Summary of Sentiment Analysis of Online Product Reviews

Chaoyuan Guo*

Dalian university of technology, Dalian 116000, China.

*gcy0610@mail.dlut.edu.cn

Abstract

In e-commerce services, the influence of online comments of users on consumers' purchase decisions and business behaviors is increasingly prominent. How to use this important online text data to mine users' emotional tendencies has become the focus of academic and industry attention. After sorting out the literature, this paper combs out a fine-grained hierarchical emotional analysis business process of online comments, based on this process, the research and development status in this field is analyzed, which provides reference for future research.

Keywords

Sentiment Analysis; Online Shopping; Product Reviews.

1. Introduction

With the rapid development of e-commerce platform for online shopping, more and more consumers share their views on products online. These product reviews express the customers' attitude and evaluation of a large amount of emotional information contained in product performance or quality, and have an significant impact on customers' cognition and emotion. How to identify the praise and criticism of users' comment information, so as to present users' attitude towards products more intuitively. Emotional analysis(SA) is required.

Emotional analysis of product reviews is a computing technology that automatically extracts subjective information, that is, customers' opinions on products. It came into being for quickly processing massive data.

Traditional statistical classification methods can't deal with a large amount of product review information, so automatic and intelligent methods are needed to analyze the review information. Using certain processing methods to mine and analyze these product reviews, the data obtained can help consumers realize the product quality and other information. Besides, it can also help manufacturers to improve the quality of product. Merchants are able to adjust sales strategies and improve service quality according to product evaluation, which is of great importance to e-commerce, manufacturers and potential consumers.

2. Data preprocessing

Review data obtained by researchers from third-party platforms such as ecommerce websites should go through preprocessing stages, including word splitting, part-of-speech tagging, linguistic feature processing and so on. Tools such as NLTK, SpaCy, StanfordCoreNLP are mainly used to split words. Part-of speech tagging is the process of determining whether each word in the sentence is a noun, verb, adjective or other part of speech. Linguistic feature processing refers to the further improvement of product reviews after word segmentation and part-of-speech tagging according to linguistic features. It solves the individual product features that cannot be identified by word segmentation tools.

3. Attribute extraction

After data preprocessing, it is necessary to further extract product attributes from the preprocessing results, which are often the object of user evaluation in comment texts. For example, "the style of this skirt is very good" and "the fabric of the skirt is great". The product attributes in these two comments are "style" and "fabric" respectively. The task of attribute extraction is to extract such attributes from online comment texts.

There are two forms of product attributes, one is explicit attribute, which can be obtained directly from the text. The other is implicit attribute, which depends on reference analysis of the context of the current statement.

3.1. Explicit attribute extraction

Explicit attribute extraction tasks can be divided into supervised learning methods, unsupervised learning methods and semi-supervised learning methods according to different learning methods.

3.1.1. Supervised learning method

Supervised learning method uses labeled sample data sets to learn and complete the extraction of explicit attributes. At present, the widely used methods include Naive Bayesian Model, Support Vector Machine, Gaussian Mixture Model, Artificial Neural Network and so on.

Jin et al. [1] put forward a HMM model which is incorporated into vocabulary, which extracts opinions from product reviews. Firstly, a word set composed of different words and their corresponding parts of speech is established to manually mark the aspects and opinions of the review text, and then the marked data is sent to HMM for training, and further sentences containing attributes and opinions are obtained, considering the role of negative words in them.

3.1.2. Unsupervised learning method

Unsupervised learning method uses unlabeled sample data sets for training, and commonly used unsupervised learning methods include clustering and association rules.

Samha et al. [2] extract attributes by querying the names and synonyms of related fields in WordNet dictionary. This dictionary-based method is simple and fast, but the problem of limited dictionary coverage leads to poor performance when it involves multi-domain or complex structure texts.

3.1.3. Semi-supervised learning method

Semi-supervised learning method is a combination of supervised learning and unsupervised learning, which focuses on adding unlabeled samples into supervised classification algorithm and using marked and unlabeled mixed sample data sets to achieve semi-supervised classification.

Wang et al. [3] proposed two semi-supervised models, which used seed words extracted from e-commerce product description to classify comments in advance, thus guiding the models to learn more key words.

3.2. Implicit attribute extraction

Hai et al. [4] proposed a method based on ABOOT model to extract implicit attributes. Firstly, the extracted explicit attributes are clustered to construct an aspect set, and then the semantic relation correspondence set is established by using the relationship between the aspect set and the corresponding viewpoints. Finally, the new viewpoints are located to the corresponding aspect set through the semantic relation set.

Tang et al. [5] proposed a sentiment analysis method based on special product reviews, and proved that this method can solve the problem of implicit attribute extraction. Firstly, the feature ontology is constructed, and the features are classified by using the feature ontology.

Then, the implicit attributes are identified by calculating the collocation weights of opinion words and features. Sun et al. [6] locates implicit attributes through the existing relationship between explicit attributes and opinion words. Firstly, it identifies the relationship between attributes and opinion words, then looks for implicit attributes to generate candidate sets, and further accurately filters out implicit attributes through scoring mechanism.

Xu et al. [7] uses LDA and SVM to extract implicit attributes. Firstly, explicit attributes are extracted by LDA, and two constraint and relevance prior parameters that guide the establishment of LDA are input into SVM as features, and SVM classifies and recognizes implicit attributes and explicit attributes.

4. Analysis of text emotional polarity intensity

4.1. Purpose of text emotional polarity strength analysis

The analysis of text emotional polarity strength mainly judges the subjective text emotional polarity strength, which is generally divided into five categories: strongly derogatory, generally derogatory, objectively treated, generally praised and strongly praised. Even though both comments express a favorable impression on commodities, different expressions will also express different favorable degrees.

4.2. Calculation of emotional intensity

In order to evaluate the emotional polarity of a product feature, we evaluate the emotional score of each phrase in which a feature appears and calculate its average value (formula (1)). Do this for every product feature. These characteristics are classified as positive, neutral or negative scores:

$$\text{sentiment\_score (feature)} = \frac{\sum_{p \in P} \text{sentiment\_score}(p)}{|P|} \tag{1}$$

$$\text{where } P = \{p \mid p \text{ is a phrase of review } \wedge \text{ feature in } p\}$$

5. Conclusion

In the related research, many advanced technologies of text classification have been tried, and many different features have been tested. Sentiment analysis of product reviews has developed from the simple discrimination of positive and negative polarity of text to the identification of specific emotions, from a single emotion category to the inference of multiple emotions. The future will still be a research direction full of difficulties and challenges.

The following describes the conceivable hot spots of sentiment analysis of product reviews in the future:

(1) How to solve the growing diversity of emotional research objects and the complexity of emotional tasks. With the persistent expansion of application fields, emotional objects change from favorable comments on products and services to more diverse expressions of users and topics in social media, and the research contents will change accordingly, including paying more attention to users' information and changing users' emotions in response to events in social media. Nowadays, some small samples and transfer learning in the field of machine learning may become the research direction to solve the problem.

(2) How to deal with the complicacy of emotional expression in commodity reviews. Customers' emotional expressions in comments, particularly in short commentary texts, are diversified, including straightforward expressions or rhetorical devices and even irony. The application of neural network learning technology with more intricate networks and the constructing of emotional common sense database may become the method to work out.

References

[1] Jin W, Ho H H. A novel lexicalized HMM-based learning framework for web opinion mining NOTE FROM ACM: A Joint ACM Conference Committee has determined that the authors of this article violated ACM's publication policy on simultaneous submissions. Therefore ACM has shut of [M]. 2009. PP 465-472

[2] Samha A K, Li Y, Zhang J. ASPECT-BASED OPINION EXTRACTION FROM CUSTOMER REVIEWS[J]. International Journal of Computer Science & Information Technolo, 2014.

[3] Bo W, Wang H. Bootstrapping Both Product Features and Opinion Words from Chinese Customer Reviews with Cross-Inducing 1. 2008.

[4] Zhen, Hai, Kuiyu, et al. An Association-Based Unified Framework for Mining Features and Opinion Words [J]. ACM Transactions on Intelligent Systems and Technology (TIST) -Special Section on Visual Understanding with RGB-D Sensors and Regular Papers, 2015. PP 1-21

[5] Tang X, Lan Y. Sentiment Analysis of Microblog Product Reviews Based on Feature Ontology[J]. Library and Information Service, 2016.

[6] Sun L, Sheng L, Li J Y, et al. A novel context-based implicit feature extracting method [C]// International Conference on Data Science and Advanced Analytics. 0.

[7] Hua X, Fan Z, Wei W. Implicit feature identification in Chinese reviews using explicit topic mining model [J]. Knowledge-Based Systems, 2015, 76(mar.):166-175.