A fuzzy-oriented sentic analysis to capture the human emotion in Web-based content

Vincenzo Loia, Sabrina Senatore *

Dipartimento di Informatica, Universitá degli Studi di Salerno, via Giovanni Paolo II, 132, 84084 Fisciano, SA, Italy

ARTICLE INFO

ABSTRACT

Capturing the sentiments and the emotional states enclosed in textual information is a critical task which embraces a wide range of web-oriented activities such as detecting the sentiments associated to the product reviews, developing marketing programs that would be attractive for users, enhancing customer service with respect to its expectation until to identifying new opportunities and financial market prediction, besides managing reputations. Opinions and the emotions that are embedded in them, play a key role in decision-making processes, with different effects depending on the negative or positive valence of the mood. When the choice depends on some important features (i.e., time, money, reliability/efficacy, etc.) and on other opinions (which come from previous experience), could be crucial to make the best decision.

Inferring opinions and emotions enclosed in the written language is a complex task which cannot rely on body languages (posture, gestures, vocal inflections), rather than discovering concepts with an affective valence. The role of opinions extracted by the social content is crucial to support consumers' decision process; in addition, thanks opinions and emotions, it is possible to evidence improvements on existing decision supports and show how the opinion-mining techniques can be incorporated into these systems.

This paper presents a tentative contribution that addresses this issue: it introduces a framework for extracting the emotions and the sentiments expressed in the textual data. The sentiments are expressed by a positive or negative polarity, the emotions are based on the Minsky's conception of emotions, that consists of four affective dimensions, each one with six levels of activations [1]. Sentiments and emotions are modeled as fuzzy sets; particularly, the intensity of the emotions has been tuned by fuzzy modifiers, which act on the linguistic patterns recognized in the sentences. The approach has been tested on some sets of documents categories, revealing interesting performance on the global framework processing.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

The growing popularity of Web 2.0 strictly depends on the acquired awareness of being active part of the global network: the web users interact, share and collaborate through social networks, online communities, blogs, wikis, feeds and chats. In the last years, there has been a gradual and evident transformation from read-only to read-write Web, from informative to social and interactive content that has made web-users more inclined to express their opinions, sentiments about daily life, movies, commercial assets and products, etc. At the same time, the widespread of "e-services" (i.e., e-commerce, e-tourism, e-business, etc.) provides a crucial role to users' opinion and judgment which are often more listened to than experts' viewpoints. The importance of the opinions with the related emotions in the decision making processes is undiscussed. Users spend some time on the Web to search for user-generated reviews in order to estimate the actual utility value of products [2], often considering user opinions as a feasible source of information about products.

Introducing the human dimension into text comprehension guarantees to guess the user feeling, expectation and preferences, suggests ad hoc advertisement and addresses profitable market analysis.

In an effort to explore potential resolutions to this issue, the recent research proceeds towards the "distillation" of the social knowledge, in order to capture the moods that govern the users' actions and behavior on the Web, for a wide range of decision-making activities.

Companies exploit sentiment analysis and opinion mining tools for their marketing strategies, in order to evaluate the public attitudes towards their brands; social events, political actions are directed by collective moods and sentiments; even recommendation systems begin considering opinions and sentiments as additional aspects to take into account, besides the traditional positive or negative feedback.

* Corresponding author. Tel.: +39 089968233.
  E-mail address: ssenatore@unisa.it (S. Senatore).

But, the tricky understanding of the natural language text makes the interpretation of human feelings a very complicated task. Sentiments and opinions indeed are concealed in the sentences, typically associated to adjectives and verbs; then the intrinsic meaning of some textual expressions is not amenable to rigid linguistic patterns.

Recent trends [1] emphasize the role of the emotions in the text understanding, although the emotions are not clearly visible as facial expression, vocal inflections, body movement and gesture. *Sentic Computing*, is a novel paradigm that aims at grabbing the sentiments in text. It is from the Web Science [3], a new interdisciplinary area which explores the development of the Web in the everyday life and gathers evidence on human behavior about how it will affect the Web technological progress. Sentic Computing exploits the common sense in the natural language for capturing and processing human feelings in web resources, typically, webposts i.e. short texts such as blog posts, forum entries, RSS feeds, tweets and instant messages [1].

Sentic Computing is defined as a paradigm which combines Opinion Mining and Sentiment Analysis, by exploiting Artificial Intelligence and Semantic Web techniques to recognize, interpret and process opinions and sentiments in natural language text [4]. It captures the human "sentire" (root of words like sentiment and sensation [5]), in decision making processes which are driven by emotions, sentiments and sensations.

This work has been inspired by this new paradigm. We present a framework that elaborates text in natural language in order to grab the mood expressed in the content. We consider the "mood" as a mix of sentiments, emotions, feelings that move the author of a certain text to write that comment, observation, criticism, etc. The approach herein presented, exploits WordNet [6], a well known lexical database and some its derivative tools, to retrieve the meaning of each word in the text and the associated sentiment polarity, i.e. if the word expresses a positive or negative sentiment. Then, a variant of the Plutchik's wheel of emotions [7] allows to us recognize the emotional states associated to these words. Then we present a fuzzy modeling (even it is still embryonal) of the sentiments and emotions, able to capture richer interpretations of the text when linguistic patterns are identified, in addition to the processing of the single words. The emotions are modeled as triangular fuzzy sets and (fuzzy) linguistic modifiers act to intensify/reduce each emotion identified in some linguistic expression.

The use of Fuzzy logic in the modeling of the emotions is a remarkably simple way to process vague, ambiguous or imprecise information. It provides a mathematical tool where vague, linguistic expressions, conceptual emotions can be rigorously modeled. Therefore, the fuzziness as a means of modeling linguistic uncertainty lends itself well to analyze emotions by providing a more realistic interpretation of the text.

This "mood"-oriented text analysis can be considered a tentative contribution to make decisions in different domains: e-commerce, marketing and advertising, opinion-tracking, collaborative filtering, etc. The final result is twofold: (1) to analyze the sentiment polarity of the text; and (2) to extract a range of emotional states that describe the author's mood expressed in the text.

The paper is developed as follows: Section 2 presents an overview of the current topics and techniques related to the sentiments or in general, the emotions analysis; then the workflow of the framework is introduced in Section 3, by describing the major steps in the whole process and the roles of the different components of the working flow. In particular, Section 4 introduces the external resources employed in the process, while the main phases of the workflow are detailed in Section 5. Experimental results and conclusion close the paper.

2. Related work

Recent research trends aim at grasping the emotional and affective information associated with the natural language opinions [8]. Capturing emotions that are behind the opinions is assuming a considerable importance with the growth of social Web (i.e., reviews, forum discussions, blogs, Twitter, and social networks). The Web is becoming a big data repository with a huge volume of opinionated data to analyze. Opinion mining is a wide area of interest that gathers all the methodologies and techniques aimed at extracted opinions, sentiments, attitudes from written language. Emotion Extraction is a specific subfield of Opinion Mining that aims at interpreting the emotions and moods from the natural language. A deeper description of these areas are given in the following sections and some relevant works are presented.

2.1. Sentiment analysis and opinion mining

Sentiment analysis and opinion mining are two interrelated fields of growing interest. They refer interchangeably to the same area of study [8]. More specifically, the sentiment analysis relates to techniques to automatically classify sentiments (in terms of positive or negative polarity) for entire documents. The opinion mining considers approaches for capturing opinions, moods and human feelings for the webposts, reviews or documents in general.

Sentiment Analysis (SA) has been extensively studied in recent years. Our approach achieves phrase-level SA, i.e., the identification of sentiments expressed in phrases, rather than document-level SA, i.e., the detection of sentiments representing the whole document and its polarity classification. The need of deep syntactic analysis for the phrase-level SA, in order to extract more different sentiments from a document, has been outlined in [9]. In fact, identifying the local sentiment is more reliable than the global document sentiment: the approach described in [10] presents a system for discovering semantic relationships between the sentiment expressions and the subject/topic, by identifying the polarity of the sentiment expressions and then, determining the subjects of these sentiments, through a merge of semantic analysis, syntactic parsing and sentiment lexis. In [11], a machine-learning method applies text-categorization techniques to some subjective parts of the document, by means of techniques for finding minimum cuts in graphs. In [12], a system consisting of a sentiment identification phase for large corpus of news and blogs is proposed: some lexicons are defined for seven sentiment dimensions (general, health, crime, sports, business, politics, media) by recursively querying for synonyms using WordNet.

The phrase-level SA makes sense for the classification of Twitter messages, due to the character limitations on tweets, even the informal language and the nature of microblogging domain make the identification of sentiments a different task [13]. In fact, in [14] the microblogging features such as re-tweets, hashtags, replies, punctuations, and emoticons have been taken into account, instead of using n-grams on tweet data that may hinder the classification performance because of the large number of infrequent words in Twitter [15].

Existing work mainly concentrates on the use of three types of features: lexicon features, POS features, and microblogging features for sentiment analysis and opinion mining [15]. Some approaches combine these three types of features, giving major emphasis to POS tags with or without word prior polarity involved [16], others explore the use of microblogging features [13,14].

Mining opinions needs an accurate understanding of the natural language that takes into account explicit and implicit, regular and irregular, syntactic and semantic language rules that part of the NLP unresolved problems. Thus, a positive document does not