

Detection of Peculiar Word Sense by Distance Metric Learning with Labeled Examples

Minoru Sasaki, Hiroyuki Shinnou

Department of Computer and Information Sciences,
Faculty of Engineering, Ibaraki University
4-12-1, Nakanarusawa, Hitachi, Ibaraki, 316-8511, Japan
msasaki@mx.ibaraki.ac.jp, shinnou@mx.ibaraki.ac.jp

Abstract

For natural language processing on machines, resolving such peculiar usages would be particularly useful in constructing a dictionary and dataset for word sense disambiguation. Hence, it is necessary to develop a method to detect such peculiar examples of a target word from a corpus. Note that, hereinafter, we define a peculiar example as an instance in which the target word or phrase has a new meaning. In this paper, we proposed a new peculiar example detection method using distance metric learning from labeled example pairs. In this method, first, distance metric learning is performed by large margin nearest neighbor classification for the training data, and new training data points are generated using the distance metric in the original space. Then, peculiar examples are extracted using the local outlier factor, which is a density-based outlier detection method, from the updated training and test data. The efficiency of the proposed method was evaluated on an artificial dataset and the Semeval-2010 Japanese WSD task dataset. The results showed that the proposed method has the highest number of properly detected instances and the highest F-measure value. This shows that the label information of training data is effective for density-based peculiar example detection. Moreover, an experiment on outlier detection using a classification method such as SVM showed that it is difficult to apply the classification method to outlier detection.

Keywords: Peculiar Word Sense Detection, Semi-Supervised Outlier Detection, Distance Metric Learning

1. Introduction

In everyday life, we often encounter examples of words used in an unknown sense, including those that may not even be listed in the dictionary. For natural language processing on machines, resolving such peculiar usages would be particularly useful in constructing a dictionary and dataset for word sense disambiguation (WSD). Hence, it is necessary to develop a method to detect such peculiar examples of a target word from a corpus. Note that, hereinafter, we define a peculiar example as an instance in which the target word or phrase has a new meaning.

As one approach, we consider the outlier detection methods used in data mining. However, although outlier detection is used to detect anomalous observations from data, it is generally unsupervised and so is unable to incorporate sense information for the detection of peculiar examples.

To solve this problem, here we propose a new approach using distance metric learning from labeled example pairs. First, distance metric learning is performed by large margin nearest neighbor (LMNN) classification (Weinberger and Saul, 2009) for the training data, and new training data points are generated using the distance metric in the original space. Then, peculiar examples are extracted using the local outlier factor (LOF) (Breunig et al., 2000), which is a density-based outlier detection method, from the updated training and test data.

In this paper, we present the results of experimental evaluations of the proposed method using an artificial dataset and the Semeval-2010 Japanese WSD task dataset (Okumura et al., 2010). The proposed method proved to be effective on both datasets in comparison with the LOF and the one-class support vector machine (SVM) (Schölkopf et al., 2001). Moreover, we present the results of an experi-

ment on outlier detection using a classification method such as an SVM. The results show that it is difficult to apply the classification method for outlier detection.

2. Outlier Detection

Many methods have been proposed for detecting outlier instances, such as distance-based methods (Orair et al., 2010), probabilistic methods (Kriegel et al., 2009), and density-based methods. Here we briefly explain the LOF algorithm, a local density-based method for outlier detection, and the one-class SVM (Cortes and Vapnik, 1995), an unsupervised segmentation method based on machine learning.

2.1. LOF

LOF is a well-known outlier detection method for unlabeled data sets. This method specifies the degree of outlieriness, determined from the difference in density between a data object and its neighborhood. Outliers are objects that have high LOF values; in other words, objects that have low LOF values are likely to be normal with respect to their neighborhood.

The first step in computing the LOF value of data object x is to compute its k -distance(x), where k is an arbitrary positive constant. The k -distance(x) of object x in a dataset D is defined as the distance $d(x, y)$ between two objects as follows:

1. $d(x, y') \leq d(x, y)$ for at least k objects $y' \in D \setminus \{x\}$,
2. $d(x, y') < d(x, y)$ for at most $k - 1$ objects $y' \in D \setminus \{x\}$.

In other words, the k -distance(x) represents the distance between the object x and the k -th nearest object from x .

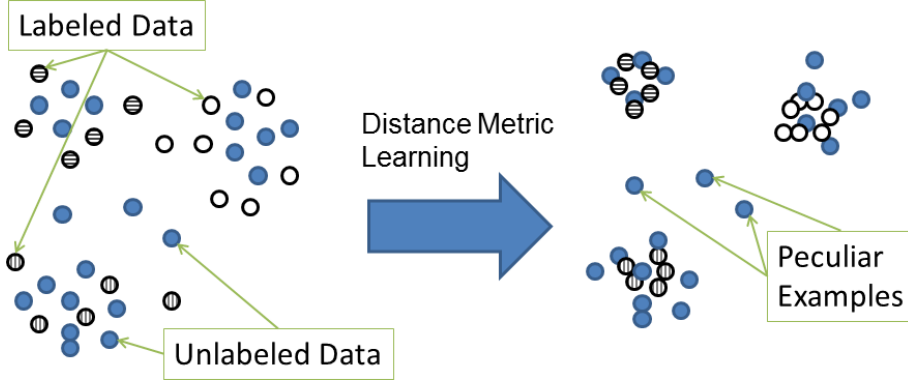


Figure 1: Peculiar Example Detection

Then, the second step is to find the set of k nearest neighbors $N_k(x)$ of x .

$$N_k(x) = \{y \in D \setminus \{x\} | d(x, y) \leq k\text{-distance}(x)\} \quad (1)$$

The third step is to compute the reachability distance $\text{reach-dist}_k(x, y)$ of an object x with respect to object y and the local reachability density $\text{lrd}_k(x)$ of x .

$$\text{reach-dist}_k(x, y) = \max\{d(x, y), k\text{-distance}(y)\} \quad (2)$$

$$\text{lrd}_k(x) = \frac{|N_k(x)|}{\sum_{y \in N_k(x)} \text{reach-dist}_k(x, y)} \quad (3)$$

Finally, we compute the local outlier factor $\text{LOF}(x)$ of x ,

$$\text{LOF}(x) = \frac{1}{|N_k(x)|} \sum_{y \in N_k(x)} \frac{\text{lrd}_k(y)}{\text{lrd}_k(x)} \quad (4)$$

$\text{LOF}(x)$ is the average of the ratios of the local reachability density of x and those of its k -nearest neighbors. The LOF of an object is based on the single parameter k , which is the number of nearest neighbors. In this paper, this parameter k is set to 4.

2.2. One-Class SVM

A one-class SVM is one of the most popular unsupervised learning methods to identify outliers in a dataset (Schölkopf et al., 2001). This method can be considered as a popular two-class SVM, where all data in a dataset lie in the class “+1” and the origin is the class “-1.” This method maps the input data into a higher-dimensional space using a kernel function and constructs a separating hyperplane to best separate class “+1” data from class “-1” data with the maximum margin. Then, by using soft margin classification, both the origin and the data that are close to the origin belong to the class “-1” and these data are classified as outliers.

In this method, we need to select a kernel function and set the parameter as a misclassification cost using the LIBSVM tool¹ (Chang and Lin, 2011). In the experiments, we used libsvm with a linear kernel and set the parameter “nu” to 0.02.

¹<http://www.csie.ntu.edu.tw/~cjlin/libsvm/>

3. Proposed Peculiar Example Detection Method

As mentioned above, the LOF and one-class SVM are unsupervised outlier detection techniques. In general, outlier detection is difficult to use with supervised or semi-supervised frameworks because we can not establish the definitions of outliers and normal classes. Therefore, when we apply the LOF to peculiar example detection, we cannot use only the data for which sense labels are not assigned. Furthermore, if we apply the LOF to the whole dataset, it is possible that training data will be identified as peculiar examples.

In peculiar example detection, however, an outlier is an instance of the usage of a target word whose sense is not listed in the dictionary. Hence, we can define the outlier and normal classes clearly and assign the sense label to the subset of examples.

For this reason, we propose a new peculiar example detection method using distance metric learning from labeled example pairs. As shown in Figure 1, the distance metric is learned from the given training data such that the same class’s data points are close to each other and those in different classes are separated by a large margin. This method affords a high-density region for each sense and thus allows effective detection of peculiar examples using density-based outlier detection. In this paper, we use the LMNN as the distance metric learning method to extract peculiar examples using the LOF.

3.1. Distance Metric Learning by LMNN

LMNN is a method for learning a distance metric such that data points in the same class are close to each other and those in different classes are separated by a large margin, as shown in Figure 2.

In this method, the k neighbors of data \mathbf{x}_i are the k nearest neighbors that share the same label y_i , and the matrix η is defined as $\eta_{ij} = 1$ if the input \mathbf{x}_j is a target neighbor of input \mathbf{x}_i and 0 otherwise. From these definitions, the cost function of LMNN is given by,

$$\begin{aligned} \varepsilon(\mathbf{A}) = & \sum_{ij} \eta_{ij} \|\mathbf{A}\mathbf{x}_i - \mathbf{A}\mathbf{x}_j\|^2 + c \sum_{ijl} \eta_{ij}(1 - \eta_{il}) \\ & [1 + \|\mathbf{A}\mathbf{x}_i - \mathbf{A}\mathbf{x}_j\|^2 - \|\mathbf{A}\mathbf{x}_i - \mathbf{A}\mathbf{x}_l\|^2]_+, \end{aligned} \quad (5)$$

where $[\cdot]_+$ denotes the positive part, for example, $[a]_+ = a$

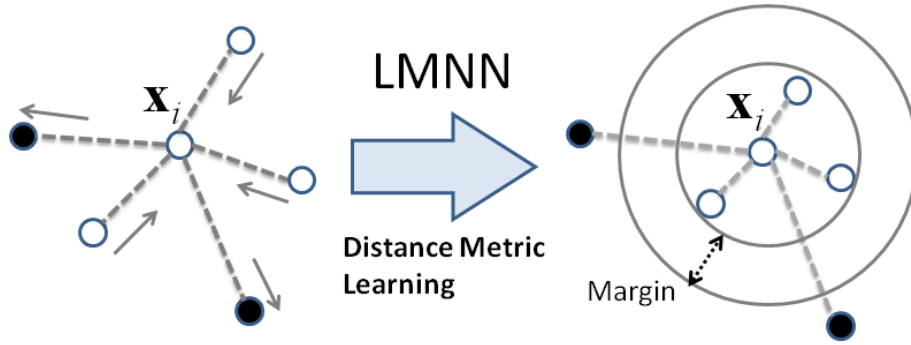


Figure 2: Large Margin Nearest Neighbor

Table 1: Experimental Results for Artificial Dataset

Method	Number of Extracted Data	Number of Correct Outliers	F-Value
LOF	20	12	0.600
OCS	30	12	0.480
Proposed Method	20	15	0.750

Table 2: Experimental Results for Semeval-2010 Japanese WSD Task Dataset

Method	Number of Extracted Data	Number of Correct Outliers	F-Value
LOF	960	3	0.006
OCS	1150	3	0.005
Proposed Method	960	5	0.012

if $a > 0$, and 0 otherwise, and c is some positive constant.

4. Experiments

4.1. Outlier Detection on a Small Dataset

The artificial data set in this experiment consisted of three five-dimensional Gaussian mixture models. From each class distribution, 20 instances were generated for training data and 180 instances for test data. In this data, there was no correlation between each element of the vector. Additionally, 20 outliers were generated with a uniform distribution in the range between the maximum and minimum value of the obtained data.

For this dataset, the proposed method converts the raw features into the learned metric space by learning a distance metric from the 60 training instances and uses these to detect outliers using the LOF. Moreover, to perform a comparison with the conventional method, outlier detection was also carried out using the LOF and one-class SVM.

In the proposed method, the highest 20 data in terms of LOF values were extracted as candidate outliers, and in the one-class SVM, 30 data were extracted as candidate outliers, as shown in Figure 1.

4.2. Peculiar Example Detection on Japanese Dataset

4.2.1. Data

We used the Semeval-2010 Japanese WSD task, which includes 50 target words comprising 22 nouns, 23 verbs, and 5 adjectives. Although there are 50 training and 50 test instances for each target word, we selected only 48 because some instances of the words “可能 (kanou; possible)” and “入る (hairu; enter)” have a “new word sense” tag in the training instances, bringing the total to 16 peculiar examples in this dataset: one example of the word “意味 (imi; meaning),” three examples of the word “手 (te;

hand),” seven examples of the word “前 (mae; front, before),” one example of the word “求める (motomeru; require, call in),” two examples of the word “あげる (ageru; give, get up),” and two examples of the word “始める (hajimeru; start, open up).”

4.2.2. Implementation

To implement these methods, we extracted features from the training and test data of a target word. In this method, we consider the words that appear before and after the target word and extract its word strings, part-of-speech tags, and thesaurus ID numbers—features that are mostly beneficial for high-performance supervised WSD (Shinnou and Sasaki, 2008). In the LOF and one-class SVM, peculiar examples were extracted from the whole dataset.

In the proposed method, first, distance metric learning was performed by LMNN for the training data and new training data points were generated using the distance metric in the original space. Then, peculiar examples were extracted from the updated training data and the test data using the LOF. In the LOF and the proposed method, the top 20 examples were extracted as candidate peculiar examples, as shown in Figure 2.

5. Experimental Results

As shown in Tables 1 and 2, the proposed method has the highest number of properly detected instances and the highest F-measure value. Thus, the label information of the training data is an effective tool for density-based peculiar example detection. Thus, in the experiment using the artificial dataset, the proposed method detected more outliers and achieved the higher F-measure value.

Table 3: Effect of Distance Metric Learning for Each Peculiar Example

Word	Example ID	Ranking Before	Ranking After	
意味 (imi)	31	84	12	
	手 (te)	32	14	14
		33	20	18
前 (mae)	34	37	19	
	34	54	64	
	35	51	60	
	36	53	63	
	38	9	17	
	46	50	59	
	47	59	66	
48	98	95		
求める (motomeru)	31	67	46	
あげる (ageru)	40	89	65	
	41	93	55	
始める (hajimeru)	48	59	42	
	49	57	38	

Figure 3 shows the change in the LOF values of peculiar examples before and after the implementation of metric learning. For most of the target words, the LOF value increased after learning, showing that the density change with metric learning is effective for the peculiar example detection.

For the target word “前 (mae),” all peculiar examples are represented as a clipped word for “午前 (gozen; a.m.).” In this case, the target word co-occurs with Japanese numerals such as “前十時 (10 a.m.)” in the peculiar examples. There are the words “十年前 (junenmae; decade ago)” in the examples in the training data, suggesting that the Japanese characters “(ju; ten)” and “年 (nen; year)” cause the LOF value to decrease with metric learning.

Moreover, we conducted an experiment on outlier detection using a classification method (SVM). We set three classes A, B, and C in the artificial dataset and implemented a one-versus-rest classification for each class to identify instances classified as “not A, not B, and not C” as outliers. As a result, we found three outliers correctly, while 173 instances were extracted (the F-measure value was 0.031). This experimental result shows that it is difficult to apply the classification method for outlier detection.

6. Conclusion

In this paper, we proposed a new peculiar example detection method using distance metric learning from labeled example pairs. The efficiency of the proposed method was evaluated on an artificial dataset and the Semeval-2010 Japanese WSD task dataset. The results showed that the proposed method has the highest number of properly detected instances and the highest F-measure value. This shows that the label information of training data is effective for density-based peculiar example detection. Moreover, an experiment on outlier detection using a classification method such as SVM showed that it is difficult to apply the classification method to outlier detection.

7. References

- Markus M. Breunig, Hans-Peter Kriegel, Raymond T. Ng, and Jörg Sander. 2000. Lof: identifying density-based local outliers. *SIGMOD Rec.*, 29:93–104, May.
- Chih-Chung Chang and Chih-Jen Lin. 2011. Libsvm: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.*, 2:27:1–27:27, May.
- Corinna Cortes and Vladimir Vapnik. 1995. Support-vector networks. *Machine Learning*, 20(3):273–297.
- Hans-Peter Kriegel, Peer Kröger, Erich Schubert, and Arthur Zimek. 2009. Loop: local outlier probabilities. In *Proceeding of the 18th ACM conference on Information and knowledge management, CIKM ’09*, pages 1649–1652, New York, NY, USA. ACM.
- Manabu Okumura, Kiyooki Shirai, Kanako Komiya, and Hikaru Yokono. 2010. Semeval-2010 task: Japanese wsd. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pages 69–74, Uppsala, Sweden, July. Association for Computational Linguistics.
- Gustavo H. Orair, Carlos H. C. Teixeira, Wagner Meira, Jr., Ye Wang, and Srinivasan Parthasarathy. 2010. Distance-based outlier detection: consolidation and renewed bearing. *Proc. VLDB Endow.*, 3:1469–1480, September.
- Bernhard Schölkopf, John C. Platt, John C. Shawe-Taylor, Alex J. Smola, and Robert C. Williamson. 2001. Estimating the support of a high-dimensional distribution. *Neural Comput.*, 13:1443–1471, July.
- Hiroyuki Shinnou and Minoru Sasaki. 2008. Division of example sentences based on the meaning of a target word using semi-supervised clustering. In *Proceedings of the Sixth International Language Resources and Evaluation (LREC’08)*, Marrakech, Morocco, may.
- Kilian Q. Weinberger and Lawrence K. Saul. 2009. Distance metric learning for large margin nearest neighbor classification. *The Journal of Machine Learning Research*, 10:207–244, June.