Supervised Learning and Knowledge-Based Approaches Applied to Biomedical Word Sense Disambiguation

Abstract:
Word sense disambiguation (WSD) is an important step in biomedical text mining, which is responsible for assigning an unequivocal concept to an ambiguous term, improving the accuracy of biomedical information extraction systems. In this work we followed supervised and knowledge-based disambiguation approaches, with the best results obtained by supervised means. In the supervised method we used bag-of-words as local features, and word embeddings as global features. In the knowledge-based method we combined word embeddings, concept textual definitions extracted from the UMLS database, and concept association values calculated from the MeSH co-occurrence counts from MEDLINE articles. Also, in the knowledge-based method, we tested different word embedding averaging functions to calculate the surrounding context vectors, with the goal to give more importance to closest words of the ambiguous term. The MSH WSD dataset, the most common dataset used for evaluating biomedical concept disambiguation, was used to evaluate our methods. We obtained a top accuracy of 95.6% by supervised means, while the best knowledge-based accuracy was 87.4%. Our results show that word embedding models improved the disambiguation accuracy, proving to be a powerful resource in the WSD task.

Keywords: Biomedical text mining, information extraction, word embeddings
DOI: 10.1515/jib-2017-0051
Received: August 17, 2017; Revised: September 9, 2017; Accepted: November 11, 2017

1 Introduction

Nowadays new biomedical entities such as proteins, genes, mutations, and diseases are constantly being discovered, which leads to the growth of the biomedical lexicon. However, a new discovered biomedical entity is not always assigned to a new biomedical term, causing the biomedical term to have several possible senses. These ambiguous terms hinder the automatic extraction of biomedical information. Word sense disambiguation (WSD) is responsible for solving these ambiguities in textual documents, having the automatic capability to assign the correct meaning to ambiguous words given the surrounding textual context. WSD is a challenging artificial intelligence problem that has been studied for the last years [1]. Particularly in the biomedical field, there are much more ambiguous terms increasing the WSD difficulty [2], [3]. For assessing the biomedical WSD systems, some datasets containing biomedical ambiguous terms were proposed [4], [5], being the currently most used the MSH WSD dataset that was proposed by Jimeno-Yepes et al. [6]. This last dataset was created by automatic means using the unified medical language system (UMLS) Metathesaurus [7] and the medical subject headings (MeSH) [8] indexing of MEDLINE articles.

In a previous work [9] we applied WSD supervised and knowledge-based methods in a subset of the MSH WSD dataset, achieving top accuracies of 94.7 and 85.1% respectively. In this paper we extended our previous work by (1) testing more supervised classifiers, (2) combining bag-of-words as local features and word embeddings as global features in the supervised approach, (3) using different word embedding averaging functions, in the knowledge-based method, to calculate the surrounding context vectors, (4) extracting concept definitions for every ambiguous term in the MSH WSD making possible to apply our knowledge-based method to all the ambiguous terms in the MSH WSD dataset.
2 Related Works

Schuemie et al. [2] makes an overview of WSD in the biomedical domain until 2005. In [10] the authors showed that metadata information and well structured ontologies can play an important role to improve disambiguation. In 2011, Jimeno-Yepes et al. [6] proposed the MSH WSD dataset achieving a supervised accuracy of 93.9 %, and a knowledge-based accuracy of 83.8 %. The results of the following works are with respect of this same dataset. Some knowledge-based WSD approaches use semantic similarity measures from the UMLS achieving accuracies of 80.7 % [11] and 75.0 % [12]. Another knowledge-based method uses the UMLS semantic network achieving an accuracy of 60.3 % [13]. McInnes and Stevenson [3] explored supervised and knowledge-based WSD methods achieving a supervised state-of-the-art accuracy around 97.0 % and a knowledge-based accuracy around 78.0 %. Their supervised system relies in a vector space model, calculating the cosine between the vector representing the ambiguous term and each of the vectors representing the possible senses. Another knowledge-based method is presented in [14] achieving an accuracy of 89.1 %, where the authors developed a method to generate word-concept probabilities from a knowledge-base. The state-of-the-art accuracy by knowledge-based means is 92.2 %, and it was obtained using a method proposed by Sabbir et al. [15] that uses neural concept embeddings. Another supervised state-of-the-art accuracy of 96.0 % was achieved using a combination of unigrams and word embeddings with a SVM classifier by Jimeno-Yepes [16].

As far as we know, Iacobacci et al. [17] were the first to weight word embeddings considering the absolute word distance between a specific word and the ambiguous term in the problem of disambiguation. Word embeddings are a recent technique that maps words to high dimensional numeric vectors, being generated from unlabeled training data [18]. These models have shown to improve text mining tasks, such as named entity recognition [19] and word sense disambiguation [20], [21].

3 Implementation

In this work we applied supervised and knowledge-based methods to biomedical WSD. Bag-of-words features were used only in the supervised setting. Although, in both approaches, we used word embedding models generated by unlabeled MEDLINE abstracts. These word embeddings were used to calculate embedding vectors of the surrounding contexts of the ambiguous terms, which we will denominate as context vectors or context embeddings. For the knowledge-based approach we extracted concept unique identifier (CUI) textual definitions from the UMLS Metathesaurus, and calculate CUI embedding vectors, which we will denominate as concept vectors or concept embeddings. To find the most plausible meaning for a specific ambiguous term, our knowledge-based method calculates similarities between the context vector and the concept vectors weighted by CUI-CUI association values. Each step is explained more detailed below.

3.1 Dataset

We evaluated our proposed methods in the MSH WSD dataset [6], which is the most adopted for assessing biomedical WSD methods. This dataset is composed by a total of 203 ambiguous entities, of which 88 are regular terms, 106 are abbreviations, and 9 are a mix of both. Most of the ambiguous entities have only two possible senses, where a minor part of 14 terms have from three up to five senses. For each possible sense there is a maximum of 100 instances. Each instance is a MEDLINE abstract where the ambiguous term occurs. The dataset has a total of 37,090 distinct MEDLINE abstracts.

3.2 Word Embeddings

The word embedding models were generated using PubMed articles which are biomedical domain-specific. MEDLINE abstracts corresponding to the years 1900 to 2015 were used, which contained around 15 million documents involving a total of around 800 thousand unique words. Six word embedding models were trained using windows of 5, 20, and 50 words, and vector sizes of 100 and 300. For generating the word embedding vectors we used the continuous-bag-of-words model proposed by Mikolov et al. [18], implemented in the Gensim framework [22].

The word embedding models were used to calculate the context embedding vectors and the CUI embedding vectors, with the last ones only being used in the knowledge-based approach.
3.3 Context Embeddings

The context embeddings are vectors that represent the surrounding contexts of the ambiguous terms. Each surrounding context of an ambiguous term is composed by the words of the respective abstract (excluding the ambiguous term occurrences) in the MSH WSD dataset. All the context vectors were weighted using the inverse document frequency (IDF) scheme and normalized.

3.3.1 Supervised

In the supervised setting the term frequency (TF) component was added to the IDF weighting. However, since the cross-validation technique was used, these TF-IDF weights were fitted to a linear regression (from the labels of the current training fold), estimating new weights for each word. These final weights were the ones used to weight the word embeddings in the test fold.

3.3.2 Knowledge-Based

In the knowledge-based method we tested five different word embedding averaging functions: the TF-IDF weighting scheme, and four word distance decay functions using also the IDF scheme. The objective of using decay functions was to give greater importance to closest words of the ambiguous terms. The absolute word distance \( d \) between some specific word and the closest occurrence of an ambiguous term was defined as being the input of the decay function. Summarily the weighting schemes used were (IDF weighting included):

- Term frequency;
- No decay: \( f(d) = 1 \);
- Fractional decay: \( f(d) = 1/d \);
- Exponential decay: \( f(d) = \exp(-d) \);
- Logarithmic decay: \( f(d) = 1/\ln(1 + d) \).

3.4 Supervised Learning Classification

We tested five machine learning classifiers from the scikit-learn framework [23]: decision tree (DT), k-nearest neighbor (k-NN, \( k = 5 \)), logistic regression (LR), multi-layer perceptron (MLP), and support vector machine (SVM). To train the classifiers, bag-of-words features (unigrams, bigrams) and the context embeddings were used.

3.5 Knowledge-Based Method

3.5.1 Concept Embeddings

CUI textual definitions were extracted from UMLS knowledge sources\(^1\) to create the concept embedding vectors weighted by the TF-IDF scheme. All the concept vectors were normalized.

3.5.2 CUI-CUI Association Values

We calculated CUI-CUI association values as normalized Pointwise Mutual Information (nPMI) from the MeSH co-occurrence counts in MEDLINE articles.\(^2\) The nPMI values are between 0 and 1, with 0 representing no association, and 1 a perfect association. So, a concept in relation to himself has a value of 1. Since there are many CUIs, and consequently much more CUI-CUI relations, we considered only the nPMI values greater or equal than 0.3.
3.5.3 Method

Our knowledge-based method came from the idea to compare the surrounding contexts of the ambiguous terms with the concept textual definitions, in order to find the most similar concept (meaning) given a specific context. With that in mind, we extended this baseline approach calculating a score for each possible CUI (meaning) of an ambiguous term as shown in equation (1).

\[
\text{score}(\text{CUI}) = \frac{1}{N} \sum_{j} \text{nPMI}(\text{CUI}, \text{CUI}_j) \cdot \text{CS}(t, \text{CUI}_j)
\]

(1)

Accordingly to equation (1), CUI represents the target meaning, \( \text{CUI}_j \) represents any other related concept, \( t \) is the context vector, and \( \text{CUI}_j \) is the concept vector of the related concept \( \text{CUI}_j \). Each context \( t \) is compared to the concept textual definitions \( \text{CUI}_j \) by their cosine similarities \( \text{CS}(t, \text{CUI}_j) \), which are weighted by their nPMI(\( \text{CUI}, \text{CUI}_j \)) association values. The value \( N \) is the total number of relations considered, that is the number of non-zero nPMI values, and it is used to normalize the final score. For each possible CUI is calculated a score, and the one who get the highest score is considered the correct meaning.

4 Results

All the results, presented in Table 1–Table 8, were obtained dividing the dataset into five folds. The average accuracies and the respective standard deviations across the five folds are shown.

Table 1: Supervised learning WSD accuracies (standard deviations) with bag-of-words as local features.

|     | U     | B     | U + B  |
|-----|-------|-------|--------|
| DT  | 0.9067 (0.0030) | 0.8335 (0.0045) | 0.9019 (0.0018) |
| kNN | 0.9324 (0.0017) | 0.8850 (0.0043) | 0.9354 (0.0019) |
| LR  | 0.9205 (0.0025) | 0.8704 (0.0018) | 0.9101 (0.0024) |
| MLP | 0.9401 (0.0013) | 0.9224 (0.0010) | 0.9445 (0.0022) |
| SVM | 0.9511 (0.0013) | 0.9253 (0.0028) | 0.9552 (0.0022) |

Accuracies are the average across five folds. Five classifiers were tested. U, unigrams; B, bigrams; DT, decision tree; kNN, k-nearest neighbor (k = 5); LR, logistic regression; MLP, multi-layer perceptron; SVM, support vector machine. The top accuracy is shown in bold.

Table 2: Supervised learning WSD accuracies (standard deviations) with word embeddings as global features.

|     | S100 | S300 |
|-----|------|------|
|     | W5   | W20  | W50  | W5   | W20  | W50  |
| DT  | 0.9219 (0.0017) | 0.9185 (0.0030) | 0.9194 (0.0033) | 0.9186 (0.0013) | 0.9186 (0.0025) | 0.9166 (0.0017) |
| kNN | 0.9452 (0.0024) | 0.9452 (0.0024) | 0.9447 (0.0017) | 0.9449 (0.0019) | 0.9444 (0.0023) | 0.9441 (0.0025) |
| LR  | 0.9500 (0.0013) | 0.9495 (0.0008) | 0.9495 (0.0011) | 0.9505 (0.0012) | 0.9508 (0.0008) | 0.9509 (0.0013) |
| MLP | 0.9503 (0.0011) | 0.9498 (0.0016) | 0.9501 (0.0012) | 0.9503 (0.0010) | 0.9508 (0.0014) | 0.9514 (0.0016) |
| SVM | 0.9449 (0.0018) | 0.9452 (0.0026) | 0.9431 (0.0012) | 0.9452 (0.0025) | 0.9446 (0.0012) | 0.9444 (0.0008) |

Accuracies are the average across five folds. Five classifiers were tested. S, Size; W, window; DT, decision tree; kNN, k-nearest neighbor (k = 5); LR, logistic regression; MLP, multi-layer perceptron; SVM, support vector machine. The top accuracy is shown in bold.

Table 3: Supervised learning WSD accuracies (standard deviations) with unigrams (bag-of-words) as local features and word embeddings as global features.
|       | W5    | W20   | W50   |
|-------|-------|-------|-------|
| DT    | 0.9244 (0.0018) | 0.9215 (0.0031) | 0.9229 (0.0038) |
| kNN   | 0.9464 (0.0024) | 0.9468 (0.0026) | 0.9467 (0.0022) |
| LR    | 0.9515 (0.0013) | 0.9514 (0.0010) | 0.9515 (0.0008) |
| MLP   | 0.9557 (0.0010) | 0.9556 (0.0006) | 0.9555 (0.0003) |
| SVM   | 0.9490 (0.0008) | 0.9486 (0.0011) | 0.9481 (0.0015) |

Accuracies are the average across five folds. Five classifiers were tested. S, Size; W, window; DT, decision tree; kNN, k-nearest neighbor (k = 5); LR, logistic regression; MLP, multi-layer perceptron; SVM, support vector machine. The top accuracy is shown in bold.

Table 4: Knowledge-based WSD accuracies (standard deviations) using CUI association values (MeSH term co-occurrences), CUI definitions (UMLS), and word embeddings.

|       | W5    | W20   | W50   |
|-------|-------|-------|-------|
| DT    | 0.8144 (0.0012) | 0.8254 (0.0010) | 0.8321 (0.0026) |
| kNN   | 0.8132 (0.0014) | 0.8243 (0.0011) | 0.8314 (0.0024) |
| LR    | 0.8005 (0.0041) | 0.8152 (0.0038) | 0.8197 (0.0038) |
| MLP   | 0.8430 (0.0022) | 0.8573 (0.0022) | 0.8600 (0.0010) |
| SVM   | 0.8164 (0.0024) | 0.8286 (0.0011) | 0.8341 (0.0026) |

TF-IDF word embedding averaging was used to calculate the surrounding context vectors. Accuracies are the average across five folds. S, Size; W, window; CS, cosine similarity between term context vector and concept vector only; nPMI, normalized pointwise mutual information; nPMI ≥ thresh, cosine similarity plus related concepts with a nPMI value higher than the threshold. The top accuracy is shown in bold.

Table 5: Knowledge-based WSD accuracies (standard deviations) using CUI association values (MeSH term co-occurrences), CUI definitions (UMLS), and word embeddings.

|       | W5    | W20   | W50   |
|-------|-------|-------|-------|
| CS    | 0.8164 (0.0024) | 0.8286 (0.0011) | 0.8341 (0.0026) |
| nPMI ≥ 0.8 | 0.8154 (0.0024) | 0.8277 (0.0008) | 0.8334 (0.0020) |
| nPMI ≥ 0.5 | 0.8019 (0.0043) | 0.8178 (0.0031) | 0.8236 (0.0040) |
| nPMI ≥ 0.3 | 0.8458 (0.0023) | 0.8600 (0.0018) | **0.8635** (0.0007) |

IDF word embedding averaging with no decay, \( f(d) = 1 \), was used to calculate the surrounding context vectors. Accuracies are the average across five folds. S, Size; W, window; CS, cosine similarity between term context vector and concept vector only; nPMI, normalized pointwise mutual information; nPMI ≥ thresh, cosine similarity plus related concepts with a nPMI value higher than the threshold. The top accuracy is shown in bold.

Table 6: Knowledge-based WSD accuracies (standard deviations) using CUI association values (MeSH term co-occurrences), CUI definitions (UMLS), and word embeddings.

|       | W5    | W20   | W50   |
|-------|-------|-------|-------|
| CS    | 0.8164 (0.0024) | 0.8286 (0.0011) | 0.8341 (0.0026) |
| nPMI ≥ 0.8 | 0.8154 (0.0024) | 0.8277 (0.0008) | 0.8334 (0.0020) |
| nPMI ≥ 0.5 | 0.8019 (0.0043) | 0.8178 (0.0031) | 0.8236 (0.0040) |
| nPMI ≥ 0.3 | 0.8458 (0.0023) | 0.8600 (0.0018) | **0.8635** (0.0007) |

IDF word embedding averaging with no decay, \( f(d) = 1 \), was used to calculate the surrounding context vectors. Accuracies are the average across five folds. S, Size; W, window; CS, cosine similarity between term context vector and concept vector only; nPMI, normalized pointwise mutual information; nPMI ≥ thresh, cosine similarity plus related concepts with a nPMI value higher than the threshold. The top accuracy is shown in bold.
IDF word embedding averaging with fractional decay, $f(d) = 1/d$, was used to calculate the surrounding context vectors. Accuracies are the average across five folds. S, Size; W, window; CS, cosine similarity between term context vector and concept vector only; nPMI, normalized pointwise mutual information; nPMI $\geq$ thresh, cosine similarity plus related concepts with a nPMI value higher than the threshold. The top accuracy is shown in bold.

### Table 7: Knowledge-based WSD accuracies (standard deviations) using CUI association values (MeSH term co-occurrences), CUI definitions (UMLS), and word embeddings.

|       | S100          |          | S300          |          |
|-------|---------------|----------|---------------|----------|
|       | W5 W20 W50    | W5 W20 W50 |               |          |
| CS    | 0.8415 (0.0022) | 0.8473 (0.0018) | 0.8502 (0.0033) |          |
| nPMI $\geq$ 0.8 | 0.8395 (0.0022) | 0.8459 (0.0024) | 0.8493 (0.0032) | 0.8438 (0.0019) |
| nPMI $\geq$ 0.5 | 0.8234 (0.0013) | 0.8348 (0.0012) | 0.8376 (0.0015) | 0.8267 (0.0023) |
| nPMI $\geq$ 0.3 | 0.8617 (0.0017) | 0.8720 (0.0016) | **0.8744** (0.0021) | 0.8622 (0.0020) |

IDF word embedding averaging with exponential decay, $f(d) = \exp(-d)$, was used to calculate the surrounding context vectors. Accuracies are the average across five folds. S, Size; W, window; CS, cosine similarity between term context vector and concept vector only; nPMI, normalized pointwise mutual information; nPMI $\geq$ thresh, cosine similarity plus related concepts with a nPMI value higher than the threshold. The top accuracy is shown in bold.

### Table 8: Knowledge-based WSD accuracies (standard deviations) using CUI association values (MeSH term co-occurrences), CUI definitions (UMLS), and word embeddings.

|       | S100          |          | S300          |          |
|-------|---------------|----------|---------------|----------|
|       | W5 W20 W50    | W5 W20 W50 |               |          |
| CS    | 0.8259 (0.0013) | 0.8270 (0.0031) | 0.8278 (0.0036) |          |
| nPMI $\geq$ 0.8 | 0.8236 (0.0011) | 0.8255 (0.0032) | 0.8264 (0.0033) | 0.8255 (0.0021) |
| nPMI $\geq$ 0.5 | 0.8057 (0.0022) | 0.8137 (0.0007) | 0.8150 (0.0027) | 0.8092 (0.0035) |
| nPMI $\geq$ 0.3 | 0.8378 (0.0022) | 0.8458 (0.0032) | 0.8459 (0.0030) | 0.8404 (0.0029) |

IDF word embedding averaging with logarithmic decay, $f(d) = 1/\ln(1 + d)$, was used to calculate the surrounding context vectors. Accuracies are the average across five folds. S, Size; W, window; CS, cosine similarity between term context vector and concept vector only; nPMI, normalized pointwise mutual information; nPMI $\geq$ thresh, cosine similarity plus related concepts with a nPMI value higher than the threshold. The top accuracy is shown in bold.
4.1 Supervised

Supervised learning results were tested using five distinct classifiers as described in Section 3.4. Table 1 shows the results using only bag-of-words features (unigrams, bigrams) with a best accuracy of 95.5% using a SVM classifier. Table 2 shows the results using only word embeddings with a best accuracy of 95.1% using a MLP classifier. The combination of unigrams and word embeddings (Table 3) improved the individual accuracies achieving a best accuracy of 95.6% using a MLP classifier. One can see that the differences of using different word embedding models are not significant.

4.2 Knowledge-Based

Knowledge-based results were tested using five distinct word embedding averaging functions as described in Section 3.3.2 (Table 4 – Table 8). Different thresholds (0.3, 0.5, 0.8, 1.0) for the nPMI values were imposed to filter out the most weighty relations. The threshold 1.0 is the particular case of the baseline scenario where only the cosine similarity between the context vector and the possible CUI (meaning) vector is computed. One can see that, in all the word embedding averaging functions, the threshold 0.3 produced the best accuracies proving that the addition of more related concepts leads to a better score refinement. The fractional decay averaging function obtained the highest results (Table 6), while the exponential decay averaging function obtained the lowest results (Table 7) even when compared to the baseline TF-IDF weighting (Table 4). Also, the word embedding models with higher windows achieved slightly higher accuracies. The top accuracy, 87.4%, is presented in Table 6 and it was achieved using the fractional decay averaging function, the nPMI threshold set to 0.3, and the word embedding model with size of 100 and window of 50.

5 Discussion

In this paper we extended our previous work [9] by applying more settings to the supervised and knowledge-based approaches. Furthermore, we extracted textual definitions for every CUI included in the MSH WSD dataset, making it possible to apply our knowledge-based method to the entire dataset.

As expected, the supervised classifiers obtained the highest results with a top accuracy around 95.6%, while on the other hand our knowledge-based approach obtained a best accuracy around 87.4%. Our supervised accuracy is very close to the state-of-the-art accuracy of 96.0%, which was also obtained using a combination of unigrams and word embeddings with a SVM classifier [16].

Our knowledge-based method and results are comparable with other proposed knowledge-based approaches. In [6], Jimeno-Yepes et al. proposed the MSH WSD dataset, and tested four knowledge-based methods, where the automatic extracted corpus (AEC) method obtained the best accuracy around 84.5%. McInnes and Pedersen [12] developed a knowledge-based method based on semantic similarity measures between UMLS concepts, and obtained an accuracy of 75% in the same dataset. In [14], the authors used word-concept probabilities achieving a knowledge-based accuracy around 89%. Our method is similar to the one proposed by Tulkens et al. [24], which also compared concept representations with the representations of the context of ambiguous terms, who obtained an accuracy of 84.0% on the same dataset. As far as we know, the knowledge-based state-of-the-art accuracy in the MSH WSD dataset is 92.2%, and it was obtained using a method proposed by Sabbir et al. [15] that uses neural word/concept embeddings.

Our work showed that the word embeddings and their averaging function plays a key role in the WSD problem.

Acknowledgements

This work was supported by Portuguese national funds through FCT - Foundation for Science and Technology, in the context of the project IF/01694/2013/CP1162/CT0018. Sérgio Matos is funded under the FCT Investigator Programme.

Conflict of interest statement: Authors state no conflict of interest. All authors have read the journal’s publication ethics and publication malpractice statement available at the journal’s website and hereby confirm that they comply with all its parts applicable to the present scientific work.
Notes

1 https://www.nlm.nih.gov/research/umls/licensedcontent/umlsknowledgesources.html
2 https://ii.nlm.nih.gov/MRCOC.shtml

References

[1] Navigli R. Word sense disambiguation: a survey. ACM Comput Surv 2009;41:10:1–69.
[2] Schuemie MJ, Kors JA, Mons B. Word sense disambiguation in the biomedical domain: an overview. J Comput Biol 2005;12:554–65.
[3] McInnes BT, Stevenson M. Determining the difficulty of word sense disambiguation. J Biomed Inform 2014;47:83–90.
[4] Weeber M, Mork JC, Aronson AR. Developing a test collection for biomedical word sense disambiguation. In: Proceedings of the AMIA symposium. American Medical Informatics Association, 2001:746–750. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2243574/.
[5] Moon S, Pakhomov S, Liu N, Ryan JO, Melton GB. A sense inventory for clinical abbreviations and acronyms created using clinical notes and medical dictionary resources. J Am Med Inform Assoc 2014;21:299–307.
[6] Jimeno-Yepes A, McInnes BT, Aronson AR. Exploiting MeSH indexing in MEDLINE to generate a data set for word sense disambiguation. BMC Bioinformatics 2011;12:223.
[7] Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. Nucleic Acids Res 2004;32(Suppl 1):D267.
[8] Lipscomb CE. Medical Subject Headings (MeSH). Bull Med Libr Assoc 2000;88:265–66.
[9] Antunes R, Matos S. Biomedical word sense disambiguation with word embeddings. In: 11th International Conference on practical applications of Computational Biology & Bioinformatics. Springer International Publishing, 2014:273–9.
[10] Alexopoulou D, Andreopoulos B, Dietze H, Doms A, Candon F, Hakenberg J, et al. Biomedical word sense disambiguation with ontologies and metadata: automation meets accuracy. Bioinformatics 2009;10:28.
[11] Carla VN, Brandt C. Knowledge-based biomedical word sense disambiguation: an evaluation and application to clinical document classification. J Am Med Inform Assoc 2013;20:882.
[12] McInnes BT, Pedersen T. Evaluating measures of semantic similarity and relatedness to disambiguate terms in biomedical text. J Biomed Inform 2013;46:1116–24.
[13] El-Rab WG, Zaiane OR, El-Hajj M. Biomedical text disambiguation using UMLS. In: Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. Niagara, Ontario, Canada, ACM, 2013:943–947.
[14] Jimeno-Yepes A, Berlanga R. Knowledge based word-concept model estimation and refinement for biomedical text mining. J Biomed Inform 2015;53:300–7.
[15] Sabbir AKM, Jimeno-Yepes A, Kavuluru R. Knowledge-based biomedical word sense disambiguation with neural concept embeddings. 17th IEEE International Conference on Bioinformatics and BioEngineering (BIBE), 2017.
[16] Jimeno-Yepes A. Word embeddings and recurrent neural networks based on long-short term memory nodes in supervised biomedical word sense disambiguation. J Biomed Inform 2017;73(Supplement C):137–47.
[17] Jacobacci I, Pilehvar MT, Navigli R. Embeddings for word sense disambiguation: an evaluation study. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Berlin, Germany: Association for Computational Linguistics, 2016:897–907.
[18] Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. arXiv e-print. 2013; Available from: https://arxiv.org/abs/1301.3781.
[19] Habibi M, Weber L, Neves M, Wiegandt DL, Leser U. Deep learning with word embeddings improves biomedical named entity recognition. Bioinformatics 2017;33:137–48.
[20] Taghipour K, Ng HT. Semi-supervised word sense disambiguation using word embeddings in general and specific domains. In: Proceedings of the North American Chapter of the Association for Computational Linguistics — Human Language Technologies (NAACL-HLT 2015). Denver, Colorado, USA, 2015:314–323.
[21] Wu Y, Xu J, Zhang Y, Xu H. Clinical abbreviation disambiguation using neural word embeddings. In: Proceedings of the 2015 Workshop on Biomedical Natural Language Processing (BioNLP 2015). Beijing, China: Association for Computational Linguistics, 2015:171–176.
[22] Rehürek R, Sojka P. Software framework for topic modelling with large corpora. In: Proceedings of the LREC 2010 Workshop on NewChallenges for NLP Frameworks. Valletta, Malta, 2010:45–50.
[23] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. J Mach Learn Res 2011;12:2825–30.
[24] Tulkens S, Suster S, Daeleman W. Using distributed representations to disambiguate biomedical and clinical concepts. In: Proceedings of the 15th Workshop on Biomedical Natural Language Processing. Berlin, Germany: Association for Computational Linguistics, 2016:77–82.