Current Apprises of Opinion Mining Methods

B. Manjula, Ameen Abdullah Aqlan, R. Lakshman Naik

Abstract — increasingly, the data is increasing day by day and storage capacity is expanding more and more, this allowing the field of SA to growing and expanding faster in research and prospecting for different opinions and emotions to be combed and technically treated to be more accurate. In our present, data can be a wealth where major global companies and development, research and crime detection centers benefit from it. In this paper we focused on the current apprises of research in this field which contributed to various improvements in the field of sentiment analysis. We have tackles comprehensive overviews for different fields which related to the Sentiment Analysis (Transfer Learning (TL), Building Resource (BR), Emotion Detection (ED)) which have the popularity of researchers has gained in recent times and attracted them. We have the aim of this survey which is to give a clear and accurate picture about the techniques of analyzing emotions and related fields.

Keywords- Sentiment Analysis; NL Process; Emotion Detection; Data Mining; Building Resources

I. INTRODUCTION

The area of Sentiment Analysis has received a lot of attention recently due to the huge amount of data circulating in different areas such as social networks, aviation, climate, health and other insurance data and social media, etc. There's no difference between SA and opinion mining, that's why it's called sentiment analysis in opinion mining. Today we live in the era of "big data" in many areas, such as meteorological data, genetics, complex physical simulation, biological and environmental research, machine learning, sensing, imaging, social networking science, computational Biology, and others, he said. These large and complex data are constantly increasing and bring with them a wide range of new computing challenges and opportunities that can provide new solutions to address them by introducing innovative improvements using modern algorithms and including sophisticated computational tools such as distributed computing, Parallel, randomization, concurrency, and analysis. Big data is of high importance, it offers a high competitive advantage to companies if they can benefit from it and address It. it helps them understand and know what customers are looking for, in the meantime the company is working to transfer these responses and study them and take appropriate measures to succeed in their work and gain the friendliness of the customers, profit and minimize losses. Studies and researches under the umbrella of sentiment analysis are of great interest to experts, officials and decision makers, according to our study, this area has seen a remarkable development since 2010.

As the attention and focus of this study has increased, this attention has made the field of emotion analysis one of the fastest growing areas[64]. Because this paper focuses on approaches classification of sentiment analysis, we have conducted research on the public's interest in this field at Google. Through this research, we have shown how important this subject is to the public. Fig 1 show the tremendous increase in search by Google's engine.

![Figure 1. Popularity of Sentiment Analysis and Customer Feedback in Google Search Engine](image)

The world is moving very fast, there is a revolution in big and open data. IDC converse in an important report sponsored by EMC on the great interest in the field of big data, in 2012 the British government identify the big data as one of eight future technologies. China also has a special strategy for big data and provided all possibilities to lead the world in this field. European Union He stressed companies that rely on big data in their decisions will reduce their costs 3.6% annually almost. Intel corporation reported that the digital data until 2002 is approximately 5 Exabyte, from 2002 to 2012, this data has increased 500 times and reached almost 2.7 Zettabyte, From 2012, we produce the same amount every ten minutes, according to Eric Schmidt the former CEO of Google. According to the latest statistic in 2018, we have approx. 4.4 Zettabyte, Estimates say 2020 will be 44 Zettabyte. The data and information available on the Internet is enormous, but it is difficult to make full use of it because of the lack of analysis. Governments, companies and others are benefiting from the data available on the Internet almost 1%. Modern science is in dire need of many experts in field of data analysis, for example, Students data be able to develop education, disease cases data enabling to development of health, transport data to be able to develop the means of transportation. Using Big data analysis techniques and tools, Wal-Mart has managed to improve its online product search results by 10-15%, while in a report to McKinsey—a leading business consulting firm-the U.S.
health sector if it uses big data analysis techniques. Effectively and efficiently, more than 300 million US dollars had been produced as an annual health budget surplus of two-thirds due to reduced spending costs. Because of the sheer volume of data that has been developed, the term "big data" has begun to spread, increasing in size so that it is difficult to address now using a single program or device, or using traditional data processing applications. Here, technology companies have started to develop new assistance programs and hardware that can help to analyze that huge data. Here are some examples of companies that analyze the reviews of customer: Amazon analyzes and processes millions of operations per day to meet the wishes of its customers, in addition to answering queries from more than half a million sellers per day, so Amazon owns the world’s top 3 databases. On the other hand, Wal-Mart handles more than one million business operations per hour, which are stored in databases containing more than 2.5 beta bytes (2500 terabytes) of data. Global demand increases for information management and processing, and free enterprises, through which they can make decisions according to the results of the analysis. As for the new oil in this era. In this paper, the data will be categorized in a comprehensive manner as follows figure1 [1-2]. The propose of analysis reviews and data is to classify them into two phases, either positive or negative. [3], The data sources used in data analysis can be considered a very important problem in such a field, since most of these extracted data result in important ideas that benefit the owners of industrial and business and free enterprises, through which they can make decisions according to the results of the analysis [4]. There are many important enhancements and applications that have occurred on science algorithms in recent years, our study in this paper aims to take a careful look at the overall applications and improvements and make a clear and explicit classification of high value studies and articles presented in this field. Our research will focus on other new fields related to Sentiment Analysis (SA) which has attracted the attention of many researchers lately. These fields are Transfer Learning (TL), Building Resource (BR), Emotion Detection (ED), and Sentiment Classification (SC).

The Transfer of Learning aims to analyze the data across only one domain and then use the results across another target domain, Emotion Detection focuses on the extraction, detection and analysis of emotions, and so that we are not shrouded in any ambiguity, emotions may exist express or implied in sentences, The purpose of Building Resources is to create (Lexicon, Corpora) and in this case all comments are explained according the polarity, Sentiment Classification aims to classify the expressive comments contained in the text according to the polarity of the opinions contained [5].

II. METHODOLOGY

In the 52 article has presented in this studies paper are summarized in the following table1. This table1 contains fifty-two high-value research papers in the field of sentiment analysis. In order to be able to learn about recent research in the fields related to SA, we have focused on research published in the last few years, specifically from 2014 to 2018. We’ll put the objectives of this article in the third column of table 1, it is divided into five categories those are (SA, SC, ED, TL and BR). The BR is categorized to lexica it’s called currently in dictionaries as well as corpora. These dictionaries contain many words which are categorized into positive, negative and neutral. By the way, most authors have categorized articles that help solve the problems of rating emotions in SC, and other articles that help solve problems of sentiment analysis in general have been categorized as SA. Finally, Other categories of articles that belong to Sentiment analysis related fields are categorized as shown in table 1.

| Emotion Detection (ED), Transfer Learning (TL), Sentiment Classification (SC) and Building Resource (BR). | Table 1 contains eight columns in the 1st column specifying the References number, while the sec column specifies the year of publish the article, in the third column as we mentioned earlier the objectives specifies, the fourth column shows the algorithm used, and also determines its characteristics and method of execution because the articles mentioned has used different algorithms. In the fifth column specifies either the article used by the emotion analysis technique for general analysis of the text (G) or use the binary classification to solve problems (Pos|Neg). The column of sixth explain the scope of data targeted and used to evaluate the algorithm used in the article, and this data can be obtained from networking sites, small blogs and news articles. In the seventh column we mention the sources of information or known data used in the article if the reader has access to it, because some authors in their articles do not address the sources of data, in this situation we write (N/A) Not Available, so when the data sources are known they may give a clear impression to the reader if he is interested in it. |
We conducted a study on various languages to achieve better results, we focused on the famous objectives that are relevant in the analysis of emotions, and then we explained their contribution and how to use them to solve the problem in SA.

The primary purpose of this survey is to create a clear and uniquely important classification of all articles that have a strong relationship in the field of sentiment analysis.

### III. RELATED WORKS

Current data analysis studies are among the most advanced recent studies and it get interested by decision makers, but it is natural to start with the most important research papers published in 2004 for the scholar Hu M, and Liu B. His paper considers one of the most important in this field and has a great respect [58]. His paper has got the highest number of citations in this area; more than 5867 are cited in this paper. This means that most of the authors have cited. Outperforms all other papers and got the first rank in the list with the most research paper cited to date [59]. Second rank was the share of Pang B, and Lee. In a research paper published in 2008, where it outperformed other fields in numbers of research and development, followed in the second place both the sentiment classification and emotion detection. We also note that other emerging areas have received fewer researches as TL and BR, in the last few months there has been a growing interest with them, and in the future, they will be almost more widespread and growing.

In Figure 3, the percentage and number of articles that use Lexicon-based Approach and Machine Learning techniques, where show us that a large percentage of researchers prefer to use the Lexicon-based Approach instead of the ML because it supports them to solve the difficulties and tasks of Sentiment Analysis despite their complications. On the other hand, the ML still an open field for researchers and has received increasing attention in the last two years.

### IV. RESULT

In this section we will be clearer about the results of the research analysis mentioned in table 1. In this section we have focused on several aspects, in the following we will explain in detail the research that contributed to the analysis of emotions through years.

In Figure 2 it was remarkable that the Sentiment Analysis remained of interest to many researchers often, where it outperformed other fields in numbers of research and development, followed in the second place both the sentiment classification and emotion detection. We also note that other emerging areas have received fewer researches as TL and BR, in the last few months there has been a growing interest with them, and in the future, they will be almost more widespread and growing.

In Figure 3, the percentage and number of articles that use Lexicon-based Approach and Machine Learning techniques, where show us that a large percentage of researchers prefer to use the Lexicon-based Approach instead of the ML because it supports them to solve the difficulties and tasks of Sentiment Analysis despite their complications. On the other hand, the ML still an open field for researchers and has received increasing attention in the last two years.
A lot of data used in articles has taken almost from products reviews; Figure 4 shows the total number of each type. Many of data has used from several sites, but what has been used frequently especially in recent years the reviews in social network (commercial sites), it’s due to the abundance credible data that reflects the opinions of a wide cross-section of intellectuals. Other types of data came in this order social media, news, web blogs.

SA recently has interest of many languages around the world, as shown in the Figure 5. It’s included other non-English languages such as Arabic language, Chinese language, Hindi, Thailand, Japanese, Turkish, and Spanish. But the English language still the number one in use that due to resources and references provided in this language. This is giving the priority to English because many of the research paper and electronic libraries are spread in English.

In the past five years, the overall disaggregation of data has been more general than those who have used the binary (positive or negative) classification, which means that the general classification still trusts by most researchers and decision-makers. The binary classification is also still in the process of growing, but it is characterized by the ease of word processing and the detection of problems clearly and this is due to its clear analytical nature. We can say binary classification is used in constant increase with time.

V. CONCLUSION AND FUTURE WORK

The literature paper outlined an overview of the SA algorithms and recent updates over the past five years.
This paper has made a general contribution to many of the relevant areas of SA that use the applications and Techniques of SA. Clearly, this paper has been given a clear result Where illustrated the improvements in SC and BR. It is clear that SVM techniques and NB techniques more used in ML algorithms to solve several problems, most importantly the SC, these two methods are an important model in the development area of SA and reference to many researchers and writers. There is a noticeable interest in the recent years in other languages, But English language still the first used in field of SA; they face many difficulties such as the lack of research and development centers related to non-English language. This aspect must be given great attention so that the resources used to SA are available for many other languages. We noticed that a large segment of people has tended to use small blogs as well as forums and news sources recently on a large scale. With great progress in social networking sites, played an important role in attracting people to express their feelings and opinions on different topics or a product. Relying on social networking sites for data collection or as a source requires deeper and more accurate analysis. Understanding the context of text and knowing the feeling of the user is extremely important in the shadow of the sheer amount of data across the communication sites and this is reason enough for further development and research on the analysis of emotions application other techniques such as TL.

REFERENCES

1. Chih Yu, and other, “Using a contextual entropy model to expand emotion words and their intensity for the sentiment classification of stock market news, Science Direct, elsevier (2013).
2. Michael h, Liebmann, DirkN, Automated news reading: Stock price prediction based on financial news using context- capturing features, Science Direct, (2013).
3. Theresa, J Wiebe, and other, Recognizing Contextual Polarity in Phrase-LevelSA, doi:10.3115/1220575.1220619, (2005).
4. Isa M, and other, A lexicon model for deep SA and opinion mining applications, DSS (2012).
5. Pang and L Lee, Shivakumar Vaithyanathan, “Thumbs up? Sentiment Classification using Machine Learning Techniques”, doi:10.3115/1118693.1118704, (2002).
6. Jitrlada ROJIRATANAVIJ, Preecha VICHITTHAMAROS and Sukanya PHONGSUPHAP, “Acquiring Sentiment from Twitter using SL and Lexicon-based Techniques” (2016).
7. Fan, and other, “Apply Word Vectors for Sentiment Analysis of APP Reviews” DOI: 10.1109/ICSAW2016.7811108, (2016).
8. V Ikoro and other,” AS Expressed on Twitter by UK Energy Company Consumers” DOI: 10.1109/SNAMS.2018.8554619, (2018).
9. Z Lumin, JIA and other, “User-Level Sentiment Evolution Analysis in Microblog” DOI: 10.1109/CC.2014.7019849, (2014).
10. Kim S and F Frasincar,” Survey on Aspect-Level SA IEEE, (2016).
11. Ying F, and other “Multi-Strategy SA of Consumer Reviews Based on Semantic Fuzziness” IEEE (2018).
12. Xing F and Zhan,” Sentiment analysis using product review data” DOI 10.1186/s40537-015-0015-2, (2015).
13. Naaima Band other “SA in Arabic: A review of the literature” doi: org/10.1016/j.asoc.2017.04.007, (2018).
14. Cumali T urkenkoglu and Ahmet C uneyd Tantug,” Sentiment Analysis in Turkish Media” DOI:10.13140/2.1.502.1125., (2014)
15. Mario Andrés Paredes-Valverde and Ricardo Colomo-Palacios,” Sentiment Analysis in Spanish for Improvement of Products and Services: A Deep Learning Approach” doi.org/10.1105/2017/1329281, (2017).
16. T Mohammed, “Review of Sentiment Analysis for Classification Arabic Tweets” IJETAE (2016).
17. Hongwei W and other,” Sentiment classification of Chinese online reviews: a comparison of factors influencing performances”DOI: 10.1080/17517575.2014.947635, (2016).
18. Deepali Mishra, Manju Venugopalan and DeepaGupta” ICACC 2016, Cochin, India” (2016).
19. Ms. Sneha Mulakar, Prof. Varunakshi Bhojane, “Sentiment Classification in Hindi” DOI: 10.9790/66611-174100102, (2015).
20. N Mittal, and other “Discourse Based Sentiment Analysis for Hindi Reviews” doi:10.1007/978-3-642-45062-4_102, (2013).
21. Zhang, Mamoru K, “Japanese SC with Stacked Denoising Auto-Encoder using Distributed Word Representation” (2018).
22. LIU Lizhen, and other “A Novel Feature-based Method for Sentiment Analysis of Chinese Product Reviews” DOI: 10.1109/CC.2014.6825268, (2014).
23. Kai Zhao, YaohongLin, “A Hybrid Method for Sentiment Classification in Chinese Movie Reviews Based on Sentiment Labels” DOI: 10.1109/IAF.2015.7451538, (2015).
24. Shenghua Liu and other “TASC-Topic-Adaptive Sentiment Classification on Dynamic Tweets” DOI:10.1109/TKDE.2014.2382600, (2015).
25. Lin Zhang and other, “Sentiment Analysis on Reviews of Mobile Users” Science directed, (2014).
26. Cagatay CATAL, “A Sentiment Classification Model Based On Multiple Classifiers” doi:10.1016/j.asoc.2016.11.022, (2016).
27. Doaa Hussein, A survey on sentiment analysis challenges, JKSUES (2016).
28. Vimalkumar B. “Analysis of Various Sentiment Classification Techniques”, DOI:10.5120/jica2016909259, (2016).
29. D. RAMESH and others; “A Comparative Analysis of Classification Algorithms on Weather Dataset Using Data Mining Tool”, doi.org/10.13005/ojct/10.04.13, (2017).
30. P Tarnowski and others,“Emotion recognition using facial expressions” ScienceDirect, (2017).
Current Apprises of Opinion Mining Methods

47. YILONG YANG and others, “Medshare: A Novel Hybrid Cloud for Medical Resource Sharing Among Autonomous Healthcare Providers”, DOI: 10.1109/ACCESS.2018.2865535, (2018).
48. S.M. Waseem and Afroz, “Test Scheduling with Built in Logic Block Observer for NoC Architecture”, DOI: 10.1109/ICMIA.2017.7975601, (2017).
49. Mikhail L. and others, “Translated Subtitles Language Learning Method: a New Practical Approach to Teaching English”, Elsevier Journal, (2015).
50. El Barbary, A. Salama, “Feature selection for document classification based on topology” Egyptian Informatics Journal (2017).
51. Said Bahassine and others, “Feature selection using an improved Chi-square for Arabic text classification” of King Saud University Journal (2018).
52. Marwa Trabelsia and others, “A New Feature Selection Method for Nominal Classifier based on Formal Concept PCS Journal Elsevier, (2017).
53. Shahana P.H.and others, “Evaluation of Features on Sentimental Analysis” PCS Journal Elsevier, (2015).
54. Tuba P., S. Ozeli, “A New Feature Selection Method for Sentiment Analysis of Turkish Reviews” (2016).
55. Oğuz Kaynar and others, “Feature Selection Methods in Sentiment Analysis” (2017).
56. Ankit A., and others “Sentiment Classification System of Twitter Data for US Airline Service Analysis” DOI: 10.1109/COMPSAC.2018.00114, (2018).
57. A.S. Altheneayen, M.B. Menai, “Naïve Bayes classifiers for authorship attribution of Arabic texts” KSU Journal (2014).
58. Minghu Hu and others “Mining and Summarizing Customer Reviews” book, (2004).
59. O Ahlgren, “Research On Sentiment Analysis: The First Decade” DOI 10.1109/ICDMW.2016.94, (2016).
60. B. Pang and L. Lee, “A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts”
61. Azwa Abdul Aziz, Andrew Starkey and Marcus Campbell Bannerman, "Evaluating Cross Domain Sentiment Analysis using Supervised Machine Learning Techniques" (2017).
62. Danushka Bollegala and others, “Cross-Domain Sentiment Classification using a Sentiment Sensitive Thesaurus” doi:10.1109/TKDE.2012.103, (2014).
63. H. Ehrig and J. Bross, “Automatic Construction of Domain and Aspect Specific Sentiment Lexicons for Customer review Mining”, doi:10.1145/2505515.2505574.
64. Ameen Abdullah Qaid Aqlan, B. Manjula, R. Lakshman Naik. “A Study of Sentiment Analysis: Concepts, Techniques, and Challenges”, Springer Science and Business Media LLC, 2019

AUTHORS PROFILE

Dr. B. Manjula received the BCA in Computer Application and M.Sc. in Information System from Osmania University, Hyderabad, India. She completed the Ph.D in Computer Science from Kakatiya University, Warangal, India. She is currently working as Assistant Professor in Department of Computer Science, Kakatiya University, Warangal, India. She is carrying out her research on Data Analytics, Big Data Analysis, Sentiment Analysis and Optimization Techniques. She has published over 18 research papers in international journals and presented more than 15 research papers in international conferences. She is a member of ACM, IAENG, CSTA and other Professional Organizations.

Ameen Abdullah Qaid Aqlan received the BCA in Computer Application and M.Sc. in Information System from Osmania University, Hyderabad, Telangana State, India. He is currently Research Scholar in Department of Computer science, Kakatiya University, Warangal, India. His research area is Sentiment Analysis Using Big Data Techniques. He is Members of IAENG and other professional bodies. He published 2 research papers in international journals.

R. Lakshman Naik received the B.Tech in Electronics & Communication Engineering, M.Tech in Computer Science & Engineering from Jawaharlal Nehru Technological University, Hyderabad, India and M.Tech in Digital Communication from Kakatiya University, Warangal, India. He served as a Systems Engineer in Wipro Technologies. He is a Member of ACM, AIRCC, IAENG, IAEME and other Professional Organizations. He published 20 research papers in international journals and presented more than 12 research papers international conference. He is carrying out his research in area of Computer Networks, IoT, Data Analytics and Mobile Cloud Computing.

Retrieval Number: B3906129219/2019©BEIESP DOI: 10.5940/ijeat.B3906.129219