Knowledge-Enhanced Named Entity Disambiguation for Short Text

Zhifan Feng, Qi Wang, Wenbin Jiang, Yajuan Lyu, Yong Zhu
Baidu Inc., Beijing, China
{fengzhifan,wangqi31,jiangwenbin,lvyajuan,zhuyong}@baidu.com

Abstract

Named entity disambiguation is an important task that plays the role of bridge between text and knowledge. However, the performance of existing methods drops dramatically for short text, which is widely used in actual application scenarios, such as information retrieval and question answering. In this work, we propose a novel knowledge-enhanced method for named entity disambiguation. Considering the problem of information ambiguity and incompleteness for short text, two kinds of knowledge, factual knowledge graph and conceptual knowledge graph, are introduced to provide additional knowledge for the semantic matching between candidate entity and mention context. Our proposed method achieves significant improvement over previous methods on a large manually annotated short-text dataset, and also achieves the state-of-the-art on three standard datasets. The short-text dataset and the proposed model will be publicly available for research use.

1 Introduction

Name entity disambiguation (NED) aims to associate each entity mention in the text with its corresponding entity in the knowledge graph (KG). It plays an important role in many text-related artificial intelligent tasks such as recommendation and conversation, since it works as a bridge between text and knowledge. In decades, researchers devoted their efforts to NED in many ways, including the rule-based methods (Shen et al., 2014), the conventional statistic methods (Shen et al., 2014) and the deep learning methods (Octavian-Eugen Ganea, 2017). On formal text, state-of-the-art methods achieve high performance thanks to the well-written utterance and rich context. However, experiments show that the performance of these methods degrades dramatically on informal text, for example, the short text widely used in many real application scenarios such as information retrieval and human-machine interaction. It is difficult for existing methods to make decisions on the non-standard utterance without adequate context.

The discrimination procedure of NED depends on sufficient context in the input text, which is usually noisy and scarce in the short text used in information retrieval and human-machine interaction. For example, an analysis based on search engine logs demonstrates that a search query contains 2.35 words on average (Yi Fang, 2011). Such short text could not provide adequate context which is necessary for NED models. In recent years, many efforts improve NED by exploiting more powerful models and richer context information (Shen et al., 2014). These methods mainly focus on the better utilization of existing context. Therefore, they can not improve NED effectively on short text since the problem of information shortage still exists. Intuitively, it is hard for NED to achieve essential improvement on short text if it can not exploit external information to enhance the recognition procedure.

In information scarce situations, human beings can still perform recognition by association with related external information, such as commonsense or domain-specific knowledge. It inspires us that, NED on short text could be improved if appropriate external knowledge can be retrieved and considered. We propose a novel knowledge-enhanced NED model, where the prediction procedure of NED is enhanced by two kinds of knowledge formalized as two different KGs. The one kind is conceptual knowledge formalized as a conceptual KG, it is used to augment the representation of the entity mention by giving the mention a concept embedding. The other kind is factual knowledge formalized as a factual KG, it is used to augment the representation of each candidate entity by giving the entity an entity embedding. The augmented

Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing, pages 735–744
December 4 - 7, 2020.

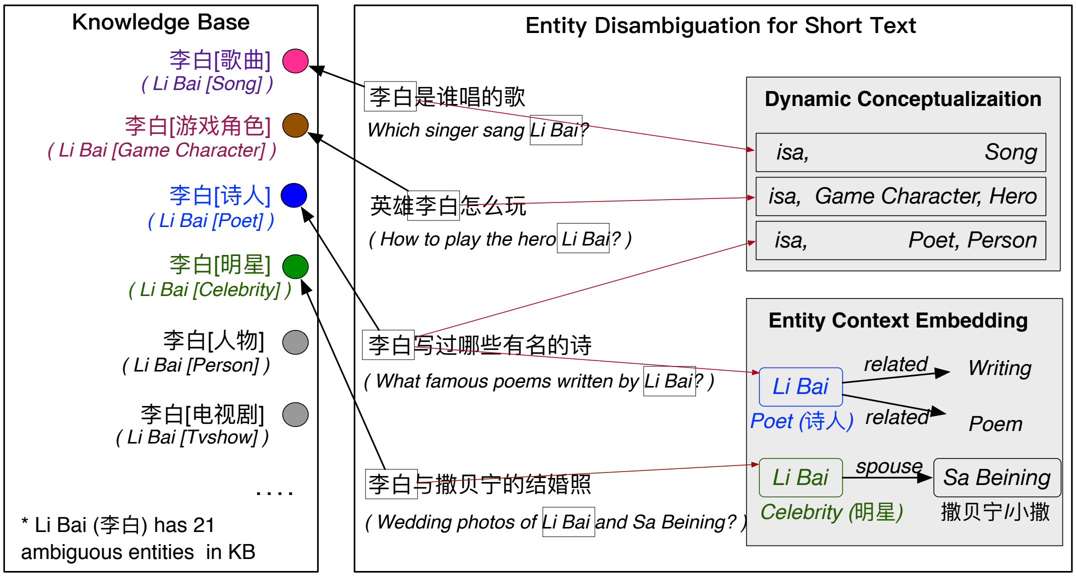

Figure 1: Short text entity disambiguation and our method. We solve the problem of entity disambiguation of sparse short texts through dynamic conceptualization and entity context embedding.

representations of the mention context and the candidate entities are used in a matching network for better NED prediction.

We validate the knowledge-enhanced NED model on three public NED datasets for short text (NEEL, KORE50 and FUDAN) as well as our new dataset (DUEL), which is constructed for information acquiring scenarios and will be publicly available for research use. Experiments show that the knowledge-enhance NED model performs significantly better than previous methods. It shows the effectiveness of external knowledge in improving the prediction of NED in information scarce situations. The contribution of our work includes two aspects. First, we introduce conceptual and factual knowledge to improve NED for short text for the first time, and achieve significant improvement. Second, we release a large-scale good-quality NED dataset for short text for information acquisition scenarios, which is complementary existing datasets.

The rest of this paper is organized as follows. We first introduce the NED task and the baseline method (section 2), and then describe the architecture and details of our knowledge-enhanced model (section 3). After giving the detailed experimental analysis (section 4), we give the related work (section 5) and conclude the work.

2 Task Definition and Baseline Model

NED is a fundamental task in the area of natural language processing and knowledge base. It aims to associate each entity mention in the given text with its corresponding entity in the given KG. Formally, given a KG $\mathcal{G}$ and a piece of text $\mathcal{T}$, it assigns each mention $m \in \mathcal{T}$ with an entity $e \in \mathcal{G}$ indicating that m refers to e, or with the symbol ϕ indicating that there is no corresponding entity.

The disambiguation procedure can be formalized as matching between the context of the mention and each candidate entities.

$$\mathbf{f}(m) = \begin{cases} \arg\max_{e \in \mathbf{e}(m)} (\mathbf{s}(e, \mathbf{c}(m))), & \mathbf{e}(m) \neq \emptyset \\ \phi, & \textbf{otherwise} \end{cases} \quad (1)$$

Here, the function $\mathbf{e}$ returns the entity candidate set for a given mention, and the function $\mathbf{c}$ returns the context of the given mention. The function $\mathbf{s}$ is used to evaluate the matching degree between context and candidate, and is usually implemented as matching networks. If the entity candidate set is empty or the highest matching score is bellow a given threshold, the function $\mathbf{f}$ returns ϕ for the given mention. The conditions $\mathcal{G}$ and $\mathcal{T}$ in the functions $\mathbf{e}$ and $\mathbf{c}$ are omitted in the equation for simplicity.

We adopt the deep structured semantic model (DSSM) (Huang et al., 2013) as the baseline model for NED(Nie and Pan, 2018). Based on a self-attention matching network, DSSM maps the candidate entities and the context to the same semantic space, and finds the candidate entity that best semantically matches the context. The representation learning for both entities and contexts is enhanced

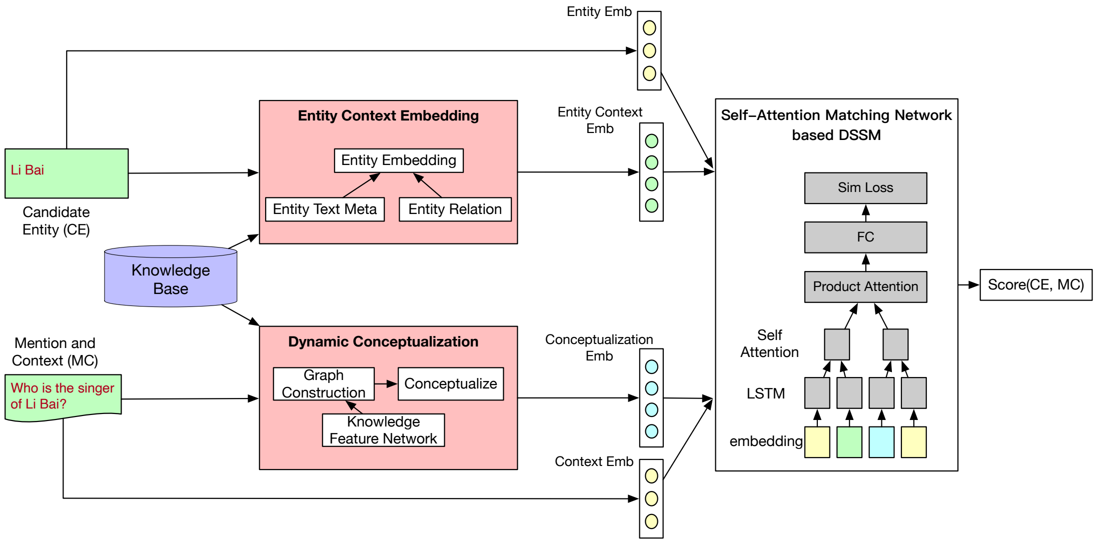

Figure 2: The architecture of our proposed entity disambiguation model K-NED.

by word2vec (Tomas Mikolov, 2013). In this work, we focus on the problem of NED itself, that is, predicting the right entity candidate for each given entity mention. The entity mentions needed for NED are simply derived from the results of named entity recognition (NER).

3 Knowledge-Enhanced NED Model

For short text used in information acquisition scenarios such as information retrieval and question answering, the lack of both lexical and syntactical information obstacles the precise disambiguation of entity mentions. For human beings, however, the problem of information scarcity does not hinder the disambiguation procedure. This is because there are many implicit assumptions and apriori knowledges in these information acquisition scenarios, which can be effectively considered by association and imagination during disambiguation procedure. Inspired by this, we propose a knowledge-enhanced NED model (K-NED) for short text, where two kinds of knowledge are introduce to provide additional information for better disambiguation performance. Figure 2 gives the overall architecture of the model.

The overall procedure of the K-NED model is a pipeline including feature extraction and semantic matching, where the former is composed of two sub-procedures, taking charge of feature extraction for mention context and candidate entity, respectively. Rather than considering only the utterance of the input text, the feature extraction procedure also considers external knowledge for better representation. In details, the feature extraction sub-procedure for mention context is enhanced by conceptual knowledge formalized as a conceptual KG, which augments the representation of the mention context by giving each word a concept embedding; while the sub-procedure for candidate entity is enhanced by factual knowledge formalized as a factual KG, which augments the representation of each candidate entity by giving the entity an entity embedding. The augmented representation is used in the following semantic matching procedure for better prediction.

The major difference of the K-NED model is the introduction of external knowledge in the representation learning procedure. For the representation learning procedure, it simply uses the pre-trained word2vec language model to take charge of the conventional utterance representation learning. For the semantic matching procedure, it directly adopt the self-attention matching network based on DSSM. Given a mention m and a candidate entity e, the word2vec-based module gives two representation vectors, $\mathbf{r}_m^{lm}$ and $\mathbf{r}_e^{lm}$, while the KG-based modules give another two representation vectors, $\mathbf{r}_m^{kg}$ and $\mathbf{r}_e^{kg}$. The concatenation of the four representation vectors is fed into the matching network to obtain the matching degree. Based on the utterance of m or e, the word2vec-based representation vector $\mathbf{r}_m^{lm}$ or $\mathbf{r}_e^{lm}$ is obtained by averaging the hidden representations for the words or characters in the utterance.

We omit the detailed descriptions of the word2vec-based feature extraction and the DSSM-based semantic matching owing to space limitations. In the following subsections, we describe in details the computation procedures for the KG-based feature extraction.

3.1 KG-enhanced Representation of Mention

The concepts in a conceptual KG can be treated as upper classes of the entities in the factual KG. A concept is a name or label representing a concrete or material existence such as a person, a place or a thing. For example, the entity apple, maybe corresponds to the concept of fruits, companies and songs. For a mention, we label the mention word with a concept and use the concept representation as additional feature representation of the mention. Intuitively, the concept labeling procedure works as a semantic bridge between the mentions and the entities.

Different from traditional methods where mentions are classified into coarse-grained entity types, the concept labeling procedure in our work classifies the mentions into fine-grained concepts, which can better utilize the context of the mentions and provide more information for disambiguation. We adopt a graph-based labeling algorithm for concept labeling, as shown in Figure 3. Given a short sentence, it first builds a knowledge feature network (KFN) based on the short sentence and reference conceptual/factual KGs, and then searches for the appropriate concept for the mention by a random walking algorithm. The KFN is built according to the correspondence between the symbols in the short sentence and the reference KGs. The symbols include words, entity mentions and candidate concepts, where the words and mentions are obtained by lexical analysis and entity recognition, and the concepts are obtained by matching on the reference KGs. The KFN describes three kinds of relationships, that is, the entity-concept relationship, the concept-concept relationship and the word-concept relationship.

The concept-entity relationship is represented by the generation probability from concept c to entity e. The probability $p(c|e)$ is calculated based on the page-view (PV) statistics of the Wikipedia entity pages:

$$P(c|e) = \frac{N_{PV}(e)}{\sum_{e' \in c} N_{PV}(e')} \tag{2}$$

The concept-concept relationship is represented by the transition probability between two concepts,

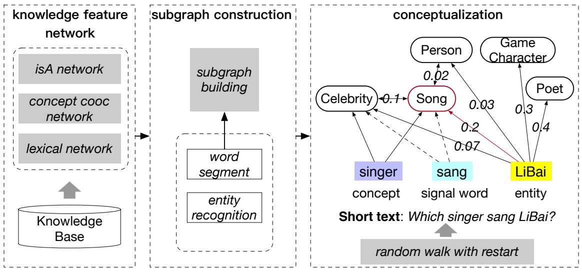

Figure 3: Architecture of fine-grained conceptualization, which consists of three parts: (a) Knowledge Feature Network. (b) Sub-graph construction. (c) Conceptualization.

c_i and c_j. The probability $P(c_i|c_j)$ is calculated based on the co-occurrence frequencies of the entities under the two concepts:

$$P(c_i|c_j) = \frac{\sum_{e_j \in c_j, e_i \in c_i} N(e_j, e_i)}{\sum_{c \in C} \sum_{e_j \in c, e_i \in c} N(e_j, e_i)} \tag{3}$$

where the co-occurrence frequency $N(e_j, e_i)$, is calculated based on the statistics of anchor links of Baidu Encyclopedia, and w is the size of the window that counts the co-occurrences frequencies of the entity pair in Baidu Encyclopedia. In this paper, w is set to 25.

$$N(e_j, e_i) = freq_w(e_j, e_i) \tag{4}$$

The word-concept relationship is represented by the labeling probability between the word w and the related concept c. The probability is calculated based on the word frequency and word-concept co-occurrence frequency:

$$P(c|w) = \frac{N(c, w)}{N(w)} \tag{5}$$

We perform a random walk algorithm (Jia-Yu Pan, 2004) on the KFN to get the appropriate concept of entity mention. First, we initialize the weights of the nodes and the edges by:

$$\mathbf{E}^0(e) = \begin{cases} P(c|t) & \textbf{if } e \textbf{ is } c \to t \\ P(c_i|c_j) & \textbf{if } e \textbf{ is } c_j \to c_i \end{cases} \tag{6}$$

$$\mathbf{N}^0(n) = \begin{cases} 1/|T| & \textbf{if } n \textbf{ is } entity \\ 0 & \textbf{if } n \textbf{ is } concept \end{cases} \tag{7}$$

Second, we iteratively update the node and edge by:

$$\mathbf{N}^k = (1-\alpha)\mathbf{E}' \times \mathbf{N}^{k-1} + \alpha \mathbf{N}^0 \tag{8}$$

$$\mathbf{E}^k \leftarrow (1-\beta)\mathbf{N}^k + \beta\mathbf{E}^k \tag{9}$$

where α and β are hyper-parameters tuned on developing sets. Finally, we normalize the edge weights and obtain the concept type with the highest weight:

$$\begin{aligned} c^* &= \arg\max_c P(c|t) \\ &= \arg\max_c \frac{\mathbf{E}(t \rightarrow c)}{\sum_{c_i} \mathbf{E}(t \rightarrow c_i)} \end{aligned} \tag{10}$$

3.2 KG-enhanced Representation of Entity

The conventional representation for an entity is the textual representation of the entity. Inspired by the wide usage of distributed representation of KG entities in many NLP applications, we think that such knowledge representation is also helpful in NED. In this work, we use both textual and knowledge representation to better represent the semantics of candidate entities. We propose a novel representation learning method which can simultaneously learns both kinds of knowledge. Based on the related textual context and other information of the entities, it uses the CBOW model with a sigmoid layer to generate the distributed representation of the entities.

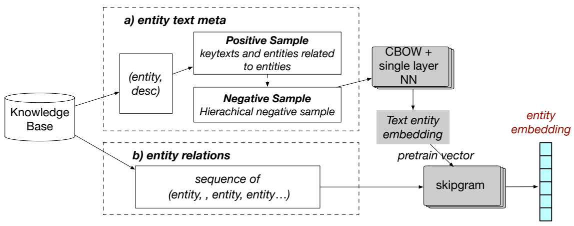

Figure 4: Entity context embedding architecture which combines entity relations and the entity context.

The detailed training process for the two models will now be introduced. Figure 4 shows that the entity e and its description generate the entity embedding . First, a positive sample is generated by entity description from KB (Wikipedia and Baidu Encyclopedia), and then word segmentation is applied to the entity description text. We have counted the word frequency in positive samples, and negative samples are generated by band-frequency random sampling. In order to learn the relationship between entities and enhanced entity representations, we use entity co-occurrence data and KB S-P-O data to generate training samples:

- entity co-occurrence sequence $\{e_1, e_2, \cdots, e_n\}$, which are extracted from KB hyperlinks.
- S-P-O triples from KB, which are extracted from the key-value block of Wikipedia and Baidu Encyclopedia.

We obtained entity sequences as training sample, where each entity has an entity embedding . Then we updated the entity embedding representation with Skip-Gram Model to enhance the inter-entity relationships. Finally, we obtain as the final entity embedding. Entity embedding vector are input as feature representations of entities into an KG-enhanced entity disambiguation network, as shown in Figure 2.

4 Experiments and Analysis

In this section, we first introduce the experimental dataset, and construction methods of the dataset we published with this paper. Then, we present evaluation metrics, the experiments conducted for both the English and the Chinese datasets with existing approaches, and we analyze the experimental results in detail.

4.1 Datasets

We have experimented on both Chinese and English datasets. For the English experiment, we use Wikipedia with a release time of 202003 as KB and apply the framework to NEEL and KORE50 datasets. For the Chinese experiment, due to the lack of large-scale short text entity disambiguation datasets, we constructed a dataset called DUEL and use it as the Chinese experiment dataset alongside the FUDAN dataset (Xu et al., 2017).

4.1.1 English Datasets

Most of the existing datasets on NED are based on long text, which are not suitable for our task. Two English datasets could be found that were suitable for short text entity disambiguation. Because KORE50 only has test data, but no training data, we use the training samples of NEEL as the training samples of KORE50 as well, to compare their performances.

- NEEL(Rizzo et al., 2017): The training dataset consists of 6,025 tweets, the validation dataset consists of 100 tweets, and the testing dataset consists of 300 tweets.

- KORE50(Hoffart et al., 2012): It contains 50 short sentences with highly ambiguous mentioned entities. It is considered to be among the most challenging for NED. Average sentence length (after removing stop words) is 6.88 words per sentence and each sentence has 2.96 mentioned entities on average.

4.1.2 Chinese Datasets

The typical size of existing Chinese NED datasets is about a few thousand annotated words (Rizzo et al., 2017; Hoffart et al., 2012). Because there is a lack of existing data sets for short text NED, we manually construct the largest available human annotated Chinese dataset, and we have released it to the global research community, please refer to this (DUEL) for more data details.

4.1.3 Construction of Our Dataset

Our dataset provides a high-precision manually-annotated entity disambiguation dataset consistin of 100,000 short texts. The text corpus consists of queries and web page titles. The annotated entities are in the general domain, including instances (e.g. Barack Obama) and concepts (e.g. Basketball player). Table 1 and 2 depict the statistical data of the KB and the annotated text.

Table 1: Statistics of knowledge base in our dataset. AvgNumOfEntityProperties is the average number of attributes for all entities.

Statistic	KB
#Entities	398082
#SPO	3564565
#EntitiyDesc	361778
#AvgNumOfEntityProperties	9

Data Annotation Method: We annotated the entire short text in the dataset by crowd-sourcing. The same data was repeatedly labeled by three domain experts, then reviewed and released by additional experts. The average precision of annotating entities is about 95.2%. The evaluation method of dataset is as follows: given an input of a short text q, the annotated entities is $E'_q = e'_1, e'_2, e'_3, ...$. By comparing the outputs E'_q with additional experts-annotated set $E_q = e_1, e_2, e_3, ...$, precision P is defined as follows.

$$P = \frac{\sum_{q \in Q} E_q \cap E'_q}{\sum_{q \in Q} E_q} \quad (11)$$

Comparison with previous datasets: As summarized in Table 2, FUDAN is a representative evaluation dataset for Chinese short text entity disambiguation, which consists of manually annotated short text. Both FUDAN and DUEL consist of entities in various domains (including instances and concepts), such as persons, movies, and general concepts. However, DUEL is much larger.

4.2 Results and Analysis

4.2.1 Evaluation Metrics

We directly use the gold standard in mentioned entities - the NER results in the dataset, and choose standard micro F1 score as our performance metric for NED task (aggregated over all mentions).

4.2.2 Performance Comparison with Other Approaches

In order to verify the enhancement of different methods used in NED, we compare the proposed method with several state-of-the-art approaches both for the Chinese and the English datasets. All of these methods are effective and comparable in the case of short text. Our method is called knowledge-enhanced NED (K-NED).

- FEL(Blanco et al., 2015): A toolkit for training models to link entities to KB in documents and queries. And we use DSSM model to use this entity embedding for comparing. We experiment with default parameters.

- NTEE (Yamada et al., 2017): A neural network model that learns embedding of texts and Wikipedia entities, and then use them in entity linking task. We experiment with default parameters.

- Mulrel-nel (Le and Titov., 2018): A python implementation of multi-relational NED. We experiment with default parameters.

- Fudan (Xu et al., 2017): Entity linking of Fudan University which is a Chinese entity linking service API.

As summarized in Tables 3, the experimental results indicate that our approach K-NED outperforms existing state-of-the-art methods such as FEL, NTEE, Mulrel-nil and Fudan on Chinese and English datasets except on KORE50. In particular, our method disambiguate to all correct result of the examples in Figure 1. We found that 72%

Table 2: Comparisons between DUEL and the FUDAN dataset. AvgLen is the average length of the annotated text. AvgNumEntity is the average number of entities in the annotated text.

Statistic	**DUEL**	**FUDAN**	**NEEL**	**KORE50**
#Train	90000	-	6025	-
#Dev	10000	-	100	-
#Test	10000	1037	300	50
#AvgLen	21.73	23.38	16.5157	6.88
#AvgNumEntity	3.43	2.08	2.1	2.96
#Accuracy	95.2%	-	-	-

Table 3: F1 scores on Chinese and English datasets.

Method	Datasets	
	Chinese datasets	
	DUEL	FUDAN
Fudan	0.861	0.945
Mulrel-nel	0.889	0.893
K-NED(Ours)	**0.897**	**0.947**
	English datasets	
	NEEL	KORE50
FEL	0.601	0.360
NTEE	0.748	0.618
Mulrel-nel	0.805	**0.625**
K-NED(Ours)	**0.811**	0.544

of the types of annotated entities in the KORE50 dataset belong to the category ”Person”, and so it is possible that this dataset distribution is biased. Compared to KORE50, NEEL, FUDAN and DUEL datasets are more consistent with the entity type distribution of practical scenarios. NTEE and FEL use representational learning to improve performance, Mulrel-nel relied on supervised systems or heuristics to predict these relations and treat relations as latent variables in neural entity disambiguation model, and our approach uses knowledge enhancement to improve the performance without using other complex features. Data analyses demonstrate that each short text contains 3 mentioned entities on average, each of which includes 20 ambiguous entities to be linked, and the context is sparse. Experiments demonstrate that knowledge enhancement is helpful for short text entity disambiguation.

4.2.3 Performance of Knowledge-Enhancement Components

In order to gain a deeper understanding of the various components of our model, we compare the difference in performance after removing two components separately, where all models are trained using the same settings.

Table 4: F1 scores of each component on Chinese and English datasets.

Feature	DUEL	NEEL
K-NED	0.897	0.811
K-NED -DC	0.804	0.755
K-NED -ECE	0.874	0.779
K-NED -DC-ECE	0.759	0.577

As listed in Table 4, K-NED is the result of our complete model. DC represents the fine-grained dynamic conceptualization component, and ECE represents entity context embedding components. We find that dynamic conceptualization exhibits a 10.36%improvement in performance, and entity context embedding exhibits a 2.56% improvement in performance on our Chinese dataset: DUEL. By the analysis of examples in Figure 1, we find that dynamic conceptualization can mark the concepts of ”Li Bai” in ”Who is the singer of Li Bai?”, ”How to play the hero Li Bai?” and ”Which famous poems are written by Li Bai?” as ”songs”, ”game characters” and ”poets” respectively. The correct conceptualization greatly facilitates the entity disambiguation. On the other hand, however, although we successfully mark the concept of ”Li Bai” in ”Wedding photos of Li Bai and Sa Beining” as ”person”, it still disambiguates incorrectly without the help of entity context embedding, which indicates that dynamic conceptualization and entity context embedding can be complementary in NED.

This result demonstrates that the conceptualization of entities is more direct and effective for the semantic disambiguation in short text entity disambiguation. In addition, we find that fine-grained conceptualization plays a significant role in dynamic conceptualization.

5 Related works

Many efforts have been devoted to NED in recent years. Some methods (Shen et al., 2014; Ratinov et al., 2011; Shen et al., 2012b,a; Han, 2015) exploit the Learning To Rank framework (LTR) (Liu, 2009) to rank the candidate entities, taking advantage of the relationships between all candidates. Most commonly used ranking models are the pairwise framework (Perceptron (Shen and Joshi, 2005), RankSvm(Chingpei Lee, 2014)) and the listwise framework (ListNet (Cao et al., 2007)). (Bao-Xing et al., 2014) proposed a named entity linking method based on a probabilistic topic model(Blei, 2012), which employs the conceptual topic model to map words and mentioned entities into the same topic space. (Nakashole et al., 2013) used a graph-based collaborative entity linking model. (Bilenko et al., 2003) proposed using random walks for entity linking. Some models choosed to rely solely on the context of the links to learn entity representations, such as (Lazic et al., 2015), and some methods used a pipeline of existing annotators to filter entity candidates such as (Ling et al., 2015). Different from these conventional work, we use multiple sources of information and a deep structured semantic model to achieve better NED performance.

Many efforts have been devoted to NED for queries, such as (Hasibi et al., 2015) and (Hasibi et al., 2017). Some approaches try to solve NED by making extensive use of deep neural networks (Globerson et al., 2016), or by adopting distributed representations of words or entities (Yamada et al., 2016, 2017). Other existing approaches take advantage of global context, which captures the coherence between mapped entities of the related keywords in a document (Cucerzan, 2007; Han et al., 2011). In (Globerson et al., 2016), the neural network model uses attention mechanism to focus on the contextual entities to be disambiguated. In (Yamada et al., 2016, 2017), the distributed representation of contexts models the relationships between words and entities or between documents and entities, where the distances between various vectors provides useful information for disambiguation. Different from these work where complicated techniques or features are used, we adopt external knowledge including factual and conceptual knowledge graphs for better NED performance. (Radhakrishnan and Varma, 2018) proposes a method to train entity embedding for entity similarity, but this method relies on a dense knowledge map, we use the text and relationship information of entity to model the similarity between entity and context to improve the effect of NED.

There are also previous work using concept or type information to improve NED performance. The models of (Hua et al., 2015; Wang et al., 2015; Priya Radhakrishnan, 2018; Isaiah Onando Mulang, 2020) try to map short text to a concept space, and then generate comprehensive concept vectors to represent the short text. (Raiman and Raiman, 2018) constructs a type ontology and a type classifier to map entities to a closed type ontology. (Chen and Xiao, 2018) proposes to modeling context explicitly by entity concept, but we use a complementary way of coarse-grained and fine-grained to dynamically predict the concept according to the context and improve the effect of disambiguation. (Derczynski et al., 2015) studies named entity recognition (NER) and named entity linking (NEL) for tweets. Unlike these work, our method uses fine-grained entity concepts and predicts concepts more accurately by using an advanced knowledge feature network.

6 Conclusion

We propose a knowledge-enhanced approach to short text entity disambiguation. Through bridging and facilitating semantic understanding of the fine-grained concept associated to a mentioned entity and entity embedding, the performance of entity disambiguation can be significantly improved. The experimental results demonstrate that our approach outperforms existing SOTA methods on English and Chinese datasets for this task. At the same time, we constructed a large-scale manual-annotated Chinese dataset for short text entity disambiguation, which has been released with the paper for use by researchers. As a future direction of research, we plan to explore better conceptualization and semantic understanding methods, and further improve the performance of the short text entity disambiguation task. We intend to continue to update our Chinese dataset.

In the future, We will use more modern embedding(such as BERT) or encoder(such as transformer) to obtain better embedding, and we also plan to conduct experiments to verify the effectiveness of our methods in other tasks related to semantic understanding such as Q&A, Dialogue, etc.

References

HUAI Bao-Xing, BAO Teng-Fei, ZHU Heng-Shu, and LIU Qi. 2014. Topic modeling approach to named entity linking. Journal of Software.

M. Bilenko, R. Mooney, W. Cohen, P. Ravikumar, and S. Fien-berg. 2003. Adaptive name matching in information integration. IEEE Intelligent Systems.

Roi Blanco, Giuseppe Ottaviano, and Edgar Meij. 2015. Fast and space-efficient entity linking for queries. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, pages 179–188. ACM.

David M. Blei. 2012. *Probabilistic Topic Models*. Magazine Communications of the ACM, ACM New York, NY, USA.

Z. Cao, T. Qin, T.-Y. Liu, M.-F. Tsai, and H. Li. 2007. Learning to rank: from pairwise approach to listwise approach. ICML.

Liang J. Xie C. Chen, L. and Y. Xiao. 2018. Short text entity linking with fine-grained topics. pages 457–466. CIKM.

Chihjen Lin Chingpei Lee. 2014. Large-scale linear ranksvm. pages 781–817. Neural Computation.

S Cucerzan. 2007. Large-scale named entity disambiguation based on wikipedia data. EMNLP CoNLL.

Leon Derczynski, Diana Maynard, Giuseppe Rizzo, Marieke Van Erp, Genevieve Gorrell, Raphaël Troncy, Johann Petrak, and Kalina Bontcheva. 2015. Analysis of named entity recognition and linking for tweets. *Information Processing & Management*, 51(2):32–49.

A. Globerson, N. Lazic, S. Chakrabarti, A. Subramanya, M. Ringaard, and F.n Pereira. 2016. Collective entity resolution with multifocal attention. ACL.

Wei Shen; Jianyong Wang; Jiawei Han. 2015. Entity linking with a knowledge base: Issues, techniques, and solutions. pages 443 – 460. IEEE Transactions on Knowledge and Data Engineering.

Xianpei Han, Le Sun, and Jun Zhao. 2011. Collective entity linking in web text: a graph-based method. page 765–774. SIGIR.

Faegheh Hasibi, Krisztian Balog, and Svein Erik Bratsberg. 2015. Entity linking in queries: Tasks and evaluation. In *Proceedings of the 2015 International Conference on The Theory of Information Retrieval*, pages 171–180. ACM.

Faegheh Hasibi, Krisztian Balog, and Svein Erik Bratsberg. 2017. Entity linking in queries: Efficiency vs. effectiveness. In *European Conference on Information Retrieval*, pages 40–53. Springer.

J. Hoffart, S. Seufert, D.B. Nguyen, M. Theobald, and G. Weikum. 2012. Kore: Keyphrase overlap relatedness for entity disambiguation. pages 545–554. ACM international conference on Information and knowledge management.

Wen Hua, Zhongyuan Wang, Haixun Wang, Kai Zheng, and Xiaofang Zhou. 2015. Short text understanding through lexical-semantic analysis. IEEE 31st International Conference on Data Engineering.

Po-Sen Huang, Xiaodong He, Jianfeng Gao, Li Deng, Alex Acero, and Larry Heck. 2013. Learning deep structured semantic models for web search using clickthrough data. pages 2333–2338. CIKM.

Akhilesh Vyas Saeedeh Shekarpour Maria Esther Vidal Soren Auer Jens Lehmann Isaiah Onando Mulang, Kuldeep Singh. 2020. Encoding knowledge graph entity aliases in attentive neural network for wikidata entity linking. page 15. WISE.

Christos Faloutsos Pinar Duygulu Jia-Yu Pan, Hyung-Jeong Yang. 2004. Automatic multimedia cross-modal correlation discovery. pages 653–658. KDD.

Nevena Lazic, Amarnag Subramanya, Michael Ringgaard, and Fernando Pereira. 2015. Plato: A selective context model for entity resolution. TACL.

Phong Le and Ivan Titov. 2018. Improving entity linking by modeling latent relations between mentions. ACL.

Xiao Ling, Sameer Singh, and Daniel S Weld. 2015. Design challenges for entity linking. TACL.

T.Y. Liu. 2009. Learning to rank for information retrieval. Foundations and Trends in Information Retrieval.

N. Nakashole, T. Tylenda, and G. Weikum. 2013. Fine-grained semantic typing of emerging entities. ACL.

Zhou S.-Liu J. Wang J. Lin C.Y. Nie, F. and R. Pan. 2018. Aggregated semantic matching for short text entity linking. pages 476–485. CoNLL.

Thomas Hofmann Octavian-Eugen Ganea. 2017. Deep joint entity disambiguation with local neural attention. page 2619–2629. ACL.

Vasudeva Varma Priya Radhakrishnan, Partha Talukdar. 2018. Elden: Improved entity linking using densified knowledge graphs. page 1844–1853. NAACL.

Talukdar P. Radhakrishnan, P. and V. Varma. 2018. Elden: Improved entity linking using densified knowledge graphs. pages 1844–1853. NAACL.

Jonathan Raiman and Olivier Raiman. 2018. Deeptype: Multilingual entity linking by neural type system evolution. AAAI.

L. Ratinov, D. Roth, D. Downey, and M. Anderson. 2011. Local and global algorithms for disambiguation to wikipedia. ACL.

G. Rizzo, B. Pereira, A. Varga, M. van Erp M, and A.E. Cano Basave. 2017. the named entity recognition and linking (neel) challenge series. Semantic Web Journal (in press).

L. Shen and A. K. Joshi. 2005. Ranking and reranking with per-ceptron. Mach. Learn.

W. Shen, J. Wang, P. Luo, and M. Wang. 2012a. Liege: Link entities in web lists with knowledge base. SIGKDD.

W. Shen, J. Wang, P. Luo, and M. Wang. 2012b. Linden: linking named entities with knowledge base via semantic knowledge. WWW.

Wei Shen, Jianyong Wang, and Jiawei Han. 2014. Entity linking with a knowledge base: Issues, techniques, and solutions. pages 443–460. IEEE Transactions on Knowledge and Data Engineering.

Kai Chen Greg Corrado Jeffrey Dean Tomas Mikolov, Ilya Sutskever. 2013. Distributed representations of words and phrases and their compositionality. pages 3111–3119. NIPS.

Zhongyuan Wang, Kejun Zhao, Haixun Wang, Xiaofeng Meng, and Ji-RongWen. 2015. Query understanding through knowledge-based conceptualization. IJCAI.

Bo Xu, Yong Xu, Jiaqing Liang, Chenhao Xie, Bin Liang, Wanyun Cui, and Yanghua Xiao. 2017. Cndbpedia: A never-ending chinese knowledge extraction system. In *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*, pages 428–438. Springer.

I. Yamada, H. Shindo, H. Takeda, and Y. Takefuji. 2017. Learning distributed representations of texts and entities from knowledge base. arXiv preprint.

I. Yamada, H. Shindo, and H. Takedaand Y. Takefuji. 2016. Joint learning of the embedding of words and entities for named entity disambiguation. arXiv preprint.

Luo Si Jeongwoo Ko Aditya P. Mathur Yi Fang, Naveen Somasundaram. 2011. Analysis of an expert search query log. pages 1189–1190. SIGIR.