



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 4, April 2014

An Approach for Word Sense Disambiguation using modified Naïve Bayes Classifier

Abhishek Fulmari¹, Dr. Manoj B. Chandak²

M.Tech Student, Dept. of CSE, Shri Ramdeobaba College of Engineering & Management, Nagpur, India¹

Professor, Dept. of CSE, Shri Ramdeobaba College of Engineering & Management, Nagpur, India²

ABSTRACT: Word Sense Disambiguation is a difficult task to perform in Natural Language Processing and it is the process of identifying the correct sense or correct meaning about a word. This is basically used in application like information retrieval, machine translation, information extraction because of its semantics understanding. There are various methods available in supervised approach to perform the disambiguity of word. This paper describes a Naïve Bayes method of Supervised Learning Approach which is based on probabilistic method. The main emphasis of this paper is feature selection for disambiguation purpose. Here we have just tried to remove the stop words while selecting the feature set.

KEYWORDS: Word Sense Disambiguation, Supervised Learning Approach, Naïve Bayes Method, Feature Vector.

I. INTRODUCTION

In today's world, many refers web for searching purpose and sometimes it was observed that the result of search is not appropriate. The reason behind this is because of ambiguity in the words. To solve the above problem of ambiguity Word Sense Disambiguation (WSD) is introduced. WSD is an important task in Natural Language Processing. Word Sense Disambiguation is the process where the system has to find the exact meaning of word. For the application like machine translation the word should give proper meaning, then only one can say that the resulted output will be similar to that of expected output.

To solve the problem of WSD basically four approaches are available supervised learning approach, semi-supervised approach, knowledge based approach and unsupervised approach. In supervised learning approach, the dataset used for classification was already divided into classes. Feature of input are compared with the data available with each classes and according to highest value the input sentences will be classified in that class. In semi-supervised learning approach, some data are classified and some are raw data. In knowledge based approach, the data for classification are taken from resources like wordnet or semcor. The data in this resources are basically of kind like having some relationship among the word or in the form of glosses, which will help to get the sense for the target word. And in unsupervised learning approach complete dataset is of raw data. Firstly the clusters are formed by checking some similarities among raw data and then the task of classification is performed.

Supervised learning approach having different method and they are Naïve Bayes, Decision Tree, Decision List, Neural Network, Exemplar-Based learning [2] method. Naïve Bayes algorithm is based on conditional probability, which calculate probabilities for each senses and assign the sense to target word that have maximum probability value. Decision List method is based on some set of rules. The decision is taken based on if-then-else conditions. Decision Tree [4] is a method which predication based method. Here the classification rules are is of the form yes-no. Neural network is also an approach in supervised learning which is interconnection of artificial neurons.

The results from different approach had been seen and they clearly indicate that supervised learning approach perform well as compared to other approaches. The result being so because of use of already classified data for training purpose. The drawback of other approaches were like dictionary not have sufficient material to train the classifier and for unsupervised approach some of the instances of training data may not be assigned correct sense.

This paper includes Naïve Bayes method for classification. This paper is organised as related work, about Naïve Bayes algorithm, our proposed approach, future work and conclusion.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 4, April 2014

II. RELATED WORK

As according to the authors [2] explained about the feature selection for the training and testing purpose in machine learning technique. There they have told about different feature selection.

Example: “An electric guitar and **bass** player stand off to one side, not really part of scene”

Here bass is the word which has to be disambiguate. The collocation feature says that it is a collection of word from each side of the target word or collocation of part-of-speech of those words. They mentioned that if multiple feature are mixed together to form single feature they would be help in better result.

When combining the collocation of word and part of speech the resulted feature vector is
guitar, NN, and, CJC, player, NN, stand, VVB

This approach is been tested and found that the result will be better approximate to 90.4%.

Here they also explained about another type of feature selection which is co-occurrences. Co-occurrences is about counting of words which are around the target words in the context. The result of disambiguation having co-occurrence vector as feature set is about 86.13%.

In the algorithm proposed by Pederson [5] a simple Ensembles Naïve Bayes classifier is used, which showed that result of WSD can be improved by combining number of classifiers. The author tested the classifier using different sizes of left and right windows of context (0, 1, 2, 3, 4, 5, 10, 20, 50) and got accuracy of 89% and 88% on two dataset from senseval of the word interest and line respectively.

There are total six supervised methods are proposed by Mooney and it concluded that Naïve Bayes Classifier produced the best result among all and the feature set used by them are the words surrounding the ambiguous words.

III. NAÏVE BAYES ALGORITHM

Naïve Bayes[1] classifier is based on Baye’s theorem. It is simple conditional probabilistic approach to disambiguate the ambiguous word in proper sense. For the purpose of disambiguation it is considered that all the feature selected for the classification are independent of each other. Here probability for each feature as an individual is calculated for a class (sense) and finally the product of them is taken. This product represents a probability of occurrences of target word in that sense.

This algorithm is worked as first applying the bayes theorem

$$P(\text{sense}|\text{feature}) = \frac{P(\text{sense}) * P(\text{feature}|\text{sense})}{P(\text{feature})}$$

Then applying the naïve assumption, the final equation reduce to:

$$\operatorname{argmax}_{S_i \in \text{Senses}_D(w)} P(S_i) \prod_{j=1}^m P(f_j | S_i),$$

Here f_j represent the feature of feature vector and S_i represent the sense of word.

The sense which give the highest probability value will be consider as the correct sense for the word.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 4, April 2014

IV. PROPOSED METHOD FOR DISAMBIGUATION HAVING DIFFERENCE IN FEATURE SELECTION

In this paper, we tried to perform the disambiguation using the Naïve Bayes method, but with different feature set. As mentioned above, in most of the papers it has been seen that feature vector are formed by taking the words around the ambiguous word along with their part-of-speech tags according to the window size. When taking the surrounding words there are words which are not that much useful such as of, is, not, we, I, etc.

Here we tried to create a feature set which will not include such words and then performing the process of disambiguation. So when creating feature set firstly such stop words are removed and then feature set is created, which train the classifier and then new sentences are tested.

In this approach the feature set extracted will be
electric, JJ, guitar, NN, player, NN, stand, VVB

in the above example the word 'and' was not considered in the feature set, as it is a stop word we removed it and then created a feature set. Instead of selecting the word 'and' in the feature set, the word next to it in the sequence was selected in order to get more useful or meaningful related word in context of target word.

A. Algorithm for Disambiguation:

```
//training
1. From all sentences extract features.
2. Train the classifier with the features extracted above.
//calculation of prior probabilities
3. Calculating the prior probabilities.
//disambiguation
4. Select some words around the target word.
5. Compare these words with the word of feature set in classes.
6. Calculating the probabilities for each sense.
//assignment of sense
7. Comparing the result of each sense and assign sense of maximum sense.
```

The algorithm works as firstly the classifier was trained with some data. Means forming classes of different senses having similar set of features in it. After feature extraction part, the classifier first provide the prior probabilities of all the classes which was calculated as number of training sample of same class divide by the total number of training samples.

The disambiguation part work as first we select the same feature from the testing sample and then start counting that how much time that feature appears in training set. And calculate the probability for the same as number of time that feature appear in set of particular class (sense) divided by the number of training sample in the class (sense). Such probability values of feature of testing sample are then multiplied together to get the actual value for a particular class (sense). After performing the same operation with different class (sense) and getting some values, these are compared with each other to know that in which class (sense) the testing sample is classified.

V. FUTURE WORK

This approach is just an attempt to improve the efficiency of classification with the accuracy more the 89% that was of Naïve Bayes algorithm proposed by Pederson. This approach is yet to be tested on dataset of Senseval 2 and SemCor 3.0. Will try to find the accuracy by changing the window size of feature vector. The pseudo-code is under development for above algorithm.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 4, April 2014

VI. CONCLUSION

From the result of previous approaches we come to know that the accuracy was about 89%, but while selecting the feature set for classification they includes each and every words with their part-of speech according to the window size. Also from the above experiment perform we have come to know that by combining the feature set will improve the result of classification. So, here we try to remove the stop words such as is, or, not, I, an, was, etc. and try to improve the overall result of the disambiguation algorithm.

REFERENCES.

1. CuongAnh Le and Akira Shimazu, 'High WSD accuracy using Naive Bayesian classifier with rich features', Japan Advanced Institute of Science and Technology (JAIST), pp.105-114, 2004.
2. Gerard Escudero, Lluís M'arquez and German Rigau, 'Naïve Bayes and Exemplar-based approaches to Word Sense Disambiguation Revisited', arXiv:CS/0007011v1, 2000.
3. Abhishek Fulmari and Manoj B. Chandak, 'A Survey on Supervised Learning for Word Sense Disambiguation', International Journal of Advanced Research in Computer & Communication Engineering, Vol 2, Issue 12, pp. 4667-4670.
4. L. Y. Keok., 'An empirical evaluation of knowledge sources and learning algorithms for word sense disambiguation', EMNLP, Philadelphia, pp 41-48, 2002.
5. Pedersen, T., 'A Simple Approach to Building Ensembles of Naive Bayesian Classifiers for Word Sense Disambiguation', Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL), pp. 63-69, 2000.
6. Mooney, R. J., 'Comparative Experiments on Disambiguating Word Senses: An illustration of the role of bias in machine learning', Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 82-91, 1996.
7. Navigli, Roberto, 'Word Sense Disambiguation: a Survey', ACM Computing Surveys, 41(2), ACM Press, pp. 1-69, 2009.