Symbiosis

Research Article | Journal of Computer Science Applications and Information Technology | Open Access

Received: June 25, 2017
Accepted: July 15, 2017
Published: July 24, 2017

Word Sense Disambiguation Based On Global Co-Occurrence Information Using Non-Negative Matrix Factorization

Minoru Sasaki*

Department of Computer and Information Sciences, College of Engineering, Ibaraki University, Japan

***Corresponding author:** Minoru Sasaki, Lecturer, Department of Computer and Information Sciences, Faculty of Engineering, Ibaraki University, 4-12-1, Naka-narusawa, Hitachi, Ibaraki, Japan, Tel: +81-294-38-5159; Fax: +81-294-38-5159, E-mail: minoru.sasaki.01@vc.ibaraki.ac.jp

1 Abstract

In this paper, I propose a novel word sense disambiguation method based on global co-occurrence information using Non-negative Matrix Factorization (NMF). When I calculate the dependency relation matrix, the existing method tends to produce very sparse co-occurrence matrix from a small training set. Therefore, the NMF algorithm sometimes does not converge to desired solutions. To obtain a large number of co-occurrence relations, I propose to use co-occurrence frequencies of dependency relations between word features in the whole training set. This enables us to solve data sparseness problem and induce more effective latent features. To evaluate the efficiency of the method of word sense disambiguation, I make some experiments to compare with the result of the two baseline methods. The results of the experiments show this method is effective for word sense disambiguation in comparison with the all baseline methods. Moreover, the proposed method is effective for obtaining a stable effect by analyzing the global co-occurrence information.

2 Keywords

Word sense disambiguation; Global co-occurrence information; Dependency relations; Non-negative matrix factorization;

3 Introduction

Natural language processing (NLP) is a field of computer science designed to interpret and process natural language, in either textual or spoken form. It aims to design computer systems that can understand human language. These techniques allow computers to understand natural language to perform various tasks. Word Sense Disambiguation (WSD) is one of a fundamental problem in the natural language processing. The problem of WSD is defined as the task of finding the most appropriate meaning for a polysemous word within a given context.

The most successful approaches of WSD employ supervised machine learning techniques to extract linguistic knowledge from natural language data automatically. Supervised learning for WSD requires large amounts of labelled training data which consist sense-annotated instances for a specific word to construct a classifier. Then, the obtained classifier is used to identify the appropriate sense for new examples. A typical method for this approach is the classical Bag-Of-Words (BOW) approach, where each document is represented as a feature vector counting the number of occurrences of different words as features [9]. By using such features, it becomes easy to adapt many existing supervised learning methods such as Support Vector Machine (SVM) for the WSD task. However, when the general vector space model, in which a document is represented as a vector using term frequency based weighting methods, is applied to the WSD, the local context words are typically used as features and the global co-occurrence information without dictionary information is not employed in the previous research [1].

In this paper, I propose a novel WSD method based on the global co-occurrence information using Non-Negative Matrix Factorization (NMF). Previous study proposes a novel WSD method of particular word instances using the automatically extracted sense information [5]. When I calculate the dependency relation matrix, the existing method tends to produce very sparse co-occurrence matrix from a small training set. Therefore, the NMF algorithm sometimes does not converge to desired solutions. To avoid this problem and to obtain more effective co-occurrence relations, I propose to use co-occurrence frequencies of dependency

Citation: Sasaki M (2017) Word Sense Disambiguation Based On Global Co-Occurrence Information Using Non-Negative Matrix Factorization. J Comp Sci Appl Inform Technol. 2(3):1-4

relations between word features in the large document set. This enables us to solve data sparseness problem and induce more effective latent features.

The organization of residual of the paper is as follows. Related works Section is devoted to the introduction of the related work in the literature. Wsd Using Global Co-Occurrence Information Section describes the proposed WSD system based on global co-occurrence information using Non-negative Matrix Factorization. In Experiment Section, I describe an outline of experiments. Experimental results and discussions are presented in Experimental Results and Discussions Section. Finally, Conclusion Section concludes this paper.

4 Related works

WSD is the process for identification of appropriate meaning of polysemous words for a particular context. This process is based on the distributional hypothesis that words that occur in the same context tend to have similar meaning [3]. Thus, many approaches to WSD have focused on the contexts formed by the words surrounding the target polysemous word. By using a set of features that represent the contexts of the target word (various combinations of collocations and bag-of-words, etc.), the extracted features are represented in a multidimensional feature vector.

The feature vector such as a bag-of-words model is a simple but effective representation used in natural language processing. This feature space model represents words in a vector space where words that are close in meaning are mapped to nearby points. Latent Semantic Analysis (LSA) is one of the most popular techniques to transform the original feature space to a semantic space of low dimensionality by analysing relationships between a set of documents and the words they contain [2]. In this paper, I employ the modified Non-Negative Matrix Factorization approach to find a low dimensional semantic space.

The proposed method presents a novel approach to capture global co-occurrence information as well as local information for WSD. The local context is a window of words that occur around the target polysemous word in the sentence and includes information about word order, collocation and syntactic structure [4]. The earliest use of local context for WSD was proposed by Lin [7]. This method uses syntactic dependencies to resolve word sense ambiguity.

5 Wsd Using Global Co-Occurrence Information

5.1 System Overview

A WSD system is used to select the appropriate sense for a target polysemous word in context. WSD can be viewed as a classification task in which each target word should be classified into one of the predefined existing senses. In this paper, supervised classification is employed for this WSD task. This supervised method requires a corpus of manually labelled training data to construct classifiers for every polysemous word. Then, each obtained classifier is applied to a set of unlabeled examples.

5.2 Non-Negative Matrix Factorization

Non-Negative Matrix Factorization (NMF) is a popular decomposition method for multivariate data [4]. NMF decomposes the m×n non-negative matrix X to the m×k matrix W and the k×n matrix H, while these matrixes have no negative elements. Usually, k is chosen to be smaller value than n and m.

$$X \approx WH \qquad (1)$$

Using the NMF for a term-document matrix X, the matrix H represents the induced result with k topics.

For quantifying the quality of this approximation, cost functions based on Kullback-Leibler divergence is used and minimized using iterative update rules as follows:

$$W_{ij} \leftarrow W_{ij} \frac{(XH)_{ij}}{(WHH^T)_{ij}} \qquad (2)$$

$$H_{ij} \leftarrow H_{ij} \frac{\left(X^T H\right)_{ij}}{(HW^T W)_{ij}} \qquad (3)$$

Where W_{ij} and H_{ij} indicate the i-th row and the j-th column element respectively. These matrices W and H are initialized randomly with non-negative data and these above update rules are iteratively applied until the max iteration number (or convergence) is reached.

5.3 Latent Semantic WSD Using Local Co-occurrence Information

In previous research, proposes a WSD method of particular word instances using the automatically extracted sense information [3]. This method induces latent features for three matrices. The first n×t matrix A, where n is the number of examples that contain the target word and t is the number of words that have dependency relation with the target word, contains co-occurrence frequencies of the target word cross-classified by dependency relations. The second n×s matrix B, where s is the number of words that appear in the context window, contains term frequencies of words that appear in the context window. The third s×t matrix C contains co-occurrence frequencies of words that the co-occurring context words of the target word co-occur with. Then, NMF is applied to the three matrices to factorize each matrix into two non-negative matrices, while the former results are used to initialize the next factorization, as shown in Figure 1. By inducing these latent factors such as dependency relations of words, context words and co-occurrence frequencies, this method obtains rich context information of the target word in comparison to co-occurrence matrix from training examples used in general WSD methods.

Given a non-negative matrices A, B and C in the beginning of this method, matrices W, H, G and F are initialized randomly with non-negative values. Then, it decomposes the matrix A into the two matrices $W_{n\times k}$ and $H_{k\times t}$ using NMF such that

$$A \approx W_{n\times k} H_{k\times t} \qquad (4)$$

Where k is the number of topics, $W_{n\times k}$ is the n×k word-topic matrix and $H_{k\times t}$ is the k×t topic-word matrix. In the decomposi-

Citation: Sasaki M (2017) Word Sense Disambiguation Based On Global Co-Occurrence Information Using Non-Negative Matrix Factorization. J Comp Sci Appl Inform Technol. 2(3):1-4

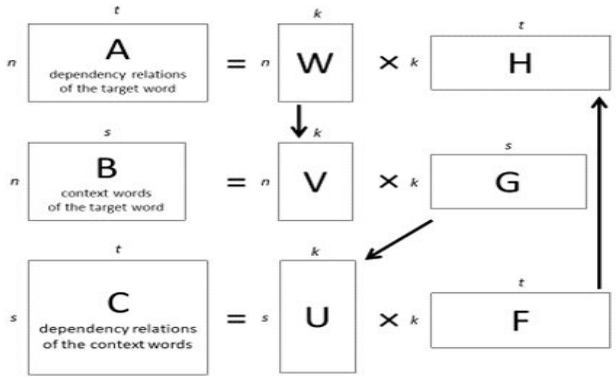

Figure 1: Interleaved NMF algorithm for Latent Semantic WSD

tion of the matrix B, the updated matrix W is copied to matrix V and the updated matrices $V_{n\times k}$ and $G_{k\times s}$ is computed using NMF such that

$$B \approx V_{n\times k} G_{k\times s} \qquad (5)$$

Where $V_{n\times k}$ is the n×k word-topic matrix and $G_{k\times s}$ is the k×s topic-word matrix. In the decomposition of the matrix C, the transpose of the updated matrix G is copied to matrix U and the updated matrices $U_{s\times k}$ and $F_{k\times t}$ are obtained using NMF such that

$$C \approx U_{s\times k} F_{k\times t} \qquad (6)$$

Where $U_{s\times k}$ is the s×k word-topic matrix and $F_{k\times t}$is the k×t topic-word matrix. At the last step of the iteration, the matrix F is copied to matrix H. This iteration is repeated until the maximum number of iterations is reached or the objective function of all NMF decomposition no longer improves.

In order to perform this method for WSD, it needs to fold each sense of the target word into semantic space using the matrix H. For each sense label l in training data, the centroid vector cl is calculated from the mean of the vectors in the sense l and this centroid is mapped into the semantic space using the matrix H as follows:

$$b = c_l H^T \qquad (7)$$

For test data of the target word, its context words are extracted to construct a vector f and the vector f is also mapped into the same semantic space using the matrix G as follows:

$$d = f G^T \qquad (8)$$

Then, cosine similarity between the vector d and each of the sense vectors b are calculated and the sense that is the largest cosine similarity is selected.

5.4 Latent Semantic WSD Using Global Co-occurrence Information

This previous latent semantic WSD method is efficient for finding a reduced semantic space. However, problem arises when I apply this method. When I calculate the third matrix C, this method tends to produce very sparse co-occurrence matrix from a small training set. In this case, the NMF algorithm does not converge onto a unique solution properly.

To obtain a large number of co-occurrence relations, I propose a new WSD method based on the global co-occurrence information using NMF, as shown in Figure 2. The proposed method also

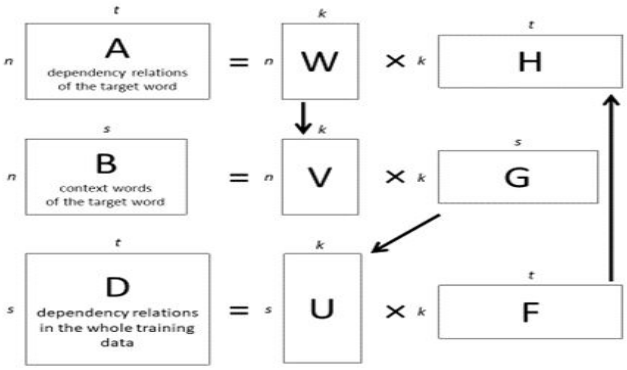

Figure 2: The proposed algorithm based on global co-occurrence information

induces latent features for three matrices. The first n×t matrix A, where n is the number of examples that contain the target word and t is the number of words that has dependency relation with the target word, contains co-occurrence frequencies of the target word cross-classified by dependency relations. The second n×s matrix B, where s is the number of words that appear in the context window, contains term frequencies of words that appear in the context window. The third s×t matrix D contains co-occurrence frequency of context words that co-occur in dependency relations to context words in a large document set. The proposed method induces latent features for these three matrices A, B and D. By using the co-occurrence frequencies of dependency relations in the larger document set, this enables us to solve data sparseness problem and induce more effective latent features.

6 Experiment

To evaluate the efficiency of the proposed WSD method using the global co-occurrence information, I conduct some experiments to compare with the result of the existing methods. In this section, I describe an outline of the experiments.

6.1 Data

I used the Semeval-2010 Japanese WSD task data set, which includes 50 target words comprising 22 nouns, 23 verbs, and 5 adjectives [2]. In this data set, there are 50 training and 50 test instances for each target word. Moreover, to obtain a large number of co-occurrence relations, I use 22,832 documents chosen from the Japanese BCCWJ corpus[1].

6.2 Evaluation Method

6.2.1 Baseline System 1

As the first baseline method, I only use the first matrix A described in Latent Semantic WSD Using Global Co-occurrence Information section. To construct the matrix A, I represent each sentence with the target word in the training set as a high-

Citation: Sasaki M (2017) Word Sense Disambiguation Based On Global Co-Occurrence Information Using Non-Negative Matrix Factorization. J Comp Sci Appl Inform Technol. 2(3):1-4

dimensional vector where each component represents the co-occurrence frequency of the target word in the sentence. Then, NMF is applied to the matrix A to factorize each matrix into two non-negative matrices W and H. Each vector is tagged with the sense of the target word in that sentence. So centroid ci of the co-occurrence vectors that are assigned the same sense i is calculated and each centroid ci is mapped into the semantic space using the matrix H as follows:

$$b_i = c_i H^T \qquad (9)$$

For input example of the target word, its context words are extracted to construct a vector f and the vector f is also mapped into the same semantic space using the matrix H as follows:

$$d = f H^T \qquad (10)$$

Then, cosine similarity between the vector d and each of the sense vectors b are calculated and the sense that is the largest cosine similarity is selected.

6.2.2 Baseline System 2

In the second baseline system, I use the latent semantic WSD using local co-occurrence information described in System Overview Section. I construct the three matrices A, B and C to induce latent semantic dimensions using NMF.

7 Experimental Results and Discussions

Figure 3 shows the experimental results of the baseline methods and the proposed method. The number of topics k is k=30 in this experiment. Computational experience reported shows that the choice of initial point is quite important to the NMF's goal. In practice, the algorithms are run several times with different initial points and the NMF is chosen as the feasible solution. In my experiments, each method is executed three times and average precision of all execution is calculated.

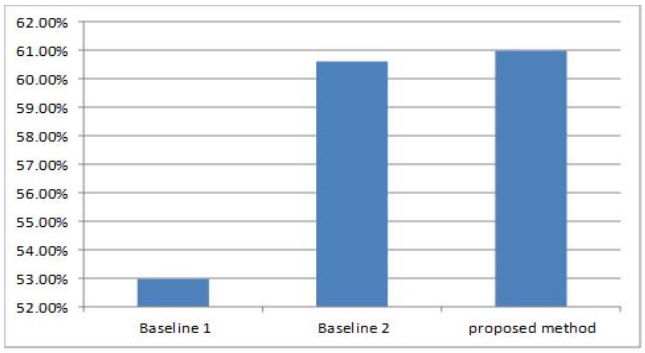

Figure 3: Highest Precision of Each system

In this Figure 3, the proposed method shows higher precision than the other baseline methods so that this approach is effective for WSD. In comparison with the baseline system 1, the proposed method can obtain better precision so that it is effective for WSD to use context information and co-occurrence information. In comparison with the baseline system 2, the proposed method provides slightly better precision than the baseline system 2. As shown in Table 1, the proposed method can obtain the highest precision and can be stable at high precision value. However, the baseline system 2 cannot achieve stable precision because of the lack of the number of co-occurrence information. Therefore, the proposed method is effective for obtaining a stable effect by analyzing the global co-occurrence information.

Table 1 Experimental Results of Each Execution
(highest average precision rates are written in bold font)

System	Run 1	Run 2	Run 3	Average
Baseline System 1	53.28%	53.88%	51.80%	52.99%
Baseline System 2	59.68%	61.08%	61.08%	60.61%
Proposed Method	60.44%	61.48%	61.04%	**60.99%**

8 Conclusion

In this paper, I propose a novel word sense disambiguation method based on the global co-occurrence information using NMF. To evaluate the efficiency of the method of WSD, I conduct some experiments to compare with the result of the two baseline methods. The results of the experiments show this method is effective for WSD in comparison with the all baseline methods. Moreover, the proposed method is effective for obtaining a stable effect by analyzing the global co-occurrence information.

Further work would be required to consider a larger sized training data to obtain a large amount of co-occurrence information.

9 References

1. Cortes C, Vapnik V. Support-Vector Networks. Machine Learning. 1995;20(3):273-297.
2. Deerwester S, Dumais ST, Landauer TK, Furnas GW, Harshman RA. Indexing by latent semantic analysis. Journal of the Society for Information Science. 1990;41(6):391-407.
3. Harris ZS. Distributional Structure. Word. 1954;10(3):146-162. DOI: 10.1080/00437956.1954.11659520
4. Ide N, Veronis J. Word Sense Disambiguation: The State of the Art. Computational Linguistics. 1998;24(1):1-40.
5. Van de Cruys T, Apidianaki M. Latent Semantic Word Sense Induction and Disambiguation. Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies. 2011;1:1476-1485.
6. Lee DD, Seung HS. Algorithms for Non-negative Matrix Factorization. Advances in Neural Information Processing Systems 13, MIT Press. 2001:556-562.
7. Lin D. Using syntactic dependency as local context to resolve word sense ambiguity. Proceedings of the eighth conference on European chapter of the Association for Computational Linguistics (EACL '97). 1997:64-71.
8. Okumura M, Shirai K, Komiya K, Yokono H. SemEval-2010 task: Japanese WSD. Proceedings of the 5th International Workshop on Semantic Evaluation. 2010:69-74.
9. Witten IH, Moffat A, Bell TC. Managing Gigabytes: Compressing and Indexing Documents and Images. Second ed. Morgan Kaufmann Publishers Inc. 1999.

Citation: Sasaki M (2017) Word Sense Disambiguation Based On Global Co-Occurrence Information Using Non-Negative Matrix Factorization. J Comp Sci Appl Inform Technol. 2(3):1-4