Annals of Library and Information Studies
Vol. 65, September 2018, pp. 196-209

Role of sentiment classification in sentiment analysis: a survey

Pavan Kumar M R[a] and Prabhu J[b]

[a]Research Scholar, School of Computer Science & Engineering, VIT, Vellore, Tamilnadu, India,
E-mail: marabattinasivapavan@gmail.com

[b]Associate Professor, School of Information Technology & Engineering, VIT, Vellore, Tamilnadu, India,
E-mail: j.prabhu@vit.ac.in

Received: 10 May 2018; revised and accepted: 04 September 2018

Through a survey of literature, the role of sentiment classification in sentiment analysis has been reviewed. The review identifies the research challenges involved in tackling sentiment classification. A total of 68 articles during 2015 – 2017 have been reviewed on six dimensions viz., sentiment classification, feature extraction, cross-lingual sentiment classification, cross-domain sentiment classification, lexica and corpora creation and multi-label sentiment classification. This study discusses the prominence and effects of sentiment classification in sentiment evaluation and a lot of further research needs to be done for productive results.

Keywords*:* Accuracy; Binary Classification; Dataset; Lexicon; Machine learning; Sentiment analysis

Introduction

With the increase in usage of internet, the volume of data being generated has considerably increased. The generation of data comes from various sources viz., social media (Twitter, Facebook, LinkedIn, Instagram etc.), micro-blogging (Twitter.com, Tumblr.com, foursquare.com etc.), e-commerce (taobao.com, amazon.com, ealmart.com, etc.), online transactions, discussion forums, surveillances, click streams, news reports etc,. People share their opinion about a product, film, event, news, politics, etc., in the form of short to moderately sized text messages. These reviews can be analysed to get useful insights for consumers and businesses. Product reviews are helpful to identify the satisfactory level of the customer about the product[1]. Movie databases can be used to predict the commercial success of a movie[2]. Opinions involved in news articles can be extracted to identify the mood of the public towards a political party or government schemes[3]. Blog sources can be used to detect the sentiment towards financial companies[4]. Public emotion towards a political party can be predicted using tweets[5]. The mood of the stock market can also be predicted based on public opinion available online[6]. In digital libraries, by analysing the past usage history of users a new book recommender system can be built for personalized recommendations.

In general, the sentiment orientation of a user towards an entity (product, movie, event or person) could be either in favour of the entity (positive polarity) or against the entity (negative polarity). Sometimes, other sentiment orientation is also possible i.e., neutral (user is not either in favour of or against an entity). The process of computing polarity from the text available online is said to be polarity classification or sentiment classification and is a basic step of sentiment analysis.

Polarity detection can be done at different levels: document, sentence and feature. The basic unit in document level sentiment analysis is the whole document (on a single topic). The sentences that are beside the point/topic in the document should be removed before processing. The entire document is classified as either positive or negative. The basic unit in sentence level sentiment analysis is each sentence in the document. Each sentence can be either subjective or objective. Objective sentence has facts and thus contains no polarity. For example, *John is studying sixth grade*. There is no opinion involved in it. Whereas, subjective sentence contains opinion words. Depending on those opinion words each subjective sentence is classified as either positive or negative based on identified key opinion/sentiment words. In feature level sentiment analysis, identifying the features of item is the first step. Then, we find out the sentiment of the feature directly. For example, *weight of the laptop is low*. In this short review '*weight*' is the item feature and '*low*' is the opinion (adjective) of that feature.

Most of the research on sentimental analysis in classification of text is done based on three approaches: lexicon based approaches, machine learning based approaches and hybrid approaches. Ensemble approaches are recently gaining much importance in solving text classification problems.

Lexicon based method first identifies a set of sentiment words, also called sentiment lexicons from the textual content. Identification of sentiment words can be done using three approaches: manually, dictionary-based and corpus-based. The manual approach of identifying sentiment words is effective but time consuming. Dictionary-based approaches first identifies the sentiment words manually and then expands the list using the synonyms of the identified sentiment words in the corpora (WordNet, SentiWordNet, HowNet, SenticNet etc.,) already available. Corpus-based approaches are used in finding the sentiment words in context-specific classification. Here, a set of sentiment words are identified with known polarity. This set is expanded by exploiting the syntactic patterns of co-occurrence patterns using a corpus. Then they apply predefined rules related to the language of the textual content to classify the sentiments.

Machine learning approaches build a sentiment classifier using a set of features selected from the textual content with the help of labelled corpora. The task of feature selection can be solved using either lexicon based or statistical methods. Lexicon based methods need human annotation, whereas statistical methods are automatic. The most popularly used feature selection algorithms are Information Gain (IG), Document Frequency (DF), CHI Statistics (CHI) and Gain Ratio (GR). The most used traditional classifiers are Support Vector Machines (SVM), Naïve Bayes (NB), Decision Trees (DT), K-Nearest Neighbour (KNN), Artificial Neural Network (ANN), Random Forest, Linear Regression, Logistic Regression, etc. Semi-supervised learning and un-supervised learning are also used in sentiment classification in the absence of labelled data for training a classifier.

In hybrid approaches, combinations of techniques are used from both lexicon approaches and machine learning approaches to tackle sentiment classification. Recently ensemble methods have become the research area of interest for sentiment classification. In this array of classifiers is used to improve the classification accuracy of sentiment classification. Knowing the importance of analysing the sentiment, it is important to introduce the approaches, techniques, tools, issues and applications involved in the process to the novice researchers, to address the challenges to give a better new solution.

Objectives of the study

- To present the importance of sentiment classification in real world applications;
- To introduce the approaches, techniques, tools and applications involved in the process of sentiment classification; and
- To identify the research challenges involved in tackling sentiment classification.

Methodology

A search was performed for the period 2015-2017 on the major database and journal platforms viz., IEEE, ScienceDirect, Springer and ACM using the search strings 'sentiment analysis', 'opinion analysis', 'sentiment classification' and 'supervised learning'. Papers not relatec to were excluded from the study.

The current review on the role of sentiment classification in sentiment analysis is conducted in six broad dimensions viz., Sentiment Classification, Feature Extraction, Cross-Lingual Sentiment Classification, Cross-Domain Sentiment Classification, Lexica and Corpora Creation and Multi-Label Sentiment Classification tasks. All the 68 articles studied are categorized into these six dimensions. The dimensions are the tasks that have to be performed in sentiment analysis. For each article methodology involved, issues addressed, classifiers used, datasets used and the performance of each technique are given using the metric classification accuracy. A number of research challenges have been identified from this study, exemplified and listed in research challenges subsection. For reader convenience, a quick review of what happened during 2017 in the domain of sentiment analysis in the perspective of sentiment classification is given in the form of Annexures 1 and 2. Annexure 1 lists technique used, issues addressed, datasets used and classification accuracy. Annexure 2 lists the corresponding applications, limitations identified and classifier used in the same order of preference shown in the Annexure 1.

Research challenges

1. Most research in SA is to identify the sentiment polarity hidden in the text which is written in one language. The research to identify the sentiment appears in the text written in multiple languages is under explored.
2. The process of identifying the proper meaning of a word that has a different sense in a different context is said to be Word Sense Disambiguation (WSD). For example, the size of laptop is small; in this review, small is having positive polarity. The size of room is small; in this review small is having negative polarity in case of hotels. Thus, understanding the right sense of the word depending on the context is very important to increase classification accuracy.
3. Labelled sentiment data is limited mostly to use languages like English, Chinese, Spanish and Arabic. A serious challenge is how to leverage labelled sentiment data in one language to classify the sentiments in other language.
4. Extracting polarity from comparative sentences is a serious challenge. For example, the processing speed of mobile X is faster than mobile Y. In this review faster is a positive sentiment, but identifying the objective of the opinion holder is difficult to estimate.
5. Most research in sentiment analysis approaches in text classification are oriented towards binary classification (i.e., "positive" and "negative") or ternary classification (i.e., 'positive", "negative" and "neutral"). The classification accuracy tends to decrease, along with the increase in the subdivision of binary and ternary classes (eg., especially on twitter data). Sophisticated methods are required to classify the text into multiple classes.
6. Sentiment classification suffers from occurrence of negation words in sentences. For example the size of mobile X is not good, but works better. In this review the scope of the word not is limited to the immediate next word. Consider another review Sensors of the tap do not give better lifetime. In this review the scope of the word not is throughout the sentence after it.
7. Detecting sarcasm in sentences to analyse sentiment expressed in the text is a challenge. For example, you better leave this place for my happiness. In this review, better is a positive word in negative context.
8. The sentiment analysis in naturalistic audio (extracting sentiment from audio sources like YouTube and UT-opinion opinion audio archive) is an area where research is under explored.

Sentiment classification

A SentiWordNet based Vocabulary (SWN-V) framework was built for improved opinion analysis and sentiment classification based on SWN-V. Experimental results on Cornell movie review dataset, large movie review dataset and multi-domain dataset (books, clothes, dvd, health and video) using a SVM classifier showed an improvement of 13.4% increase in accuracy compared to baseline methods[7]. An unsupervised approach was discussed[8] to identify the polarity of tweets, reviews and online texts based on dependency parsing. Experimental results proved the effectiveness of the proposed approach on OMD dataset, Cornell movie review dataset and SemEval-2015 Task 10 dataset. Chen et al.[9] explored a new term weighting scheme Term Frequency & Inverse Gravity Moment (TF & IGM) along with its variants. Besides that, a statistical model was also built that describes the inter-class distribution of each term in the text corpus. It also computes the class distinguishing power of a term in the text corpus. Junejo et al.[10] constructed a term-based discriminative information space for classification of text based on discriminative term weights that compute the power of each term and linear opinion pooling that accumulates term discriminative powers of each document to generate discriminative information scores.

DaSilva et al.[11] presented a framework for sentiment classification that combines unsupervised information with a classifier. The proposed approach outperforms the baseline methods in tweet sentiment classification. Wu et al.[12] addressed the issue of sentiment classification using a framework: structured micro-blog sentiment classification. This framework combines textual content information with social context information to improve the classification accuracy of micro-blog text messages. The performance of proposed approach was investigated on OMD twitter dataset and HCR twitter dataset. Rao et al.[13] described topic-level maximum entropy model for the classification of social opinion in short texts. They combined topic indicators with word tokens to address the over-fitting problem of maximum entropy principle. The proposed approach outperforms the baseline models in terms of accuracy on a dataset that contain short documents. Wang et al.[14] addressed the issue of noisy labels on sentiment classification using a probabilistic frame work. The framework consists of a hidden de-noising classification model that explores the relation between user and labels. The effectiveness of the proposed model is evaluated on Stanford twitter sentiment corpus (STS) and International survey on emotion antecedents and Reactions (ISEAR).

Perikos et al.[15] proposed an ensemble classifier that consists of two statistical classifiers: naïve bayes, maximum entropy and one knowledge based classifier tool. The statistical learners were trained on ISEAR and Affective Text datasets. Knowledge-based classifier tool is used to identify word dependencies and to spot words that convey the sentiment. The performance of the system was evaluated using articles, social media posts and news headlines. Zhang et al.[16] notified a framework that learns topic hidden in text based on the corpus, learns vector representation of word/topic for feature representation and builds a classifier for short text classification. The proposed approach was evaluated on web search snippets database and proved effective. Onan et al.[17] investigated the performance of five statistical keyword extraction methods on text classification algorithms and ensemble methods for classification of a scientific text document. The keyword extraction methods used were a TextRank algorithm, term frequency-inverse sentence frequency based keyword extraction, a most frequent measure based keyword extraction, eccentricity based keyword extraction and co-occurrence statistical information based keyword extraction. They compared Random forest, naïve bayes, support vector machine and logistic regression classification algorithms with five ensemble methods AdaBoost, dagging, majority voting, random subspace and stacking. Saif et al.[18] introduced an approach SentiCircle, to capture the contextual semantics of words that adjust polarity and pre-defined strength of a sentiment in sentiment lexicon. They evaluated the proposed approach at the entity level and tweet level on OMD, HCR and STS-Gold twitter data sets. Experimental results showed that the proposed approach outperformed the baseline approaches.

Muhammad et al.[19] designed a system for sentiment classification: SmartSA. To improve the classification accuracy SmartSA considered contextual polarities of terms. They also presented an approach using distant supervised learning to bridge the vocabulary differences in the general specific lexicon and its domain. Tripathy et al[20] classified movie reviews using different n-gram techniques on Maximum Entropy (ME), Naïve Bayes (NB), Support Vector Machine (SVM) and Stochastic Gradient Descent (SGD) classifiers. Experimental results showed that classification accuracy is more for lesser values of n and less for bigger values of n. Xia et al[21] described a framework that deals with polarity shift in sentiment classification at document level. First, a hybrid model was designed to detect polarity shifts that employ rule-based and statistic-based methods. Second, a novel method that eliminates polarity shifts in negations called antonym reversion is introduced. Finally, a weighted ensemble sentiment classifier was built for sentiment classification. The framework was evaluated using three classifiers: linear SVM, Naïve Bayes and logistic regression on product reviews taken from multiple domains. Ren et al[22] proposed a recursive auto-encoder for twitter sentiment analysis that encodes topic information into word embedding. Further, they used traditional features to develop a universal classifier for tweet classification. Experimental results on SemEval-2014 dataset proved that the proposed word embedding method outperforms other methods for word representation.

Altinel et al.[23] notified a Hybrid Class Semantic Classifier which is a semi-supervised algorithm for text classification. This approach makes use of weights of terms and class-based meaning values. Experimental results showed that HCSC outperforms the baseline methods on document datasets. Appel et al[24] demonstrated a hybrid approach for sentence level sentiment analysis problem. This approach combines the components: enriched sentiment lexicons, semantic rules for modelling and fuzzy sets to compute sentiment polarity. Experimental results on Twitter dataset A and Twitter dataset B obtained an accuracy of 88.02% and 75.85%. Onan et al[25] explored an ensemble method based on static classifier. This method incorporates Logistic regression, bayesian logistic regression, linear discriminant analysis, naive bayes and support vector machines as base learners. Experiments were conducted by applying a weighted voting scheme on learning algorithms and nine public sentiment analysis data sets and achieved an accuracy of 98.86%. Hu et al[26] discussed the usage of the results obtained in active learning process in the classification applications using different classifier types: Support Vector Machines, K-Nearest Neighbour and Naïve Bayes. Experimental results on datasets showed that Naïve Bayes classifier is the best to use as a selection strategy algorithm in active learning and the same classifier in sample reuse.

Korkontzelos et al[27] notified an improved ADRmine approach with sentiment features that extracts adverse drug reactions associated with negative sentiments. Investigations on DailyStrength posts and tweets showed that the proposed approach improves identification of ADR mention in health related forum messages and tweets. Lochter et al[28] described an ensemble model that combines text processing techniques with conventional classification approaches to extract the sentiment in short texts. The effectiveness of the proposed method is evaluated using nine data sets: STS-Test, HCR, OMD, SS-Tweet, Sanders, UMICH, Iphone6, Archeage and Hobbit using variants of Naïve Bayes, Support Vector Machine, Decision Trees, logistic regression and K-Nearest Neighbour. Kauer et al[29] proposed an Information Retrieval based method SABIR for polarity identification. Lima et al[30] discussed a hybrid framework to automate the process of tweet polarity classification. In this framework, a lexicon based approach that contains sentiment words and emoticons is combined with machine learning approach to automate tweet classification process. Experimental results on four datasets: OMD, SS-Twitter, Sanders, SemEval and four classifiers: Naïve Bayes, Maximum Entropy, Decision Trees(J48) and Support Vector Machine classified twitter messages with 80% accuracy.

Hogenboom et al[31] showed the usage of new structure based features that allow a better way of representing natural language text. They evaluated the performance of structure based features on English movie reviews and reviews of different products viz. books, electronics, dvds and kitchen appliances using machine learning approach. Zhang et al[32] explored that finding semantic features between words gives better classification accuracy. They used word2vec and SVM^{perf} to classify the Chinese comments on clothing products. They used word2vec to compute the semantic relationship between words in the document. Lexicon-based approach and POS based approaches are used for feature selection and then SVM^{perf} classifier is used for training and classification. The measures used to evaluate the proposed method are Precision, Recall and F-value and accuracy. Experimental results showed that their proposed method sentiment classification based on word2vec and SVM^{perf} reaches over 90% accuracy.

Wang et al[33] presented a sentiment classification method based on part-of-speech analysis using random subspace ensemble method. They employed two parameters to control the balance between diversity of base learners and accuracy: content lexicon subspace rate and function lexicon subspace rate. Experiments were conducted on proposed method using ten publicly available data sets on camera, camp, doctor, drug, laptop, layer, music, movie, radio and TV. Experimental results showed that proposed method produced best results and simultaneously reduced bias and variance. Colace et al[34] discussed a probabilistic method to classify the sentiment of textual documents. They learned mixed Graph Term structures (mGTs) from training text documents having a sentiment orientation. They investigated the performance of proposed approach on movie reviews data set and obtained an accuracy of 88.5%. Nguyen et al[35] presented a method that predicts stock market using sentiment topic. To capture the sentiment topic association two sentiment models were used: JST model based on latent topic and Aspect-based model that relies on proposed method. SVM with a linear kernel is used as the prediction model.

Mohammad et al[36] used US presidential elections 2012 tweets to annotate text for sentiment. They developed automatic supervised classifiers to detect emotion stimulus, emotional state and intent of the tweets. Williams et al[37] investigated the role of idioms as features in sentiment analysis. A corpus was also built that constitutes idioms in sentences to measure the performance of sentiment analysis.

Feature extraction

Uysal et al[38] integrated a one-sided feature selection approach with a global feature selection method to design an Improved Global Feature Selection Scheme (IGFSS). In this scheme each class in the dataset is equally represented by the feature set. The proposed scheme was evaluated on three datasets: Reuters-21578 ModApte Split, WenKB, Classic3 using SVM and NaïveBayes classifiers. Macro-F1, Micro-F1 were used as performance metrics. Bandhakavi et al[39] presented a new feature extraction method based on domain-specific lexicons to quantify neutrality and emotionality of words that represent documents along sentiment concepts. The proposed method was examined using a multi-class SVM classifier on four datasets: blogs, incident reports, news headlines and tweets.

Cross-lingual sentiment classification

Lin et al[40] presented an unsupervised framework for aspect based cross-lingual topic modelling in sentiment classification. They incorporated Joint Sentiment Topic, Aspect and Sentiment Unification model into the framework to build Cross-Lingual JST and Cross-Lingual ASUM. They investigated the framework on hotel reviews and product reviews collected from popular websites in different languages: English, Chinese, Spanish, German, French, Italian and Dutch. Hajmohammadi et al[41] notified a model for sentiment classification that combines semi-supervised self-training and uncertainty-based active learning to perform sentiment classification across languages. Self-training is an iterative process that labels unlabelled data and adds high confidence selected examples to the training dataset. Informative examples are selected using the density of the examples and active learning. Experimental results inferred that proposed method outperforms traditional methods on book review datasets in three languages.

Cross-domain sentiment Classification

Franco Salvador et al[42] designed a Stacked Generalization meta-learning scheme for cross domain sentiment classification. This scheme contains a bag of words classifier, word n-gram classifier, lexical resource-based classifier, vocabulary expansion-based classifier and word sense disambiguation-based classifier. The latter two classifiers were trained on knowledge graphs using BabelNet multilingual sentiment network. This scheme produced stable results on domains like books and dvds. Zhang et al.[43] presented TPF: Transferring the polarities approach for sentiment classification in cross domains. They transferred the polarities of independent features in the source domain to target domain to solve feature divergence and polarity divergence. They investigated the approach on RevDat dataset that has amazon reviews on four products: books, dvds, electronics and kitchen appliances.

Lexicon & corpora creation

Wu et al[44] derived a sentiment lexicon for Chinese micro-blogs by embedding three types of sentiment knowledge: word-sentiment knowledge, sentiment similarity knowledge and prior sentiment knowledge. They also presented a new approach to improve the coverage of sentiment lexicon by identifying popular terms mentioned by users in micro-blogs. Park et al[45] designed a method to extract sentiment lexicon specific to a domain. This method addressed two issues: converting the sentiment representation at document level to word level using a probabilistic algorithm that uses Bayesian estimation and efficiently selecting the document representation to minimize the document representation effort. The proposed method is evaluated by comparing the extracted lexicon with SentiWordNet and other baselines and it produced better F1 scores. Qi et al[46] proposed a method to compute the subjective well-being of Chinese context. This method used five weighted emotions in positive affect and negative affect. To compute the Chinese people subjective well-being they built a unique lexicon Ren-CECps-SWB 2.0. The proposed method is validated on panel data from Sina.com.

Multi-label classification

Liu et al[47] notified a multi-label classification prototype for sentiment classification of micro-blogs.

The three components of the proposed prototype are text segmentation using Jieba python package, feature extraction & representation and multi-label classification. As many as 11 methods were compared on two data sets (Huangpu River dead pigs and Influenza a virus subtype H7N9) with 8 evaluation metrics. The sentiment dictionaries used for polarity identification are DUTSD, NTUSD and HD. Authors considered 5 sentiment labels under HR data set and 10 sentiment labels under IA data set. The results showed that Dalian University of Technology Sentiment Dictionary (DUTSD) outperformed the other two dictionaries.

Conclusion

This paper presents a rigorous review on the research work done in the area of different aspects of sentiment classification during 2015 and 2017. This paper presents the approaches to build recommender systems like online movie recommender system, digital library recommender systems, e-commerce recommender systems, by analysing the sentiments and past history of the users. This paper is reviewed on six broad dimensions viz. Sentiment classification, Feature extraction, Cross-lingual sentiment classification, Cross-domain sentiment classification, lexicon & corpora creation and multi-label classification. Various research challenges and techniques involved in the process of tackling sentiment analysis have been discussed at layman level. A quick review of research on sentiment analysis techniques is also given in the form of tables. This paper also presents a number of publicly available datasets to the reader.

References

1. Dave K, Lawrence S, Pennock D M, Mining the peanut gallery: Opinion extraction and semantic classification of product reviews, In Proceedings of the 12th international conference on World Wide Web ACM, 2003, p. 519-528. 10.1145/775152.775226.
2. Mishne G, Glance NS, Predicting Movie Sales from Blogger Sentiment, In AAAI spring symposium: computational approaches to analysing weblogs, 2006, p. 155-158.
3. Godbole N, Srinivasaiah M, Skiena S, Large-Scale Sentiment Analysis for News and Blogs, ICWSM, 2007, p. 219-222.
4. O'Hare N, Davy M, Bermingham A, Ferguson P, Sheridan P, Gurrin C, Smeaton AF. Topic-dependent sentiment analysis of financial blogs[C]. *In* Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion ACM, 2009, p. 9-16. 10.1145/1651461.1651464.
5. Bakliwal A, Foster J, van der Puil J, O'Brien R, Tounsi L, Hughes M, Sentiment analysis of political tweets: Towards an accurate classifier, *Association for Computational Linguistics*, 2013.
6. Bollen J, Mao H and Zeng X, Twitter mood predicts the stock market, *Journal of Computational Science*, (2011) 1-8. http://dx.doi.org/10.1016/j.jocs.2010.12.007.
7. Khan FH, Qamar U and Bashir S, eSAP: A decision support framework for enhanced sentiment analysis and polarity classification, *Information Sciences*, 367 (2016) 862-873. http://dx.doi.org/10.1016/j.ins.2016.07.028
8. Fernández-Gavilanes M, Álvarez-López T, Juncal-Martínez J, Costa-Montenegro E and González-Castaño FJ, Unsupervised method for sentiment analysis in online texts, *Expert Systems with Applications*, 58 (2016) 57-75. https://doi.org/10.1016/j.eswa.2016.03.031
9. Chen K, Zhang Z, Long J and Zhang H, Turning from TF-IDF to TF-IGM for term weighting in text classification, *Expert Systems with Applications*, 66 (2016) 245-260. http://dx.doi.org/10.1016/j.eswa.2016.09.009
10. Junejo KN, Karim A, Hassan MT and Jeon M, Terms-based discriminative information space for robust text classification, *Information Sciences*, 372 (2016) 518-538. https://doi.org/10.1016/j.ins.2016.08.073
11. da Silva NF, Coletta LF, Hruschka ER and Hruschka Jr ER, Using unsupervised information to improve semi-supervised tweet sentiment classification, *Information Sciences*, 355 (2016) 348-365. http://dx.doi.org/10.1016/j.ins.2016.02.002
12. Wu F, Huang Y and Song Y, Structured microblog sentiment classification via social context regularization, *Neurocomputing*, 175 (2016) 599-609. http://dx.doi.org/ 10.1016/j.neucom.2015.10.101.
13. Rao Y, Xie H, Li J, Jin F, Wang FL and Li Q, Social emotion classification of short text via topic-level maximum entropy model, *Information & Management*, 53 (8) (2016) 978-986. http://dx.doi.org/10.1016/j.im.2016.04.005
14. Wang Y, Rao Y, Zhan X, Chen H, Luo M and Yin J, Sentiment and emotion classification over noisy labels, *Knowledge-Based Systems*, 111 (2016) 207-216. http://dx.doi.org/10.1016/j.knosys.2016.08.012
15. Perikos I and Hatzilygeroudis I, Recognizing emotions in text using ensemble of classifiers, *Engineering Applications of Artificial Intelligence*, 51 (2016) 191-201. http://dx.doi.org/10.1016/j.engappai.2016.01.012.
16. Zhang H and Zhong G, Improving short text classification by learning vector representations of both words and hidden topics, *Knowledge-Based Systems*, 102 (2016) 76-86. http://dx.doi.org/10.1016/j.knosys.2016.03.027.
17. Onan A, Korukoğlu S and Bulut H, Ensemble of keyword extraction methods and classifiers in text classification, *Expert Systems with Applications*, 57 (2016) 232-247. http://dx.doi.org/10.1016/j.eswa.2016.03.045.
18. Saif H, He Y, Fernandez M and Alani H, Contextual semantics for sentiment analysis of Twitter, *Information Processing & Management*, 52 (1) (2016) 5-19. http://dx.doi.org/10.1016/j.ipm.2015.01.005.
19. Muhammad A, Wiratunga N and Lothian R, Contextual sentiment analysis for social media genres, *Knowledge-based*

Systems, 108 (2016) 92-101. http://dx.doi.org/10.1016/j.knosys.2016.05.032.

20. Tripathy A, Agrawal A and Rath SK, Classification of sentiment reviews using n-gram machine learning approach, *Expert Systems with Applications*, 57 (2016) 117-126. http://dx.doi.org/10.1016/j.eswa.2016.03.028.
21. Xia R, Xu F, Yu J, Qi Y and Cambria E, Polarity shift detection, elimination and ensemble: A three-stage model for document-level sentiment analysis, Information Processing & Management, 52 (1) (2016) 36-45. http://dx.doi.org/10.1016/j.ipm.2015.04.003.
22. Ren Y, Wang R and Ji D, A topic-enhanced word embedding for twitter sentiment classification, *Information Sciences*, 369 (2016) 188-198. http://dx.doi.org/10.1016/j.ins.2016.06.040.
23. Altınel B and Ganiz MC, A new hybrid semi-supervised algorithm for text classification with class-based semantics, *Knowledge-Based Systems*, 108 (2016) 50-64. http://dx.doi.org/10.1016/j.knosys.2016.06.021.
24. Appel O, Chiclana F, Carter J and Fujita H, A hybrid approach to the sentiment analysis problem at the sentence level, *Knowledge-Based Systems*, 108, (2016) 110-124. http://dx.doi.org/10.1016/j.knosys.2016.05.040.
25. Onan A, Korukoğlu S and Bulut H, A multiobjective weighted voting ensemble classifier based on differential evolution algorithm for text sentiment classification, Expert Systems with Applications, 62 (2016) 1-6. http://dx.doi.org/10.1016/j.eswa.2016.06.005.
26. Hu R, Mac Namee B and Delany SJ, Active learning for text classification with reusability, *Expert Systems with Applications*, 45 (2016) 438-449. http://dx.doi.org/10.1016/j.eswa.2015.10.003.
27. Korkontzelos I, Nikfarjam A, Shardlow M, Sarker A, Ananiadou S and Gonzalez GH, Analysis of the effect of sentiment analysis on extracting adverse drug reactions from tweets and forum posts, *Journal of Biomedical Informatics*, 62 (2016) 148-158. http://dx.doi.org/10.1016/j.jbi.2016.06.007.
28. Lochter JV, Zanetti RF, Reller D and Almeida TA, Short text opinion detection using ensemble of classifiers and semantic indexing, *Expert Systems with Applications*, 62 (2016) 243-249. http://dx.doi.org/10.1016/j.eswa.2016.06.025.
29. Kauer AU and Moreira VP, Using information retrieval for sentiment polarity prediction, *Expert Systems with Applications*, 61 (2016) 282-289. http://dx.doi.org/10.1016/j.eswa.2016.05.038.
30. Lima AC, de Castro LN and Corchado JM, A polarity analysis framework for Twitter messages, *Applied Mathematics and Computation*, 270 (2015) 756-767. http://dx.doi.org/10.1016/j.amc.2015.08.059.
31. Hogenboom A, Frasincar F, De Jong F and Kaymak U, Polarity classification using structure-based vector representations of text, *Decision Support Systems*, 74 (2015) 46-56. http://dx.doi.org/10.1016/j.dss.2015.04.002.
32. Zhang D, Xu H, Su Z and Xu Y, Chinese comments sentiment classification based on word2vec and SVMperf, *Expert Systems with Applications*, 42 (4) (2015) 1857-1863. http://dx.doi.org/10.1016/j.eswa.2014.09.011.
33. Wang G, Zhang Z, Sun J, Yang S and Larson CA, POS-RS: A Random Subspace method for sentiment classification based on part-of-speech analysis, *Information Processing & Managemen*t, 51 (4) (2015) 458-479. http://dx.doi.org/10.1016/j.ipm.2014.09.004.
34. Colace F, Casaburi L, De Santo M and Greco L, Sentiment detection in social networks and in collaborative learning environments, *Computers in Human Behaviour*, 51 (2015) 1061-1067. http://dx.doi.org/10.1016/j.chb.2014.11.090.
35. Nguyen TH, Shirai K and Velcin J, Sentiment analysis on social media for stock movement prediction, *Expert Systems with Applications*, 42 (24) (2015) 9603-9611. http://dx.doi.org/10.1016/j.eswa.2015.07.052
36. Mohammad SM, Zhu X, Kiritchenko S and Martin J, Sentiment, emotion, purpose, and style in electoral tweets, Information Processing & Management, 51 (4) 2015 480-499. http://dx.doi.org/10.1016/j.ipm.2014.09.003.
37. Williams L, Bannister C, Arribas-Ayllon M, Preece A and Spasić I, The role of idioms in sentiment analysis, *Expert Systems with Applications*, 42 (21) (2015) 7375-7385. http://dx.doi.org/10.1016/j.eswa.2015.05.039.
38. Uysal AK, An improved global feature selection scheme for text classification, *Expert Systems with Applications*, 43 (2016) 82-92. http://dx.doi.org/10.1016/j.eswa.2015.08.050.
39. Bandhakavi A, Wiratunga N, Padmanabhan D and Massie S, Lexicon based feature extraction for emotion text classification, *Pattern Recognition Letters*, 93 (2017) 133-142.http://dx.doi.org/10.1016/j.patrec.2016.12.009.
40. Lin Z, Jin X, Xu X, Wang Y, Cheng X, Wang W and Meng D, An unsupervised cross-lingual topic model framework for sentiment classification, *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 24 (3) (2016) 432-444. http://dx.doi.org/10.1109/TASLP.2015.2512041.
41. Hajmohammadi MS, Ibrahim R, Selamat A and Fujita H, Combination of active learning and self-training for cross-lingual sentiment classification with density analysis of unlabelled samples, *Information Sciences*, 317 (2015) 67-77. http://dx.doi.org/10.1016/j.ins.2015.04.003.
42. Franco-Salvador M, Cruz FL, Troyano JA and Rosso P, Cross-domain polarity classification using a knowledge-enhanced meta-classifier, *Knowledge Based Systems*, 86 (2015) 46-56. http://dx.doi.org/10.1016/j.knosys.2015.05.020.
43. Zhang Y, Hu X, Li P, Li L and Wu X, Cross-domain sentiment classification-feature divergence, polarity divergence or both?, *Pattern Recognition Letters*, 65 (1) (2015) 44-50. http://dx.doi.org/10.1016/j.patrec.2015.07.006.
44. Wu F, Huang Y, Song Y and Liu S, Towards building a high-quality microblog-specific Chinese sentiment lexicon, *Decision Support Systems*, 87 (2016) 39-49. http://dx.doi.org/10.1016/j.dss.2016.04.007.
45. Park S, Lee W and Moon IC, Efficient extraction of domain specific sentiment lexicon with active learning, *Pattern Recognition Letters*, 56 (2015) 38-44.http://dx.doi.org/10.1016/j.patrec.2015.01.004.
46. Qi J, Fu X and Zhu G, Subjective well-being measurement based on Chinese grassroots blog text sentiment analysis, *Information & Management*, 52 (7) 2015 859-869. http://dx.doi.org/10.1016/j.im.2015.06.002.

47. Liu SM and Chen JH, A multi-label classification based approach for sentiment classification, *Expert Systems with Applications*, 42 (3) (2015) 1083-1093. http://dx.doi.org/10.1016/j.eswa.2014.08.036.

48. Tellez ES, Miranda-Jiménez S, Graff M, Moctezuma D, Siordia OS and Villaseñor EA, A case study of Spanish text transformations for twitter sentiment analysis, *Expert Systems with Applications*, 81 (2017) 457-471. http://dx.doi.org/10.1016/j.eswa.2017.03.071.

49. Deng S, Sinha AP and Zhao H, Adapting sentiment lexicons to domain-specific social media texts, *Decision Support Systems*, 94 (2017) 65-76. http://dx.doi.org/10.1016/j.dss.2016.11.001.

50. Jha V, Savitha R, Shenoy PD, Venugopal KR and Sangaiah AK, A novel sentiment aware dictionary for multi-domain sentiment classification, *Computers & Electrical Engineering*, (2017) In Press. http://dx.doi.org/10.1016/j.compeleceng.2017.10.015.

51. Onan A, Korukoğlu S and Bulut H, A hybrid ensemble pruning approach based on consensus clustering and multi-objective evolutionary algorithm for sentiment classification, *Information Processing & Management*, 53 (4) (2017) 814-833. http://dx.doi.org/10.1016/j.ipm.2017.02.008.

52. Keshavarz H and Abadeh MS, ALGA: Adaptive lexicon learning using genetic algorithm for sentiment analysis of microblogs, *Knowledge-Based Systems*, 122 (2017) 1-6. http://dx.doi.org/10.1016/j.knosys.2017.01.028.

53. Catal C and Nangir M, A sentiment classification model based on multiple classifier, *Applied Soft Computing*, 50 (2017) 135-141. http://dx.doi.org/10.1016/j.asoc.2016.11.022.

54. Wu F, Huang Y and Yuan Z, Domain-specific sentiment classification via fusing sentiment knowledge from multiple sources, Information Fusion, 35 (2017) 26-37. http://dx.doi.org/10.1016/j.inffus.2016.09.001.

55. Gui L, Zhou Y, Xu R, He Y and Lu Q, Learning representations from heterogeneous network for sentiment classification of product reviews, *Knowledge-Based Systems*, 124 (2017) 34-45. http://dx.doi.org/10.1016/j.knosys.2017.02.030.

56. Liu Y, Bi JW and Fan ZP, Multi-class sentiment classification: The experimental comparisons of feature selection and machine learning algorithms, *Expert Systems with Applications*, 80 (2017) 323-339. http://dx.doi.org/10.1016/j.eswa.2017.03.042.

57. Lin J, Mao W and Zeng DD, Personality-based refinement for sentiment classification in microblog, *Knowledge-Based Systems*, 132 (2017) 204-214. https://doi.org/10.1016/j.knosys.2017.06.031.

58. Yu D, Mu Y and Jin Y, Rating prediction using review texts with underlying sentiments, *Information Processing Letters*, 117 (2017) 10-18. http://dx.doi.org/10.1016/j.ipl.2016.08.002.

59. Hung C, Word of mouth quality classification based on contextual sentiment lexicons, *Information Processing & Management*, 53 (4) (2017) 751-763. http://dx.doi.org/10.1016/j.ipm.2017.02.007.

60. Lee S and Kim W, Sentiment labeling for extending initial labeled data to improve semi supervised sentiment classification, *Electronic Commerce Research and Applications*, 26 (2017) 35-49. https://doi.org/10.1016/j.elerap.2017.09.006.

61. Bouazizi M and Ohtsuki T, A Pattern-Based Approach for Multi-Class Sentiment Analysis in Twitter, *IEEE Access*, 5 (2017) 20617-20639. http://dx.doi.org/10.1109/ACCESS.2017.2740982.

62. Kaushik L, Sangwan A and Hansen JH, Automatic Sentiment Detection in Naturalistic Audio, *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 25 (8) (2017) 1668-1679. http://dx.doi.org/10.1109/TASLP.2017.2678164.

63. Xia R, Jiang J and He H, Distantly Supervised Lifelong Learning for Large-Scale Social Media Sentiment Analysis, *IEEE Transactions on Affective Computing*, 8 (4) (2017) 480-491. http://dx.doi.org/10.1109/TAFFC.2017.2771234.

64. Liu Y, Bi JW and Fan ZP, A method for multi-class sentiment classification based on an improved one-vs-one (OVO) strategy and the support vector machine (SVM) algorithm, *Information Sciences*, 394 (2017) 38-52. http://dx.doi.org/10.1016/j.ins.2017.02.016.

65. Amplayo RK and Song M, An adaptable fine-grained sentiment analysis for summarization of multiple short online reviews, *Data & Knowledge Engineering*, 110 (2017) 54-67. http://dx.doi.org/10.1016/j.datak.2017.03.009.

66. Huang F, Zhang S, Zhang J and Yu G, Multimodal learning for topic sentiment analysis in microblogging, *Neurocomputing*, 253 (2017) 144-153. http://dx.doi.org/10.1016/j.neucom.2016.10.086.

67. Liu Y, Bi JW and Fan ZP, Ranking products through online reviews: A method based on sentiment analysis technique and intuitionistic fuzzy set theory, *Information Fusion*, 36 (2017) 149-161. http://dx.doi.org/10.1016/j.inffus.2016.11.012.

68. Daniel M, Neves RF and Horta N, Company event popularity for financial markets using Twitter and sentiment analysis, *Expert Systems with Applications*, 71 (2017) 111-124. http://dx.doi.org/10.1016/j.eswa.2016.11.022.

69. Giatsoglou M, Vozalis MG, Diamantaras K, Vakali A, Sarigiannidis G and Chatzisavvas KC, Sentiment analysis leveraging emotions and word embeddings, *Expert Systems with Applications*, 69 (2017) 214-224. http://dx.doi.org/10.1016/j.eswa.2016.10.043.

70. Pandey AC, Rajpoot DS and Saraswat M, Twitter sentiment analysis using hybrid cuckoo search method, *Information Processing & Management*, 53 (4) (2017) 764-779. https://doi.org/10.1016/j.ipm.2017.02.004.

71. Akhtar MS, Gupta D, Ekbal A and Bhattacharyya P, Feature selection and ensemble construction: A two-step method for aspect based sentiment analysis, *Knowledge-Based Systems*, 125 (2017) 116-135. http://dx.doi.org/10.1016/j.knosys.2017.03.020.

72. Chan SW and Chong MW, Sentiment analysis in financial texts, *Decision Support Systems*, 94 (2017) 53-64. http://dx.doi.org/10.1016/j.dss.2016.10.006.

73. Vilares D, Alonso MA and Gómez-Rodríguez C, Supervised sentiment analysis in multilingual environment, *Information Processing & Management*, 53 (3) (2017) 595-607. http://dx.doi.org/10.1016/j.ipm.2017.01.004.

74. Chen T, Xu R, He Y and Wang X, Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN, *Expert Systems with Applications*, 72 (2017) 221-230. http://dx.doi.org/10.1016/j.eswa.2016.10.065.

Annexure 1

Comparison of methodology and approach

Reference	Technique Used	Issues Addressed	Dataset Used	Accuracy (in %)
[48]	Combination of text transformation techniques, tokenizers, and token weighting schemes. Combination of word-based n - grams and character-based q-grams.	Text Modelling	INEGI Spanish tweets TASS'15 Spanish tweets	65.00 63.00
[49]	An automated approach to incorporate content domain opinions and language domain opinions in a sentiment lexicon.	Learning context-specific sentiment	Stock market tweets. Political Tweets.	-
[50]	POS tagging, lemmatization, Word filtering, creating unigram and bigram list.	Construction of sentiment-aware dictionary from multiple domain data.	Reviews from kitchen appliances, DVDs, electronics and books	76.00
[51]	Clustering and randomized based ensemble pruning algorithm, consensus clustering scheme	Text sentiment classification	12 public sentiment analysis datasets named camera, camp, drug, doctor, lawyer, laptop, TV and music.	90.00
[52]	Using lexicon and n-gram features besides created lexicon features.	Identifying optimum lexicons.	6 Twitter Data sets. Sanders OMD Strict OMD HCR SemEval 2013 Stanford	 85.71 82.90 85.85 82.61 93.70 88.40
[53]	Multiple classification system model with 3 classifiers and parameter optimization technique.	Usage of meta classifiers, Parameter optimization.	Turkish reviews on Movies(Beyazperde), Shopping(Hepsiburada), Books(Antologi)	 83.68 79.96 86.13
[54]	A framework to learn target domain by fusing four kinds of sentiment information.	Integrating other domain sentiment information into target domain sentiment classifier.	Amazon Product Review Data set. Book DVD Electronics Kitchen Twitter Data Set	 79.15 81.23 85.01 87.22 77.35

Reference	Technique Used	Issues Addressed	Dataset Used	Accuracy (in %)
[55]	Incorporating users, products, reviews, and polarities from a heterogeneous network into a single framework.	To link users, reviews, products using the words that occur in product reviews from the heterogeneous network.	IMDB reviews Yelp Data set 2013 Yelp Data set 2014	50.90 65.60 66.20
[56]	A multi-class classification framework using feature selection algorithm and machine learning algorithm.	Classification of online texts using 3-clasess, 4-classes, and 5-classes	Online texts from 3 public data sets.	DT-52.53 NB-60.46 SVM-68.14 RBFNN-63.3 KNN-50.70
[57]	Extracting features by considering Big Five Model to build a meta classifier.	Considering behavioural traits of opinion holders to extract features from micro-blog.	968,854 tweets from SinaWeibo micro-blog.	-
[58]	Considered different aspects in review text about an item along with ratings. Mapped user rating to sentiment distribution probability space.	Cold-Start Problem: lack of available ratings.	Amazon review data set.	-
[59]	Vector modelling method	Computing the quality of word of mouth document.	IMDb data set hotels.com data set	SVM Linear kernel-96.88 SVM RBF kernel-92.11 J48 DT-96.28 Naïve bayes-95.23 Note: accuracy for IMDb Bigram tokenization
[60]	Sentiment labelling and document embedding	-Incorrect labelling. -Insufficient amount of labelled data.	Movie reviews in korean language data set.	63.2 61.77 62.5
[61]	SENTA Tool	-Classifying twitter texts into multiple classes.	Twitter tweets	60.2(for multiple classes) 81.3(for binary classes) 70.1(for ternary classes)
[62]	-Spotting key word approach.	- sentiment extraction from audio sources.	-UT-Opinion -YouTube Videos	70.8 85.1
[63]	-Collecting text data over time and partition them. Each partition is considered as past data set. -Each new task will be compared to past data. Based on similarity the new task is either merged with past data or added to the past data set.	-training a sentiment classifier for massive data that contains different domains in social media.	-English Twitter data set. -Chinese Weibo data set.	72.95 79.06 (for english kitchen data set) 81.09 81.32 (for chinese hotel data set

Reference	Technique Used	Issues Addressed	Dataset Used	Accuracy (in %)
[64]	- Extracting features from online texts using information gain. -Computing sentiment classes of texts based on improved OVO strategy and SVM algorithm.	-classifying online texts into multiple classes.	-Movie review data set	75.61(average accuracy for k =20)
[65]	-Three-level classification approach for sentiment classification. -biterm topic model(eBTM) for aspect extraction.	-Analyzing short texts in social media	-Naver korien movie data set. -Douban chinese movie data set. -Rotten tomatoes English movie data set. -Amazon(English Products)	86 75 86 98
[66]	LDA and an extra sentiment layer.	-detection of topic -analysis of short messages in a microblog.	3 data sets from SinaWeibo	70.75 66.81 69.15
[67]	Combination of sentiment analysis and intuitionistic fuzzy set theory	Using online reviews to rank alternative products.	Online automobile reviews in chinese language	-
[68]	Defining financial community and extracting tweets. Filtering tweets. Performing sentiment analysis. Manual detection of events.	Detecting popularity of events in financial area in twitter.	Tweets	-
[69]	Usage of hybrid vectors for prediction of sentiment.	Multi-lingual text classification.	Movie and IMDb user reviews data set in english E-gadget user review data set in greek	74.49 87.80 78.57 83.60
[70]	A metaheuristic approach based on cuckoo search and k-means.	Classification of twitter sentiment data	Four Twitter data sets	68
[71]	Identifying the feature set for aspect term extraction. Applying sentiment classification on the feature set.	Feature extraction and sentiment classification	SemEval-2014 data set Restaurant data Laptop data	80.07 75.22
[72]	Feature extraction. Phrase recognizer. Parse tree resolution. Decision tree classification.	To identify the trends of SAE.	Movie review data set.	67.9 80.6
[73]	Constructing a code switching corpus. Soft-data fusion for sentiment analysis.	To identify the sentiment appear in a text written in multiple languages.	SemEval 2014 TASS 2014	-
[74]	Identifying target sentence group using a neural network. Feeding 1-D CNN with each sentence group.	Sentence level sentiment classification.	MR v1.0. SST-1 SST-2 CR	84.5 50.9 89.6 88.4 *For non-target sentence group.

Annexure 2

Lists applications, limitations and classifiers used

Reference	Applications	Limitations	Classifier Used
[48]	Micro-blogging platforms (Tumblr, FriendFeed, Twitter, etc.)	More memory requires along with an increase in q in q-grams.	SVM
[49]	Business intelligence.	Only positive and negative polarities considered. Neutral polarity not considered.	Lexicon based approach
[50]	Market analysis in unknown target domains.	Not suitable for context based domains.	-
[51]	Business Intelligence, Micro-blogging platforms	Need to check whether it is suitable for other type of sentiment classification.	Ensemble learning. (Combinations of classifiers)
[52]	Micro-blogging platforms	-Creation of lexicon is time consuming. -Difficult to build a lexicon that works in all domains.	-Lexicon + Corpus based approach
[53]	e-commerce, opinion polls	Did not discuss about the approach in multiple domains.	Combination of multiple classifiers.
[54]	Online Shopping, Micro-blogging.	-	-
[55]	e-commerce	-	Convolutional neural network.
[56]	Health care	SVM classifiers with different kernels not discussed.	Decision Trees, Naïve Bayes, Support Vector Machines. Radial bias function neural network, K-Nearest neighbour
[57]	Social management Public security Business security	-	Meta classifier
[58]	Recommender Systems	-Cold-start problem alleviated but continues to exist.	Aspect and Sentiment Unification model.
[59]	E-commerce	-doesn't discuss about the classification of the sentiment polarity of WOM	SVM Linear kernel SVM RBF kernel J48 Decision trees Naïve bayes
[60]	E-commerce	-Informative instances are not considered for sentiment labelling. -Experiments conducted with JST models (unsupervised) only.	Logistic regression
[61]	Micro-blogging	- Accuracy low for multi-class classification when compared to binary, ternary classification with the same approach.	Random forest
[62]	Social media	-	Maximum entropy

Reference	Applications	Limitations	Classifier Used
[63]	Social media	-	Naïve bayes. Logistic regression.
[64]	Micro-blogging	-number of classes is limited to quaternary.	SVM Classifier
[65]	-	- not suitable for languages that do not support tokenizers and POS tagging. - topic labelling problem.	A three level classifier. i. NLP classifier. ii. SVM iii. Feed forward neural network
[66]	Micro-blogging	-	Multimodal joint sentiment topic model
[67]	e-commerce	-	-
[68]	Twitter microblog.	Restricted to one domain.	Lexicon based approach.
[69]	Business intelligence	-	SVM linear kernel.
[70]	Social media	-	Clustering.
[71]	e-commerce and entertainment	-	Maximum Entropy Conditional Random Field Support Vector Machine
[72]	Financial markets and business intelligence.	-	Decision tree (C4.5)
[73]	Global marketing systems.	-limited to two languages English and Spanish.	-
[74]	e-commerce	Only one target expression detection technique is used.	1-D Neural Network.