Knowledge-Based Systems 000 (2016) 1–13

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Game theory based emotional evolution analysis for chinese online reviews

Zhan Bu[a,*], Huijia Li[b], Jie Cao[a], Zhiang Wu[a], Lu Zhang[a]

[a] Jiangsu Provincial Key Laboratory of E-Business, Nanjing University of Finance and Economics, Nanjing 210003, China
[b] School of Management Science and Engineering, Central University of Finance and Economics, Beijing 100080, China

ARTICLE INFO

Article history:
Received 12 October 2015
Revised 25 February 2016
Accepted 24 March 2016
Available online xxx

Keywords:
Online reviews
Affective computing
Game theory
Emotional evolution
Pairwise happiness-popularity Coordination

ABSTRACT

Sentiment analysis has become one of the mainstream researches in social network analysis. Its impact can be seen in many practical applications, ranging from public opinion analysis to marketing of public praise and information prediction. However, most of the existing research has been performed in the sentiment classification for subjective text, the emotional evolution analysis for complex interactive text (e.g., online reviews) has not yet been thoroughly targeted by the research community. This paper is concerned on short-text Chinese online reviews collected from Tianya forum. First, an efficient affective computing framework is proposed to capture the underlying emotions of Chinese online reviews. It can accurately calculate the semantic orientation of the entire review, without requiring expensive manual labeling of seed words. As users' attitudes might influence with each other, predicting their future emotional behaviors that only relying on the emotional values of historical reviews is very one-sided. Therefore, we propose a game theory based emotional evolution prediction algorithm combining the affective computing, in which the mixed nash equilibrium strategies are calculated as the future emotional behavior of interactive users. Then, experimental results on the large-scaled review dataset are provided to demonstrate the effectiveness and accurateness of our approaches. Finally, by applying the research results on the pairwise happiness-popularity coordination evaluation, we have revealed some interesting phenomenon on the "World View" board in Tianya forum.

1. Introduction

Since the Computer-Mediated Communication (CMC) [1] first appeared two decades ago, it has changed the way we share information. Internet citizens do not only consume the available content, but in turn, actively annotate the content, and generate new pieces of information. There exist many mediums, such as blogs, forums and social networks, where users can post information, give opinions and get feedback from other users. In other words, users can play the role of writers/information-producers as well as of readers/information-consumers during their lifetime in an online community. In this context, textual affective computing and emotional evolution analysis are becoming increasingly popular topics. The problems of affective computing [2] and emotional evolution analysis [3] have slightly different notions. The former originates from machine learning and natural language processing, and aims at extracting and further quantifying users' opinions about products, movies, or other entities. Emotional evolution analysis, on the other hand, focuses on measuring the underlying emotion in complex interactive text (e.g., online reviews) and the evolution over time. Nevertheless, these two problems are similar in their essence, and fall under the scope of sentiment analysis [4,5].

The task of sentiment analysis can be roughly divided into three sub-phases: determining subjectivity [6,7], determining orientation [8,9], and determining the strength of orientation [10,11]. From the perspective of data granularity, most of the studies focus on investigating the sentiment orientations of words, phrases, and documents. In recent years, a great deal of research has been done along this line, aiming at analyzing the sentiment expressed in the static text, which are collected during a particular period [12,13], and does not consider the change in sentiment when the collection of corpus evolve over time. The problem gets even harder, when we analyze the underlying sentiments in the interactive online reviews, in such context, the mutual attitudes might be infected by each other. For example, some individuals use multiple usernames or copycat/forge other users (the so-called lurkers) to communicate with others. Those lurkers are fake identities through which members of Internet community praise or create the illusion of support for the product or one's work, pretending to be

* Corresponding author. Tel.: +8613645172764.
E-mail address: buzhan@nuaa.edu.cn (Z. Bu).

http://dx.doi.org/10.1016/j.knosys.2016.03.026
0950-7051/

a different person [14,15]. In this respect, Tagarelli et al. [14] first introduced a computational approach for lurker ranking, and they further addressed time-evolving dynamics of lurkers [16]. While Anderson et al. [17] focuses on active engagement of users in social networks. Most of above mentioned researches are mainly done on English text, however the application of sentiment analysis is more difficult in Chinese than in Indo-European languages, due to the compounding nature of Chinese words and phrases, and relatively lack of reliable emotional resources in Chinese. Existing studies that analyze static English text fail to capture the real emotions in Chinese online reviews, especially the short interactive reviews.

In this paper, we focus on the emotional evolution analysis for Chinese online reviews. In summary, there are three key problems that need to be addressed:

- **P1)** `Affective computing framework:` How to calculate the emotional value of a Chinese online review?
- **P2)** `Emotional interaction model:` How to model the emotional interactions between online users?
- **P3)** `Emotional evolution prediction:` How to predict the mutual attitudes in the further interactions?

To solve the above problems effectively, we attempt to survey online interactions from an emotional game perspective: 1) a subsupervised affective computing framework is designed for Chinese review emotion calculation; 2) a game model is proposed to model emotional interactions between online users, in which the satisfaction or disappointment of interacted users are quantified by predefined utility functions; 3) an emotional evolution prediction algorithm is applied on each interactive user pair to predict their mutual attitudes in the future. The theoretical analysis and experiments prove that the proposed approaches can effectively deal with the existing problems.

To conclude, we make the following contributions

- **1)** We design an efficient affective computing framework which includes three phases: feature selection, semantic orientation study, and comment emotional calculation. The proposed method can generate domain-dependent Chinese morphemes directly from large dataset, without requiring expensive manual labeling of sentimental seeds; and the obtained emotion value of a review can be seen as the probability that the generator sends a positive comment to the recipient.
- **2)** We employ the game theory to model the emotional interactions between online users, and carefully define a set of utility functions to measure the degree of "happiness" of interacted users.
- **3)** To predict the mutual attitudes in the further, we summarize three criterions: 1) if one user has a dominant strategy, he/she will adopt this strategy; 2) if only his/her parter has a dominant strategy, he/she will adopt the best response paired with that dominant strategy; 3) if neither user has a dominant strategy, he/she will adopt the mixed strategy under the "Nash Equilibrium" state. With these three criterions, we further propose an emotional evolution prediction algorithm (EEPA) for attitude prediction.
- **4)** Through the theoretical analysis and experiments, the advantages of our presented methods are validated: 1) only 533 high-frequency words are selected as the emotional features, and an indirect evaluation method proves that the proposed affective computing framework is effective; 2) a group of test sets are manually built to verify the performance of the emotional evaluation prediction algorithm, and the average precision of EEPA runs up to 80.5%.
- **5)** We extract the largest connected component from the Tianya reciprocal network and further define the pairwise happiness-popularity coordination. Some interesting phenomena have been observed, and the details can be found in Section 7.

The remainder of the paper is organized as follows: Section 2 presents the related work about sentiment analysis. In Section 3, we introduce the dataset collection. Section 4 provides the affective computing framework to measure Chinese online reviews. In Section 5, we present the algorithm for emotional evolution prediction. Section 6 provides the evaluation of the method and discussion about the results. In Section 7, we reveal some interesting phenomenon in the dataset, and finally conclude this paper in Section 8.

2. Related work

Here we discuss related work from two areas: affective computing, and emotional evolution analysis.

2.1. Affective computing

With the development of CMC, a large number of people are writing reviews online. For example, in the electric business platform, when trying to locate information on a product, a general web search would retrieve a large collection of documents; however, getting an overall sense of the reviews can daunting and time-consuming. To solve these problems, recent years have witnessed a growing interest in affective computing, whose objective is to find opinions, feelings, and attitude expressed in text, rather than facts. In the literature, affective computing also goes under various names, such as opinion mining [18–20], sentiment extraction [21,22], etc. Its related work may come from both computer science and linguistics, and its immediate applications may involve data mining, market intelligence, and customer relationship management.

In the early literature, Hatzivassiloglou and MacKeown [8] presented and evaluated a method that automatically retrieves semantic orientation information using indirect information collected from a large corpus. As the method relies on the corpus, it extracts domain-dependent information and automatically adapts to a new domain when the corpus is changed. Although they only focused on the adjectives, it can be directly applied to other word classes. Some other work along this line can be found in [7,11,23,24]. Specifically, Turney and Littman [11] proposed an unsupervised method, which is based on the Point Mutual Information (PMI) to measuring semantic orientation from semantic association. The method can be used to identify the semantic orientations of any etymologies, and is easy to implement. Poria et al. introduced a novel paradigm to concept-level sentiment analysis that merges linguistics, common-sense computing, and machine learning for improving the accuracy of tasks such as polarity detection [24].

Some other work tried to extract the semantic association from the corpus dictionaries (e.g., WordNet [25] and HowNet [26]), to determine the semantic orientation of a word. For example, Kamps and Marx [27] used WordNet to evaluate the semantic distance from a word to `good`/`bad`. They first defined a graph on the adjectives appeared in both the WordNet and the target term list. If two adjectives in WordNet display a synonymy relation, a link will be added between them, in turn, the semantic orientation of a word is decided by its relative distance to `good` and `bad`. However, not all the languages have the rich emotional resources as English does, Mihalcea et al. [28] tried to translate the subjective words of WordNet into other languages (i.e., Chinese) for cross-lingual analysis, however, experiments showed that some words have changed their original semantic orientations after the translation.

Aside from the explicit two-class classification problem, some work tended to determine the authors' opinion with different rating scales (e.g., the number of stars). Pang et al. [29] employed three machine learning approaches (Naive Bayes, Maximum Entropy, and Support Vector Machine) to label the polarity of IMDB (Internet Movie Database) movie reviews. The reviews

Please cite this article as: Z. Bu et al., Game theory based emotional evolution analysis for chinese online reviews, Knowledge-Based Systems (2016), http://dx.doi.org/10.1016/j.knosys.2016.03.026