

Title	構文構造を用いた感情極性分類の精度向上
Author(s)	中山, 貴樹
Citation	
Issue Date	2012-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/10421
Rights	
Description	Supervisor: 東条敏, 情報科学研究科, 修士

Improvement of accuracy for sentiment classification with syntactic structure

Yoshiki Nakayama (1010045)

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

February 6, 2010

Keywords: machine learning, clustering, sentiment classification, text classification.

Emotion Information Processing is a technique that aimed to objectively evaluate the human emotion by analyzing of text and conducting a comprehensive research. The sentiment classification is the technique that automatically determines whether the human written text is written in positive way or negative way. It may be thought as negative if it has a word such as ‘ugly’ in context. Positive or negative word can be used to the sentiment classification.

In order to automatically classify the human written text as positive or negative, the text classification will use Bag-of-Words approach. Bag-of-Words approaches use the dimension of the feature space that is equal to the number of different words in all of the documents. A common example is the the sentiment classification use the bag-of-words model. The model is at machine learning approach which uses words in the context as features.

However, It is difficult to increase the accuracy with the bag-of-words method. In most cases, several words are written in a positive or negative text which leads difficulty of measuring accuracy of the semantic orientation. In addition, some word can invert the orientation.

For example, the semantic orientation of multiple words namely ‘not

beautiful' could be divided into 'not' and 'beautiful'. 'Beautiful' is positive orientation. While 'not beautiful' is counted as negative orientation. Conjunction is one of the important factors of a whole text. Conjunction is part of parse that around the context. By conjunction, a text takes a hold to reverse semantic orientation, said to be important for the sentiment classification.

In this paper, we focus on the modification relation on the syntactic structure. Word of combination, the combination of the semantic orientation word and specific word, is used in machine learning on the assumption of reversing semantic orientation word. By learning the word of combination, the presence or absence of reverse of semantic orientation words could be known. In addition, by using the differences of the Bag-of-Words models extracted and the word of combination, it is possible to measure the total number of words in the original semantic orientation words of the text. For the text written by the user, by using the Cabocha, to analyze the syntactic structure of the text. The results of that analysis have been used to extract a set of the semantic orientation word and specifics word. The training data was trained by Support Vector Machine learning. Then, this adapts the text to be used in the sentiment classification, it is possible to extract the number of semantic orientation words in each reversed text. The results of acting on the total number of semantic orientation words was extracted by using Bag-of-Word model. We can assume the total number of semantic orientation words could be found in the text's original. And, the following techniques were used in our experiment. The First technique is N-gram model. Second is counting the total number of positive and negative words. The training data was implemented by combining this total number with N-gram model. Finally, a training data combined with the total number of semantic orientation words by modification relation with N-gram model. Support vector machine for these data was used to execute the classification of semantic orientations. The sentiment classification were executed using these methods of the reviews. The experimental results in the method showed high classification accuracy for training data including a combination of bigram and modification relation. As a results show it is effective to use modification relation to improve the sentiment classification.