Sentiment Analysis Main Tasks and Applications: A Survey
Sara Tedmori* and Arafat Awajan*

Abstract
The blooming of social media has simulated interest in sentiment analysis. Sentiment analysis aims to determine from a specific piece of content the overall attitude of its author in relation to a specific item, product, brand, or service. In sentiment analysis, the focus is on the subjective sentences. Hence, in order to discover and extract the subjective information from a given text, researchers have applied various methods in computational linguistics, natural language processing, and text analysis. The aim of this paper is to provide an in-depth up-to-date study of the sentiment analysis algorithms in order to familiarize with other works done in the subject. The paper focuses on the main tasks and applications of sentiment analysis. State-of-the-art algorithms, methodologies and techniques have been categorized and summarized to facilitate future research in this field.

Keywords
Feature Selection, Opinion Mining, Sentiment Analysis, Sentiment Analysis Applications, Sentiment Classification, Sentiment Visualization, Social Media Monitoring

1. Introduction

Sentiment analysis (SA) also known as opinion mining is a sub-division of data mining. SA refers to the practice of applying text analysis and natural language processing (NLP) for the purpose of identifying, extracting, and analyzing subjective information from textual sources. SA focuses on the task of classifying a given input text by the polarity of its sentiment as being positive, negative or neutral. More advanced SA techniques look at whether the textual sources have associations with emotional states such as fear, anger, happiness, and sadness. Alternatively, instead of classifying text as being either positive, negative, or neutral, the text could be associated with a number on a pre-defined scale (e.g., -5 to +5).

Subjectivity/objectivity identification is another important research topic in SA [1,2], which involves classifying a given textual source as being either objective or subjective. When compared with the task of polarity classification, the task of subjectivity/objectivity identification is more challenging [3]. The reason behind this challenge is that the subjectivity of words and phrases may well be dependent on the context and an objective document may possibly enclose subjective sentences and vice-versa. Furthermore, results are largely reliant on the definition of subjectivity chosen when annotating texts [4].

\[1\] This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
Manuscript received March 14, 2016; first revision August 26, 2016; second revision November 23, 2016; accepted December 6, 2016.
Corresponding Author: Sara Tedmori (S.Tedmori@psut.edu.jo)
* Dept. of Computer Science, King Hussein Faculty of Computing Sciences, Princess Sumaya University for Technology, Amman, Jordan ((S.Tedmori, awajan}@psut.edu.jo)
Research has shown that eliminating objective sentences from a document prior to polarity classification has a positive effect on the performance of the sentiment classification task [3].

A more challenging, but preferable SA task, is feature/aspect-based SA, which refers to the task of classifying the sentiments expressed on the different features or aspects of the entities. Feature/aspect-based SA involves several sub-tasks, including entity identification, entity features/aspects extraction, feature/aspect opinion identification, and feature/aspect opinion classification.

Every year, many researchers publish articles in the area of SA. Hence, there have been many attempts by researchers to summarize the recent research developments and directions of SA. A survey paper by Medhat et al. [5] focuses on the different SA techniques and provides a comprehensive overview of sentiment analysis. Fifty-four articles published between 2010 and 2013 covering a wide variety of SA fields are summarized and categorized according to the techniques used. Their survey paper also includes a discussion of related fields.


This paper aims is to provide a comprehensive overview of SA. The details reported in this paper have noteworthy differences to those reported in the studies referenced above and which focus on only one or two of the main SA tasks. This paper links to new references and includes a detailed discussion of four main SA tasks. The paper also provides an overview of the diverse array of SA applications.

The paper is organized as follows. Section 2 provides an overview of SA and its categories. Section 3 provides details related to the main subtasks of SA. Section 4 provides a comprehensive overview of its common applications. Section 5 offers some discussion and challenges for future work. Finally, section 6 ends the paper.

2. Sentiment Analysis

SA refers to the process of detecting and extracting opinions, feelings, attitudes, views, and evaluations from a specific input data. There have been many efforts to resolve the general SA problem; some of which cast it as a text classification task in which the focus is to annotate the polarity of input text as positive or negative depending on the feelings the text conveys and that can be expressed in any form or language. Standard approaches to SA have been classified in many different ways. The most fundamental classification is into the three broad categories of word spotting, lexical affinity, and statistical methods [10]. In word spotting, text is classified into affect categories (e.g., happy, sad, angry, bored) based on the existence of unambiguous affect words. Lexical affinity, unlike word spotting, does not detect clear affect words, but assigns words a probability of association with a particular emotion. Statistical methods
employ machine learning techniques such as latent semantic analysis (LSA), support vector machines (SVMs), and Bayesian inference, with a large set of affectively annotated text, known as a corpus, as training data. Statistical methods, similar to word spotting, are used to learn the emotional value of affect words, taking into consideration other features such as the emotional value of other arbitrary keywords, punctuation, and word co-occurrence [10]. Cambria et al. [10] point out that although keyword spotting is a very popular and simple SA approach; however, there are two weaknesses associated with it. First, in the case of negation, the recognition of affect is poor. Second, the method is dependent on the presence of obvious affect words. Lexical affinity, on the other hand, outperforms keyword spotting, but similarly has its own weaknesses. Lexical affinity is word level based and hence will not operate properly in the presence of negation. In addition, the performance of lexical affinity is dependent on the employed linguistic corpora, making the development of a reusable, domain-independent model difficult.

More recent approaches to SA, lead to the introduction of a fourth category: namely, the concept level approach [14] which relies on ontologies and semantic networks that take into account the implicit affective information expressed when utilizing natural language. Concept-level based approaches are capable of detecting subtly expressed sentiments. Hybrid approaches to sentiment analyses that combine two or more of the afore-mentioned approaches have also been proposed [15].

In 2010, Qiu et al. [16] tackled the problem of contextual advertising by proposing a strategy named DASA (dissatisfaction-oriented advertising based on sentiment analysis) that helps optimize advertisement relevance and user experience. The authors use a rule-based approach along with syntactic parsing and a sentiment dictionary to extract the topic words from negative subjective sentences.

Maks and Vossen [17], Cao et al. [18], Zhou et al. [19], Zhang et al. [20], and Pai et al. [2] exploit the semantic characteristics of the text to help perform the SA task at hand. Maks and Vossen [17] employ a syntactic semantic approach to build a SA lexicon model whose aim is to portray the subjectivity associations found between the actors in a sentence conveying the distinct attitudes of each actor. Cao et al. [18] explore the semantic characteristics of reviews and show that they are more significant than other characteristics in affecting the volume of helpfulness votes reviews garner. Similarly, Pai et al. [2] put forward a method for analyzing the content of product/service reviews for the purpose of assisting consumers in their decision making process. Zhou et al. [19] present an unsupervised technique that utilizes the semantic sequential representations to recognize the rhetorical structure theory based discourse relations for the purpose of diminishing polarity ambiguities inside the sentence. Zhang et al. [20] present a system that uncovers products’ weakness from Chinese reviews. The authors use statistical methods to identify implicit feature words and semantic methods to group together explicit feature words.

Min and Park [21] apply NLP techniques and use linguistic hints in another effort to rank the helpfulness of product reviews and introduce a new measure referred to as ‘mentions about experiences’ that discovers the expressions of time associated with the use of products and product entities over various purchasing periods. Hu et al. [22] employ a statistical method to detect the manipulation of online reviews, assess how consumers interact with the products that these manipulated reviews target, and examine whether these reviews have been manipulated through ratings and/or sentiments. Moreo et al. [23] propose a system that uses a news analysis taxonomy model to assess user opinions on topics expressed in news items. NLP techniques were used to build the taxonomy model. Other researchers employ supervised learning models in order to conduct the SA task at hand [24-26].

In addition to the SA works discussed above, Table 1 summarizes other SA works; it is organized in seven columns. The third column specifies whether the approaches detailed in the articles are domain
### Table 1. Summary of related SA works

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Year</th>
<th>DO</th>
<th>Algorithms used</th>
<th>Polarity</th>
<th>Data scope</th>
<th>Data set/sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>[35]</td>
<td>2008</td>
<td>N</td>
<td>SVM, Naïve Bayes (NB), Alternating decision tree (ADTree)</td>
<td>GA</td>
<td>Movie reviews</td>
<td>IMDB</td>
</tr>
<tr>
<td>[36]</td>
<td>2008</td>
<td>Y</td>
<td>SVM, Similarity measure (SIM), Maximum entropy (ME), Support vector regression (SVR)</td>
<td>GA</td>
<td>Game software review documents</td>
<td>N/A</td>
</tr>
<tr>
<td>[37]</td>
<td>2009</td>
<td>N</td>
<td>SVM</td>
<td>GA</td>
<td>Online forums</td>
<td>N/A</td>
</tr>
<tr>
<td>[38]</td>
<td>2009</td>
<td>N</td>
<td>NB, Dictionary</td>
<td>BC</td>
<td>Blogs, movie reviews</td>
<td>IMDB</td>
</tr>
<tr>
<td>[40]</td>
<td>2010</td>
<td>N</td>
<td>Multinomial Naïve Bayes (MNB), Stochastic Gradient Descent (SGD), Hoe ding Tree (HT)</td>
<td>BC</td>
<td>Micro-blogs</td>
<td>Twitter</td>
</tr>
<tr>
<td>[41]</td>
<td>2010</td>
<td>N</td>
<td>MNB</td>
<td>GA</td>
<td>Micro-blogs</td>
<td>Twitter</td>
</tr>
<tr>
<td>[16]</td>
<td>2010</td>
<td>Y</td>
<td>Rule-based</td>
<td>GA</td>
<td>Web forums</td>
<td>automotivieforums.com</td>
</tr>
<tr>
<td>[18]</td>
<td>2011</td>
<td>N</td>
<td>Semantic</td>
<td>GA</td>
<td>Software programs users’ feedback</td>
<td>N/A</td>
</tr>
<tr>
<td>[42]</td>
<td>2011</td>
<td>N</td>
<td>CRF</td>
<td>GA</td>
<td>Mobile customer reviews</td>
<td>amazon.com, epinions.com, blogs, and emails</td>
</tr>
<tr>
<td>[24]</td>
<td>2011</td>
<td>N</td>
<td>Multi-class SVM</td>
<td>GA</td>
<td>Digital cameras, MP3 reviews</td>
<td>N/A</td>
</tr>
<tr>
<td>[22]</td>
<td>2012</td>
<td>N</td>
<td>Statistical</td>
<td>GA</td>
<td>Book reviews</td>
<td>amazon.com</td>
</tr>
<tr>
<td>[23]</td>
<td>2012</td>
<td>N</td>
<td>Lexicon-based NLP</td>
<td>GA</td>
<td>News</td>
<td>N/A</td>
</tr>
<tr>
<td>[20]</td>
<td>2012</td>
<td>N</td>
<td>Pointwise mutual information (PMI), semantic</td>
<td>GA</td>
<td>Product reviews</td>
<td>N/A</td>
</tr>
<tr>
<td>[21]</td>
<td>2012</td>
<td>N</td>
<td>NLP</td>
<td>GA</td>
<td>Product reviews</td>
<td>amazon.com</td>
</tr>
<tr>
<td>[43]</td>
<td>2012</td>
<td>N</td>
<td>Sentiment Hyperspace Analogue to Language (S-HAL), Semantic Orientation inference from PMI (SO-PMI)</td>
<td>GA</td>
<td>News pages</td>
<td>Sogou CS corpus</td>
</tr>
<tr>
<td>[26]</td>
<td>2012</td>
<td>N</td>
<td>SVM, K-nearest neighbor, Radial Basis Function (RBF), Bayesian Network (BN), Decision Trees (DT), a propositional rule learner (JRip)</td>
<td>BC</td>
<td>Media</td>
<td>Media-analysis company</td>
</tr>
<tr>
<td>[44]</td>
<td>2013</td>
<td>Y</td>
<td>Context-based method NLP</td>
<td>GA</td>
<td>Restaurant reviews</td>
<td>N/A</td>
</tr>
<tr>
<td>[45]</td>
<td>2013</td>
<td>Y</td>
<td>Formal Concept Analysis (FCA)</td>
<td>GA</td>
<td>Smartphones, Tweets</td>
<td>Twitter</td>
</tr>
<tr>
<td>[2]</td>
<td>2013</td>
<td>N</td>
<td>Semantic</td>
<td>BC</td>
<td>Fast food reviews</td>
<td>N/A</td>
</tr>
<tr>
<td>[46]</td>
<td>2013</td>
<td>Y</td>
<td>n-gram, Dynamic ANN</td>
<td>GA</td>
<td>Tweets</td>
<td>Twitter</td>
</tr>
<tr>
<td>[47]</td>
<td>2013</td>
<td>Y</td>
<td>ML, NLP Tools</td>
<td>BC</td>
<td>Advertising strategies</td>
<td>Twitter</td>
</tr>
<tr>
<td>[48]</td>
<td>2014</td>
<td>N</td>
<td>Supervised term weighting scheme</td>
<td>GA</td>
<td>Product reviews</td>
<td>Cornell and Stanford movie review, Amazon product review</td>
</tr>
<tr>
<td>[49]</td>
<td>2014</td>
<td>Y</td>
<td>Lexicon-based, ML</td>
<td>BC</td>
<td>SentBuk on Facebook</td>
<td>Facebook</td>
</tr>
<tr>
<td>[50]</td>
<td>2014</td>
<td>Y</td>
<td>Lexicon-based, ML</td>
<td>BC</td>
<td>Tweets</td>
<td>Twitter</td>
</tr>
<tr>
<td>[51]</td>
<td>2014</td>
<td>Y</td>
<td>NB, SVM, Random Forest, Logistic Regression</td>
<td>GA</td>
<td>Tweets</td>
<td>Twitter</td>
</tr>
<tr>
<td>[52]</td>
<td>2015</td>
<td>N</td>
<td>SVM</td>
<td>GA</td>
<td>Tweets</td>
<td>Twitter</td>
</tr>
<tr>
<td>[53]</td>
<td>2015</td>
<td>N</td>
<td>Linguistic feature scoring</td>
<td>GA</td>
<td>customer reviews</td>
<td>Multi-domain SA dataset</td>
</tr>
<tr>
<td>[54]</td>
<td>2015</td>
<td>N</td>
<td>Latent Dirichlet allocation</td>
<td>BC</td>
<td>Mobile phone producer, Tweets</td>
<td>Movie reviews dataset</td>
</tr>
<tr>
<td>[55]</td>
<td>2016</td>
<td>Y</td>
<td>n-gram, Dynamic ANN</td>
<td>GA</td>
<td>Tweets</td>
<td>Twitter</td>
</tr>
</tbody>
</table>
oriented (DO) or not (Y=domain oriented, N=not domain oriented). The fourth column lists the algorithms used. The fifth column indicates whether each of SA techniques detailed in the articles performs general text analysis (GA) or tackles the task as binary classification (BC) (positive/negative). The sixth column highlights the scope of the data used for evaluation. The seventh column lists the benchmark dataset or data source used, if available.

Fig. 1. Generic SA architecture.

SA involves the main subtasks of opinion identification (identifying the opinionative words or phrases), feature extraction (identifying the features required for the analysis), sentiment classification (SC) (classifying the polarity as positive, negative or neutral), and summarization of results [27]. These subtasks are depicted in Fig. 1. Web users use various blogs, forums, social networking sites, and review sites to express their personal feelings, opinions, emotions, and feedback in relation to a particular product, service, individual, or event. As presented in Fig. 1, the method begins with data collection, which involves collecting data from opinion sources. To facilitate SA, from these opinion rich sources, only data related to the search topic under study is collected. The input document is then preprocessed. Opinion identification is then performed which involves determining whether the textual source being analyzed contains opinions. Only textual sources containing opinions undergo further processing. Additional refinements can be made at this stage (e.g., negation removal, word disambiguation). Feature extraction then follows. The feature extraction subtask, which follows opinion identification, is a subtask of SC task that aims to identify the important words in the inputted text document. Examples of text features include part-of-speech (POS), term presence and frequency, negations, and opinion words and phrases [5]. Feature extraction subtask will reduce the amount of data that needs to be considered in the subsequent SC subtask. As the extracted features will be fed to the classifier, it is vital to extract the correct features which will contribute to the success of the machine learning technique that will be used in the
SC subtask that will follow. The aim of the SC task is to classify the sentiment polarity. Finally, the summarization of results subtask involves presenting the user with a summary of the results of the analysis. This summary can range from a simple list of positive and negative evaluations to more advanced graph based summarization and summary models. The majority of the works in SA usually focus on only one or two of these four main tasks. A detailed discussion of the SA subtasks highlighted above is provided in the subsections that follow.

3. Sentiment Analysis Subtasks

3.1 Opinion Identification

A crucial step towards SA is the identification of words and phrases that express opinions. Bethard et al. [28] tackle this task by extracting propositional opinions and their holders. A propositional opinion is an opinion that is located in the propositional argument of specific verbs, such as “believe”. The authors make the assumption that opinion is a sentence, or a fragment of a sentence that answers the question “How does X feel about Y?”. Although, sentences may contain multiple sentiment expressions, the authors identify within each sentence only one verb with a propositional argument. Breck et al. [29] propose an opinion identification approach that uses conditional random fields (CRF). Other opinion identification works include the works of Wiebe et al. [30], Munson et al. [31], and Wilson et al. [32].

Kim and Hovy [33] present an approach that helps identify opinions, opinion holders, and topics from sentences. The approach has a few steps, the first of which is the identification of opinion-bearing words which are in essence opinion-bearing verbs and adjectives classified into positive, negative, and neutral classes. Htay and Lynn [34] propose an aspect/feature based approach that focuses of extracting for each aspect, from customer reviews, opinion words or phrases. The authors make use of a POS tagger to identify opinion phrases (adjectives, adverbs, verb, and nouns).

3.2 Feature Extraction

Feature extraction methods can be classified as lexicon-based or statistical based. Lexicon-based approaches use a small set of `seed’ words and then grow this set using synonyms or online resources. Such approaches can be time intensive. In addition, some terms may not occur frequently enough for proper classification. Statistical methods are fully automated methods, which work by extracting linguistic rules from domain-oriented corpus to detect candidate sentiment terms and structures. Feature extraction techniques deal with documents either as a bag of words or as a sequence of words. The bag of words approach is more popular due to its simplicity [5].

Feature extraction involves a preprocessing step that includes the application of NLP techniques such as POS, stemming, stop words removal, and noun identification to help in the identification of features. Examples of features include opinion words and phrases, terms presence and frequency, POS, n-grams, negations, etc. Many studies exist that have addressed feature selection. A study of Pang et al. [56] addresses the problem of classifying documents according to the overall sentiment they score. A few machine learning algorithms were analyzed with different feature sets (e.g., unigrams, bigrams, POS, position, feature frequency, feature presence) on a movie review dataset.
Statistical techniques are used as feature extraction methods in order to discriminate between features that need to be considered for classification and those that shouldn’t [57]. In general, these feature selection techniques fit into one of two categories, filter approach or wrapper approach. In filter approach, the feature subset selection is performed where for each given feature a measure (often statistical) is calculated. Only features with a measure greater than a threshold are classified. The most frequently used filters are mutual information, chi-square, frequency-based, and latent semantic indexing. Other statistical measures used in feature selection include hidden Markov model (HMM), latent Dirichlet allocation (LDA), weight by correlation, information gain, Gini index, and many more. In wrapper approach, the given features are evaluated by training a classifier.

There have been various attempts by researchers to address the feature selection problem using different statistical measures [58-61]. The PMI which is a measure of association/dependence between any two objects was exploited by Yu et al. [58], who developed a contextual entropy model that employs sentiment annotation to add to a list of seed words created from a small corpus of stock market news. The model calculates the similarity between two words by using an entropy system to compare their relative contextual distributions, leading to the discovery of new words related to the seed words. Both set of words are then used in the classification process. Results show that the proposed technique is capable of discovering useful emotion words, hence improving the classification results. The authors of [59] propose a machine learning approach that selects semantically relevant features minimizing the risk of over-fitting. In their work, the authors examined whether it is possible to improve stock price predication from financial news as earlier approaches have resulted prediction accuracies near guessing likelihood. The authors used expressive features to represent text and market in order to enhance the existing text mining methods. Robust feature selection allows a significant improvement in classification accuracies when used in combination with complex feature types. Duric and Song [60] used a model that combines both content and syntax and automatically learns from a review a set of features. The model distinguishes between the entities being reviewed and the subjective expressions that describe those entities in terms of polarities. By concentrating only on the subjective expressions, it was possible to choose more salient features for document-level sentiment analysis. Reyes and Rosso [61] tackled the problem of irony detection by analyzing a set of ironic customer reviews collected from social and mass media. The purpose of their research is to collect a set of discriminating elements that represent irony. The authors built a data set with ironic reviews collected from Amazon and provided valuable insights into language issues that can encounter sentiment analysis, opinion mining and decision-making tasks.

Table 2 contains details of the approaches proposed in the field of feature extraction.

<table>
<thead>
<tr>
<th>Ref</th>
<th>Year</th>
<th>DO</th>
<th>Algorithms used</th>
<th>Polarity</th>
<th>Data scope</th>
<th>Data set/sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>[60]</td>
<td>2012</td>
<td>N</td>
<td>Statistical (HMM - LDA) ME</td>
<td>BC</td>
<td>Movie reviews</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Bi-normal separation (BNS), SVM</td>
<td></td>
<td>Customer reviews</td>
<td></td>
</tr>
<tr>
<td>[59]</td>
<td>2013</td>
<td>Y</td>
<td>Chi-square, Bi-normal separation (BNS), SVM</td>
<td>GA</td>
<td>Stock market</td>
<td>DGAP, EuroAhoc</td>
</tr>
<tr>
<td>[58]</td>
<td>2013</td>
<td>Y</td>
<td>PMI-based</td>
<td>GA</td>
<td>Stock news</td>
<td>N/A</td>
</tr>
<tr>
<td>[62]</td>
<td>2016</td>
<td>Y</td>
<td>Gini index</td>
<td>GA</td>
<td>Movie reviews</td>
<td>N/A</td>
</tr>
</tbody>
</table>
3.3 Sentiment Classification

SC, as mentioned previously, is a fundamental task in the SA field which refers to the task of automatically categorizing a piece of text into classes such as, “positive”, “negative” or “neutral”. Most of the approaches to SC can be classified as being machine learning approach, lexical based approach, or hybrid approach. The machine learning approach combines linguistic features and machine learning algorithms. SC methods that use machine learning can be roughly divided into supervised, unsupervised, and semi-supervised methods. Supervised methods depend on the availability of large annotated training data. When an annotated training data is not available, an unsupervised method can be an attractive alternative. Semi-supervised method can be used when only a small amount of labeled data is available. The lexical-based approach uses a sentiment lexicon. A sentiment lexicon is a compiled set of affect words and phrases. A positive or negative score is associated with each of these affect words or phrases to reflect its sentiment polarity and strength. Lexical-based approaches can be further divided into two categories: dictionary-based and corpus based. The dictionary-based approach often starts with a small set of seed words and then grows this seed word list by iteratively searching the dictionary of their synonyms and antonyms. The corpus based approach employs statistical or syntactic methods to mine opinion words in a large corpus. The hybrid approach, a combination of machine learning and lexical based approaches, is the most popular. There have been many studies on sentiment polarity classification. These studies are summarized in Table 3.

Table 3. Summary of related sentiment classification works

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Year</th>
<th>DO</th>
<th>Algorithms used</th>
<th>Polarity</th>
<th>Data scope</th>
<th>Data set/sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>[63]</td>
<td>2009</td>
<td>Y</td>
<td>Semantic</td>
<td>BC</td>
<td>Blogs</td>
<td>TREC</td>
</tr>
<tr>
<td>[65]</td>
<td>2009</td>
<td>Y</td>
<td>Dictionary</td>
<td>GA</td>
<td>Blogs</td>
<td>N/A</td>
</tr>
<tr>
<td>[66]</td>
<td>2009</td>
<td>N</td>
<td>NB, ME, SVM</td>
<td>BC</td>
<td>Micro-blogs</td>
<td>Twitter</td>
</tr>
<tr>
<td>[67]</td>
<td>2010</td>
<td>N</td>
<td>Graph-based approach, NB, SVM</td>
<td>BC</td>
<td>Camera reviews</td>
<td>Chinese Opinion Analysis Domain (COAE)</td>
</tr>
<tr>
<td>[68]</td>
<td>2011</td>
<td>Y</td>
<td>Graph-based approach</td>
<td>BC</td>
<td>Movie, Product reviews</td>
<td>N/A</td>
</tr>
<tr>
<td>[69]</td>
<td>2011</td>
<td>Y</td>
<td>Weakly and semi supervised classification</td>
<td>BC</td>
<td>Movie reviews multi-domain sentiment data set</td>
<td>IMDB, Amazon.com</td>
</tr>
<tr>
<td>[70]</td>
<td>2011</td>
<td>N</td>
<td>CRFs</td>
<td>BC</td>
<td>Car, Hotel, Computer reviews</td>
<td>N/A</td>
</tr>
<tr>
<td>[71]</td>
<td>2011</td>
<td>N</td>
<td>Markov Blanket, SVM, NB, ME</td>
<td>BC</td>
<td>Movie reviews, News articles</td>
<td>IMDB</td>
</tr>
<tr>
<td>[72]</td>
<td>2011</td>
<td>N</td>
<td>Lexicon-based semantic</td>
<td>BC</td>
<td>Movie reviews</td>
<td>IMDB</td>
</tr>
<tr>
<td>[73]</td>
<td>2011</td>
<td>N</td>
<td>Markov logic network (MLN), SVM</td>
<td>BC</td>
<td>Product reviews</td>
<td>amazon.com</td>
</tr>
<tr>
<td>[74]</td>
<td>2012</td>
<td>N</td>
<td>NB, SVM</td>
<td>BC</td>
<td>Restaurant reviews, N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>[75]</td>
<td>2012</td>
<td>N</td>
<td>Semi supervised, BN</td>
<td>GA</td>
<td>Artificial data sets</td>
<td>N/A</td>
</tr>
<tr>
<td>[76]</td>
<td>2012</td>
<td>Y</td>
<td>NB, SVM rule-based</td>
<td>GA</td>
<td>2 sided debates</td>
<td>convinceme.net</td>
</tr>
<tr>
<td>[77]</td>
<td>2013</td>
<td>N</td>
<td>Unsupervised, LDA</td>
<td>GA</td>
<td>Social reviews</td>
<td>2000-SINA blog data set, 300-SINA, Hownet lexicon</td>
</tr>
<tr>
<td>[78]</td>
<td>2013</td>
<td>N</td>
<td>FCA, Fuzzy FCA (FFCA)</td>
<td>BC</td>
<td>Movie, eBook reviews</td>
<td>Reuters 21578</td>
</tr>
<tr>
<td>[79]</td>
<td>2013</td>
<td>N</td>
<td>SVM, Artificial neural network (ANN)</td>
<td>BC</td>
<td>Movie, GPS, Camera, Books reviews</td>
<td>amazon.com</td>
</tr>
<tr>
<td>[80]</td>
<td>2013</td>
<td>N</td>
<td>SVM, NB, C4.5</td>
<td>BC</td>
<td>Film reviews</td>
<td>MC, MCE corpus</td>
</tr>
<tr>
<td>[81]</td>
<td>2013</td>
<td>N</td>
<td>NB, SVM</td>
<td>BC</td>
<td>Movie reviews, Tweets</td>
<td>Twitter</td>
</tr>
<tr>
<td>[82]</td>
<td>2013</td>
<td>N</td>
<td>SVM</td>
<td>GA</td>
<td>Tweets</td>
<td>Twitter</td>
</tr>
<tr>
<td>[83]</td>
<td>2013</td>
<td>Y</td>
<td>Taxonomy-based, corpus-based</td>
<td>BC</td>
<td>Headphones, Car, Hotel reviews</td>
<td>epinions.com</td>
</tr>
<tr>
<td>[85]</td>
<td>2015</td>
<td>N</td>
<td>ISVM</td>
<td>GA</td>
<td>Micro-blogs</td>
<td>Micro-blogs</td>
</tr>
</tbody>
</table>
3.4 Sentiment Summarization

The widespread use of social media platforms such as Facebook, Twitter, Tumblr, Instagram, and YouTube have provided individuals with first-hand access to the sentiments and reactions of millions of people. Analyzing such data, entails high computational power, before the results can be summarized and efficiently presented to the intended users in a concise understandable form. It is essential for the users when they look at the results to get a general understanding of peoples’ sentiments towards an item/product or a specific feature. In the case where the volume of data is large, presenting this data in a graphical manner can be more efficient than presenting it in basic tabular or numerical formats [86].

Some researchers present opinions in the form of an aspect based summary [87-93]. When presenting the summary of opinion, Hu and Liu [87] employ two numbers for each aspect; one shows the number of people who have positive feelings towards this aspect and the other shows the number of people who have negative feelings towards the same aspect. Users can click links to access the actual sentences that express these sentiments. Ku et al. [88] and Mei et al. [90], on the other hand, make use of a summary timeline to highlight opinion changes over time. Lu et al. [89], Popescu and Etzioni [91], and Titov and McDonald [92] in their summarization technique show the most occurring phrase in each aspect. Zhuang et al. [93] provide a statistical summary showing the sentiment distribution of each aspect along with the corresponding sentences for each aspect and sentiment.

Fukuhara et al. [94] propose a system for temporal SA that produces two kinds of graphs represented by a line diagram. The first graph is a topic graph that shows over a period of time the temporal change of topics related to a particular sentiment (e.g., happy, sad, anger) and the second graph is a sentiment graph that shows over a period of time the temporal change of sentiments related to a particular topic. Visualizations by Fukuhara et al. [94] are limited to showing a range of sentiments associated with a single event or a range of events associated with a single sentiment. Havre et al. [95] propose ThemeRiver, a system that uses a river metaphor to visualize thematic changes in documents over time. The river in the graph is pictured flowing from left to right, altering in width to mirror changes in thematic strength. Colored “currents” flowing within the river narrow or widen to designate decreases or increases in the strength of a single topic or set of topics in the related documents.

Marcus et al. [96] present TwitInfo, a system that visualizes and summarizes events on Twitter. TwitInfo extracts the tweets that match a keyword and arranges them on a timeline that highlights peaks of high tweet activity. TwitInfo utilizes information retrieval techniques to assign peaks with meaningful labels using text from the tweets. TwitInfo uses only two categories of sentiment (positive and negative). Vox Civitas [97] is another visualization tool for Twitter. Similar to TwinInfo, Vox Civitas uses temporal visualization. Vox Civitas is used by journalists to support the exploration of tweets about specific events. Vox Civitas makes the use of a sentiment bar. The color of the bar for each time interval represent the polarity (positivity, negativity, or controversial) of the aggregate of all messages in that interval. Towards the bottom of the Vox Civitas interface, a set of keywords appears along the sentiment time-line.

The VISA [98] and the TriVis [99] systems also employ visualization techniques to summarize SA results and interactively present them to the user. VISA has three different visualization views. The first view is the sentiment trend view which shows chronologically, for the different categories/topics, the sentiment dynamics and sentiment comparisons. The second view employs chart visualization to highlight the associated structured facets. Finally, the third view shows the snippet/document panel which provides details of the documents and the context of sentiment [98].
TriVis [99] employs the three sentiment classes (positive, neutral and negative) to visualize multivariate social sentiment data. TriVis combines sentiment and volume in one graph. Trivas uses color to distinguish between multiple topics. In addition, TriVis uses a modified scatter plot with three axes to allow the user to easily interpret and understand multivariate data. Other systems employ simpler visualization techniques. The Semantize [100] system uses font styles and background colors to mark the sentiments found in documents. Pino et al. [101] provide a tool that visualizes sentiments related to major events in geographically confined populations on interactive maps.

4. Applications of Sentiment Analysis

SA has found its way in many applications across multiple domains. This section discusses some of the common applications. The examples presented in this section are not complete but exist to provide a snapshot of the possibilities.

People are sensitive to sentiment; hence, it is only natural from them to always want to find out what others’ opinions are on, amongst many, a specific product, application, service, or political figure. Classically, when someone needed to know others’ opinions, he/she would ask his/her friends, family, and/or others for their opinion. Moreover, when an organization needed to know public or customer opinions, it conducted surveys and/or focus groups.

Recent years have witnessed the global tremendous growth of e-commerce and social media. This growth, combined with their sentiment rich nature, has led to an increasing use of the content available in these sites for purposes of decision-making. Unfortunately, the large number of diverse sites available coupled with the large volume of information contained in them, makes the task of finding and monitoring opinions very difficult. Hence, the need for automated SA systems has dramatically increased.

Documents containing opinions do not only exist on the web. Many companies possess internal sentiment rich sources. Examples of such sources include customer feedback collected via email, surveys, and/or over the telephone.

The applications of SA are numerous. In order to best understand their varying applications, they are grouped into three categories: applications for business analytic, applications for political/government/public analytic, and other types of applications.

4.1 Applications for Business Analytic

Mining sentiments from public sources can be particularly effective in the area of business analytics. The most common SA applications in this area center around reviews of consumer products and services. Hence, many works have been published in literature detailing the design of such applications. Liu et al. [102] in their paper propose, ARSA, a sentiment-aware model that uses blogs to predict sales performance. McGlohon et al. [103] exploit the sentiments expressed in reviews to rank products and merchants.

Some works exploit sentiment rich sources for the purpose of predicting movies box-office revenues [104-106], whilst others [107] test the effectiveness of their approach in improving the predication quality of box-office revenues.

Another widespread set of applications is in stock market analysis. Bollen et al. [108] exploited the sentiments in tweets for purposes of performing stock market analysis. Bar-Haim et al. [109] and
Feldman et al. [110] propose a framework for identifying expert investors, and then use this information in models that exploit micro blogging messages for the purpose of predicting stock rise and fall.

Twitter and Facebook are a central source of many sentiment analysis applications in this area, of which brand reputation monitoring is the most common. One application that allows companies to explore consumers’ opinions about their brands or products is Trackur [111]. Many other brand monitoring applications exist, including BrandsEye [112], TweetFeel [113], Twittratr [114], and Trendrr [115].

4.2 Applications for Public Opinion Analytic

Other SA research efforts and applications utilize political information sources, such as political blogs, tweets with political content, candidates’ web sites/Facebook pages, and/or online news.

Hong and Skiena [116] study the relationships between the National Football League betting line and public opinions in blogs and Twitter and show that a significant correlation exits between the public sentiments (expressed in blogs and tweets) and the betting line. Tumasjan et al. [117] were the first researchers who published their efforts detailing the application of SA on twitter data for purposes of predicting the national German Federal election results, reporting a strong correlation between the data embedded in 140 character tweets and the election results. Subsequent efforts include the works of DiGrazia et al. [118], Franch [119], Ceron et al. [120], Caldarelli et al. [121], and Burnap et al. [122]. Chen et al. [123] use an opinion scoring model to explore political standpoints. Applications under this category also include those that try to look at the public image and reputation of a specific population (e.g., gender, race, social class, and religious minority) [124-126]. Kwon and Lee [127] exploit personal and social opinions for the purpose of bias detection.

Other applications are those that target the sentiments expressed in news articles. Some of these applications are interested in the polarity of these articles towards the subject of interest [126], whilst others are interested in the attitude that these articles impart [128]. Applications targeting sentiment in political campaign discourse [129] and campaign publications and news releases [130-132] also fall under this category.

SA has also been explored in the area of eRulemaking, for the purpose of analyzing the opinions of people towards a specific policy or government regulation [133-135]. Other efforts reported exploit “blawgs”, weblogs devoted to legal tissues [136].

4.3 Other Types of Applications

To better depict the diversity of SA applications, this subsection includes some additional examples of such applications. Groh and Hauffa [137] employ NLP based SA techniques on a collection of email messages to characterize social relations, whilst Mohammad and Yang [138], apply SA on emails to uncover the emotional variations between genders.

Sakunkoo and Sakunkoo [139], in an attempt to answer the question of whether and how previous opinions on a certain item affect the opinions that come after, study the social influences in online book reviews.

Visual SA is another area that has recently attracted some attention. Wang et al. [140] propose USEA, an unsupervised SA framework that predicts the sentiments of social media images. Tedmori and Al-Lahaseh [141] propose an addition to social networks that automatically analyses selfies for their underlying sentiments. The system then automatically generates sentiment bearing hashtags.
5. Discussions and Challenges

The general SA task and the SC subtask are and will continue to be the most attractive opinion mining research areas. In regards to SC, machine learning algorithms are the most popular due to their simplicity, and the fact that these algorithms learn from training data which make them particularly useful for domain adaptability. The lexicon-based algorithms, on the other hand are very popular in addressing the general SA problems. Lexicon-based algorithms are not only scalable, but also simple and computationally efficient [5]. In relation to feature extraction, the use of unigram is simple and helps improve the performance of the SC. However, looking beyond unigrams, into bigrams and trigrams can boost SC accuracy considerably. In addition, POS tagging has also been shown to also improve the SC performance. Moreover, the correct ensemble of feature selection, suitable to the task at hand, should improve further the accuracy. As stated by many, SVM performs better than other classifiers including Naïve Bayes for sentiment analysis. It is worth however, to continue experimenting with other classifiers or with an ensemble of classifiers in order to significantly improve SC.

Although a variety of research can be found employing lexicon-based approaches to tackle SA as a general analysis of text using, it still remains largely unsolved. General sentiment is an immensely more difficult task than binary polarity classification (positive/negative classification). Moreover, it’s only natural that more accurate results can be obtained when using domain-specific data instead of domain-general data. According to [5], and as can be seen in Tables 1–3, more researchers have worked and are continuing to work on domain-general data, making domain-specific SA an ongoing field of interest. Moreover, the majority of works detailed in this research target the English language. Other languages remain under-explored. Another main problem in SA for non-English is a significant lack of resources. Hence, some researchers can focus their efforts on building resources such as lexica, corpora and dictionaries, and other necessary resources.

6. Conclusions

SA has found its way in many applications across multiple domains such as the political, government, public and business analytic. SA refers to the process of identifying and extracting opinions, emotions, and evaluations from a specific input. Standard approaches to SA have been classified in many different ways. The most fundamental classification is into the three broad categories of: keyword spotting, lexical affinity, and statistical methods. SC, a fundamental task in the SA field, refers to the to the task of automatically categorizing a piece of text into classes such as, “positive”, “negative” or “neutral”. SC approaches can be categorized as being either machine learning based, lexical based approach, or hybrid. Moreover, the methods of feature extraction, another vital SA task, can be classified into two categories: namely, lexicon-based methods that need human annotation, and statistical methods which are fully automatic methods.

The aim of this paper is to provide an in-depth up-to-date study of the SA techniques in order to familiarize with other works done in the subject. The paper focuses on the main tasks and applications of sentiment analysis. The majority of the works detailed in this survey target the English language. Hence, various English language lexica, corpora and dictionaries are freely available for research purposes. However, SA for non-English languages including Arabic, Spanish, Italian, German, Dutch, Chinese,
Japanese, and Taiwanese has only recently attracted researchers’ attention paving way to new challenges including building language specific resources. Not only micro-blogs, blogs and forums are popular data sources for sentiment analysis, social networks can also be used to identify authors’ sentiments on brands, products, or services. The majority of researchers use different datasets for their different testing objectives. This calls for the need for a single dataset that can be used to accurately test and compare the performance of these algorithms.

References


S. M. Mohammad and T. W. Yang, “Tracking sentiment in mail: how genders differ on emotional axes,” in Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis, Portland, OR, 2011, pp. 70-79.


**Sara Tedmori** https://orcid.org/0000-0002-0391-0160

In 2001, she received her B.Sc. degree in Computer Science from the American University of Beirut, Lebanon. In 2003, she obtained her M.Sc. degree in Multimedia and Internet Computing from Loughborough University. In 2008, she received her Engineering Doctorate in Computer Science from Loughborough University, UK. Currently she is an associate Professor in the Computer Science Department at Princess Sumaya University of Technology, Jordan. Her research interests include natural language processing, sentiment analysis, data mining, image processing, video coding, cryptography, social networks, web technologies, and classification.

**Arafat Awajan** https://orcid.org/0000-0002-7067-5658

He is a full professor at Princess Sumaya University for Technology (PSUT). He received his Ph.D. degree in computer science from the University of Franche-Comte, France in 1987. He held different academic positions at the Royal Scientific Society and Princess Sumaya University for Technology. He was appointed as the chair of the Computer Science Department (2000–2003) and the chair of the Computer Graphics and Animation Department (2005–2006) at PSUT. He had been the dean of the King Hussein School for Information Technology from 2004 to 2007, the Dean of Student Affairs from 2011–2014 and the director of the Information Technology Center in the Royal Scientific Society from 2008–2010. He is currently the dean of the King Hussein School for computing Sciences and leads a research group on natural language processing and arabic text mining.