

SENTIMENT ANALYSIS A DEEP SURVEY ON METHODS AND APPROACHES

Reena G. Bhati^{#1}

Computer Science Department, Tilak Maharashtra Vidyapeeth, Pune, India

Abstract

In this paper we present an in-depth analysis of sentiment. It also offers a survey and discusses the latest papers and techniques relevant to strategies for lexicon-based, conventional machine-based learning, deep learning-based, & hybrid sentiment study. The SA is used to forecast & examine sentiment polarities. The purpose of this research work is to study the existing systems which focus on sentiment analysis. Also the expansion of an integrated system for information discovery various social media, which permits the survey of person's individual opinion towards the any incident or things related to that person or towards society with the help of data analysis techniques.

Keywords- Sentiment Analysis, Natural Language Processing(NLP), Lexicon based methods.

I. INTRODUCTION

Recent years have seen growing research interest in analyzing the content of academic publications, specifically the citation context [1]. Mining authors' thoughts or sentiments in a large number of publications for different topics published every day in academic libraries is useful for improving citation analysis methods by considering the negative weight in its scheme. In earlier era, Newspapers, magazines & articles, blog are used to express people's reviews towards any incident or any other kind of event. Analysis of customer sentiment is not only useful for business but it has widened the scope in health care, sports, politics, recommendation system & various domains are using such techniques in direct or indirect mode. Nowadays, attractiveness of sentiment analysis is rapidly increasing as the no. of viewers is increased due to the sharing of their views. These reviews can be associated to an occasion, brand, person or product. Though, with the improvement in technology the peoples start writing & showing their emotions & what they are feeling towards any particular incident on different social networking medias like Facebook, Twitters, Instagram & Whatsapp etc. The Sentiment analysis is mainly classified into the three subtypes are as- Negative, Neutral & positive. The aim of this direction of research work is the perception & understanding what users' opinions are about particular any instance & what the effects, if a any person makes a positive & negative comment about that instant.

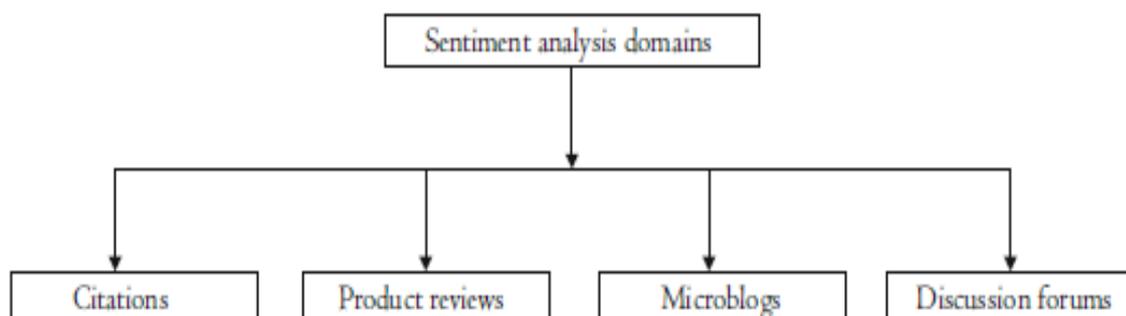


Figure no.1: Various Sentiment Analysis Domains [2]

The above fig 1 show the various sentiment analysis domains used for detecting the discussion forums, citations, product reviews & discussion forums[2]. The citation is the fresh & innovative research topic nowadays.

The rest of paper is organized as follows: section I is includes introduction, section II gives the idea about Literature Survey & Section III gives the information of sentiment classification & their approaches. Last section is conclusion.

II. LITERATURE SURVEY

This section gives idea of the literature scope, search methodology, search databases, various search terms used & elaborate the different the research methodology adopted for conducting the review study. For detection the scope of negation & determine the most accurate techniques with the help of various algorithms, comparison, developing the new prototype for the SA analysis technique. Sentiment analysis (SA)[9] is a developing field of research nowadays & many researchers choose are this topic as their research studies , like as computational linguistics, text mining, natural language processing & machine learning. The SA is nothing but the analyzing peoples sentiments, views attitudes, emotions & subjectivity of thoughts & process them into particular method. This study of SA helps to achieve many goals like projecting view regarding any movie, counting the customer satisfaction regarding any products & analyzing market survey by people's view[10,11,12].Huang et.al.[7] proposed neural language model that used local & global document contexts for better semantic word representations. These representations can be used to represent contexts for grouping word instances which are used in multi-prototype version of the model i.e. used for words with multiple senses. The ample of assessment of SA s given by Pang & Lee [15] in their research work. In the research study, mainly focused on challenges of SA, different tasks based on SA & their respective applications. Along with this SA types are also studied & gives enough information about types. The different tasks give like polarity determination, sentiment classification, sentiment extraction & opinion summarization. Cambria [16] represents the emotion recognition, polarity detection, opinion target identification, multi modal fusion & subjectivity detection as a research topic in [14]. This paper summarizes the techniques based on statistical methods, hybrid approaches, knowledge-based techniques & gives the ideas about different categories telling tat how to handle the different task of SA with their respective advantages & disadvantages.

Priyanka Tyagi et.al [2], used the Twitter data as input for performing sentiment analysis. In this technique firstly, preprocessing of Twitter data, tokenization, feature extraction & classification are carried out over twitter data after that calculating the polarity of data are done for this purpose SVM classifier are used. The following figure gives the short idea about main steps of SA. The reason behind using SVM classifier is it gives accuracy approx. 80% for SA. For increasing accuracy of SVM, naïve bayes & KNN approaches are used [13]. The python languages are used as coding language.

Basically the primary step of SA is data is collected from Twitter i.e. tweets that peoples are posting related to any kind information. Then Twitter tweets are extracted from public API & implement in PHP after that output query is generated by using MySQL database. This database is related to the information carried out from Twitter. The data processing is done by using NLP. Then Creation of Dataset, Preprocessing of Tweets, Creation of Feature Vector & Sentiment Classification are used for data analyzing.

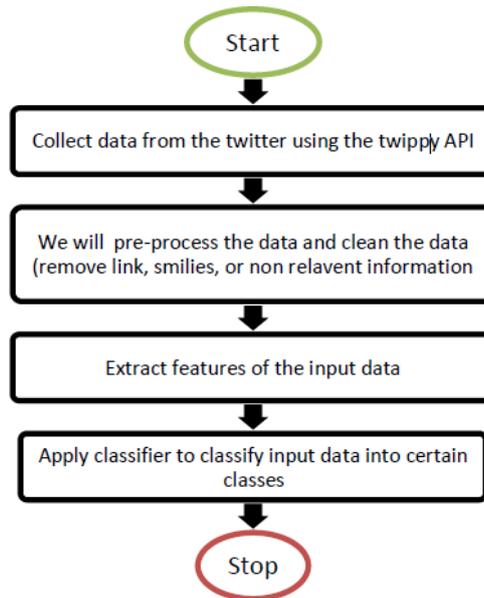


Figure no.2: Flowchart for sentiment analysis [2]

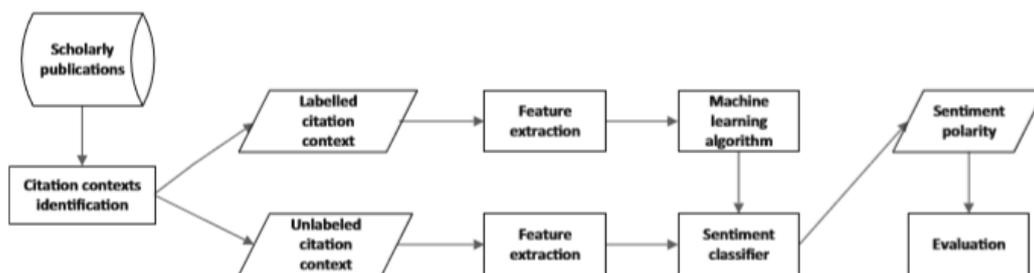


Figure no.3: The Process of SCSA[3]

In past recent era, huge demand of digital library is becomes very popular. The digital library used by many researchers for their studies. Hence, Abdallah Yousif et.al [2], used the research method for scientific citation sentiment analysis(SCSA). The following figure explains the basic process for scientific citation sentiment analysis. In this process, extract the information of citation of articles or IEEE papers which is used for research purpose. After this, the primary dataset is made up from extracted information which used as sentiment citation the machine learning approaches are used for performing SCSA.As figure showing, the extracted the context named as positive, negative or neutral. This named citation symbolized as training dataset. The Recognizing exact citation context dataset should be executed very sensibly. The Recognizing exact citation context dataset is very difficult task. In the second step, by using machine learning classifier the feature extraction are carried out. The appropriate feature set & choosing accurate improves the performance of classifier. In third step, labeled citation context as an input using machine learning classifier for categorizing model into positive, neutral or negative. In the last step, information of SCSA are recovered from evaluation methods like accuracy metrics, recall & precision, recall [11].

Usually, sentiment analysis is accomplished over a single data source, for instance, online product reviews or Tweets. Because of this, the necessity of development a more precise & more comprehensive result is arises. The multi-source sentiment analysis is nothing but performing a

sentiment analysis on datasets obtained from multiple data sources. The Above diagram fig no 3 shows what are the challenges occurred in sentiment analysis. This thesis is proposed by author Nor Aniza Abdullah et.al in [4]. The researchers gone thorough understanding on both multi-source & multi-domain sentiment analysis & deliberated the solutions over the problems.

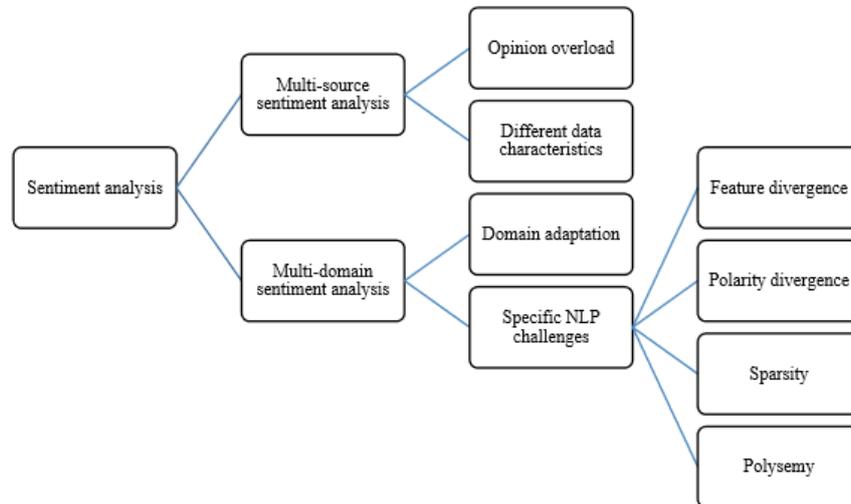


Figure no.4: Classification of Challenges in multi-source & multi-domain sentiment analysis [4]

The multi domain SA is done only when any similarities in characteristics are finding between any two domains whereas, dataset obtained from multiple data source is called as multisource SA & this SA is used for the observing of brands & products in social media networks [2-4]. The fig no describes the classification of challenges occurred in multi-domain & multi-source SA. These two SA care categorized into four main categories: domain adaptations, opinion overload, NLP challenge & different data characteristics. The following tables narrates overview the advantages & disadvantages of different approaches for different domain adaption challenges [4].

Table no.1 The advantages & disadvantages of different approaches for different domain adaption challenges [4]

Approaches	Advantages	Disadvantages
Transfer learning model	A well-established approach for cross domains sentiment analysis.	Require a new transfer model for every new domain, hence limits generalization.
Multi-task learning	It does not require to build a model prior to analysing new domain.	If the number of tasks is very huge, optimization algorithm is required to speed up the learning process.
Word embeddings	It provides explicit representation of words in regards to the context they appear, in which words with common contexts appear near to one another in the vector space.	The word-to-vector representation does not take into account the semantic meaning of words.
Sentiment lexicon	Does not need to build a transfer model prior to analysing any new domain.	A predefined set of sentiment lexicon is required.

The below figure 5 describes the procedure of SA which is developed by Claudia Diamantini et.al[5]. The diagram of process elaborates the typical NLP tasks, namely tokenization, POS tagging & lemmatization [5].

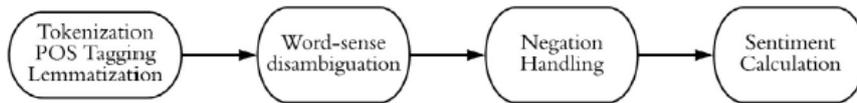


Figure no.5: The Sentiment Analysis Process [5]

After that one algorithm is used for polysemous words named as a word sense disambiguation algorithm which is listed below in fig no 5. As the name of the algorithm indicated that a single word having no. of meanings called as polysemous words. This algorithm chooses the correct meaning of word after analyzing the context of word & avoids the ambiguous word. The third step of this procedure negation handling is done.

Algorithm 1 Word Sense Disambiguation

```

    Let  $w_1$  be the term to be disambiguated,
    let  $w_2$  the term used for disambiguation,
    let  $S(x)$  be the set of polysemous variants of term  $x$ ,
    let  $G(y)$  be the set of terms in the gloss of the variant  $y$ ,
    let  $sent(y)$  be the function that returns the sentiment score
    for variant  $y$ ,
    let  $anc(x)$  be the ancestor of  $x$  which is child of  $w_1$ ,
    let  $Q$  be an empty queue of terms.
    1: function WSD( $w_1, w_2, depth$ )
    2:    $Q.push(w_1)$ 
    3:   while  $Q$  not empty &&  $depth \geq 1$  do
    4:      $t \leftarrow Q.pop()$ 
    5:     for each  $x$  in  $S(t)$  do
    6:       for each  $j$  in  $G(x)$  do
    7:         if  $w_2 == j$  then
    8:           return  $sent(anc(j))$ 
    9:          $Q.push(j)$ 
    10:     $depth \leftarrow (depth - 1)$ 
    11:  return null
  
```

Figure no.6: Word sense Disambiguation [5]

The negation handling is the way of altering the polarity of certain words that creates the negation word i.e. negative words. Claudia Diamantini et.al [5] proposed the algorithm based on depth-first search (DFS) strategy. This proposed algorithm used branching node method for finding negation word only which affects the terms by creating negative polarity. If any negation word is found then the algorithm flip-flops the polarity of its following sibling nodes. At the last score of sentiment calculation is done with the help of post or tweet of peoples by taking mean value of word's sentiment score. This paper represents a innovative set of structures that improve the classification accuracy of state-of-the-art noise detection algorithms for Twitter [7]. The Disambiguation of polysemous words & the correct handling of negated sentences these two algorithms give the superior results over the old-style the lexicon-based sentiment analysis. Eeshita Biswas et.al [10], innovates the system process for software domain specific word embeddings learned from Stack Overflow (SO) posts & studied the effect of software domain-specific word embeddings. The effect of these words embedding compared with the generic word embeddings learned from Google News. By comparing these the result found that are in some cases embeddings learned from the Google News data mostly similar with SO post

& sometimes embeddings learned from SO posts are good than Google News data. The comparison is based on up-sampling & over-sampling these two machine learning techniques.

Sentiment Classification Approaches

SCSA has been employed at three different levels: document, sentence, & entity levels. The analysis of document level aims to recognize the sentiment polarity conveyed in the entire document (the document, here is a citation context, & should be a paragraph). The sentence level seeks to classify each citation sentence as positive or negative whereas the entity-level finds the sentiment polarity of a particular entity of a specific object. There are four techniques for solving SCSA challenges that can be identified in the literature of SCSA, which are machine learning approaches, lexicon-based approaches, deep learning approaches, & keyword based approach.

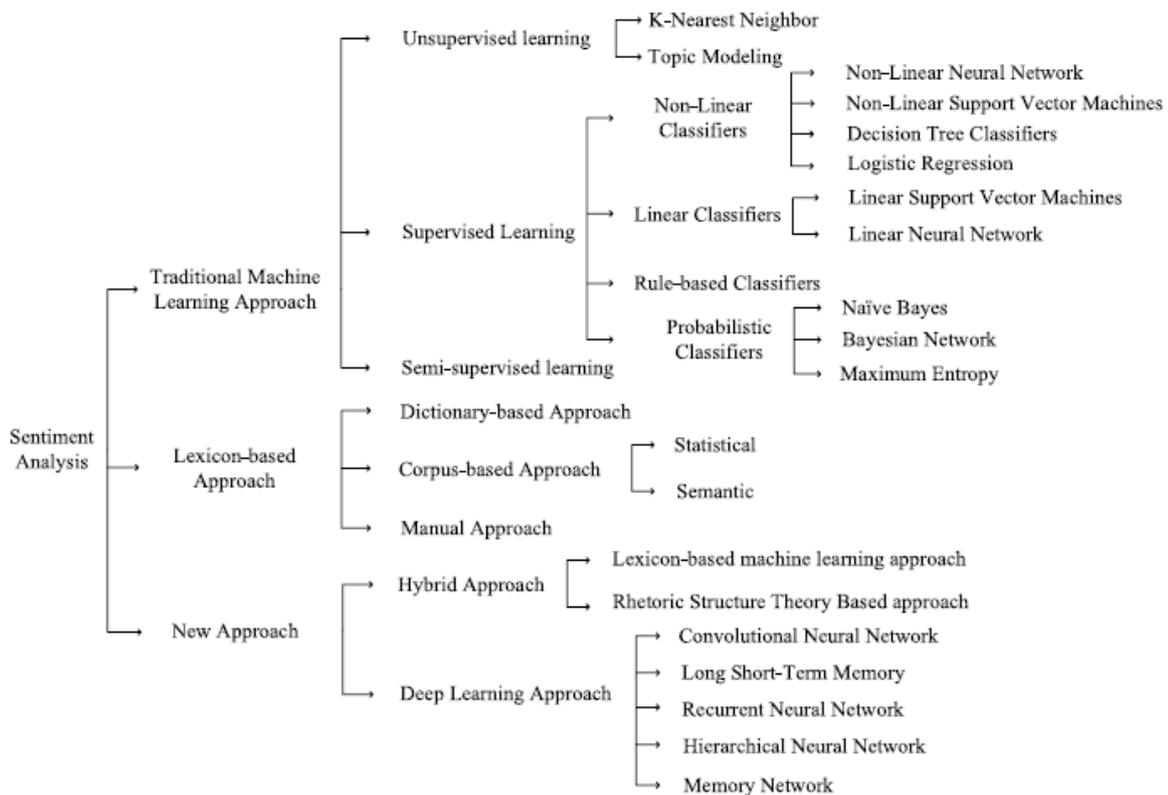


Figure no.6: Map of Sentiment Classification Techniques [6]

The above figure narrates the different approaches of sentiment analysis classification. These SA classifications are categorized on basis of various researchers. Mostly, SA are classified into traditional machine learning, lexicon based method & new approaches developed by researchers. For machine learning approaches, effective linguistic features is necessary in other hand high-quality sentiment lexicons is required for Lexicon-based approaches. By referring fig no 6 corpus-based, manual approaches & dictionary-based approaches these are types of lexicon-based approaches. Out of these methods corpus-based method is the favorite one of researchers. The lexicon-based approach obtains the sentiment of the analyzed text by examining the frequencies, & the polarities of the negative & positive words used in the text this approach is also unable to detect non-standard abbreviations, which are commonly used in posts published on the social media platforms, such as Twitter [17]. The machine learning-based approach of sentiment analysis caters to a standard classification problem [18]. Therefore, it is more suitable for extracting sentiments from unstructured

contents, & less formal texts, such as tweets from Twitter. It also eliminates pre-defined lexicons, thereby providing room for greater flexibility to be applied in any domain [18]. However, as this approach entails the construction of a text classification model to train a sentiment classifier, large labeled training dataset are necessary for the sentiment analysis to be effective [19] [20]. Unfortunately, there are always instances of insufficiently labeled data, & manual annotating is laborious, costly, & time consuming [20] [21]. Moreover, a sentiment classifier trained for a specific domain may not produce accurate result when it is directly applied to different domains [22]. The third approach is the hybrid approach which involves using the lexicon database & machine learning together. [8] Introduced a hybrid sentiment analysis which employed a machine learning theory, & a method based on polarity lexicon for analyzing Chinese sentiment phrases.

A. Sentiment Classifiers

Following are the some types of Sentiment Classifiers:

- **SentiWordNet-Based Classifier (SWN).** The SentiWordNet-based classification approach is constructed following the approach proposed by [26], which used the updated SentiWordNet3.0 dictionary [27]. In addition, authors use the ‘geometric’ weighting strategy that considers the word frequency to compute the prior polarity of each sentiment lexicon. The sentiment score is obtained by averaging the sentiment score of words in each user’s comments.
- **SVM-Based Classifier (SVM).** Following the experimental setup of Pang et al. [24], authors implement an SVM-based classifier, using the labeled word frequency vector for each review, and trained using a linear kernel.
- **CNN-Based Classifier (CNN).** A CNN-based classifier [23] for sentiment classification is used. In addition, authors also follow the ‘CNN-Static’ model setup in [23], which reported a good performance without the need for tuning the word embedding vectors.
- **LSTM-Based Classifier (LSTM).** We deploy an LSTM-based classifier [25], which obtained the top performance in the sentiment classification competition in SemEval 2017. Authors follow the experimental model construction process & configuration described by Baziotis et al. [25].

III. CONCLUSION

By taking reference of articles published in past few years, some conclusions can be made using study of various approaches of sentiment analysis with the help of literature review. The SA which depends on single data source gives sometimes erroneous forecast results or partial conclusion. Hence study of SA has increasing from single data source to multiple data source.

REFERENCES

- [1] Bhati, Reena. (2019). A Survey on Sentiment Analysis Algorithms and Datasets. Review of Computer Engineering Research. 6. 84-91. 10.18488/journal.76.2019.62.84.91.
- [2] Priyanka Tyagi, Sudeshna Chakraborty, R.C Tripathi, Tanupriya Choudhury, “Literature Review of Sentiment Analysis Techniques for Microblogging Site”, Information Systems

- &eBusiness Network (ISN) ,ICAESMT-19, DOI:<http://ssrn.com/link/ICAESMT-2019.html=xyz>
- [3] Abdallah Yousif,Zhendong Niu1,John K. Tarus, Arshad Ahmad,“A survey on sentiment analysis of scientific citations”, © Springer Science+Business Media B.V., part of Springer Nature 2017,<https://doi.org/10.1007/s10462-017-9597-8>.or
- [4] Aniza Abdullah, Ali Feizollah, Ainin Sulaiman, Nor Badrul Anuar, “Challenges and Recommended Solutions in Multi-Source and Multi-Domain Sentiment Analysis”, Citation information: DOI 10.1109/ACCESS.2019.2945340, IEEE Access
- [5] Claudia Diamantini, Alex Mircoli, Domenico Potena, Emanuele Storti, “Social information discovery enhanced by sentiment analysis Techniques” Dipartimento di Ingegneria dell’Informazione, Università Politecnica delle Marche, via Brecce Bianche, 60131 Ancona, Italy, 25 January 2018
- [6] Yong Shi,Luyao Zhu, Wei Li, Kun Guo and Yuanchun Zheng,“Survey on Classic and Latest Textual Sentiment Analysis Articles and Techniques”, International Journal of Information Technology & Decision Making, Vol. 18, No. 4 (2019) 1243–1287 World Scientific Publishing Company, DOI: 10.1142/S021962201930001
- [7] HuangEH,SocherR,ManningCD,NgAY(2012),”Improvingwordrepresentationsviaglobalcontextandmultiplewordprototypes”.In:Paperpresentedatthe50thannualmeetingoftheassociationforcomputational linguistics, ACL 2012, Jeju Island, Korea, Republic of, 8–14 July 2012, pp 873–882
- [8] Heredia B, Khoshgoftaar TM, Prusa J and Crawford M (2016), "Integrating multiple data source to enhance sentiment prediction", Proceedings of the IEEE 2nd International Conference on Collaboration and Internet Computing (CIC), Pennsylvania, USA, pp. 285-291.
- [9] Ducange P, Fazzolari M, Petrocchi M and Vecchio M (2019), "An effective Decision Support System for social media listening based on cross-source sentiment analysis models", Engineering Applications of Artificial Intelligence, Vol. 78, pp. 71-85[2-4].
- [10] Eeshita Biswas, K. Vijay-Shanker, Lori Pollock, “Exploring Word Embedding Techniques to Improve Sentiment Analysis of Software Engineering Texts”,
- [11] Tripathy A, Agrawal A, Rath SK (2016), “Classification of sentiment reviews using n-gram machine learning approach”, Expert System Application” 57:117–126. <https://doi.org/10.1016/j.eswa.2016.03.028>
- [12] Kataria S, Mitra P, Bhatia S (2010) ,“Utilizing context in generative bayesian models for linked corpus.” In: Paper presented at the 24th AAAI conference on artificial intelligence and the 22nd innovative applications of artificial intelligence conference, AAAI-10/IAAI-10, Atlanta, GA, United states, 11–15 July 2010, pp 1340–1345
- [13] Kim IC, Le DX, ThomaGR(2014) Automated method for extracting citation sentences from online biomedical articles using SVM-based text summarization technique. In: Paper presented at the IEEE international conference on systems, man, and cybernetics (SMC)2014 San Diego, CA, USA, 5–8 Oct 2014, pp 1991–1996
- [14] Li S-T, Tsai F-C (2013) A fuzzy conceptualization model for text mining with application in opinion polarity classification. Knowledge Based System 39:23–33. <https://doi.org/10.1016/j.knosys.2012.10.005>

- [15] Pang B, Lee L (2008) Opinion mining and sentiment analysis. *Found Trends Inf Retr* 2:1–135. [https://doi.org/ 10.1561/1500000011](https://doi.org/10.1561/1500000011)
- [16] Cambria E (2016) Affective computing and sentiment analysis. *IEEE Intell Syst* 31:102–107. [https://doi.org/ 10.1109/MIS.2016.31](https://doi.org/10.1109/MIS.2016.31)
- [17] Cambria, B. B. Schuller, Y. Xia and C. Havasi, New avenues in opinion mining and sentiment analysis, *IEEE Intelligent Systems* 28 (2013) 15–21, doi: 10.1109/MIS.2013.30.
- [18] Katz G, Ofek N and Shapira B (2015), "ConSent: Context-based sentiment analysis", *Knowledge-Based Systems*, Vol. 84, pp. 162-178.
- [19] Wu F, Wu S, Huang Y, Huang S and Qin Y (2016), "Sentiment domain adaptation with multi-level contextual sentiment knowledge", *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, Indianapolis, USA, pp. 949-958.
- [20] Yuan Z, Wu S, Wu F, Liu J and Huang Y (2018), "Domain attention model for multi-domain sentiment classification", *Knowledge-Based Systems*, Vol. 155, pp. 1-10.
- [21] Zhang Y, Hu X, Li P, Li L and Wu X (2015), "Cross-domain sentiment classification-feature divergence, polarity divergence or both?", *Pattern recognition letters*, Vol. 65, pp. 44-50.
- [22] Wu F and Huang Y (2016), "Sentiment domain adaptation with multiple source", *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, Berlin, Germany, pp. 301-310.
- [23] Kim, Y.: Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882* (2014)
- [24] Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up?: sentiment classification using machine learning techniques. In: *Proceedings of ACL* (2002)
- [25] Baziotis, C., Pelekis, N., Doulkeridis, C.: Datastories at SemEval-2017 task 4: deep LSTM with attention for message-level and topic-based sentiment analysis. In: *Proceedings of SemEval* (2017)
- [26] Guerini, M., Gatti, L., Turchi, M.: Sentiment analysis: how to derive prior polarities from sentiwordnet. In: *Proceedings of EMNLP* (2013)
- [27] Baccianella, S., Esuli, A., Sebastiani, F.: SentiWordNet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In: *Proceedings of LREC* (2010)