❒   2818

Insights to Problems, Research Trend and Progress in Techniques of Sentiment Analysis

Kumar P. K.[1], Nandagopalan S.[2]
[1]Department of MCA, Post Graduate Studies, VTU, MysoreRegion, Mysuru, India
[2]Department of Computer Science & Engineering, Bangalore Institute of Technology, Bangalore, India

Article Info	ABSTRACT
	The research-based implementations towards Sentiment analyses are about a decade old and have introduced many significant algorithms, techniques, and framework towards enhancing its performance. The applicability of sentiment analysis towards business and the political survey is quite immense. However, we strongly feel that existing progress in research towards Sentiment Analysis is not at par with the demand of massively increasing dynamic data over the pervasive environment. The degree of problems associated with opinion mining over such forms of data has been less addressed, and still, it leaves the certain major scope of research. This paper will brief about existing research trends, some important research implementation in recent times, and exploring some major open issues about sentiment analysis. We believe that this manuscript will give a progress report with the snapshot of effectiveness borne by the research techniques towards sentiment analysis to further assist the upcoming researcher to identify and pave their research work in a perfect direction towards considering research gap.

Corresponding Author:

Kumar P. K. ,
Department of MCA, Post Graduate Studies,
VTU, Mysore Region, Mysuru, India
Email: kumar.pk.krushna@gmail.com

1. INTRODUCTION

The process of sentiment analysis can be said to be a form of application that combines applies the concept of text analytics, computational linguistics as well as natural language processing [1]. It is also called as opinion mining and is referred to a method of determining as well as extracting the subjective information about the source materials [2]. The problems associated with this area are also referred to as multi-disciplinary intelligence problems that target to create a communication bridge between computer and human [3]. This field of study can also be said to be using both electronic intelligence as well as human intelligence for the purpose of extracting knowledge as well as categorizing different forms of discrete sentiments [4]. As the numbers of social network applications are increasing, it gave birth to the sentiment analysis. The users are now more interested to share their opinion on the internet using ratings, reviews, and a suggestion with diversified forms of user's expression. These opinions are used by the stakeholders to understand the user's requirement as well as flaws/success factor involved in their process management towards product/service design. As nowadays, complete business process is on the verge of automation, so it is quite eminent that there are all the possibilities of noisy data or certain unscrupulous data that could overall reduce the genuine factor about the user reviews or opinion. Sentiment analysis also suffers from the bigger set of problems.

The significant problem about it is the usage of very simplified terms to represent express emotions (or sentiment) related to specific service/product. Apart from this, various forms of linguistic-related factors, cultural scales, and dynamic context make the process of sentiment analysis further challenging. It is a

computationally challenging system for an algorithm to come to a conclusion about the actual score of sentiment if the sentence is very small or very vague or written out of the grid of basic grammar. However, existing investigation has shown that smaller sentences are no more problems as there is various implementation of opinion mining towards social network data where the shared data is usually smaller in size (e.g. tweets). At present, there are already certain applications where tweets are used for political surveys. The level of accuracy of existing mechanism of sentiment analysis heavily depends upon the synchronicity with the judgment factor of the human. Technically, the performance parameter used for this purpose is precision and recall rate [5].

Therefore, this paper presents a discussion of sentiment analysis, its problems, its trends, and measuring the effectiveness of some significant recent research works. Section 1.1 briefly discusses the background of the study followed by problem identification in Section 2. Section 3 briefs up the proposed solution while existing research trend is discussed in Section 4. Discusses brief highlights of open research issues in Section 5. Finally, concluding remarks about the findings is made in Section 6.

1.1. Background

There has been a certain amount of good ground work being carried out towards sentiment analysis in recent times. We find that significant approaches of enhancing the performance of sentiment analysis are through i) N-Gram model, ii) subjective lexicon, and iii) machine learning [6]. Most recently, Schouten and Frasincar [7] have discussed their review findings about sentiment analysis that discusses various forms of aspect-level based approaches. Sentiment Analysis is also investigated on Indian Language [8], Chinese language [9], Arabic Language [10] apart from the English language. At present, existing techniques towards sentiment analysis is focused on using lexicon generation in text-based processing [11], subjectivity detection [12], sentiment polarity detection [13], sentiment structurization [14], summarization-visualization tracking [15], etc. Apart from this, adopt of sentiment analysis is very much frequent for analyzing social network data from Facebook, Twitter, and Google [16].

2. PROBLEM IDENTIFICATION

This section briefs up the significant research problems that have been observed after reviewing the existing techniques:

- In presence of massive number of interactive data shared by social network, it is very hard to differentiate legitimate and illegitimate data e.g. counterfeited opinion, spam broadcast, viral messages, etc. We do not have any standard threshold in sentiment analysis that could detect such outliers pertaining to the identify the reputation of users.
- Adoption of the filtering technique that works on collaborative approach just filters the frequently shared messages and ignores the discrete message thereby showing that existing opinion mining lacks richness in intelligence.
- At present, the data is quite dynamic and massively growing and is taking the shape of Big Data which has certain specific characteristics. We don't find much implementation work on opinion mining or sentiment analysis pertaining to big data approach. It thereby shows sentiment analysis has to be more investigated in order to cope up with existing demands and futuristic technologies.
- Sentiment or emotional analysis will need more research and improvement regarding behavioral analysis from text-related data to cope upmost with dynamic Internet-of-Things (IoT) based applications in future.

Hence, there are various fields of problems where sentiment analysis will need to work for e.g. i) improved exploring process of complex contents using correlation, ii) performing assessment with real-time data, iii) evaluating and improving bipolar approach in opinion mining, iv) investigating on constructing uniform platform to assess on different language, v) labeling of the resources and sharing using collaborative means, vi) ensuring accuracy and precision of the system generated recommendation system. Therefore, there are multiple directions of research work to be directed to addressing the diversified problems associated with sentiment analysis. The next section briefs up the solution towards addressing such research problems in opinion mining as well as sentiment analysis.

3. SOLUTION

The solution to the research problems as discussed in prior section is to perform further investigation towards the existing techniques. Studies towards sentiment analysis are a decade old, and hence it is essential to explore fair possibilities of recently implemented techniques. While looking for the

effectiveness of recent implementation towards sentiment analysis, following aspects will be a factor of success i.e. i) accuracy level and its enhancement techniques, ii) minimization of human-effort, iii) precise classification ability for fake and legitimate review, iv) computational efficiency with respect to time and space complexity, etc. Hence, the prime objective of the proposed system is to explore the level of effectiveness of recently implemented techniques and to explore research gap. To do this, we initiate our study by observing the existing research trend as discussed in next section.

4.   EXISTING RESEARCH TREND

In this section, we will look into the existing research trends towards sentiment analysis. We look into the research papers published during 2010 to till date from IEEE Xplore to understand the frequencies of problems being addressed in the area of sentiment analysis. We find that from 2010, there are total of 1832 conference papers, 128 Journals, 8 early access articles, and 5 books that have been published in sentiment analysis.

(a)Twitter data   (b)Semantics

(c)Ontology   (d)Semantic & ontology
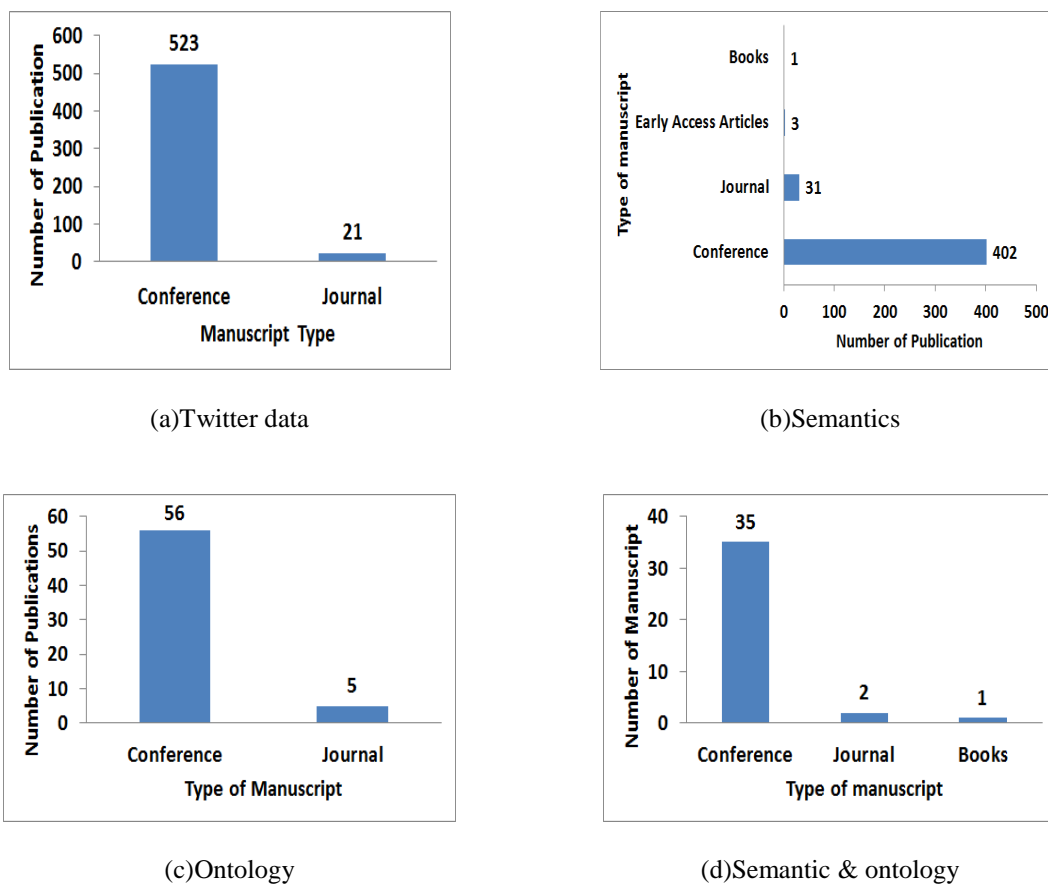Figure 1 Research trends towards problems in sentiment analysis

Figure 1 (a) shows that there are 21 journals dedicated to investigating sentiment analysis about Twitter data. Usage of semantics (Figure 1 (b)) is also found to be higher. However, there is less focus on ontology-based approach as shown in Figure 1(c), while approaches that combine uses semantics and ontology is extremely less to find in the existing system. Overall observation is that investigation towards sentiment analysis is witnessed with less number of research papers as compared to research problems in other streams of data analysis or mining approaches. Figure 1 shows the frequently addressed problems in the area of sentiment analysis. The next section discusses the recent techniques of sentiment analysis.

Apart from the above-mentioned techniques and line of research in sentiment analysis, it was found that sentiment analysis was also implemented for analyzing social network applications [17-18], financial predictions [19], recommendation system [20], knowledge discovery of social emotion [21]. Hence, it can be seen that there are various techniques and approaches towards addressing diversified problems in sentiment

analysis. All the methodologies presented by the authors have their advantageous points as well as limiting points. The majority of the limiting points are related to lack of benchmarked research implementation. The next section discusses the open research issues in the area of sentiment analysis.

5.   OPEN RESEARCH ISSUES
The open research issues in Sentiment Analysis are:-

5.1. Less Novelty in the Approaches
Majority of the existing approaches towards sentiment analysis are more or less related to analyzing public-related sentiments from a text, social network applications, etc. However, there is comparatively less number of work where semantics and ontology are found to have significant contribution in sentiment analysis.

5.2. Less Focus on Complex Cases
At present, existing research techniques are more or less tested over datasets of public reviews. However, such reviews are not assessed in depth with respect to sentiments. Merely classifying the sentiment in positive and negative doesn't exhibit the actual and hidden meaning behind the original context of the sentiments. Moreover, there are certain sentences which are very confusing in nature and cannot be classified as precisely to positive or negative.

5.3 Consistent Carry over Limitation
There are various forms of limitations associated with e.g. evaluating sentiment with much clarity in the review process, investigating sentiments from documents with multiple themes, etc. Existing techniques focuses on macro-issues and not on micro issues for which reason accuracy is not found to be optimized.

5.4 Less Benchmarked Studies
It has been found that there is less number of research approaches that are found to have been benchmarked. Apart from benchmarking, there is less number of the studies whose outcome is found to be less compared with appropriate technique.

5.5 Less focus on Computational Complexity
Analysing a simple text of smaller dimension may consume fewer resources. However, accumulation of sentiments from the collaborative platform will require good number of computational resources. Unfortunately, in existing studies we don't find many evidences of computationally cost effective algorithms to perform sentiment analysis for larger data.

6.   CONCLUSION
The study outcome directly shows that existing research techniques have less emphasized on the new forms of the dynamic data that is characterized by higher complexity level in the form of structure, heterogeneity, uncertainty, etc., There are various significant research problems e.g. computational efficiency, unstructured data, adoption of complex cases of public reviews, etc, which is received a very less attention from the research community. Studies towards for larger and unstructured data will require more attention in the area of sentiment analysis to ensure the adoption of sentiment analysis for upcoming communication-related technologies. Therefore, our future work will be in the direction of addressing such open research issues. We will initiate our investigation using unstructured data and introduce a technique to countermeasure the complexity involved in opinion mining from such complex data.

REFERENCES
[1] J.Kantola, W. Karwowski, "Knowledge Service Engineering Handbook," CRC Press, Business & Economics, pp. 599, 2016
[2] P.K. Singh and M.S.Husain, "Methodological Study of Opinion Mining and Sentiment Analysis Techniques," International Journal on Soft Computing, Vol.5, No. 1, pp.11, 2014
[3] A. Kaur and V. Gupta, "A Survey on Sentiment Analysis and Opinion Mining Techniques," Journal of Emerging Technologies in Web Intelligence, Vol.5, No. 4, pp. 367-371, 2013
[4] B. Liu, "Sentiment Analysis and Opinion Mining," Synthesis lectures on human language technologies, Vol.5, No. 1, pp. 1-167, 2012

[5]   B.L. Pang and S. Vaithyanathan, "Thumbs up?: Sentiment Classification Using Machine Learning Techniques," In Proceedings of the ACL-02 conference on Empirical methods in natural language processing, Vol.10, pp. 79-86, 2002

[6]   W. Lincy and N. Kumar, "A Survey on Challenges in Sentiment Analysis," Retrieved, 16-Feb-2017

[7]   T Sutikno, D Stiawan, IMI Subroto, "Fortifying big data infrastructures to face security and privacy issues," *TELKOMNIKA Telecommunication Computing Electronics and Control*., vol. 12, no. 4, pp. 751-752, 2014.

[8]   A. Gelbukh, "Computational Linguistics and Intelligent Text Processing: 14th International Conference, CICLing 2013, Karlovasi, Samos, Greece, March 24-30, 2013, Proceedings, Part 2," Springer, pp. 576, 2013

[9]   S.M. Vohra and J. B. Teraiya, "A Comparative Study of Sentiment Analysis Techniques," Journal JIKRCE, Vol.2, No. 2, pp. 313-317, 2013

[10] M. Biltawi, W. Etaiwi, S. Tedmori, A. Hudaib, and A. Awajan, "Sentiment Classification Techniques for Arabic Language: A survey," In Information and Communication Systems (ICICS), 7th International Conference, pp. 339-346, 2016.

[11] O. Kolchyna, T.TP. Souza, P.Treleaven, and T. Aste, "Twitter Sentiment Analysis: Lexicon Method, Machine Learning Method and Their Combination," arXiv preprint arXiv: 1507.00955, 2015

[12] M.Graña, C. Toro, "Advances in Knowledge-based and Intelligent Information and Engineering Systems," Volume 1", IOS Press, pp. 2273, 2012

[13] "Sentiment analysis", https://en.wikipedia.org/wiki/Sentiment_analysis, Retrieved, 16-Feb-2017

[14] W. Medhat, A. Hassan, and H. Korashy, "Sentiment Analysis Algorithms and Applications: A Survey," Ain Shams Engineering Journal, Vol.5, No. 4, pp.1093-1113, 2014

[15] A. Das, S. Banyopadhyay and B. Gambäck, "The 5W Structure for Sentiment Summarization Visualization-Tracking," ERCIM, Retrived, 16-Feb-2017

[16] A. Selamat, H. Fujita, H. Haron, "New Trends in Software Methodologies, Tools and Techniques: Proceedings of the Thirteenth SoMeT_14," IOS Press, pp. 1128, 2014

[17] H. Saif, M. Fernandez, Y. He, and H. Alani, "Semicircles for Contextual and Conceptual Semantic Sentiment Analysis of Twitter," *In European Semantic Web Conference*, Springer International Publishing, pp. 83-98, 2014.

[18] D. D. Wu, L. Zheng and D. L. Olson, "A Decision Support Approach for Online Stock Forum Sentiment Analysis," in *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 44, no. 8, pp. 1077-1087, Aug. 2014.

[19] Y. Zhang, G. Lai, M. Zhang, Y. Zhang, Y. Liu, and S. Ma, "Explicit Factor Models for Explainable Recommendation Based on Phrase-level Sentiment Analysis," In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, pp. 83-92, 2014.

[20] S. Bao *et al*., "Mining Social Emotions from Affective Text," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, no. 9, pp. 1658-1670, Sept. 2012.

[21] H. Saif, Y. He, and H. Alani, "Semantic Sentiment Analysis of Twitter," *In International Semantic Web Conference*, Springer Berlin Heidelberg, pp. 508-524, 2, 2012

BIOGRAPHIES OF AUTHORS

Kumar P K., Department of MCA, Post Graduate Studies, VTU, Mysore Region, Mysuru, India. Currently pursuing PhD under Visvesvaraya Technological University, Belagavi. My research interest is in image analysis, data mining. I have around 4 years of Experience in teaching and 6 years of experience in industrial field. Also, I have been published 02 International Journal Articles and 2 National Conference paper.

Dr. Nandagopalan S., Department of Computer Science & Engineering, Bangalore Institute of Technology, Bangalore, India. My research interests are in Image analysis, data mining, Medical images. I have around 30 years of Experience in teaching. Also, I have been published 06 International Journal Articles (Peer Reviewed) and 09 Refereed International Conference Proceedings paper.