yaghoobzadeh et al. (2019) probe the aggregated word embeddings of polysemous words for semantic classes. They created a WIKI-PSE corpus, where word and semantic class pairs are annotated using Wikipedia anchor links, e.g., “apple” has two semantic classes: *food* and *organization*. A separate embedding for each semantic class was learned based on the WIKI-PSE corpus. They found that the linearly aggregated embeddings of polysemous words represent well their semantic classes.

The most similar work is that of Gupta et al. (2017), but there are many differences: (i) they use the FIGER (Ling and Weld, 2012) type taxonomy that contains manually curated 112 types organized into 2 levels; we employ over 3000 vocabulary words as type, and we treat them as a flat list; (ii) they mapped the Freebase types to FIGER types, but this method is less credible, as noted by Daniel Gillick et al. (2014); we extract type words directly from Wikipedia articles, which is more reliable. (iii) their entity vectors and type vectors are learned jointly on a limited corpus. Ours are linear aggregations of existing entity vectors, and

word vectors learned from a large corpus, such fine-grained semantic word embeddings are helpful for capturing informative context.

2.3 Motivation

Coarse-grained semantic types (e.g. *person*) have been used for candidate selection (Ganea and Hofmann, 2017). We observe that fine-grained semantic words appear frequently as apposition (e.g., *Defense contractor Raytheon*), coreference (e.g., *the company*) or anonymous mentions (e.g., *American defense firms*). These fine-grained types of entities can help capture local contexts and relations of entities.

Some of these semantic words have been used for learning entity embeddings, but they are diluted by other unimportant or noisy words. We reinforce entity embeddings with such fine-grained semantic types.

3 Extracting Fine-grained Semantic Types

We first create a dictionary of fine-grained semantic types, then we extract fine-grained types for each entity.

3.1 Semantic Type Dictionary

We select those words that can encode the hallmarks of individual entities. Desiderata are as follows:

- profession/subject, e.g., *footballer, soprano, biology, rugby*.
- title, e.g., *president, ceo, head, director*.
- industry/genre, e.g., *carmaker, manufacturer, defense contractor, hip hop*.
- geospatial, e.g., *canada, asian, australian*.
- ideology/religion, e.g., *communism, buddhism*.
- miscellaneous, e.g., *book, film, tv, ship, language*.

We extract noun frequency from the first sentence of each entity in the Wikipedia dump. Then some seed words are manually selected from frequent nouns. We use word similarity to extend these seed words and finally got a dictionary with 3,227 fine-grained semantic words.

Specifically, we use `spaCy` to compute the similarity between words. For each seed word, we find the top 100 similar words that also appear in Wikipedia articles. We then manually select semantic words from these extended words.

3.2 Extracting Semantic Types

For each entity, we extract at most 11 dictionary words (phrases) from its Wikipedia article. For example, “Robert Mueller” in Figure 1 will be typed as [*american, lawyer, government, official, director*].

3.3 Remapping Semantic Words

For some semantic words (e.g., *conchologist*) or semantic phrases (e.g., *rugby league*), there are no word embeddings available for generating the semantic entity embeddings. We remap these semantic words to semantically similar words that are more common. For example, the *conchologist* is remapped to *zoologist*, and *rugby league* is remapped to *rugby_league*.

4 FGS2EE: Injecting Fine-Grained Semantic Information into Entity Embeddings

FGS2EE first uses semantic words of each entity to generate semantic entity embeddings, then combine them with existing entity embeddings to generate semantic reinforced entity embeddings.

4.1 Semantic Entity Embeddings

Based on the semantic words of each entity, we can produce a semantic entity embedding. We treat each semantic word as a sense of an entity. The embedding of each sense is represented by the Word2Vec embedding of the semantic word. Suppose we only consider T semantic words for each entity, and the set of entity words of entity e is denoted as S_e . Then the semantic entity embedding e^s of entity e is generated as follows:

$$e^s = \frac{1}{T} \sum_{i=1}^T e_{w_i} \quad (1)$$

where $w_i \in S_e$ is the i th semantic word, e_{w_i} is the Word2Vec embedding¹ of semantic word w_i . If $|S_e| < T$, then $T = |S_e|$.

4.2 Semantic Reinforced Entity Embeddings

We create a semantic reinforced embedding for each entity by linearly aggregating the semantic entity embeddings and Word2Vec style entity embeddings (Ganea and Hofmann, 2017) (hereafter referred to as “Wikitext entity embeddings”).

¹<https://code.google.com/archive/p/word2vec/>

Entity Embeddings	Linking Methods	AIDA-B	MSNBC	AQUAINT	ACE2004	CWEB	WIKI	Avg
<i>Wikipedia</i>								
-	(Milne and Witten, 2008)	-	78	85	81	64.1	81.7	77.96
-	(Ratinov et al., 2011)	-	75	83	82	56.2	67.2	72.68
-	(Hoffart et al., 2011)	-	79	56	80	58.6	63	67.32
-	(Cheng and Roth, 2013)	-	90	90	86	67.5	73.4	81.38
-	(Chisholm and Hachey, 2015)	84.9	-	-	-	-	-	-
<i>Wiki + Unlabelled documents</i>								
-	(Lazic et al., 2015)	86.4	-	-	-	-	-	-
(Ganea and Hofmann, 2017)	(Le and Titov, 2019)	89.66±0.16	92.2±0.2	90.7±0.2	88.1±0.0	78.2±0.2	81.7±0.1	86.18
$T = 6, \alpha = 0.1$	(Le and Titov, 2019)	89.58±0.2	92.3±0.1	90.93±0.2	87.88±0.17	78.47±0.11	81.71±0.21	86.26
$T = 11, \alpha = 0.2$	(Le and Titov, 2019)	89.23±0.31	92.15±0.24	91.22±0.18	88.02±0.15	78.29±0.17	81.92±0.36	86.32
<i>Wiki + Extra supervision</i>								
-	(Chisholm and Hachey, 2015)	88.7	-	-	-	-	-	-
<i>Fully-supervised(Wiki+ AIDA train)</i>								
-	(Guo and Barbosa, 2016)	89.0	92	87	88	77	84.5	85.7
-	(Globerson et al., 2016)	91.0	-	-	-	-	-	-
(Yamada et al., 2016)	(Yamada et al., 2016)	91.5	-	-	-	-	-	-
(Ganea and Hofmann, 2017)	(Ganea and Hofmann, 2017)	92.22±0.14	93.7±0.1	88.5±0.4	88.5±0.3	77.9±0.1	77.5±0.1	85.22
(Ganea and Hofmann, 2017)	(Le and Titov, 2018)	93.07±0.27	93.9±0.2	88.3±0.6	89.9±0.8	77.5±0.1	78.0±0.1	85.5
(Ganea and Hofmann, 2017)	DCA (Yang et al., 2019)	93.73±0.2	93.80±0.0	88.25±0.4	90.14±0.0	75.59.5±0.3	78.84±0.2	85.32
$T = 6, \alpha = 0.1$	(Le and Titov, 2018)	92.29±0.21	94.1±0.24	88.0±0.38	90.14±0.32	77.23±0.18	77.16±0.43	85.33
$T = 11, \alpha = 0.2$	(Le and Titov, 2018)	92.63±0.14	94.26±0.17	88.47±0.23	90.7±0.28	77.41±0.21	77.66±0.23	85.7

Table 1: F1 scores on six test sets. The last column is the average of F1 scores on the five out-domain test sets.

Our semantic entity embeddings tend to be homogeneous. If we average them with the Wikitext embeddings, the aggregated embeddings would be homogeneous too. Thus the entity linking model would not be able to distinguish between those similar candidates. Our semantic reinforced entity embedding is a weighted sum of semantic entity embedding and Wikitext entity embedding, similar to (Yaghoobzadeh et al., 2019). We use a parameter α to control the weight of semantic entity embeddings. Thus the aggregated (semantic reinforced) entity embeddings achieve a trade-off between homogeneity and heterogeneity.

$$e^a = (1 - \alpha) e^w + \alpha e^s \quad (2)$$

where e^w is the Wikitext entity embedding of entity e .

5 Experiments

5.1 Datasets and Evaluation Metric

We use the Wikipedia dump 20190401 to extract fine-grained semantic type dictionary and semantic types for entities. We use the Wikitext entity embeddings shared by Le and Titov (2018, 2019). For entity linking corpora, we use the datasets shared by Ganea and Hofmann (2017) and Le and Titov (2018, 2019).

We use the standard micro F1-score as evaluation metric. Our data and source code are publicly available at [github](https://github.com/fhou80/EntEmb)².

5.2 Experimental Settings

The parameters T in Equation (1) and α in Equation (2) are critical for the effectiveness of our se-

²<https://github.com/fhou80/EntEmb/>

semantic reinforced entity embeddings. We got two sets of entity embeddings with two combinations of parameters: $T = 6, \alpha = 0.1$ and $T = 11, \alpha = 0.2$

To test the effectiveness of our semantic reinforced entity embeddings, we use the entity linking models **mulrel** (Le and Titov, 2018) (ment-norm $K = 3$) and **wnel** (Le and Titov, 2019) that are publicly available. We do not optimize their entity linking code. We just replace the entity embeddings with our semantic reinforced entity embeddings.

Similar to Ganea and Hofmann (2017) and Le and Titov (2018, 2019), we run our system 5 times for each combination of entity embeddings and linking model, and report the mean and 95% confidence interval of the micro F1 score.

5.3 Results

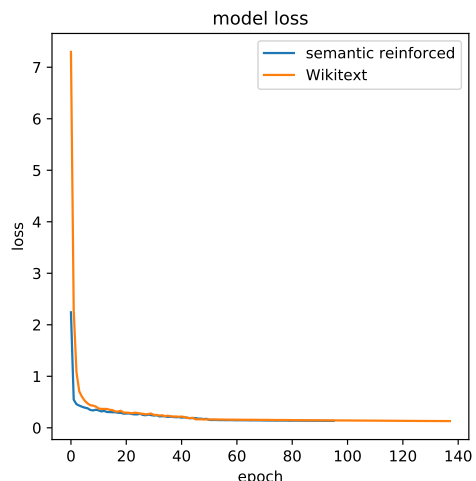


Figure 2: Learning curves of **mulrel** (Le and Titov, 2018) using two different sets of entity embeddings.

The results on six testing datasets are shown in Table 1. For the **mulrel** model, our entity embed-

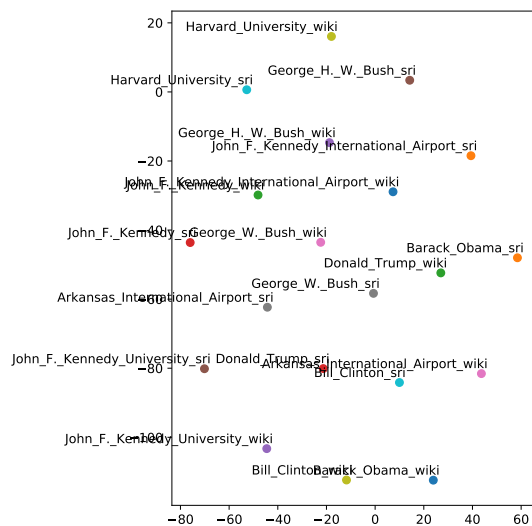


Figure 3: T-SNE visualization of two sets of entity embeddings. Suffix “_wiki” denotes the Wikitext entity embeddings, while suffix “_sri” denotes the semantic reinforced entity embeddings ($T = 11, \alpha = 0.2$).

dings ($T = 11, \alpha = 0.2$) improved performance drastically on MSNBC, ACE2004 and average of out-domain test sets. Be aware that CWEB and WIKI are believed to be less reliable (Ganea and Hofmann, 2017). For the **wnel** model, our both sets of entity embeddings are more effective for four of the five out-domain test sets and the average.

Our entity embeddings are better than that of Ganea and Hofmann (2017) when tested on the **mulrel** (Le and Titov, 2018) (ment-norm $K = 3$) and **wnel** (Le and Titov, 2019) entity linking models. Ganea and Hofmann (2017) showed that their entity embeddings are better than that of Yamada et al. (2016) using the entity relatedness metrics.

One notable thing for our semantic reinforced entity embeddings is that the training using our entity embeddings converges much faster than that using Wikitext entity embeddings, as shown in Figure 2. One reasonable explanation is that the fine-grained semantic information lets the linking models capture the commonality of semantic relatedness between entities and contexts, hence facilitate the training.

The properties of two different sets of entity embeddings can be visually manifested in Figure 3. Our semantic reinforced entity embeddings draw entities of similar types closer, and entities of different types further. For example, our semantic reinforced embeddings of “John F. Kennedy University” and “Harvard University” are closer than the Wikitext embeddings, while our embeddings of

“John F. Kennedy International Airport” and “John F. Kennedy” are further. We believe this property contributes to the faster convergence.

6 Conclusion

In this paper, we presented a simple yet effective method, FGS2EE, to inject fine-grained semantic information into entity embeddings to reduce the distinctiveness and facilitate the learning of contextual commonality. FGS2EE first uses the word embeddings of semantic type words to generate semantic embeddings, and then combines them with existing entity embeddings through linear aggregation. Our entity embeddings draw entities of similar types closer, while entities of different types are drawn further. Thus can facilitate the learning of semantic commonalities about entity-context and entity-entity relations. We have achieved new state-of-the-art performance using our entity embeddings.

For the future work, we are planning to extract fine-grained semantic types from unlabelled documents and use the relatedness between the fine-grained types and contexts as distant supervision for entity linking.

References

- Sanjeev Arora, Yuanzhi Li, Yingyu Liang, Tengyu Ma, and Andrej Risteski. 2018. [Linear algebraic structure of word senses, with applications to polysemy](#). *Transactions of the Association for Computational Linguistics*, 6:483–495.
- Rajarshi Bhowmik and Gerard de Melo. 2018. [Generating fine-grained open vocabulary entity type descriptions](#). In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*.
- Xiao Cheng and Dan Roth. 2013. [Relational inference for wikification](#). In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1787–1796.
- Andrew Chisholm and Ben Hachey. 2015. [Entity disambiguation with web links](#). *Transactions of the Association for Computational Linguistics*, 3:145–156.
- Eunsol Choi, Omer Levy, Yejin Choi, and Luke Zettlemoyer. 2018. [Ultra-fine entity typing](#). In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, pages 87–96. Association for Computational Linguistics.
- Daniel Gillick, Nevena Lazic, Kuzman Ganchev, Jesse Kirchner, and David Huynh. 2014. [Contextdependent fine-grained entity type tagging](#). *arXiv preprint arXiv:1412.1820*.

- Wei Fang, Jianwen Zhang, Dilin Wang, Zheng Chen, and Ming Li. 2016. [Entity disambiguation by knowledge and text jointly embedding](#). In *Proceedings of The 20th SIGNLL Conference on Computational Natural Language Learning*, pages 260–269, Berlin, Germany. Association for Computational Linguistics.
- Octavian-Eugen Ganea and Thomas Hofmann. 2017. [Deep joint entity disambiguation with local neural attention](#). In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2619–2629, Copenhagen, Denmark. Association for Computational Linguistics.
- Amir Globerson, Nevena Lazic, Soumen Chakrabarti, Amarnag Subramanya, Michael Ringgaard, and Fernando Pereira. 2016. [Collective entity resolution with multi-focal attention](#). In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 621–631.
- Zhaochen Guo and Denilson Barbosa. 2016. Robust named entity disambiguation with random walks. *Semantic Web (Preprint)*, 9(4):459–479.
- Nitish Gupta, Sameer Singh, and Dan Roth. 2017. [Entity linking via joint encoding of types, descriptions, and context](#). In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2681–2690, Copenhagen, Denmark. Association for Computational Linguistics.
- Johannes Hoffart, Mohamed Amir 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 2011 Conference on Empirical Methods in Natural Language Processing*, pages 782–792. Association for Computational Linguistics.
- Nevena Lazic, Amarnag Subramanya, Michael Ringgaard, and Fernando Pereira. 2015. [Plato: A selective context model for entity resolution](#). *Transactions of the Association for Computational Linguistics*, 3:503–515.
- Phong Le and Ivan Titov. 2018. [Improving entity linking by modeling latent relations between mentions](#). In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1595–1604, Melbourne, Australia. Association for Computational Linguistics.
- Phong Le and Ivan Titov. 2019. [Boosting entity linking performance by leveraging unlabeled documents](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1935–1945, Florence, Italy. Association for Computational Linguistics.
- Xiao Ling and Daniel S. Weld. 2012. Fine-grained entity recognition. In *Proceedings of Association for the Advancement of Artificial Intelligence*.
- Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems*, pages 3111–3119.
- 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.
- Jiaqi Mu, Suma Bhat, and Pramod Viswanath. 2017. Geometry of polysemy. In *Proceedings of the 5th International Conference on Learning Representations*.
- Lev Ratinov, Dan Roth, Doug Downey, and Mike Anderson. 2011. [Local and global algorithms for disambiguation to wikipedia](#). In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, pages 1375–1384. Association for Computational Linguistics.
- Ruili Wang, Wanting Ji, Mingzhe Liu, Xun Wang, Jian Weng, Song Deng, Suying Gao, and Chang-an Yuan. 2018. [Review on mining data from multiple data sources](#). *Pattern Recognition Letters*, 109:120–128.
- Yadollah Yaghoobzadeh, Katharina Kann, T. J. Hazen, Eneko Agirre, and Hinrich Schütze. 2019. [Probing for semantic classes: Diagnosing the meaning content of word embeddings](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5740–5753, Florence, Italy. Association for Computational Linguistics.
- Ikuya Yamada, Hiroyuki Shindo, Hideaki Takeda, and Yoshiyasu Takefuji. 2016. [Joint learning of the embedding of words and entities for named entity disambiguation](#). In *Proceedings of The 20th SIGNLL Conference on Computational Natural Language Learning*, pages 250–259, Berlin, Germany. Association for Computational Linguistics.
- Xiyuan Yang, Xiaotao Gu, Sheng Lin, Siliang Tang, Yueting Zhuang, Fei Wu, Zhigang Chen, Guoping Hu, and Xiang Ren. 2019. [Learning dynamic context augmentation for global entity linking](#). In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 271–281, Hong Kong, China. Association for Computational Linguistics.
- Stefan Zwicklbauer, Christin Seifert, and Michael Granitzer. 2016. Robust and collective entity disambiguation through semantic embeddings. In *Proceedings of the 39th International ACM SIGIR conference*, pages 425–434. ACM.