Chapter 17
Automatic Sentiment Analysis on Web Texts for Competitive Intelligence

Stanley Loh
Lutheran University of Brasil (ULBRA), Brazil & Technology Faculty Senac Pelotas, Brazil

Fabiana Lorenzi
Lutheran University of Brasil (ULBRA), Brazil

Paulo Roberto Pasqualotti
University FEEVALE, Brazil

Sabrina Ferreira Rodrigues
Lutheran University of Brasil (ULBRA), Brazil

Luis Fernando Fortes Garcia
Lutheran University of Brasil (ULBRA), Brazil & Dom Bosco Faculty of Porto Alegre, Brazil

Valesca Persch Reichelt
Lutheran University of Brasil (ULBRA), Brazil

ABSTRACT

This chapter presents a software tool that helps the Competitive Intelligence process by collecting and analyzing texts published on the Internet. The goal is to automatically analyze indicators of the sentiment present in texts about a certain theme, whether positive or negative. The sentiment analysis is made through a probabilistic process over keywords present in the texts, using as reference a task ontology with positive and negative words defined with a degree of confidence.

INTRODUCTION

Nowadays organizations have to evaluate their competitive environment, collecting and analyzing external information. This information may be collected from different sources. Among these, Internet is an easy and broad source of information. In this way, organizations and people are measuring the public opinion (repercussion) about facts related to their market or actions by analyzing information published in the Internet. One example is the analysis of news. News are formal sources while there are other many ways for people externalizing their opinions in informal sources, like blogs, forums, twitter, communities and social networks. For organizations, it is important to identify the main informal actors, their sentiment about the facts and how they influence

DOI: 10.4018/978-1-4666-2494-8.ch017

other people. However, the success depends on the capacity and velocity of collecting, selecting, and analyzing information that can lead to strategic and valuable decisions.

This work presents a software tool that helps people and organizations in collecting and analyzing texts published in the Web about certain subjects. From a given theme (represented by keywords), texts are collect in different sources as blogs, news, websites and twitter. After that, the tool performs a sentiment analysis over the texts, that is, it analyzes keywords present in the text and compares with a task ontology in order to identify indicators of positive or negative sentiment. Using a probabilistic reasoning process, the tool determines the main sentiment of the text (whether positive, negative, or neutral). The tool also allows selecting the kind of source to be analyzed (or all kinds).

This chapter is structured as follows. Section 2 presents some related work about sentiment analysis. Section 3 presents the software tool, its functionalities, methods, and techniques. Section 4 describes the process of creating the task ontology for the purpose of sentiment analysis, besides explaining the structure of the task ontology if one wishing to apply it on other subject. Section 4 presents an experiment carried out with the tool and the evaluation of the process. Section 5 finishes with concluding remarks and future work.

RELATED WORK

Automatic Sentiment Analysis is the analysis of affective states made by machines over documents written in natural language, based on emotion categories like positive and negative (Gregory, et al., 2006). Other authors utilize synonyms as "opinion mining" (Pang & Lee, 2008). This kind of work is applied for analysis of customers´ satisfaction about products and services, for identification of attitudes on reports of employees and so on.

Ma et al. (2005) presents experiments with sentiment analysis over chat messages. Emotions are estimated based on keywords present in the text. The work of Genereux and Evans (2006) classifies weblogs according to their affective content, using a two-dimensional space with axes "positive or negative" and "active or passive."

Godbole et al. (2007) assigns scores indicating positive or negative opinion to each distinct entity in texts from newspapers and blogs. They expand a small candidate list of positive and negative words into a sentiment dictionary using path-based analysis of synonyms and antonyms in WordNet.

Pang and Lee (2008) presents a survey about the main techniques used for opinion mining and sentiment analysis in texts, besides concerning other issues in this theme. They expose the difficulty of analyzing sentiments based only on words present in the text. They report an accuracy of 60% for the use of lists of keywords created by humans within a straightforward classification policy. In contrast, "word lists of the same size but chosen based on examination of the corpus' statistics achieves almost 70% accuracy" and "applying machine learning techniques based on unigram models can achieve over 80% in accuracy."

Ortony, Clore, and Colins (Ortony, et al., 1988) presents a psychological and cognitive model of emotions, containing emotions resulting from the description of cognitive processes and real world interpretations by a human subject. The model contains 22 kinds of emotions: "happy for," "resentment," "gloating," "pity," "joy," "distress," "pride," "shame," "admiration," "reproach," "love," "hate," "hope," "fear," "satisfaction," "fears-confirmed," "relief," "disappointment," "gratification," "remorse," "gratitude," and "anger." Table 1 shows the emotions and their corresponding adjectives.