# Implicit aspect extraction in sentiment analysis: Review, taxonomy, oppportunities, and open challenges

Mohammad Tubishat[a], Norisma Idris[a,*], Mohammad A.M. Abushariah[b]

[a] *Department of Artificial Intelligence, Faculty of Computer Science and Information Technology, University of Malaya, 50603 Kuala Lumpur, Malaysia*
[b] *Computer Information Systems Department, King Abdullah II School of Information Technology, The University of Jordan, Amman, Jordan*

## ARTICLE INFO

## ABSTRACT

Sentiment analysis is a text classification branch, which is defined as the process of extracting sentiment terms (i.e. feature/aspect, or opinion) and determining their opinion semantic orientation. At aspect level, aspect extraction is the core task for sentiment analysis which can either be implicit or explicit aspects. The growth of sentiment analysis has resulted in the emergence of various techniques for both explicit and implicit aspect extraction. However, majority of the research attempts targeted explicit aspect extraction, which indicates that there is a lack of research on implicit aspect extraction. This research provides a review of implicit aspect/ features extraction techniques from different perspectives. The first perspective is making a comparison analysis for the techniques available for implicit term extraction with a brief summary of each technique. The second perspective is classifying and comparing the performance, datasets, language used, and shortcomings of the available techniques. In this study, over 50 articles have been reviewed, however, only 45 articles on implicit aspect extraction that span from 2005 to 2016 were analyzed and discussed. Majority of the researchers on implicit aspects extraction rely heavily on unsupervised methods in their research, which makes about 64% of the 45 articles, followed by supervised methods of about 27%, and lastly semi-supervised of 9%. In addition, 25 articles conducted the research work solely on product reviews, and 5 articles conducted their research work using product reviews jointly with other types of data, which makes product review datasets the most frequently used data type compared to other types. Furthermore, research on implicit aspect features extraction has focused on English and Chinese languages compared to other languages. Finally, this review also provides recommendations for future research directions and open problems.

## 1. Introduction

In the last decade, the research community has witnessed various technological improvements and increase in the internet activities such as e-commerce, discussion forums, chatting, merchant's/manufacture's websites, social media communications, and many other online activities, which were able to provide positive impact on many research initiatives (Ravi & Ravi, 2015).

Sentiment Analysis (SA) or Opinion Mining (OM) is among the fields that benefited from these technological advancements and the internet, which is generally defined as the computerized process of recognizing, detecting, and determining the orientation of human opinion or emotion, which is directed towards different entities. Individuals normally post their opinions or emotions for

products, services, hotels, movies, restaurants, political issues, or any other entity of their interest. Opinions, sentiments, and emotions can be captured using the individual's writings, facial expressions, speech, and many other media (Yadollahi, Shahraki, & Zaiane, 2017)

From applications perspective, there are many applications of SA in our daily life, where SA can be used for monitoring and analyzing the public opinions regarding political issues. SA can also be used in market intelligence (Li & Li, 2013), measuring the degree of user satisfaction on products or services and improving their weaknesses (Kang & Park, 2014), forecasting of price changes according to news sentiments, developing new products and services, and promoting and improving products according to customers' reviews. Being more trustworthy products and services' reviews as posted by their users compared to the vendor's reviews, many individuals rely on these reviews to make their decisions about the products, services, and other entities (Ravi & Ravi, 2015).

According to the previous studies (Medhat, Hassan, & Korashy, 2014; Rana & Cheah, 2016; Yadollahi et al., 2017), SA can be classified into three main levels, which are document level, sentence level, and aspect level. In document and sentence level the main aim to find the sentiment of overall document or sentence, while in aspect level the task to find the sentiment of each aspect as a single unit. In order to develop and evaluate SA at aspect level, features extraction is a crucial process, which can either be implicit or explicit.

This paper presents a review that aims to provide a comprehensive overview on different studies done on implicit aspect extraction, which to the best of our knowledge, serves as the first comprehensive review for implicit aspect extraction in SA. This review classifies and summarizes all the proposed research works on implicit aspect extraction and provides researchers with the current state-of-the-art in this field in order to assist them for further findings, and possible directions for new algorithms and improvements to the available works. Various taxonomies, open challenges, and future directions on implicit aspect extraction in SA are also highlighted in this review.

The rest of the review is organized as follows: in Section 2, we present the taxonomy of SA and background information towards the review. In addition, we provide a summary of different implicit aspects extraction techniques according to the extraction method used, which are either unsupervised, semi-supervised, or supervised. In Section 3, we present the results and discuss the analysis of the results and highlight some future researches and open problems in Section 4. We finally present the conclusions in Section 5.

## 2. Taxonomy of sentiment analysis

From the previous studies (Medhat et al., 2014; Rana & Cheah, 2016; Yadollahi et al., 2017), sentiment analysis can be classified into document, sentence, and aspect levels as illustrated in Fig. 1.

From document level perspective, the task is to extract all opinion words inside the whole document, which can either be long or short, in order to determine the polarity of the overall document without finding the polarity of each feature as a single case. The result in this case will be the overall opinion of the document. In the document level, SA considers the whole document as one topic and decides whether the overall opinion of the document is a positive opinion or a negative opinion based on some opinion words. This type of SA is very crucial for applications within social and psychological studies in social networks, consumer satisfaction, analyzing patients in medical settings, and many others (Yadollahi et al., 2017).

In addition, from sentence level perspective, the process is to find the polarity at the overall sentence without considering each feature as a single case and provide the opinion at the overall sentence level. As early as possible, it is important to identify whether the target sentence is subjective or objective and decide whether the overall opinion of the sentence is a positive opinion or a negative opinion for subjective sentences that are considered small documents. This type of SA is normally influenced by the surrounding context of the sentence, and is considered very crucial for applications that deal with tweets, Facebook posts and comments, short messages, and many others.

Finally, the aspect level that is also known as feature level is a fine-grained model of SA that deals with determining the opinion intended by people to a specific feature (aspect) of a product, service, or any entity (Medhat et al., 2014; Rana & Cheah, 2016; Yadollahi et al., 2017). In order to conduct a SA task based on aspect level, it is essential to extract the entities and their corresponding aspects/features or also referred to as opinion targets and their opinion words from given opinionated reviews. Thereafter,
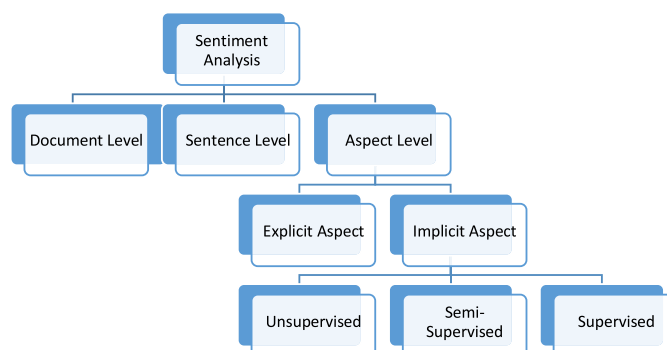


**Fig. 1.** Taxonomy of sentiment analysis.