Chapter 48

Chinese Text Sentiment Analysis Utilizing Emotion Degree Lexicon and Fuzzy Semantic Model

Xing Wu
Shanghai University, China

Shaojian Zhuo
Shanghai University, China

ABSTRACT

Text on the web has become a valuable source for mining and analyzing user opinions on any topic. Non-native English speakers heavily support the growing use of Network media especially in Chinese. Many sentiment analysis studies have shown that a polarity lexicon can effectively improve the classification consequences. Social media, where users spontaneously generated content have become important materials for tracking people's opinions and sentiments. Meanwhile, the mathematical models of fuzzy semantics have provided a formal explanation for the fuzzy nature of human language processing. This paper investigated the limitations of traditional sentiment analysis approaches and proposed an effective Chinese sentiment analysis approach based on emotion degree lexicon. Inspired by various social cognitive theories, basic emotion value lexicon and social evidence lexicon were combined to improve sentiment analysis consequences. By using the composite lexicon and fuzzy semantic model, this new sentiment analysis approach obtains significant improvement in Chinese text.

INTRODUCTION

Text sentiment analysis technology has been applied to many fields. For example, Pulse, a business intelligence system developed by Microsoft can extract the user view from comment text by using the text clustering technology. Opinion Observer system can comment on the subjective content from customer reviews on the Internet, and extract product features and consumers reviews. Many current sentiment

analysis approaches are mainly focus on emotion tendency analysis. Texts are usually classified into three categories: positive text, negative text and neutral text. The technology of text sentiment analysis generally consists of subjective classification, emotion polarity, semantic orientation, opinion mining, opinion extraction, emotion analysis and emotion summarization.

Text sentiment analysis is applied to find user reviews and emotion polarity from text. User reviews can help users make decisions or get the product feedback. It can also make prediction for political elections and some other major events. Otherwise, the technology of sentiment analysis also contributes to the research of other natural language processing fields. In sentiment analysis fields, there are two widely used approaches: combining rules with emotional dictionary and machine learning technology. In combining rules technology, texts are classified by using positive emotional words and negative emotional words. Machine learning often uses Naive Bayes, Max Entropy or Support Vector Machine to classify texts. Most sentiment analysis researches are concentrating on English texts. Recent studies have shown that non-native English speakers heavily support the growing use of Network media. And, Chinese text is growing fast at Internet in recent years. But there does not exist an effective approach in Chinese text sentiment analysis.

Fuzzy semantics comprehension plays a crucial role in thinking, perception and problem solving. The fuzzy nature of linguistic semantics stems from inherent semantic ambiguity, context variability, and individual perceptions. Almost all problems in natural language processing and semantic analyses are constrained by these fundamental issues. The mathematical structure of fuzzy concepts and fuzzy semantics enables cognitive machines and fuzzy systems to mimic the human fuzzy inference mechanisms in cognitive linguistics, fuzzy systems, cognitive computing, and computational intelligence. Yingxu W demonstrated that fuzzy semantic comprehension is a deductive process, where complex fuzzy semantics can be formally expressed by algebraic operations on elementary ones with fuzzy modifiers.

Natural language processing research, in fact, primarily depends on the availability of resources like lexicons and corpora. These utilities are still very limited for Chinese text sentiment analysis. Cambria E, et al. developed a Chinese common and common sense knowledge for sentiment analysis by blending the largest existing taxonomy of English common knowledge. By using machine translation techniques, they can effectively translate its content into Chinese. However, English grammar is different from Chinese. Chinese grammar is enormous, so the approach proposed by Cambria E can't effectively classify Chinese text. The polarity of the sentiment words which are not in lexicon cannot be calculated and classified effectively by lexicon-based classifiers. Miao Y proposed the EM-SO algorithm based on expectation maximization model for constructing and updating sentiment lexicon. Experiments showed that the EM-SO algorithm and designed components outperform SO-CAL for the calculation performance of the polarity and strength of sentiment words on review sets. Another model called Chinese sentiment expression model has been proposed by Liang Y. It can effectively improve the accuracy of emotion classification.

Emotion degree lexicon composes of emotional word and corresponding metric value. Each word of the lexicon works as a basic semantic unit of linguistics. Present emotion degree lexicon usually derived from artificial marking which is hard to extensible and low reliability. Social cognition theories provide a strong theoretical basis to construct an emotion lexicon.

Related Content

Evaluation Challenges for Computer-Aided Diagnostic Characterization: Shape Disagreements in the Lung Image Database Consortium Pulmonary Nodule Dataset
William H. Horsthemke, Daniela S. Raicu, Jacob D. Furstand Samuel G. Armato (2013). *Data Mining: Concepts, Methodologies, Tools, and Applications (pp. 1794-1818).*
www.irma-international.org/chapter/evaluation-challenges-computer-aided-diagnostic/73523

Big Data and Data Modelling for Manufacturing Information Systems
Norman Gwangwava, Khumbulani Mpofuand Samson Mhlanga (2016). *Big Data: Concepts, Methodologies, Tools, and Applications (pp. 116-138).*
www.irma-international.org/chapter/big-data-and-data-modelling-for-manufacturing-information-systems/150162

A Predictive Modeling of Retail Satisfaction: A Data Mining Approach to Retail Service Industry
M. Hemalatha (2013). *Data Mining: Concepts, Methodologies, Tools, and Applications (pp. 1276-1290).*
www.irma-international.org/chapter/predictive-modeling-retail-satisfaction/73495

Big Data Computing and the Reference Architecture
M. Baby Nirmalaand Pethuru Raj (2016). *Big Data: Concepts, Methodologies, Tools, and Applications (pp. 56-72).*
www.irma-international.org/chapter/big-data-computing-and-the-reference-architecture/150158

A BPMN-Based Design and Maintenance Framework for ETL Processes
Zineb El Akkaoui, Esteban Zimányi, Jose-Norberto Mazónand Juan Trujillo (2013). *International Journal of Data Warehousing and Mining (pp. 46-72).*
www.irma-international.org/article/bpmn-based-design-maintenance-framework/78375