A survey on opinion mining and sentiment analysis: Tasks, approaches and applications

Article *in* Knowledge-Based Systems · June 2015

DOI: 10.1016/j.knosys.2015.06.015

1 author:

Kumar Ravi
Institute for Development & Research in Banking Technology

PRE-PRINT VERSION

A survey on opinion mining and sentiment analysis: tasks, approaches and applications

Kumar Ravi[1,2] and Vadlamani Ravi[1,*]

[1]Center of Excellence in CRM and Analytics,
Institute for Development and Research in Banking Technology,
Castle Hills Road No. 1, Masab Tank, Hyderabad - 500057 (AP), India
[2]School of Computer & Information Sciences,
University of Hyderabad, Hyderabad-500046 (AP), India
(ankitaravi.ravi00@gmail.com; rav_padma@yahoo.com)

Abstract: With the advent of Web 2.0, people became more eager to express and share their opinions on web regarding day-to-day activities and global issues as well. Evolution of social media has also contributed immensely to these activities, thereby providing us a transparent platform to share views across the world. These electronic Word of Mouth (eWOM) statements expressed on the web are much prevalent in business and service industry to enable customer to share his/her point of view. In the last one and half decades, research communities, academia, public and service industries are working rigorously on sentiment analysis, also known as, opinion mining, to extract and analyze public mood and views. In this regard, this paper presents a rigorous survey on sentiment analysis, which portrays views presented by over one hundred articles published in the last decade regarding necessary tasks, approaches, and applications of sentiment analysis. Several sub-tasks need to be performed for sentiment analysis which in turn can be accomplished using various approaches and techniques. This survey covering published literature during 2002-2015, is organized on the basis of sub-tasks to be performed, machine learning and natural language processing techniques used and applications of sentiment analysis. The paper also presents open issues and along with a summary table of a hundred and sixty one articles.

Key Words: Opinion Mining, Sentiment Analysis, Social Media, Micro Blog, Lexica Creation, Machine Learning, Ontology

[*] Corresponding Author, Phone: +91-40-23294042; FAX: +91-40-23535157

Knowledge-Based Systems 89 (2015) 14–46
http://dx.doi.org/10.1016/j.knosys.2015.06.015

1. Introduction

Ever increasing use of Internet and online activities (like chatting, conferencing, surveillances, ticket booking, online transactions, e-commerce, social media communications, blogging and micro-blogging, clicks streams, etc.) leads us to extract, transform, load, and analyze very huge amount of structured and unstructured data, at a fast pace, referred to as Big Data. Such data can be analyzed using a combination of Data Mining, Web Mining and Text Mining techniques in various real life applications. Huge amount of information related to customer opinions/reviews is quite cumbersome to analyze and needs extant approaches to get a generalized opinion summary. Numerous forums, blogs, social networks, e-commerce web sites, news reports and additional web resources serve as platforms to express opinions, which can be utilized for understanding the opinions of the general public and consumers on social events, political movements, company strategies, marketing campaigns, product preferences, and monitoring reputations [26]. To accomplish these tasks, research communities and academicians are working rigorously on sentiment analysis for last one and half decade. Sentiment analysis (SA) is a computational study of opinions, sentiments, emotions, and attitude expressed in texts towards an entity [138]. Sentiment analysis (also called opinion mining, review mining or appraisal extraction, attitude analysis) is the task of detecting, extracting and classifying opinions, sentiments and attitudes concerning different topics, as expressed in textual input [84]. SA helps in achieving various goals like observing public mood regarding political movement, market intelligence [90], the measurement of customer satisfaction [158], movie sales prediction [131] and many more.

Sentiments, evaluations, and reviews are becoming very much evident due to growing interest in e-commerce, which is also a prominent source of expressing and analyzing opinions. Nowadays, customers on e-commerce site mostly rely on reviews posted by existing customers and, producers and service providers, in turn, analyze customers' opinions to improve the quality and standards of their products and services. For example opinions given on e-commerce sites like Amazon, IMDb, epinions.com etc can influence the customers' decision in buying products and subscribing services [18].

In developing countries, online and social media is taking the place of offline media swiftly, which encourages common people to involve in political discussions and enable them to put across unilateral thoughts on Global issues interactively. Online media provides the platform for wide sharing of ideas and encouraging public for group discussions with open views. Online media provides better means to get quick response and feedback on different Global issues and entities in the form of textual posts, news, images, and videos. Thus, it can be utilized to analyze peoples' opinions for learning the behaviors of consumer, patterns market, and trends of society [206]. Twitter has 255 million monthly active users and it oversees 500 million tweets every day[1]. Thus, it serves as a good resource to extract heterogeneous opinions published by people from diverse societies for different purposes like improvement of quality of products and services, prediction of consumers' demand and taste etc.

Online media and social networking sites (SNS) are used to express and share public experiences in the form of product reviews, blogs, and discussion forums. Collectively, these media contain highly unstructured data combining text, images, animations and videos that are useful in making public aware of various issues.

1.1 Earlier reviews

Pang et al. [1] performed an extensive survey of more than three hundred papers by covering applications, common challenges for sentiment analysis, major tasks of opinion mining viz., opinion extraction, sentiment classification, polarity determination, and summarization. Then, Tang et al. [16] discussed four problems related to opinion mining, i.e., subjectivity classification, word sentiment classification, document sentiment classification and opinion extraction. For subjectivity classification, they highlighted some approaches like similarity dependent, NB classifier, Multiple NB classifier, and cut-based classifier.

O'Leary et al. [89] presented a survey on blog mining, which includes introduction on blog search and mining, type of blogs to be analyzed, unit and type of opinions to be extracted from blogs, and their applications. Montoyo et al. [84] listed some open issues along with achievements obtained thus far in the area of subjectivity analysis and sentiment analysis. Tsytsarau and Palpanas [137] presented a survey on SA by focusing on opinion mining, opinion aggregation including spam detection and contradiction analysis. They compared opinion mining methods, which were employed on some common dataset.

Liu [181] presented different tasks possible and works published in SA and opinion mining. Major tasks listed are subjectivity and sentiment classification, aspect-based SA, sentiment lexicon generation, opinion summarization, analysis of comparative opinions, opinion search and retrieval, opinion spam detection and quality of reviews. Cambria et al. [15] pointed out complexities involved in SA with respect to current demand along with possible future research directions. Recently, Feldman [14] focused on five specific problems in the field of SA: Document-level SA, sentence-level SA, aspect-based SA, comparative SA and, sentiment lexicon acquisition. They also listed some open issues like SA of composition statement, automatic entity recognition, discussion on multi-entity in same review, sarcasm detection and subjectivity classification at finer level.

Most recently, Medhat et al. [138] presented a survey on feature selection and sentiment classification methods. A very brief description is presented about feature selection methods (mainly pointwise mutual information and Chi-square) and a detailed discussion is presented on sentiment

[1] https://about.twitter.com/company.

http://dx.doi.org/10.1016/j.knosys.2015.06.015

classification methods and related papers. They summarized fifty-four articles listing out task accomplished, domain, algorithm utilized, polarity, data scope, data source, and type of language. The authors' major concern is to discuss the techniques applied in surveyed papers.

Along with these surveyed papers, a considerable amount of work has been reported in this area and a number of lexica have been created by research community to evaluate new devised sentiment analysis algorithm. Especially in the last four years, major concerns of researchers are micro-blogs, which have been successfully applied for market prediction [18], social advertising [43], and box-office prediction [51]. Tsytsarau and Palpanas [137] presented very limited discussion on this major domain and its applications. In addition to that, several other issues are reported in recently published papers. Therefore, there is an urgent need to focus on several other issues raised in currently published papers, which were not the part of the extant surveys.

This survey work differs from existing literature surveys in various ways (i) we classified existing studies on the basis of opinion mining tasks, approaches and applications as presented in Figure 1, (ii) this paper presents articles related to tasks and major issues pointed out by existing articles like subjectivity classification, sentiment classification from coarse-grained to fine-grained level, review usefulness measurement, opinion spam detection, lexicon creation, and opinion word and product aspect extraction as presented throughout the paper (iii) we summarized each of surveyed articles in four aspects viz. problem addressed, exploited dataset details, feature representation and selection method (if applied), techniques applied, obtained results, and indicated future directions along with our views, (iv) we included some recently proposed feature selection techniques for SA, (v) we provided a detailed list of online available datasets, (vi) classification of articles on the basis of SA performed at various granular levels as presented in Table 1, (vii) the exploited lexica are listed in Table 10, and (viii) summary of one hundred and sixty one articles is presented in Table 10 before concluding the paper.

Therefore, works addressing these issues are considered for this survey. This rest of the paper is organized as follows: Section 2 presents background information related to the survey. Section 3 presents state-of-the art discussion on SA covering common issues listed in previous paragraph. Detailed discussion on the existing work, open issues, and possible applications of sentiment analysis is presented in section 4. The paper is concluded in section 5.

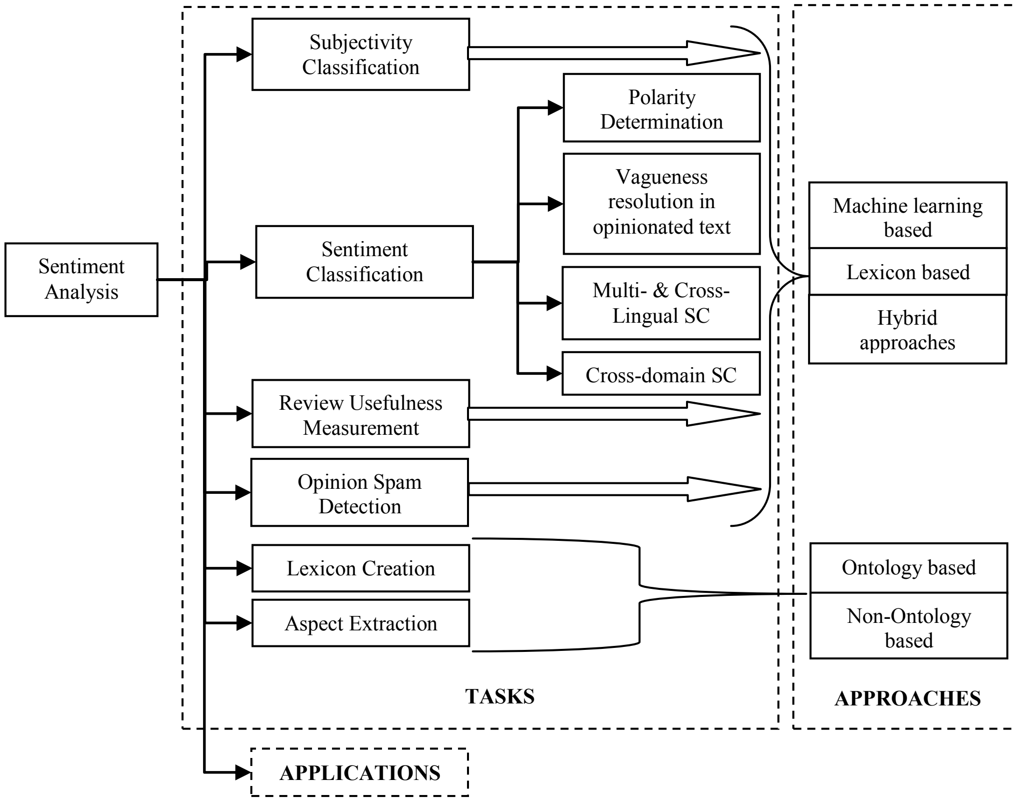

Figure 1. Organization of the review

Table 1. Distribution of articles based on the granularity of sentiment analysis

Level of analysis	#Articles	Articles' References
Document level	73	[13], [18], [22], [32], [33], [36], [40], [43], [45], [48], [50], [51], [53], [54], [61], [64], [66], [77], [81], [80], [85], [88], [90], [91], [94], [96], [101], [111], [117], [121], [123], [130], [131], [132], [148], [155], [156], [157], [158], [167], [168], [169], [175], [176], [177], [179], [180], [182], [194], [195], [197], [200], [203], [205], [206], [207], [209], [210], [211], [212], [217], [220], [221], [222], [223], [224], [225], [226], [227], [228], [229], [231], [232]
Word level	25	[26], [37], [45], [55], [57], [61], [67], [69], [73], [74], [86], [97], [99], [109], [110], [112], [114], [116], [118], [127], [163], [165], [166], [180], [184]
Aspect level	23	[8], [25], [35], [41], [44], [56], [59], [62], [63], [67], [68], [76], [93], [107], [126], [154], [185], [186], [189], [193], [218], [240], [241], [243]
Sentence level	20	[24], [29], [46], [58], [100], [108], [124], [125], [160], [168], [171], [173], [174], [178], [183], [186], [190], [191], [193], [213]
Concept level	9	[21], [23], [30], [52], [95], [98], [202], [214], [216]
Phrase level	3	[12], [162], [172]
Link based	3	[11], [49], [70]
Clause level	2	[29], [170]
Sense level	1	[75]

2. Preliminary steps of sentiment analysis

Sentiment analysis, opinion mining and subjectivity analysis are interrelated areas of research which use various techniques taken from Natural Language Processing (NLP), Information Retrieval (IR), structured and unstructured Data Mining (DM). Major part of data available worldwide, being unstructured (such as text, speech, audio, video etc.), poses important research challenges. To deal with such unstructured text data, traditional methods of NLP i.e. information retrieval and information extraction came into existence [84]. In order to get a sense of the extracted text, numerous research efforts have been witnessed in recent years leading to automated SA, an extended NLP area of research [23].

Sentiment analysis is not a single problem; instead it is a multi-faceted problem [134]. Various steps are needed to perform opinion mining from given texts, since texts for opinion mining is coming from several resources in diverse format. Data acquisition and data preprocessing are most common sub-tasks required for text mining and SA, which are discussed in this section.

2.1 Data acquisition

Due to wide availability of various online resources, data acquisition is highly subjective to the type of media, data format supported by media, and the type of analysis needed to perform. Some micro-blogging sites like Twitter, Sina-Wiebo etc made available their Application Programming Interface (API) to collect public data from their sites. Twitter has provided Twitter REST API to get static data like user profile information, and Streaming API[2] to get streaming data like tweets [19]. Twitter4J API[3] has been exploited by [85, 96] to extract streaming tweets. Similarly, Facebook and Sina-Wiebo have made available Facebook Graph API[4] and Tancent API[5] respectively. These APIs help us extract posts and other information from their site as well and are exploited in [49, 123, 190, and 209]. Xu et al. [70] collected information about 5,012 members and 23,507 friendships from a product review social network "UrCosme.com" and tried to measure the social influence of one user to other. Google Scholar and Technorati is most commonly utilized blog search engine and have been exploited in [48].

2.2 Preprocessing

Raw data acquired from various sources often needs to be preprocessed before launching a fully fledged analysis. Some popular preprocessing steps are: tokenization, stop word removal, stemming, parts of speech (POS) tagging, and feature extraction and representation. Tokenization is used to break a sentence into words, phrases, symbols or other meaningful tokens by removing punctuation marks. Stop words do not contribute to analysis and hence aredropped during preprocessing step. Stemming is the process to bring a word into its root form, while ignoring other POS of the word. POS tagging is performed to recognize different parts of speech in the text, which is quite essential for natural language processing. Some of the publicly available tools for different preprocessing tasks are listed in Table 2. Due to sparseness and extreme noise in textual data, it often requires extreme level of feature extraction, which is also one of the important preprocessing steps. Aside from feature extraction, feature selection is also critical to the success of any analysis. A study of sentiment analyses reported in [138] highlights different

[2] https://dev.twitter.com/docs/api/1.1.
[3] http://twitter4j.org/en/javadoc.html.
[4] https://developers.facebook.com/docs/graph-api.
[5] http://dev.datasift.com/docs/sources/public-sources/tencentweibo.

feature selection techniques like Pointwise Mutual Information (PMI), chi-square, and latent semantic indexing. In addition to these, some more statistical feature extraction methods, proposed in recent literature, are summarized here.

Table 2. Available tools for text preprocessing

Reference	Name of the tool	Purpose
[251]	TweetMotif	Tokenization of tweets
[204], [250]	POS tagger	Twitter POS tagger
[244]	TweetNLP[6]	Twitter natural language processing
[245]	Lancaster stemming algorithm	Stemmer
[246]	GNU Aspell	Spell Checker
[247]	Snowball	English stemmer
[248]	Stanford Log-linear Part-Of-Speech Tagger	POS tagger
[249]	TweeboParser	Tweet Dependency parser

Wang et al. [44] performed subjectivity classification by considering improved *Fisher's discriminant ratio* based feature selection method. Experiments were performed on two Chinese corpora, multi-domain reviews; COAE2008s, and 11 different car brand reviews; BOACAR. The proposed feature sets along with words appearing in positive (+ve) and negative (-ve) texts were used for training Support Vector Machine (SVM), which yielded sentiment classification accuracy of 86.6%. Accuracy can be improved by training the classifier on richer dictionary. Then, Vinodhini & Chandrasekaran [197] employed *principal component analysis* for dimension reduction and ensembled hybrid techniques for sentiment classification. They experimented with 500 (250 +ve & 250 -ve) reviews on digital camera. For feature representation, they tried different combination of unigram, bi-gram and tri-gram representation. SVM and Naive Bayes (NB) were compared against bagged SVM and Bayesian boosting. Bayesian boosting outperformed all methods using hybrid (i.e. unigram + bigram) feature representation by yielding the highest precision of 83.3%. In future, more novels, diverse and powerful hybrids/ensembles of feature reduction techniques can be employed.

Abbasi et al. [160] considered multi-lingual (viz. English and Arabic) sentiment classification of forums. Sentiment classification was performed on the basis of syntactic and stylistic features such as word-length distributions, vocabulary richness measures, character- and word-level lexical features, and special-character frequencies. Features were extracted using newly proposed *Entropy Weighted Genetic Algorithm* (EWGA). The first experiment, conducted on movie review dataset [33, 167] achieved the highest accuracy of 87.95% with features selected using Stylistic + Syntactic techniques. The EWGA outperformed all other feature engineering techniques by yielding an accuracy of 91.70% using SVM. The second experiment was performed on 1000 English web forums messages on US extremist forums and 1000 Arabic web forums messages. EWGA improved the classification accuracy up to 92.84% on English dataset, and 93.84% on Arabic dataset. In future, the proposed classification approach is to be

[6] Tweet NLP:http://www.ark.cs.cmu.edu/TweetNLP/.

http://dx.doi.org/10.1016/j.knosys.2015.06.015

applied on sentence- and phrase-level SA. Moreover, EWGA can be employed for topic, genre, and style classification.

Along with these notable feature selection based studies, we surveyed others involving syntactic [92, 130, 171, 174, 178, 190], semantic [32, 92, 48, 60, 69, 109, 155, 169, 172, 190], stylistic [116, 117, 118, 127], and information gain [75, 165] based methods.

3. Review methodology

As mentioned earlier, the current review is conducted in six broad dimensions viz. *subjectivity classification, sentiment classification, review usefulness measurement, lexicon creation, opinion word and product aspect extraction*, and *various applications* of opinion mining as presented in Figure 1. First five dimensions represent *tasks* to be performed in the broad area of SA. In turn, sentiment classification is further divided into four categories: *polarity determination*, *vagueness resolution in opinionated text, multi-lingual and cross-lingual SA*, and *cross-domain sentiment classification*. These sub-categories considered as *sub-tasks*. Distribution of reviewed articles with respect to different tasks and sub-tasks is presented in Table 3. Here, subsection 3.1-subsection 3.5 discuss different *tasks* and *sub-tasks* along with applied *approaches* and *techniques*. Applied *approaches* are broadly classified into three categories viz. *machine learning, lexicon based,* and *hybrid approaches* for subjectivity classification, sentiment classification, review helpfulness measurement, and opinion spam detection as presented in Figure 1. Further, *ontology* and *non-ontology based approaches* are considered for *lexicon creation* and *aspect extraction*. All attempted *approaches and techniques* are discussed along with review of each article. For detailed discussion on various techniques applied for SA, an interested reader can refer Medhat et al. [138]. The last subsection 3.6 presents various existing *applications* of SA, whereas possible future applications are discussed in the next section.

Table 3. Distribution of articles based on tasks and applications

S#	Tasks and applications	#Articles	References
1	Subjectivity Classification	6	[44], [75], [110], [163], [167], [174]
2	Polarity determination	43	[12], [26], [29], [32], [33], [35], [40], [45], [48], [50], [54], [57], [66], [85], [95], [96], [108], [109], [112], [114], [123], [126], [154], [156], [157], [160], [162], [165], [166], [168], [169], [170], [171], [172], [176], [177], [178], [179], [180], [203], [205], [206], [209]
3	Vagueness in opinionated text	5	[22], [41], [86], [216], [217]
4	Multi- & cross-lingual SA	6	[46], [88], [94], [115], [148], [173]
5	Cross-domain SA	4	[36], [98], [99], [121]
6	Review usefulness measurement	13	[76], [78], [81], [130], [221], [222], [223], [224], [225], [226], [227], [228], [229]
7	Opinion spam detection	7	[199], [200], [212], [216], [220], [231], [232]
8	Lexica and corpora creation	22	[21], [23], [24], [30], [52], [55], [56], [69], [74], [97], [106], [111], [116], [117], [118], [127], [136], [202], [207], [211], [213], [214]
9	Opinion word and aspects extraction, entity recognition, name disambiguation	36	[8], [11], [25], [27], [35], [37], [59], [60], [61], [62], [63], [67], [68], [92],[93], [100], [101], [102], [107], [125], [132], [175], [182], [185], [186], [189], [190], [191], [193], [194], [195], [196], [218], [240], [241], [243]
10	Applications of SA	21	[13], [18], [43], [47], [49], [51], [53], [58], [64], [73], [77], [79], [80], [90], [91], [124], [131], [155], [158], [183], [184]
	Total	**163**	

http://dx.doi.org/10.1016/j.knosys.2015.06.015

The important aspects considered for the reviews of each article are *problem addressed*, *exploited dataset details*, *feature representation and selection method* (if applied), *obtained results*, and *indicated future directions* by author and/or us. Most of the studies do not have uniformly same experimental setup. Therefore, results of different studies are not directly comparable. Still, we tried our best to compare some related works wherever possible.

3.1 Subjectivity classification

Subjectivity classification deals with the detection of "private states" - a term that encloses sentiment, opinions, emotions, evaluations, beliefs and speculations [84]. Maks and Vossen [24] claimed that subjectivity can be expressed in different ways, at various text levels, and using several types of expressions. Subjectivity is the property associated with words and word sense. Some more complex opinionated texts are news, political documents, and online debates, which requires identification of the attitude holder and topic also. Subjectivity analysis is defined as the recognition of opinion-oriented language in order to distinguish it from objective language. Subjectivity and SA are challenging tasks, spanning over many different areas and applications. Although a considerable number of studies have been reported in this field in the past decade, much remains to be done in order to create systems that can be reliably utilized in real-life applications. The problem of distinguishing subjective versus objective instances has often proved to be more difficult than subsequent polarity classification. Therefore, improvements in subjectivity classification promise to positively impact on sentiment classification. It has been performed using machine learning [44, 167] as well as lexicon based approaches [75, 110, 163, 174], whereas Banea et al. [75] and Molina-González et al. [110] undertaken cross-lingual approaches. Some of the subjectivity classification studies carried out on a common dataset provided by Pang and Lee [167] are presented in Table 4 for comparison purpose.

Table 4. Subjectivity classification accuracy reported on common datasets

S#	Dataset	Article	Accuracy achieved	Approaches
1	[167]	[160]	91.70	Hybrid
2		[167]	92	Machine learning
3		[174]	92.1	Lexicon based
4	[33]	[92]	87.5	Machine learning

Pang and Lee [167] utilized physical proximity between the items to be classified, where cut-based classification has been applied for subjectivity classification. Cut-based subjectivity detectors determine the subjectivity status of all sentences of the document using per-item and pair-wise relationship information. 5,000 movie-reviews snippets from www.rottentomatoes.com, for subjective sentences, and 5,000 sentences from plot summaries available from www.imdb.com, for objective sentences, have been considered in this study. NB and SVM yielded average 10-fold cross validation (FCV) performance on the subjectivity dataset accuracy of 92% and 90% respectively. Possible development claimed as inclusion of parameter-selection techniques and incorporation of other contextual cues besides sentence proximity. Xuan et al. [174] achieved slighted better accuracy than that of Pang and Lee [167]. They constructed in total twenty-two syntactic patterns over adjectives, adverbs, verbs, and noun to determine the subjectivity of the text. Maximum entropy was employed on the Pang and Lee [167] dataset and achieved up to an accuracy of 92.1%, which turned out to be better than selected three baseline methods. The proposed approach should be tested on other languages also. Bravo-Marquez et al. [163] utilized the diverse features of several lexica together for subjectivity and sentiment classification like polarity, strength and emotion. They experimented with Go et al. [164] dataset, Sanders7, and SemEval8 and improved accuracy by 5% compared to individual classifier. Due to streaming texts, online algorithms should be experimented for real time subjectivity and sentiment classification.

Molina-González et al. [110] presented a new Spanish lexicon (SOL-Spanish Opinion Lexicon) comprising opinion words for performing opinion mining. They utilized the Bing Liu English Lexicon [38] and automatically translated it into Spanish language to get Spanish Opinion Lexicon (SOL). Sentiment classification was performed on 3,878 movie reviews selected from MuchoCine Corpus (MC) in Spanish language and achieved an accuracy of 63.16%. In future work, random walk algorithm can be utilized for building domain-oriented sentiment lexicons for Spanish language. Banea et al. [75] performed sense level multilingual subjectivity classification for English and Romanian languages. They have started with intersection of words from English WordNet [133] and Romanian WordNet to get initial senses as seeds. Multilingual bootstrapping was applied to explore a list of senses. The obtained list yielded subjectivity classification accuracy of 73.98% for both languages. The quantitative study was also performed by applying information gain based feature selection. In future, the proposed approach can be applied in various other languages and a list of senses can be explored further.

3.2 Sentiment classification

Sentiment classification is the determination of orientation of sentiment of given text in two or more classes. Sentiment classification has been performed in various classes like binary, ternary, n-ary in the form of stars [130], and “thumbs up” or “thumbs down” [32, 33] etc. Sentiment classification can be performed using machine learning as well as lexicon-based approaches as reported in [138]. Machine learning yields maximum accuracy while semantic orientation provides better generality. Machine learning can be further divided into supervised and unsupervised approaches. For supervised approaches, we need two sets of annotated data, one each for training and testing. Some of the commonly applied classifiers for supervised learning were Decision Tree (DT), SVM, Neural Network (NN), Naïve Bayes, and Maximum Entropy (ME). Under lexicon based approaches, one can use either dictionary or corpus based approach. Dictionary based approach will use an existing dictionary, which is a collection of

[7] http://www.sananalytics.com/lab/twitter-sentiment/.
[8] 16 http://www.cs.york.ac.uk/semeval-2012/.

http://dx.doi.org/10.1016/j.knosys.2015.06.015

opinion words along with their positive (+ve) or negative (-ve) sentiment strength. In turn, dictionaries were created with/without using ontology. Corpus based approach relies on the probability of occurrence of a sentiment word in conjunction with positive or negative set of words by performing search on very huge amount of texts like Google search, AltaVista search etc.

3.2.1 Polarity determination

Sentiment classification is concerned with determining polarity of a sentence, whether a sentence is expressing positive, negative or neutral sentiment towards the subject. Hence, Sentiment classification is also termed as polarity determination. Polarity determination has been performed for product reviews, forums, blogs, news articles, and micro-blogs. Due to the word limit of 140 words, micro-blogs do not contain full sentences. Moreover, micro-blogs often contain abbreviations and noisy texts. Therefore, it needs high level preprocessing as well as more intelligent techniques for analysis. Micro-blogs has been proved useful for various applications [11, 18, 43, 49, 51, 80, and 131]. Study carried out under polarity determination can be grouped under machine learning based [26, 33, 45, 50, 57, 66, 114, 156, 157, 162, 168], lexicon based [12, 32, 154, 165, 166, 169, 170, 171, 172, 178], and hybrid based [29, 48, 54, 95, 108, 109, 112, 180, 209] approaches. Some of these studies exploited common datasets, which are presented in Table 5 for global comparison.

Table 5. Sentiment classification accuracy reported on common datasets

S#	Dataset	Articles	Obtained result
1	Pang and Lee [167]	[156]	92.70% accuracy
2		[112]	90.45% F_1
3		[169]	90.2% accuracy
4		[35]	89.6% accuracy
5		[54]	87.70% accuracy
6		[46]	87.4% accuracy
7		[50]	86.5% accuracy
8		[26]	85.35% accuracy
9		[162]	81% F_1
10		[124]	79% accuracy & 86% F_1
11		[61]	76.6% accuracy
12		[69]	76.37% accuracy
13		[48]	75% precision
14		[98]	79% precision
15	Pang et al. [33]	[109]	approx. 90% accuracy
16		[165]	88.5% accuracy
17		[172]	87% accuracy
18		[33]	82.9% accuracy
19		[156]	78.08% accuracy
20		[180]	75% accuracy
21		[48]	60% precision
22		[195]	86.04%
23	Blitzer et al. [149]	[45]	84.15% accuracy
24		[99]	80.9% (Avg.) accuracy
25		[54]	85.15% (Avg.) Max. 88.65% accuracy on Kitchen reviews
28		[165]	88.7% accuracy
29		[61]	71.92% accuracy

http://dx.doi.org/10.1016/j.knosys.2015.06.015

3.2.1.1 Machine learning based approaches

Pang and Lee [33] pioneered in applying machine learning viz. NB, Maximum Entropy (ME), and SVM for binary sentiment classification of movie reviews. For experiments, they collected movie reviews from IMDb.com. They experimented with various feature engineering, where SVM yielded the highest accuracy of 82.9% with unigrams features. They claimed that discourse analysis, focus detection, and co-reference resolution could improve the accuracy. McDonald et al. [168] presented a structured model for jointly classifying the sentiment at sentence and document level. They represented a document and its sentences in the form of cliques. They assumed that labels of sentences and documents were interdependent. They applied Viterbi algorithm (a.k.a. linear-chain models) for inference [103]. MIRA algorithm was employed for classification purpose [10, 20]. For feature selection, unigram, bigram, and trigrams conjoined with POS, and their back-offs have been considered. Experiments were performed on 600 online product reviews from amazon.com in 3 different domains (a) car seats for children (b) fitness equipment (c) MP3 player. Accuracy reported as 62.6% at sentence and 82.8% at document level using 10-FCV. The proposed model can be modified for longer than 2 levels in similar fashion for finer level sentiment classification.

Dang et al. [45] classified sentiments using SVM by using different feature selection methods. Experiments were performed on two corpora (a) 305 positive reviews and 307 negative reviews on digital camera and (b) Blitzer et al. [149] multi-domain dataset. SVM was trained on three collections of features set based on domain free, domain dependent, and sentiment features. Information Gain (IG) was applied to reduce the number of features for different combination of features. The reduced features set performed better on multi-domain dataset than digital camera dataset, and yielded an accuracy of 84.15% for kitchen appliance. The proposed feature selection methods should be tested on bigger dataset and compared with other statistical based feature selection method. Saleh et al. [26] carried out twenty seven sentiment classification experiments using SVM with various feature selection methods. Experiments were performed on (a) Pang and Lee [167] dataset, b) Taboada and Grieve [65] multi-domain corpora, and c) SINAI corpus of digital cameras. Using 10-FCV, the best classification accuracy reported were 85.35%, 73.25%, and 91.51% for pang corpus using binary occurrences and trigrams, Taboada corpus using term frequency-inverse document frequency (TF-IDF) and trigrams, SINAI corpus using TFIDF and bigrams respectively. Further experiments can be performed to observe the results affected by rating reviews. Bai [156] projected a Tabu Search-Markov-Blanket Classifier (TS-MBC). Here, MB learns conditional dependencies among the words and encodes them into a Markov Blanket-Directed Acyclic Graph of sentiment words. In the next step, the author applied a Tabu search meta-heuristic strategy to fine-tune the MB-DAG to improve the accuracy. The proposed Tabu search-enhanced Markov blanket model provides a vocabulary in order to extract sentiments. Using Tabu search–MB selects 35 relevant words out of 7716 words in the vocabulary. TS-MBC yielded accuracy of 78.08%, 92.70%, and 73.21% for Pang and Lee [33] dataset, Pang and Lee [167] dataset, and 600 online news articles respectively. Validation is required on the larger data set with more sentiment categories for more fine-grained level SA.

Zhang et al. [114] classified sentiment using machine learning (NB and SVM) for restaurant reviews written in Cantonese. They studied the effects of feature representations and feature size on the classification performance. Experiments were performed on 1500 +ve and 1500 -ve reviews. They experimented with different feature representations like unigram, unigram_freq, bigram, bigram_freq, trigram, and trigram_freq and varying number of features in the range of 50 to 1,600 features. The highest accuracy reported was 95.67% using NB for 900–1100 features. Future work can be integration of automatic review mining technologies to search engines. Tan et al. [162] proposed an automatic approach in deriving polarity pattern rules to detect sentiment polarity at the phrase level. Class sequential rules were utilized to automatically learn the typed dependency patterns. Experiments were performed on Pang and Lee [167] dataset and achieved the highest up to 85.37% average F-measure for adjectival modifier over other feature selection methods. In future, more complex relationships between phrases are to be considered. More in-depth analysis is required on the possible phrase-level influences that would improve overall sentiment polarity. Wang et al. [66] compared the performance of three popular ensemble methods viz. bagging, boosting, and random subspace based on five base learners namely NB, ME, Decision Tree, KNN, and SVM for sentiment classification. They experimented with ten different datasets and reported better accuracy over base learners at the cost of computational time. Ensemble method can be further validated on large dataset, and feature construction based on linguistic method can also be considered. Moraes et al. [50] compared SVM and NB with ANN-based approach for sentiment classification. Experiments were performed on both balanced and unbalanced dataset. Four datasets were chosen for this purpose benchmark movies review dataset [167] and reviews on three distinct product domains: GPS, Books, and Cameras. For unbalanced dataset, performances of both classifiers ANN and SVM were affected. Information gain as feature selection method did not help yield good accuracy for more than 1,000 features. Therefore, these classifiers should be tested on given dataset using different feature selection methods.

Basari et al. [157] formed a hybrid method of Particle Swarm Optimization (PSO) and SVM for sentiment classification of movie reviews. PSO was applied for the selection for the best parameters in order to solve dual optimization problem. Experiments were performed on EMOT[9] dataset and they achieved an accuracy of 76.20% after data cleansing. In future, multi-class sentiment classification and more combinations of feature weighting and n-grams can be tried to improve the classification accuracy. Ghiassi et al. [57] utilized supervised feature reduction using n-grams and statistical analysis to create

[9] Available at: http://cs.stanford.edu/people/alecmgo/trainingandtestdata.zip.

Twitter-specific lexicon for SA. They considered six features viz. information loss, bias, noise, collision, difference in scale, and overfitting. They collected tweets on then the largest Twitter account Justin Bieber. Manual annotations were performed for 3,440 tweets in the scale of 5 values. The proposed system utilized 755 of the most frequently occurring terms and emoticons along with other features like contradictory terms inclusion, synonyms, feature boundary calculation using affinity to include bi-grams and trigrams, negation etc. The Dynamic Artificial Neural Network (DAN2) and SVM employed as multi-class classifiers. DAN2 was designed as to contain multiple hidden layers with four hidden nodes per layer. DAN2 outperformed SVM with an accuracy of 71.3% for strongly positive, 66.7% for mildly positive, 89.9% for mildly negative, and 95.1% for strongly negative. The proposed system can be improved by adopting ontological approach at the creation of the lexicon.

3.2.1.2 Lexicon based approaches

Turney [32] utilized a set of patterns of tags for extracting two-word phrases from reviews. They employed PMI-Information Retrieval (PMI-IR) to determine semantic orientation of review by issuing queries to a search engine, where "excellent" and "poor" have been taken as boundary for positive and negative reference words. Experiments have been carried out on 410 reviews on different domains from Epinions.com. The highest accuracy of 84.0% was achieved on the smallest dataset of 75 reviews on automobiles and the lowest accuracy of 65.83% was achieved on movies reviews. Further enhancement can be made by tagging of sentences on the basis of whole or partial element of discussion. Nasukawa and Yi [171] performed subject favorability determination by creating a sentiment lexicon of 3513 sentiment terms. They considered the syntactic dependencies among the phrases and subject term modifiers. Experiments have been conducted on (a) a multi-domain corpus of 175 cases of subject terms within the contexts, (b) 2000 cases related camera reviews. The proposed system has been evaluated on 552,586 web pages and 230,079 news articles to extract sentiments on an organization by defining 13 subject terms, which yielded precision of 86% and 88% respectively. Sentiment extraction about ten products performed on 476,126 web pages in pharmaceutical domain, and achieved a precision of 91%. The discourse processing, anaphora resolution and indirect expressions are required to be considered to improve the efficiency.

Mullen and Collier [172] proposed a hybrid model comprising PMI-IR approach [32], Osgoodian semantic differentiation with WordNet [9], and topic proximity and syntactic-relation features [171]. Osgoodian semantic gave the potency (strong and weak), activity (active and passive), and the evaluative factor (good or bad) for all adjectives. They experimented with the dataset presented in [33], and 100 reviews on media. The hybrid SVM with PMI/Osgood and lemmas feature selection yielded the highest accuracy of 87% using 10-FCV. Incorporation of domain context to AltaVista search results and dependency relation based feature selection can improve the performance of classification. Whitelaw et

http://dx.doi.org/10.1016/j.knosys.2015.06.015

al. [169] semi-automatically created a lexicon of 1,329 adjectives and modifiers. A detailed semantic analysis of attitude expression was performed by considering appraisal groups. They emphasized the importance of the appraiser (speaker), attitude, target (the appraised), and the orientation. The adjectival appraisal group has been labeled with four features: attitude, orientation, graduation, and polarity. SVM with BOW+G:AO feature selection, out of various feature selection methods, yielded accuracy of 90.2% on Pang and Lee [167] dataset for 10-FCV. Further attempts can be made for accurate identification of appraisal expression including opinion actor (appraiser) and opinion target (appraised).

Wilson et al. [12] identified contextual polarity using machine learning and various sentiment expression based features. They claimed that the prior polarity of the phrase changes with respect to context of discussion. They exploited Multi-Perspective Question Answering (MPQA)[10] opinion corpus to add contextual polarity judgments to sentiment expressions. Experiments performed on a lexicon of over 8,000 subjectivity clues, where 92.8% of them were marked as either positive (33.1%) or negative (59.7%). Twenty eight features were proposed in different categories for neutral-subjectivity classification, which yielded an accuracy of 75.9%. Ten features were proposed in two categories for polarity classification, which yielded an accuracy of 65.7%. Kanayama and Nasukawa [170] utilized two types of contextual coherency like intra-sentential and inter-sentential context in terms of polarity. They extracted lexical entries from the noisy clues. The proposed method was divided into 3 steps (a) sentence delimitation, (b) proposition detection at 'Clause-level', and (c) 3-ary polarity assignment. They considered verb, adjective, argument coincidence with the proposition, and negation. English sentiment lexicon [7] was used to extract various polar atoms. Experiments were performed on digital cameras, movies, mobile phones, and car dataset containing 1757917, 637054, 609072, 959831 sentences respectively. Evaluation was performed on four domain corpora by manually selecting 100 random polar atoms per domain, which yielded a precision of 90% for all four domains.

Benamara et al. [178] composed three Adverb-Adjective Combinations (AAC) scoring methods viz. *variable scoring* for adjective, *adjective priority scoring* for adjective with relevance to adverb, and *adverb first scoring* for adverb with relevance to adjective. AAC was mainly based on the intensity of *adverb of degree* at five linguistic level and assigned score between 0 and 1. For the experiment, annotations have been performed by 10 annotators to 200 BBC news articles. The correlation between the proposed algorithms performance and human subjects was reported. Human subjects correlation yielded r=0.35 for Pearson's coefficient. *Adjective priority scoring* was reported as the best algorithm, which implies that adjective was more important than adverb. More adverb scoring axioms and syntactic constructions can be explored like adverb of time or frequency, adverb and verb combination etc. Lu et al. [166] calculated sentiment polarity strength of a review by multiplying the strength of used adjectives and

[10] http://nrrc.mitre.org/NRRC/publications.htm.

adverbs. The strength of an adjective was calculated using progressive relation rules of adjectives and link analysis (propagation algorithm). A total of 3497 adjective and 100 adverb words from Chinese lexicon were considered to represent progressive relation rules. The strength of the selected adverbs was calculated, which ranges from +1 to -1. Sentiment classification was performed on 2,000 positive and 2,000 negative hotel reviews in Chinese and yielded up to a precision of 71.65%. More experiments are required to ensure the effectiveness the proposed method.

Eirinaki et al. [154] developed a two-step based feature-level opinion mining and ranking algorithm which was deployed in a search engine AskUs [159]. In the first step, High Adjective Count algorithm identifies mostly discussed nouns in the reviews and respective opinion score. The opinion score of each noun was the number of adjectives used along with a noun. In the second step, Max Opinion Score algorithm determined opinion score of each opinion word along with negation by assigning a value in the range of [-4,+4]. And, it also determines score of each feature by summing up opinion scores of utilized opinion words with a feature. The highest scored feature was ranked as the highest and so on. Experiments were performed on Liu [38] dataset and accuracy achieved on vacuum dataset was the best at 97% and the worst on DVD Player was 87%. The proposed method should be compared with some standard lexicon based sentiment calculation approach. Deng et al. [165] devised a supervised term weighting scheme based on *importance of a term in a document* (ITD) and *importance of a term for expressing sentiment* (ITS) with the help of 7 statistical feature selection methods viz. document frequency (DF), information gain (IG), mutual information (MI), odds ratio (OR), chi-square statistic (CHI), Weighted Log Likelihood Ratio (WLLR) and Weighed Frequency and Odds (WFO). They experimented with proposed weighting schemes using SVM on Pang et al. [33] dataset, multi-domain dataset[11], and Maas et al. [39] dataset and achieved accuracy of 88.5%, 88.7%, and 88.0%, respectively. Further accuracy can be improved by combining proposed method with more unsupervised approaches.

Agarwal et al. [177] performed sentiment classification of tweets. They experimented with 11,875 manually annotated tweets. They employed five different combinations of features over unigram, senti-features, and tree kernel. SVM was used for 2-way and 3-way classification tasks. For 2-way classification tasks, unigram+senti-features outperformed all other feature representation with an accuracy of 75.39%. For 3-way classification tasks, tree kernel + senti-features outperformed all other feature representation with an accuracy of 60.83%. In future, richer linguistic analysis can be tried. Taddy [40] came up with an algorithm for maximizing sampling efficiency, which was required in selection of a sub-sample of representative posts for sentiment scoring. He predicted both generic and subject-specific document sentiment through the use of variable interactions in *multinomial inverse regression*. For the experiments, tweets on different USA politicians have been collected during Jan 27 to Feb 28, 2012. They

[11] http://www.cs.jhu.edu/~mdredze/datasets/sentiment/.

http://dx.doi.org/10.1016/j.knosys.2015.06.015

classified collected tweets into positive, negative and neutral category for each politician. Proposed approach can be tried on long text also. Khan et al. [85] applied 3 methods in tandem: Enhanced Emoticon Classifier (EEC), Improved Polarity Classifier (IPC), and SentiWordNet based classifier (SWNC) to classify tweets. For the classification, at first EEC was applied using a total of 145 emoticons (70 positive and 75 negative). Neutral tweets yielded by previous step were given as input to IPC, which uses Opinion Lexicon [38] and Bill McDonald dictionaries. Neutral tweets classified by IPC were further classified using SentiWordNet dictionary. For the experiment, 2,116 tweets have been collected on six personalities to create 6 datasets viz. Imran khan, Nawaz Sharif, Dhoni, Tom Crouise, Pakistan, and America. Proposed method achieved average accuracy of 85.7% over all datasets. In future, validation can be performed based on supervised learning algorithms.

Mostafa [126] analyzed consumers' sentiments towards major worldwide brands such as IBM, Nokia, DHL etc using qualitative and quantitative methodology. He addressed issues like detection of hidden patterns in consumers' sentiments towards Global brands. And, he also assessed the importance of blogosphere for companies to redesign their marketing and advertising campaigns. Relative frequency word counts were generated to get insights and predict characteristics of a particular topic. The proximity graph plotted the frequency of tweets for different brands. Multidimensional scaling (MDS) technique was employed to draw 3-D concept map, which shows the tendency of co-occurrence of phrases or words. 3,516 tweets for sixteen brands from July 18, 2012, to August 17, 2012 have been collected for analysis purpose. To determine the polarity of tweets, the dictionary proposed by Hu and Liu [38] has been utilized. The qualitative part was carried out using QDA Miner 4.0 software package. For future work, the most representative topics discussed behind each sentiment is required. Moreover, extra effort is required to identify vendor's interest towards products and services.

Kontopoulos et al. [96] analyzed sentiment of tweets using ontological engineering. After preprocessing of 667 tweets on mobile, they created domain ontology of tweets using a proposed algorithm. They employed a web service Opendover[12] in order to tag the opinions and get intensity of the sentiment expression. Evaluations were performed on two versions of the proposed architecture (a) the full-fledged ontology-based semantically-enabled system (SEM), and (b) the same system without synonym/hyponym augmentation. The former has yielded better recall and the degree of synchronization. In future, Opendover can be replaced by some advanced algorithmic approach for sentiment determination of each tweet. Bell et al. [203] exploited twitter as a platform to instruct an educational tool called Finch Robot to take a picture and reading temperature, where both instructions could be given in any order. Tweets were preprocessed and POS tagged using tweet tagger [204]. They devised 15 rules on the basis of POS to extract different events, which were verified using ConceptNet. Evaluation of the

[12] OpenDover sentiment tagging web service: http://opendover.nl/.

developed system was carried out by a total of 11 participants. And authors reported efficiency, effectiveness, and satisfaction on the basis of users' experiences. In future, the proposed architecture can be modified in order to carry out some other works like vacuum cleaning robot etc. And additional information like spatial, temporal, physical, social, and psychological information can be exploited for better operations.

Popescu & Strapparava [206] studied diachronic phenomena of two different domains namely socio-political and sports on Google N-grams corpora. Analysis was performed 761 and 34 words from Socio-political domain and sport respectively. Epoch delimitation was performed on the basis of word distribution over certain periods of time. They also analyzed the opinion change phenomenon using the covariance between the frequencies of two or more terms over a certain period of time. Eight types of emotions were decided for 14000 words using WNA based NRC Word-Emotion Association Lexicon (WNANRC) and Semeval 2007 Affective Text (SAT). The proposed methodology can be extended to predict future changes in society like the covariance between socialism and capitalism etc.

3.2.1.3 Hybrid approaches

Sindhwani and Melville [180] developed an unsupervised and a semi-supervised sentiment classification algorithm. For unsupervised method, they utilized a lexicon containing human-labeled 2,968 words to achieve domain adaptability. For semi-supervised lexical classification, they utilized bipartite representation of the data i.e. document-word bipartite and regularization operator graph. Due to sparseness of data, they used Regularized Least Squares (RLS) classification algorithm. They incorporated prior knowledge of sentiment-laden terms directly into the model using Lexical-RLS. Generality of the approach have been justified by experimenting on three, very different, domains (a) Pang et al. [33], (b) lotus blogs, and (c) political blogs using 10-FCV. They compared the lexical RLS and semi-supervised lexical RLS with three baseline methods and performed well. More feature extraction methods along with non-linear kernel based classifier can be tested to improve the performance. Narayanan et al. [29] performed linguistic analysis of conditional sentences and sentiment classification using supervised learning. Different linguistic features viz. opinion words, POS tags of sentiment words, words without opinion, tense patterns, special characters, conditional connectives, length of condition and consequent clauses, and negation words were exploited to determine the orientation of a sentence. SVM with Gaussian kernel yielded F-score 75.6% by training on manually annotated 1,378 sentences taken from forums on cellphone, automobile, LCD TV, audio systems, and medicine. More rigorous experiments should be performed to ensure the efficiency of the proposed system.

Prabowo and Thelwall [112] developed different hybrid classifiers over five classifiers viz. a) General Inquirer based classifier (GIBC) [105] (b) Rule based classifier (RBC) (c) Statistics based classifier (SBC) (d) Induction rule-based classifier (IRBC) (like ID3 and Ripper), and e) SVM.

Experiments were performed on Pang and Lee [167][13], Movie Review-2 (100 +ve and 100 -ve reviews from previous dataset), product reviews (180 +ve and 180 -ve), MySpace comments (110 +ve and 110 -ve). They reported the highest micro averaged F_1 90.45% by applying RBC, SBC, GIBC, SVM in tandem on MySpace comments. In future, the performance of the proposed hybrid classifier should be tried on bigger and complex data set. Abbasi et al. [109] framed a feature relation network (FRN), which was rule-based, multi-variate, n-gram feature selection technique. It incorporates semantic information derived from existing lexical resources, enabling augmented ranking of n-gram features. Sentiment classification has been performed on 2000 reviews on digital cameras from epinions.com, 2000 reviews on automobiles from edmunds.com, and Pang et al. [33] dataset. FRN outperformed log-likelihood ratio (LL), chi-square, word n-grams/LL, and bag of words/LL for all three test beds with considerably better accuracy on virtually all feature subset sizes. A lot of scope is there for future improvement like augmenting FRN with occurrence frequency and positional/distributional feature, and incorporating FRN in conjunction with other multivariate selection techniques.

Xia et al. [54] ensembled feature sets and machine learning for sentiment classification. Two types of feature sets namely "POS based features" and "the world-relation based feature sets" have been designed. These feature selection methods were ensembled with NB, ME, and SVM using 3 techniques viz. fixed combination, weighted combination and meta-classifier combination. Word relation based weighted classifier yielded accuracy of 87.7% and average 85.15% on Pang and Lee [167] and Blitzer et al. [149] dataset respectively. Feature selection based on syntactic relations can improve accuracy. Chen et al. [48] classified sentiment of blogs on products. Classification was performed using hybrid of semantic orientation and back-propagation neural network. They proposed 4 different types of semantic orientation (SO) indexes as the input neurons viz. *SO-PMI (AND), SO-PMI (NEAR), semantic association, and latent semantic analysis*. Experiments were performed on Blogs (data size: 353, +ve: 157, -ve: 186), MP3 (Data size: 579 +ve: 235, -ve: 344), EC (Data size: 386 -ve: 249 -ve: 137), Movie-001, and Movie-002 (+ve: 500, -ve: 500) and yielded 87.5%, 73.5%, 81.2%, 75%, and 67.8% recall respectively. The proposed approach should be tested on different corpora like Twitter, Plurk, Facebook etc employing different machine learning techniques.

Balahur [95] extended her own proposed method to detect emotion using commonsense knowledge. For the experiment, 1,081 examples from ISEAR have been selected for the test set A, and 895 examples were selected out of 1,081 for test set B. She applied different combination of supervised Sequential Minimal Optimization-SVM technique and linguistic features. The proposed EmotiNet based technique outperformed some other supervised and lexical-knowledge based techniques. Future work was claimed as to extend of EmotiNet, test new methods to assign affective value to the concepts, and expand

[13]Available at: http://www.cs.cornell.edu/people/pabo/movie-review-data/.

EmotiNet to other language and domain as well. Abdul-Mageed et al. [108] performed subjectivity and sentiment analysis (SSA) of social media for a morphologically-rich language. Experiments were performed on 2,798 chat turns, 3,015 Arabic tweets, 3,008 sentences from 30 modern standard Arabic Wikipedia Talk pages, and 3,097 web forum sentences. For the subjectivity analysis, each of the sentences of collected data sets was labeled manually for the subjectivity and objectivity. 3,982 Arabic adjectives were collected for binary polarity classification of the sentences. SVM outperformed baseline methods for subjectivity classification, and yielded an accuracy of 73% for tweets and 84.36% for forums. SVM outperformed baseline method for sentiment classification with an accuracy of 70.30% for chat turns. Error analysis and irony detection can play significant role in the improvement of accuracy.

Ortigosa et al. [209] performed sentiment classification and sentiment change detection on Facebook comments using lexicon and machine learning based approach. They developed a sentiment lexicon on the basis of Spanish Linguistic Inquiry and Word Count (LIWC) and slang found in the comments. In order to evaluate the proposed lexicon, they employed C4.5, NB, and SVM to classify 3000 status messages (1000 for each class: positive, negative and neutral) and yielded an accuracy of 83.17, 83.13, and 83.27% respectively. Future improvement would be to calculate contextual sentiment score.

Jiang et al. [176] proposed a target-dependent and contextual based approach to perform sentiment classification of tweets. Subjectivity and sentiment classification was performed using SVM. The graph-based optimization was applied on related tweets to boost the performance. PMI was used to identify top-k nouns and noun phrases associated with the target. For the experiment, they collected 400 English tweets on each of the query like Obama, Google, iPad, Lakers, Lady Gaga and achieved accuracy up to 68.2% for subjectivity classification and up to 85.6% for sentiment classification. In future, Twitter's user profile based sentiment classification can be undertaken. Kouloumpis et al. [179] classified sentiment expressed in tweets. They used a list of features viz. 1000 n-gram, MPQA subjectivity lexicon, POS, micro-blogging emoticons, and abbreviations features. Adaboost.MH was trained on HASH dataset containing 222,570 tweets, and EMOT dataset containing 381,381 tweets and emoticons. Testing was performed on 4,015 tweets from HASH+EMOT dataset. The micro-blogging feature, n-grams+lex+twit, was reported as the best feature combination and yielded average accuracy of 75.0% on test dataset.

Li and Xu [123] enacted a novel method to classify emotions using the emotions cause extraction technique. The emotion cause extraction helped in removal of unnecessary features. Further chi-square method was employed to remove irrelevant features. A list of 1845 emotion words and short phrases has been collected to identify emotion words from tweets. To test the proposed lexicon, 16,485 posts from Sina-Wiebo have been collected. Support vector regression (SVR) with linear kernel was utilized for multi-class classification and yielded a precision of 75.70%. The obtained precision can be improved by considering more linguistic patterns, elements and factors. Automatic generation and modification of the pattern set will reduce the manual effort. Rill et al. [205] proposed early detection of twitter trend, which was faster than Google Trends. They considered temporal changes in the number of tweets to decide the emerging topic. Polarity of tweets were decided using sentiment lexica like SenticNet3[14], SWN etc., where polarity of novel words were determined by plotting a relational graph of emerging political tweets at different time periods. They collected 4,000,000 tweets during Germany parliamentary election on nine political parties to test the proposed system. In future emerging patterns of controversial topics can be explored.

3.2.2 Vagueness resolution in opinionated text

Ambiguity and vagueness have been considered as major issues since user reviews are often written using a loose style than standard texts, and often express sarcasm (mock or convey or irony), rhetoric or metaphor. Political discussion and extreme often include irony and sarcastic words; detection of such expression is a challenging task in opinion mining area. The fuzzy approach is quite useful to represent such feelings and expressions. In this regard, some initiatives taken in sentiment analysis area are summarized in this section.

Li and Tsai [41] framed a classification framework on the basis of fuzzy formal concept analysis (FFCA). For feature representation and dimensionality reduction, TF-IDF, Inverted Conformity Frequency (ICF) and Uniformity (Uni), and the relation between documents and terms have been considered using context matrices. The attribute lattice set was represented using terms of the document. The TF-IDF value represented the degree of significance of each term in the lattice object set, which were given to the classifier. Experiments were performed on (a) Reuters-21578 containing 21,578 economic news stories, (b) Pang and Lee [167] dataset, (c) Kindle eBook reviews (1000 +ve and 1000 -ve). Fuzzy Formal Concept analysis Model (FFCM) outperformed J48, k-NN, SVM and BN, and reported a precision of 96.20%, 88.75%, 95.09% on the three datasets respectively. Topics such as Multi-class classification, novel similarity measures, and computational complexity of FFCM are the future research directions.

Reyes and Rosso [86] composed six notable features to deal with verbal irony. Those features were n-grams, POS n-grams, funny profiling, +ve/-ve profiling, affective profiling, and pleasant profiling. Among these features, first and second were common for preprocessing purpose. The third feature gives humorous property of the sentence. The fourth one deals with SA and utilized the Macquarie Semantic Orientation Lexicon (MSOL) [139]. Fifth feature uses WordNet-Affect (WNA) to categorize affective content into 11 different classes. And, final one assigns a score of pleasantness to approximately 9000 English words. For the experiments, 3163 ironic reviews on five products from Amazon.com and three

[14] http://sentic.net/senticnet-3.0.zip.

http://dx.doi.org/10.1016/j.knosys.2015.06.015

negative sets containing 3000 document each from Amazon.com (AMA), Slashdot.com (SLA), and TripAdvisor.com (TRI) have been collected. For the classification NB, SVM, and DT were trained on 5861 instances (2861 +ve and 3000 -ve) using 10-FCV. SVM classified AMA with best performance 75.75%, NB classified SLA with best performance 75.19%, and DT classified TRI with best performance 89.05%. In future, more rigorous validation of result is required. Bosco et al. [22] developed a corpus, Senti-TUT, of irony words used in political discussion in Italy. The proposed corpora exploited two twitter corpora TWNews (3,288 posts) and TWSpino (1,159 posts) of Italian political discussions during 6 October 2001 to 3 February 2012 for Mario Monti nominations. Each tweet was annotated one of the following sentiment tags: positive, negative, humor, mixed (positive and negative both), none (objective). Five annotators employed to annotate tweets and inter-annotator agreement was determined by Cohen's k score=0.65. To apply polarity reverser, 723 ironic tweets from TWNews have been classified by human annotator and Blogmeter classifier (BC)[15]. BC is a rule-based approach which relies on a sentiment lexicon and sentiment grammar expressed by compositional rules. To determine emotion level of tweets, BC was employed to annotate ironic tweets into six different categories. Future work is required to consider a formal account and a measure of these phenomena, and a finer granularity of text analysis.

Justo et al. [217] proposed to classify sarcasm and nastiness available in dialogic language on the web. They integrated statistic, linguistic, semantic and emotional features to train a RBC and NB classifier. Statistic features were dictated by n-grams. The linguistic features involved word and character counts etc. Semantic and emotional features were obtained by SenticNet 3.0. Chi-square feature selection was employed to select most prominent features. Experiments were carried out on two subsets of the IAC[16] sarcasm data set consisting of 6,461 instances and nastiness data set consisting of 2765 instances balanced between nasty and not nasty posts and achieved an F-measure of 70.7% and 79.7% respectively. The proposed approach can be enhanced in order to achieve better F-measure. Weichselbraun et al. [216] proposed context based polarity ambiguity removal. They employed ConceptNet based vector space similarity, and WordNet based graph-based similarity to determine the polarity of a concept. They evaluated the proposed system on 2000 reviews on each of Amazon electronics, Amazon software, IMDb comedy, IMDb crime, and IMDb drama, which outperformed baseline methods. The proposed system could be evaluated on larger dataset and can be developed at sentence level.

3.2.3 Multi-lingual and cross-lingual sentiment analysis

Different languages across the world have different degree of expressive power regarding sentiments. An important body of research has tried to address this issue at some extent. Although, higher accuracy is not achieved in this regard, unexplored corners of several approaches can positively influence the accuracy.

[15] www.blogmeter.eu.
[16] https://nlds.soe.ucsc.edu/iac.

http://dx.doi.org/10.1016/j.knosys.2015.06.015

Cross-lingual sentiment analysis can be performed using two different approaches: (a) Lexicon-based approach, where a target-language subjectivity classifier is generated by translating an existing lexicon into another idiom; (b) Corpus-based approach, where a subjectivity-annotated corpus for the target language is built through projection, training a statistical classifier on the resulting corpus [82]. For multi-lingual and cross-lingual sentiment analysis, following study is divided into two categories: machine learning [46, 88, 115] and hybrid approach [94, 148, 173].

Hiroshi et al. [173] developed Japanese Sentiment Analysis system (JSA) to extract sentiment units from Japanese corpus. JSA system performs Japanese to English translation using transfer-based machine translation engine containing 3 parts: (a) a source language syntactic parser, (b) a bilingual transfer which handles syntactic tree structures, and (c) a target language generator. They also employed full syntactic parser and top-down pattern matching. Opinion word sense disambiguation in machine translation and aggregation was performed to extract frequently mentioned opinions. Experiments were conducted on 200 digital camera review sentences to extract sentiment units. Weak and strong precision were claimed as 100% and 89% respectively by the proposed system. The proposed system outperformed lexicon only system for sentiment unit extraction and measurement for the appropriateness of the sentiment scope. Some other machine translation, such as word sense disambiguation, anaphora resolution, and automatic pattern extraction etc. can improve the efficiency of the system. Boiy and Moens [46] performed multilingual and multiple domain sentiment classification. They utilized different feature representations like unigrams, negation, discourse features, compound words and verbs etc. They utilized cascaded approach with 3 single classifiers viz. MNB, ME, SVM for experiments using various combination. They reported the best results using MNB for English, SVM (linear kernel) for Dutch, and ME for French. For experiments, they collected 750 positive sentences and 750 negative sentences for each language; evaluation was performed on Pang et al. [167] dataset and reported up to accuracy of 87.40%. Irony and sarcasm detection will be the major issue for cross-lingual sentiment classification.

Emirtas [148] performed cross-lingual SA over English and Turkish language by applying machine translation techniques from Turkish to English using Google Translator API. Experiments were performed on Pang and Lee [198], Blitzer et al. [149], Turkish movie reviews and Turkish multi-domain product reviews [148]. NB, ME and SMO-SVM were used for classification purpose. ME yielded the best performance for Blitzer et al. [149] dataset. Martin-Valdivia et al. [94] carried out Spanish sentiment classification using a hybrid meta-classifier. Spanish MuchoCine (MC)[17] corpus was automatically translated into MuchoCine English (MCE) corpus. MC contains 3,878 movie reviews collected from the MuchoCine website. For supervised learning, SVM and NB were utilized for both corpora. As an unsupervised approach, SentiWordNet was used for classification of MCE corpus. Finally, both

[17]http://www.muchocine.net.

approaches were combined to get meta-classifier. TF-IDF, Term Frequency (TF), Term Occurrence (TO), and Binary Occurrence (BO) were considered as feature representation schemes. SVM outperformed NB for both corpora. TF-IDF was reported as better representation scheme. SVM using TF-IDF without stopper and stemmer yielded the best precision 0.8773 and 0.8776 on MC and MCE respectively. Stacking ensembling yielded a precision of 88.58%. In future, sentiment classification of MCE can be performed using other emotion rich dictionaries like WordNet-Affect, SenticNet etc.

Seki et al. [88] presented multilingual opinion holder identification using author and authority viewpoints. They employed three annotators to annotate opinionated sentences and opinion holder with three opinion types. They experimented with four sample topics of NTCIR-6 Opinion Corpus for Japanese data and MPQA for English data. Classification was performed using SVM for Japanese corpus, and SVM, Conditional Random Field (CRF), and Logistic Regression (LR) for English corpora. For Japanese corpora, SVM yielded 60.9% recall after performing named entity extraction. The proposed method can be applied for opinion holder identification from multilingual blogs. Wang et al. [115] classified sentence level subjectivity at the first step. At the next step, sentiment score was calculated using three techniques viz. *equal weights, correlation degree,* and *sentiment conditional probability*. Finally, sentence level sentiment aggregation was performed over sentences using weighted average. Experiments were performed in four aspects viz. sentiment classification at sentence level, determining the importance of a sentence to a document, sentiment classification of a document, and comparative experiment of document-level sentiment classification. Information Gain (IG) was applied to select initial 300 text features. SVM was utilized for sentence level classifier. Selected dataset were 4000 Chinese reviews on hotel, 4000 Chinese and 4000 English reviews on cell phone. The accuracy of sentiment classification at document level declines from English to Chinese online reviews and from cell phone to hotel reviews. Removal of manual treatment of negation, and strengthening and weakening of sentiment due to exclamation symbols are issues to be addressed in future work.

3.2.4 Cross-domain sentiment classification

Of late, cross-domain sentiment analysis became an interesting research problem to work upon. Due to high variation of subjectivity across domains, it is a challenging task. Cross-domain requires at least two domains: source domain on which a classifier is to be trained on, and target domain on which testing is to be performed. Work carried out in this area can be classified into two groups; the first group requires initial training set from source domain as well as from target domain [121]. The learners in the second group of study are trained on source domain and tested on target domain [98, 99]. These studies were carried out using lexicon based [36, 98], machine learning based [121], and hybrid approaches [99].

Tan et al. [121] applied supervised learning approach for cross-domain sentiment classification. They proposed an effective measure i.e. Frequently Co-occurring Entropy (FCE), to select generalizable features that occur frequently in both old-domain data and the unlabeled new-domain data. They employed weighted expectation-maximization based Adapted Naïve Bayes (ANB) to train a classification model for the new domain. For the experiment, education reviews (1012 -ve and 254 +ve), stock reviews (683 -ve and 364 +ve) and computer reviews (390 -ve and 544 +ve) were considered. The proposed approach outperformed NB, Expectations Maximization-NB (EM-NB), and Naive Bayes transfer classifier and yielded average 82.62% micro F_1 (i.e. for common categories) and 79.26% macro F_1 (i.e. for rare categories). FCE can be replaced with other techniques to pick out better generalizable features. Weichselbraun et al. [98] created contextualized, cross-domain lexicons that can be integrated into a wide range of opinion mining and decision support applications. Disambiguation and contextualization have been considered to get better result on cross-domain SA. Ontological lexicon was created in three phases (a) ambiguous term detection with the help of existing pre-labeled corpora, (b) calculating sentiment score based on the probabilities of co-occurring contextual terms, and finally (c) SA by combining polarity values for unambiguous and ambiguous terms. Evaluation was performed on 2500 reviews from Amazon.com, 1800 hotel reviews from TripAdvisor.com and Pang and Lee [167] movie review dataset and yielded precision of 76.5%, 82%, 79% respectively. Extension of the proposed lexicon was required by using grounded concepts from SenticNet [106], ConceptNet [120], Freebase [128], DBPedia [129] etc.

Bollegala et al. [99] proposed automatic sentiment sensitive thesaurus creation and feature expansion for cross-domain sentiment classification. Novelty of the proposed method was the exploitation of the created thesaurus to expand feature vectors at training and testing part of a binary classifier. Sentiment score was calculated based on PMI between a sentiment element and feature vectors. Experiments were performed on the Blitzer et al. [149] dataset and yielded average accuracy of 80.9%. For each domain, a sentiment sensitive thesaurus was created using labeled data from source domain and unlabeled data from source and target domains. Each thesaurus was utilized to expand the labeled feature vectors from the source domains and train an L1 regularized logistic regression-based binary classifier (Classias[18]). These four thesauri were compared against three baseline methods and yielded better performance over all methods. Future work needs to ensure wide applicability in other domains as well.

Cho et al. [36] suggested cross-domain sentiment classification by integrating multiple sentiment dictionaries viz. WordNet-Affect [4], SentiWordNet [5, 6], WordNet [133], Opinion Lexicon (OL) [38], AFINN [55], SO-CAL [69], Subjectivity Lexicon [104], General Inquirer [105], SenticNet [106], SentiSense [113], and Micro-WNOp [119] together. These were exploited either after removal of some entries or shifting polarity of some of the entries according to a given domain. They proposed an algorithm to remove and/or shift polarity of a sentiment word. For the experiments, 17,500, 35,000, and

[18] Available at: http://www.chokkan.org/software/classias/.

90,000 reviews were collected from Amazon.com for smartphones, movies, and books to build a positive/negative review dataset. They achieved 82.6%, 80.1%, and 81.8% accuracies for smartphone, movie, and book reviews, respectively. The proposed approach can be utilized to create a custom dictionary, which may yield promising result for cross-domain sentiment classification.

3.3 Review usefulness measurement

In information age, market managers are becoming more interested in promotion of their products and services. In order to promote their products some of them are hiring fake reviewers to write fake reviews [137, 181]. Although these fake reviews do not have long time effect on product sale, it can amplify initial sale. Thus, opinion spam detection and review usefulness measurement gained so much attention by research communities [181]. These two sub-tasks of SA sound similar but they are the two different side of the same coin. In most of the cases, a review spam usually concerns a good quality review, while a bad quality review need not tend to be a review spam. This is because on one hand in the most of the cases, a review spam is written very intelligently either to boast the product or discredit the product. On the other hand a bad quality review is written by an honest consumer. Opinion spam detection is briefly presented in the next section. The quality of reviews is discussed briefly in [181]. Therefore, we considered some of recent advancements made by existing study for review usefulness measurement. The study followed machine learning [76, 81, 222, 223, 224, 225, 226, 227, and 228] and hybrid approaches [78, 130, 221, and 229].

Ghose et al. [81] identified several features to measure helpfulness of a review like subjectivity levels, various measures of readability and extent of spelling errors to identify important text-based features. Some more features related to multiple reviewer features like average usefulness of past reviews, the self-disclosed identity measures of reviewers have been considered. Observation was that subjectivity, informativeness, readability, and linguistic correctness in reviews matter in influencing sales and perceived usefulness. Experiments have been performed on product reviews collected over 15 months from Amazon.com on audio and video player, digital camera, and DVD. SVM and Random forest (RF) were utilized for predicting review helpfulness, where RF outperformed SVM in all cases. They concluded that only subjective or objective reviews have good impact on product sales in comparison to highly subjective or highly objective reviews. In future, helpfulness can be calculated on the basis other factors like considering type of users and the context etc. Liu et al. [76] developed an algorithm in two phases to measure the helpfulness of a review. In the first phase, product feature extraction was performed using PMI based document profile model. For the helpfulness prediction, bootstrap aggregating algorithm along with decision tree outperformed MLP, simple linear regression (SLR), and sequential minimal optimization-support vector regression (SMOreg). In second phase, principal component analysis (PCA), feature-instance similarity, and mutual information (MI) based feature selection methods were applied to extract generic features to determine helpfulness. Experiments were performed on 1000 reviews collected from Amazon.com on mobile phone, and achieved MAE 0.599 for training in the first phase, and MI and PCA outperformed for 3 datasets. Further experiments can be carried out on new domain.

Racherla et al. [222] established that review helpfulness was correlated with a combination of review and reviewer's characteristics. Credibility of a review was proven by establishing linear relationship among reviewer's identity, expertise, and reputation along with review elaborateness, review valence, and review helpfulness using ordinary least square regression method. They experimented with 3000 reviews on three different service categories taken from www.yelp.com. Consequently, most of the characteristics listed above were found to be effective except reviewers' identity. Future study can include brand, cost of services, number of review reads, reviews from different reviews site and category etc. to determine the credibility of a review. Mudambi and Schuff [223] applied regression model to measure helpfulness of a set of reviews. They considered type of products viz. experience or search, number of votes to a review, number of people found a review to be helpful, number of stars, and word count. Experiments were carried out on 1608 reviews on six product items taken from www.amazon.com. They reported that product type affects rating and review length. Moreover, review length positively affects the helpfulness of a review. This study is further extended by Huang & Yen [224] and proposed two modified regression equations. They experimented with 2209 reviews in the same category of products but slightly new version of products. The proposed method yielded up to 15.7% of variation in review helpfulness. In future, reviewer’s information, product metadata and subjectivity can be considered to improve the performance.

Chen and Tseng [225] proposed information quality framework on the basis of nine dimensions of review features. The prominent dimensions included believability, objectivity, reputation, relevancy, timeliness, completeness, appropriate amount of information, ease of understanding, and concise representation. These nine review dimensions gave fifty-one features in order to classify useful and non-useful reviews using SVM. On the basis of proposed method, they developed a quality-based review retrieval system. The proposed system was used to extract top 150 reviews on each of ten categories of digital camera and mp3 players. Evaluation of the result was carried out by employing 3 human experts. The proposed framework can be extended to other domain like forums, blogs etc. as well. Lee & Choeh [226] employed MLP to measure helpfulness of a review, which outperformed the linear regression analysis. Required features considered for the model were: product related information, linguistic features, and metadata of a review. They predicted helpfulness of 28,299 reviews on different products taken from www.amazon.com using in total of 20 features. Training and testing have been performed 12 times on 12 different test sets using MLP yielding the least mean-squared error of 0.122. Statistical significance test conducted using Wilcoxon test indicated that the results of different test sets are different. In future, temporal information about review can be incorporated to improve the effectiveness of the approach.

Ngo-Ye & Sinha [227] employed vector space model and RFM (Recency, Frequency, and Monetary Value) score to predict helpfulness of a review. In RFM, recency referred to the difference between current review post date and last review post date, frequency represented the number of reviews written by a reviewer, and monetary value was calculated on the basis of average number of helpful votes a reviewer received on his/her entire written reviews. In total 28 different types of models have been proposed and experimented with 584 Amazon book reviews and 7465 Yelp restaurant reviews. Out of 28 models, modified bag of words (BOW')+RFM score based model yielded the best helpfulness prediction using SVR model. Here, BOW' represented subset of features obtained by applying correlation-based feature selection on the features obtained in document-term matrix using TF-IDF. And RFM score gave three features in the form of Recency, Frequency, and Monetory value. The study suggested more reviewers' information to be included for better helpfulness prediction and proposed method should be tested on various other domains. Krishnamoorthy [228] developed a predictive model to measure helpfulness of a review. He mainly considered linguistic features, review metadata, readability and subjectivity features to estimate helpfulness. Among linguistic features like adjectives, state verbs, action verbs, he emphasized action verbs due to its concreteness, verifiability, indisputability, and being informative. They experimented with the dataset of Blitzer et al. [149] and 1653 reviews on different products reporting an F-measure of 87.21% and 81.33% respectively using RF. RF outperformed NB and SVM in this study. He compared different categories of features involving linguistic, review metadata, readability and subjectivity features and found linguistic features to the best. Study indicated several interesting future directions to work further.

Hu et al. [78] devised two discretionary accruals based models, one for helpfulness measurement and other for validation. The model considered 7 variables to decide the helpfulness of a review. Considered variables were viz. bestseller dummy, time dummy, mean and variance helpful vote dummy, high-average-helpful vote dummy, high-variance-helpful vote dummy, popularity, and price. Experiments were performed on 100 consumer ratings on 1851 books from amazon.com during July, 2005. With the help of the second model, they found that the bestsellers were less likely to engage in review manipulation, and review manipulation decrease over time. Future studies require observing the behavior of review manipulator over time and across different categories. Zhang [130] aggregated online product reviews by weighing stars in two orthogonal dimensions: polarity extraction, and usefulness scoring using regression analysis. He utilized 3 features (a) lexical similarity features, (b) shallow syntactic features, and (c) lexical subjectivity clues to identify useful reviews and usefulness scoring. In order to measure the lexical similarity between customer reviews and product specification, cosine similarity with TF-IDF

http://dx.doi.org/10.1016/j.knosys.2015.06.015

were utilized. Shallow syntactic features and lexical subjectivity clues have been retrieved using 4 different existing works by Hatzivassiloglou [31], Wiebe [87], Rilo et al. [153] i.e. MetaBoot algorithm, and Thelen et al. [83] i.e. Basilisk algorithm. For the experiment, Amazon Web Services[19] was used to collect reviews on three different domains: electronics (Sony and Canon), video, and books. The author assessed the influence of the review author's profile using analysis of variance between the usefulness score. They inferred that the full-regression model (ε-SVR and ν-SVR algorithms) outperform a uni-variate polarity-based model. The regression models significantly outperform the length-based and rating-based baselines. The proposed model can be applied for different text genres like blogs and forums on events, public figures, and social movements etc.

Min et al. [221] argued that a review written by an experienced customer was more important than a professional reviewer. In order to measure the helpfulness of a review they considered the duration of product use, the number of products used from the same brand, and temporal detailed description about product use. They extracted the duration of product use using TERN evaluation[20] and time tagger[21]. In addition to that discussed entities are extracted using shallow parsing on the basis of deictic expressions or time expression. Experiments are carried out on the dataset of Blitzer et al. [149] and 3345 reviews on apparel and beauty from amazon.com. They reported F_1 up to 88.58% for entity detection and an accuracy of 83.33% for identifying product-referring terms using RBC, where RBC outperformed DT and SVM. Evaluation was performed by employing eight human evaluators to rate the helpfulness of 96 reviews. Assumption needs to be more rigorously evaluated on larger set and some other user mentioned cues can be explored.

Purnawirawan et al. [229] purposed to use the balance (the ratio of positive and negative reviews) and sequence (the order in which the reviews are presented) to measure the usefulness of a review set. They further investigated various effects of review usefulness on review readers. Regarding balance, they argued that either positively or negatively inclined reviews were more useful compared to neutral reviews. In terms of sequence, they found that wrapping one polarity based reviews into other polarity based reviews was more useful. Experiments were carried out on 8 positive and 8 negative reviews on 3 (Balance: positive, neutral, negative) × 4 (Sequence: positive/negative, negative/positive, positive/negative/positive, negative/positive/negative). The proposed method needs to be extended over larger corpora and different domains in future.

[19] http://aws.amazon.com/ecs.
[20] The Time Expression Recognition and Normalization, 2004 http://www.itl.nist.-gov/iad/mig/tests/ace/2004/.
[21] http://fofoca.mitre.org/taggers/timex2_taggers.html.

http://dx.doi.org/10.1016/j.knosys.2015.06.015

3.4 Opinion spam detection

With the growing popularity of e-commerce and online reviews, an individual used to actively engage professionals in writing false reviews. Here, an individual may be marketing personnel, a manufacturer, a service provider, a leader, or a movie producer etc. And a false review can also be referred as a fake review, an opinion spam, a fraudulent review or a non-genuine review. The person who writes a fake review is called as a spammer or fake reviewer. If a group of people get involved in the process, then they are referred to as group spammer [199]. A spammer used to post fake reviews either to promote a low quality product or to discredit a good quality product.

Most of the opinion spam detection technique depends on three types of features related to a fake review, which includes content of review, meta-data of review, and real-life knowledge about the product [181]. The first and foremost source of information of spam detection is the text written in a review. The text can be analyzed using natural language processing and machine learning techniques in order to uncover deception and lies hidden in the text [232]. Meta-data of a review includes various attributes viz. user-id, IP address, geo-location, date and time of writing, number of stars etc [231]. But in the most of the cases all required meta-data is not easily accessible to spam detector, which makes review detection a challenging task. Third attribute i.e. real-life knowledge about the product can play a major role in identifying an opinion spam. For instance, a particular brand has very good reputation and an inferior brand has been shown as superior to that in some reviews posted during specific period, in such cases a review can be suspected as a fake review.

Opinion spam detection need to be performed for review spam detection, spammer detection and group spammer detection [199]. Review spam detection is the process of determining whether a review is written by an authorized buyer with honest intention or not. Next, Spammer detection identifies an un-authorized buyer, who wrote reviews on products or services without experiencing them. And, group spammer detection discovers whether more than one person is involved in spamming either from same or different geographical location. Heydari et al. [199] reviewed a host of methods proposed for the three types of opinion spam detection. Some promising review spam detection methods included duplicate finding methods [234], concept similarity based method [235], content based method [200, 210], and review and reviewer oriented features based method [236] etc. Spammer detection techniques included graph-based method [237], temporal activity-based method [238] etc. Graph based method builds a graph of review, reviewer and product in order to capture trustworthiness of the reviewer, the honesty of review and the reliability of the product, sentiment of the review etc. Group spammer detection technique involved pattern mining based method [239]. Apart from these techniques, some other recently proposed studies are broadly classified into machine learning based approaches [200, 216, and 232] and hybrid

http://dx.doi.org/10.1016/j.knosys.2015.06.015

approaches [212, 220, and 231]. Interested reader can refer to Ref. [137, 181, and 199] for detailed reviews on opinion spam detection.

Ott et al. [212] performed opinion spam detection on the basis of artificially developed positive review spam dataset. The dataset contained 400 positive truthful reviews which are taken from TripAdvisor.com and having rating star 5. Amazon's Mechanical Turkers were employed to write 400 positive deceptive reviews. In order to detect a spam, a) they performed genre identification based on POS tag, b) psycholinguistic deception detection using LIWC software, and c) NB and SVM based text categorization using different n-gram based representation. LIWC+Bigram representation based SVM yielded the best F-measure of 89.8% for both truthful and deceptive classification. They observed several relationships between emotion and deceptive reviews. In other study, Ott et al. [200] developed a negative deceptive opinion dataset and performed spam classification using SVM. The dataset contained 400 negative truthful and 400 negative deceptive reviews. Truthful reviews were collected from 6 review sites having rating stars 1 or 2. Deceptive reviews were written by workers employed by AMT service, where each review is written by different worker. So, the deceptive dataset contained 20 reviews on each of the 20 hotels situated at Chicago. They experimented with the developed 800 negative reviews and 800 positive reviews taken from Ott et al. [212] on unigram and bigram term-frequency representation. Linear SVM yielded an F-measure of 86.1% and 89.3% respectively under 5-FCV framework. They explored relationships between sentiment and deception. Moreover, they found to fake negative reviews were linguistically different from truthful and fake positive reviews in terms of first person singular use, verbs use etc. At the beginning of a launched product item review manipulation occurs more frequently and it decreases over time [78].

Hu et al. [231] proved that review manipulation is a monotonically decreasing function of the product's true quality or the mean consumer rating of that product using regression. Regression included some important features of a review like age of the previous review, age difference between two successive reviews, average rating, popularity, constant to control the product category etc. They experimented with reviews collected from Amazon and Barners & Noble on books, DVD, and videos collected over several months at approximately three-day intervals. The study mainly relied on rating based features and content based features can be considered in future. Banerjee and Chua [220] considered some linguistic features viz. readability, genre, and writing style in order to predict review spam. First, they computed readability using Gunning-Fog Index (FOG), Coleman-Liau Index, (CLI), Automated-Readability Index (ARI), and Flesch- Kincaid Grade Level (FKG). Second, the review genre included distribution of different POS. Third, writing style calculated ratio of positive cues, perceptual words, and future tense using LIWC software. In total, they considered 13 features involving two readability metrics, the eight POS tags, and the three writing style metrics. The variance inflation factors

(VIF) were calculated in order to check the existence of multi-co-linearity. Then, they classified with Ott et al. [212] dataset using Binomial logistic regression and yielded a sensitivity of 71.75% and a specificity of 70.75%, where deceptive review spam comes under positive class. The proposed approach can be employed to other negative review spam review detection, different domain, and big data as well.

Fusilier [232] developed a modified PU-learning (Positive and Unlabeled samples based learning) [233] in order to detect positive and negative deceptive opinions. PU-learning is a semi-supervised kind of learning, which performs binary classification using positive samples and unlabelled samples. They experimented with Ott et al. [200] dataset, which contains 400 deceptive and 400 truthful reviews on each positive and negative category. The modified PU-learning model yielded an F-measure of 78.0% on positive spam vs. positive genuine and 65.7% on negative spam vs. negative genuine dataset respectively. In future, they would like to apply the proposed approach in the detection of online sexual predators and lies. Costa et al. [210] performed spam classification of tips written about different places. They experimented with 7076 tips (3538 spam and 3538 non-spam) which were collected from a Brazilian Location Based Social Networks using Apontador API. Spam and non-spam labels were manually assigned by three annotators. Classification was performed on the basis of *Variable Importance Measures* and ranked 60 attributes comprising content, user, place, social attributes etc. The flat and hierarchical classifications had been performed using SVM and RF. The flat classification was performed into four classes: non-spam, local marketing spam, bad-mouthing spam and pollution spam. The hierarchical classification was performed at two levels. The first level classification was non-spam vs. spam. The second level classification was performed for spam group consisting of three classes: local marketing, bad-marketing and pollution. For both classifiers, flat classification dominated hierarchical classification for non-spam and local marketing determination and hierarchical classification dominated flat classification for bad-marketing and pollution detection. They analyzed the impact of different subsets of attributes of size 10 and all subgroups of attributes were found to outperform the baseline method. The proposed approach can be extended to find spam group as well.

Therefore, in summary, the key challenges in this area include lack of proper review spam dataset in order to compare effectiveness of two or more techniques and no access to spammers' identity to analyst. There are mainly three techniques to collect spam reviews, which have been exploited in different opinion spam detection studies. First option is to collect suspected spam reviews from different review sites. Second alternative is to synthesize a list of non-real reviews from existing real reviews by taking different combination of sentences. Third solution is to employ human resource from Amazon Mechanical Turk[22] (AMT) to write fake reviews, but that cannot make a real challenging task for analyst. That happens due to psycholinguistic state of mind of an AMT employer will be very different from a

[22] https://www.mturk.com/mturk/welcome.

http://dx.doi.org/10.1016/j.knosys.2015.06.015

fake reviewer. Moreover, a fake reviewer can write a fake review, which can be quite similar to an original review.

3.5 Lexica and corpora creation

A lexicon is a vocabulary of sentiment words with respective sentiment polarity and strength value. The lexicon creation starts with an initial list of words also known as *seed words* and the list is extended using synonym and antonym of seed words. Synonyms and antonyms words were taken from WordNet [133] dictionary. This process is repeated until extension of the list is not stopped. Lexicon can be broadly divided into two categories: non-ontology based [6, 24, 55, 56, 69, 74, 97, 111, 116, 117, 118, 127] and ontology based [21, 23, 30, 52, 106, 120, 199, and 211]. Majority of these study considered here are based on English language, except [26, 111, and 211].

3.5.1 Non-ontology based approaches

This category includes lexica created using machine learning, lexicon based and hybrid approaches.

Thelwall et al. [116] produced a lexicon, SentiStrength, for sentiment classification. The lexicon was created by identifying sentiment expressed in a range of recognized *nonstandard spellings* and *other common textual methods* found on MySpace. For the evaluation of the proposed lexicon, MySpace[23] 2,600 comments for the learning and 1,041 comments for the evaluation were considered. The effectiveness of the proposed lexicon based sentiment classification was compared against a host of classifiers. The proposed method predicted positive emotion with an accuracy of 60.6% and negative emotion with an accuracy of 72.8%. The performance of the system can be improved through linguistic processing. Thelwall et al. [117] analyzed the sentiment strength of public opinion associated with popular events. Observations on top 30 events gave strong evidence that popular events were normally associated with increases in negative sentiment strength. For the experiments, part of spinn3r datasets, 34,770,790 English language tweets from 2,749,840 different accounts were considered. Sentiment classification is performed using a dictionary, SentiStrength [116], which gave sentiment strength either positive or negative of each tweet. The hourly average positive and negative sentiment strength scores have also been computed for each topic. They came up with six different conclusions regarding average sentiment strength difference between different events.

Working further on SentiStrength, Thelwall et al. [118] assessed the extended version of SentiStrength viz., SentiStrength 2. Evaluations was performed on six diverse social web datasets viz. BBC forum, Runners World Marathon discussion forum, MySpace comments, Tweets, YouTube comments, Digg News identification site. SentiStrength 2 performed well on all datasets except BBC news discussion forums. Machine-learning approach can outperform SentiStrength 2 if discourse features

[23] www.myspace.com.

http://dx.doi.org/10.1016/j.knosys.2015.06.015

were indirectly associated with sentiment rather than by directly identifying sentiment. Logistic regression was more suitable than SentiStrength 2, if sufficient human-coded data were available particularly for news discussions. The lexicon-based approach was more robust and relatively domain-independent. For future enhancement of SentiStrength, discourse analysis, irony and sarcasm detection should be considered. Nielsen [55] extended an existing sentiment lexicon namely AFINN-96. Extension was performed using a list of existing lexica and corpora. Lexica and corpora include the Compass DeRose uide to Emotion Words[24], Urban dictionary, Wiktionary, Original Balanced Affective Word List[25], tweets on the United Nation Climate Conference etc. AFINN-96 outperformed Affective Norms for English Words (ANEW) for twitter data but not SentiStrength. Further extension is required to improve its effectiveness.

Thelwall and Buckley [127] classified sentiments using SentiStrength in two phases: mood setting and lexicon extension. These functions can improve the accuracy of topic-specific lexical sentiment strength detection. Thus, for sentiment classification, SentiStrength lexicon was extended for the particular topic with respect to mood of social web. The proposed method was applied for tweet corpora: the UK riot rumors corpus and the AV referendum corpus. Evaluation was performed on six corpora generated in [118]. The proposed method was not performing well on news discussion forums. In future, evaluation is required on larger corpora.

SentiWordNet 3.0 is the one of the most promising lexical resource for sentiment score determination of an opinion word [6]. It defines three types of sentiment scores (Obj(s), Pos(s), and Neg(s), which range from 0.0 to 1.0 and in sum equal to 1.0) to each noun, verb, adjective and adverb given in WordNet [133]. Score was decided using eight ternary classifiers, which were trained on glosses defined for WordNet synsets. But, 93.75 percent synset of SentiWordNet 3 are objective type. Hung et al. [97] considered these objective words to further classify into binary polarity. Sentiment orientation of an objective word was decided on the basis of other common subjective words used in the sentence. If positive score of a sentence was higher than negative score, then objective words were assigned positive polarity otherwise negative polarity. And, sentiment value of the objective word was computed using pairwise mutual information from two words being extended to a word and its associated sentence. Evaluation of the proposed method was performed on 27,886 movie review articles. Sequential Minimal Optimization-SVM with a poly kernel was used for classification and yielded the average accuracy of 76.02%, which was 4.13 percent more than traditional SentiWordNet classifier. In future, semantic viewpoint based sentiment extraction and document summarization at different level can be tried.

[24] http://www.derose.net/steve/resources/emotionwords/ewords.html.
[25] http://www.sci.sdsu.edu/CAL/wordlist/origwordlist.html.

http://dx.doi.org/10.1016/j.knosys.2015.06.015

Further advancement in SentiWordNet can be seen in Neviarouskaya et al. [74], who reported better performance over SentiWordNet. Initially, they created a lexicon SentiFul which was created by exploiting WordNet-Affect database containing 2,438 direct and indirect emotion-related entries based on nine different emotions. SentiFul has been further extended using SentiWordNet and polysemy. New lexemes have been devised by compounding of words, changing bases and affixes of opinion words, where suffixes played pivotal role. Evaluation of the lexicon was performed in two phases: in the first phase two annotators were employed to annotate 200 terms from each of the five lists created by different methods. The second phase evaluation was performed using General Inquirer (GI) and accuracy achieved was 94.1% and 86.3% for *SentiFul core* and *SentiFul* respectively. The proposed dictionary can be extended further using more linguistic features like modal verbs and adverbs. It should then be evaluated using machine learning algorithm.

Taboada et al. [69] extended their own proposed dictionary Semantic Orientation-Calculator (SO-CAL), which gives polarity and strength of an opinion word. They computed semantic orientation using a simple aggregate-and-average method: the total score of all adjectives was divided by the total number of adjectives in the document. To develop the dictionary, different corpora described in Taboada and Grieve [71] and Taboada et al. [72], consisting of a 400-text collection of epinion.com reviews i.e. Epinions 1, were utilized. They considered a list of opinion words taken from Taboada and Grieve [71] and Taboada et al. [72], Epinions 1 Polarity Dataset, and General Inquirer (GI) [105] dictionary to prepare the dictionary. The dictionary was extended by considering intensification, negation, irrealis (non-sentiment word indicator) blocking etc. The performances have been tested on four dataset: (a) Epinions 1, (b) Epinions 2, (c) Pang and Lee [167] dataset, d) Camera: A 2,400-text corpus of camera, printer, and stroller reviews. The highest classification accuracy reported for Computers in Epinions 1 and Epinions 2 was 0.94 and 0.90 respectively. The proposed dictionary was also evaluated across domain using different dataset for short texts. In future, contextual information, word disambiguation, and discourse analysis can be considered.

In the view of several names for a feature of the product, Zhai et al. [56] developed an automatic system for synonym grouping. They applied semi-supervised learning, which starts with a small set of seeds i.e. labeled data, and expanded with other synonyms. Expectation maximization was applied to assign soft constrained labels to a set of unlabeled feature expressions. WordNet was utilized to calculate lexical similarity between different classes of features. Bayesian learning was applied to each feature expression of the document. Evaluation was performed on five domains: home theater, insurance, mattresses, cars, and vacuums, which included reviews, feature expressions, and groups. The proposed soft constrained-expectation maximization method outperformed 16 baseline methods. In future, natural-language knowledge is to be exploited at more semantic level.

Kang et al. [111] created a new senti-lexicon[26] for SA of restaurant review, which was based on unigram + bigram, negation and intensifiers. Two improved Naïve Bayes methods were proposed to map the gap between positive and negative classification accuracy, and to improve the average classification accuracy. Sentiment classification has been performed on approximately 70,000 review documents from different restaurant search sites. The proposed NB outperformed baseline methods, and yielded up to accuracy of 81.2%. In future, performance of the proposed algorithm is to be tested using more experiments and for different languages.

Maks and Vossen [24] extended an existing Dutch lexicon, i.e. Cornetto. The lexicon combines two resources with different semantic organizations: the Dutch WordNet and the Dutch Reference Lexicon. They emphasized on the extent to which a speaker is able to interpret the meaning. For this purpose, they considered different type of subjectivity like *epistemic subjectivity* (how certain or uncertain a speaker is about a given state-of-affairs), *speaker subjectivity*, *character subjectivity*, where existing literature concerns more about the presence of a speaker. For evaluation, annotation was performed on 574 verbs, 595 nouns, and 609 adjective lexical units, yielded inter-annotator agreement 66%, 74%, and 79% respectively. They created a gold standard of approximately 600 items for each part-of-speech in order to build a rich subjectivity lexicon of Dutch.

3.5.2 Ontology based approaches

An ontology is an explicit specification of a conceptualization [192]. It provides formal representation of knowledge, which enables reasoning. It is better than taxonomy or relational database management system, since it captures semantic association between concepts and relationships as well. Therefore, SA community swiftly moving towards ontological based approaches to represent commonsense knowledge base [95, 106, and 120].

ConceptNet is a freely available large-scale knowledge base with an integrated natural-language-processing tool-kit that supports many practical textual-reasoning tasks over real-world documents [120]. Tsai et al. [21] developed sentiment dictionary based on ConceptNet using iterative regression model. The regression model was built upon concept, concept sentiment value and polarity, and features of neighboring concepts. Values generated using regression model were used as the starting values for the proposed random-walk method with in-link normalization. 4-ary concept polarity and concept-pair ranking evaluation was performed by knowledge workers, Amazon Mechanical Turk[27], whose annotation quality was assessed using ANEW dictionary. SVM was used for iterative regression model, which was trained on SenticNet and ANEW, and tested on *Outside dataset* (unique concepts of ConceptNet not found in SenticNet and ANEW). The proposed method achieves the highest ratios in both datasets, which

[26] Available at: http://irlab.sejong.ac.kr/res-sentiwordnet/.
[27] www.mturk.com/mturk/welcome.

http://dx.doi.org/10.1016/j.knosys.2015.06.015

ensures the fewer concepts have insignificantly low sentiment values.. The study was further extended by Wu and Tsai [207]. Wu and Tsai [207] added two new steps viz. relation selection and bias correction. They selected 33 relation types of ConceptNet and applied sequential forward search to filter relations in order to modify existing lexicon. Bias correction was carried out using zero-alignment, and mean & variance alignment. Evaluation of polarity accuracy determination and concept-pair ranking were performed by AMT worker. Moreover, inter-annotator agreement as measured by by Fleiss' kappa value, turned to be 0.74. Experiments were carried out on the datasets of Pang and Lee [167] and Rajagopal et al. [208] for sentiment classification. Effectiveness of the lexicon can be ensured by performing sentiment classification on different domains.

As regards ontological development of sentiment knowledge base, Cambria et al. [106] developed SenticNet 1.0. They exploited ConceptNet [120] and WordNet-Affect[28] (WNA) [4] to build a n-dimensional vector space of affective knowledge viz., AffectiveSpace [141]. AffectiveSpace was obtained by applying singular value decomposition in order to select 100 principal components representing common sense concepts and emotions. In order to decide positive and negative valence, they considered positive and negative eigen components. Unlike 3 sentiment score values in SentiWordNet, single sentiment score was assigned to each concept on the basis of hourglass of emotions. Manual evaluation was performed on 2,000 patient opinions and yielded better F-measure over SentiWordNet i.e. 67% versus 49%. SenticNet 1.0 was further extended by Poria et al. [23] with emotion label taken from WNA [4]. They used supervised machine-learning to assign WNA emotion labels to SenticNet's concepts. WordNet 3.0 [142] was exploited for lexical induction to SenticNet 1.0 WNA and International Survey of Emotion Antecedents and Reactions (ISEAR) dataset[29] was used to extract classification features, which yielded 40 ISEAR dataset attributes and corpus-based similarity measures. Concept co-occurrence similarity measure was calculated using SenticNet score-based similarity, WordNet distance-based similarity, PMI, emotional affinity, and ISEAR text-distance similarity. ISEAR-based features were considered for classification. SVM yielded the best accuracy of 88.64%, which was better than that of NB (71.20%) and MLP (74.12%). The proposed lexicon can be extended with new concepts taken from new monolingual or multilingual corpora.

Meanwhile, Cambria et al. [136] proposed SenticNet 2.0, which was based on sentic computing and principles taken from computer and social sciences. It provides the semantics and sentics (that is, the cognitive and affective information) associated with over 14,000 concepts. In order to capture common and common-sense knowledge base, Cambria et al. [202] developed SenticNet 3, which was an ensemble of several knwoledge base viz. ConceptNet, DBPedia, WordNet etc. SenticNet 3 knowledge base was

[28]www.cse.unt.edu/~rada/affectivetext/data/WordNetAffectEmotionLists.tar.gz.

[29]www.affective-sciences.org/system/files/page/2636/ ISEAR.zip and www.affective-sciences.org/researchmaterial

developed in two stages. In the first stage, it builds Resource Description Framework (RDF) triplets, which were then were inserted into a graph through the energy-based knowledge representation (EBKR) formalism. Cambria et al. [52] proposed a novel cognitive model to capture multi-word expressions by exploiting ConceptNet and WordNet-Affect. AffectiveSpace [141] was able to grasp the semantic and affective similarity between concepts by plotting them into a multi-dimensional vector space. Principal component analysis was applied for feature selection. KNN and K-medoids were applied to determine semantically related concepts. In order to determine the level of affective valence of concept 'discrete' neural network (DNN) and 'continuous' neural network (CNN) were designed. Testing was performed on the benchmark for affective common-sense knowledge (BACK) [140] using 10-FCV. DNN yielded the highest strict accuracy of 46.9% and CNN yielded the highest relaxed accuracy of 84.3% over K-NN, and k-medoids, and Random classifier. For practical environment, testing was performed by embedding ANNs into an opinion mining engine [140] to infer cognitive and affective information associated with natural language on a patient opinion database[30] and CNN outperformed baseline methods. In future, more feature selection methods can be tested.

Along the same line, Poria et al. [213] introduced a hybrid approach comprises linguistics, common-sense computing, and machine learning for concept-level sentiment analysis. In order to extract concepts and events from natural language text, they proposed POS-based bigram algorithm and event concept extraction algorithm. Concept polarity and intensity were calculated using AffectiveSpace and the Hourglass of Emotions, where AffectiveSpace was obtained by applying truncated singular value decomposition on matrix representation of AffectNet. Extreme Learning Machine (ELM) was applied to cluster AffectiveSpace with respect to the Hourglass model. A list of linguistic patterns was developed to deal with various natural language complexities like coordination, discourse and dependency rules based structures. Experiments were carried out on Socher et al. [201] and Blitzer et al. [149] and yielded a precision of 86.21% and 87% respectively. The proposed approach also outperformed state-of-the-art approach in sentiment classification of sentences with conjunctions and comparisons. In other study, Poria et al. [214] developed EmoSenticSpace on the basis of SenticNet, EmoSenticNet and AffectiveSpace, where EmoSenticNet was created by merging WNA and SenticNet. They assigned one of the six emotion label of WordNetAffect viz. anger, fear, disgust, sadness, surprise, and joy to each concept of SenticNet. In order to do that, they applied supervised classification using SVM. The number of features was identified in two stages. In the first stage, in order to obtain membership value of concept to six emotion categories, they applied Fuzzy C-means clustering (FCM). FCM was applied on 16 ISEAR data columns and 13 similarity measures calculated using WordNet 3.0, SenticNet score and co-occurrence. In the second stage, membership values belonging to six emotion categories were considered.

[30] http://patientopinion.org.uk.

SVM yielded 92.15% accuracy, which was applied to top two fuzzy clusters. Evaluation was performed on sentiment classification of Stanford Twitter Dataset [169], emotion recognition on the ISEAR dataset, and personality detection from Matthews et al. [215] dataset. Some possible improvements have been suggested in the literature. The proposed lexicon can be applied to different domains.

Balahur et al. [30] constructed a knowledge base EmotiNet for representing and storing affective reaction to real-life contexts. They exploited various existing dictionaries and corpora like SentiWordNet [5], ConceptNet [120], VerbOcean [146], and ISEAR [147] to extract emotion expressing words. They evaluated the proposed ontology on three different size of dataset created using ISEAR and found promising result. In future, EmotiNet can be improved by adding more affective properties to the concepts. Montejo-Ráez et al. [211] developed a corpora of emotional words crawled from streaming data to perform sentiment classification. They extended an existing project WeFeelFine[31] for Spanish language. WeFeelFine contained 2178 different feelings collected from 2 million sentences. Out of 2178 feelings, the 200 most frequent ones had been considered for the study. They crawled 1,863,758 tweets to generate MeSientoX corpus in Spanish. Opinion words were extracted on the presence of ‘‘me siento’’ (‘‘I feel’’) and the polarity was decided using SenticNet 3 and three manual annotators. The whole process yielded 201 opinion words (the most frequently occurring ones), of which 84 were considered as positive and 117 as negative opinion words. In order to perform explicit sentiment analysis, a document was projected onto a space of feelings, and latent semantic analysis based cosine similarity was calculated to determine the polarity. Evaluations were performed on Emoticon data set and SFU Review corpus and yielded an accuracy of 71.86% and 69.75% respectively. They indicated to use on-line learning algorithms and evolving training data in future work.

3.6 Opinion feature extraction and product aspects extraction

In order to perform fine grained level SA, we need to identify people's opinion on various parts of a product. Sentiment score of different aspect will have varying degree of affect on aggregate opinion on a product. Thus, mostly discussed as well as important aspect should be extracted from feedback text. Data collected from diverse domain often contain so much noise that it makes opinion word and aspect extraction very hard. Therefore, research attempts were made to address these issues.

3.6.1 Opinion feature extraction

It has been observed that opinion feature may appear either in text format or social media links. Therefore, this sub-section is divided into two groups namely *opinion word extraction* and *opinion extraction from social networking*.

31 http://wefeelfine.org.

3.6.1.1 Opinion word extraction

Most effective part of speech for opinion words is adjective, adverb, verb, and noun. These kinds of part-of-speech are known as *opinion or sentiment word*. Thus, extracting opinion words also follows machine learning based [61, 63, 92, 100], lexicon based [68, 132, 175] and hybrid approaches [8, 102].

Zhu et al. [63] developed aspect-based sentence segmentation model and performed aspect-based opinion polling. A multi-aspect bootstrapping (MAB) method was used to learn aspect related terms from unlabelled data. MAB was started with 5 seeds per aspect from 500 frequent nouns. Experiments were performed on 3,325 Chinese restaurant reviews collected from the DianPing.com and achieved 75.5% accuracy for opinion polling. Domain adaptation was a critical issue for opinion polling. Liu et al. [100] presented opinion extraction and product features extraction, which was applied for recommender system. To extract an opinion word, association rule mining was applied wherever an adjective follows an adverb and a product feature was extracted by taking n-gram window around an opinion word. For experiments, they downloaded 54,208 Chinese reviews from (beijing.koubei.com), yielded 60.23% recall for opinion feature extraction which was more than Hu and Liu [38] method. Context-aware recommendation system can also be tried.

Lin et al. [61] proposed Joint Sentiment-Topic (JST) and Reverse-JST. Both models were based on modified Latent Dirichlet Allocation (LDA) with 4 levels i.e. four layer hierarchical Bayesian model. Models can extract sentiment as well as positive and negative topic from the text. JST and RJST outperformed baseline methods. JST yielded accuracy of 76.6% and 72.5% on Pang and Lee [167] dataset, and Blitzer et al. [149] respectively. Similarly, RJST yielded accuracy of 76.6% and 71.92% on Pang and Lee [167] dataset, and Blitzer et al. [149] respectively. Other supervised information can be incorporated into LDA to improve the efficiency. Duric and Song [92] suggested three different methods for sentiment word extraction; (a) Hidden Markov Model - Latent Dirichlet Allocation (HMM-LDA) to get syntactic classes of features, (b) syntactic and semantic classes based features, and (c) max scores based features. Experiments were carried out on Pang and Lee [33] dataset. Maximum Entropy (ME) yielded up to accuracy of 87.5% classification using max scores. All proposed scheme outperformed POS-based feature selection schemes. The proposed techniques should be tested on other domain and other purposes like aspect extraction as well.

Zhan et al. [132] applied text summarization approach to gather customer concern. They proposed a new algorithm based on frequent word sequences and equivalent classes to extract candidate sentences for summarization. Experiments were performed on five sets from Hu and Liu [135] corpus and three datasets from Amazon.com. The proposed approach yielded better results than opinion mining and clustering summarization approach. The proposed approach should be tested on other corpora as well. Wang et al. [175] applied dictionary based approach (viz., Hownet) to extract opinion phrases from 61

Chinese blog posts on digital camera. They employed window based opinion extraction method, which considers same polarity for words utilized along with other opinion words in the same window. To improve the effectiveness of the proposed approach further, shallow parsing techniques can be applied. Cruz et al. [68] defined a set of domain-specific resources to extract the opinion words. Domain-specific resource consists of a set of features like feature-taxonomy, feature cues, and dependency patterns. They utilized dictionary based approach like WordNet-, PMI-, and SentiWordNet-based classifier for sentiment classification. The lexicon was expanded using random walk algorithm. They created a dataset[32] for the experiment. For the opinion extraction, resource based extraction outperformed resource free extraction and achieved up to a precision of 71.69%. An accuracy of 93.45% was achieved for sentiment classification. The domain adaptability is still a major issue.

Li et al. [8] devised a new product feature ranking method DPLR-R (Double Propagation-Based Linear Regression with Rules). They proposed page ranking algorithm based feature opinion rank to rank the product aspects. Rank was based on the number of opinion words utilized along with a feature and/or vice-versa. Experiments were performed on 6000 phone, laptop and hotel reviews and achieved better results than baseline methods. The challenge lies in extraction of useful training features for these ranking models. Zhai et al. [102] proposed to extract different opinion features like sentiment term, substring, substring-group and key-substring group features. Experiments were carried out on Hotel[33], Product[34], Ctrip[35], Chinese IT-product website-IT168[36] datasets. SVM was trained on selected features and yielded up to accuracy of 91.9%. The proposed feature selection techniques can be tried for other languages as well.

3.6.1.2 Opinion extraction from social networking

On social networking, if someone is linked to/followed/liked a social networking group or community or public figure, it implies that the person has either positive or negative feelings towards that entity. Thus, social networking provides a new research challenge for SA community to figure sentiment from such links. To the best of our knowledge, very few studies have been carried out on this issue. Thus, some of such rare initiatives using hybrid approaches are reviewed here.

Rabelo et al. [11] utilized link mining and user-centric approach to get opinion expressed by social network users. Link was utilized for assigning classes to objects based on the relationships among objects. For the experiment, they collected 8000 posts containing selected hashtags to get 97000 nodes and almost 1 million edges for the graph to apply collective classification algorithm. They applied a link

[32] http://www.lsi.us.es/_fermin/index.php/Datasets.
[33] http://www.searchforum.org.cn/tansongbo/corpus/ChnSentiCorp_htl_ba_4000.rar.
[34] http://wanxiaojun1979.googlepages.com/PKU-ICST-ProductReviewData.rar.
[35] http://www.ctrip.com/.
[36] http://www.it168.com.

http://dx.doi.org/10.1016/j.knosys.2015.06.015

mining technique and assigned classes to objects based on the relationships among objects. The proposed approach can be applied for group detection. Camp and Bosch [125] performed classification of personal relationships in biographical texts. In order to do that, they performed induction of social networks using social network extraction and SA. They utilized two types of feature sets viz. lexical features and co-occurrence features to train SVM, NB, Ripper, and KNN classifiers. For the experiment, they selected 574 Biographical articles from Biographical Dictionary of Socialism and the Labor movement in the Netherlands (BWSA) in Dutch language. Evaluation of the system's performance was performed by the expert in the field of Dutch social history. Induction of named entity recognition and disambiguation will help to extract more accurate relations.

Sobkowicz et al. [27] framed an opinion formation framework using agent technology, where agents were able to capture opinions based on effects of leadership, dynamic social structure, and effects of social distance. Bayesian framework was suggested for opinion formation forecasting. Application scenarios were recommended for political discussions in Poland, governance of Java standard, and BP oil spill emails.

3.6.2 Product aspect extraction for aspect level sentiment analysis

For fine-grained comparison of two or more products of similar categories, we need to figure out pros and cons of various components and features (aka aspects). Let us consider a review on Apple iPhone from amazon.com.

> "One thing is reading specs, and another is using it. The new iphone is much better in my hand than it could be forecasted from the specs. It feels *much lighter*, with the *nice bigger screen*. And it's *faster* and takes *better pictures*, despite camera having the same specs!"

In this review, positive opinions are expressed on various aspects and components of an iPhone. Here, *"much ligther"* refers to *"Weight"*, *"nice bigger"* is used for *"screen", "faster"* refers to *"processor speed"* and *"better pictures"* for *"camera quality"*. So, notable aspects discussed are *weight, screen, processor,* and *camera*. In order to extract aspects and aspect level analysis, aspect level SA came into picture. During the last few years, some of the researchers have focused on this issue. Product aspect extraction approaches are divided into two categories: (a) non-ontological approach [25, 37, 59, 67, 107, 185, 186, 189, 190, 191, 193, 218, 240, 241, and 243], (b) ontological approach [35, 62, 182, 194, 195, and 196].

3.6.2.1 Non-ontological approach

Non-ontological approaches can further be divided into topic modeling based [107, 185, 189, 191, 240] and non-topic modeling based approaches [25, 37, 59, 67, 186, 190, 193, 218, 241, 243].

Moghaddam and Ester [185] devised factorized LDA (FLDA) to extract aspects and estimate aspect rating from cold-start problem. The cold-start term was taken from recommender system, where on each product item, small amount of reviews were available. They considered reviews influence too for both purposes. They worked on multi-domain reviews taken from epinions.com, amazon.com, tripadvisor.com etc. FLDA-SVM yielded up to accuracy of 86% for item categorization and an accuracy of 74% for review rating on cold-start items from TripAdvisor. In future, item categorization can be performed hierarchically. Wang et al. [189] projected two semi-supervised model viz. Fine-grained Labeled LDA (FL-LDA) and Unified Fine-grained Labeled-LDA (UFL-LDA) to extract aspects from reviews. The first model was incorporated with products seeding aspects taken from e-commerce site. The second model was incorporated with unlabeled documents to consider high-frequency words. Experiments were performed on Wang et al. [187]. The qualitative results were evaluated using cosine similarity between seeding aspect and generated aspects. And, the quantitative results were evaluated using RandIndex, Entropy, and Purity. In future, incorporating conceptual knowledge into a topic model was required to capture aspect hierarchies.

Zheng et al. [191] incorporated appraisal expression pattern into LDA (AEP-LDA) to extract aspect and sentiment. Appraisal expression pattern were captured using the shortest dependency path between POS. The meaningfulness of pattern was decided using confidence score. Gibbs sampling was used for inference. Hotel, restaurant, MP3 player and Camera reviews were collected for experiments. AEP-LDA outperformed some other variants of LDA for aspect identification. AEP-LDA outperformed PMI and nearest neighbor method for sentiment word identification. The proposed model can be extended for aspect-based review summarization and clause level topic modeling. Xueke et al. [107] came up with a Joint Aspect/Sentiment model (JAS) to extract aspects. They developed aspect dependent sentiment lexicons. The lexicon has been applied to a series of aspect-level opinion mining tasks viz. implicit aspect identification, aspect-based extractive opinion summarization and classification. Sentiment prior was induced to Latent Dirichlet Allocation (LDA) for topic extraction and subjectivity classification. Experiments were performed on restaurant reviews and hotel reviews. MPQA, SentiWordNet, and union of them were utilized as opinion lexicon. The proposed model outperformed other two models like MaxEnt-LDA and Aspect and Sentiment Unification Model (ASUM). SVM with linear kernel yielded promising result tor sentiment classification. The proposed opinion lexicon can be extended with embedded composite opinion words, synonyms and antonyms etc. Xu et al [240] developed an implicit aspect extraction model, which used LDA-based explicit topic model and SVM. LDA was used to extract aspects, where a few explicit aspects were assigned to different topics before applying topic modeling. Explicit topic model was guided by two constraints and relevance-based prior knowledge, which yielded training attributes for SVM. Finally SVM was trained to classify implicit and explicit features. They experimented with 14218 Chinse sentences from www.360buy.com and yielded an F-measure of 77.78%. The proposed method should be tested on bigger corpora and be developed for different domains and languages.

Kim et al. [186] formed a hierarchical aspect sentiment model (HASM) to represent hierarchical structure of aspect-based sentiments. They extracted structure and parameters of the tree using a Bayesian non-parametric model viz., recursive Chinese Restaurant Process. Gibbs sampling was used for inference. Experiments were performed on reviews on different product and services. The proposed model outperformed three other models for small scale and large scale dataset for classification purpose. They found better hierarchical affinity and aspect-sentiment consistency than baseline methods, thereby demonstrating the power of HASM. The proposed model can further be utilized to discover a set of topics with shared features in a hierarchical structure. Zhang et al. [190] suggested context-based sentiment classification method. They incorporated CRF with several syntactic and semantic features like similarity to neighboring sentences, word positions, comparative and superlative POS, conjunction terms, +ve/-ve emotions etc. Experiments were performed on 300 digital camera reviews, 300 TV reviews, and 500 Facebook comments from 13 walls. The proposed approach outperformed rule based algorithm, SVM, LR, and HMM for camera reviews with an accuracy of 72%. Sarcastic sentences can be included in the scope in future.

Chen et al. [37] compared the Conditional Random Fields (CRF) based opinion mining method to four typical related methods: (a) model based methods such as Lexicalized Hidden Markov Model (L-HMM), (b) statistical methods like association rule mining based (ASM) techniques, (c) ASM + Linguistics (d) rule-based method on the basis of several opinion mining units: basic product entities, opinions, intensifiers, phrases, infrequent entities, and opinion sentences. Experiments were performed on Hu and Liu [38] dataset containing 821 reviews on 9 digital cameras and in addition to that authors collected 187 reviews of four extra digital cameras. It has been observed that CRF-based learning method was more suitable for mining aspects, opinions and intensifiers (including phrases) in comparison to L-HMMs based and statistical methods. Visualization of opinion mining (OM) results and evaluation of opinion extraction is required. García-Moya et al. [59] introduced a language modeling framework for aspect-based summarization of reviews. The framework combines a probabilistic model of opinion words and a stochastic mapping between words. It estimates a unigram language model of product features. Expectation-maximization (EM) was utilized to minimize the cross entropy, which was based on background language model of the source language. To retrieve the product features, iterative strategy was followed, which starts with an initial list of features and expanded using bottom-up strategy. A kernel-based density estimation approach was utilized to learn the model of opinion words, which started with a list of seed words from SenticNet. Experiments were performed on 3 datasets: Hu and Liu [38],

http://dx.doi.org/10.1016/j.knosys.2015.06.015

TBOD, TripAdvisor. Training for opinion words was performed on Hu and Liu [38][37] for English and the fullStrengthLexicon for Spanish. The proposed system outperformed two baseline methods (a) Double Propagation, and (b) Hyperlink-Induced Topic Search (HITS-based) algorithm for aspects extraction. Integration of the proposed model into a probabilistic topic-modeling framework can be performed.

Hai et al. [193] considered domain relevance of an opinion feature to identify opinion feature. They suggested intrinsic and extrinsic domain relevance, which relies on threshold for dispersion and deviation calculated on the basis of TF-IDF weight. The domain relevance formula was used for measuring inter-corpus statistics disparity. Experiments were performed on Chinese 10,073 cellphone and 6,313 hotel reviews. The proposed approach outperformed other six methods for opinion feature extraction as well as for feature-based opinion mining. In future, the proposed approach can be modified to deal with infrequent features, non-noun opinion feature, and multi-lingual corpora. Miao et al. [25] developed a sentiment mining and retrieval system viz., AMAZING. AMAZING uses a novel ranking mechanism Temporal Opinion Quality (TOQ) and relevance. This method can rank review sentences, visualize opinion trends, and evaluation of a particular feature. POS tagging was performed using NLProcessor. Rule association miner CBA was applied to extract frequent nouns aka features. Opinions were extracted with respect to each feature using dictionary based approach, which also yielded polarity and strength. Lucene API was utilized to create index file of review sentences. Review sentence rank was created using TOQ and Lucene Rank, which was based on Vector Space Model. Experiments were performed on product reviews on 20 digital cameras, 20 cell phones, 20 laptops and 20 MP3 players taken from Amazon.com. Average precision 87.4% was achieved on queries supplied by 8 participants. In future, implicit opinion and opinions expressed with verbs are required to be extracted.

Bagheri et al. [218] proposed an unsupervised and domain-independent model to extract explicit and implicit multi- and single word aspects. In order to get candidate explicit aspect list, they exploited four POS patterns based on nouns, adjectives, determiners, and verb gerunds. The aspect list was further expanded using bootstrapping algorithm. They ranked the obtained aspects list by applying extended FLR (Frequencies and Left and Right of the current word) algorithm. Pruning was performed using PMI and frequency based A-scoring method, two heuristic rules, subset- and superset-support based rules. Moreover, graph based scoring method were employed to identify implicit words, where a pre-defined aspects and related opinion words were represented using nodes. They experimented with more than 2400 reviews on five gadgets collected from amazon.com and cnet.com and yielded an average precision of 84.52%. The proposed approach outperformed Wei et al. [219] and Hu and Liu [38]. They proposed to augment the proposed method with clustering in order to extract explicit and implicit aspects and opinions.

[37] www.cs.uic.edu/~liub/FBS/opinion-lexicon-English.rar

http://dx.doi.org/10.1016/j.knosys.2015.06.015

Quan and Ren [67] coined a novel similarity measure, PMI-TFIDF, to identify association between products and its aspects. They performed aspect-level SA based on aspect-opinion pair extraction and aspect oriented opinion lexicon generation. Experiments were carried out on 10 datasets, where 4have been taken from Hu and Liu [135]. For feature extraction, the proposed method outperformed that of Hu and Liu [135], Popescu and Etzioni [143], Qiu et al. [145] for 4 datasets and yielded more than 89% recall. For sentiment classification, three hybrid models were proposed, where 2 hybrid models were based on PMI-TF-IDF that outperformed third model GI-SVM (with polynomial kernel). Aspect based ontology generation is claimed as future work. Yan et al. [241] developed EXPRS (An Extended PageRank algorithm enhanced by a Synonym lexicon) to extract product features. In order to extract features, they extracted nouns/noun phrases first, and then extracted dependency relations between nouns/noun phrases and associated sentiment words. Dependency relations included subject-predicate relations, adjectival modifying relations, relative clause modifying relations, and verb-object relations. The list of product features was extended using its synonyms. Non-features nouns were removed on the basis of proper nouns, brand names, verbal nouns and personal nouns. They experimented with 8901, 11291, and 9000 reviews on Samsung GalaxyNote II, Canon EOS 600D, and Philips DVP3600 respectively. In terms of results, they achieved an F-measure of 76.2%, 73.0%, and 72.7% on Samsung GalaxyNote II, Canon EOS 600D, and Philips DVP3600 respectively and in the process outperformed Eirinaki et al. [154] and HITS based algorithm Zhang et al. [242]. Their method does not deal with implicit features, on which sentiments are expressed without specifying a product feature name in the reviews. Li et al. [243] augmented frequency-based extraction and PMI-IR in order to extract product aspects. In order to identify frequent aspects from the dataset, they employed Apriori algorithm and they removed useless rules using compactness and redundancy rules. To finalize the number of relevant aspects they applied location based ordering and similarity i.e. PMI-IR based filtering. In order to classify the extracted aspects into aspects and non-aspects category, they applied RCut with threshold value of 1 on PMI-IR score. They experimented with 120 reviews for nine categories of products, achieved an average F-score of 73.3% and outperformed Popescu and Etzioni [143]. In future, extraction of infrequent aspects can be performed.

3.6.2.2 Ontological approaches

Quality of a product is fully dependent on its aspects' quality. In turn, aspects will be qualified by their various attributes. Hence, we can say that products and aspects are hierarchically related. In order to capture this relationship, ontological approaches proved helpful. And, it yielded better sentiment classification accuracy over non-ontological approaches [35]. Therefore, some of initiatives taken by SA community are presented in this section.

http://dx.doi.org/10.1016/j.knosys.2015.06.015

Liu et al. [62] automatically constructed fuzzy domain sentiment ontology tree (FDSOT) based on the product features and sentiment words. FDSOT was used further for feature based sentiment classification. They exploited 2400 product reviews for laptops from 360buy.com to create FDSOT using double propagation method. They achieved a precision of 72.4% compared to 58.4% without FDSOT. To improve the effectiveness of the proposed method, feature relation, discourse analysis can be considered. Lau et al. [182] enacted an ontology learning system for contextual aspect-level SA. They generated ontology in two phases: in the first phase implicit and explicit aspects ontology was generated. In the second phase contextual sentiment words for each aspect was presented in the ontology form. They extracted implicit and explicit aspects using LDA and Gibbs sampling methods. They created ontology for several domains. The proposed ontology quality was compared against Text-to-Onto based ontology. They reported 11.6% better sentiment classification accuracy than that obtained by OpinionFinder. In future aspect-oriented SA can be conducted.

Mukherjee and Joshi [194] focused on sentiments aggregation technique using ontology considering 4 level hierarchical relationships. They assigned high weight to the aspect occurring at higher level and low weight at lower level. Ontology was constructed with the help of ConceptNet 5 ontology. Liu [38] lexicon was used to decide the subjectivity of a word. They collected 584, 986, and 1000 reviews on automobile, camera and software respectively. They further devised a Phrase annotated Author Specific Sentiment Ontology Tree (PASOT) based on an existing movie ontology [195]. Ontology was mapped to WordNet based similarity measure. They aggregated the node polarities using bottom-up approach to get overall review polarity. They experimented with Pang and Lee [33] dataset. The proposed approach outperformed other 11 approaches and achieved an accuracy of 7.55% over SVM classifier. Outlined technique can be applied for other domain as well.

Zhou and Chaovalit [196] manually developed a *movie ontology* for polarity mining. Conceptual properties were derived by two persons by analyzing 180 movie reviews from IMDb.com. They applied maximum likelihood estimation for bi-gram modeling. The performance of bi-gram modeling were compared with that of GI based technique. Further experiments should be carried out on new domains and larger dataset. Peñalver-Martinez et al. [35] enacted a methodology to perform aspect-based SA of movie reviews. To extract the movie features from the reviews, they utilized a domain ontology viz., Movie Ontology. Different relative sentiment score has been assigned to a movie feature according to its appearance in the review viz. beginning, middle, and end. SentiWordNet was utilized to calculate the sentiment score. For the experiment, they exploited Pang and Lee [167] movie dataset to create the domain ontology. The proposed method yielded classification accuracy of 89.6%.

3.6.3 Relation Extraction

Detection of relation between entities is useful for the competitive intelligence. Xu et al. [93] developed a CRF based model to extract comparative relations from customer reviews. Relation extraction was performed in three stages: (i) entities viz., product name and attributes were recognized using a developed lexicon of mobile phone names and attributes (ii) a graphical model was developed using CRF to model the dependencies between relations and entities (iii) then, the belief propagation algorithm was applied with unfixed interdependencies. Experiments were performed on 1348 reviews on 33 types of mobile phones collected from various sources. CRF with/without interdependencies outperformed multi-class SVM. The proposed model should be tested on large scale dataset on different domains. The model can be extended to jointly recognize the comparative relations so as to reduce the errors accumulated in the pipeline errors.

3.6.4 Named entity recognition and name disambiguation in micro-blogs

Named entity recognition and name disambiguation are widely studied on long text, while micro-text based study is also gaining importance. Noisy and short contents of micro-blog make this task harder. Hence, it is an essential task to recognize an entity discussed in the comment. In this regard, Jung [60] considered named entity recognition (NER) system for streaming micro-text. It was based on different contextual associations like semantic association, temporal association, and social association between micro-text clusters. Four types of entity classes namely persons, organizations, locations, and digital IDs were exploited. Maximum entropy approach-based NER method has been applied to determine the micro text clusters. On the basis of manually collected training sets, the proposed NER system detected the named entities from the text streams online.

Spina et al. [101] developed an algorithm for *company name disambiguation* from tweets. They extracted filter keywords from micro-blogs that did not use any previously annotated data about the target company. A positive/negative filter keyword, if present in a tweet, indicates high probability that the tweet was related. Filtering was performed manually as well as using C4.5 decision tree classifier. For automatic filter keyword identification, three hybrid feature selection methods like collection-based, web-based and features expanded with co-occurrence have been applied Experiments were performed on WePS-3 Online Reputation Management Task dataset i.e. TREC Micro blog Track[38] using NN, C4.5 and CART Decision Trees, Linear Support Vector Machines and NB. In future, cross-lingual strategies should be employed.

[38] https://sites.google.com/site/microblogtrack.

3.7 Various applications of opinion mining

SA is in its infancy stage, as far as the real life applications are concerned. It has been tried in various areas like business intelligence, market prediction, attrition prediction etc.,which will be reviewed now.

3.7.1 Market and FOREX rate prediction

Bollen et al. [18] studied market prediction by augmenting twitter sentiment to financial data. For the analysis of the twitter content, two mood tracking tools were utilized, namely OpinionFinder (OF) [39] [34] that measures positive vs. negative mood, and Google-Profile of Mood States (GPOMS) that measures mood in terms of 6 dimensions viz. Calm, Alert, Sure, Vital, Kind, and Happy. In total 7 public mood time series were utilized. For the analysis, 9,853,498 tweets by approximately 2.8 million users during 28 Feb 2008 to Dec 2008 and a time-series of daily Dow Jones Industrial Average closing-values from Yahoo! Finance were collected. To evaluate the effectiveness of OF and GPOMS time series, experiments have been performed on U.S Presidential election and Thanksgiving and which yielded promising result. So, the daily time series obtained by OF and GPOMS were applied on DJIA using econometric technique of Granger causality analysis, which shows that there was predictive relation between certain mood dimensions and DJIA. The relation between public mood and stock market values was almost certainly non-linear, while Granger causality analysis was based on linear regression. Therefore, to address this issue better, Self-organizing Fuzzy Neural Network (SOFNN) was utilized to predict DJIA values on the basis of two sets of inputs: (a) 3 days of DJIA values, and (b) the same combined with various performance of the obtained time series.

Qiu et al. [80] exploited wisdom of crowds approach to predict the market. It was based on the assumption that information and knowledge in social systems frequently exist only as dispersed opinions, and that aggregating dispersed information. A theoretical model of network prediction market was set up, which frames a two-stage network game. In the first stage, information acquisition was performed by an agent. In the second stage, agent makes use of her information to choose the optimal demand for the risky asset. Random social network with 100 agents in the form of 100 x 100-dimensional matrix was utilized. Pari-mutuel betting mechanism was utilized for twitter based market prediction. Future work is claimed as to test whether participants in the same local structure exhibit similar bets and prediction performance.

Yu et al. [124] investigated the effect of social media and conventional media, their relative importance, and their interrelatedness on firm performances. For the experiments, the financial-statement and financial-market data for the 824 companies from COMPUSTAT and the Center of Research in Security Prices (CRSP) was recorded. These data help them to construct measures of abnormal returns and cumulative abnormal returns. Naïve Bayes was trained on Pang and Lee [167] dataset for binary

[39]http://www.cs.pitt.edu/mpqa/opinionfinderrelease/

http://dx.doi.org/10.1016/j.knosys.2015.06.015

sentiment classification system and achieved up to 86% F_1. They found several observations like overall social media has a stronger relationship with firm stock performance than conventional media, both media were interrelated etc. Future work could be to explore the tone of the messages in various media sources, and to study the extent of sentiment among the general public as compared to sophisticated media practitioners. Nassitoussi et al. [183] proposed to predict intraday directional movements of a currency-pair in the foreign exchange market. Contributions were made in many folds for text mining and market prediction. They used SentiWordNet to get sentiment score, WordNet to get hypernyms, TF-IDF to get weight of a term and SVM for sentiment classification of news headlines. They collected financial news from MarketWatch.com and foreign exchange historic data in the form of Euro/USD. Evaluation was performed for heuristics-hypernym feature-selection, TF-IDF sumscore, synchronous targeted feature-reduction, and machine learning algorithm. SVM outperformed NB and KNN and yielded an accuracy of 83.33%. The proposed method can be tried for other market prediction purposes as well.

Li et al. [184] performed intra-day market prediction. They represented news articles in the form of vector space model, which was multiplied by sentiment word matrix. Each headline was labeled with intra-day return. They performed 3-class classification of news (+ve, -ve, and neutral) using SVM. For the experiment, news articles were collected from Hong Kong financial news archive during January 2003 to March 2008. And stock daily quotes were collected from Yahoo! Finance for the same period. Comparisons were performed between stock, index and sector level for the validation and testing purposes. Accuracy achieved up to 69.68% for testing. In future, relationship between news impact and intra-day stock price return can be analyzed. Geva and Zahavi [53] performed a systematic evaluation of the effectiveness of augmenting market-data with textual news-based data for intraday stock recommendation decisions. Dataset were obtained from the New York Stock Exchange (NYSE) Trades Quotes (TAQ) database for 50 sell and purchase companies, and 51,263 news items from 'Reuters 3000'. A feed forward neural network algorithm, decision trees involving a genetic algorithm, and stepwise logistic regression were trained on dataset. They utilized double scoring approach for positive and negative model to reduce prediction biases and rule-based expert system for feature selection. Some other classifiers and hybrid models can be tried on textual and market data for better accuracy.

3.7.2 Box office prediction

Yu et al. [155] mentioned a novel approach to sentiment mining based on Probabilistic Latent Semantic Analysis (PLSA), which was called as Sentiment PLSA (S-PLSA). Autoregressive (AR) model was employed to capture the past performance of same movie. A combined form of S-PLSA and AR was proposed and known as Autoregressive Sentiment and Quality Aware (ARSQA) model for sales prediction. For the purpose of feature selection, the authors took 2030 appraisal words from the Whitelaw et al. [169] lexicon. To determine the helpfulness of a review, $\in$-Support Vector Regression with Radial

http://dx.doi.org/10.1016/j.knosys.2015.06.015

Basis Function (RBF) kernels was utilized to learn the relationship between vector of features and the quality of reviews. Experiments were performed on movie blogs along with helpfulness score collected on two movies, one month box office revenue data from IMDb, and 45,046 blog entries. The proposed model yielded the least Mean Average Percentage Error (MAPE) values in comparison to other competitive model. S-PLSA can be tried for clustering and classification of reviews based on sentiments. It can be used to track and monitor sentiment pattern expressed online.

Du et al. [51] accomplished micro-blogs based box office prediction using two features: count and content based features. Count based features considers several factors like quantity and quality of users, influence of expert users to other users, and time based factors. The content based features label the comments into 3 categories: beneficial, harmful and neutral. 3000 Chinese micro-blogs on 17 movies were labeled for the training purpose. Experiments were performed on 24 movies' box offices collected from Entgroup box office and 120,413 microblogs from 68,269 users from Tencent micro-blog. Different combinations of 5 feature selection methods were applied, where chi-square test plus subsequence using unigram representation yielded the best result. NN outperformed linear model and SVM with linear kernel for classification with lowest MAPE value. Improvement can be made to the proposed method by considering the relationship among users. Rui et al. [131] analyzed twitter sentiment to predict the movies sales. They introduced the pre-consumption concept of product sales, instead of post-consumption. Experiments were performed on movie sales data from BoxOfficeMojo.com, 4166623 tweets on 63 movies along with different author's profile. SVM was utilized for subjectivity classification and NB was utilized for sentiment classification of tweets. Training for both classifiers was performed on 3000 tweets, which were manually labeled. For movie sales prediction, intention tweets ratio (%), positive tweets ratio (%), negative tweets ratio (%) were utilized as explanatory variables. The number of followers of a Twitter user was considered as a variable, which should be replaced by Twitter user's personal influence.

3.7.3 Business analytics

Coussement and Van den Poel [77] considered emotions expressed in client/company emails to be incorporated in a *churn prediction model* for a newspaper subscription business. For the experiment, 18,331 e-mails from a Belgian newspaper have been purchased. Linguistic Inquiry and Word Count (LIWC) [28] was utilized to create a list of 690 positive words and 1,347 negative words, that helps in classifying the contents of emails either in positive or negative category. Logistic Regression (LR), Support Vector Machine (SVM), and RF had been utilized for sentiment classification. RF outperformed other two classifiers. In future, the proposed framework can be utilized for different kind of business applications like cross- and up-sell applications, customer acquisition in wide range of industries, e.g. retail, finance, services, e-commerce etc.

http://dx.doi.org/10.1016/j.knosys.2015.06.015

Kang et al. [158] assessed and visualized the *customer satisfaction* in mobile service using a combination of SA and VIKOR approaches. Experiments were performed on 1,487 reviews from AppStore HQ collected on 8 mobile application services i.e. alternatives. Dictionary based approach (proposed in [13]) was utilized for sentiment score calculation with respect to common eight aspects or criteria of mobile applications. Here, VIKOR multi-criteria decision making method was utilized for ranking of different mobile services. Maximum group utility and individual regret values were plotted to visualize customer satisfaction towards various alternatives and criteria. In future, basic SA technique should be replaced by some more advanced techniques. Comparative analysis with some other multi-criteria decision making methods can be carried out.

3.7.4 Recommender System

Li and Shiu [43] developed a *recommender system* for social advertising over microblogs. They exploited various features like structure of relationship, content popularity, social activeness, social interactions, social similarity, click through rate etc. These features helped them to get a list of targeted users for each endorser at the current stage, and suitable paths for information diffusion. On the basis of that, they produced a recommended list of targeted users for endorsing an advertisement. For the experiment, they collected 247,099 users and their 1,969,253 plurks, responses and interactions. Incorporating other tangible factors and dynamic modules into the mechanism may improve the quality of the system.

Garcia-Cumbreras et al. [79] projected a novel application of SA in recommender system by categorizing users according to the average polarity of their comments. At first they justified the relation between comments and ratings. And, they incorporated this knowledge to the recommender system. Two groups of users, optimists and pessimists, have been created for collaborative filtering. SVM outperformed on K-nearest neighbor (KNN) for rating calculation and KNN with k=80 has yielded the least error for rating prediction. For the experiment, they collected user's ratings and user's comments on 2,713 movies, 54,112 users, and 80,848 opinions from IMDb. The proposed system can be improved with a higher level of integration, changing the between-items and between-users similarity in the core of collaborative filtering algorithms. Bao et al. [49] proposed a temporal and social probabilistic matrix factorization model to predict user's future interests. They integrated users' friendships and users' historical interests in micro-blogging sites. They utilized Sina-wiebo posts on 2,788 topics by 1,170 users posted during 15 days, where the last day's data was utilized for testing the model. Automatic tuning of parameters like number of users can be performed using MCMC algorithm. Some other effective factors can enhance the performance of the proposed model like re-tweeting relationship and discussions among users.

http://dx.doi.org/10.1016/j.knosys.2015.06.015

3.7.5 Marketing Intelligence

Market intelligence is designed to fulfill four needs of business managers like (a) Opportunities and threat determination, (b) Knowing competitors, (c) Help preempt competitors' actions, and (d) aid effective marketing decision making. In this regard, Li and Li [90] framed a framework for market intelligence with the help of different tasks like trendy topic detection, opinion classification, credibility assessment, and numeric summarization. The first task was performed by assigning topic tendency score with the help of term frequency, inverse document frequency, and patterns appear in a term. The second task was performed in two phases (a) subjectivity classification based on self developed lexicon (b) sentiment classification using SVM with RBF Kernel and NB with emoticons, unigram, bigram, unigram + bigram, and subjective word set as feature representation. Third task was credibility calculation based on follower-followee ratio. Finally, semantic score and the credibility score have been aggregated under numeric summarization task. For the experiment, follower and followee relationship and tweets on three brands viz. Google, Microsoft, Sony, for three products namely iPhone, iPad, and Mac book have been collected from Twitter. For the training, 11,929 tweets have been collected with two emoticons ":)" and ":(", SVM and NB have been trained on 9,165 preprocessed tweets. SVM outperformed NB and unigram feature representation dominated other alternatives. So, improvement can be made by considering meronym patterns, PageRank consideration for the credibility assessment, etc.

3.7.6 Other applications

Fortuny et al. [47] analyzed the *media coverage in time of political crisis* on various topics in Belgium. They analyzed 68,000 articles covered in 8 newspapers. Pattern mining module of Python was utilized as subjective lexicon containing 3000 Dutch adjectives with polarity values. To calculate sentiment score for each opinion word related to political party, a window of 5 sentences have been analyzed. The frequency of occurrence and the tone of articles have been considered for bias and SA. The amount of ranking difference of each newspaper was calculated using Hamming distance, which gives the degree of disagreement between the consensus rankings. Sentiment of newspapers towards a political party was visualized, where 30-40% coverage was negative. The proposed approach can be tried for different dataset. Moreo et al. [13] devised a linguistic modularized model with low-cost adaptability for *news focus detection*. It can deal with views expressed in non-standard language and target detection in a multi-domain scenario. They created a taxonomy of objects and features by analyzing 250 news items. Here, each word has been characterized using polarity and strength. They divided the whole task into three parts (a) preprocessing, (b) focus identification and sentiment classification, and (c) summary mining. For focus identification, they considered the most recurrent discussion topics. Sentiment classification module exploited colloquial language and multi-word expressions. Experiments were performed on randomly selected 500 current news items on different topics achieving an accuracy of 89%. Though the proposed taxonomy is in initial stage, it can be easily extended with the concepts of new corpora.

Mohammad [64] studied the *distribution of emotion words in different texts* like e-mails (love letters, hate mail, and suicide notes), books, fairy tales, novels, etc. For this, he exploited the Roget thesaurus to create word-emotion association lexicon NRC Emotion Lexicon. Relative Salience (RS) was calculated to compare emotion expressed by different terms. RS is a ratio between the frequencies of words associated with a particular emotion to the total number of emotion words in whole text. Four e-mail corpora (a) love letters corpus (LLC) v 0.1[40], (b) hate mail corpus (HMC) v0.1[41], (c) the suicide notes corpus (SNC) v 0.1[42], d) Enron email corpus[43] were exploited. For novels and fairy tales Corpus of English Novels (CEN) and the Fairy Tale Corpus (FTC)[44] have been exploited. This study can help users to understand writing style of different authors in emotional point of view. Li and Wu [91] performed online *forums hotspot detection* and forecast in two stages. At the first stage, emotional polarity was computed. At the second stage, K-means clustering and SVM have been applied to group the forums into various clusters. For the experiment 220,053 posts have been collected on 31 different topics from Sina sports forums[45]. To determine the polarity of the forums, HowNet[46]; a Chinese sentiment dictionary have been exploited. With the help of the dictionary eight word lists with different emotional intensifiers have been created: positive, negative, privatives, and five lists of modifiers. K-means clustering have been applied with K=31, i.e. 31 leaf forums, while SVM has been applied in the sliding time window manner. For each SVM, the input was a matrix containing 31 leaf forums' representation vectors, and the output was a vector containing 31 integer values either 1 (hotspot) or 0 (non-hotspot). SVM achieved highly consistent results with K-means clustering. And, the top 10 hotspot forums listed by SVM forecasting resembles 80% of K-means clustering results. In future, topic extraction is to be included to pop the question what event or topic triggered the user attention.

Desmet and Hoste [58] performed *emotion detection in suicide notes*. They detected fifteen types of emotions expressed in suicide notes. Experiments were performed on 1319 suicide notes collected between 1950 and 2011 (total 900 notes: 600 for training and 300 for testing). SVM was used for binary classification. Bootstrap sampling was performed to determine threshold to maximize F-score for each classifier. Suicide note content analysis can help in suicide prevention and forensic linguistics and can be applied for prediction of suicidality in the text. Favorability Analysis (FA) determines *how favorable an article* is with respect to an entity and *how much a client interested* in that entity. FA determines

[40] LLC: http://www.lovingyou.com/content/inspiration/loveletters-topic.php?ID=loveyou.
[41] HMC: http://www.ratbags.com.
[42] SNC: http://www.well.com/art/suicidenotes.html?w.
[43] http://www-2.cs.cmu.edu/enron.
[44] CEN:https://perswww.kuleuven.be/~u0044428/cen.htm. FTC:https://www.l2f.inesc-id.pt/wiki/index.php/Fairy_tale_corpus.
[45] http://bbs.sports.sina.com.cn/treeforum/App/list.php?bbsid=33&subid=0.
[46] http://www.keenage.com/download/sentiment.rar

http://dx.doi.org/10.1016/j.knosys.2015.06.015

sentiment polarity along with favorable objective mentions of entities. Lane et al. [73] applied opinion mining for favorability analysis (FA). Data were represented by unigrams, bigrams, trigrams, entity words and dependency words. Data imbalance was managed at training time using under sampling and at evaluation time by modifying the output threshold. Three datasets collected from newspapers and magazines were utilized for the experiment; out of them two were on high-tech companies and one was on a charity trust. Different classifiers viz. NB, SVM, JRip, RBF, and Random were trained with different settings. Experiments were performed for pseudo-subjectivity and pseudo-sentiment classification tasks. For pseudo-subjectivity, SVM outperformed all other alternatives with cross validation accuracy of 91.2%. For pseudo-sentiment NB was reported as the best classifier. In future, proposed approach can be applied on social media.

4. Discussion

This study reviews many interesting and useful works regarding the state-of-the-art in SA. The paper is organized on the basis of tasks, approaches and applications of SA as presented in Figure 1. If we consider granule based sentiment analysis, most of the studies focused on document and sentence level as presented in Table 1. Here, document refers to product reviews, blogs, forums, biography, news, news comments, tweets, plurks, Facebook and Sina-Wiebo comments etc. Concept refers to a class or category in ontological engineering. Phrase is a special combination of two or more words. Link based studies were carried out in opinion extraction from social networking while clause level studies dealt with conditional sentences. In order to compare two or more products we need to perform fine-grained sentiment analysis. Table 1 indicates that more investigation is required at finer-grained level SA.

Subjectivity classification had been addressed in very few studies as presented in Table 2 and Table 3. Out of various sub-tasks of sentiment analysis, subjectivity classification is more challenging task than sentiment classification. The highest subjectivity classification accuracy achieved on Pang and Lee [167] dataset was 92.1% by Xuan et al. [174], which was minor improvement over 92% by Pang and Lee [167] as presented in Table 4. Here, Xuan et al. [174] followed lexicon based approach and Pang and Lee [167] adopted machine learning based approach. Therefore, we infer that both approaches are competitive to each other for subjectivity classification. Recently, Maks and Vossen [24] came up with different kind of subjectivity classification such as *epistemic subjectivity, speaker subjectivity*, and *character subjectivity.* They employed a professional linguist to detect such kind of subjectivity. For this purpose, machine learning can be employed to automate the process.

Table 3 indicates that majority of works have been carried out in sentiment classification. In polarity determination, machine learning based approach [156] dominated other hybrid [112] and lexicon based approaches [169] as presented in Table 5. Similarly, for Pang et al. [33] dataset, hybrid approach [109] dominated lexicon based approaches [165]. For sentiment classification, various types of syntactic, semantic, and statistical approaches were employed to select most relevant features. These approaches have been proved to be powerful and consequently, sentiment classification accuracy has improved to 92% [109] from 88.9% [33]. A lot of work still remains to be done for polarity determination. Moreover, we found very limited amount of study on other sentiment classification sub-tasks like vagueness resolution in opinionated text, multi-lingual and cross-lingual sentiment classification, and cross-domain sentiment classification. Machine translation based approaches are successful for cross-lingual sentiment classification, which requires some extra effort and time [148, 173]. These sub-tasks need to be investigated at a wider scale in order to have global discussion on common issues and exchange of views worldwide.

In the cut-throat competition of e-commerce industry, nowadays, customer reviews are playing a quintessential role. The promotion of quality product can be affected by bad quality reviews and opinion spam. Projecting quality reviews, whether positive or negative, can save customers' time and effort to get quality products. For review helpfulness measurement, various features like lexical similarity features, shallow syntactic features, helpfulness votes, subjectivity clues have been exploited. From the current study, we found that machine learning models yielded better results than vote count based measures. Moreover, content based review quality measurement is more reliable than vote counts. As regards lexicon creation, non-standard text and spelling are in rampant use for sentiment expression [116, 117, 118, 127]. At another level, sentiment lexicons were constructed using ontology based approaches [21, 23, 30, 52, 106, 120 and 199] to capture semantic relationships between opinion words and represent commonsense knowledge [95, 106, and 120]. Aspect level SA requires capturing the opinion effect of different aspects at various levels. Therefore, some initiatives were taken to address hierarchical relationship among product and its aspects [186, 195, and 197]. Further, ontology proved to be useful for the same purpose [35, 62, 182, 194, 195, and 197].

From Table 6, it is glaringly evident that SVM is the most often applied technique for various tasks. SVM yielded the best accuracy in most of the studies, except in [52, 54, 57, and 197]. This is because SVM is eminently suitable for solving high dimensional problems compared to other techniques. However, there are a few exceptions as follows: Neural network outperformed SVM in [52, 54, and 57], while Bayesian boosting outperformed SVM in [197]. As regards kernel of SVM, linear kernel was found to be widely employed except in [29, 67, 90, 97, and 155], where, Gaussian kernel was employed in [29]; Polynomial kernel was used in [67, 97] and RBF kernel was used in [90, 155]. Further, a tree kernel was developed in [177]. Naive Bayes with little modification performed well for document classification [44, 121]. A sentiment lexicon is quite relevant for sentiment classification. It yielded promising performance for sentiment analysis [13, 158]. Among non-parametric approaches, modified latent dirichlet allocation

http://dx.doi.org/10.1016/j.knosys.2015.06.015

(LDA) has been mostly employed for aspect extraction. This helps us draw several conclusions as presented in the *conclusions* section.

Table 6. Distribution of articles based intelligent techniques applied

Applied Techniques	#Articles	Articles' References
SVM	55	[21], [26], [29], [33], [44], [45], [46], [50], [51], [53], [54], [57], [58], [66], [67], [73], [76], [77], [86], [88], [90], [91], [94], [95], [97], [101], [108], [109], [111], [114], [116], [118], [125], [131], [148], [157], [160], [163], [165], [167], [169], [172], [176], [177], [183], [195], [197], [200], [209], [210], [212], [214], [225], [228], [240]
Dictionary based approaches (DBA)	41	[13], [18], [23], [24], [25], [35], [36], [47], [55], [64], [67], [68], [69], [85], [96], [110], [112], [117], [126], [127], [158], [170], [171], [175], [183], [193], [194], [195], [196], [202], [203], [206], [207], [209], [210], [213], [216], [218], [220], [229], [241]
NB	28	[33], [46], [50], [54], [56], [73], [80], [86], [90], [94], [98], [101], [111], [114], [116], [118], [121], [124], [125], [131], [148], [156], [163], [167], [197], [209], [217], [228]
NN	11	[48], [50], [51], [52], [53], [57], [76], [101], [116], [213], [226]
DT	9	[53], [66], [73], [76], [86], [94], [116], [118], [209]
Maximum Entropy	8	[33], [46], [54], [60], [63], [66], [148], [156], [174]
Logistic Regression	9	[53], [77], [88], [99], [116], [118], [163], [197], [220]
Linear Regression	8	[8], [18], [70], [75], [222], [223], [224], [231]
Ontology	8	[30], [41], [62], [98], [182], [194], [195], [196]
LDA	8	[61], [92], [107], [185], [189], [191], [182], [240]
Random Forest	4	[77], [81], [228], [210], [228]
SVR	5	[118], [123], [130], [155], [227]
CRF and rCRP	5	[37], [88], [93], [186], [190]
Boosting	4	[12], [118], [179], [197]
SVM-SMO	4	[76], [97], [118], [169]
Fuzzy Logic	3	[23], [41], [62], [213]
Rule Miner	4	[37], [100], [112], [217]
EM	3	[56], [59], [155]
K-medoids	1	[52]
RBF NN	1	[130]

Table 7 presents year wise distribution of articles considered for this systematic literature review. From Table 7, it is quite evident that that research in SA proliferated during 2011-2015. The number of articles published in some notable journals is presented in Table 8, which indicates that Journals such as "Expert Systems with Applications" and "Decision Support Systems" published maximum number of research papers since research in SA became popular. Table 9 presents a list of available public dataset for sentiment analysis. Table 10 listed some attributes of one hundred and sixty one articles which were considered for review. The first column contains the reference number. The second column represents techniques utilized, while the third column represents the degree of polarity (if applicable) considered, the fourth column shows the language of corpora exploited for experiments, the fifth column shows the name of the corpora, and the last column presents the name of lexica (if utilized). Some abbreviations used in the Table 10 are: English (E), Dutch (D), Chinese (C), Spanish (S), Turkish (T), Cantonese (CT), Romanian (R), Belgian (B), Arabic (A), Japanese (J), Italian (I), French (F), German (GE), Product Reviews (PR), Movie Reviews (MR), Micro-blog (MB), Global domain (G), Restaurant Reviews (RR), Dictionary Based Approach (DBA), SenticNet (SN), WordNet (WN), ConceptNet (CN), WordNet-Affect (WNA), and Bing Liu Opinion Lexicon (OL).

Table 7. Year wise distribution of articles

Year	#Articles	Articles' References
2002	2	[32], [33]
2003	1	[171]
2004	3	[167], [172], [173]
2005	2	[12], [169]
2006	2	[170], [178]
2007	1	[168]
2008	4	[130], [160], [180], [196]
2009	8	[25], [29], [46], [77], [88], [112], [121], [132]
2010	5	[45], [91], [116], [166], [223]
2011	25	[18], [26], [30], [44], [48], [54], [55], [63], [69], [74], [78], [81], [89], [93], [109], [114], [117], [156], [175], [176], [177], [179], [212], [225], [231]
2012	27	[8], [11], [13], [24], [27], [37], [43], [47], [56], [60], [61], [62], [64], [73], [86], [92], [95], [111], [118], [125], [154], [155], [162], [174], [221], [222], [229]
2013	42	[21], [22], [23], [40], [41], [49], [50], [51], [52], [53], [57], [58], [59], [66], [68], [70], [76], [80], [85], [90], [94], [96], [97], [98], [99], [100], [101], [107], [110], [123], [124], [126], [127], [131], [148], [157], [158], [185], [186], [200], [218], [224]
2014	33	[35], [36], [67], [75], [108], [163], [165], [182], [183], [184], [189], [190], [191], [193], [194], [195], [197], [202], [203], [205], [206], [207], [209], [210], [211], [213], [214], [216], [217], [220], [226], [227], [232]
2015	5	[199], [228], [240], [241], [243]

Table 8. Number of articles published (and reviewed here) in different journals

S#	Name of Journals	#articles
1	Expert Systems with Applications	33
2	Decision Support Systems	28
3	Knowledge-based systems	17
4	IEEE Intelligent Systems	12
5	IEEE Transactions on Knowledge and Data Engineering	6
6	IEEE Transactions on Affective Computing	3
7	Information Sciences	3
8	Information Processing and Management	3
9	Computer Speech and Language	2
10	Communications of the ACM	2
11	Journal of Computer Science and Technology	2
12	Journal of Informetrics	2
13	Information Retrieval	2
14	Computer Speech and Language	2
15	Inf. Retrieval	1

Table 9. List of publicly available datasets

S#	Data Set	Type	Lang.	Web Resource	Details
1	Stanford large movie data set	Movie Reviews	English	http://ai.stanford.edu/~amaas/data/sentiment/	Movie Reviews
2	COAE2008	Product Reviews	Chinese	http://ir-china.org.cn/coae2008.html	2739 documents for movie, education, finance, economics, house, computer, mobile phones etc. 1525 +ve, 1214 -ve
3	Boacar	Car Reviews	Chinese	http://www.riche.com.cn/boacar/	11 type of car TradeMarks and total review 1000 words, having 578 POS, 428 -ve reviews
4	[187]	Reviews, forums	English	http://sifaka.cs.uiuc.edu/~wang296/Data/	Accessed: 27 Aug, 2014
5	[188]	Reviews	English	http://uilab.kaist.ac.kr/research/WSDM11	Aspect oriented dataset. Accessed: 18 Dec, 2014
6	Movie-v2.0	Movie Reviews	English	http://www.cs.cornell.edu/people/pabo/movie-review-data/	Data size:2000 Positive: 1000 Negative: 1000
7	Multi-domain	Multi-domain	English	http://www.cs.jhu.edu/~mdreze/datasets/sentiment	
8	SkyDrive de Hermit Dave	Spanish Word Lists	Spanish	https://skydrive.live.com/?cid=3732e80b128d016f&id=3732E80B128D016F%213584	
9	TripAdvisor	Reviews	Spanish	http://clic.ub.edu/corpus/es/node/106	18000 customer reviews on hotels and restaurants from Hopinion

http://dx.doi.org/10.1016/j.knosys.2015.06.015

10	[38]	Multi-Domain	English	www2.cs.uic.edu/~liub/FBS/sentiment-analysis.html	6800 opinion words on 10 different products
11	TBOD [144]	Reviews	English		Product Review on Cars, Headphones, Hotels
12	[68]	Product Reviews	English	http://www.lsi.us.es/_fermin/index.php/Datasets	Product Reviews from Epinion.com on headphones 587 reviews, hotels 988 reviews and cars 972 reviews.
13	[148]	Movie Reviews	Turkish	http://www.win.tue.nl/~mpechen/projects/smm/#Datasets	5331 positive and 5331 negative reviews on movie
14	[148]	Product Reviews	Turkish	http://www.win.tue.nl/~mpechen/projects/smm/#Datasets	700 +ve &700 -ve reviews on books, DVD, electronics, kitchen appliances
15	ISEAR	English sentences	English	www.affective-sciences.org/system/files/page/2636/ISEAR.zip	The dataset contains 7,666 such statements, which include 18,146 sentences, 449,060 running words.
16	[149]	Product Reviews	English	http://www.cs.jhu.edu/~mdredze/datasets/sentiment/	Amazon reviews on 4 domain (books, DVDs, electronics, kitchen appliances)
17	DUC data, NIST	Texts	English	http://www-nlpir.nist.gov /projects/duc/data.html, http:// www.nist.gov/tac/data/index.html	Text summarization data
18	[70]	Restaurant and Hotel Reviews	English	http://uilab.kaist.ac.kr/research/WSDM11	Restaurant and Hotel Reviews from Amazon and Yelp
19	[114]	Restaurant Reviews	Cantonese	http://www.openrice.com	Reviews on restaurant
20	[125]	Biographical Articles	Dutch	http://www.iisg.nl/bwsa	574 Biographical articles
21	Spinn3r dataset	Multi-Domain	English	http://www.icwsm.org/2011/data.php	
22	[86]	Ironic Dataset	English	http://users.dsic.upv.es/grupos/nle/	3163 ironic reviews on five products
23	HASH [179]	Tweets	English	http://demeter.inf.ed.ac.uk	31,861 Pos tweets, 64,850 Neg tweets, 125,859 Neu tweets
24	EMOT [179]	Tweets and Emoticons	English	http://twittersentiment.appspot.com	230,811 Pos & 150,570 Neg tweets
25	ISIEVE [179]	Tweets	English	www.i-sieve.com	1,520 Pos tweets, 200 Neg tweets, 2,295 Neu tweets
26	[177]	Tweets	English	e-mail: apoorv@cs.columbia.edu	11,875 tweets
27	[52]	Opinions	English	http://patientopinion.org.uk	2000 patient opinions
28	[96]	Tweets	English	http://goo.gl/UQvdx	667 tweets
29	[39]	Movie Reviews	English	http://ai.stanford.edu/~amaas/data/sentiment/	50000 movie reviews
30	[164]	Tweets	English	http://cs.stanford.edu/people/alecmgo/trainingandtestdata.zip	
31	[210]	Spam Reviews	English	http://myleott.com/op_spam	400 deceptive and 400 truthful reviews in positive and negative category. Last Accesed by: 12 April, 2015
32	[230]	Sarcasm and nasty reviews	English	https://nlds.soe.ucsc.edu/iac	1,000 disscussions, ~390,000 posts, and some ~73,000,000 words

Table 10. Summary of reviewed articles

Ref.	Concepts and Techniques Utilized	P	L	Type of data	Dictionary
[8]	Page rank, Gradient descent, Linear regression	2	E	PR	
[11]	Link mining, Collective classification	NA	E	MB	
[12]	AdaBoost.HM	2	E	G	GI
[13]	DBA	5	E	News Comments	New Lexicon
[18]	DBA, SOFNN, Linear regression	2,7	E	MB, DJIA data	OF, GPOMS
[21]	Regression, Random walk, SVM	4,2	E		ANEW, CN
[22]	Cohen's K coefficient	6,2	I	MB	SN
[23]	Fuzzy clustering, PMI, DBA	6,2	E	G	WNA, SN, WN.
[24]	DBA	NA	D	G	Dutch WN
[25]	Association Miner CBA, DBA	2	E	PR	WN
[26]	SVM	2	E	PR	
[27]	Markov-Chain Monte Carlo (MCMC)	NA	E	Online discussion	
[29]	SVM with Gaussian Kernel	3, 2			MPQA

[30]	Ontology, K-means	2	E		ReiAction[47], Family Relation[48]
[32]	PMI-IR	2	E	Multi-domain	
[33]	NB, SVM, ME	2	E	MR	
[35]	Ontology, DBA	2	E	MR	SWN
[36]	New Algorithm, DBA	2	E	MR, Book, Mobile	11 dictionaries
[37]	CRF	NA		PR	
[40]	Multinomial inverse regression	3	E	MB	
[41]	FFCA, Lattice	2	E	PR	
[43]	Analytic hierarchy process	NA	C	MB	
[44]	Fisher's discriminant ratio, SVM	2	C	PR	
[45]	Semantic orientation, SVM	3, 2	E	PR	SWN
[46]	MNB, ME, SVM	3, 2	E, D, F	Forum, Blog, PR	
[47]	DBA	2	D, E	News	
[48]	Semantic orientation and BackProp	2	E	Blogs, PR	
[49]	Probabilistic Matrix Factorization	NA	C	MB	
[50]	NB, SVM, NN	2	E	PR	
[51]	SVM, NN	NA	C	MB	
[52]	DNN, CNN, K-medoids, KNN	NA	E	G	CN, WNA, AffectiveSpace
[53]	SVM, NN, MLP, DT, GA, Stepwise LR, RBC	2	E	News	
[54]	NB, ME, SVM	2	E	PR	
[55]	DBA	5, 2	E	MB	
[56]	NB, EM	NA	E	PR	WN
[57]	SVM, NN	5, 2	E	MB	
[58]	SVM	NA	E	Suicide Notes	WN, SWN.
[59]	EM	NA	E, S	PR	fullStrengthLexicon[49]
[60]	ME	NA	E	MB	
[61]	Bayesian Model, LDA	2	E	PR	MPQA, Appraisal Lexicons[50]
[62]	Fuzzy Set, Ontology	2	C	PR	
[63]	ME, Bootstrapping, IG	3, 2	C	PR	Hownet, NEUCSP[51]
[64]	DBA	NA	E	e-mail, book	Roget Thesaurus[52]
[66]	NB, ME, DT, KNN, SVM	NA	C, E	PR, Forums	
[67]	SVM, DBA	2	E	PR	GI
[68]	DBA, Random walk algorithm	2	E	PR	
[69]	DBA	2	E	PR	
[70]	Linear Regression	NA	C	PR, social network	
[73]	BayesNet, J48, Jrip, SVM, NB, ZeroR, Random	5, 2	E	News, Magazine	
[74]	Semantic relationships	2	E		SWN, GI
[75]	Multilingual bootstrapping and cross-lingual bootstrapping, linear regression, IG	NA	E, R		WN
[76]	Bootstrapping, DT, MLP, PCA, SLR, SMO-SVM	2	E	Phone Reviews	WN
[77]	LR, SVM, RF	2	B	e-mails	
[78]	Discretionary accrual model	NA	E	Book Reviews	
[80]	Bayes-Nash equilibria	NA	E	MB	
[81]	RF	NA	E	PR	
[85]	DBA	3, 2	E	MB	SWN
[86]	Semantic, NB, SVM, DT	NA		PR	WN, MSOL, WNA
[88]	SVM, LR, CRF	NA	E	PR	
[90]	SVM, NB	NA	E	MB	
[91]	K-means, SVM	NA	C	Forums	
[92]	HMM-LDA	NA	E	PR	
[93]	Two level CRF	NA	E	PR	
[94]	Corpus based approach, SVM, NB, C4.5, BBR	5, 2	E, S	PR	SWN, Tree Tagger
[95]	SVM	NA	E		WNA, LIWC, VerbOcean, CN
[96]	DBA, Ontology	2	E	MB	
[97]	SMO-SVM, DBA	2	E	MR	SWN, WN
[98]	NB and Ontology	2	E	PR, MR	WN
[99]	Cosine similarity, L1 regularized logistic	2	E	PR	WN and SWN

[47] http://cyc.com/cyc/opencyc/overview.
[48] http://www.ontologyportal.org/index.html.
[49] https://github.com/maksim2042/snowwhite/blob/master/snowwhite/data/fullStrengthLexicon.txt.
[50] http://lingcog.iit.edu/arc/appraisal_lexicon_2007b.tar.gz.
[51] NEUCSP is a Chinese word segmentation and POS tagging tool at (http://www.nlplab.com/chinese/source.htm).
[52] http://thesaurus.com/Roget-alpha-index.html.

	regression				
[100]	Association miner CBA	NA	C	PR	
[101]	NN, C4.5, CART, SVM, NB	2	E	MB	
[102]	SVM	2	C	HR, PR	TU lexicon[53]
[107]	LDA, DBA	2	E	RR, HR	MPQA, SWN
[108]	SVM	2	A	Dialects, MB, Wiki Talks, Forums	
[109]	Rule-based multivariate features, SVM	2	E	MR, PR, Automobile	
[110]	DBA	2	S	MR	BLEL, WN
[111]	NB, SVM	2	E	RR	SWN
[112]	DBA, RBC, SVM	2	E	MR, Product, MySpace texts	WN, GI
[114]	IG, DBA	2	CT	RR	
[115]	SVM, Statistical approach	2	E, C	HR, Mobile	
[116]	DBA, SVM, NB, LR, J48, Jrip, AdaBoost, Decision Table, MLP, NB.	2	E	MySpace	SentiStrength
[117]	DBA	2	E	MB	SWN
[118]	SMO-SVM, LR, AdaBoost, SVR, DT, NB, J48, Jrip	2	E	Social Media	SentiStrength
[121]	Adaptive-NB	NA	C	PR	
[123]	SVR	6, 2	C	Sina-Wiebo	
[124]	NB	2	E	Social & Mass media	
[125]	Lexical features, NB, Linear SVM, Jrip, K-NN	2	D	Biographies	Brouwers thesaurus
[126]	DBA	2	E	MB	OL
[127]	DBA	5, 2	E	G	SentiStrength
[130]	SVR, RBF	NA			
[131]	SVM, NB	3	E	MB, PR	
[132]	New Algorithm	NA		PR	
[148]	SVM, NB, ME	2	E, T		
[154]	New algorithm, Lexical features	3	E	PR	
[155]	SP-LSA, AR, EM, ε-SVR	2	E	MR	2,030 appraisal words
[156]	Tabu search, MB, NB, SVM, ME	2	E	MR and News	
[157]	PSO and SVM	2	E	MB	
[158]	DBA	3, 2	E	Mobile Reviews	Moreo et al. [13]
[160]	EWGA, SVM, Bootstrapping	2	E, A	Forums	
[162]	Class sequential rules	3	E	MR	SWN
[163]	DBA, SVM, NB, Logistic, NN	2	E	MB	10 dictionaries
[165]	Semantic, GI, Chi-square, SVM	2	E	MR and PR	
[166]	Semantic	2	C	HR	
[167]	NB, SVM, Min.-cut in the graph	2	E	MR	
[168]	Linear classifiers, Clique, MIRA classifier	2	E	PR	
[169]	DBA, SVM, and SMO-SVM	2	E	MR	WN
[170]	DBA	3	J	MR and PR	Yi et al. [7] lexicon
[171]	DBA	2	E	Web pages, News	
[172]	SVM, Osgoodian values, PMI	2	E	MR	WN
[173]	Transfer-based machine translation	2	J	Camera Review	
[174]	ME	2	E	MR	
[175]	DBA, Sigmoid scoring	2	C	Blogs	Hownet
[176]	SVM, PMI	2	E	MB	GI
[177]	Convolution kernels [152], SVM, DBA	2, 3	E	MB	WN, DAL [151]
[178]	Statistical method of OASYS [8]	C	E	News articles	OASYS
[179]	Boosting, SVM	3	E	MB	MPQA, NetLingo
[180]	Bipartite graph, Regularization operator	2	E	Blogs	
[182]	LDA, Ontology, MCMC	2	E	Multi-domain	OF
[183]	SVM, TF-IDF	2	E	News headlines, Forex Rate	SWN
[184]	Vector space model	3	E	News articles	Harvard IV
[185]	Modified LDA	5	E	PR	
[186]	Recursive Chinese Restaurant Process	2	E	PR	
[189]	LDA incorporated with domain knowledge	NA	E	Camera and HR	
[190]	CRF, syntactic and semantic features	2	E	PR, Facebook text	
[191]	LDA, Appraisal expression pattern	NA	E	HR, RR, PR	
[192]	PMI, TF-IDF	2	E	PR	GI
[193]	TF-IDF, Domain relevance	2	C	HR, Cellphone	

[53] Available at: http://nlp.csai.tsinghua.edu.cn/_lj/downloads/sentiment.dict.v1.0.zip.

[194]	Ontology	2	E	Automobile, PR, SW	SWN, GI, OL
[195]	Ontology	2	E	MR	WN
[196]	Ontology, Maximum-Likelihood	2	E	MR	GI
[197]	PCA, SVM, LR, Bayesian Boosting, Bagged SVM	2	E	PR	
[200]	SVM	2	E	PR	
[202]	DBA, Graphical Techniques	2	E	G	CN, DBPedia, WN
[203]	DBA	2	E	MB	CN, WN, JMDict, Verbosity
[205]	Graphical techniques	2	GE	MB	SWN, SN 3
[206]	DBA	8	E	Google n-grams	SN 3, WNANRC, SAT
[207]	Ontology, DBA	4	E	PR, MR	CN
[209]	SVM, NB, J48	3	S	Facebook text	Spanish LIWC
[210]	SVM, RF	3	S	Apontador	
[211]	DBA	2	S	MB	SN 3, WeFeelFine
[212]	NB, SVM, DBA	2	E	PR	LIWC
[213]	Ontology, DBA, ELM	2	E	G	AffectiveSpace
[214]	Ontology, DBA, SVM, FCM	2	E	G	SN 3, WNA, AffectiveSpace
[216]	DBA, Ontology	2	E	PR, MR	WN, CN
[217]	Rule base classifier, NB	2	E	Dialogue	SN 3
[218]	Bootstrapping, PMI, DBA	NA	E	PR	
[220]	DBA, Binomial LR	NA	E	PR	LIWC
[221]	Product, Review & Reviewer Information	NA	E	PR	
[222]	Linear Regression	2	E	PR	
[223]	Linear Regression	NA	E	PR	
[224]	Linear Regression	NA	E	PR	
[225]	SVM	NA	E	PR	
[226]	MLP	NA	E	PR	
[227]	RFM, SVR	NA	E	PR	
[228]	RF, NB, SVM	NA	E	PR	
[229]	DBA	2	E	PR	
[231]	Linear Regression	NA	E	PR	
[232]	PU-learning	NA	E	PR	
[240]	LDA, SVM, PMI	NA	C	PR	
[241]	PageRank algorithm, DBA	NA	C	PR	
[243]	PMI-IR, RCut, Apriori Algo.	NA	C	PR	

4.1 Open issues/future directions

We now list out the open issues in several aspects of sentiment analysis.

- Data collected from various resources are often so much noisy, wrongly spelt and unstructured. There is no automatic system for spelling correction and noise removal. Preprocessing step consumes maximum time to convert raw data into a structured format. For instance, tweets and plurks often contain abbreviations with high percentage. Even though we have a sort of acronym dictionary like netlingo[54], twittonary[55], urbandictionary[56] etc. to deal with acronyms, they do not contain emerging slangs. So, data preprocessing is the most time consuming activity and prone to less accuracy.
- There is a lack of universal opinion grading system across sentiment dictionaries.
- Online discussion and political discussions often contain irony and sarcastic sentences [181]. Moreover, Irony and sarcastic words vary from language to language. Little or no studies have

[54] http://www.netlingo.com/.
[55] http://www.twittonary.com/.
[56] http://www.urbandictionary.com/.

been devoted in this regard. In order to deal with irony and sarcasm detection, more computational approaches are needed to be developed on the basis of appraisal theory.

- For better product comparison, we should compare a set of products with respect to their common aspects (also known as product features). For that, we need to identify aspects discussed in the given text and explicit opinionated sentences written for the corresponding aspect. Topic modeling based approaches address this issue to some extent but it requires bigger corpora to be trained on. Therefore, a lot of work remains to be done along this line.
- Machine learning provides automated learning systems to get insights of data. Then, ontology based studies proved to be successful in representing, visualizing, and determining sentiment units [35, 96, and 182]. Exploitation of the use of ontology can resolve scalability and vagueness issues in sentiment analysis. Thus applying machine learning together with ontology on linked data is still an open research problem.
- A very few attempts were made to utilize the potential of optimization techniques for feature selection. Therefore, various hybrids of machine learning and optimization techniques can be developed for feature subset selection.
- There is a lack of opinion mining system in non-English languages. A great deal of study has been carried out in SA in English language. Consequently, substantial amount of resources have been generated for the English language. Thus, these resources can be mapped to other languages in order to perform cross-lingual sentiment analysis.
- Cross-domain SA is still a major challenge. Cross-domain SA needs to address three issues: the first issue, opinion expressed for one domain will be reverted for other domain. For instance, polarity of a sentence e.g. "Screen is curved." is positive for TV but negative for mobile. The next issue is the difference in sentiment vocabularies across different domains should be considered, and the third issue is to objectively assign a strength marker to each sentiment word.
- Aspect level SA is very much required for comparative visualization of similar kind of products.
- The main challenge lies in review helpfulness is the validation of the proposed method. In order to validate the proposed method, researcher should have access to sale information about the product along with date of sale. Such kind of information can be very useful to correlate the helpfulness of reviews to sale of the product, estimation of effect of reviews on sale of products.
- The lack of proper review spam dataset is a major issue in order to perform opinion spam detection. Because AMT developed dataset and artificially synthesized dataset cannot contain psycholinguistic features like an original spammer. Therefore, comparison of effectiveness of two or more techniques is barely reliable.

- Finally, some more open issues are contextual SA, intrinsic feature based SA, SA on Facebook posts, sense level subjectivity classification, SA of streaming text, and suitable automatic SA system with respect to different media.

4.2 Other Possible Applications

In addition to the existing applications studied in the literature, we foresee possible applications of SA at two levels viz. global level and business level. The former includes rumor detection, SA on streaming data, study the trend of sentiment propagation over different media on some special event like general election, and study the flow of emotion during chatting etc. The latter involves celebrity recommendation for specific brand, comparison of celebrity popularity, measurement of celebrity effect on the sale of products, decision making for candidates based on recommendation made by previous employer, and appraisal preparation for an employee based on customer feedback in service industry (a customer-centric approach) etc.

5. Conclusions

This paper presents a comprehensive, state-of-the-art review on the research work done in various aspects of SA during 2002-2014. The paper is reviewed in six broad dimensions viz. subjectivity classification, sentiment classification, review usefulness measurement, lexicon creation, opinion word and product aspect extraction, and various applications of opinion mining. These six dimensions refer to *tasks* to be accomplished for SA. All tasks and sub-tasks are reviewed in four aspects viz. problem addressed, dataset exploited, features selected (if applicable), approaches and techniques employed, result, and indicated future directions by author and/or us. We draw some important conclusions from this review paper regarding the applied techniques. Apart from SVM, NN and lexicon based approaches; we found that some of the intelligent techniques have not been exploited exhaustively like random forest, evolutionary computation, association rule mining, fuzzy rule based systems, rule miner, conditional random field theory (CRF), formal concept analysis, radial basis function neural network (RBFNN), and online learning algorithms. Rule miner can help figure out common opinion words used together. Selective attempts were made to exploit evolutionary computation in feature selection, while its major potential is still unutilized. Since, sentiment is often found in vague form, fuzzy logic is eminently suitable to model the vagueness in more robust way. CRF can be augmented with domain information in better extraction of aspects. RBFNN and online learning algorithms can be very much useful especially in big data scenario like sentiment analysis on streaming text etc. Ontology can be useful in globalizing the measurement standard of sentiments.

One of the main contributions of the paper is to present a list of available public datasets for SA. The most interesting part of the survey is to present claimed future enhancement in surveyed articles. That

http://dx.doi.org/10.1016/j.knosys.2015.06.015

will be helpful to a novice researcher to address new challenges very precisely and find out the most common challenges to look forward for a new solution. Overall, sentiment analysis has found various promising applications like market prediction, political sentiment determination, equity value prediction, box office prediction etc. But, a lot work still remains to be done and it is a fertile area.

References:

[1] B. Pang, L. Lee, Opinion mining and sentiment analysis, Foundations and Trends in Information Retrieval 2 (2008) 1–135.
[2] A. Balahur, Methods and Resources for Sentiment Analysis in Multilingual Documents of Different Text Types, PhD Thesis, University of Alicante, Spain, 2011, 273 pages.
[3] I. Niles, A. Pease, Linking Lexicons and ontologies:mapping WordNet to the Suggested Upper Merged Ontology, Proceedings of the 2003 International Conference on Information and Knowledge Engineering (IKE 03), Las Vegas, 2003, pp. 23–26.
[4] C. Strapparava, A. Valitutti, WordNet-Affect: an affective extension of WordNet, Proceedings of LREC, vol. 4, 2004, pp. 1083–1086.
[5] A. Esuli, F. Sebastiani, SENTIWORDNET: a publicly available lexical resource for opinion mining, Proceedings of the 5th Conference on Language Resources and Evaluation LREC-06, Genoa, Italy, 2006, pp. 417–422, See also: http://sentiwordnet.isti.cnr.it/.
[6] S. Baccianella, A. Esuli, F. Sebastiani, SENTIWORDNET 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining, Proceedings of LREC-10, Malta, 2010, pp. 2200–2204.
[7] J. Yi, T. Nasukawa, R. Bunescu, W. Niblack, Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques. In Proceedings of the Third IEEE International Conference on Data Mining, 2003, pages 427–434.
[8] S.K. Li, Z. Guan, L.Y. Tang et al., Exploiting consumer reviews for product feature ranking, Journal of Computer Science and Technology 27(3): 635-649 May 2012. DOI 10.1007/s11390-012-1250-z.
[9] J. Kamps, M. Marx, Words with attitude, Proceedings of the 1st International WordNet Conference, Mysore, India, 2002, pp. 332–341.
[10] K. Crammer, Y. Singer, Ultraconservative online algorithms for multiclass problems, JMLR, 2003.
[11] J.C.B. Rabelo, R.B.C. Prudêncio, F.A. Barros, Using link structure to infer opinions in social networks. IEEE International Conference on Systems, Man, and Cybernetics (SMC 2012), IEEE, 2012.
[12] T. Wilson, J. Wiebe, P. Hoffmann, Recognizing contextual polarity in phrase-level sentiment analysis, Proceedings of HLT/EMNLP-05, Vancouver, Canada, 2005.
[13] A. Moreo, M. Romero, J.L. Castro, J.M. Zurita, Lexicon-based Comments-oriented News Sentiment Analyzer system, Expert Systems with Applications 39 (2012) 9166–9180.
[14] R. Feldman, Techniques and Applications for Sentiment Analysis, Review Articles, Communications of the ACM, Vol. 56 No. 4, Pages 82-89, April 2013.
[15] E. Cambria, B. Schuller, and Y.-Q. Xia. New avenues in opinion mining and sentiment analysis (extended abstract). In: Proceedings of IJCAI, Buenos Aires (2015)
[16] H. Tang, S. Tan, X. Cheng, A survey on sentiment detection of reviews, Expert Systems with Applications 36 (2009) 10760–10773.
[17] H. Chen, D. Zimbra, AI and opinion mining. IEEE Intelligent Systems, 25, (2010) 74-80.
[18] J. Bollen, H. Mao, X. Zeng, Twitter mood predicts the stock market, Journal of Computational Science 2 (2012) 1-8
[19] S. Kumar, F. Morstatter, H. Liu, Twitter Data Analytics, August 19, 2013, Springer.
[20] R. McDonald, K. Crammer, F. Pereira, Online large-margin training of dependency parsers. In Proc. ACL, 2005.
[21] A. C.-R. Tsai, C.-E. Wu and R. T.-H. Tsai, J. Y.-j. Hsu, Building a Concept-Level Sentiment Dictionary Based on common sense knowledge, Knowledge-Based Approaches to Concept-Level Sentiment Analysis, IEEE Intelligent Systems, 2013.
[22] C. Bosco, V. Patti, A. Bolioli, Developing Corpora for Sentiment Analysis: The Case of Irony and Senti-TUT, Knowledge-Based Approaches to Concept-Level Sentiment Analysis, IEEE Intelligent Systems, 2013.
[23] S. Poria, A. Gelbukh, A. Hussain, N. Howard, D. Das, S. Bandyopadhyay, Enhanced SenticNet with Affective Labels for Concept-Based Opinion Mining, Knowledge-Based Approaches to Concept-Level Sentiment Analysis, IEEE Intelligent Systems, (2013) 1.
[24] I. Maks, P. Vossen, A lexicon model for deep sentiment analysis and opinion mining applications, Decision support systems 53 (2012) 680-688.
[25] Q. Miao, Q. Li, R. Dai, AMAZING: A sentiment mining and retrieval system, Expert Systems with Applications 36 (2009) 7192–7198.
[26] M.R. Saleh, M.T. Martín-Valdivia, A. Montejo-Ráez, L.A. Ureña-López, Experiments with SVM to classify opinions in different domains, Expert Systems with Applications 38 (2011) 14799–14804.
[27] P. Sobkowicz, M. Kaschesky, G. Bouchard, Opinion mining in social media: Modeling, simulating, and forecasting political opinions in the web, Government Information Quarterly 29 (2012) 470–479.
[28] J. Pennebaker, R. Booth, M. Francias, Linguistic Inquiry and Word Count. LIWC 2007, 2007.
[29] R. Narayanan, B. Liu, A. Choudhary, Sentiment analysis of conditional sentences, In Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1, pp. 180-189, Association for Computational Linguistics, 2009.
[30] A. Balahur, Jesu´s M. Hermida, and Andre´s Montoyo. Building and Exploiting EmotiNet, a Knowledge Base for Emotion Detection Based on the Appraisal Theory Model, IEEE Transactions on Affective Computing, VOL. 3, NO. 1, January-March 2012.
[31] V. Hatzivassiloglou, J. Wiebe, Effects of adjective orientation and gradability on sentence subjectivity. Proceedings of the 18th Conference on Computational linguistics (Vol. 1, pp. 299–305). Association for Computational Linguistics, 2000.
[32] P. Turney, Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews, In: Proceedings of the 40th annual meeting on association for computational linguistics ACL'02 (pp. 417–424), Stroudsburg, PA, USA: Association for Computational Linguistics, 2002.
[33] B. Pang, L. Lee, S. Vaithyanathan, Thumbs up? sentiment classification using machine learning techniques, Proceedings of the ACL-

02 conference on empirical methods in natural language processing (Vol. 10, pp. 79–86). Association for Computational Linguistics, 2002.
[34] T. Wilson, P. Hoffmann, S. Somasundaran, J. Kessler, J. Wiebe, Y. Choi ,C. Cardie, E. Riloff & S. Patwardhan , OpinionFinder: A system for subjectivity analysis, In: Proceedings of HLT/EMNLP on interactive demonstrations (pp. 34–35), Association for Computational Linguistics, 2005.
[35] I. Peñalver-Martinez et al. Feature-based opinion mining through ontologies. Expert Systems with Applications (2014), http://dx.doi.org/10.1016/j.eswa.2014.03.022.
[36] H. Cho et al., Data-driven integration of multiple sentiment dictionaries for lexicon-based sentiment classification of product reviews, Knowl. Based Syst. (2014), http://dx.doi.org/10.1016/j.knosys.2014.06.001.
[37] L. Chen, L. Qi, F. Wang, Comparison of feature-level learning methods for mining online consumer reviews, Expert Systems with Applications 39 (2012) 9588–9601.
[38] M. Hu, B. Liu, Mining and summarizing customer reviews, in: Proceedings of tenth ACM SIGKDD international conference on knowledge discovery and data mining (pp. 168–177) ACM, 2004.
[39] A.L. Maas, R.E. Daly, P.T. Pham, Dan. Huang, A.Y. Ng, C. Potts, Learning word vectors for sentiment analysis. In Proceedings of the 49th annual meeting of the association for computational linguistics (ACL 2011) Portland, Oregon, USA, (pp. 142–150) 2011.
[40] Taddy, Measuring Political Sentiment on Twitter: Factor Optimal Design for Multinomial Inverse Regression, Technometrics, 2013, DOI: 10.1080/00401706.2013.778791
[41] S.-T. Li, F.-C. Tsai, A fuzzy conceptualization model for text mining with application in opinion polarity classification, Knowledge-Based Systems 39 (2013) 23–33.
[42] F. Figueiredo, L. Rocha, T. Couto, T. Salles, M.A. Goncalves, W. Meira Jr, Word co-occurrence features for text classification, Inf. Syst. 36 (2011) 843–858.
[43] Y.-M. Li, Y.-L. Shiu, A diffusion mechanism for social advertising over microblogs, Decision Support Systems 54 (2012) 9–22.
[44] S. Wang, D. Li, X. Song, Y. Wei, H. Li, A feature selection method based on improved Fisher's discriminant ratio for text sentiment classification, Expert Systems with Applications 38 (2011) 8696–8702.
[45] Y. Dang, Y. Zhang, H. Chen, A Lexicon-Enhanced Method for Sentiment Classification: An Experiment on Online Product Reviews, Sentiment Classification, IEEE Intelligent Systems, July/August 2010.
[46] E. Boiy, M.-F. Moens, A machine learning approach to sentiment analysis in multilingual Web texts, Inf. Retrieval (2009) 12:526–558, DOI 10.1007/s10791-008-9070-z.
[47] E. J. Fortuny, T. D. Smedt, D. Martens, W. Daelemans, Media coverage in times of political crisis: A text mining approach, Expert Systems with Applications 39 (2012) 11616–11622.
[48] L.-S. Chen, C.-H. Liu, H.-J. Chiu, A neural network based approach for sentiment classification in the blogosphere, Journal of Informetrics 5 (2011) 313–322.
[49] H. Bao, Q. Li, S. S. Liao, S. Song, H. Gao, A new temporal and social PMF-based method to predict users' interests in micro-blogging, Decision Support Systems 55 (2013) 698–709.
[50] R. Moraes, J.F. Valiati, W.P. Gaviao Neto, Document-level sentiment classification: An empirical comparison between SVM and ANN, Expert Systems with Applications 40 (2013) 621–633.
[51] J. Du et al., Box office prediction based on microblog. Expert Systems with Applications (2013), http://dx.doi.org/10.1016/j.eswa.2013.08.065
[52] E. Cambria, P. Gastaldo, F. Bisio, and R. Zunino. An ELM-based model for affective analogical reasoning. Neurocomputing 149, pp. 443-455 (2015)
[53] T. Geva, J. Zahavi, Empirical evaluation of an automated intraday stock recommendation system incorporating both market data and textual news, Decision Support Systems (2013), http://dx.doi.org/10.1016/j.dss.2013.09.013.
[54] R. Xia, C. Zong, S. Li, Ensemble of feature sets and classification algorithms for sentiment classification, Information Sciences 181 (2011) 1138–1152.
[55] F.Å. Nielsen, A new ANEW: Evaluation of a word list for sentiment analysis in microblogs, arXiv preprint arXiv:1103.2903 (2011).
[56] Z. Zhai, Product Feature Grouping for Opinion Mining, Opinion Mining, IEEE Intelligent Systems, July/August 2012.
[57] M. Ghiassi, J. Skinner, D. Zimbra, Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network, Expert Systems with Applications 40 (2013) 6266–6282.
[58] B. Desmet, V. Hoste, Emotion detection in suicide notes, Expert Systems with Applications 40 (2013) 6351–6358.
[59] L. García-Moya, H. Anaya-Sánchez, R. Berlanga-Llavori, Retrieving Product Features and Opinions from Customer Reviews, Statistical Approaches to Concept-Level Sentiment Analysis, IEEE Intelligent Systems, May/June 2013.
[60] J.J. Jung, Online named entity recognition method for microtexts in social networking services: A case study of twitter, Expert Systems with Applications 39 (2012) 8066–8070.
[61] C. Lin, Y. He, R. Everson, S. Rüger, Weakly Supervised Joint Sentiment-Topic Detection from Text, IEEE Transactions on Knowledge and Data Engineering, VOL. 24, NO. 6, JUNE 2012.
[62] L. Liu, X. Nie, H. Wang, Toward a Fuzzy Domain Sentiment Ontology Tree for Sentiment Analysis, 5th International Congress on Image and Signal Processing (CISP 2012), 2012.
[63] J. Zhu, H. Wang, M. Zhu, B. K. Tsou, and M. Ma, Aspect-Based Opinion Polling from Customer Reviews, IEEE Transactions on Affective Computing, VOL. 2, NO. 1, January-March 2011.
[64] S. M. Mohammad, From once upon a time to happily ever after: Tracking emotions in mail and books, Decision Support Systems 53 (2012) 730–741.
[65] M. Taboada, J. Grieve, Analyzing appraisal automatically, In Proceedings of the AAAI spring symposium on exploring attitude and affect in text: Theories and applications, 2004, (pp. 158–161).
[66] G. Wang et al., Sentiment classification: The contribution of ensemble learning, Decision Support Systems (2013), http://dx.doi.org/10.1016/j.dss.2013.08.002
[67] C. Quan, F. Ren, Unsupervised product feature extraction for feature-oriented opinion determination, Information Sciences 272 (2014) 16-28.
[68] F.L. Cruz, J.A. Troyano, F. Enríquez, F.J. Ortega, C.G. Vallejo,'Long autonomy or long delay?' The importance of domain in opinion mining, Expert Systems with Applications 40 (2013) 3174–3184.
[69] M. Taboada, J. Brooke, M. Tofiloski, Lexicon-Based Methods for Sentiment Analysis, Computational Linguistics, Volume 37,

Number 2, 2011, 267-307.
[70] Y.C. Xu et al., Measuring product susceptibility in online product review social network, Decision Support Systems (2013), http://dx.doi.org/10.1016/j.dss.2013.01.009.
[71] M. Taboada, J. Grieve, Analyzing appraisal automatically. In Proceedings of the AAAI Spring Symposium on Exploring Attitude and Affect in Text (AAAI Technical Report SS-04-07), pages 158–161, Stanford, CA, 2004.
[72] M. Taboada, C. Anthony, K. Voll, Creating semantic orientation dictionaries. In Proceedings of 5th International Conference on Language Resources and Evaluation (LREC), pages 427–432, Genoa, 2006.
[73] P.C.R. Lane, D. Clarke, P. Hender, On developing robust models for favourability analysis: Model choice, feature sets and imbalanced data, Decision Support Systems 53 (2012) 712–718.
[74] A. Neviarouskaya, H. Prendinger, M. Ishizuka, SentiFul: A Lexicon for Sentiment Analysis, IEEE Transactions on Affective Computing, Vol. 2, No. 1, January-March 2011.
[75] C. Banea, R. Mihalcea, J. Wiebe, Sense-level subjectivity in a multilingual setting, Computer Speech and Language 28 (2014) 7–19.
[76] Y. Liu, J. Jin, P. Ji, J.A. Harding, R.Y.K. Fung, Identifying helpful online reviews: A product designer's perspective, Computer-Aided Design 45 (2013) 180–194.
[77] K. Coussement, D. von den Poel, Improving customer attrition prediction by integrating emotions from client/company interaction emails and evaluating multiple classifiers, Expert Systems with Applications 36 (2009) 6127–6134.
[78] N. Hu, I. Bose, Y. Gao, L. Liu, Manipulation in digital word-of-mouth: A reality check for book reviews, Decision support systems 50 (2011) 627-635.
[79] M.Á. García-Cumbreras, A. Montejo-Ráez, M.C. Díaz-Galiano, Pessimists and optimists: Improving collaborative filtering through sentiment analysis, Expert Systems with Applications 40 (2013) 6758–6765.
[80] L. Qiu, H. Rui, A. Whinston, Social network-embedded prediction markets: The effects of information acquisition and communication on predictions, Decision Support Systems 55 (2013) 978–987.
[81] A. Ghose and P. G. Ipeirotis, Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics, IEEE Transactions on Knowledge and Data Engineering, VOL. 23, NO. 10, OCTOBER 2011.
[82] R. Mihalcea, C. Banea, J. Wiebe, Learning multilingual subjective language via cross-lingual projections, In Annual Meeting-Association for Computational Linguistics, vol. 45, no. 1, p. 976, 2007.
[83] M. Thelen and E. Rilo, A Bootstrapping Method for Learning Semantic Lexicons Using Extraction Pattern Contexts, Proc. Conf. on Empirical Methods in Natural Language Processing (EMNLP 02), Assoc. for Computational Linguistics, 2002, pp. 214–221.
[84] A. Montoyo, P. Martínez-Barco, A. Balahur, Subjectivity and sentiment analysis: An overview of the current state of the area and envisaged developments, Decision Support Systems 53 (2012) 675–679.
[85] F.H. Khan et al., TOM: Twitter opinion mining framework using hybrid classification scheme, Decision Support Systems (2013), http://dx.doi.org/10.1016/j.dss.2013.09.004.
[86] A. Reyes, P. Rosso, Making objective decisions from subjective data: Detecting irony in customer reviews, Decision Support Systems 53 (2012) 754–760.
[87] J. Wiebe, Learning Subjective Adjectives from Corpora, Proc. 15th Nat'l Conf. Artificial Intelligence (AAAI 00), AAAI Press, 2000, pp. 735–740.
[88] Y. Seki, N. Kando, M. Aono, Multilingual opinion holder identification using author and authority viewpoints, Information Processing and Management 45 (2009) 189–199.
[89] D.E. O'Leary, Blog mining-review and extensions: "From each according to his opinion", Decision Support Systems 51 (2011) 821–830.
[90] Y.M. Li, T.-Y. Li, Deriving market intelligence from microblogs, Decision Support Systems 55 (2013) 206–217.
[91] N. Li, D.D. Wu, Using text mining and sentiment analysis for online forums hotspot detection and forecast, Decision Support Systems 48 (2010) 354–368.
[92] A. Duric, F. Song, Feature selection for sentiment analysis based on content and syntax models, Decision Support Systems 53 (2012) 704–711.
[93] K. Xu, S.S. Liao, J. Li, Y. Song, Mining comparative opinions from customer reviews for Competitive Intelligence, Decision Support Systems 50 (2011) 743–754.
[94] M.-T. Martin-Valdivia, E. Martínez-Cámara, J.-M. Perea-Ortega, L. A. Ureña-López, Sentiment polarity detection in Spanish reviews combining supervised and unsupervised approaches, Expert Systems with Applications 40 (2013) 3934–3942.
[95] A. Balahur, Jesús M. Hermida, A. Montoyo, Detecting implicit expressions of emotion in text: A comparative analysis, Decision Support Systems 53 (2012) 742–753.
[96] E. Kontopoulos, C. Berberidis, T. Dergiades, N. Bassiliades, Ontology-based sentiment analysis of twitter posts, Expert Systems with Applications 40 (2013) 4065–4074.
[97] C. Hung ,H.-K. Lin, C. Yuan, Using Objective Words in SentiWordNet to improve Word-of-Mouth Sentiment Classification, Knowledge - Based Approaches to Concept - Level Sentiment Analysis, IEEE Intelligent Systems, March/April 2013.
[98] A. Weichselbraun, S. Gindl, A. Scharl, Extracting and grounding contextualized sentiment lexicons, Knowledge - based Approaches to Concept - Level Sentiment Analysis, IEEE Intelligent Systems, March/April 2013.
[99] D. Bollegala, D. Weir, J Carroll, Cross-domain sentiment classification using a sentiment sensitive thesaurus, IEEE Transactions on Knowledge and Data Engineering, Vol. 25, No. 8, August 2013.
[100] H. Liu, J. He, T. Wang, W. Song, X. Du, Combining user preferences and user opinions for accurate recommendation, Electronic Commerce Research and Applications 12 (2013) 14–23.
[101] D. Spina, J. Gonzalo, E. Amigó, Discovering filter keywords for company name disambiguation in twitter, Expert Systems with Applications 40 (2013) 4986–5003.
[102] Z. Zhai, H. Xu, B. Kang, P. Jia, Exploiting effective features for Chinese sentiment classification, Expert Systems with Applications 38 (2011) 9139–9146.
[103] L.R. Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition, Proceedings of the IEEE, 77(2):257–285, 1989, February.
[104] E. Riloff, J. Wiebe, Learning extraction patterns for subjective expressions, in: Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing (EMNLP 2003), 2003, pp. 105–112.
[105] P.J. Stone, E.B. Hunt, A computer approach to content analysis: studies using the general inquirer system, in: Proceedings of the

Spring Joint Computer Conference (AFIPS 1963), 1963, pp. 241–256.
[106] E. Cambria, R. Speer, C. Havasi, and A. Hussain, SenticNet: A Publicly Available Semantic Resource for Opinion Mining, In AAAI Fall Symposium: Commonsense Knowledge, vol. 10, p. 02. 2010.
[107] X. Xueke, C. Xueqi, T. Songbo, L. Yue, S. Huawei, Aspect-level opinion mining of online customer reviews, Management and Visualization of User and Network Data, China Communications, March 2013.
[108] M. Abdul-Mageed, M. Diab, S. Kübler, SAMAR: Subjectivity and sentiment analysis for Arabic social media, Computer Speech and Language 28 (2014) 20–37.
[109] A. Abbasi, S. France, Z. Zhang, H. Chen, Selecting attributes for sentiment classification using feature relation networks, IEEE Transactions on Knowledge and Data Engineering, VOL. 23, NO. 3, MARCH 2011.
[110] M. D. Molina-González, E. Martínez-Cámara, M.-T. Martín-Valdivia, José M. Perea-Ortega, Semantic orientation for polarity classification in Spanish reviews, Expert Systems with Applications 40 (2013) 7250–7257.
[111] H. Kang, S. J. Yoo, D. Han, Senti-lexicon and improved Naïve Bayes algorithms for sentiment analysis of restaurant reviews, Expert Systems with Applications 39 (2012) 6000–6010.
[112] R. Prabowo, M. Thelwall, Sentiment analysis: A combined approach, Journal of Informetrics 3 (2009) 143–157.
[113] J.C. de Albornoz, L. Plaza, P. Gervas, Sentisense: An easily scalable concept-based affective lexicon for sentiment analysis, in: Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012), 2012, pp. 3562–3567.
[114] Z. Zhang, Q. Ye, Z. Zhang, Y. Li, Sentiment classification of Internet restaurant reviews written in Cantonese, Expert Systems with Applications 38 (2011) 7674–7682.
[115] H. Wang, P. Yin, L. Zheng, James N.K. Liu, Sentiment classification of online reviews: using sentence-based language model, Journal of Experimental & Theoretical Artificial Intelligence, (2013) DOI: 10.1080/0952813X.2013.782352
[116] M. Thelwall, K. Buckley, G. Paltoglou, D. Cai, and A. Kappas, Sentiment Strength Detection in Short Informal Text, Journal of the American Society for Information Science and Technology, 61(12):2544–2558, 2010.
[117] M. Thelwall, K. Buckley, and G. Paltoglou, Sentiment in Twitter Events, Journal of the American Society for Information Science and Technology, 62 (2):406–418, 2011.
[118] M. Thelwall, K. Buckley, and G. Paltoglou, Sentiment Strength Detection for the SocialWeb, Journal of the American Society for Information Science and Technology, 63(1):163–173, 2012.
[119] S. Cerini, V. Compagnoni, A. Demontis, M. Formentelli, C. Gandini, Micro- 862 WNOp: a gold standard for the evaluation of automatically compiled lexical resources for opinion mining, in: A. Sanso (Ed.), Language Resources and Linguistic Theory, Franco Angeli, 2007, pp. 200–210.
[120] K.J. Trainor et al., Social media technology usage and customer relationship performance: A capabilities-based examination of social CRM, Journal of Business Research (2013), http://dx.doi.org/10.1016/j.jbusres.2013.05.002
[121] S. Tan, X. Cheng, Y. Wang, H. Xu, Adapting Naive Bayes to Domain Adaptation for Sentiment Analysis, M. Boughanem et al. (Eds.): ECIR 2009, LNCS 5478, pp. 337–349, 2009.
[122] C. Matuszek, J. Cabral, M.J. Witbrock, J. DeOliveira, An Introduction to the Syntax and Content of Cyc, In AAAI Spring Symposium: Formalizing and Compiling Background Knowledge and Its Applications to Knowledge Representation and Question Answering, pp. 44-49. 2006.
[123] W. Li, H. Xu, Text-based emotion classification using emotion cause extraction. Expert Systems with Applications (2013), http://dx.doi.org/10.1016/j.eswa.2013.08.073
[124] Y. Yu, W. Duan, Q. Cao, The impact of social and conventional media on firm equity value: A sentiment analysis approach, Decision Support Systems 55 (2013) 919–926
[125] M. van den Camp, A. van den Bosch, The socialist network, Decision Support Systems 53 (2012) 761–769.
[126] Mohamed M. Mostafa, More than words: Social networks' text mining for consumer brand sentiments, Expert Systems with Applications 40 (2013) 4241–4251.
[127] M. Thelwall and K. Buckley, Topic-Based Sentiment Analysis for the Social Web: The Role of Mood and Issue-Related Words, Journal of the American Society for Information Science and Technology, 64(8):1608–1617, 2013
[128] K. Bollacker, C. Evans, P. Paritosh, T. Sturge, J. Taylor, FreeBase: a collaboratively created graph database for structuring human knowledge, In Proceedings of the 2008 ACM SIGMOD International conference on Management of data, pp. 1247-1250. ACM, 2008.
[129] C. Bizer, J. Lehmann, G. Kobilarov, S. Auer, C. Becker, R. Cyganiak, S. Hellmann. DBpedia-A crystallization point for the Web of Data, Web Semantics: science, services and agents on the world wide web 7, no. 3 (2009): 154-165.
[130] Z. Zhang, Weighing stars: Aggregating online product reviews for intelligent e-commerce applications, IEEE Intelligent Systems, September/October 2008
[131] H. Rui, Y. Liu, A. Whinston, Whose and what chatter matters? The effect of tweets on movie sales, Decision Support Systems 55 (2013) 863–870
[132] J. Zhan, H. T. Loh, Y. Liu, Gather customer concerns from online product reviews – A text summarization approach, Expert Systems with Applications 36 (2009) 2107–2115
[133] G.A. Miller, WordNet: a lexical database for English, Communications of the ACM 38, no. 11 (1995): 39-41.
[134] B. Liu, Sentiment analysis: A multi-faceted problem, IEEE Intelligent Systems 25, no. 3 (2010): 76-80.
[135] M. Hu, B. Liu, Mining opinion features in customer reviews, in: Proc. of the Nineteenth National Conference on Artificial Intelligence (AAAI), 2004, pp. 755–760.
[136] E. Cambria, C. Havasi, and A. Hussain, SenticNet 2: A Semantic and Affective Resource for Opinion Mining and Sentiment Analysis, Proc. 25th Int'l Florida Artificial Intelligence Research Society Conf., AAAI, 2012, pp. 202–207.
[137] M. Tsytsarau, T. Palpanas, Survey on mining subjective data on the web, Data Min Knowl Disc (2012) 24:478–514, DOI 10.1007/s10618-011-0238-6
[138] W. Medhat et al. Sentiment analysis algorithms and applications: A survey, Ain Shams Eng J (2014), http:// dx.doi.org/10.1016/j.asej.2014.04.011
[139] M. Saif, D. Cody, D. Bonnie, Generating high-coverage semantic orientation lexicons from overtly marked words and a thesaurus, Proceedings of the 2009 Conference on EMNLP, Association for Computational Linguistics, Morristown, NJ, USA, 2009, pp. 599–608.
[140] E. Cambria, A. Hussain, Sentic Computing: Techniques, Tools, and Applications, Springer, 2012.
[141] E. Cambria, J. Fu, F. Bisio and S. Poria. AffectiveSpace 2: Enabling affective intuition for concept-level sentiment analysis. In: AAAI,

pp. 508-514, Austin (2015)
[142] D. Tufi, D. Stefenescu, Experiments with a differential semantics annotation for WordNet 3.0, Decision Support Systems 53 (2012) 695–703
[143] A. Popescu, O. Etzioni, Extracting product features and opinions from reviews, in: Proc. of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing (HLT-EMNLP), 2005, pp. 339–346.
[144] F. Cruz et al., A Knowledge-Rich Approach to Feature-Based Opinion Extraction from Product Reviews, Proc. 2nd Int'l Workshop on Search and Mining User-Generated Contents, ACM, 2010, pp. 13–20.
[145] G. Qiu, B. Liu, J. Bu, C. Chen, Opinion word expansion and target extraction through double propagation, Comput. Linguist. 37 (1) (2011) 9–27.
[146] T. Chklovski, P. Pantel, VerbOcean: Mining the Web for Fine-Grained Semantic Verb Relations, Proc. Conf. Empirical Methods in Natural Language Processing, pp. 33-40, 2004.
[147] K. Scherer and H. Wallbott, The ISEAR Questionnaire and Codebook, Geneva Emotion Research Group, 1997.
[148] E. Demirtas, Cross-Lingual Sentiment Analysis with Machine Translation, Utility of training corpora and sentiment lexica, Master Thesis, University of Technology, 2013
[149] J. Blitzer, M. Dredze, F. Pereira, Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification, In Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics, ACL'07, pages 187-205, 2007. 7, 13, 29
[150] M. Collins, Discriminative training methods for hidden Markov models: Theory and experiments with perceptron algorithms, In Proc. EMNLP, 2002.
[151] C.M. Whissel, The dictionary of Affect in Language. Emotion: theory research and experience, Acad press London, 1989.
[152] D. Haussler, Convolution kernels on discrete structures. Technical report, University of California at Santa Cruz, 1999.
[153] E. Rilo, J. Wiebe, and T. Wilson, Learning Subjective Nouns Using Extraction Pattern Bootstrapping, Proc. Conf. Computational Natural Language Learning (CoNLL 03), 2003, pp. 25–32.
[154] M. Eirinaki, S. Pisal, J. Singh, Feature-based opinion mining and ranking, Journal of Computer and System Sciences 78 (2012) 1175–1184.
[155] X. Yu, Y. Liu, J. X. Huang, A. An, Mining online reviews for predicting sales performance: A case study in the movie domain, IEEE Transactions on Knowledge and Data Engineering, VOL. 24, NO. 4, APRIL 2012.
[156] X. Bai, Predicting consumer sentiments from online text, Decision Support Systems 50 (2011) 732–742.
[157] Abd. S.H. Basari, B. Hussin, I.G.P. Ananta, J. Zeniarja, Opinion Mining of Movie Review using Hybrid Method of Support Vector Machine and Particle Swarm Optimization, Procedia Engineering 53 (2013) 453 – 462.
[158] D. Kang , Y. Park, Review-based measurement of customer satisfaction in mobile service: Sentiment analysis and VIKOR approach. Expert Systems with Applications (2013), http://dx.doi.org/10.1016/j.eswa.2013.07.101
[159] S. Pisal, J. Singh, M. Eirinaki, AskUs: An Opinion Search Engine, Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on , vol., no., pp.1243,1246, 11-11 Dec. 2011, doi: 10.1109/ICDMW.2011.24
[160] A. Abbasi, H. Chen, A. Salem, Sentiment Analysis in Multiple Languages: Feature Selection for Opinion Classification in Web Forums, ACM Transactions on Information Systems, Vol. 26, No. 3, Article 12, Publication date: June 2008.
[161] H. Watanabe, A similarity-driven transfer system, In Proc. of the 14th COLING, Vol. 2, pages 770–776, 1992.
[162] L.K.W. Tan, J.C. Na, Y.L. Theng et al. Phrase-level sentiment polarity classification using rule-based typed dependencies and additional complex phrases consideration. Journal of Computer Science and Technology 27(3): 650-666 May 2012. DOI 10.1007/s11390-012-1251-y.
[163] F. Bravo-Marquez et al., Meta-level sentiment models for big social data analysis, Knowl. Based Syst. (2014), http://dx.doi.org/10.1016/j.knosys.2014.05.016
[164] A. Go, R. Bhayani, L. Huang, Twitter sentiment classification using distant supervision, Technical Report Stanford University, 2010.
[165] Z.-H. Deng, K.-H. Luo, H.-L. Yu, A study of supervised term weighting scheme for sentiment analysis, Expert Systems with Applications 41 (2014) 3506–3513.
[166] Y. Lu, X. Kong, X. Quan, W. Liu, and Y. Xu, Exploring the Sentiment Strength of User Reviews, L. Chen et al. (Eds.): WAIM 2010, LNCS 6184, pp. 471–482, 2010.
[167] B. Pang, L. Lee, A sentiment education: Sentiment analysis using subjectivity summarization based on minimum cuts, in: Proceedings of the 42nd annual meeting on Association for Computational Linguistics (p. 271), 2004, July.
[168] R. McDonald et al., Structured models for fine-to-coarse sentiment analysis, Annual Meeting-Association For Computational Linguistics, Vol. 45. No. 1. 2007.
[169] C. Whitelaw, N. Garg, S. Argamon, Using appraisal groups for sentiment analysis. In Proceedings of the 14th ACM International Conference on Information and Knowledge Management (pp. 625-631), 2005, October. ACM.
[170] H. Kanayama, T. Nasukawa, Fully automatic lexicon expansion for domain-oriented sentiment analysis. In Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, (pp. 355-363), 2006, July.
[171] T. Nasukawa, J. Yi, Sentiment analysis: Capturing favorability using natural language processing, in: Proceedings of the 2nd International Conference On Knowledge Capture, ACM, (pp. 70-77), 2003, October.
[172] T. Mullen, N. Collier, Sentiment Analysis using Support Vector Machines with Diverse Information Sources. In EMNLP (Vol. 4, pp. 412-418), 2004, July.
[173] K. Hiroshi, N. Tetsuya, W. Hideo, Deeper sentiment analysis using machine translation technology, In Proceedings of the 20th international conference on Computational Linguistics, p. 494. Association for Computational Linguistics, 2004.
[174] H.N.T. Xuan, A.C. Le, L.M. Nguyen, Linguistic Features for Subjectivity Classification, In: International Conference on Asian Language Processing (IALP), 2012, (pp. 17-20) 2012, November. IEEE.
[175] J.H. Wang, C.C. Lee, Unsupervised opinion phrase extraction and rating in Chinese blog posts. In: IEEE Third International Conference on Privacy, Security, Risk And Trust, 2011 and IEEE Third International Conference on Social Computing (SocialCom) (pp. 820-823) 2011, October. IEEE.
[176] L. Jiang, M. Yu, M. Zhou, X. Liu, T. Zhao, Target-dependent Twitter Sentiment Classification, in: ACL (pp. 151-160), 2011, June.
[177] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, R. Passonneau, Sentiment analysis of twitter data. in: Proceedings of the Workshop on Languages in Social Media, Association for Computational Linguistics, (pp. 30-38) 2011, June.
[178] F. Benamara, C. Cesarano, A. Picariello, D.R. Recupero, V.S. Subrahmanian, Sentiment Analysis: Adjectives and Adverbs are Better

than Adjectives Alone, In: ICWSM(2007, March).
[179] E. Kouloumpis, T. Wilson, J. Moore, Twitter sentiment analysis: The Good the Bad and the OMG!, In ICWSM (2011, May).
[180] V. Sindhwani, P. Melville, Document-word co-regularization for semi-supervised sentiment analysis, In: Eighth IEEE International Conference on Data Mining, (pp. 1025-1030), 2008, December.
[181] B. Liu, Sentiment analysis and opinion mining, Morgan and Claypool publishers, May 2012.
[182] R.Y.K. Lau, S.S.Y. Liao, C. Li, Social Analytics: Learning Fuzzy Product Ontologies for Aspect-Oriented Sentiment Analysis, Decision Support Systems (2014), doi: 10.1016/j.dss.2014.05.005.
[183] A.K. Nassirtoussi, S. Aghabozorgi, T.Y. Wah, D.C.L. Ngo, Text Mining of News-Headlines for FOREX Market Prediction: A Multi-layer Dimension Reduction Algorithm with Semantics & Sentiment, Expert Systems with Applications (2014), doi: http://dx.doi.org/10.1016/j.eswa.2014.08.004.
[184] X. Li et al., News impact on stock price return via sentiment analysis, Knowl. Based Syst. (2014), http://dx.doi.org/10.1016/j.knosys.2014.04.022.
[185] S. Moghaddam, M. Ester, The flda model for aspect-based opinion mining: Addressing the cold start problem, In Proceedings of the 22nd international conference on World Wide Web, pp. 909-918. International World Wide Web Conferences Steering Committee, 2013.
[186] S. Kim, J. Zhang, Z. Chen, A. Oh, S. Liu, A Hierarchical Aspect-Sentiment Model for Online Reviews, In AAAI 2013.
[187] H. Wang, Y. Lu, C. Zhai, Latent Aspect Rating Analysis on Review Text Data: A Rating Regression Approach, The 16th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD'2010), p783-792, 2010.
[188] Y. Jo, A.H. Oh, Aspect and sentiment unification model for online review analysis, In Proceedings of the fourth ACM international conference on Web search and data mining, pp. 815-824. ACM, 2011.
[189] T. Wang et al., Product aspect extraction supervised with online domain knowledge, Knowledge-Based Systems 71 (2014): 86-100.
[190] K. Zhang et al., Incorporating conditional random fields and active learning to improve sentiment identification, Neural Networks (2014).
[191] X. Zheng, Z. Lin, X. Wang, K.-J. Lin, M. Song, Incorporating appraisal expression patterns into topic modeling for aspect and sentiment word identification, Knowledge-Based Systems 61 (2014): 29-47.
[192] T.R.Gruber, Toward principles for the design of ontologies used for knowledge sharing? International journal of human-computer studies 43, no. 5 (1995) 907-928.
[193] Z. Hai, K. Chang, J.-J. Kim, C.C. Yang, Identifying Features in Opinion Mining via Intrinsic and Extrinsic Domain Relevance, IEEE Transactions on Knowledge and Data Engineering 26, No. 3 (2014): 623-634.
[194] S. Mukherjee, S. Joshi, Sentiment aggregation using ConceptNet ontology, IJCNLP, 2013.
[195] S. Mukherjee, S. Joshi, Author-Specific Sentiment Aggregation for Polarity Prediction of Reviews, Proceedings of the 9th edition of the Language Resources and Evaluation Conference (LREC 2014).
[196] L. Zhou, P. Chaovalit, Ontology-supported polarity mining, Journal of the American Society for Information Science and technology 59, no. 1 (2008): 98-110.
[197] G. Vinodhini, R. M. Chandrasekaran, Opinion mining using principal component analysis based ensemble model for e-commerce application, CSI Transactions on ICT (2014): 1-11.
[198] B. Pang and L. Lee. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, ACL'05, pages 115-124, 2005.
[199] A. Heydari, Mhd. A. Tavakoli, N. Salim, and Z. Heydari, Detection of review spam: A survey, Expert Systems with Applications 42, no. 7 (2015): 3634-3642.
[200] M. Ott, C. Cardie, & J. T. Hancock, Negative deceptive opinion spam, Proceedings of the 2013 conference of the north american chapter of the association for computational linguistics: human language technologies, June 2013 (pp. 497–501).
[201] R. Socher, A. Perelygin, J.Y. Wu, J. Chuang, C.D. Manning, A.Y. Ng, and C. Potts, Recursive deep models for semantic compositionality over a sentiment treebank, In Proceedings of the conference on empirical methods in natural language processing (EMNLP), vol. 1631, p. 1642. 2013.
[202] E. Cambria, D. Olsher, and D. Rajagopal, SenticNet 3: a common and common-sense knowledge base for cognition-driven sentiment analysis, In Twenty-eighth AAAI conference on artificial intelligence. (2014): pp. 1515-1521.
[203] D. Bell, T. Koulouri, S. Lauria, R.D. Macredie, J. Sutton, Microblogging as a mechanism for human–robot interaction, Knowledge-Based Systems 69 (2014) 64–77.
[204] K. Gimpel, N. Schneider, B. O'Connor, D. Das, D. Mills, J. Eisenstein, N.A. Smith, Part-of-speech tagging for twitter: annotation, features, and experiments, in: Carnegie-Mellon Univ Pittsburgh Pa School of Computer Science, 2010.
[205] S. Rill, D. Reinel, J. Scheidt, and R.V. Zicari, PoliTwi: Early detection of emerging political topics on twitter and the impact on concept-level sentiment analysis, Knowledge-Based Systems 69 (2014): 24-33.
[206] O. Popescu, and C. Strapparava, Time corpora: Epochs, opinions and changes, Knowledge-Based Systems 69 (2014): 3-13.
[207] C.-E. Wu, and R. T.-H. Tsai, Using relation selection to improve value propagation in a ConceptNet-based sentiment dictionary, Knowledge-Based Systems 69 (2014): 100-107.
[208] D. Rajagopal, E. Cambria, D. Olsher, K. Kwok, A graph-based approach to commonsense concept extraction and semantic similarity detection, in: Proceedings of the 22nd International Conference on World Wide Web Companion (WWW'13), Republic and Canton of Geneva, Switzerland, 2013, pp. 565–570.
[209] A. Ortigosa, J.M. Martín, and R.M. Carro, Sentiment analysis in Facebook and its application to e-learning, Computers in Human Behavior 31 (2014): 527-541.
[210] H. Costa, L.H.C. Merschmann, F. Barth, and F. Benevenuto, Pollution, bad-mouthing, and local marketing: The underground of location-based social networks, Information Sciences 279 (2014): 123-137.
[211] A. Montejo-Ráez, , M. C. Díaz-Galiano, F. Martinez-Santiago, and L. A. Ureña-López, Crowd explicit sentiment analysis, Knowledge-Based Systems 69 (2014): 134-139.
[212] M. Ott, Y. Choi, C. Cardie, and J. T. Hancock, Finding deceptive opinion spam by any stretch of the imagination, In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1, pp. 309-319, 2011.
[213] S. Poria, E. Cambria, G. Winterstein, and G.-B. Huang, Sentic patterns: Dependency-based rules for concept-level sentiment analysis, Knowledge-Based Systems 69 (2014): 45-63.

[214] S. Poria, A. Gelbukh, E. Cambria, A. Hussain, G.-B. Huang, EmoSenticSpace: A novel framework for affective common-sense reasoning, Knowledge-Based Systems 69 (2014) 108–123.
[215] G. Matthews, I.J. Deary, M.C. Whiteman, Personality Traits, Cambridge, UK, 2009, pp. 23–26.
[216] A. Weichselbraun, S. Gindl, A. Scharl, Enriching semantic knowledge bases for opinion mining in big data applications, Knowledge-Based Systems 69 (2014) 78–85.
[217] R. Justo, T. Corcoran, S.M. Lukin, M. Walker, M. InÃ©s Torres, Extracting relevant knowledge for the detection of sarcasm and nastiness in the social web, Knowledge-Based Systems 69 (2014) 124–133.
[218] A. Bagheri et al., Care more about customers: Unsupervised domain-independent aspect detection for sentiment analysis of customer reviews, Knowl. Based Syst. (2013), http://dx.doi.org/10.1016/j.knosys.2013.08.011.
[219] C.P. Wei, Y.M. Chen, C.S. Yang, C.C. Yang, Understanding what concerns consumers: a semantic approach to product feature extraction from consumer reviews, Information Systems and E-Business Management 8 (2) (2010) 149–167.
[220] S. Banerjee, and Alton YK Chua, Applauses in hotel reviews: Genuine or deceptive?, In Science and Information Conference (SAI), 2014, pp. 938-942. IEEE, 2014.
[221] H.-J. Min, J.C. Park, Identifying helpful reviews based on customer's mentions about experiences, Expert Systems with Applications 39 (2012) 11830–11838.
[222] P. Racherla, W. Friske, Perceived 'usefulness' of online consumer reviews: An exploratory investigation across three services categories, Electronic Commerce Research and Applications 11 (2012) 548–559.
[223] S.M. Mudambi, & D. Schuff, What makes a helpful online review? A study of customer reviews on amazon.com. MIS Quarterly, 34, (2010) 185–200.
[224] H. Huang & David C. Yen Predicting the Helpfulness of Online Reviews—A Replication, International Journal of Human-Computer Interaction, 29:2, (2013) 129-138, DOI: 10.1080/10447318.2012.694791.
[225] C.C. Chen, Y.-D. Tseng, Quality evaluation of product reviews using an information quality framework, Decision Support Systems 50 (2011) 755–768.
[226] S. Lee, J.Y. Choeh, Predicting the helpfulness of online reviews using multilayer perceptron neural networks, Expert Systems with Applications 41 (2014) 3041–3046.
[227] T.L. Ngo-Ye, A.P. Sinha, The influence of reviewer engagement characteristics on online review helpfulness: A text regression model, Decision Support Systems 61 (2014) 47–58.
[228] S. Krishnamoorthy, Linguistic features for review helpfulness prediction, Expert Systems with Applications 42 (2015) 3751–3759.
[229] N. Purnawirawan, P. De Pelsmacker & N. Dens, Balance and Sequence in Online Reviews: How Perceived Usefulness Affects Attitudes and Intentions, Journal of Interactive Marketing 26 (2012) 244–255.
[230] M.A. Walker, P. Anand, J.E. Fox Tree, R. Abbott, J. King, A Corpus for Research on Deliberation and Debate. LREC 2012.
[231] Nan Hu, Ling Liu, V. Sambamurthy, Fraud detection in online consumer reviews, Decision Support Systems 50 (2011) 614–626.
[232] Hernández Fusilier, D., et al. Detecting positive and negative deceptive opinions using PU-learning. Information Processing and Management (2014), http://dx.doi.org/10.1016/j.ipm.2014.11.001.
[233] B. Liu, Y. Dai, X.L. Li, W.S. Lee, & Y. Philip, Partially supervised classification of text documents. In ICML 2002, Proceedings of the nineteenth international conference on machine learning, July 2002 (pp. 387–394).
[234] Jindal & Liu, Opinion spam and analysis, In Proceedings of the international conference on web search and web data mining. ACM.
[235] S.P. Algur, A.P. Patil, P.S. Hiremath, and S. Shivashankar, Conceptual level similarity measure based review spam detection, In Signal and Image Processing (ICSIP), 2010 International Conference on, pp. 416-423. IEEE, 2010.
[236] F. Li, M. Huang, Y. Yang, and X. Zhu, Learning to identify review spam, In IJCAI Proceedings-International Joint Conference on Artificial Intelligence, vol. 22, no. 3, p. 2488. 2011.
[237] G. Wang, S. Xie, B. Liu, and P.S. Yu., Review graph based online store review spammer detection, In Data Mining (ICDM), 2011 IEEE 11th International Conference on, pp. 1242-1247. IEEE, 2011.
[238] E.-P. Lim, V.-A. Nguyen, N. Jindal, B. Liu, and H.W. Lauw, Detecting product review spammers using rating behaviors, In Proceedings of the 19th ACM international conference on Information and knowledge management, pp. 939-948. ACM, 2010.
[239] A. Mukherjee, B. Liu, and N. Glance, Spotting fake reviewer groups in consumer reviews, In Proceedings of the 21st international conference on World Wide Web, pp. 191-200. ACM, 2012.
[240] H. Xu et al., Implicit feature identification in Chinese reviews using explicit topic mining model, Knowl. Based Syst. (2014), http://dx.doi.org/10.1016/j.knosys.2014.12.012.
[241] Z. Yan, et al., EXPRS: An extended pagerank method for product feature extraction from online consumer reviews, Inf. Manage. (2015), http://dx.doi.org/10.1016/j.im.2015.02.002.
[242] C. Zhang, D. Zeng, J. Li, F.-Y. Wang, W. Zuo, Sentiment analysis of chinese documents: from sentence to document level, J. Am. Soc. Inform. Sci. Technol. 60 (12), 2009, pp. 2474–2487.
[243] S. Li, L. Zhou, Y. Li, Improving aspect extraction by augmenting a frequency-based method with web-based similarity measures, Information Processing and Management 51 (2015) 58–67.
[244] O. Owoputi, B. O'Connor, C. Dyer, K. Gimpel, N. Schneider, and N.A. Smith, Improved Part-of-Speech Tagging for Online Conversational Text with Word Clusters, In *HLT-NAACL*, pp. 380-390. 2013.
[245] C. Paice, Another stemmer, SIGIR Forum 24 (3) (1990) 56–61.
[246] K. Atkinson, Gnu Aspell 0.60.4 (2006).
[247] M.F. Porter, Snowball: A language for stemming algorithms, (2001).
[248] K. Toutanova, D. Klein, C. Manning, and Y. Singer, Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network, In *Proceedings of HLT-NAACL 2003*, pp. 252-259.
[249] L. Kong, N. Schneider, S. Swayamdipta, A. Bhatia, C. Dyer, and N.A. Smith, A dependency parser for tweets, InProceedings of the Conference on Empirical Methods in Natural Language Processing, Doha, Qatar, to appear. 2014.
[250] L. Derczynski, A. Ritter, S. Clarke, and K. Bontcheva. 2013, Twitter Part-of-Speech Tagging for All: Overcoming Sparse and Noisy Data, In Proceedings of the International Conference on Recent Advances in Natural Language Processing, ACL.
[251] B. O'Connor, M. Krieger, D. Ahn. TweetMotif: Exploratory Search and Topic Summarization for Twitter. ICWSM-2010.

http://dx.doi.org/10.1016/j.knosys.2015.06.015