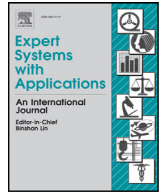




ELSEVIER

Contents lists available at ScienceDirect

## Expert Systems With Applications

journal homepage: [www.elsevier.com/locate/eswa](http://www.elsevier.com/locate/eswa)

## Opinion mining for multiple types of emotion-embedded products/services through evolutionary strategy

Heng-Li Yang\*, Qing-Feng Lin

Department of Management Information Systems, National Cheng-Chi University, Taipei, Taiwan

## ARTICLE INFO

## Article history:

Received 23 July 2017

Revised 14 January 2018

Accepted 15 January 2018

## Keywords:

Chinese corpus

Evolutionary strategy

Multiple polarities

Opinion mining

Optimization

Sentiment analysis

## ABSTRACT

Since the advent of blogging, microblogging, and social networking sites, researchers and practitioners have been increasingly concerned with the problem of obtaining useful evaluations from web-based opinion articles in a process known as opinion mining or sentiment analysis. In this study, we focused on reviews based on highly emotion-embedded products/services, such as movies, music, and drama. Furthermore, we tried to solve the multiple polarities problem for the same review word for multiple types of product/service. First, we collected text written in Chinese from a Taiwanese movie forum. In our proposed approach, we applied an evolutionary strategy algorithm to optimize the weight tables corresponding to two different types of movies: horror and drama movies. The experimental results indicated that the proposed method performed better than conventional methods when considering only one generalized type. Further, we employed a new multi-class support vector machine approach for predicting opinions at the document level. We used seven measures to describe the characteristics of an overall document, including the central tendency, dispersion, and shape of the predicted sentence value distribution, where the fluctuations in these values corresponded to their positions in the document. We also demonstrated the effectiveness of this approach for identifying opinions at the document level.

© 2018 Elsevier Ltd. All rights reserved.

## 1. Introduction

Since the development of blogging, microblogging, and social networking sites (SNSs), large numbers of personal opinions have been accumulating on the web each day. Companies can use a web spider to collect large amounts of opinion data regarding brands and products, as well as news, discussions, and even rumors. Thus, there is a growing need to obtain targeted information from this large repository of stored data.

Opinion mining, sentiment analysis, and sentiment classification are used to extract or classify the sentiment polarity regarding product attributes by analyzing sentiment-oriented text. Unfortunately, many previous studies have used the same processing methods regardless of the specific context, but the sentiment polarity of any given sentence can easily be reversed in different types. For example, let us consider the following comment on a movie blog: "From the beginning, many of the scenes made me scared and disturbed." In general, a reader could infer that this as a negative opinion. However, it is more likely to be a positive opinion in the context of a horror movie. Understanding the true opinion of the author requires that the movie type is

classified in an appropriate manner. Lin and He (2009) also noted that topic/feature detection and sentiment classification are often performed separately, which ignores their mutual dependence. They also gave an example where the phrase "unpredictable steering" is negative in an automobile review but positive in a movie review. Similar to other researchers (Choi, Kim, & Myaeng, 2009; Jo & Oh, 2011; Wang, Chen, & Liu, 2016), they proposed several unsupervised joint approaches for detecting topic and sentiment pairs. However, without prior knowledge regarding the type, it might be easy to distinguish the topics "automobile" and "movie"; but difficult to clearly distinguish the reviews of some sub-types of products/services, especially those specified by authors or in cases where some fuzzy words appear in the reviews. If the reviews given by authors specify the sub-types of their reviews and the reviews are uploaded to certain categories in blogs or forums, then their written intention should be respected when researchers conduct text mining. Thus, in the present study, we developed a new supervised approach for solving the multiple polarities problem for the same review word for multiple types of product/service.

The products/services considered in the present study are highly emotion-embedded products/services. In previous studies, researchers have investigated affective product design and tried to design attributes of products that trigger "happiness" in one's mind or that help to communicate positive emotions (Demirbilek & Sener, 2003; Seva, Duh, & Helander, 2007). Barnes and Lillford

\* Corresponding author.

E-mail address: [yanh@nccu.edu.tw](mailto:yanh@nccu.edu.tw) (H.-L. Yang).

presented a decision support framework to assist the development of emotionally appealing products by eliciting the needs of users early in the development process (Barnes & Lillford, 2009). However, some art products/services are naturally embedded with emotions, such as movies, drama, dance, and acrobatics, or physical products, e.g., antique and crafts. In the present study, we focused on movies because there are many movie reviews on the Internet.

In order to exploit the characteristics of the embedded emotions in art products/services, we propose a new machine learning approach based on ConceptNet and an evolutionary strategy. We assume that the positive/negative implications of various types of emotion (e.g., happiness, anger, and fear) in the review texts on a blog/forum for art products/services will remain constant for some time. Each emotion polarity in an art product/service type can be obtained using an evolutionary strategy. Thus, the analysis of each art product/service on a blog/forum can be performed only once and re-used for a period of time. The implied emotions in all of the review sentences are identified using ConceptNet. Furthermore, we propose a new approach for document-level opinion mining.

The remainder of this paper is organized as follows. In Section 2, we provide a brief review of related research. In Section 3, we propose our approach for mining opinions in multiple art product/service types. In Section 4, we present our approach for document-level opinion mining. The results of two experiments are presented in Section 5. Finally, we give our conclusions and suggestions for future research in Section 6.

## 2. Related work

The opinions of others are important references for decision making. Opinion mining, sentiment analysis, and sentiment classification are automatic text classification technologies based on the emotional polarity of the author, such as “good,” “bad,” or “mixed” (Dave, Lawrence, & Pennock, 2003; Pang & Lee, 2008). Opinion mining is widely applicable to formal text, such as newspapers (Allahyari & Kochut, 2015; Perikos & Hatzilygeroudis, 2016) or technology articles (Chen et al., 2016), but also to informal text, such as user comments on the Internet regarding topics including electronic items (Turney & Littman, 2003), movies (Chaovalit & Zhou, 2005; Ye, Shi, & Li, 2006), books, hotels (Yin, Han, Huang, & Kumar, 2014), or restaurants (Yan, Wang, & Chau, 2015). The rapid development of SNSs has provided a novel medium that allows people to express their emotions, and studies of emotion analysis based on posts (e.g., tweets or short reviews) in SNSs have become popular in recent years (Goudas, Louizos, Petasis, & Karkaletsis, 2015; Kuo, Fu, Tsai, Lee, & Chen, 2016; Mostafa, 2013). For example, Nguyen, Phung, Adams, and Venkatesh (2014) applied a set of psychology-inspired features and psycholinguistic features to conduct a large-scale mood analysis in social media texts (blog posts). In the following section, we review related research into the level of mining granularity, analytical techniques, Chinese text processing, the applications of evolutionary algorithms to opinion mining, and the problem of interpreting multiple types.

### 2.1. Level of mining granularity

Previous studies have tackled the problem of opinion mining at three levels of granularity according to their objectives: phrase level, sentence level, and document level. The identification of contextual information is even important in the phrase level. For example, Wilson, Wiebe, and Hoffmann (2005) presented an approach that begins with a lexicon of words with established prior polarities, before identifying the contextual polarity of phrases where instances of these words appear in a corpus. Some studies that focused on the sentence level tried to enhance the accuracy of assessments by analyzing the locations and density

of words as well as the relationships with subjective words in a sentence (Eskimez, Sturge-Apple, Duan, & Heinzelman, 2016; Missen, Boughanem, & Cabanac, 2013). Using sentence-level sentiment analysis, researchers have identified the positive/negative orientations of opinions about certain aspects/features (e.g., lens of a digital camera) of products/services (Hu & Liu, 2004; Lerman, Arora, Gallegos, Kumaraguru, & Garcia, 2016; Yang, Jiang, Wang, & Xie, 2014). For example, Liu, Liu, Zhang, Kim, and Gao (2016) proposed a lifelong learning-based approach for extracting aspects of a product/service by using semantic similarity and aspect associations. Sauper and Barzilay (2013) suggested a model for simultaneously identifying the underlying set of ratable aspects presented in the reviews of a product (e.g., chicken, dessert, and drinks) and determining the corresponding sentiment for each aspect. At the document level, early attempts identified the polarity of an article based simply on the occurrence of subjective words throughout the entire document (Pang, Lee, & Vaithyanathan, 2002). In the present study, our proposed method focuses on mining at the sentence level and document level, but we do not explore detailed aspects of a product/service.

Two approaches have been developed for fine-to-coarse (i.e., sentence-to-document) opinion mining: cascaded (Zhang, Zeng, Li, Wang, & Zuo, 2009) and joint (Chen et al., 2016; Kaminka, 2016; Li & Lu, 2017). The cascaded approach begins by computing the polarity of each sentence in an article and then using a model to summarize opinions at the document level (Mao & Lebanon, 2006). The joint approach combines sentence-level and document-level analysis within a single joint model to determine the overall polarity (McDonald, Hannan, Neylon, Wells, & Reynar, 2007). The joint approach can effectively classify opinions at the document level by using trained output classifiers, but the classifier training process requires large training data sets at the sentence and document levels, and creating these data sets can be costly and time consuming. The cascaded approach is more intuitive and easier to explain, but operating at the sentence-level makes it computationally intensive. In the present study, our proposed method uses the cascaded approach.

### 2.2. Analytical techniques

The analytical techniques used for opinion mining can be categorized as rule-based and learning-based. Most rule-based approaches require a predefined dictionary of subjective words for use when analyzing the polarity of phrases, sentences, or documents (Wilson et al., 2005). Thus, it is necessary to scan articles to detect subjective words in order to identify trends in the opinions expressed.

By contrast, learning-based approaches do not require predefined dictionaries and they can provide predictive models by learning from labeled training data sets. This approach involves the construction of a self-adjusting learning model using a set of initial parameters, before several iterations of a process comprising data input, object value calculation, model adjustment, and checking the termination conditions. Finally, the trained model is validated by testing (Lin & He, 2009; Pang et al., 2002). The learning-based approach requires more training time and a larger labeled data set, but there is no need for a predefined dictionary. As a result, researchers are increasingly using the learning-based approach (Kontopoulos, Berberidis, Dergiades, & Bassiliades, 2013). In this study, we propose a new learning-based approach for analyzing the sentiment of sentences.

### 2.3. Opinion mining in Chinese articles

Many studies of opinion mining have focused on the English language, but several recent investigations have addressed the

Download English Version:

<https://daneshyari.com/en/article/6855124>

Download Persian Version:

<https://daneshyari.com/article/6855124>

[Daneshyari.com](https://daneshyari.com)