Applying Convolutional Neural Network Model and Auto-expanded Corpus to Biomedical Abbreviation Disambiguation

Ren Kai[1,2,*] and Wang Shi-Wen[3]

[1] Computer School, Wuhan University, Wuhan, 430072, China
[2] College of Computer Science, South-Central University for Nationalities, Wuhan, China
[3] Confucius Institute, New Jersey City University, 2039 Kennedy Blvd, Jersey City, NJ, 07305, United States

___

Abstract

The polysemy phenomenon of abbreviations in the medical domain generates a prodigious effect on the accuracy of computer auto text analysis. Hence, abbreviation disambiguation has been extensively studied in recent years. A large quantity of manually labelled corpuses is required in existing methods for training models, thereby restricting the application range of abbreviation disambiguation. This study proposes an abbreviation disambiguation method based on the convolutional neural network (CNN) to solve the abbreviation disambiguation problem in the biomedical field when no labelled corpus exists. First, the full name of the ambiguous abbreviation was taken as the keyword to obtain a large quantity of texts on Medline as the training corpus. The corpus was then applied to the improved CNN model, through which each abbreviation was mapped onto the corresponding sense to complete the abbreviation disambiguation. A test was conducted on 103 common biomedical abbreviations. Results show that the method obtained an average of 90.1% accuracy, which is significantly higher than the other unsupervised abbreviation disambiguation methods. This study provides a basis for effectively improving the accuracy of abbreviation disambiguation in the biomedical field without a large labelled corpus and for increasing the accuracy of follow-up work, such as information retrieval and relation extraction. Thus, the proposed method can be applied to computer analytical research on real-time updated medical big data.

Keywords: Abbreviation disambiguation, Convolutional Neural Network, Medline

___

1. Introduction

Mass data and literature related to biological information are available on the Internet. Doctors expect to apply large amounts of data information to provide auxiliary guidance in clinical treatment to a large extent; as such, doctors have an urgent demand for biomedical information technology. Bioinformatics attempts to use technologies related to computers and informatics for the auto excavation of effective information from mass information to provide assistance and guidance for clinical medicine.

In bioinformatics texts, the "Synonym" and "Homograph" phenomena are relatively common because of the particularity of the biomedical field. This problem is divided into the word sense disambiguation problem in the natural language processing field. A particular polysemy phenomenon called abbreviation polysemy exists besides the ordinary polysemy of medical terms. Many important concepts and terms are represented by abbreviations in biomedical texts. For example, AA can represent amino acids or alcoholics anonymous in biomedical texts. Furthermore, RA has three extended forms: Refractory anemia, radium and rheumatoid arthritis. The accurate recognition of abbreviation sense is largely significant to understand and analyze biomedical data.

The full-name expanded form of an abbreviation in biomedical texts is labeled when it is utilized in an article for the first time. The abbreviation is then directly applied in subsequent instances. Many of these abbreviations are relatively explicit in a specific context because of the particularity of the biomedical field, and the author directly utilizes abbreviations without providing their expanded forms. Thus, two problems in the computer recognition of abbreviations in the biomedical field emerge. First, the abbreviation contained in the study to be analyzed does not appear in the study for the first time, so it does not include its expanded form and the sense is uncertain. Second, the author believes that the sense of abbreviation can be directly inferred according to the context of the text to be analyzed, but abbreviation sense is uncertain for the computer. The two problems largely affect the accurate recognition of abbreviations in the biomedical field because conventional computer auto text processing does not have the medical science background and common sense of ordinary readers. Computer recognition also exhibits many limitations in sensing the disambiguation of abbreviations.

Among the past cases in abbreviation disambiguation in the biomedical field, most of the adopted supervised learning methods require many manually annotated corpuses, which have a relatively favorable effect. Manual corpus annotation has high cost and low efficiency, so applying this method has several drawbacks. The present study utilizes a method

___

for the acquisition of training corpus on Medline through the expanded full-name forms of abbreviations together with a method based on the convolutional neural network (CNN) to conduct abbreviation disambiguation. The proposed method attempts to complete abbreviation disambiguation work that is independent of many manually annotated corpuses.

2. Related work

Word sense disambiguation work generally requires a certain amount of manually labeled corpuses to be utilized for training models. However, manually labeled corpus is usually difficult to obtain and consume a lot of manpower and time. For this problem, several methods attempt to supplement data utilized for the training model with many unlabeled external resources. Although these methods are generally inferior to the supervised methods of the effects, they do not require manual annotation or only require minimal manually labeled data to complete the task of word sense disambiguation. Thus, these methods have high values in practical application.

Representative methods utilizing external resources are discussed as follows. Navigli et al. [1] utilized external resources of page classification in Wikipedia, which were applied as features to improve word sense disambiguation effects. Wikipedia, as online encyclopedia, covers a wide range of information, but it still encounters the problem when some professional detailed knowledge is not covered by the biomedical field, which affects its application. McInnes et al. [2] concentrated more on the specialty of biomedicine; their word sense disambiguation system applied the Unified Medical Language System (UMLS) as an external resource to improve the accuracy of word sense disambiguation in the biomedical field. The said study reached advanced levels among the past study achievements. Jimeno-Yepes et al. [3][4] constructed a standard corpus dataset biomedical field word sense disambiguation dataset(MSH-WSD)), so that a unified standard was utilized to evaluate subsequent study. In their representative work from follow-up research on word sense disambiguation based on external knowledge, the said researchers adopted Automatic Extracted Corpus (AEC) and Machine Readable Dictionary (MRD) to make relevant word sense disambiguation work by combining collocation features and assess the MSH-WSD dataset. AEC and MRD were also adopted as baseline approaches to compare with the proposed method in our study.

Chasin et al. [5] compared a graph-based algorithm with the unsupervised word sense disambiguation method based on the topic model; which was also a typical concept of the unsupervised method. Agirre et al. [6] presented a graph-based method constructed utilizing ontological knowledge—WordNet. This method did not provide special attention to features of the biomedical field, and its effect was inferior to those of previous methods specially optimized for this field. However, this universal method is noteworthy for reference and learning. Pedersen [7] compared several high-order matrixes to represent vectors and performed word sense disambiguation work, which was realized utilizing a Sensecluster tool. Thus, the said researcher exploited the idea for unsupervised word sense disambiguation. The results of the Sensecluster method are reported in the current study and compared with those of our proposed method. The

aforementioned methods were basically word sense disambiguation methods transited from the traditional field, and no special optimization was implemented by directing at the features of ambiguous terms in the biomedical field. A specialized exploration on unsupervised word sense disambiguation for ambiguous terms in the medical field was conducted at an early stage; the unsupervised method based on kernel function ambiguity C-means clustering was utilized at the early-stage study [8] on perform the disambiguation of ambiguous terms in the biomedical field. Given that this method does not have especial advantages in abbreviation disambiguation, no comparison with this method was particularly conducted in the current study.

Okazaki et al. [9] conducted relevant work on the correlation between abbreviations in the biomedical field and their expanded forms to facilitate follow-up word sense disambiguation work; their work provided the idea for us to extract learning corpuses. The complete expanded forms of abbreviations were utilized as keywords to extract many training corpuses on Medline as part of training corpuses in the later word embedding and initial training corpuses of CNN.

The rest of this study is organized as follows. Section 3 introduces the proposed method in detail and discusses the acquisition of the online training corpus, word feature training based on Word2vec, and how CNN is utilized to conduct word sense disambiguation work. Section 4 discusses the experiment and analysis of the comparison results. Section 5 presents the conclusion of this study.

3. Methodology

3.1 Extraction of training corpus

A total of 103 abbreviations commonly used in the biomedical field and included in the MSH-WSD corpus were selected. MSH-WSD is a manually labeled corpus, in which the sense of each abbreviation included approximately 100 contexts that contain ambiguous terms. This labeled corpus was selected as our test set. A specialized program was compiled for abbreviations in the biomedical field to construct the training set. The first 5,000 search results that took each abbreviation and all expanded forms as keywords were automatically extracted from Medline. This corpus was processed into each abbreviation form, and each expanded form independently had one text file. A total of 103 abbreviations and 1.26 GB of raw text were obtained. These grabbed data were utilized as the training corpus in the subsequent word embedding and CNN model training.

Our corpus extraction process is shown in Fig. 1. First, the expanded full-name forms of the 103 ambiguous abbreviations were acquired by inquiring through the Medical Subject Headings. These ambiguous abbreviations and their expanded full names were taken as the keywords to obtain 5,000 titles and abstracts on Medline. These texts were then combined as the training corpus of the word2vec toolkit. For the inquiry results of the 103 expanded full names, the sentences where the expanded forms of the keyword were located were extracted for each result. The inquiry result of each expanded keyword was saved line by line. This work generated 103 independent text files, which were utilized as the training and development sets in the later-stage training of the CNN sense disambiguation model.
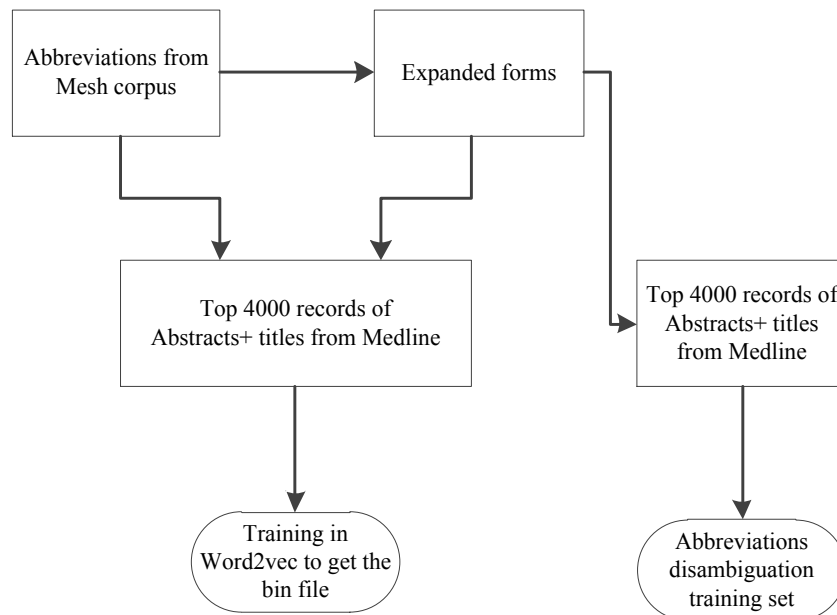
Fig. 1. Preparation of the training set and Word2vec training data

3.2 Word Embedding

With the development of semantic representation technology, the present word vector employed by semantic representation is no longer the traditional One-Hot Representation type of word vector, but the Distributed Representation type of word vector.

Building a vector in the distributed semantic representation model is necessary to represent context information, and context is generally determined through the context window or N-gram form. Context information is commonly represented in the frequency manner, such as the commonly applied term frequency–inverse document frequency method. Its basic idea is to map each word by training the K-dimensional vector (K is generally the hyper-parameter in the model), and their semantic similarity is judged through the distance between words (such as cosine similarity and Euclidean distance). The Word2vec word vector utilizes a three-layer neural network composed of an input layer, a hidden layer, and an output layer. Its core view is to utilize Huffman coding to encode all occurring words by word frequency to make the contents of all words of similar word frequency activated in the hidden layer basically identical. Thus, a fewer number of activated hidden layers are observed for words with a higher occurrence frequency, which can effectively decrease the computing complicacy. The reason for the extensive application of Word2vec is its high efficiency, as stated in the in-depth research of Mikolov [10].

Obtaining representation of one word in the vector space through the three layers of the neural network is the goal of utilizing Word2vec technology. Unlike the classical processes of Latent Semantic Index and Latent Dirichlet Allocation (LDA), Word2vec employs the context of the word, so the semantic information is richer.

Word2Vec is composed of two different methods, namely, CBOW and Skip-gram. The objective of CBOM is to predict the probability of the present work according to the context. By contrast, Skip-gram predicts the probability of the context according to the present word (Fig. 2). Both methods utilize the artificial neural network as their classification algorithm. Each word is initially a random N-dimensional vector. The algorithm then utilizes CBOW or

the Skip-gram method to obtain the optimal vector of each word after training.

Given that the distributed representation mode is the neural network distributed representation, it is generally referred to as Word Embedding. This technology implements modeling according to the complicated relationship between the context and target word through the neural network. The neural network itself has multiple categories and parameter variation. When the neural network makes a representation, it can perform modeling of the complicated context, including richer semantic features. With the increase in representation length, the parameters only increase linearly and not at the geometric level.

When performing word embedding, a special consideration was given to the particularity of the abbreviation corpus. The expanded form of the abbreviation generally contained the abbreviation itself, which was also added to our word embedding features.
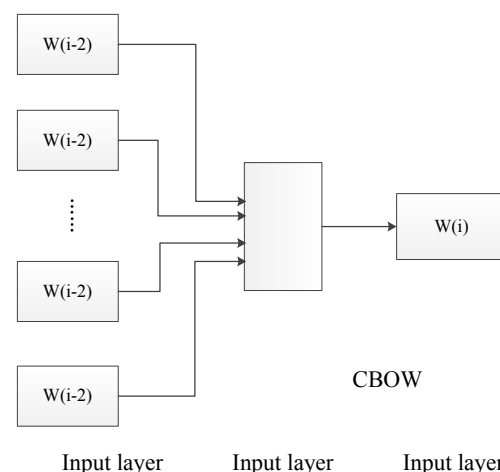

Fig. 2. Continuous Bag of Words (CBOW) method of the Word2vec model

3.3 CNN

CNN was proposed by Fukushima [11] in 1980 at the earliest, and its core ideas are local perception and weight sharing. Each node in the hidden layer of the general feed forward neural network was in full joint with the nodes in

the input layer. By contrast, each node in the hidden layer of CNN is only jointed with a region of a fixed size (window size) in the input layer. From the fixed region to the sub-network of the hidden layer, all regions in the input layer were underweight sharing. The formula from the input layer to the hidden layer was formalized as Eq. (1). For the semantic vector representation method of the words and documents based on the neural network, CNN generally utilized pooling technology after several hidden layers to compress the hidden layers of uncertain lengths into hidden layers of fixed lengths. Mean-pooling and max-pooling [12] were employed in the current study, whose formula is expressed in Eq. (3):

$$x_i = \left\lfloor e(\omega_{i-\lfloor \omega in/2 \rfloor}) \ldots e(\omega_i) \ldots e(\omega_{i+\lfloor \omega in/2 \rfloor}) \right\rfloor \qquad (1)$$

$$h_i^{(1)} = \tanh(W_{x_i} + b) \qquad (2)$$

$$h^{(2)} = \max_{i=1}^{n} h_i^{(1)} \qquad (3)$$

CNN can model the local information of each part in the text through its convolution kernel. This network can integrate the full-context word sense from different local information sources through its pooling layer, and the overall complicacy of the model is O (n). CNN has relatively extensive applications. Collobert [13] applied CNN to the process semantic role labeling task in the natural language field for the first time; the system performance effectively improved. Kalchbrenner [14] et al. and Kim [15] published their respective theses in 2014 that apply CNN to perform

text classification. Zeng et al. [16] proposed utilizing CNN to perform the relation classification task and obtained a favorable effect.

Based on the work conducted by Kim, improved CNN was employed in the current study to perform semantic classification work and realize word sense disambiguation. The sentence length decided the vector length. The input structure is shown in Fig. 3, and the convolutional layer of CNN is expressed in Eq. (2). A window parameter h existed, and then consecutive h words in each group constituted a convolutional layer feature. One feature matrix was finally obtained by combining all features. Max-over-time pooling operation, which is commonly described as max-pooling, was adopted in the next layer as expressed in Eq. (3). Max-pooling selected the maximum eigenvalue as the eigenvalue in the said figure. An abstract understanding of the core idea could be that the most marked value would be selected to represent the feature of each convolutional layer. This pooling operation basically solved the problem of non-uniform sentence lengths, and then the output lengths could be normalized.

In the improved CNN model in this study, each filter generated one feature, and the quantity of the model filters was decided by the selected window length. Multi-module filters corresponded to the multiple features during the feature extraction process from the filters. These features were outputted through the fully connected softmax layer from the second layer. The classification results were then obtained. This model would finally realize a multi-label text classification system based on the full-text features.
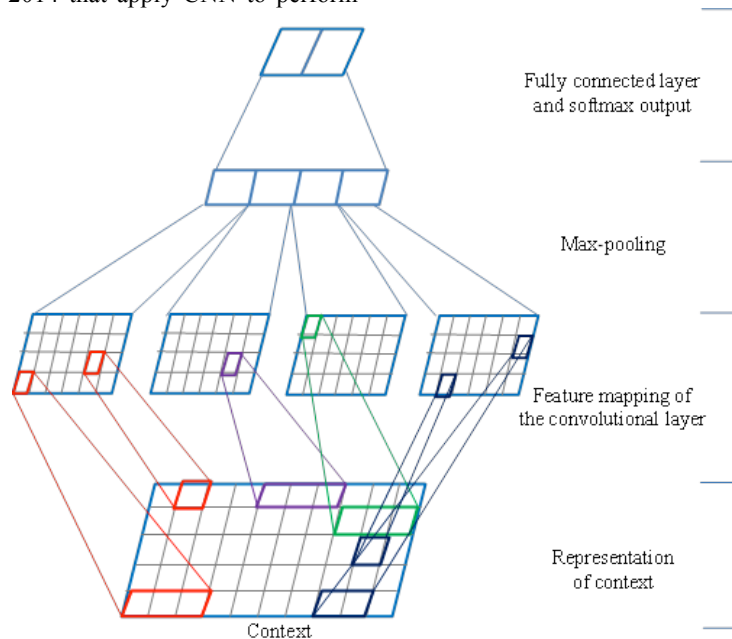

Fig. 3. CNN model for context classification

3.4 Abbreviation disambiguation

Word sense disambiguation is generally based on clustering or the classification method. Nearly all machine learning methods can be currently used for text classification. Similar to clustering and LDA, the class name and relationship between classes cannot be directly described after clustering is completed. This indeterminacy can cause difficulties in the usage and optimization of the WSD. A significant advantage of the classification method over the topic model

or clustering method is that the category system is determinate.

As a rule, our training data were manually labeled training data. Manually labeled corpuses have high accuracy but also a large workload cost. Unlabeled data were utilized and semi-supervised learning was proposed to lower the annotation cost. A semi-supervised algorithm mainly considers how to employ a small quantity of labeled samples and a large quantity of non-labeled samples to perform

training and classification. A common method was the self-learning method.

The self-learning algorithm was executed as follows:
Input: Two sample sets were and unlabeled.
Output: labeled data sets
Step 1) The labeled sample set was adopted to generate the classification strategy F.
Step 2) F was applied to classify the unlabeled samples and calculate errors.
Step 3) Subset u with a small error in the unlabeled sample set was selected and added to the labeled sample set.
Step 4) The aforementioned steps were repeated.

The basic idea of this algorithm is as follows: when classifier was being constructed, several features were first selected for each category. Large-scale unlabeled texts were then classified according to these features. More feature words were extracted from the text, which definitely belonged to this category and added to the original feature word list. Classification was then supplemented by a certain artificial verification.

The proposed method, which refers to the self-learning idea, utilized the expanded form of abbreviations as extra features of the word sense classification. The overall workflow is shown in Fig. 4. Given that the quantity of labeled training corpuses was limited, supplemented auto-labeled corpuses were required to complete this work and better meet the CNN requirements. Multiple expanded forms that automatically matched for abbreviations were taken as keywords, and many training corpuses were grabbed on Medline. The CNN classifier was trained through this corpus, which was employed to realize the word sense disambiguation. Abbreviation disambiguation could be realized in the biomedical field in this manner when no manually labeled corpus was available.
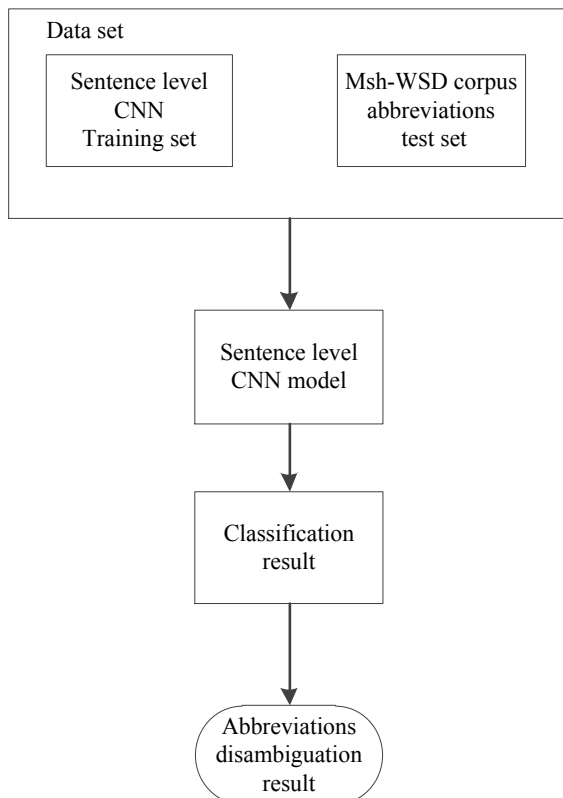

Fig. 4. Workflow of the CNN abbreviation disambiguation model

4 Experiments

4.1 Experimental environment
OS： Ubuntu 14.04LTS
CPU: Intel Core i7-6700k @4.0GHz * 8
RAM: 16GB 2400 MHz DDR4
GPU： NVIDIA GeForce GTX 960 4G
Python：2.7.6
Theano：0.7.0
CUDA：7.5.17

4.2 Experimental corpuses

The MSH-WSD of NLM included 106 abbreviations, 88 ordinary ambiguous terms, and 9 ambiguous words of the mixed type. A total of 106 abbreviations in this corpus were adopted in the present study to conduct the experiment. The corpus in this study has two parts. First, each ambiguous term was utilized as the keyword (appearing in the title and abstract) through the Medline interface provided by NLM. Full-text records of the titles and abstracts of the first 4,000 relevant results (if less than 4,000 retrieval results were available, then all retrieved titles and abstract records were utilized) were obtained from Medline. These acquired texts had a total size of 800MB. Second, the expanded forms of the 106 abbreviations were utilized as the keywords (appearing in titles and abstracts). Full-text records of the headings and abstracts of the first 4,000 relevant texts from Medline were obtained (if less than 4,000 retrieval results were available, then all retrieved headings and abstract records were utilized). This part had 1.09 GB of raw texts. The texts in the first and second parts were combined. The relevant representation vector (vector.bin) files were generated with the Word2Vec toolkit and utilized for the later CNN training and learning. The corpuses in the second part were then utilized as the training sets of the next step of the CNN algorithm and as the training corpus of the CNN classified disambiguation system.

In the Word2Vec training, the parameters adopted in this study were as follows: cbow 1 -size 200 -window 8 -negative 25 -hs 0 -sample 1e-4 -threads 20 -binary 1 -iter 15. One vector.bin file was obtained by utilizing the Word2Vec toolkit, and this file was applied in the subsequent representation of the input word vector.

4.3 Experimental results
Extracted training corpus, together with the 106 abbreviation datasets applied by CNN in the MSH-WSD, was adopted to conduct the word sense disambiguation test of abbreviations. Obtained results are listed in the following tables:

Tab. 1 shows that unlike several other common supervised abbreviation disambiguation methods, the proposed method without a manually labeled corpus was only slightly lower than AEC method. However, its accuracy was significantly higher than those of the MRD, 2-MRD, and SenseRelate methods based on UMLS. Given that our method did not require a manually labeled corpus, expanded and automatically generated learning corpuses were directly utilized to perform training. The applicability of the method in this study was improved compared with the previous methods. No other unsupervised methods or semi-supervised methods have independently tested this part (abbreviation ambiguous terms) in recent studies, so all of our comparisons were conducted with supervised methods. On the condition that no labeled corpus as utilized, the proposed

method obtained a 90.02% accuracy, which was superior to that of most common supervised word sense disambiguation methods.

For further comparison, a group of comparative experiments was conducted between the proposed method and supervised methods with the same CNN model. The obtained results are listed in Tab. 2.

The comparative experiment shows that the results of the supervised CNN word sense disambiguation method, which utilized manually labeled corpuses, was 94.91%. As expected, the accuracy of the CNN word sense disambiguation method, which adopted manually labeled training sets, relatively improved unlike that of the CNN method without manually labeled dataset. This result can be attributed to the natural difference between supervised learning and unsupervised learning. Our method without a

labeled training set obtained a 90.02% result, which has reached the application standard in practical engineering.

Fig. 5 shows that the average accuracy of our method was lower than that of the supervised CNN method by approximately 5%. The reasons for the difference were analyzed as follows. First, no labeled corpus was employed, so several effective features were lost. Second, the expanded forms of abbreviations were adopted as keywords to extract the training set. The senses of several abbreviations and their expanded forms were relatively similar, so our training set was did not have a high distinction degree. Third, the expanded forms of several abbreviations extracted a small scale of corpus from Medline, which directly resulted in an unbalance in the training data. The training effect was largely affected.

Table 1. Disambiguation results of the CNN method and other supervised methods

Methods	AEC	MRD	2-MRD	UMLS SenseRelate	CNN Without annotated data
Abbreviation set (accuracy %)	90.90	87.59	85.01	83	90.02

Table 2. Comparison of the CNN-based supervised model and without human annotated data

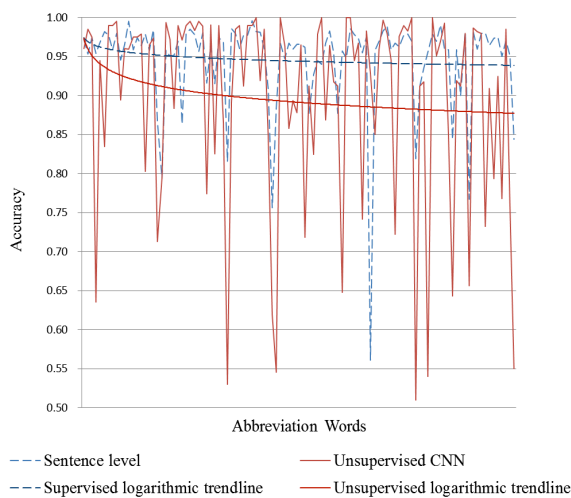| Methods	CNN sentence S200	CNN Fulltext S200	CNN sentence S200 without annotated data
Abbreviation set (accuracy %)	94.91	96.40	90.02

Fig. 5. Accuracy analysis of the supervised and unsupervised methods

5. Conclusions

An abbreviation disambiguation method without any need of manually labeled corpus was proposed in this study to solve the abbreviation ambiguity problem in mass unlabeled biomedical texts. The training corpus for abbreviation disambiguation was automatically extracted, and this corpus trained the CNN model and completed the abbreviation

disambiguation work. The following conclusions were drawn:

(1) The training corpus extracted from Medline according to the expanded form of the abbreviation could be the training corpus to be applied to the word sense disambiguation of this abbreviation. Its accuracy even exceeded that of most traditional supervised methods. Results indicate that the abbreviation had many context features similar to those of the corresponding expanded full name. These features could improve the accuracy of word sense disambiguation.

(2) CNN had a suitable effect on discovering the hidden features of the contexts. Experimental results indicate that under the same conditions, the accuracy of the neural network method was evidently higher than those of other traditional methods. These results verify that CNN could sufficiently discover hidden features in the contexts.

An externally acquired auto-labeled corpus was mainly employed in this study to train the CNN model and realize the abbreviation disambiguation work. A new feasible approach was probed for abbreviation disambiguation without a labeled corpus. We will attempt to optimize and improve the neural network structure and further improve the accuracy of abbreviation disambiguation in the biomedical domain in subsequent research work.

References

1   Navigli, R., Faralli, S., Soroa, A., de Lacalle O., Agirre, E., "Two birds with one stone: learning semantic models for text categorization and word sense disambiguation". In: *Proceedings of the 20th ACM international conference on Information and knowledge management,* Glasgow, United Kingdom: ACM, 2011, pp: 2317–2320.

2. McInnes, B. T., Pedersen, T., Liu, Y., Melton, G. B., Pakhomov, S. V., "Knowledge-based method for determining the meaning of ambiguous biomedical terms using information content measures of similarity". In: *Proceedings of the American Medical Informatics Association Annual Symposium,* Washington, DC, USA: AMIA, 2011, pp: 895-904.

3. Yepes, A. J., McInnes, B. T., Aronson, A. R., "Exploiting mesh indexing in Medline to generate a data set for word sense disambiguation". *BMC bioinformatics*, 12 (1), 2011, pp: 223-223.

4. Yepes, A. J., McInnes, B. T., Aronson, A. R., "Collocation analysis for UMLS knowledge-based word sense disambiguation". *BMC Bioinformatics,* 12(3), 2011, S4.pp:1471-2105.

5. Chasin, R., Rumshisky, A., Uzuner, O., Szolovits, P., "Word sense disambiguation in the clinical domain: a comparison of knowledge-rich and knowledge-poor unsupervised methods". *Journal of the American Medical Informatics Association*, 21 (5), 2014, pp: 842–849.

6. Agirre, E., Soroa, A., Stevenson, M., "Graph-based word sense disambiguation of biomedical documents". *Bioinformatics*, 26 (22), 2010, pp: 2889–2896.

7. Pedersen, T., "The effect of different context representations on word sense discrimination in biomedical texts", In: *Proceedings of the 1st ACM international health informatics symposium,* Arlington, VA, USA: ACM, 2010, pp: 56–65.

8. REN, K., REN, Y F., "Kernel Fuzzy C-Means Clustering for Word Sense Disambiguation in Biomedical Texts". *Journal of Digital Information Management*, 13(6), 2015, pp: 411-420

9. Okazaki, N., Ananiadou, S.,Tsujii, J., "Building a high-quality sense inventory for improved abbreviation disambiguation". *Bioinformatics* 26(9), 2010, pp: 1246-53.

10. Mikolov, T., Chen, K., Corrado, G., Dean, J., "Efficient estimation of word representations in vector space". *In: Proceedings of the International Conference on Learning Representations Workshop*, Scottsdale, Arizona, USA: ICLR, 2013, pp: 1-12.

11. Kunihiko, F., "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position". *Biological cybernetics*, 36(4), 1980, pp: 193–202.

12. Elman, J. L., "Finding structure in time". *Cognitive science*, 14(2), 1990, pp: 179–211

13. Collobert, Ronan, et al. "Natural language processing (almost) from scratch". *Journal of Machine Learning Research*, 12.Aug, 2011, pp: 2493-2537.

14. Kalchbrenner, N., Grefenstette, E., Blunsom, P., "A convolutional neural network for modelling sentences", In: *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, Baltimore, Maryland, USA: ACL, June. 2014, pp: 655–665.

15. Kim, Y., "Convolutional neural networks for sentence classification", In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, Doha, Qatar: ACL, October.2014, pp: 1746–1751.

16. Zeng, D., Liu, K., Lai, S., Zhou, G., & Zhao, J., "Relation classification via convolutional deep neural network", In: *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistic*, Dublin, Ireland: ACL, August.2014. pp: 2335-2344.