Available online at www.sciencedirect.com

ScienceDirect

Procedia - Social and Behavioral Sciences 220 (2016) 373 – 381

Procedia
Social and Behavioral Sciences

19th International Conference Enterprise and Competitive Environment 2016, ECE 2016, 10–11 March 2016, Brno, Czech Republic

Mobile application for customers' reviews opinion mining

Jan Přichystal[a,*]

[a]*Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic*

Abstract

Nowadays people often form their opinion about various subjects according to the reviews they read on the Internet. Measuring the quality of products or services is a complicated task based on revealing customers' satisfaction or sentiment. However, having the large amounts of data, it is hardly possible to process it manually and it also takes a long time to read all product reviews. This paper presents the analysis and design of the application for mobile devices that aims at automatic discovery of human feelings hidden in textual messages that customers produce.

Peer-review under responsibility of the organizing committee of ECE 2016

Keywords: Product evaluation; customers' reviews; opinion mining; word cloud; mobile device

1. Introduction

In the modern era of information and communication technologies, it has become quite common that customers create their opinion not just talking to their friends or reading expert reviews in magazines but also reading reviews of other customers on the Internet. Majority of such data is in the form of textual documents. Their sources include electronic markets, recommender systems, social networks, blogs, discussion boards and others. These documents potentially contain important knowledge that could be useful for customers to make a right decision to choose a suitable product. Reviews might by helpful in a situation the lone customer is choosing between different products in a shop.

To obtain information about the product a user can use a mobile device and search reviews on the Internet. But collecting and analyzing reviews from many sources about selected products brings a problem. The volume of the

* Corresponding author.
E-mail address: jan.prichystal@mendelu.cz

Peer-review under responsibility of the organizing committee of ECE 2016
doi:10.1016/j.sbspro.2016.05.511

data is huge and people cannot process them manually within an acceptable time. The extensive set of data is not a problem for a computer represented by a special application. What is the problem, computers cannot understand text written in a natural language and cannot realize the opinion or reveal the sentiment of reviews. Despite these difficulties, it is possible – and in situation there are a lot of reviews for a certain product – to use computers for analyzing and uncovering the hidden knowledge. Then it could be presented to a customer in a specific form helping him to understand and make a decision.

This paper presents a method of opinion mining from the online data by a mobile application. In the following sections methods for automatic discovery of customers' opinion are discussed. Based on this the specific approach is selected. The process of application for mobile devices design is described and experiments with freely written reviews containing customers' subjective evaluation of products conducted.

2. Related work

Nowadays, the topic of opinion mining from customers' reviews is frequently discussed. There are a lot of articles with various approaches and methods focusing on this topic. To analyze the sentiment of text documents, lexicon based or machine learning based techniques are often used. The former methods require a lexicon where the words are mapped to their semantic value (Moreo et al., 2012). The problem is that for many tasks there are not enough existing examples, the labeling process is very demanding, or that the lexicons are domain dependent. In order to extract relevant aspects of a certain product or service, some kind of ontologies might be used as well (Iwashita et al., 2011). A rather simple, machine learning based approach combining both unsupervised and supervised techniques was proposed by Žižka & Dařena (2013). The presented method enabled automatic extraction of terms characterizing groups of topics related to a service. A method that aims at automatic discovery of sources of human feelings hidden in textual messages that clients of medical facilities produce was proposed by Žižka, Dařena & Přichystal (2014).

In order to reveal topics that are sources of dissatisfaction with a service, Tsujii et al. (2013) built a synonym dictionary, specified an evaluation expression, and in order to judge affirmative and negative polarity, built an evaluation-expression dictionary. This was, however, a demanding process requiring detailed knowledge of the given domain.

Finding how people perceive individual aspects of a product or service requires identification of these aspects. This might require a linguistic analysis of the document, e.g., finding nouns or noun phrases (Popescu & Etzioni, 2005). Alternatively, clustering might be used in order to automatically reveal the aspects (Dařena, Žižka & Burda, 2012). However, there exists an additional problem with their validity that must be often confirmed by a human expert.

The reason such methods are omitted in the process of data analysis in our application is they are time consuming and the result is not presented to user in within a reasonable time. These methods focus more on accuracy of the result but we require speed and settle for less accurate because the judgment makes a customer himself.

When there exists a set of documents related to one topic, a multi-document summarization approach to find the most important parts of the documents can be used. For example, Britsom, Bronselaer & Tré (2012) used a multiset merging approach to summarize a small set of strongly coherent newspaper articles. But such approach could by unnecessarily complicated. The type of data our application analyzes consists of strongly related texts focused on customers' opinion of certain product. The summarization often involves extraction of sentences with words occurring frequently (Haghighi & Vanderwende, 2009) which sometimes doesn't bring acceptable results but in our situation could be acceptable. This approach requires having the group of strongly related documents sharing one common topic or sentiment.

Because we focus on analysis of customers' products reviews a study of this topic on Amazon.com presented by Mudambi & Schuff (2010) or Liu (2012) provide us an important approach. Several novel techniques to perform mining product features that have been commented on by customers and identifying opinion sentences in each review and deciding whether each opinion sentence is positive or negative propose Hu and Liu (2004).

A survey covering the techniques and methods in sentiment analysis and challenges appearing in the field present Vinodhini & Chandrasekaran (2012).