Graph-Based Chinese Word Sense Disambiguation with Multi-Knowledge Integration

Wenpeng Lu[1, *], Fanqing Meng[2], Shoujin Wang[3], Guoqiang Zhang[4], Xu Zhang[1], Antai Ouyang[5] and Xiaodong Zhang[6]

Abstract: Word sense disambiguation (WSD) is a fundamental but significant task in natural language processing, which directly affects the performance of upper applications. However, WSD is very challenging due to the problem of knowledge bottleneck, i.e., it is hard to acquire abundant disambiguation knowledge, especially in Chinese. To solve this problem, this paper proposes a graph-based Chinese WSD method with multi-knowledge integration. Particularly, a graph model combining various Chinese and English knowledge resources by word sense mapping is designed. Firstly, the content words in a Chinese ambiguous sentence are extracted and mapped to English words with BabelNet. Then, English word similarity is computed based on English word embeddings and knowledge base. Chinese word similarity is evaluated with Chinese word embedding and HowNet, respectively. The weights of the three kinds of word similarity are optimized with simulated annealing algorithm so as to obtain their overall similarities, which are utilized to construct a disambiguation graph. The graph scoring algorithm evaluates the importance of each word sense node and judge the right senses of the ambiguous words. Extensive experimental results on SemEval dataset show that our proposed WSD method significantly outperforms the baselines.

Keywords: Word sense disambiguation, graph model, multi-knowledge integration, word similarity.

1 Introduction

The ambiguous words are ubiquitous in human languages, which leads to a huge confusion for natural language processing (NLP). Word sense disambiguation is to

---

[1] School of Computer Science and Technology, Qilu University of Technology (Shandong Academy of Sciences), Jinan, 250353, China.

[2] School of Information Science and Engineering, Zaozhuang University, Zaozhuang, 277160, China.

[3] Department of Computing, Macquarie University, Sydney, NSW 2109, Australia.

[4] Centre for Audio, Acoustics and Vibration, University of Technology Sydney, Sydney, NSW 2006, Australia.

[5] School of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao, 266590, China.

[6] Jinan Intellectual Property Information Center, Jinan, 250099, China.

[*] Corresponding Author: Wenpeng Lu. Email: lwp@qlu.edu.cn.

determine the meaning of a word according to its context, which is a fundamental task in NLP that directly affect the upper applications, e.g., machine translation, information retrieval, text categorization and automatic summarization [Raganato, Camacho-Collados and Navigli (2017); Lu, Wu, Jian et al. (2018); Xiang, Li, Hao et al. (2018)].

The existing WSD methods are divided into three categories: supervised, unsupervised and knowledge-based methods. Supervised method trains the classifiers with machine learning on sense-annotated corpus, which are utilized to judge the senses of new instances [Raganato, Bovi and Navigli (2017)]. Though supervised method can achieve the best disambiguation performance, its effectiveness depends on the size and quality of the sense annotated corpus. Due to the limitation of annotated corpus, supervised method is hard to be applied on a large-scale WSD task. Unsupervised method distinguishes the categories of word senses according to their context with clustering technology, which can only different sense categories instead of senses and can not annotate each instance with its accurate sense [Panchenko, Ruppert, Faralli et al. (2017)]. Knowledge-based method judges the sense of each instance according to its context and various knowledge bases. Though the performance of knowledge-based method is not better than that of supervised one, it can utilize all kinds of existing knowledge bases and can achieve a better coverage [Raganato, Camacho-Collados and Navigli (2017)]. Knowledge-based method is the unique method which is available on large-scale WSD tasks and has achieved better performance in SemEval [Moro and Navigli (2015); Navigli and Ponzetto (2012); Raganato, Camacho-Collados and Navigli (2017); Chen, Liu and Sun (2014)]. The existing knowledge bases contain abundant semantic relationships, which can form a huge semantic graph and are beneficial to WSD. Graph-based WSD is a representative one of knowledge-based methods, which is the most popular method and has attracted more and more attention in NLP field [Dongsuk, Kwon, Kim et al. (2018); Duque, Stevenson, Martinez-Romo et al. (2018); Meng, Lu, Zhang et al. (2018)]. Graph-based WSD constructs the disambiguation graph according to semantic knowledge relationships, whose performance is affected greatly by the size and quality of knowledge resources. Knowledge acquisition bottleneck is the key factor that limits its development, which is more serious in Chinese due to the rareness of Chinese semantic knowledge resources [Lu (2018)].

The traditional graph-based Chinese WSD method usually utilizes one kind of Chinese knowledge resource, which is extremely troubled with the problem of knowledge bottleneck [Lu, Huang and Wu (2013); Yang and Huang (2012)]. Compared with knowledge resources in Chinese, those in English are more mature and abundant. If we can integrate various Chinese and English knowledge resources together, which can complement each other, we can fully exploit all kinds of disambiguation knowledge. This shows the potential to significantly improve the performance of Chinese WSD.

Apparently, how to integrate the existing Chinese and English knowledge resources is actually highly challenging, as the senses of them are not mapped to each other. Besides, how to evaluate the overall similarities of sense pairs is difficult, as the relative importance of each knowledge resource is unknown for us. Inspired by the significant progress made on representation learning and optimization algorithm in various tasks such as sentence representation [Mikolov, Sutskever, Chen et al. (2013); Subramanian,

Trischler, Bengio et al. (2018)] and simulated annealing optimization [Mafarja and Mirjalili (2017); Mamano andHayes (2017)], this work integrates the existing English and Chinese knowledge resources and optimizes their weights to construct a knowledge graph so as to disambiguate the ambiguous words in Chinese. The main idea and contributions are as follows:

- We propose a novel knowledge integration method, which merge the English and Chinese knowledge resources together by sense definition alignment with the help of sentence representation. The method is flexible, which can integrate various knowledge conveniently.

- We propose a simulated annealing algorithm to optimize the weights of various knowledge. With the optimized weights, the semantic relationships between senses are evaluated to construct an overall knowledge graph.

- To the best of our knowledge, this is the first work on graph-based Chinese WSD with multiple-knowledge integration. This work maps and integrates a variety of English knowledge resources into Chinese, and optimizes their weights with simulated annealing algorithm to compute similarities of sense pairs. According to the senses and their similarities, an overall knowledge graph is constructed, where the graph scoring algorithm evaluates the importance of the sense nodes to judge the right sense.

Extensive experiments on SemEval WSD task are conducted to evaluate our proposed method. The result shows that our method substantially defeat the existing methods, with at least 2.4% improvement.

The rest of this paper is organized as follows: Section 2 discusses the related work and gives a brief summary in WSD. Section 3 details the proposed graph-based Chinese WSD with multi-knowledge integration, where each key module is described. Section 4 provides the empirical results by comparing our method with the baselines. Finally, we conclude this work and provide future work in Section 5.

2 Related work

Graph-based WSD methods are inspired by the lexical chain, which refers to a sequence of semantic related words in a given text, that are linked together by lexical semantic relations, e.g., eat → apple → fruit → banana. Graph-based WSD is the most popular method in knowledge-based WSD, which constructs a knowledge graph with senses as nodes and semantic relations as edges. Based on the structure of knowledge graph, the right sense is selected [Dongsuk, Kwon, Kim et al. (2018)].

Galley et al. [McKeown and Galley (2003)] have proposed a WSD method based on lexical chain, introduced as follows. Firstly, when constructing the disambiguation graph, all possible sense is added to the graph as nodes, then the words in the ambiguous sentence are processed one by one. If there exists a semantic relationship between the current word and the processed ones, this relationship is added to the graph as an edge, which is assigned a weight according to the type of relationship and the distance. After the graph is constructed, the weights of sense nodes of ambiguous words are summed and the sense with the greatest weight is selected as the right sense. The method achieves

62.1% accuracy on SemCor noun dataset.

Mihalcea [Mihalcea (2004)] has proposed a WSD method based on PageRank algorithm, which takes all the senses of the words as the nodes and the semantic relationships between the words as the edges, to construct the disambiguation graph. Pagerank algorithm is applied on the graph to evaluate the importance of each sense node to judge the right sense. Agirre et al. propose personalized PageRank for WSD [Agirre and Soroa (2009)], which pays more attention on some words and improves the evaluation of sense importance.

Navigli et al. [Navigli and Velardi (2005)] propose a structural semantic interconnections (SSI) algorithm for WSD, which creates structural specifications of the possible senses for each word and constructs the grammar rules to describe the interconnection relations. The most suitable sense is selected according to the grammar. SSI achieves the best performance in Senseval-3 and SemEval-2007.

Yang et al. [Yang and Huang (2012)] propose a graph-based WSD method based on word distance, which strengthens the influence of near words and weakens that of far words when evaluating the importance of sense nodes in the graph. Lu et al. [Lu, Huang and Wu (2014)] propose a graph-based WSD method based on domain knowledge, which integrates domain knowledge into the disambiguation framework and improves multiple graph scoring algorithms.

Traditional graph-based method try to construct the subgraph of all words in a sentence, which may induce some noise information [Navigli and Lapata (2010)]. To avoid the problem, Dongsuk et al. [Dongsuk, Kwon, Kim et al. (2018)] propose a WSD method based on subgraph reconstruction, where context words of an ambiguous word for constructing the subgraph are selected with a word similarity threshold. The word similarity is computed based on an embedding generated by Doc2Vec [Le and Mikolov (2014)], which encodes information of the semantic relational path of words in BabelNet [Navigli and Ponzetto (2012)].

The above existing graph-based WSD methods construct the disambiguation graph according to some lexical knowledge resources, e.g., WordNet, BabelNet and HowNet [Miller (1995); Navigli and Ponzetto (2012); Zhendong and Qiang (2006)]. Most of them only utilize one kind of knowledge resource. As the limitation of size and quality of the resource, the graph-based methods are confused with knowledge bottleneck. Apparently, the knowledge resources are different and complementary. It is necessary to integrate as many as existing resources to strengthen the ability of WSD systems. Comparing with English, the available Chinese semantic resources are more rare, which makes the problem more critical. How to integrate the existing various semantic resources to improve the performance of Chinese WSD is an important issue that is waiting to be solved.

3 The proposed WSD method

In this section, we describe the framework of graph-based WSD method and its key modules in detail. With the framework, for sense pairs of Chinese words, the English knowledge resources are utilized to compute their similarities together with the Chinese resources. The weights of the similarities are optimized with simulated annealing algorithm. The disambiguation graph is constructed with senses as nodes, semantic

relations as edges and similarities as their weights, where graph algorithm is utilized to score each sense node and select the right sense. The framework and its key modules are introduced as follows.
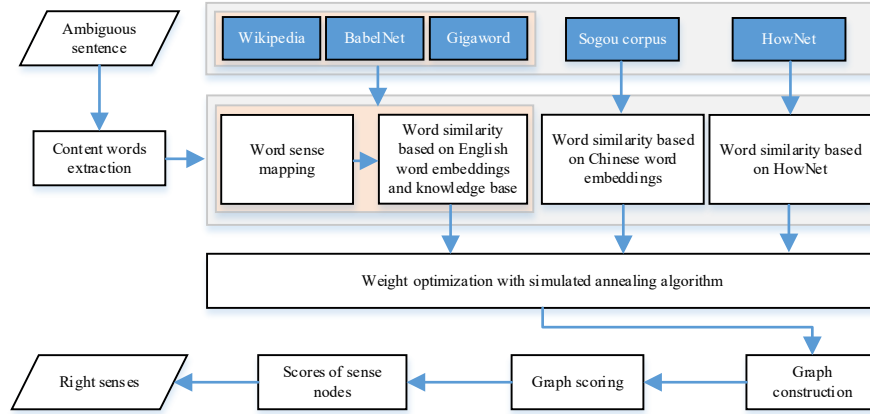
Figure 1: Model architecture of sentence matching

3.1 Framework of the WSD method

The framework of our proposed graph-based WSD with multi-knowledge integration is shown in Fig. (1). The content words in a Chinese sentence are extracted and mapped into English words with BabelNet. By the mapping, the resources in English would be available for Chinese words. Then, according to English and Chinese knowledge resources, three kinds of word similarity are computed, whose weights are optimized with simulated annealing algorithm so as to obtain overall similarities to construct the disambiguation graph. The importance score of each sense node in the graph is evaluated to select the right senses of ambiguous words. The detailed framework is described as follows:

(1) Extract the content words after preprocessing the Chinese ambiguous sentence.

(2) Map Chinese word senses into English ones [Meng, Lu and Xue (2017)].

(3) Compute word similarity based on English word embeddings and knowledge bases, e.g., Wikipedia, BabelNet, Gigaword [Parker, Graff, Kong et al. (2011)].

(4) Compute word similarity based on Chinese word embeddings trained on Sogou corpus3.

(5) Compute word similarity based on HowNet [Zhendong and Qiang (2006)].

(6) Optimize the relative weights of the above three kinds of word similarity with simulated annealing algorithm so as to obtain an overall similarities.

(7) Take word senses as nodes, semantic relations as edges, overall similarities as weights of edges, to construct the disambiguation knowledge graph.

(8) Evaluate the importance of each sense node in the graph with graph scoring algorithm to select the right sense.

As shown in Fig. 1. sense mapping module, three modules of word similarity, weight optimization module, graph construction and scoring module are the key components of our proposed methods, which are explained in the following subsections.

3.2 Word sense mapping

As the rareness of Chinese semantic knowledge resources, mapping Chinese word sense to English ones and utilizing English resources to compensate the deficiency of Chinese resources is a practicable solution. In order to map the senses in Chinese and English semantic resources, we have proposed a method to map the senses between Chinese and English with BabelNet and English-Chinese dictionary [Meng, Lu and Xue (2017); Navigli and Ponzetto (2012); Ke (2011)].

For each English sense, BabelNet has provides a detailed definition with several short examples. Beside, the English-Chinese dictionary, i.e., Collins COBUILD Advanced Learner's English-Chinese Dictionary, has provided the detailed bilingual definitions with bilingual examples. That is, both BabelNet and Collins dictionary have provided English description for each sense, and the latter also provides the corresponding Chinese sense annotation. If an English sense is corresponding with a Chinese sense, the meaning of their English definitions or examples should be similar. This is an important and key clues to find and verify the mapping relations among Chinese and English senses.

With this in mind, we generate embedding representations for the English definitions and examples. According to their cosine similarities, we find the corresponding relationships among BabelNet and Collins definitions in English. Then, as Collins English-Chinese dictionary provides English and Chinese definition simultaneously, we can further obtain the mapping relations between English and Chinese. The detailed implementation is introduced as follows.

Firstly, for each Chinese sense, its possible candidate English senses are prepared according to HowNet or Chinese-English dictionary. Secondly, for the candidate English senses,we get their definitions and examples according to BabelNet, and collect the bilingual definitions and example according to an English-Chinese bilingual dictionary. Thirdly, inspired by related work with Word2vec [Mikolov, Sutskever, Chen et al. (2013); Le and Mikolov (2014)], we generate embedding representation for each sentence in definitions and examples. Finally, the cosine similarities among embedding representations are computed to find the corresponding English sense for each Chinese sense. Once the senses are mapped between Chinese and English knowledge resources, we can utilize English semantic resources to assist Chinese WSD tasks, which will provide great convenience for Chinese WSD. Our another paper has described the above procedures carefully, whose F1-measure reaches 75.75% [Meng, Lu and Xue (2017)].

3.3 Word similarity based on English word embeddings and knowledge base

After the mapping processing in last subsection, the Chinese and English senses are mapped each other. Then, the English knowledge resources can be utilized to disambiguate the words in Chinese. When the disambiguation graph is constructed, each semantic relation between sense nodes need to be assigned a reasonable weight, which should consider as more as information. In this subsection, the information from English knowledge resources is considered.

We have realized a method for word similarity computation based on English word embeddings and knowledge base, as described in Meng et al. [Meng, Lu, Zhang et al.

(2017)]. The method has participated in SemEval-2017 Task 24, i.e., multilingual and cross-lingual semantic word similarity [Camacho-Collados, Pilehvar, Collier et al. (2017)]. In the competition, our method has reached 0.778 on official evaluation measure, which wins the second place on English monolingual word similarity subtask [Meng, Lu, Zhang et al. (2017)].

In SemEval 2017 Task 2[5], i.e., multilingual and cross-lingual semantic word similarity, we have proposed a word similarity based on English word embeddings and knowledge base [Meng, Lu, Zhang et al. (2017)], whose performance reaches 0.781, which wins the second place on English monolingual word similarity subtask. Since the competition system has achieved the excellent performance, we integrate it into our proposed WSD framework, and utilize it to compute the word similarity based on English knowledge resources, which are introduced as follows.

The method is a combination method consisting of two basic modules, which are the similarity based on word embedding and the similarity based on knowledge base, i.e., BabelNet. For the former, Word2Vec toolkit[6] is used to train word embedding on English wikipedia corpus [Mikolov, Sutskever, Chen et al. (2013)]. With the embeddings of each word pair, their cosine similarity is computed. For the latter, BabelNet[7] contains a large number of concepts and semantic relations, such as synonymy, hypernymy and meronymy.

With BableNet API, we can obtain all of the semantic relations among two words. According to the shortest path, the similarity of word pair is computed. The similarity based on word embedding and the similarity based on knowledge base are linearly weighted accumulated as the overall similarity based on English knowledge resources. Our another paper has introduced the implementation in detail [Meng, Lu, Zhang et al. (2017)]. The method is flexible, which can combine more knowledge resources.

3.4 Word similarity based on Chinese word embeddings

In the last subsection, with the support of word sense mapping, word similarity based on English word embedding is integrated into our WSD framework. Since we aim at the disambiguation problem in Chinese, word similarity based on Chinese word embeddings is crucial and necessary.

As Word2Vec has demonstrated a powerful ability in various tasks [Mikolov, Sutskever, Chen et al. (2013)], we continue to utilize it to generate Chinese word embeddings, which is trained on Sogou news corpus[8] . With the Chinese word embeddings, we compute their cosine similarity as the word similarity based on Chinese word embeddings.

3.5 Word similarity based on HowNet

For word similarity of Chinese words, besides word embedding method in last subsection,

---

[5] http://alt.qcri.org/semeval2017/task2/

[6] https://code.google.com/p/word2vec/

[7] https://babelnet.org/

[8] http://www.sogou.com/labs/resource/ca.php

HowNet[9] also provides the API interface to compute word similarity [Zhendong and Qiang (2006)].

HowNet is a common semantic knowledge base, which describes the concepts in Chinese and English and the relationships among concepts and their attributions. There are about 800 lexemes in HowNet, which is the basic and smallest unit of meaning that can not be divided further. All concepts in HowNet are described with the basic lexemes.

HowNet is widely applied in Chinese NLP field, which provides a convenient API interface, i.e., Hownet_GET_Concept_Similarity, to compute the semantic similarity between two concepts [Yang and Huang (2012)]. The similarity has considered multiple relationships from HowNet, including four kinds of primitive lexeme similarities [Qun and Sujian (2002)], which are computed according to their path distance on HowNet hierarchical structure.

3.6 Weight optimization with simulated annealing algorithm

The above three similarity methods compute word similarities with different semantic knowledge resources, which are complementary each other. In order to fully utilize their respective advantages, we propose a weight optimization algorithm with simulated annealing algorithm to automatically decide the weight parameters of the three similarities, which are used to linearly combine them so as to obtain a more reasonable overall similarity. The procedure to optimize the weight parameters is shown in Algorithm 1. The core of simulated annealing algorithm for weight optimization is described as:

$$p = \begin{cases} 1, if\ result\left(x_{new}\right)\ \geq\ result\left(x_{old}\right) \\ \exp(\dfrac{result(\text{x}_{new})-result(\text{x}_{old})}{\delta t}), if\ result\left(x_{new}\right) > result\left(x_{old}\right) \end{cases}, \tag{1}$$

where *result(x)* is the target function, i.e., the disambiguation accuracy, δ is cooling rate, t is the temperature. If the result of new parameters x_{new} is better than that of x_{old}, the new parameters would be selected with a probability value of 1. Otherwise, the new parameters would be selected with a probability value of $\exp(\dfrac{result(\text{x}_{new})-result(\text{x}_{old})}{\delta t})$.

Algorithm 1 Weight optimization with simulated annealing algorithm.

Input:

The initial weight values, *x, y, z*

Output:

The best optimized weight values, *x, y, z*

1: Initialize t, t_{min}, δ, k, y;

2: while $t > t_{min}$ do {

---

[9] http://www.keenage.com/

```
3:     for i = 0 to k do {
4:         x = a randomly generated double value from 0 to 1-y
5:         z = 1 − x − y
6:         result = getEvalResult(x, y, z)
7:         x_new = x + (Math.random() × 2 − 1) × (t/100)
8:         if x_new >= 0 and x_new <= 1 − y then {
9:             result_new = getEvalResult(x_new, y, z)
10:            if result_new ≥ result then
11:                x = x_new
12:            else {
13:                p = Math.exp((result_new − result)/t)
14:                if Math.random() < p then
15:                    x = x_new
16:            }
17:        }
18:    }
19:    t = δ × t
20: }
21: return the best optimized weight values of x, y, z
```

In Algorithm 1, the parameters x, y, z is the weights of three kinds of word similarity, which need to be optimized. Line 1 is initialization operation, which sets the initial temperature t as 100, the minimal cooling temperature t_{min} as 0.001, the cooling rate δ as 0.98, the maximum iterations k as 100 in the experiments. Line 4-5 select a random double value to x, which affects the value of z. In the Line 6, *getEvalResult* is the target function, which returns the disambiguation accuracy, given the parameters x, y, z. Line 7 generates an updated value to x_{new} from the neighbourhood of x. Line 8-18 decide whether the parameter x is update with the new x_{new}, as described in Eq. (1). Line 19 changes the value of t with the cooling rate δ. We obtain the three optimized weight parameters by running Algorithm 1 twice. In the first run, we set the value of y as 1/3, and get the optimized weights x, z. Then, we keep the minor one of them as the final weight parameter, and run the algorithm again to get the other two weights. The parameters satisfy that $x + y + z = 1, x \geq 0, y \geq 0, z \geq 0$.

After the weight parameters are optimized, the final overall word similarity is decides as:

$$sim(ws, ws') = x \times sim_{en} + y \times sim_{vec} + z \times sim_{how}, \tag{2}$$

where ws and ws' are two senses, sim_{en} is the word similarity based on English word embeddings and knowledge base, sim_{vec} is the word similarity based on Chinese word embeddings, sim_{how} is the word similarity based on HowNet, their optimized weight parameters are x, y, z, respectively.

3.7 Disambiguation graph construction

In order to construct the disambiguation graph, we take word senses as nodes, semantic relationships as edges, the overall word similarities as the weights of edges.

As we utilizes Chinese and English knowledge resources, the senses are represented with a triple, i.e., *Word(ID, Sword, Enword)*. *ID* is the ID of a sense or concept. *Sword* is the first primitive lexeme of concept definition in HowNet. *Enword* is its corresponding description in English, i.e., the mapping from Chinese to English. With the representation form of triples, we can easily integrate the three kinds of word similarity. For example, "中医" has two senses, which can be represented as "中医 (157329, 人, practitioner of Chinese medicine)" and "中医 (157332, 知识, traditional Chinese science)", whose word similarity with other senses can be computed with Eq. (2).

3.8 Graph scoring algorithm

PageRank algorithm is selected to evaluate and score the importance of each sense node in the disambiguation graph. If a sense node connects with more nodes with higher importance, its importance is more higher, which means that the sense is more related with the context words. With the algorithm iterations, the importance of each node is gradually differentiated. The sense node with maximum importance would be selected as the right sense of ambiguous word. The importance of each node is updated with the following equation:

$$PR(v) = \frac{1-\alpha}{N} + \alpha \sum_{in(v) \ni \mu} \frac{PR(v)}{|\operatorname{out}(\mu)|} \ , \tag{3}$$

where *v* refers to a sense node, α indicates the probability to continue the current Markov chain, 1-α indicates the probability of randomly selecting another node instead of continuing current Markov chain, N is the total numbers of sense nodes, *|out(u)|* refers to the out-degree of node *u*, *in(v)* indicates the set of all nodes that link with the node *v*.

4 Experiments

4.1 Data sets and evaluation measure

The benchmark dataset is SemEval task#5, i.e., multilingual Chinese English lexical sample task [Jin, Wu and Yu (2007)], which consists of 19 nouns and 21 verbs. Both training and test corpus are provides. The detail information of this dataset is shown in Tab. 1. Our proposed WSD method is unsupervised, which only utilizes test instances instead of training ones.

Table 1: Summary of SemEval dataset

Pos	#word-types	#average senses	#training instances	#test instances
Noun	19	2.58	1019	364
Verb	21	3.57	1667	571

Macro-average p_{mar} is selected to evaluate the performance of WSD methods, which is defined as:

$$p_{mar} = \sum_{i=1}^{N} \frac{P_i}{N}, P_i = \frac{m_i}{n_i} \ , \tag{4}$$

where N is the number of all word-types, m_i is the number of disambiguated correctly to one specific word-type, n_i is the number of all test instances of this word-type.

4.2 Baselines

Our proposed graph-based WSD with multi-knowledge integration is compared with four different types of representative unsupervised WSD methods, i.e., TorMd, HowGraph, EnGraph, SogouGraph [Jin, Wu and Yu (2007); Meng (2018)].

- **TorMd:** An unsupervised WSD method proposed by the University of Toronto, which wins the first place in this SemEval competition.
- **HowGraph:** This is an abridgment version of our proposed method, which only utilizes the word similarity based on HowNet.
- **EnGraph:** This also is an abridgment version of our proposed method, which only utilizes the word similarity based on English word embeddings and knowledge base.
- **SogouGraph:** This is another abridgment version of our proposed method, which only utilizes the word similarity based on Chinese word embeddings trained on Sogou corpus.
- **MultiGraph:** This is the full version of our proposed method, which integrates three kinds of word similarity. In the experiments, for word similarity based on HowNet, English word embeddings and knowledge base, Chinese word embeddings, the optimized weight parameters are 0.028, 0.336 and 0.636, respectively.

4.3 Results and analysis

4.3.1 Comparison of overall performances

The performance comparison of all methods is shown in Tab. 2. All the graph-based methods, including our proposed method and its three abridgment versions, outperforms TorMd method, which demonstrates that graph-based methods are potential to achieve a significant improvement.

MultiGraph shows 6.1%, 2.4%, 4.1% and 3.6% improvement over the TorMd, HowGraph, EnGraph, SogouGraph. The results demonstrate that utilizing single knowledge resource will hurt the performance and show that our proposed WSD framework integrating multiple resources is powerful. In order to achieve a satisfied performance, it is necessary to integrate as many as knowledge resources in graph-based WSD.

Table 2: Comparison of overall performances. Imp refers to the improvement made by MultiGraph over the corresponding method

	TorMd	HowGraph	EnGraph	SogouGraph	MultiGraph
P_{mar}	0.431	0.468	0.451	0.456	0.492
Imp(%)	6.1	2.4	4.1	3.6	-

4.4 Comparison of noun performances

The performance comparison on nouns is shown in Tab. 3. MultiGraph achieves the best performance, which demonstrates 4.5%, 4.4%, 3.5%, 4.3% improvement over TorMd, HowGraph, EnGraph and SogouGraph. As shown in Tab. 3, HowGraph, EnGraph and SogouGraph have different advantages on different words, while MultiGraph integrates their advantages so as to improve the noun performance greatly.

Table 3: Comparison of noun performances. For each word, its best performance is boldfaced

Nouns	#Senses	TorMd	HowGraph	EnGraph	SogouGraph	MultiGraph
本	3	0.72	0.64	**0.76**	0.36	0.64
表面	2	0.556	0.444	0.611	0.556	**0.667**
菜	2	0.474	0.684	0.579	0.684	**0.789**
长城	3	0.429	0.238	0.143	**0.476**	0.381
单位	2	**0.706**	0.529	0.529	0.529	0.588
道	3	0.5	0.5	**0.556**	0.5	**0.556**
队伍	3	0.318	0.5	0.409	0.455	**0.545**
儿女	2	0.5	0.65	0.5	**0.7**	0.5
机组	2	0.643	0.571	0.714	0.929	**1**
镜头	2	0.467	0.467	**0.6**	**0.6**	**0.6**
面	3	0.348	0.348	0.609	0.13	**0.652**
牌子	2	0.353	0.353	**0.588**	0.353	0.294
旗帜	3	0.5	0.556	**0.611**	0.389	0.556
气息	2	**0.857**	0.5	0.571	0.571	0.5
气象	2	0.438	0.562	0.375	**0.688**	0.625
日子	3	0.281	**0.531**	0.469	0.469	0.375
天地	3	0.56	**0.76**	0.4	0.6	0.4
眼光	2	**0.714**	0.357	0.286	0.357	0.357
中医	2	0.438	0.625	**0.688**	0.5	0.625
P_{mar}		0.516	0.517	0.526	0.518	**0.561**
Imp(%)		4.5	4.4	3.5	4.3	-

4.5 Comparison of verb performances

The performance comparison on verbs is shown in Tab. 4. MultiGraph still achieves the best performance, which demonstrates 9.4%, 0.5%, 4.6%, 3.0% improvement over TorMd, HowGraph, EnGraph and SogouGraph. The significant improvements demonstrate that the effectiveness of our proposed graph-based WSD method with multi-knowledge integration, i.e., MultiGraph.

Table 4: Comparison of verb performances

Verbs	#Senses	TorMd	HowGraph	EnGraph	SogouGraph	MultiGraph
补	3	**0.55**	0.45	0.5	0.25	0.3
成立	3	0.481	0.556	0.259	**0.519**	0.37
吃	4	0.174	0.478	0.13	**0.478**	**0.478**
出	9	0.169	0.221	0.091	**0.26**	0.169
带	8	0.119	0.194	0.284	0.194	**0.299**
动	4	0.3	**0.45**	0.2	0.2	0.2
动摇	2	0.5	0.25	**0.625**	0.5	**0.625**
发	5	0.25	**0.5**	0.361	0.222	0.222
赶	3	0.389	0.389	0.333	0.389	**0.444**
叫	4	0.256	0.333	0.333	**0.487**	0.462
进	5	0.25	0.136	0.159	**0.318**	0.273
开通	2	**0.5**	**0.5**	**0.5**	**0.5**	**0.5**
看	4	0.294	0.412	0.265	**0.471**	0.441
平息	2	0.375	0.375	**0.5**	**0.5**	0.5
使	2	0.563	0.562	**0.625**	0.375	0.438
说明	2	0.444	**0.611**	0.556	0.333	0.389
挑	2	0.143	**0.5**	0.357	0.357	0.429
推翻	2	0.3	0.7	**0.8**	0.6	**0.8**
望	2	0.462	0.615	0.231	0.615	**0.692**
想	4	0.216	0.243	0.216	0.243	**0.27**
震惊	2	**0.714**	0.429	**0.714**	0.571	**0.714**
P_{mar}		0.335	0.424	0.383	0.399	**0.429**
Imp(%)		9.4	0.5	4.6	3.0	0

5 Conclusion

This work proposes a novel graph-based Chinese WSD method with multi-knowledge integration. Different from the existing knowledge-based methods, our methods utilize Chinese and English semantic knowledge simultaneously to disambiguation words. Three different kinds of word similarity from various knowledge resources are optimized with simulated annealing algorithm, and integrated to compute an overall similarity. With sense as nodes, semantic relations as edges and the overall similarities as weights of edges, the disambiguation graph is constructed, which is evaluated with graph scoring algorithm to select the right senses. Extensive experiments on SemEval dataset shows that the proposed method significantly outperforms four baselines. In this work, we only use three kind of knowledge resources, which only makes a trial to integrate multi-language knowledge resources. Our future work is to find more semantic resources and design more sophisticated integration methods for graph-based WSD.

References

Agirre, E.; Soroa, A. (2009): Personalizing pagerank for word sense disambiguation. *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 33-41.

Camacho-Collados, J.; Pilehvar, M. T.; Collier, N.; Navigli, R. (2017): Semeval-2017 task 2: multilingual and cross-lingual semantic word similarity. *Proceedings of the 11th International Workshop on Semantic Evaluation*, pp. 15-26.

Chen, X.; Liu, Z.; Sun, M. (2014): A unified model for word sense representation and disambiguation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pp. 1025-1035.

Dongsuk, O.; Kwon, S.; Kim, K.; Ko, Y. (2018): Word sense disambiguation based on word similarity calculation using word vector representation from a knowledge-based graph. *Proceedings of the 27th International Conference on Computational Linguistics*, pp. 2704-2714.

Duque, A.; Stevenson, M.; Martinez-Romo, J.; Araujo, L. (2018): Co-occurrence graphs for word sense disambiguation in the biomedical domain. *Artificial Intelligence in Medicine*, vol. 87, pp. 9-19.

Jin, P.; Wu, Y.; Yu, S. (2007): Semeval-2007 task 05: multilingual chinese-english lexical sample. *Proceedings of the 4th International Workshop on Semantic Evaluations*, pp. 19-23.

Ke, K. (2011): *Collins Cobuild Advanced Learner's English-Chinese Dictionary*. Beijing: Foreign Language Teaching and Research Press & Harper Collins Publishers Ltd.

Le, Q.; Mikolov, T. (2014): Distributed representations of sentences and documents. *International Conference on Machine Learning*, pp. 1188-1196.

Lu, W. (2018): Word sense disambiguation based on dependency constraint knowledge. *Cluster Computing*, pp. 1-9.

Lu, W.; Huang, H.; Wu, H. (2013): Word sense disambiguation based on dependency fitness with automatic knowledge acquisition. *Ruanjian Xuebao/Journal of Software*, vol. 24, no. 10, pp. 2300-2311.

Lu, W.; Huang, H.; Wu, H. (2014): Word sense disambiguation with graph model based on domain knowledge. *Acta Automatica Sinica*, vol. 40, no. 12, pp. 2836-2850.

Lu, W.; Wu, H.; Jian, P.; Huang, Y.; Huang, H. (2018): An empirical study of classifier combination based word sense disambiguation. *IEICE Transactions on Information and Systems*, vol. 101, no. 1, pp. 225-233.

Mafarja, M. M.; Mirjalili, S. (2017): Hybrid whale optimization algorithm with simulated annealing for feature selection. *Neurocomputing*, vol. 260, pp. 302-312.

Mamano, N.; Hayes, W. B. (2017): Sana: simulated annealing far outperforms many other search algorithms for biological network alignment. *Bioinformatics*, vol. 33, no. 14, pp. 2156-2164.

McKeown, K.; Galley, M. (2003): Improving word sense disambiguation in lexical chaining. *Proceedings of 18th International Joint Conference on Artificial Intelligence*, 1486-1488.

Meng, F. (2018): *Research on Chinese Word Sense Disambiguation Method Based on Graph Model (Ph. D. Thesis).* Qilu University of Technology.

Meng, F.; Lu, W.; Xue, R. (2017): Mapping senses in babelnet to chinese based on word embedding. *10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics*, pp. 1-6.

Meng, F.; Lu, W.; Zhang, X.; Cheng, J. (2018): Word sense disambiguation method based on hownet and graph model. *Journal of Qilu University of Technology*, vol. 32, no. 6, pp. 66-73.

Meng, F.; Lu, W.; Zhang, Y.; Jian, P.; Shi, S. et al. (2017): Qlut at semeval-2017 task 2: word similarity based on word embedding and knowledge base. *Proceedings of the 11th International Workshop on Semantic Evaluation*, pp. 235-238.

Mihalcea, R. (2004): Co-training and self-training for word sense disambiguation. *Proceedings of the Eighth Conference on Computational Natural Language Learning.*

Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G. S.; Dean, J. (2013): Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems*, pp. 3111-3119.

Miller, G. A. (1995): Wordnet: a lexical database for English. *Communications of the ACM*, vol. 38, no. 11, pp. 39-41.

Moro, A.; Navigli, R. (2015): Semeval-2015 task 13: multilingual all-words sense disambiguation and entity linking. *Proceedings of the 9th International Workshop on Semantic Evaluation*, pp. 288-297.

Navigli, R.; Lapata, M. (2010): An experimental study of graph connectivity for unsupervised word sense disambiguation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 4, pp. 678-692.

Navigli, R.; Ponzetto, S. P. (2012): Babelnet: the automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial Intelligence*, vol. 193, pp. 217-250.

Navigli, R.; Velardi, P. (2005): Structural semantic interconnections: a knowledge-based approach to word sense disambiguation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 7, pp. 1075-1086.

Panchenko, A.; Ruppert, E.; Faralli, S.; Ponzetto, S. P.; Biemann, C. (2017): Unsupervised does not mean uninterpretable: The case for word sense induction and disambiguation. *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, vol. 1, pp. 86-98.

Parker, R.; Graff, D.; Kong, J.; Chen, K.; Maeda, K. (2011): English Gigaword fifth edition. Linguistic Data Consortium. https://catalog.ldc.upenn.edu/LDC2011T07.

Qun, L.; Sujian, L. (2002): Word similarity computing based on hownet. *Computational Linguistics Chinese Language Processing*, vol. 7, no. 2, pp. 59-76.

Raganato, A.; Bovi, C. D.; Navigli, R. (2017): Neural sequence learning models for word sense disambiguation. *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pp. 1156-1167.

Raganato, A.; Camacho-Collados, J.; Navigli, R. (2017): Word sense disambiguation: A unified evaluation framework and empirical comparison. *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, vol. 1, pp. 99-110.

Subramanian, S.; Trischler, A.; Bengio, Y.; Pal, C. J. (2018): Learning general purpose distributed sentence representations via large scale multi-task learning. arXiv:1804.00079.

Xiang, L.; Li, Y.; Hao, W.; Yang, P.; Shen, X. (2018): Reversible natural language watermarking using synonym substitution and arithmetic coding. *Computers, Materials & Continua*, vol. 55, no. 3, pp. 541-559.

Yang, Z. Z.; Huang, H. Y. (2012): Graph based word sense disambiguation method using distance between words. *Ruanjian Xuebao/Journal of Software*, vol. 23, no. 4, pp. 776-785.

Zhang, Y.; Lu, W.; Ou, W.; Zhang, G.; Zhang, X. et al. (2019): Chinese medical question answer selection via hybrid models based on CNN and GRU. *Multimedia Tools and Applications*.

Zhendong, D.; Qiang, D. (2006): Hownet and the Computation of Meaning (With CD-ROM). *World Scientific.*