ARTICLE IN PRESS

Computers in Human Behavior xxx (2013) xxx–xxx

Contents lists available at ScienceDirect

Computers in Human Behavior

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

Sentiment analysis in Facebook and its application to e-learning

Alvaro Ortigosa *, José M. Martín, Rosa M. Carro

Department of Computer Science, Universidad Autónoma de Madrid, Francisco Tomás y Valiente 11, 28049 Madrid, Spain

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Sentiment analysis
Social networks
User modeling
Adaptive e-learning

ABSTRACT

This paper presents a new method for sentiment analysis in Facebook that, starting from messages written by users, supports: (i) to extract information about the users' sentiment polarity (positive, neutral or negative), as transmitted in the messages they write; and (ii) to model the users' usual sentiment polarity and to detect significant emotional changes. We have implemented this method in SentBuk, a Facebook application also presented in this paper. SentBuk retrieves messages written by users in Facebook and classifies them according to their polarity, showing the results to the users through an interactive interface. It also supports emotional change detection, friend's emotion finding, user classification according to their messages, and statistics, among others. The classification method implemented in SentBuk follows a hybrid approach: it combines lexical-based and machine-learning techniques. The results obtained through this approach show that it is feasible to perform sentiment analysis in Facebook with high accuracy (83.27%). In the context of e-learning, it is very useful to have information about the users' sentiments available. On one hand, this information can be used by adaptive e-learning systems to support personalized learning, by considering the user's emotional state when recommending him/her the most suitable activities to be tackled at each time. On the other hand, the students' sentiments towards a course can serve as feedback for teachers, especially in the case of online learning, where face-to-face contact is less frequent. The usefulness of this work in the context of e-learning, both for teachers and for adaptive systems, is described too.

1. Motivation

The use of computers in education has meant a great contribution for students and teachers. The incorporation of adaptation methods and techniques allows the development of adaptive e-learning systems, where each student receives personalized guidance during the learning process (Brusilovsky, 2001). In order to provide personalization, it is necessary to store information about each student in what is called the student model (Kobsa, 2007). The specific information to be collected and stored depends on the goals of the adaptive e-learning system (e.g., preferences, learning styles, personality, emotional state, context, previous actions, and so on).

In particular, affective and emotional factors, among other aspects, seem to affect the student motivation and, in general, the outcome of the learning process (Shen, Wang, & Shen, 2012). Therefore, in learning contexts, being able to detect and manage information about the students' emotions at a certain time can contribute to know their potential needs at that time. On one hand, adaptive e-learning environments can make use of this information to fulfill those needs at runtime: they can provide the user with recommendations about activities to tackle or contents to interact with, adapted to his/her emotional state at that time. On the other hand, information about the student emotions towards a course can act as feedback for the teacher. This is especially useful for online courses, in which there is little (or none) face-to-face contact between students and teachers and, therefore, there are fewer opportunities for teachers to get feedback from the students.

Knowing the users' emotions is useful not only in the educational context but also in many others (e.g., marketing, politics, online shopping, and so on) (Feldman, 2013). In general, in order for a system to be able to take decisions based on information about the users, it is necessary for it to get and store information about them. One of the most traditional procedures to obtain information about users consists of asking them to fill in questionnaires. However, the users can find this task too time-consuming. Recently, non intrusive techniques are preferred (de Montjoye, Quoidbach, Robic, & Pentland, 2013). We also think that information for student models should be obtained as unobtrusively as possible, yet without compromising the reliability of the model built (Ortigosa, Carro, & Quiroga, 2013).

When reflecting about potential sources of information regarding user sentiment, we looked for digital places in which the users

* Corresponding author. Tel.: +34 91 4972271.
E-mail addresses: alvaro.ortigosa@uam.es (A. Ortigosa), joseph.martin@estudiante.uam.es (J.M. Martín), rosa.carro@uam.es (R.M. Carro).

0747-5632/$ - see front matter © 2013 Published by Elsevier Ltd.
http://dx.doi.org/10.1016/j.chb.2013.05.024

express themselves frequently and naturally. Nowadays, the number of users interacting with others through social networks is growing exponentially. Therefore, we focused on social networks. There exist an increasing number of online social networks available through the Web. From these applications, Facebook is the more popular around the world. On October 2012, it reached 1 billion monthly active users (that is, 1 billion users accessed the network within a month) and more than 550 million daily active users (Kiss, 2012). Besides its popularity, Facebook provides a distinctive advantage for this research: it is a network of friends. That is, whilst other social networks focus on professional relationships or serve, mostly, as sources of information, people make use of Facebook mainly to share and communicate with friends; in fact, the acquaintances or relationships between users are called "friends". Messages in Facebook are spontaneous and users express their emotions more naturally. It was because of all these reasons that we chose Facebook as the development platform. In Facebook, the "wall" is the space where the users publish their own messages, contents and so on. Regarding text messages, there are several categories: status messages (each user writes them in his/her own wall), posts in others' walls, and comments to either one's or others' publications. Typically, a user's wall is visible to his/her friends, and they can make comments or express that they "like" a particular message or post.

When dealing with users and sentiments, it is useful to know the users' emotional state at a certain time (positive/neutral/negative), in order to provide each of them with personalized assistance accordingly. Moreover, it is also interesting to know whether this state corresponds to their "usual state" or, on the contrary, a noticeable variation might have taken place. Behavior variations, as detected in the messages written by a user (when sentiment histories are available), can indicate changes in the user's mood, and specific actions could be potentially needed or recommended in such cases.

With the purpose of extracting information about users' sentiments from the messages they write in Facebook and detecting changes, we have developed a new and non-intrusive method for sentiment analysis in this social network. It consists on a hybrid approach, combining lexical-based and machine learning techniques. We have implemented this method in SentBuk, a Facebook application that retrieves the messages written by the users and extracts information about their emotional state.

This paper is organized as follows. Section 2 presents the state of the art of the research areas related to our work. Section 3 describes the new method proposed for text-based sentiment analysis. Section 4 presents SentBuk, the Facebook application in which that method has been implemented. Section 5 includes the results obtained when making use of SentBuk, as well as the analysis of these results. Section 6 presents a discussion of the proposal and shows some applications of sentiment analysis in the context of e-learning. Finally, the conclusions of the work done, along with some lines for future work, are presented in Section 7.

2. Related work

2.1. Sentiment analysis

Sentiment analysis has been defined as the computational study of opinions, sentiments and emotions expressed in texts (Liu, 2010). For the sake of simplifying the development of an emotion recognition tool, we have tried to avoid complex and potentially controversial definitions of emotions and sentiments. In this direction, we take the simplified definition of sentiment as "a personal positive or negative feeling or opinion". An example of a sentence transmitting a positive sentiment would be "I love it!", whereas "It is a terrible movie" transmits a negative one. A neutral sentiment does not express any feeling (e.g. "I am commuting to work"). Most of works in this research area focus on classifying texts according to their sentiment polarity, which can be positive, negative or neutral (Pang & Lee, 2008). Therefore, it can be considered a text classification problem, since its goal consists of categorizing texts within classes by means of algorithmic methods.

The earliest researches dealing with sentiment analysis consisted on classifying words or phrases according to semantic issues and date from the late 1990s (Hatzivassiloglou & McKeown, 1997). Linguistic heuristics or pre-selected sets of seed words were used. The results obtained in those works served as the basis for classifying entire documents, considering that the average semantic orientation of the words in a review may be an indicator of whether the text is positive or negative (Turney, 2002). The appearance of WordNet (Miller, 1995) and, in general, of annotated corpora, increased the production in this research area. On one hand, WordNet is useful because it allows knowing the semantic relationships between different words. Therefore, with a reduced set of polarity words, every word could be labeled as positive, negative or neutral through its relationships. On the other hand, corpora and, in particular, the Treebanks, are very useful. They are corpora with the syntactic structure labeled, and are of great help for training the analyzers in order to label the words automatically.

One of the first works that used the term "sentiment analysis" as we currently know it was that presented in (Das & Chen, 2001), which analyzes messages written in stock boards in order to extract the market sentiment. Currently, many of the works in this area focus on document classification based on the sentiment expressed on it. One of the best known domains is that of reviews (Pang, Lee, & Vaithyanathan, 2002) (Dave, Lawrence, & Pennock, 2003). Review websites are examples of especially useful sources for sentiment analysis, such as, e.g. Epinions (Epinions, 1999). Other application areas in which sentiment analysis can be very useful are:

– Recommendation systems (Tatemura, 2000).
– Flame detection (Spertus, 1997).
– Sensitive content detection for advertising (Jin, Li, Mah, & Tong, 2007).
– Human–computer interaction (Liu, Lieberman, & Selker, 2003).
– Business Intelligence (Mishne & Glance, 2006)
– Prediction of hostile or negative sources (Abbasi, 2007).
– Classification of citizens' opinions on a law before its approval: "eRuleMaking" (Cardie, Farina, Bruce, & Wagner, 2006).
– Broadcasting based on the receiver sentiment (Rogers, 2003).
– Dynamic adaptation of daily tools, such as e-mail (Carro, Ballesteros, Ortigosa, Guardiola, & Soriano, 2012).
– Marketing or politics (Feldman, 2013).

In general, accuracy is strongly influenced by the context in which the words are used (Turney, 2002) (Aue & Gamon, 2005) (Engström, 2004). For instance, the sentence "You must read the book" is positive in a book review but is negative if the review is about films.

Additionally, the position of words in text is an interesting factor to consider, since a word at the end of a sentence can change the polarity completely (Pang et al., 2002). For example the sentence "This book is very addictive, it can be read in one sitting, but I have to admit that it is rubbish" begins with the word "addictive" and the expression "one sitting", which are positive in the context of book reviews, but it finish with the word "rubbish" that