



ELECTRICAL ENGINEERING

Sentiment analysis algorithms and applications: A survey



Walaa Medhat ^{a,*}, Ahmed Hassan ^b, Hoda Korashy ^b

^a School of Electronic Engineering, Canadian International College, Cairo Campus of CBU, Egypt

^b Ain Shams University, Faculty of Engineering, Computers & Systems Department, Egypt

Received 8 September 2013; revised 8 April 2014; accepted 19 April 2014

Available online 27 May 2014

KEYWORDS

Sentiment analysis;
Sentiment classification;
Feature selection;
Emotion detection;
Transfer learning;
Building resources

Abstract Sentiment Analysis (SA) is an ongoing field of research in text mining field. SA is the computational treatment of opinions, sentiments and subjectivity of text. This survey paper tackles a comprehensive overview of the last update in this field. Many recently proposed algorithms' enhancements and various SA applications are investigated and presented briefly in this survey. These articles are categorized according to their contributions in the various SA techniques. The related fields to SA (transfer learning, emotion detection, and building resources) that attracted researchers recently are discussed. The main target of this survey is to give nearly full image of SA techniques and the related fields with brief details. The main contributions of this paper include the sophisticated categorizations of a large number of recent articles and the illustration of the recent trend of research in the sentiment analysis and its related areas.

© 2014 Production and hosting by Elsevier B.V. on behalf of Ain Shams University.

1. Introduction

Sentiment Analysis (SA) or Opinion Mining (OM) is the computational study of people's opinions, attitudes and emotions toward an entity. The entity can represent individuals, events or topics. These topics are most likely to be covered by reviews. The two expressions SA or OM are interchangeable. They express a mutual meaning. However, some researchers

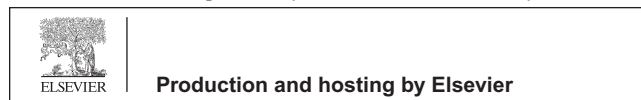
stated that OM and SA have slightly different notions [1]. Opinion Mining extracts and analyzes people's opinion about an entity while Sentiment Analysis identifies the sentiment expressed in a text then analyzes it. Therefore, the target of SA is to find opinions, identify the sentiments they express, and then classify their polarity as shown in Fig. 1.

Sentiment Analysis can be considered a classification process as illustrated in Fig. 1. There are three main classification levels in SA: document-level, sentence-level, and aspect-level SA. Document-level SA aims to classify an opinion document as expressing a positive or negative opinion or sentiment. It considers the whole document a basic information unit (talking about one topic). Sentence-level SA aims to classify sentiment expressed in each sentence. The first step is to identify whether the sentence is subjective or objective. If the sentence is subjective, Sentence-level SA will determine whether the sentence expresses positive or negative opinions. Wilson et al. [2] have pointed out that sentiment expressions

* Corresponding author. Address: School of Electronic Engineering, Canadian International College, 22 Emarate Khalf El-Obour, Masr Elgedida, Cairo, Egypt. Tel.: +20 24049568.

E-mail address: walaamedhat@gmail.com (W. Medhat).

Peer review under responsibility of Ain Shams University.



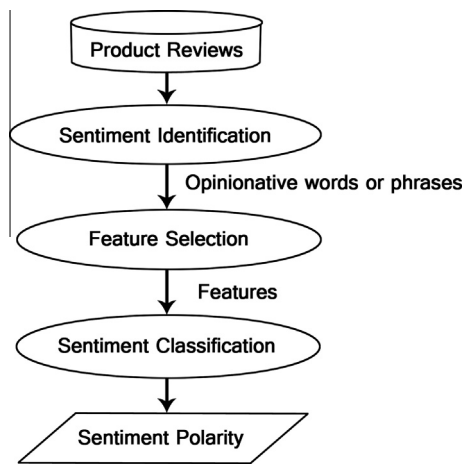


Figure 1 Sentiment analysis process on product reviews.

are not necessarily subjective in nature. However, there is no fundamental difference between document and sentence level classifications because sentences are just short documents [3]. Classifying text at the document level or at the sentence level does not provide the necessary detail needed opinions on all aspects of the entity which is needed in many applications, to obtain these details; we need to go to the aspect level. Aspect-level SA aims to classify the sentiment with respect to the specific aspects of entities. The first step is to identify the entities and their aspects. The opinion holders can give different opinions for different aspects of the same entity like this sentence “*The voice quality of this phone is not good, but the battery life is long*”. This survey tackles the first two kinds of SA.

The data sets used in SA are an important issue in this field. The main sources of data are from the product reviews. These reviews are important to the business holders as they can take business decisions according to the analysis results of users’ opinions about their products. The reviews sources are mainly review sites. SA is not only applied on product reviews but can also be applied on stock markets [4,5], news articles, [6] or political debates [7]. In political debates for example, we could figure out people’s opinions on a certain election candidates or political parties. The election results can also be predicted from political posts. The social network sites and micro-blogging sites are considered a very good source of information because people share and discuss their opinions about a certain topic freely. They are also used as data sources in the SA process.

There are many applications and enhancements on SA algorithms that were proposed in the last few years. This survey aims to give a closer look on these enhancements and to summarize and categorize some articles presented in this field according to the various SA techniques. The authors have collected fifty-four articles which presented important enhancements to the SA field lately. These articles cover a wide variety of SA fields. They were all published in the last few years. They are categorized according to the target of the article illustrating the algorithms and data used in their work. According to Fig. 1, the authors have discussed the Feature Selection (FS) techniques in details along with their related articles referring to some originating references. The Sentiment Classification (SC) techniques, as shown in Fig. 2, are

discussed with more details illustrating related articles and originating references as well.

This survey can be useful for new comer researchers in this field as it covers the most famous SA techniques and applications in one research paper. This survey uniquely gives a refined categorization to the various SA techniques which is not found in other surveys. It discusses also new related fields in SA which have attracted the researchers lately and their corresponding articles. These fields include Emotion Detection (ED), Building Resources (BR) and Transfer Learning (TL). Emotion detection aims to extract and analyze emotions, while the emotions could be explicit or implicit in the sentences. Transfer learning or Cross-Domain classification is concerned with analyzing data from one domain and then using the results in a target domain. Building Resources aims at creating lexica, corpora in which opinion expressions are annotated according to their polarity, and sometimes dictionaries. In this paper, the authors give a closer look on these fields.

There are numerous number of articles presented every year in the SA fields. The number of articles is increasing through years. This creates a need to have survey papers that summarize the recent research trends and directions of SA. The reader can find some sophisticated and detailed surveys including [1,3,8–11]. Those surveys have discussed the problem of SA from the applications point of view not from the SA techniques point of view.

Two long and detailed surveys were presented by Pang and Lee [8] and Liu [3]. They focused on the applications and challenges in SA. Cambria and Schuller et al. [9], Feldman [10] and Montoyo and Martínez-Barco [11] have given short surveys illustrating the new trends in SA. Tsytarau and Palpanas [1] have presented a survey which discussed the main topics of SA in details. For each topic they have illustrated its definition, problems and development and categorized the articles with the aid of tables and graphs. The analysis of the articles presented in this survey is similar to what was given by [1] but with another perspective and different categorization of the articles.

The contribution of this survey is significant for many reasons. First, this survey provides sophisticated categorization of a large number of recent articles according to the techniques used. This angle could help the researchers who are familiar with certain techniques to use them in the SA field and choose the appropriate technique for a certain application. Second, the various techniques of SA are categorized with brief details of the algorithms and their originating references. This can help new comers to the SA field to have a panoramic view on the entire field. Third, the available benchmarks data sets are discussed and categorized according to their use in certain applications. Finally, the survey is enhanced by discussing the related fields to SA including emotion detection, building resources and transfer learning.

This paper is organized as follows: Section 2 includes the survey methodology and a summary of the articles. Section 3 tackles the FS techniques and their related articles, and Section 4 discusses the various SC techniques and the corresponding articles. In Section 5, the related fields to SA and their corresponding articles are presented. Section 6 presents the results and discussions, and finally the conclusion and future trend in research are tackled in Section 7.

Download English Version:

<https://daneshyari.com/en/article/815568>

Download Persian Version:

<https://daneshyari.com/article/815568>

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