Word Sense Disambiguation using the EDR concept Dictionary and the Corpus

Tsunekazu Yagi (210096)

School of Information Science,
Japan Advanced Institute of Science and Technology

August 13, 2004

Keywords: Word Sense Disambiguation, Supervised Machine Learning, Definition Sentences in a Dictionary, Hypernym, Robustness.

It is important problem of the natural language processing has a meaning of a word sense disambiguation(WSD). Supervised learning is often performed as past research of word sense disambiguation. However,supervised learning is difficult to learn about a low frequency word, although a good result is obtained about the word which needs data with a correct sense and appears frequerntly in data. Although this research is premised on use by a document reading support system which supports text understanding of human. In such a case, a meaning of a word sense disambiguation to many words including the low frequency word is required for the technique of perform. In order to solve this problem, the technique of creating the classifier using the corpus with a word sense tag and the dictionary definition sentence is proposed. And the technique of supervised machine learning is used for a high frequency word, and the classifier which used the definition sentence of a dictionary is used for a low frequency word. The recall and accuracy in improvement of a WSD system are performed by two classification combining.

First, the classifier using the dictionary definition sentence is explained. For example, there is the word a “dog” and suppose that there are two sense “the men of the role of a spy” and ”an animal called a dog”. The case considerd the meaning of the “dog” of the sentence “He feeds his dog”

is judged. However, when a "dog" does not appear in a corpus with a word sense tag, the model which judges the meaning of a "dog" cannot be learned. This paper pays attention to the hypernym of the meaning of a word contained in a dictionary definition sentence. The hypernym of the meaning of a word can be taken out from the end of a dictionary definition sentence in Japanese. For example, hypernym of an "animal" can be taken out from "an animal called a dog." Here, "He feeds his elephant" and "He feeds his cat", is in the corpus, and the hypernym of an elephant and the hypernym of a cat presupposes that it is an "animal". Since it can learn that a Hypernym "animal" and "feed" coincide at this time, in the sentence "He feeds his dog", it turns out that the meaning of a "dog" is "an animal called a dog." Thus, the meaning of a word can be correctly judged also in a low frequency word by extracting the hypernym the meaning of a word from a dictionary definition sentence, and learning the collocation statistics of hypernym and a word near a forget word.

Next, the technique of performing a word sense disambiguation using a hypernym is explained. Here, the Naive Bayes model $P(s)\prod_i P(f_i|c)$ which used the hypernym is learned, and the meaning of a word from which the probability serves as the maximum is chosen. In this formula, c is a hypernym. s is the meaing of word. f_i is feature. Moreover, a surface form and a part of speech of the word which exists just before and after the word (w) which wants to decide the meaning of a word, the basic form of an content word that appears within 20 words before and after w, the basic form of the word sytactically related with w, etc. are used as feature f_i.

Next, the technique of extracting a hypernym from a dictionary definition sentence is described. Fundamentally, the word in the end of a dictionary definition sentence is taken out as a hypernym. Moreover, a hypernym may not be a word in the end of a dictionary definition sentence. The extraction pattern was created in consideration of such a case. For example, the pattern was created that takes out N (ex. "悪口 ") from the dictionary definition sentence "Nのこと"(ex."欠点を探し出していう悪口のこと") The sixty four extraction patterns were created and the a hypernym was extracted from the dictionary definition sentence of the EDR concept dictionary. The value which divided the number of the meanings

of a word which has extracted the hypernym by all the numbers of the meanings of a word in the EDR concept dictionary was 98%. Moreover, when 200 extracted a hypernym were ramdomly chosen and having been judged with the help, 185 hypernyms were suitable as a hypernym.

Next, the technique of the supervised machine learning which performs WSD of a high frequency word is explained. Here, Support Vector Machine (SVM) is used as supervised machine learning algorithm. The surface form of two words which appear as a feature used by SVM just before the word in front of w or after two or a part of speech, and w or in immediately after in addition to the feature used by the Naive Bayes model, or the group of a part of speech was added.

Finally in this research, two classifiers, the classifier by SVM for a high frequency word and the classifier using the hypernym for a low frequency word, were combined. The technique to combine is as follows. If it is beyond a threshold with the frequency of appearance in the training data of the word which wants to decide the meaning of a word, the classifier by SVM is distinguished using a Naive Bayes model except it. This threshold was set to 5 in this research.

The experiment which finally evaluates the proposed method was conducted. SVM, NB, BL (baseline model), SVM+BL (SVM and baseline model), and SVM+NB (the proposed method) were compared. Precision is inferior, although SVM+NB will exceed SVM in recall, F-measure, and applicability, if highest SVM of the three single classifier is compared with SVM+NB. Improvement in recall or applicability is especially great. It is because SVM+NB is outputting the meaning of a word also to a low frequency word to the classifier by SVM aimed only at the high frequency word in a corpus with a word sense tag. Word sense disambiguation is performed increases the word, it follow that recall and the applicability improved. On the other hand, there was no significant difference between SVM+NB and SVM+BL. SVM+NB far exceeded SVM+BL for both like the word with low frequency for a with a frequency of 20 or less word . Therefore, it became clear that the proposed method is effective in the word sense disambiguation of a low frequency word.