Received 7 April 2016
Revised 11 August 2016
11 October 2016
10 November 2016
Accepted 9 December 2016

A systematic literature review on opinion types and sentiment analysis techniques

Tasks and challenges

Atika Qazi
Department of Artificial Intelligence, University of Malaya, Kuala Lumpur, Malaysia;
Centre for Life Long Learning, University Brunei Darussalam, Gadong, Brunei Darussalam and
Faculty of Computer Science and Information Technology, COMSATS Institute of Information Technology, Islamabad, Pakistan
Ram Gopal Raj
Department of Artificial Intelligence, University of Malaya, Kuala Lumpur, Malaysia
Glenn Hardaker
Centre for Lifelong Learning and Sultan Hassanal Bolkiah Institute of Education, University Brunei Darussalam, Gadong, Brunei Darussalam, and
Craig Standing
Centre for Innovative Practice, Edith Cowan University, Perth, Australia

Abstract

Purpose – The purpose of this paper is to map the evidence provided on the review types, and explain the challenges faced by classification techniques in sentiment analysis (SA). The aim is to understand how traditional classification technique issues can be addressed through the adoption of improved methods.
Design/methodology/approach – A systematic review of literature was used to search published articles between 2002 and 2014 and identified 24 papers that discuss regular, comparative, and suggestive reviews and the related SA techniques. The authors formulated and applied specific inclusion and exclusion criteria in two distinct rounds to determine the most relevant studies for the research goal.
Findings – The review identified nine practices of review types, eight standard machine learning classification techniques and seven practices of concept learning Sentic computing techniques. This paper offers insights on promising concept-based approaches to SA, which leverage commonsense knowledge and linguistics for tasks such as polarity detection. The practical implications are also explained in this review.
Research limitations/implications – The findings provide information for researchers and traders to consider in relation to a variety of techniques for SA such as Sentic computing and multiple opinion types such as suggestive opinions.
Originality/value – Previous literature review studies in the field of SA have used simple literature review to find the tasks and challenges in the field. In this study, a systematic literature review is conducted to find the more specific answers to the proposed research questions. This type of study has not been conducted in the field previously and so provides a novel contribution. Systematic reviews help to reduce implicit researcher bias. Through adoption of broad search strategies, predefined search strings and uniform inclusion and exclusion criteria, systematic reviews effectively force researchers to search for studies beyond their own subject areas and networks.

Internet Research
Vol. 27 No. 3, 2017
pp. 608-630

1066-2243
DOI 10.1108/IntR-04-2016-0086

This project is supported by IPPP Research Grant RP026-14AET from University Malaya, Malaysia. The authors are also thankful to BRC-12 by University Brunei Darussalam, Brunei, for their support in this research.

Keywords Systematic review, Opinion mining, Sentiment analysis, Comparative sentences, Sentic computing, Suggestive sentences
Paper type Literature review

1. Introduction

Sentiment analysis (SA) and opinion mining (OM) are currently widely studied research fields. Sentiments, comprising opinions, attitudes, thoughts, judgments and emotions, are private states of individuals, even though generally demanded by conventional scientific methods (Thet *et al.*, 2010; Yu and Hatzivassiloglou, 2003). These feelings are expressed in languages by means of subjective expressions (Quigley, 2008). There are many ways to analyze sentiments, including machine learning approaches (both supervised and unsupervised learning methods), lexicon-based, keyword-based, and concept-based approaches (Cambria, 2016). So far, most of the research work carried out in the context of SA has been unimodal, i.e., related to text, but recently some works have started investigating the same issues in multiple modalities, e.g., speech and video (Poria *et al.*, 2015; Poria, Cambria and Gelbukh, 2016; Poria, Cambria, Hazarika and Vij, 2016; Poria, Cambria, Howard, Huang and Hussain, 2016; Poria, Chaturvedi, Cambria and Hussain, 2016).

SA is a "suitcase" research problem that requires to tackle many NLP subtasks, including aspect extraction (Poria, Cambria and Gelbukh, 2016), subjectivity detection (Chaturvedi *et al.*, 2016), named entity recognition (Ma *et al.*, 2016), and sarcasm detection (Poria, Cambria, Hazarika and Vij, 2016). The main goal of SA is to extract opinions about entities (such as products or services) in order to attain useful information. Moreover, its purpose is to present the information in such a way that serves the objective of both customers and manufacturers. It is in fact demonstrated that many potential buyers and manufacturers overlook detailed reviews, preferring summarized opinions. Such a summary, generally based on the concept of simple polarity of opinions (negative, positive, neutral), specifies whether opinions hold positive, negative or neutral sentiments (Cambria and White, 2014). An opinion could be simply defined as a positive or negative sentiment, view, attitude, emotion or appraisal about an entity (product, person, event, organization or topic) or an aspect of that entity from a user or group of users. SA is carried out over a huge variety of public opinions, which could be of different types.

We identify three main opinion types: regular opinions, which pertain to a single entity only; comparative opinions, which juxtapose two or more entities; and suggestive opinions that suggest a single or multiple entities (Qazi, Raj, Tahir, Cambria and Syed, 2014). The regular opinion is mostly used to find the good or bad views about a particular product (Jindal and Liu, 2006a). The comparative opinions are those that explain relations among multiple entities, where these relations can be understood with respect to some common features. There is a notable paucity of literature regarding the recognition of comparative sentences, in which comparisons between multiple entities are made. The comparative opinions are significantly utilized for competitive intelligence (Jindal and Liu, 2006a). The suggestions were recently introduced as a third type of review in the study of OM (Qazi, Raj, Tahir, Cambria and Syed, 2014). Extracting suggestive sentences from text is valuable for numerous applications in business, medical and e-learning environments, among others. These multiple types of opinions serve multiple needs of web users and can also be used to determine the perceived helpfulness of online users (Qazi *et al.*, 2016) as well as assessing consumers' satisfaction and expectations (Qazi *et al.*, 2017). This paper examines the opinion types present in the vast number of user-generated reviews available on the internet. All these types play a significant role in business decision making. In particular, an important aspect of every business is the effective analysis of information related to their product or service type, in order to prevail over competitors (Qazi *et al.*, 2013).

The wide range of internet technologies have increased the number of consumers that are using online reviews for making better and well-informed buying decisions (Lee *et al.*, 2011; Tamimi and Sebastianelli, 2015). For example, the research shows that a greater number of users are now showing a positive attitude toward online shopping in Jordan than in the recent past (Al-Debei *et al.*, 2015). The evidence of this trend is seen almost everywhere from business to other sectors such as education and even healthcare, resulting in an ever-growing availability of online reviews. A variety of research work is carried out encompassing the success of e-learning systems (Waheed *et al.*, 2016), and e-recruitment systems (Faliagka *et al.*, 2012). Online product reviews make it easier than ever before for consumers to learn more about products from other consumers for better decision making. The valence of consumer reviews such as helpfulness, star rating and facial avatar plays an important role in purchase decision making (Lee *et al.*, 2013).

As the volume of information continues to increase on the internet, it becomes increasingly difficult and time consuming to make decisions about the purchase of products and services. Opinion and SA are techniques that aim to automate the analysis of information and thereby save the user time and effort. It can be viewed as an emerging area of significance in the computing discipline, relying on a combination of new and established methods. Our study identifies recent research issues in different types of opinions in order to determine knowledge gaps and to highlight future challenges.

The existing OM approaches mostly use keyword-based reasoning and rely on the vector space representation based on text features. Although these techniques are useful when dealing with SA, presentenced domain-dependent concepts may make purely syntactical approaches ineffective. Therefore, commonsense computing techniques try to bridge the cognitive and affective gap between word-level natural language data and concept-level opinions (Cambria *et al.*, 2009). Such approaches should use new techniques capable of better grasping the conceptual rules that govern sentiment and the clues that can convey these concepts from realization to verbalization in the human mind. We are motivated to highlight this gap between keyword-based approaches and common sense knowledge for online opinions used for SA related to multiple opinion types. The purpose is to deepen the analysis of multiple opinion types, their benefits and how the issues raised from keyword-based approaches can be solved by common sense knowledge-based approaches.

This review paper is organized as follows: Section 2 presents our research questions and the method followed for the review of opinion types and techniques used for SA and Section 3 summaries the key findings of our study. We give a detailed account of the present state-of-the-art of SA, describing a large number of methods and their particular shortcomings and strengths; this is done with regard to regular, comparative and suggestive opinions as well as the employed machine learning and concept-based approaches; Section 4 provides a discussion on the presented studies; conclusions are given in Section 5, in which we reiterate outstanding challenges of this rapidly evolving field, as well as the insights into the future direction of research and applications in industry.

2. Research method

The present study consists of a systematic literature review with a specific focus on research related to opinion types and SA approaches. We follow the guidelines proposed by Keele (2007), Brereton *et al.* (2007) and Kitchenham *et al.* (2010). The purpose of a systematic review of literature is three-fold (Brereton *et al.*, 2007): to plan the review, to conduct the review, and to report the review results. This leads to the development of collective insights based on a theoretical synthesis of existing studies. The design of the systematic review reported in this paper started in September 2014. After several refinements and improvements, the publication search was started in March 2015.

2.1 Planning the review

We planned this review by proposing the research questions relevant to our research objectives. We defined the search strategy, search strings and inclusion/exclusion criteria. We present these issues in detail below.

2.1.1 Review objectives and research questions. With the rapid upsurge of web 2.0 and the increase of the use of public opinions, it has become important to study the opinion types and the approaches used for their analysis (Cambria, Wang and White, 2014). Therefore, the main goal of this work is to develop a deep understanding of the opinion types and SA approaches performed over them. This study advocates the ways opinion types are present and how Sentic computing techniques are more significant than other keyword-based approaches. Since research questions guided the design of the review process, specifying them is the most important part of any systematic review (Seuring and Müller, 2008; White and Schmidt, 2005). To fulfill these objectives, the research questions were formulated as follows:

RQ1. How many opinion types are present in online reviews according to published studies?

RQ2. Which are the tasks resolved by machine learning techniques for SA of opinions?

RQ3. Which are the Sentic computing techniques that resolved the challenges for SA of opinions?

2.1.2 Search strategy. The study provided by Keele (2007) was used as a guideline for carrying out the research. After defining our research goals and questions, we started with the formulation of a formal search strategy to analyze all available empirical materials specific to the objective of this review. The plan involved defining the search space, which included electronic databases and printed proceedings as shown in Table I. The studies were initially retrieved from the electronic databases and then analyzed to identify other meaningful studies through reference searches (snowballing). Then the inclusion and exclusion criteria were applied on the retrieved studies in two distinct rounds as explained in Section 2.1.4.

2.1.3 Search criteria. The search criteria used for this review consist of two parts C1 and C2, defined as follows:

- C1 is a string made up of keywords related to opinion types such as mining opinion types in online reviews, comparative opinion sentences, comparative opinions, comparative sentences and suggestive sentences.
- C2 is a string made up of keywords related to SA such as concept learning for SA, concept-level approaches for OM, machine learning approaches for OM, supervised machine learning for SA and unsupervised machine learning for OM.

Electronic databases	ACM IEEE Xplore Science Direct ISI Web of Knowledge Springer Link Wiley
Searched items	Journals, books and conference papers
Search applied on	Full text to find paper under scope and not miss any of the papers that do not include our search keywords in titles or abstracts
Language	English
Publication period	July 2002-October 2014

Table I.
Search sources

Boolean expression search criteria:

$$\text{C1 AND C2} \tag{1}$$

An example of a search done in the electronic databases is shown in Table II.

2.1.4 Inclusion and exclusion criteria. To determine whether a study should be included, the inclusion and exclusion criteria were chosen as follows. Inclusion criteria: (I1) the study is a peer-reviewed publication, (I2) the study is in English, (I3) the study is relevant to the search terms defined in Section 2.1.3, (I4) the study is an empirical research paper, an experience report or workshop paper, (I5) the study is published between July 2002 and November 2014. Exclusion criteria: (E1) studies that do not focus explicitly on opinion types, (E2) studies that do not discuss SA in OM, (E3) studies that do not meet the inclusion criteria, (E4) viewpoints, keynotes, discussions, editorials, comments, tutorials, prefaces and presentations in slide formats without any associated paper.

2.2 Conducting the review

In this section, we present the findings of our search and extraction of information from relevant sources and databases.

2.2.1 Study search and selection. By following the search strategy (previously explained in Section 2.1.2), the selected electronic databases were searched and the studies retrieved. In this original search, we retrieved 280 studies as shown in Table III. An extensive inspection of the studies' titles and abstracts was made by a researcher (Round 1) by applying the inclusion criteria. As a result of this first round of classification, we ended up with 49 studies. Then, in Round 2, the preselected studies were assessed by a second (one of the co-authors) and a third (independent and experienced) researcher in order to apply the exclusion criteria (E1, E2, E3 and E4). To review the agreements and disagreements raised by the researchers in their assessments, we conducted a face-to-face consensus meeting. For the papers where consensus was not reached, the three researchers read the entire paper and then excluded the studies based on the defined exclusion criteria. Therefore, the final

Table II. Keywords operationalized for search

Search clouds		
Opinion types	Sentiment analysis	Exemplary search string
Comparative sentences, regular sentences, suggestive sentences	Supervised learning, unsupervised learning, machine learning, Sentic computing	Opinion types AND (comparative OR regular OR suggestive OR comparative sentences OR suggestive sentences) AND Sentiment analysis AND (machine learning OR supervised learning OR unsupervised learning OR Sentic computing)

Table III. Number of filtered publications according to search terms

Database	Retrieved	Round 1 Included	Round 1 Excluded	Round 2 Included	Round 2 Excluded
ACM	70	20	50	5	15
IEEE Xplore	38	14	24	4	10
Science Direct	121	25	96	9	16
Wiley	15	7	8	1	6
Springer Link	16	4	12	1	3
ISI Web of Knowledge	20	5	15	4	1
Total	280	75	205	24	51

selection consisted of 24 studies (see in Table III). The complete list of studies is available in the Appendix, at the end of this paper.

2.2.2 Data extraction and analysis. According to the guidelines provided by Keele (2007), we defined a data extraction process to identify the relevant information from the 24 included primary studies that pertained to our research questions. Our data extraction process included the following: first, we set up a form to record ideas, concepts, contributions, and findings of each of the 24 studies; using this form ensured subsequent higher-order interpretation. The following data were extracted from each publication: review date, title, authors, reference; database, relevance to the theme, i.e. opinion types, SA techniques, methods and approaches; future work; comparative analysis; year of publication.

Once the extraction was completed, we used content analysis to characterize the focus of each study (Elo and Kyngäs, 2008; Hsieh and Shannon, 2005). Content analysis has been used for different studies, and more recently used for systematic review by Qazi *et al.* (2015). We assessed the results of data extraction by using inter-rater agreement between two researchers by using the κ coefficient (Fleiss *et al.*, 2004). κ is a statistical measure of agreement (Cohen, 1968) and equal to 0.5 for this study. The value shows a good agreement, where this definition derives from Landis and Koch (1977). Subsequently, independent quality assessments were conducted for 24 studies.

2.2.3 Methodological quality assessment. The study quality assessment can be used to guide the interpretation of the synthesis findings (Keele, 2007). This systematic review used the quality criterion as used by other researchers (Dybå and Dingsøyr, 2008; Inayat *et al.*, 2015; Pacheco and Garcia, 2012). The quality of each accepted study was evaluated according to the criteria shown in Table IV. With the first criterion (C1) we assessed if the authors of the study clearly state the aims and objectives of the conducted research. This question was answered positively for 91 percent of the studies. With the second criterion (C2), we asked whether the research context was properly addressed and described. This question was answered positively by 87 percent of the studies. The last question allowed us to assess if the outcome of the research was sufficient for our research purpose. The heuristic scores for the quality measures (C3) were established by a group of two experienced researchers and validated by our independent reviewer. We normalized the data for the 24 papers, combining the percentage obtained in the quality criterion (see Table V).

3. Findings of our review

In this section, we describe the findings of our review in light of our research questions.

Criteria	Response grading	Grade obtained
(C1) Is the research aim/objective clearly defined?	Yes = 1/moderately = 0.5/no = 0	22 studies 91%
(C2) Is the context of research well addressed?	Yes = 1/moderately = 0.5/no = 0	21 studies 87%
(C3) Based on the findings, for a paper, what is the acceptance quality rate?	> 80% = 1/under 20% = 0/between = 0.5	–

Table IV. Quality assessment criteria for study selection

	Quality (scores)					
	Poor (< 26%)	Fair (26-45%)	Good (46-65%)	Very good (66-85%)	Excellent	Total
Number of studies	2	1	3	5	13	24
Percentage of papers	8.4	4.1	12.5	20.8	54.1	100

Table V. Quality scores of accepted papers

3.1 Overview of studies

As previously mentioned, we identified 24 studies. Out of the 24 studies, about 33 percent (eight of them) were published in conferences, 63 percent (15 of them) in journals and 4 percent (one only) in a workshop. Regarding the focus of the 24 studies, Figure 1 presents the stages of the study selection process that was adopted for the systematic review presented in this paper. Figure 2 indicates that 38 percent regard review types. With respect

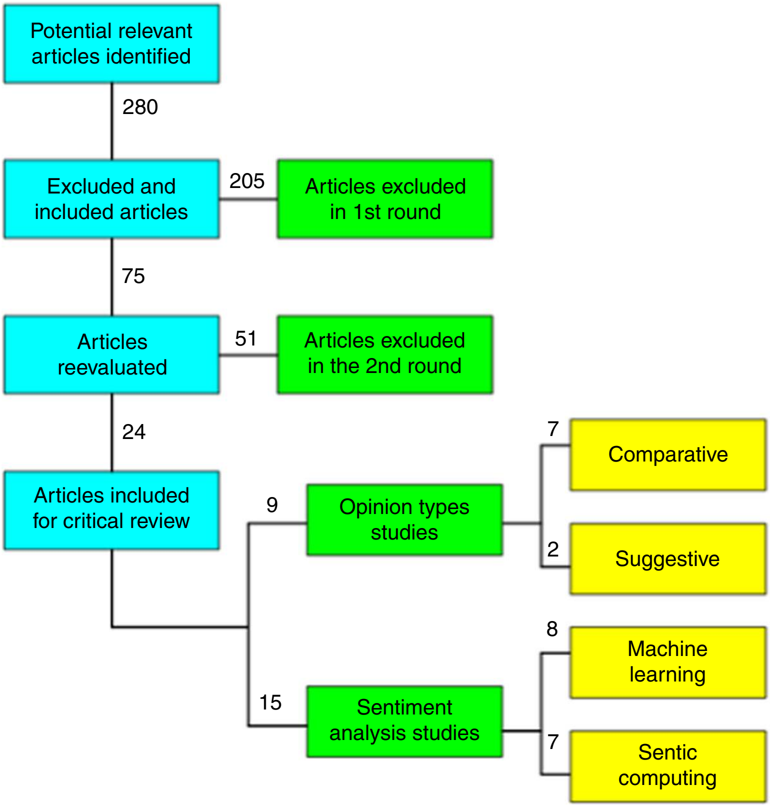

Figure 1.
Stages of the study selection process

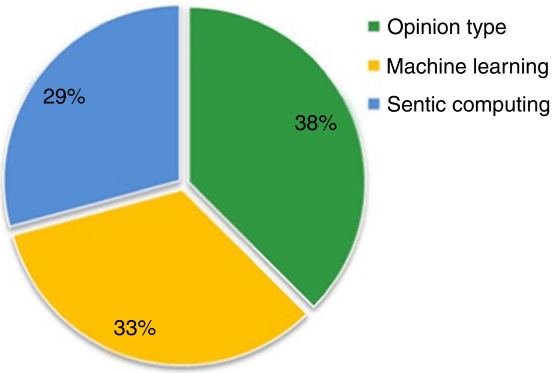

Figure 2.
Categorization of basic studies

to SA techniques, 33 percent employ machine learning techniques and 29 percent Sentic computing techniques. The distribution of the reviewed papers, which were published from 2002 to 2014, is presented in Figure 3.

These studies focus on overall opinion types (9 out of 24), on the SA task using machine learning (8 out of 24), and on Sentic computing (7 out of 24). These studies are evaluated using the precision, recall, accuracy and *F*-measure.

3.2 Evaluation criteria of the sentiment study

OM and SA are typically evaluated by experiments. The experimental evaluation of a classifier measures its effectiveness; in this case the accuracy measure of classification is carried out by different measures such as precision, recall, accuracy, correlation coefficient and relative errors. Most of the classification techniques are measured by precision and recall. However, cases vary in some scenarios. In this study, the values for precision and recall are noted at the end of each summary.

3.3 RQ1. How many opinion types are present in online reviews according to published studies?

Web 2.0 acts as an interactive platform for users to share their views, sentiments and opinions as reviews (postings). These posted data (or opinions) are generally of two types, subjective and objective, but mostly include subjective expressions (Liu, 2010). The opinion types, such as regular (A), comparative (B) and suggestive (C), are differentiated based on language constructs, in which each type expresses a different type of information (Ganapathibhotla and Liu, 2008; Jindal and Liu, 2006c; Qazi, Raj, Tahir, Waheed, Khan and Abraham, 2014). So far online reviews have been classified into these three types: regular, comparative and suggestive. The classification of opinions takes place when comparative and suggestive sentences are extracted from the bulk of opinions, e.g. regular opinions. With respect to this classification comparative and suggestive are further discussed and explained. Figure 4 presents the classified types of opinions with respect to the present literature.

The comparative, regular and suggestive opinion sentences and their classification approaches identified in our review are discussed below. Table VI summaries these studies with the research focus approach and output as a result.

Comparative sentences are identified using a class sequentail rule and a machine learning algorithm, as presented by Jindal and Liu (2006c). With this technique, comparative sentences are first categorized into gradable (e.g. greater or less than, equal to, all other types of relations) and non-gradable, and then the sentences are identified from the document. By using this approach, the comparative sentences are classified and identified, and this can help to improve business intelligence in terms of competitive comparison

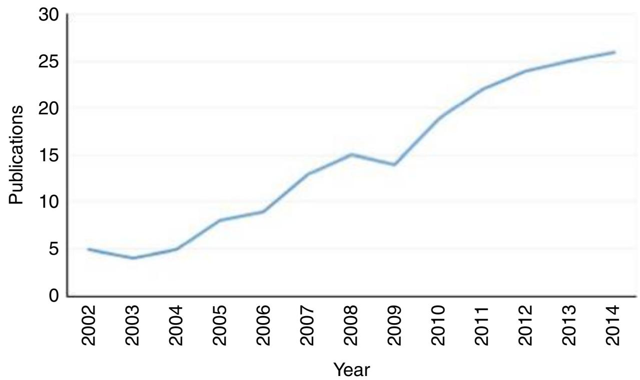

Figure 3.
Year-wise distribution of selected studies

among products. Comparative relation extraction for identified comparatives is discussed by Jindal and Liu (2006b). This method is based on two types of sequential rules: class sequential rules, used in the classification of sentences, and label sequential rules, used in relation item extraction. All the rules are then applied to match comparatives in the test data in order to extract the components of the relation. Finally, the comparative relation is extracted and used to determine the direction of a comparative. Comparative relation map is constructed for decision making, and collects reviews and then extracts simple linguistic features (Xu, Liao, Li and Song, 2011). In this way, three types of entities can be recognized: product names, attribute names, and sentiment phrases. Once entities are recognized, comparative relation extraction takes place. Finally, the comparative relation map is constructed, illustrating the comparisons among entities. Product strengths and weaknesses identification is carried out by a novel SVM-based method proposed by Xu, Wang, Ren, Xu, Liu and Liao (2011) which identifies product strengths and weaknesses by automatically identifying comparative opinions. Chinese comparative sentences within text documents are identified by using semantic role labeling technique by Hou and Li (2008). Preferred entities identification from comparative sentences is obtained by the method proposed by Ganapathibhotla and Liu (2008) that extends the lexicon-based approach. The technique proposes a set of rules to identify two types of comparative words and superlatives. Subsequently, features are recognized by an algorithm able to deal with context semantics, which determines the opinion orientations of comparatives. Finally, positive orientation is considered to be the preferred entity in making decisions, in comparison with other entities. Comparator mining from comparable questions, which are questions that intend to compare two or more entities explicitly, is described by Li *et al.* (2013). The authors propose a weakly supervised bootstrapping method to find the comparator from comparable questions. The proposed bootstrapping method was applied to source data (60M questions), where 328,364 unique comparator pairs were extracted from 679,909 automatically identified comparative questions.

Suggestive identification has been recently carried out by Qazi, Raj, Tahir, Cambria and Syed (2014) for enhancement of business intelligence. The proposed approach consists of three main steps: classify comparative and suggestive sentences; categorize suggestive sentences into different types, either explicit or implicit locutions; perform SA on the classified reviews. Experimental results for all the three tasks are obtained on a dataset of mobile phone reviews and demonstrate that extending a bag-of-words representation with suggestive and comparative patterns is ideal for distinguishing suggestive sentences. User perception and behavior intention toward using online review systems based upon a technology acceptance model and statistical measures shows that suggestive reviews plays a significant role toward using information (Qazi, Raj, Tahir, Waheed, Khan and Abraham, 2014). The results also depict that type C (suggestive reviews) could be considered a new useful review type in addition to other types, A (regular) and B (comparative).

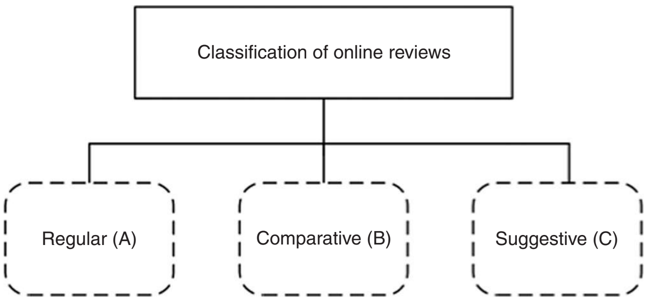

Figure 4.
Classification of online reviews

Reference	Research focus	Dataset	Technique	Precision/ recall; accuracy/ *F*-measure	Output
Jindal and Liu (2006a)	Comparative sentence identification	News articles, consumer reviews of products, internet forum postings	Rule mining and supervised learning	79% 81%	Comparative sentences
Jindal and Liu (2006b)	Identification of comparative sentences in evaluative texts, and extraction of comparative relations	Reviews, forum postings, news articles	Class sequential rules Label sequential rules	97% 88%	Comparative relations
Xu, Liao, Li and Song (2011)	Extraction and visualization of comparative relations between products from customer reviews	Amazon customer reviews	Two-level CRF with unfixed interdependencies and CRF without interdependencies	7.80% 10.21% 22.67% 15.05%	Extracted comparative relation map
Xu, Wang, Ren, Xu, Liu and Liao (2011)	Product strength and weakness	Amazon, mobile reviews	Multi-class SVM	73.15% –	Visual presentation of comparison relation map
Hou and Li (2008)	Chinese comparative sentence mining	Chinese corpus	Supervised learning	82% –	Presentation of comparative sentences
Ganapathibhotla and Liu (2008)	Identification of preferred entities in comparative sentences	Opinions	Point-wise mutual information	– 0.84%	Sentiments expressed in comparative sentences
Li *et al.* (2013)	Comparator finding	Yahoo Questions and answers	Weakly supervised	– 76.8% 82.5% 83.3	Automatic recommendation of comparable original specialized generalized entities
Qazi, Raj, Tahir, Cambria and Syed (2014)	Suggestive sentence identification	Mobile and movie reviews	Supervised learning	0.89% 0.65%	Suggestive sentences
Qazi, Raj, Tahir, Waheed, Khan and Abraham (2014)	User perception and behavior intention toward using online reviews	Questionnaire data	Quantitative method	n/a	Suggestive sentences grab more user attraction

Table VI. Summary of existing techniques for comparative opinion mining

For the functions other than the classification of multiple opinion types, such as SA, different approaches are used which are discussed and presented by research *RQ2* in the subsequent section.

3.4 RQ2. Which are the tasks resolved by machine learning techniques for SA of opinions?

Different supervised and unsupervised machine learning approaches have been used to perform OM or SA tasks. Described below are different tasks carried out by machine learning approaches.

Thumbs up/thumbs down methods have been used for polarity classification. According to this method proposed by Pang *et al.* (2002), the document classification is based on the sentiments, either positive or negative. OPINE was proposed by Popescu *et al.* (2005) and offers product features recognition, opinion identification with respect to product features, identification of opinion polarity and ranking based on opinion strength. Initially, product or service reviews are parsed by MINIPAR (Etzioni *et al.*, 2008), upon which pronoun resolution is applied in order to identify the relations between anaphoric pronouns. OPINE is compared with the technique proposed in Hu and Liu (2004), and the results are displayed in Table VII, where other techniques and results are also discussed. Opinion observer techniques are used to analyze opinions, as proposed by Liu *et al.* (2005). The proposed system is called opinion observer and provides visualization to clearly analyze the strengths and weaknesses of each product. The potential customer can have a visual side-by-side and feature-by-feature comparison of consumer opinions on these

Reference	Research focus	Dataset	Techniques	Precision/ recall; accuracy/ *F*-measure	Output
Pang *et al.* (2002)	Sentiment-based document classification	Movie reviews	Naive Bayes classification, maximum entropy classification, support vector machines	– 82.9%	Best performance with SVM achieved
Popescu *et al.* (2005)	Polarity identification	Amazon	Unsupervised information extraction system, relaxation labeling	79% 76% –	Review polarity identification and presentation
Liu *et al.* (2005)	Comparison of products and polarity identification	Amazon	Supervised pattern mining, NLP	100% 52% –	Summarized graphical presentation of opinions
Ye *et al.* (2009)	Comparison of three classification techniques for sentiment classification	Travel column of Yahoo.com	Naïve Bayes, SVM, and character based N-gram model	– > 80%	Best performance with SVM and N-gram achieved
Cruz *et al.* (2010)	Opinion recognition and classification	Opinions	Knowledge rich semi-supervised approach	0.674 0.577 –	Feature-based opinion extraction system; summarized statistics
Martínez-Cámara *et al.* (2011)	Classification of polarity of Spanish corpus	Spanish film reviews corpus	SVM, Naive Bayes, BBR, KNN, C4.5	– 87.21% 87.01% 87.21%	Successfull sentiment classification
Habernal *et al.* (2014)	Classification of polarity of Czech corpus	Czech social media corpus	Maximum entropy (MaxEnt), support vector machines	– 69.0%	Successful sentiment classification
Smailović *et al.* (2014)	Predictive sentiment analysis	Twitter feeds for selected companies	Support vector machines	– 0.645 64.05%	Forecasting of future stock prices
Popescu *et al.* (2005)	Explicit features extraction and polarity identification	Amazon	Unsupervised information extraction system, relaxation labeling	– 79% 76%	Review polarity identification and presentation

Table VII.
Summary of machine learning techniques for sentiment analysis

products, which help him to decide which product to buy. Experimental results show that the system is highly effective.

The travel blogs domain is used for sentiment classification by using machine learning approaches by Ye *et al.* (2009). Specifically, this study has applied three supervised machine learning algorithms (Naïve Bayes, SVM and the character-based N-gram model) to the classification of sentiment polarities of online reviews. Empirical findings indicated that SVM and N-gram approaches outperformed the Naïve Bayes approach. Knowledge-rich domain-specific resource-based opinion extraction system for small numbers of opinions is proposed by Cruz *et al.* (2010). In this method, initially the domain is defined, and then domain-specific taxonomy related features are considered. The feature taxonomy is described, consisting of product features; two types of features, implicit and explicit, are discussed. Implicit features, identified by some support measurements (Cruz *et al.*, 2010), refer to those that are indirectly expressed by the user. This proposed opinion extraction system is based on independent subtasks, such as explicit feature annotators, implicit feature annotators, negative expression, dominant polarity expression linkers, opinion classifiers, and an opinion extractor pipeline; these subtasks are combined in order to complete the extraction task in its entirety. SA for Spanish was carried out for the first time by Martínez-Cámara *et al.* (2011). In this work different machine learning algorithms (SVM, Naïve Bayes, BBR, KNN, and C4.5) were used to obtain the best performance. A future work is to deal with the problem from a multilingual perspective, i.e. classify the polarity without taking into account the language.

SA for Czech was carried out for the first time by Habernal *et al.* (2014). The Facebook dataset accompanied by human annotation with substantial agreement (Cohen's j 0.66), containing 10,000 posts and freely available for noncommercial purposes was used for this research. Stream-based SA was used for predicting future values of stock prices (Smailović *et al.*, 2014). This research work analyzes whether the sentiment expressed in Twitter feeds, which discuss selected companies and their products, can indicate their stock price changes. The results show that SA of public mood derived from Twitter feeds can be used for predicting forecast movements of individual stock prices. Also, the SVM neutral zone gives the way to classify tweets into the neutral category and proved to be useful for improving the predictive power.

Table VII shows the summary of the above-discussed techniques for SA.

3.5 RQ3. Which are the Sentic computing techniques that resolved the challenges for SA of opinions?

The term Sentic is derived from the Latin word sentire (root of words as sentiment and sentence) and sensus (capability of feeling as commonsense). Sentic computing stands at the junction of affective and commonsense computing. This interdisciplinary approach uses insights from the fields of computer and social sciences and attempts to facilitate processing and interpretation of sentiments and opinions given by the internet users over the web. For the sake of knowledge representation and inference, this comprehensive approach extends the techniques of AI and Semantic Web.

3.5.1 Sentic computing for cognitive and affective information. Sentic computing is required to infer cognitive and affective information from natural language text since it does not only detect the users' emotions but also the associated semantics and, hence, helps developing opinion-mining systems in fields such as Social Web, HCI, and e-health. In this book (Cambria and Hussain, 2015), commonsense computing techniques are further developed and applied to breaking the gap between cognitive and affective word-level natural language data and the concept-level opinions. The recent studies that exploit Sentic computing are discussed below.

Patient-centered applications: in this work, Cambria *et al.* (2010) use a certain number of Sentic computing techniques: Affective Space (Cambria, Fu, Bisio and Poria, 2015) for language visualization and analysis; the Hourglass of Emotions for novel emotion categorization; resources for OM based on a web ontology; the spectral association proposed by Havasi *et al.* (2010), and CF-IOF (concept frequency – inverse opinion frequency) weighting to define and find topic dependent concepts. The structured information is extracted from a patient opinion dataset through Sentic computing that proves helpful, compared to ratings of official healthcare providers. Sentic activation: in this work, Cambria, Olsher and Kwok (2012) proposed a two-level affective reasoning framework named Sentic activation, that in parallel utilize multidimensionality reduction and graph mining techniques processing and SA. The integration of conscious and unconscious reasoning is grasped, and exploited for SA at the concept level. Multidimensionality reduction is carried out for unconscious affective reasoning, where graph mining techniques are exploited for conscious reasoning (Cambria, Gastaldo, Bisio and Zunino, 2015). Brain-inspired software is also developed to test Sentic activation in real world (Cambria, Olsher and Kwok, 2012). The brain-inspired approach consists of four modules: preprocessing, semantic parser, target spotting, and affect interpreter. The results show that SA at word level is successfully performed by using the proposed method. Sentic PROM A: a system that allows to express feelings freely, and evaluate patients' health status and experience in a semi-structured way, standard PROMS (patient reported outcome measures), is proposed by Cambria, Benson, Eckl and Hussain (2012). The Sentic computing techniques and tools such as ConceptNet, CF-IOF weighting, and Spectral association are used for the extraction of semantics, where for Sentics extraction Affective Space, Hourglass of Emotions, and Sentic medoids are exploited. These techniques are used to find the cognitive and affective information associated with patient opinions to break the barriers between the structuredness of questionnaire data and the unstructuredness of natural language data. These kinds of data are different at the structured level but similar at the concept level. In order to aggregate such data and evaluate patient's health, Sentic PROMs successfully exploit the semantic and Sentics linked with patient opinions. Semantic multidimensional scaling: to overcome the gap of computerized OM systems, which mostly offer domain-dependent OM and lack common and common sense knowledge, semantic multidimensional scaling for open-domain SA is proposed by Cambria, Song, Wang and Howard (2014). In this work a leading existing taxonomy of common knowledge, ProBase (Wu *et al.*, 2012) is blended with natural language-based semantic network of commonsense knowledge called ConceptNet. The research aim is to provide a rich resource for implicit and explicit knowledge that is ordered in the human mind. This knowledge can be used to perform reasoning for OM and SA.

Biologically inspired OM engine is used as a novel cognitive model based on the combined use of multidimensional scaling and artificial neural networks (ANNs). The presented work by Cambria *et al.* (2013) involves three key steps: deconstruct natural language text into concepts, encode such concepts as coordinates of a multidimensional vector space, and infer the semantic and affective information associated with them by means of two ANNs. The study shows the effective results to better grasp the nonlinearities of the vector space of affective commonsense knowledge and, hence, improves the performance of the OM engine. EmoSenticSpace is a new framework for affective commonsense reasoning that extends WordNet-Affect and SenticNet (Poria *et al.*, 2014). In this work, EmoSenticNet is used as a dictionary, which provides average polarity and emotion labels for a large number of concepts. First, it is used to extract emotion information to be blended with ConceptNet to obtain EmoSenticSpace. Then, it is utilized to extract the polarity score and emotion category features for three applications. Another framework was introduced for extracting the emotions and the sentiments expressed in the textual data by Loia and Senatore (2014). Human moods are expressed in the form of sentiments that indicate the judgement about an item. This judgement could allow predictions of future changes in market behavior. The model consists of four affective dimensions, each one with six levels of activation. In this work, the sentiments and emotions are modeled as fuzzy sets. The approach has been tested on some sets of documents categories, revealing interesting performance on the global framework processing.

Table VIII shows the summary of the above-discussed techniques for SA.

4. Discussion of the results

The present systematic review identified 24 studies on important aspects of OM, review types, and SA. We discovered that all the studies in this review exploited an experimental research approach. Moreover, one may observe that for experimental settings most of the data are selected from commerce and healthcare sectors. However, it may be crucial to implement these ideas in other sectors such as education.

Furthermore, we have identified three review types (regular, comparative, and suggestive) in response to *RQ1*, which ensure the helpfulness of these types used by reviewers. However, comparative opinion research is still in its infancy, although its significance, and the usefulness of comparisons for consumers should not be underplayed. The comparative sentence identification proposed by Jindal and Liu (2006a) is an efficient scheme that performs well for selected datasets. However, it possesses certain

Reference	Research focus	Dataset	Technique	Output
Cambria *et al.* (2010)	Topic-dependent concepts	SenticNet	Affective space Hourglass of Emotions Sentic patterns	Reliable patient opinion rating
Cambria, Olsher and Kwok (2012)	Conscious and unconscious commonsense reasoning	LiveJournal PatientOpinion	Sentic activation	Brain-inspired computational mode
Cambria, Benson, Eckl, and Hussain (2012)	Extraction of semantic and Sentics	PatientOpinion	Sentic PROM	Semi-structured health assessment system
Cambria *et al.*(2014)	Common and commonsense knowledge base	Probase ConceptNet Isanette Open Mind corpus	Semantic multi-dimensional scaling	Open-domain opinion mining and sentiment analysis system
Cambria *et al.* (2013)	Concept-level text processing	Patient Opinion Social enterprise pioneering	Biologically inspired opinion mining engine	Opinion mining engine
Poria *et al.* (2014)	Framework for affective commonsense reasoning	International Survey of Emotion Antecedents and Reactions (ISEAR)	Fuzzy c-means clustering and support vector-machine classification	Framework outperformed the state of the art
Loia and Senatore (2014)	Framework for extracting emotions and sentiments	Twitter and Newyork times	Fuzzy logic	Prototype to detects emotions from text

Table VIII. Summary of existing Sentic computing techniques for sentiment analysis limitations: the assumption of only one relation per sentence is untrue in many cases; a manual compilation of rules is necessary; making an adaptation to new domains is difficult; directions of comparative relations are not identified, and therefore product strength cannot be measured for accurate decision making. In a subsequent study, Jindal and Liu (2006b) it is notable that only particular types of sentiment information can be extracted, whereas other types of opinion sources could be included to provide broader comparison passage among products. The research by Xu, Liao, Li and Song (2011) corrected these aforementioned shortcomings by providing a method of comparative relation extraction that not only detects the occurrence of relations, but also recognizes direction. However, the unfixed interdependencies among relations are overlooked; when addressed, better performance for relation extraction may be achieved. It is noticed that in the proposed work (Xu, Wang, Ren, Xu, Liu and Liao, 2011) more linguistic features could be explored, in order to improve the performance of their comparison relation recognition and categorization. In the proposed work Hou and Li (2008), the parse constituent structure of a sentence is quite crucial for the first dataset, and this demands further improvement. This technique Ganapathibhotla and Liu (2008) performs well for regular sentences, but efficiency decreases when increasing sentence complexity. In the proposed work (Li *et al.*, 2010) multiple and ambiguous entities are identified, so essential components in building a precise recommendation system for decision making are overlooked.

To the best of our knowledge, the comparative sentence mining concept originates from Jindal and Liu (2006a), and is then considered further in Ganapathibhotla and Liu (2008), Hou and Li (2008), Jindal and Liu (2006a), Li *et al.* (2010), Xu, Liao, Li and Song (2011), Xu, Wang, Ren, Xu, Liu and Liao (2011). We have noticed that the suggestive sentence identification is proposed for the first time by Qazi, Raj, Tahir, Cambria and Syed (2014), and further explored for preferred sentences (Qazi, Raj, Tahir, Waheed, Khan and Abraham, 2014). Hence, we feel that more research is needed to define the review types for scenarios other than commerce. An increase in empirical research studies is required to develop the area.

We have identified the task of machine learning techniques for SA in response to *RQ2*. The traditional fact-based analysis include material on summarization of evaluative text and on broader issues opinion recognition and classification, classification of documents, polarity identification, and predictive SA. Research organizations and academia have affected significant progress in the field of SA, and a variety of techniques have been summarized here. The authors Pang *et al.* (2002), Popescu *et al.* (2005), and Liu *et al.* (2005) have proposed methods for polarity classification. The comparison of these three classification techniques was proposed by Ye *et al.* (2009). Feature identification (Cruz *et al.*, 2010), polarity detection for Spanish corpus (Habernal *et al.*, 2014; Martínez-Cámara *et al.*, 2011), and forecast the stock price (Smailović *et al.*, 2014) is also observed under *RQ2*.

The technique of SA enhances business intelligence, thus attracting the attention of the business community. However, OM is a very challenging task even at its basic level of sentiment polarity classification, which is a case of binary classification. The extraction of opinion polarity from text can be performed by comparing words extracted with a set of keywords. The identification of a right set of keywords for mining opinions with a certain level of accuracy, however, is not an easy task. In addition, sentiment and subjectivity are quite context and domain dependent. As an example, the expression "go watch the movie" most likely indicates the positive sentiment for movie viewers, but negative sentiment for book readers. These kinds of tasks are not handled by traditional keyword-based machine learning techniques. To collect the opinions over the web and analyze at content/syntactic level to infer the semantic and affective information associated with opinions is a challenging task. To overcome the deficiency of keyword-based techniques, Sentic computing has recently been introduced by Cambria and Hussain (2012).

We have identified these challenges in response to *RQ3*. The works presented in recent studies (Cambria *et al.*, 2010; Cambria, Olsher and Kwok, 2012; Cambria, Song, Wang and Howard, 2014; Havasi *et al.*, 2010) based on Sentic computing developed the commonsense knowledge bases to bridge the cognitive and affective gap between word-level natural language data and concept-level opinions. When we look ahead, we can state that in future powerful intelligent systems will be able to deal with personal and social relationships and OM will have to address the priorities of each user.

The principal challenging aspects in opinion summarization are the use of different types of opinions, the complexity of sentences, emotion detection, sentiment orientation, and strength of classification, presentation of opinions and development of common and commonsense knowledge. Further challenges are posed by differences of expression between different languages, and the differences within a language with regard to the different jargon and idiomatic expressions that are used within the bounds of a certain product or service domain. From an analytical perspective, it has been observed that different authors use different datasets for the evaluation of their techniques, which may limit comparisons between works; nonetheless, most of the methods discussed here have used Amazon datasets. Also, to break the barriers of the current methods through the development of commonsense knowledge bases to link both the cognitive and affective expanse between word-level natural language data/information and the concept-level perspectives expressed, Sentic computing is required.

It is important to note that comparative opinions have not been discussed in detail by previous studies in this field. We have emphasized the logic and importance of summaries of comparative opinions, which explain individual perception in making choices about whether a product is good or bad in comparison with other products. The usefulness of all comparative, regular, and suggestive SA stems from the notion that potential buyers are averse to reading whole opinions, finding them wholly inconvenient; summaries provide a useful, time-saving facility for informed decision making. The products and services industries are becoming increasingly aware of the value of regular and comparative OM, which not only enables cross-correlation with more traditional marketing techniques, but may circumvent some of the issues that traditional market surveys create, such as the lack of candor (commonly arising out of politeness) that results from direct questioning (Cambria, Grassi, Hussain, and Havasi, 2012). Furthermore, it may help to identify the shortcomings of existing products, while facilitating the prediction of future market niches.

5. Conclusion

This paper presents a systematic review of literature on the practices and challenges of SA study. This review was conducted by following the available guidelines for conducting systematic literature reviews (Keele, 2007; Brereton *et al.*, 2007) to search and categorize all existing and available literature on OM. Out of the 280 initial papers located in well-known electronic research databases, 24 relevant papers were extracted through a multistage sifting process with independent validation in each step. These papers were then assessed for the quality of the evidence they produced and further analyzed. The findings in our research provide future dimensions to industry and research practitioners for further work on OM.

We have determined that there is a need for further research on OM with the growing field of concept learning. A holistic overview of the linguistic, learning, and statistical techniques used in the ever-expanding research field of SA has been broadly introduced here. In greater depth, we have focused on the evaluation of various techniques related to sentiment summarization of regular and comparative opinions. The opinion summarization of both regular and comparative opinions is of great significance to decision making and poses interesting opportunities for present and future development. While it is different from general text summarization or topic summarization, opinion summarization focuses on the more troublesome task of identifying sentiment perspective.

The review shows that various techniques have been developed for opinion summarization, drawing their methods from the fields of artificial intelligence, statistics and linguistics, where each technique possesses a certain focus as well as having its particular strengths and weaknesses. The range of sentiment analysis techniques is fast expanding in order to provide adequate solutions to the problems that persist in SA; although many unsolved issues are still present, a few have been discussed in the present review and this indeed represents a testament of the infancy of the field. Since it is very cumbersome and requires significant time for a human reader to find appropriate resources, extract opinion sentences, read, and then summarize them to get useful information, automated opinion detection and summarization systems are still required.

Sentiment analysis is very challenging due to limited understanding and knowledge of the problem area and its solutions. One of the obvious reasons is that it falls under the category of NLP and NLP analysis is considered to be very tough. Most of the OM research is carried out using machine learning algorithms which are not understandable by humans, yet they are beating baselines for improved accuracy. Furthermore, the overall use of SA techniques is also a promising aspect. We have addressed the implications of SA and opinion types, which can help motivate practitioners concerning the use of these techniques for making better analysis and successful policies for future decisions.

5.1 Implication of the study

This review has several implications for researchers, traders, and managers. In terms of research, the review shows that there is a need of more studies that incorporate the concept learning approaches with SA. The result of the study shows that Sentic computing techniques will continue to grow across the world. It has the ability to make computing systems process business information with human-like cognition. In this era of computing, researchers have developed new technologies that can help machines learn, reason, and efficiently process diverse data types while interacting with people in natural and familiar ways. Moreover, such approaches rely on syntactical structure of text, which is far from the way the human mind processes natural language. Therefore, a compelling need for Sentic computing is present, and commonsense computing techniques were further developed and applied to bridge the semantic gap between word-level natural language data and the concept-level opinions conveyed by these. In particular, Sentic computing performs a clause-level semantic analysis of text, which allows the inference of both the conceptual and emotional information associated with natural language opinions and, hence, a more efficient passage from (unstructured) textual information to (structured) machine-processable data.

Linking the cognitive and affective gap between word-level natural language data and the concept-level opinions by developing commonsense knowledge-based Sentic computing is playing an important role. For concept-level analysis of the text of natural language, it actually exploits affective ontologies and commonsense reasoning tools. Thus, in future, in order to better understand natural language opinions, OM has to develop better commonsense knowledge bases and better reasoning methods.

The result of this systematic review study shows that the advent of the multiple opinion types given encourage traders to make more effective new polices by keeping in view multiple types of user feedback. For business intelligence, more focused reviews (e.g. suggestive) are better able to increase the satisfaction levels, leading to successful sales. This sort of user input is valuable to designers as well as users, and it is becoming gradually more available with the rise of e-commerce and new social media including blogs and social nets. At present, there are only a few studies available on multiple opinion types such as suggestive opinions (Qazi, Raj, Tahir, Cambria and Syed, 2014). Hence, it is a potential domain for more empirical investigation of opinion types using SA techniques. Moreover, present research approaches and techniques developed for OM should be implemented and practically used for different domains of opinions other than business, e.g. health, education, and fashion. The techniques should be developed for other research fields to fulfill today's need by using these advanced approaches.

Another practical implication from the findings of this research suggests that if firms' managers desire to grab the attention of consumers' toward a certain item, they need to place the most helpful reviews on the top of reviews. They need to use the latest SA techniques to grab the helpful information. There are some limitations of the study; however, despite these important aspects review types and SA techniques, one has to be aware of other features related with opinions such as ratings, likes, and helpfulness. Our study has provided the classified types of reviews and the techniques with their pros and cons that are used to find the SA. Future research should incorporate these findings such as review types and find the effect of other aspects such as helpfulness and ratings on these types of reviews. Further investigation in the field will help to get a better understanding of the reviews related features on product purchase and sale.

References

Al-Debei, M.M., Akroush, M.N. and Ashouri, M.I. (2015), "Consumer attitudes towards online shopping: the effects of trust, perceived benefits, and perceived web quality", *Internet Research*, Vol. 25 No. 5, pp. 707-733.

Brereton, P., Kitchenham, B.A., Budgen, D., Turner, M. and Khalil, M. (2007), "Lessons from applying the systematic literature review process within the software engineering domain", *Journal of Systems and Software*, Vol. 80 No. 4, pp. 571-583.

Cambria, E. (2016), "Affective computing and sentiment analysis", *IEEE Intelligent Systems*, Vol. 31 No. 2, pp. 102-107.

Cambria, E. and Hussain, A. (2012), "Sentic album: content-, concept-, and context-based online personal photo management system", *Cognitive Computation*, Vol. 4 No. 4, pp. 477-496.

Cambria, E. and Hussain, A. (2015), *Sentic Computing: A Common-Sense-Based Framework for Concept-Level Sentiment Analysis*, Vol. 1, Springer, Cham.

Cambria, E. and White, B. (2014), "Jumping NLP curves: a review of natural language processing research", *IEEE Computational Intelligence Magazine*, Vol. 9 No. 2, pp. 48-57.

Cambria, E., Mazzocco, T. and Hussain, A. (2013), "Application of multi-dimensional scaling and artificial neural networks for biologically inspired opinion mining", *Biologically Inspired Cognitive Architectures*, Vol. 4 No. 4, pp. 41-53.

Cambria, E., Olsher, D. and Kwok, K. (2012), "Sentic activation: a two-level affective common sense reasoning framework", paper presented at the AAAI, Toronto, July 12.

Cambria, E., Wang, H. and White, B. (2014), "Guest editorial: big social data analysis", *Knowledge-Based Systems*, Vol. 69 No. 1, pp. 1-2.

Cambria, E., Benson, T., Eckl, C. and Hussain, A. (2012), "Sentic PROMs: application of sentic computing to the development of a novel unified framework for measuring health-care quality", *Expert Systems with Applications*, Vol. 39 No. 12, pp. 10533-10543.

Cambria, E., Fu, J., Bisio, F. and Poria, S. (2015), "AffectiveSpace 2: enabling affective intuition for concept-level sentiment analysis", AAAI, pp. 508-514.

Cambria, E., Gastaldo, P., Bisio, F. and Zunino, R. (2015), "An ELM-based model for affective analogical reasoning", *Neurocomputing*, Vol. 149 No. A, pp. 443-455.

Cambria, E., Grassi, M., Hussain, A. and Havasi, C. (2012), "Sentic computing for social media marketing", *Multimedia Tools and Applications*, Vol. 59 No. 2, pp. 557-577.

Cambria, E., Hussain, A., Havasi, C. and Eckl, C. (2009), "Common sense computing: from the society of mind to digital intuition and beyond", *Biometric ID Management and Multimodal Communication*, Springer, pp. 252-259.

Cambria, E., Song, Y., Wang, H. and Howard, N. (2014), "Semantic multidimensional scaling for open-domain sentiment analysis", *IEEE Intelligent Systems*, Vol. 29 No. 2, pp. 44-51.

Cambria, E., Hussain, A., Durrani, T., Havasi, C., Eckl, C. and Munro, J. (2010), "Sentic computing for patient centered applications", paper presented at the IEEE 10th International Conference on Signal Processing Proceedings, Beijing, October 24.

Chaturvedi, I., Cambria, E., Poria, S. and Bajpai, R. (2016), "Bayesian deep convolutional belief networks for subjectivity detection", ICDM, Barcelona.

Cohen, J. (1968), "Weighted kappa: nominal scale agreement provision for scaled disagreement or partial credit", *Psychological Bulletin*, Vol. 70 No. 4, pp. 213-220.

Cruz, F.L., Troyano, J.A., Enríquez, F., Ortega, F.J. and Vallejo, C.G. (2010), "A knowledge-rich approach to feature-based opinion extraction from product reviews", paper presented at the Proceedings of the 2nd International Workshop on Search and Mining User-Generated Contents, Toronto, October 30.

Dybå, T. and Dingsøyr, T. (2008), "Empirical studies of agile software development: a systematic review", *Information and Software Technology*, Vol. 50 No. 9, pp. 833-859.

Elo, S. and Kyngäs, H. (2008), "The qualitative content analysis process", *Journal of Advanced Nursing*, Vol. 62 No. 1, pp. 107-115.

Etzioni, O., Banko, M., Soderland, S. and Weld, D.S. (2008), "Open information extraction from the web", *Communications of the ACM*, Vol. 51 No. 12, pp. 68-74.

Faliagka, E., Tsakalidis, A. and Tzimas, G. (2012), "An integrated e-recruitment system for automated personality mining and applicant ranking", *Internet Research*, Vol. 22 No. 5, pp. 551-568.

Fleiss, J., Levin, B. and Paik, M. (2004), "The measurement of interrater agreement", *Statistical Methods for Rates and Proportions*, 3rd ed., John Wiley & Sons, Hoboken, NJ , pp. 598-626.

Ganapathibhotla, M. and Liu, B. (2008), "Mining opinions in comparative sentences", paper presented at the Proceedings of the 22nd International Conference on Computational Linguistics-Vol. 1, Manchester, August 18.

Habernal, I., Ptáček, T. and Steinberger, J. (2014), "Supervised sentiment analysis in Czech social media", *Information Processing & Management*, Vol. 50 No. 5, pp. 693-707.

Havasi, C., Speer, R. and Holmgren, J. (2010), "Automated color selection using semantic knowledge", *Proceedings of AAAI CSK*, *Arlington, TX*, *November 11-13*.

Hou, F. and Li, G.H. (2008), "Mining Chinese comparative sentences by semantic role labeling", paper presented at the 2008 International Conference on Machine Learning and Cybernetics, Kunming, July 12.

Hsieh, H.-F. and Shannon, S.E. (2005), "Three approaches to qualitative content analysis", *Qualitative Health Research*, Vol. 15 No. 9, pp. 1277-1288.

Hu, M. and Liu, B. (2004), "Mining opinion features in customer reviews", paper presented at the Proceedings of the National Conference on Artificial Intelligence, San Jose, CA, July 25.

Inayat, I., Salim, S.S., Marczak, S., Daneva, M. and Shamshirband, S. (2015), "A systematic literature review on agile requirements engineering practices and challenges", *Computers in Human Behavior*, Vol. 51 No. PB, pp. 915-929.

Jindal, N. and Liu, B. (2006a), "Identifying comparative sentences in text documents", paper presented at the Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Seattle, WA, August 6.

Jindal, N. and Liu, B. (2006b), "Mining comparative sentences and relations", paper presented at the Proceedings of the National Conference on Artificial Intelligence, Boston, MA, July 16.

Jindal, N. and Liu, B. (2006c), "Mining comparative sentences and relations", paper presented at the AAAI.

Keele, S. (2007), "Guidelines for performing systematic literature reviews in software engineering", technical report, Ver. 2.3 EBSE Technical Report, EBSE.

Kitchenham, B., Pretorius, R., Budgen, D., Brereton, O.P., Turner, M., Niazi, M. and Linkman, S. (2010), "Systematic literature reviews in software engineering – a tertiary study", *Information and Software Technology*, Vol. 52 No. 8, pp. 792-805.

Landis, J.R. and Koch, G.G. (1977), "The measurement of observer agreement for categorical data", *Biometrics*, Vol. 33 No. 1, pp. 159-174.

Lee, J., Park, D.-H. and Han, I. (2011), "The different effects of online consumer reviews on consumers' purchase intentions depending on trust in online shopping malls: an advertising perspective", *Internet Research*, Vol. 21 No. 2, pp. 187-206.

Lee, M., Kim, M. and Peng, W. (2013), "Consumer reviews: reviewer avatar facial expression and review valence", *Internet Research*, Vol. 23 No. 2, pp. 116-132.

Li, S., Lin, C.Y., Song, Y.I. and Li, Z. (2010), "Comparable entity mining from comparative questions", paper presented at the Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, Uppsala, July 11.

Li, S., Lin, C.-Y., Song, Y.-I. and Li, Z. (2013), "Comparable entity mining from comparative questions", *Knowledge and Data Engineering, IEEE Transactions on*, Vol. 25 No. 7, pp. 1498-1509.

Liu, B. (2010), "Sentiment analysis and subjectivity", in Indurkhya, N. and Damerau, F.J. (Eds), *Handbook of Natural Language Processing*, Print ISBN: 978-1-4200-8592-1, eBook ISBN: 978-1-4200-8593-8, 2nd ed., Vol. 2, CRC, Chapman and Hall, pp. 627-666.

Liu, B., Hu, M. and Cheng, J. (2005), "Opinion observer: analyzing and comparing opinions on the web", paper presented at the Proceedings of the 14th International Conference on World Wide Web, Chiba, May 10.

Loia, V. and Senatore, S. (2014), "A fuzzy-oriented sentic analysis to capture the human emotion in web-based content", *Knowledge-Based Systems*, Vol. 58, March, pp. 75-85.

Ma, Y., Cambria, E. and Gao, S. (2016), "Label embedding for zero-shot fine-grained named entity typing", COLING, Osaka.

Martínez-Cámara, E., Martín-Valdivia, M.T. and Ureña-López, L.A. (2011), "Opinion classification techniques applied to a Spanish corpus", *Natural Language Processing and Information Systems*, Springer, pp. 169-176.

Pacheco, C. and Garcia, I. (2012), "A systematic literature review of stakeholder identification methods in requirements elicitation", *Journal of Systems and Software*, Vol. 85 No. 9, pp. 2171-2181.

Pang, B., Lee, L. and Vaithyanathan, S. (2002), "Thumbs up? Sentiment classification using machine learning techniques", paper presented at the Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing-Vol. 10, Philadelphia, PA, July 6.

Popescu, A.-M., Nguyen, B. and Etzioni, O. (2005), "OPINE: extracting product features and opinions from reviews", paper presented at the Proceedings of HLT/EMNLP on Interactive Demonstrations, Vancouver, October 7.

Poria, S., Cambria, E. and Gelbukh, A. (2016), "Aspect extraction for opinion mining with a deep convolutional neural network", *Knowledge-Based Systems*, Vol. 108 No. C, pp. 42-49.

Poria, S., Cambria, E., Hazarika, D. and Vij, P. (2016), "A deeper look into sarcastic tweets using deep convolutional neural networks", COLING, Osaka.

Poria, S., Cambria, E., Hussain, A. and Huang, G.-B. (2015), "Towards an intelligent framework for multimodal affective data analysis", *Neural Networks*, Vol. 63 No. C, pp. 104-116.

Poria, S., Chaturvedi, I., Cambria, E. and Hussain, A. (2016), "Convolutional MKL based multimodal emotion recognition and sentiment analysis", ICDM, Barcelonoa.

Poria, S., Cambria, E., Howard, N., Huang, G.-B. and Hussain, A. (2016), "Fusing audio, visual and textual clues for sentiment analysis from multimodal content", *Neurocomputing*, Vol. 174 No. A, pp. 50-59.

Poria, S., Gelbukh, A., Cambria, E., Hussain, A. and Huang, G.-B. (2014), "EmoSenticSpace: a novel framework for affective common-sense reasoning", *Knowledge-Based Systems*, Vol. 69, pp. 108-123.

Qazi, A., Raj, R.G., Tahir, M. and Naqvi, S.G.A. (2013), "A framework of review analysis for enhancement of business decision making", paper presented at the 13th International Conference on Data Mining Workshops, Dallas, TX, December 7.

Qazi, A., Raj, R.G., Tahir, M., Cambria, E. and Syed, K.B.S. (2014), "Enhancing business intelligence by means of suggestive reviews", *The Scientific World Journal*, Vol. 2014, p. 11, Article ID 879323.

Qazi, A., Fayaz, H., Wadi, A., Raj, R.G., Rahim, N. and Khan, W.A. (2015), "The artificial neural network for solar radiation prediction and designing solar systems: a systematic literature review", *Journal of Cleaner Production*, Vol. 104, October, pp. 1-12.

Qazi, A., Raj, R.G., Tahir, M., Waheed, M., Khan, S.U.R. and Abraham, A. (2014), "A preliminary investigation of user perception and behavioral intention for different review types: customers and designers perspective", *The Scientific World Journal*, Vol. 2014, p. 8, Article ID 872929.

Qazi, A., Syed, K.B.S., Raj, R.G., Cambria, E., Tahir, M. and Alghazzawi, D. (2016), "A concept-level approach to the analysis of online review helpfulness", *Computers in Human Behavior*, Vol. 58 No. 7, pp. 75-81.

Qazi, A., Tamjidyamcholo, A., Raj, R.G., Hardaker, G. and Standing, C. (2017), "Assessing consumers' satisfaction and expectations through opinions: expectation and disconfirmation approach", *Computers in Human Behavior*.

Quigley, M. (2008), *Encyclopedia of Information Ethics and Security*, IGI Global, New York, NY.

Seuring, S. and Müller, M. (2008), "From a literature review to a conceptual framework for sustainable supply chain management", *Journal of Cleaner Production*, Vol. 16 No. 15, pp. 1699-1710.

Smailović, J., Grčar, M., Lavrač, N. and Žnidaršič, M. (2014), "Stream-based active learning for sentiment analysis in the financial domain", *Information Sciences*, Vol. 285 No. C, pp. 181-203.

Tamimi, N. and Sebastianelli, R. (2015), "The relative importance of e-tailer website attributes on the likelihood of online purchase", *Internet Research*, Vol. 25 No. 2, pp. 169-183.

Thet, T.T., Na, J.C. and Khoo, C.S.G. (2010), "Aspect-based sentiment analysis of movie reviews on discussion boards", *Journal of Information Science*, Vol. 36 No. 6, pp. 823-848.

Waheed, M., Kaur, K. and Qazi, A. (2016), "Students' perspective on knowledge quality in eLearning context: a qualitative assessment", *Internet Research*, Vol. 26 No. 1, pp. 120-145.

White, A. and Schmidt, K. (2005), "Systematic literature reviews", *Complementary Therapies in Medicine*, Vol. 13 No. 1, pp. 54-60.

Wu, W., Li, H., Wang, H. and Zhu, K.Q. (2012), "Probase: a probabilistic taxonomy for text understanding", paper presented at the Proceedings of the 2012 International Conference on Management of Data, Scottsdale, May 20.

Xu, K., Liao, S.S., Li, J. and Song, Y. (2011), "Mining comparative opinions from customer reviews for competitive intelligence", *Decision Support Systems*, Vol. 50 No. 4, pp. 743-754.

Xu, K.S.J., Wang, W., Ren, J., Xu, J.S.Y., Liu, L. and Liao, S. (2011), "Classifying consumer comparison opinions to uncover product strengths and weaknesses", *International Journal of Intelligent Information Technologies*, Vol. 7 No. 1, pp. 1-14.

Ye, Q., Zhang, Z. and Law, R. (2009), "Sentiment classification of online reviews to travel destinations by supervised machine learning approaches", *Expert Systems with Applications*, Vol. 36 No. 3, pp. 6527-6535.

Yu, H. and Hatzivassiloglou, V. (2003), "Towards answering opinion questions: separating facts from opinions and identifying the polarity of opinion sentences", paper presented at the Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing, Sapporo, July 11.

Appendix. List of reviewed studies

Cambria, E., Mazzocco, T. and Hussain, A. (2013), "Application of multi-dimensional scaling and artificial neural networks for biologically inspired opinion mining", *Biologically Inspired Cognitive Architectures*, Vol. 4 No. 2013, pp. 41-53.

Cambria, E., Olsher, D. and Kwok, K. (2012), "Sentic activation: a two-level affective commonsense reasoning framework", AAAI, Toronto, pp. 186-192.

Cambria, E., Benson, T., Eckl, C. and Hussain, A. (2012), "Sentic PROMs: application of Sentic computing to the development of a novel unified framework for measuring health-care quality", *Expert Systems with Applications*, Vol. 39 No. 2012, pp. 10533-10543.

Cambria, E., Song, Y., Wang, H. and Howard, N. (2013), "Semantic multi-dimensional scaling for open-domain sentiment analysis", *IEEE Intelligent Systems*, Vol. 10 No. 2012, pp. 1-8.

Cambria, E., Hussain, A., Durrani, T., Havasi, C., Eckl, C. and Munro, J. (2010), "Sentic computing for patient centered applications", *2010 10th International Conference on Signal Processing (ICSP)*, IEEE, pp. 1279-1282.

Cruz, F.L., Troyano, J.A., Enríquez, F., Ortega, F.J. and Vallejo, C.G. (2010), "A knowledge- rich approach to feature-based opinion extraction from product reviews", *Proceedings of the 2nd International Workshop on Search and Mining User-Generated Contents: ACM*, pp. 13-20.

Ganapathibhotla, M. and Liu, B. (2008), "Mining opinions in comparative sentences", *Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1: Association for Computational Linguistics*, pp. 241-248.

Habernal, I., Ptáček, T. and Steinberger, J. (2014), "Supervised sentiment analysis in Czech social media", *Information Processing & Management*, Vol. 50 No. 5, pp. 693-707.

Hou, F. and Li, G.-H. (2008), "Mining Chinese comparative sentences by semantic role labelling", *2008 International Conference on Machine Learning and Cybernetics*, Vol. 5, IEEE, pp. 2563-2568.

Jindal, N. and Liu, B. (2006a), "Identifying comparative sentences in text documents", *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, ACM, pp. 244-251.

Jindal, N. and Liu, B. (2006b), "Mining comparative sentences and relations", *Proceedings of the National Conference on Artificial Intelligence*, Vol. 21, AAAI Press, MIT Press, Menlo Park, CA, Cambridge, MA and London, p. 1331.

Li, S., Lin, C.-Y., Song, Y.-I. and Li, Z. (2013), "Comparable entity mining from comparative questions", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 25 No. 7, pp. 1498-1509.

Liu, B., Hu, M. and Cheng, J. (2005), "Opinion observer: analyzing and comparing opinions on the Web", *Proceedings of the 14th International Conference on World Wide Web*, ACM, pp. 342-351.

Loia, V. and Senatore, S. (2014), "A fuzzy-oriented Sentic analysis to capture the human emotion in Web-based content", *Knowledge-Based Systems*, Vol. 58, pp. 75-85.

Martínez-Cámara, E., Martín-Valdivia, M.T. and Ureña-López, L.A. (2011). "Opinion classification techniques applied to a Spanish corpus", *Natural Language Processing and Information Systems*: Springer, pp. 169-176.

Pang, B., Lee, L. and Vaithyanathan, S. (2002), "Thumbs up?: sentiment classification using machine learning techniques", *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing-Volume 10: Association for Computational Linguistics*, pp. 79-86.

Popescu, A.-M., Nguyen, B. and Etzioni, O. (2005), "OPINE: extracting product features and opinions from reviews", *Proceedings of HLT/EMNLP on Interactive Demonstrations: Association for Computational Linguistics*, pp. 32-33.

Poria, S., Gelbukh, A., Cambria, E., Hussain, A. and Huang, G.-B. (2014), "EmoSenticSpace: a novel framework for affective common-sense reasoning", *Knowledge-Based Systems*, Vol. 69, pp. 108-123.

Qazi, A., Raj, R.G., Tahir, M., Cambria, E. and Syed, K.B.S. (2014), "Enhancing business intelligence by means of suggestive reviews", *The Scientific World Journal*.

Qazi, A., Raj, R.G., Tahir, M., Waheed, M., Khan, S.U.R. and Abraham, A. (2014), "A preliminary investigation of user perception and behavioral intention for different review types: customers and designers perspective", *The Scientific World Journal*, p. 8.

Smailović, J., Grčar, M., Lavrač, N., and Žnidaršič, M. (2014), "Stream-based active learning for sentiment analysis in the financial domain", *Information Sciences*, Vol. 285, pp. 181-203.

Xu, K., Liao, S.S., Li, J. and Song, Y. (2011), "Mining comparative opinions from customer reviews for competitive intelligence", *Decision Support Systems*, Vol. 50 No. 4, pp. 743-754.

Xu, K.S.J., Wang, W., Ren, J., Xu, J.S.Y., Liu, L. and Liao, S. (2011), "Classifying consumer comparison opinions to uncover product strengths and weaknesses", *International Journal of Intelligent Information Technologies*, Vol. 7 No. 1, pp. 1-14.

Ye, Q., Zhang, Z. and Law, R. (2009), "Sentiment classification of online reviews to travel destinations by supervised machine learning approaches", *Expert Systems with Applications*, Vol. 36 Nos 3, pp. 6527-6535.

About the authors

Dr Atika Qazi serves as a Research Assistant in University Malay, Kuala Lumpur. She has worked as an Assistant Professor at the Department of Computer Science, COMSATS Institute of Information Technology, Pakistan. She has also worked as a Visiting Research Fellow in University Brunei Darussalam, Brunei. She received her PhD Degree in Computer Science from University of Malaya. Her areas of interests are opinion mining, business intelligence, information system, e-learning, clean energy, and innovative learning. She is actively involved as a Reviewer of ISI Journals such as *Journal of Cleaner Production, Internet Research Journal and Computers and Education Journal*. Dr Atika Qazi is the corresponding author and can be contacted at: atikaqazium@gmail.com

Dr Ram Gopal Raj presently works as a Senior Lecturer at the Department of Artificial intelligence, University of Malaya. He received his PhD Degree in Computer Science in 2010 from the University of Malaya, Malaysia. He has more than ten years of experience as an academician and also successfully supervised PhD and Masters students. He presently supervises five Master and six PhD students and has examined Master and PhD theses. His research area includes text classification, natural language processing, logic programming, and intelligent systems. He has published more than 30 ISI-indexed research articles.

Professor Glenn Hardaker is the Director of Centre for Lifelong Learning and a Professor of Education at the Sultan Hassanal Bolkiah Institute of Education, Universiti of Brunei Darussalam. He is also a National Teaching Fellow (Higher Education Academy, the UK). He has worked in universities in the UK for approximately 25 years in the area of technological innovation and learning. His research is focused on a blend of Islam, innovation, and learning. Glenn Hardaker is an Editor of two international publications: *Journal for Multicultural Education* and *International Journal of Information and Learning Technology*.

Professor Craig Standing is a Foundation Professor of Strategic Information Management in the School of Business and Law at the Edith Cowan University. He is a Co-leader of the Centre for Innovative Practice. Craig studied at the Lancaster University, University of Manchester, and the University of Western Australia. He has published in the top ten Information Systems journals and has presented at many international conferences. He has obtained significant research funds including the prestigious Australian Research Council Awards. Craig has authored several books on information systems that have been adopted by university courses in information systems. He has supervised numerous doctoral students, having won awards from the postgraduate association at his university for the quality of his supervision.

This article has been cited by:

1. Atika Qazi, Alireza Tamjidyamcholo, Ram Gopal Raj, Glenn Hardaker, Craig Standing. 2017. Assessing consumers' satisfaction and expectations through online opinions: Expectation and disconfirmation approach. *Computers in Human Behavior* **75**, 450-460. [Crossref]