

Study on Fine-Grained View Mining based on Network Subjective Text

Liping Wang^{1, a}, Yanling Liu^{2, b}

¹ Department of Scientific Research, Pingxiang College, Pingxiang, Jiangxi, 337055

² Department of Education, Pingxiang College, Pingxiang, Jiangxi, 337055

^a awlp8631@163.com, ^bYanling_LIU13@163.com,

Keywords: Fine-Grained View Mining, Network Subjective Text, Network

Abstract. With the rapid development of e-commerce and social media, Internet users can express their own subjective comments for commodities including online news current affairs, public figures and personal experience, including a variety of subjects and objects. Such information includes user reviews of subjective personal subjective opinions, emotions, attitudes and opinions and the like. With the participation of users continues to increase, with Web as a medium of subjective comment text information in soaring. Faced with a flood of Web comment text, how specific information needs for effective user subjective opinion mining and analysis, which reviews Web text mining observations, increasingly become a hotspot intelligent information processing, data mining and computational linguistics and other fields . Web Comments Text opinion mining technology has important theoretical and practical value, it can be widely used in information retrieval, business intelligence, social public opinion analysis.

Introduction

With the development of Internet and Web technology, the Internet has become an important way to release information, exchange of ideas and dissemination of ideas. In recent years, with the rapid development of e-commerce and social media, Internet users can freely against including online merchandise, news current affairs, public figures and personal experience, including a variety of subjects and objects to express their views and comments. It generated a lot of users that contains subjective opinions, attitudes, feelings and opinions of the comment information on the Web. These comments usually in the form of text information appear, people expressed a variety of emotional color and emotional tendencies, such as joy, anger, sadness and joy, criticism and praise and so on. Potential users can come to understand public opinion for an event or product by viewing these views have a subjective comment. However, faced with the growing proliferation of Web commentary text, only to rely on artificial methods are difficult to collect and process massive amounts of data. Therefore, how specific information needs for effective user subjective opinion analysis and opinion mining, the current Internet intelligent information processing, data mining, computing a hot research question linguistics and other fields. Comment text Opinion Mining important theoretical and practical value it has, can be widely applied to product reviews mining, business intelligence aspects, social public opinion analysis.

Sequence Labeling Element Extraction

In recent years, with the wide application of the Internet, the emergence of a large number of goods and services for the user reviews information on the Web. These comments contain information on all aspects of the product user or topic of a detailed analysis and evaluation of the different aspects of the user may make different evaluation. The overall positive comments and the sentence does not represent all the characteristics of the user reviews are positive about objects; Similarly, negative comments and sentences may also contain user for a positive attitude aspects [2]. Simply text of these comments coarse-grained primary objectivity and emotional polarity classification can not meet people's information needs. More and more e-commerce businesses to encourage consumer goods and services in all aspects of careful evaluation, consumers want to be able to understand in

detail the advantages and disadvantages of the various aspects of comparison goods, in order to make more rational purchase decision. Therefore, for the fine-grained observations comment text mining text views gradually become a hot issue, according to research dug. With traditional coarse-grained document level sentiment classification task is different, fine-grained Opinion Mining emotional information unit designed to extract valuable data from subjective texts, mainly related to the views and match elements modified relationship recognition, and emotional bias calculations. Documentation for the current level of coarse-grained and sentence sentiment classification technology is relatively mature, but the study of fine-grained views of extraction and classification techniques failed to achieve satisfactory results. In the past few years, many studies have tried to clearly define and solve this problem. Usually, however, it has a user comment great freedom and randomness, and the expression of different user habits is not the same, for example, in the digital camera field, different users may respectively "picture quality" and "image clarity" represents "photo" feature. Therefore, the user reviews the text presents the characteristics of randomness and uncertainty high. And traditional information extraction and analysis tasks compared to mining for fine-grained observations Web comment text study with greater difficulties.

Constraint-Based Evaluation Target Spectral Clustering

Evaluation objectives Clustering is one of the core user comments mining research tasks aimed at users post comments subjective evaluation objectives are classified according to the characteristics or attributes. Evaluation objectives typically include evaluation of object features, attributes and components, as well as parts of features and attributes. However, due to the different user reviews randomness and expression habits, characteristics or attributes for the same object, different commentators often describe different ways. For example, in the field of digital camera reviews, "photo", "photograph", "image" and "picture" are all used to describe the characteristics of the photo. These different but related way to describe collectively referred to as the field of synonyms, related similarity between them depends on the specific areas. For example, in the field of movie reviews, "movie" and "picture" is used to describe the film property; and in the digital camera field comment, "picture" and "photo", "movie" and "video" are used to describe the photos and video feature. Figure 3.1 depicts a schematic diagram of clustering evaluation objectives in the field of digital camera review. Clustering has become the basis of objective evaluation summary of the comments and recommendations of the feature-based. Now, with the increasing size of text comments, evaluate the number of targets to extract significant growth, using manual methods to classify their time and effort. In addition, the growing diversity of the comment field, but also makes the Rapid effective evaluation target clustering algorithm becomes a challenging problem.

Evaluation of the current research objectives are divided into clusters unsupervised learning methods and topic-based modeling-based approach. Unsupervised learning methods by calculating the similarity between the evaluation target, using clustering or classification algorithm for evaluation of the target set for automatic collation. How to calculate where the target based on the similarity between the evaluation of the key areas of impact performance of the algorithm. Similarity calculation based dictionary or glossary is divided into knowledge-based methods and areas of corpus approach. Use of dictionary-based approach lexical semantic structure artificially constructed to calculate the similarity between the evaluation target, but ignores the field dependence of the similarity evaluation objectives, but also have the problem of insufficient coverage. Use evaluation target domain corpus-based approach contextual information to calculate the similarity, although to some extent alleviated the problem of dependency evaluation target areas of similarity, but faced with the high-dimensional space is calculated and sparsity. Topic-based modeling method on the comments corpus statistical learning, training theme generation model, will generate a corresponding theme for the evaluation of the target category. However, such a method of generating themes coarse-grained and fine-grained not correspond well to the characteristics of the class. While these topics are usually a single word as a unit, the phrase can not be expressed in the form of evaluation objectives.

In response to these problems that exist in the current study, this chapter spectral clustering

method based on objective evaluation of Constraints. The method by lexical constraints between the definition and evaluation objectives and context constraints extract knowledge to enhance the correlation between the areas of evaluation objectives; the use of constraint propagation algorithm, the relationship between the local constraint propagation evaluation target to evaluate the global target space; using constrained spectral clustering algorithm, while incorporating a priori knowledge constraints, effectively reducing high-dimensional and sparse cluster of space. This chapter also discusses the contextual features further evaluation objectives extraction strategies, analyze the context and dependency syntax window adjacent to the context of the impact evaluation objectives were clustering.

Propagation Automated Build Tag based on Field Emotion Dictionary Constraint

Emotional text mining, related to words, phrases / structure, sentences, paragraphs and chapters and other languages granularity. Words and phrases which constitute the basic unit of the text, the text also contains subjective feelings with words as the basic unit of emotional expression. Words and phrases are also the expression of human emotions, attitudes, opinions basic elements. Such as "wonderful", "excellent", "ugly", "lazy" and other such words with obvious emotion tendentious called emotional words or word polarity. Emotion dictionary is an important basis for automated sentiment analysis studies [23, 79, 129-132], which contains a series of emotional words / phrases and their corresponding prior emotional tendencies. However, the text of the sentiment is not entirely dependent on the sentiment contained emotion words. For example, positive emotion words can appear in the text of the negative comments, negative emotional words can also appear in the text of positive comments. At the same time for different application areas, sentiment emotion words is not fixed, but depends on the specific areas and Context arise. For example, the word in the context of having independence sentiment may appear as a neutral word in a specific context, positive emotional words can express negative feelings in a particular context, the negative emotions in a particular word may Context express positive emotions. For example, "big" when used to describe the vehicle interior space, the expression of positive sentiment; while indicating the volume of mobile phone batteries, then the expression of negative emotional tendencies. In short, the context and the field dependence of emotional words, it is impossible to construct a complete emotional applies to all areas of the dictionary. In addition, with the participation of Internet users continues to increase, themes and areas of interest to the user increasing, build specific emotional dictionary becomes more unrealistic by experts in the field using manual methods for each of the comment field. Therefore, the target areas, Automatic / semi-automatic emotional dictionary construction method has become one of the basic tasks of fine-grained views of mining. By definition, the field of emotional dictionaries usually contain a series of specific areas of subjectivity words and phrases, as well as they have in the target areas of emotional bias. Currently emotion dictionary constructing related research work is mainly divided into a rule-based approach vocabulary knowledge base and statistical methods based on large-scale corpus. Rule-based method of vocabulary knowledge using semantic relationships that exist in the known vocabulary knowledge base entries between words and annotation information determination emotional polarity; corpus-based statistical methods between the expression of the same propensity to emotional words to each other based on common now the assumption that the use of large-scale field comments corpus statistical information calculating co-occurrence words the emotional polarity. Typically, the first specified by the user a few common emotional words as the seed, were given a positive or negative bias, then calculate sentiment candidate word based on the strength of semantic association between the words and the seed, and build emotional dictionary. However, the shortcomings of the current approach is that these studies: usually entirely dependent on the vocabulary knowledge-based approach priori semantic dictionary, can not be applied to the lack of such transcendental dictionary locale, while ignoring the dictionary has a strong emotional field dependence. The corpus based on statistical methods, while having a good field-dependent, but ignored the knowledge a priori constraints on the context and semantic level exists.

Conclusion

With the explosive growth of the Web comment text, text Opinion Mining has become intelligent information processing, data mining, computing a hot research topic in the field of linguistics. And traditional information extraction, text categorization and automatic summarization technology as compared to the text closer to the views of mining artificial intelligence targets, and can be combined with the above technique, widely used in a variety of intelligent applications in the Internet. By digging comment text subjective information, objective things quickly understand the people and themes of attitudes and opinions, provides us with automatic analysis and understanding of the Internet community intelligent way. Comment Text opinion mining technology has great social and economic value, have been widely used in the field of e-commerce, business intelligence, public opinion analysis and business management.

References

- [1] Moghaddam S, Ester M. Opinion Mining in Online Reviews: Recent Trends. Simon Fraser University Tutorial at 2013.
- [2] Hatzivassiloglou V, McKeown KR. Predicting the semantic orientation of adjectives[C]. Proceedings of the eighth conference on European chapter of the Association for Computational Linguistics, Madrid, Spain. Association for Computational Linguistics, 1997: 174-181.
- [3] Wiebe JM, Bruce RF, O'Hara TP. Development and use of a gold-standard data set for subjectivity classifications [C]. Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics. Association for Computational Linguistics, 1999: 246-253.
- [4] Hatzivassiloglou V, Wiebe JM. Effects of adjective orientation and gradability on sentence subjectivity[C]. Proceedings of the 18th conference on Computational linguistics - Volume 1, Saarbrücken, Germany. Association for Computational Linguistics, 2000: 299-305.
- [5] Wiebe J. Instructions for annotating opinions in newspaper articles. Department of Computer Science Technical Report TR-02-101, University of Pittsburgh, 2002.