Journal of King Saud University – Computer and Information Sciences (2014) **26**, 258–275

King Saud University

Journal of King Saud University – Computer and Information Sciences

www.ksu.edu.sa
www.sciencedirect.com

Mining opinion components from unstructured reviews: A review

Khairullah Khan [a,*], **Baharum Baharudin** [a], **Aurnagzeb Khan** [b], **Ashraf Ullah** [b]

[a] *Computer and Information Sciences Department, Universiti Teknologi, PETRONAS, Malaysia*
[b] *Institute of Engineering and Computing Sciences, University of Science & Technology Bannu, Pakistan*

Received 31 August 2012; revised 31 May 2013; accepted 16 March 2014
Available online 28 May 2014

KEYWORDS

Opinion mining;
Sentiment analysis;
Information retrieval;
Text mining;
Web mining

Abstract Opinion mining is an interesting area of research because of its applications in various fields. Collecting opinions of people about products and about social and political events and problems through the Web is becoming increasingly popular every day. The opinions of users are helpful for the public and for stakeholders when making certain decisions. Opinion mining is a way to retrieve information through search engines, Web blogs and social networks. Because of the huge number of reviews in the form of unstructured text, it is impossible to summarize the information manually. Accordingly, efficient computational methods are needed for mining and summarizing the reviews from corpuses and Web documents. This study presents a systematic literature survey regarding the computational techniques, models and algorithms for mining opinion components from unstructured reviews.

Contents

* Corresponding author. Tel.: +60 1609702807.
E-mail addresses: Khairullah_k@yahoo.com (K. Khan), baharbh@pertonas.edu.my (B. Baharudin), Aurangzebb_khan@yahoo.com (A. Khan).

Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

1319-1578 © 2014 King Saud University. Production and hosting by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jksuci.2014.03.009

1. Introduction

This study presents a systematic literature survey that contains a comprehensive overview of recent research trends, advances, and challenges. The aim of this study is to provide researchers and students access to the latest works in opinion mining as they frame new ideas and further develop the practice.

There has been an increase in research in this area as evidenced by the recent publication of several research survey papers in the past few years (Khan et al., 2009; Pang and Lee, 2008; Tang et al., 2009, Tsytsarau and Palpanas, 2011). Pang and Lee (2008), for example, present an extensive review of opinion mining (OM) concepts and techniques. Khan et al. (2009) provide a short overview of the published works regarding the various issues in the domain of opinion mining. Tang et al. (2009) review the techniques regarding sentiment analysis and polarity classification. Tsytsarau and Palpanas (2011) focus on summarizing opinions and analyzing contradictions. They also present a comparative analysis of machine learning algorithms for sentiment classification. This paper reviews the various advancements in OM research since 2008. Thus, the proposed work presents a review of opinion mining based on opinion component analysis of unstructured text, and accordingly, this paper differs from existing papers in several ways. In this work, we discuss citations published after 2008 that are related to the opinion component of unstructured reviews. We have divided the papers according to the sub-tasks related to opinion mining. These include subjectivity and polarity classification, opinion target extraction, opinion source identification and opinion summarization. Each section presents a comprehensive literature review about the related sub-task. Some new directions have been explored, e.g., features grouping, opinion target identification and semantic-based relevance scoring through lexical resources and concept-based analysis.

This paper is organized as follows. Section 2 presents a general overview of the opinion mining problems, its applications, and related areas. Section 3 explains technical perspectives of opinion mining based on opinion components. Section 4 presents opinion summaries, Section 5 provides an overview of challenges and issues, and Section 6 concludes the paper.

2. Opinion mining

An opinion is the private state of an individual, and as such, it represents the individual's ideas, beliefs, assessments, judgments and evaluations about a specific subject/topic/item. Liu et al. (2012) conclude that others' opinions have a great impact on and provide guidance for individuals, governments, organizations and social communities during the decision-making process. During this process, human beings require fast, accurate and concise information so they can make quick and accurate decisions. Through opinions, humans can integrate the diverse approaches, experiences, wisdom and knowledge of many people when making decisions. It is quite natural for people to participate in discussions and express their points of view. People often ask their friends, family members, and field experts for information during the decision-making process, and their opinions and perspectives are based on experiences, observations, concepts, and beliefs. One's perspective about a subject can either be positive or negative, which is referred to as the polarity of the opinion.

Opinions can be expressed in different ways. The following are examples of opinion statements.

Shahid Afridi is a good player.
She is not a good actress.
The breakfast was quite good.
The hotel was expensive.
Terrorists deserve no mercy!
Hotel A is more expensive than Hotel B.
Coffee is expensive, but tea is cheap.
This player is not worth any price, and I recommend that you not purchase it.

An opinion has three main components, i.e., the opinion holder or source of the opinion, the object about which the opinion is expressed and the evaluation, view or appraisal, that is, the opinion. For opinion identification, all of these components are important.

While opinions can be collected from different sources, e.g., individual interactions, newspapers, television, Internet etc., the Internet has become the richest source of opinion collection. Before the World Wide Web (www), people collected opinions manually. If an individual was to make a decision, he/she typically asked for opinions from friends and family members. To acquire public opinion, organizations often conducted surveys through focused groups. This type of survey, however, was expensive and laborious. Now, the Internet provides this information with a single click and at very little cost.

With the advent of Web 2.0, the Internet allows Web users to generate Web content online and post their information independently. This aspect of the Internet allows Web users to participate in collaborative global environments. Hence, the Internet has become a rich source for social networks, customer feedback, online shopping etc. According to a survey, more than 45,000 new blogs are created daily along with 1.2 million new posts each day (Pang and Lee, 2008). The information collected through these services is used for various types