様式 8

<center>論　文　内　容　要　旨</center>

報告番号	甲　先　第　　２６５　　号	氏　　名	Wang Lei

学位論文題目	A Study on Multi-label Sentiment Analysis for Chinese Text 中国語テキストにおけるマルチラベル感情分析に関する研究

内容要旨

With rapid development of the Internet, a great amount of information full of personal subjective emotions appears on the Internet, such as microblogs, news comments and product reviews. Sentiment orientation of text is studied for the purpose of exploring people's attitudes and emotional states toward entities or their attributes in the texts. Research on sentiment orientation of text has become a hot research topic in the field of natural language processing and research achievements have been widely used in many social fields, including politics, economics and management.

Over the past few years, considerable progress has been made in sentiment analysis of text. However, many difficulties and challenges have been posed on text sentiment analysis because of complicated human emotions, diversified contents and forms of texts. As a new research method in the field of artificial intelligence, granular computing is mainly used for processing a huge amount of uncertain, fuzzy and imprecise information. The basic notions and principles of granular computing is to find an ideal solution by decomposing a complex problem and analyzing data at different levels of granularity. In this paper, an attempt is made to apply the idea of granular computing to explore multi-label sentiment orientation of Chinese texts. Based on differences in linguistic granularity, research on sentiment orientation of text is studied from the perspective of word, sentence and document. Several methods of multi-label text sentiment analysis based on granular computing are put forward and a brand-new interpretation on sentiment orientation of text is given from uncertainty of sentiment orientation. The main contributions of this thesis are summarized as follows:

(1) In light of complex and diversified human emotions, multi-label sentiment orientation of Chinese text is studied. In the past, most researchers only examined complimentary and derogatory emotions on the level of words, sentences or documents. In this paper, eight categories of multi-label sentiment orientation are studied from the perspective of words, sentences and documents

in the Ren-CECps, in order to be more in line with abundant and complex human emotions in practice.

(2) Considering that sentiment orientation of words is influenced by the context, a method of word sentiment orientation based on maximum entropy is proposed. First of all, features, relationships, semantic characteristics and emotional relations of words are extracted from pertinent contexts. Subsequently, sentiment orientation of words is identified by maximum entropy model, and problems about sparse features are solved by smoothing techniques. At last, uncertainties of sentiment orientation of words are further removed based on emotional connections between words and sentences. Experiments on the Ren-CECps Chinese emotion corpus have achieved good results. The emotional relationship between words and sentences also helps to improve the accuracy of recognizing sentiment orientation of words.

(3) Viewing that a sentence is made up of diversified elements, a method is put forward to analyze sentiment orientation of sentences by integrating multiple features. Although affective words are important factors that determine sentiment orientation of sentences, sentiment orientation of sentences is affected by relationships between the sentences of a text. At first, intensity of sentiment orientation is regulated for affective words of a sentence pursuant to topical features. Besides, sentiment orientation of sentences is judged in combination with other semantic features of negations, conjunctions and adverbs of degree in a sentence. The experimental results show that semantic features in a sentences can play a positive role in the recognition of sentiment orientation of sentences, which helps to improve the accuracy of recognizing sentiment orientation of sentences.

(4) By introducing topic features into sentiment orientation of sentences, a semi-supervised multi-label emotion topic model is put forward. A sentiment topic layer is incorporated into latent Dirichlet allocation that calculates probability distribution of topics-word, document-topics and word-sentence sentiment, so as to discriminate sentiment orientation of sentences. By comparing experiments, it is proved that the method can effectively discriminate the sentiment orientation of sentences, which can improve the accuracy of recognizing sentiment orientation of sentences.

(5) In view of general and complex sentiment in document, a method is proposed

to analyze sentiment orientation of documents based on three-way decisions. Relatively great errors may be caused by directly identifying sentiment orientation of documents by words or sentences, so the method covering multiple steps and stages is adopted. Theory of three-way decision is introduced into sentiment orientation analysis in the paper. In this paper, decision-making process of sentiment orientation is divided into two stages with several steps. Firstly, a sentiment lexicon is composed by applying training data sets. Meanwhile, sentiment orientation of affective words and its intensity in testing data are identified using Tongyici Cilin and HowNet. Then, sentiment orientation of documents is preliminarily identified according to intensity of sentiment orientation of affective words by which document is then divided into positive, negative and boundary regions with appropriate thresholds. At last, sentiment orientation of documents of delayed decisions are discriminated pursuant to emotional characteristics of pertinent sentences. The experimental results show that the method can improve the accuracy of recognizing the multi-label sentiment orientation and eight basic sentiment orientation of documents.