MONITORING SOCIAL MEDIA AND RESEARCH PAPER USING NATURAL LANGUAGE PROCESSING

Bishwa Ranjan Das Department of Computer Science and Engineering, Aryan Institute of Engineering and Technology, Bhubaneswar
Amiya Kumar Sahoo Department of Computer Science and Engineering, Aryan Institute of Engineering and Technology, Bhubaneswar
Rekhanjali Sahoo GITA Autonomous College, Bhubaneswar
biswadas.bulu@gmail.com, amiya79@rediffmail.com, rekhanjalisahoo23@gmail.com,

Abstract
Natural language processing (NLP) is one of the most promising avenues for social media data processing. It is a scientific challenge to develop powerful methods and algorithms which extract relevant information from a large volume of data coming from multiple sources and languages in various formats or in free form. NLP can be used to solve a wide range of problems, from helping people search for information to summarizing text to detecting sentiment. Natural Language Processing has recently been used in a huge range of different industries for both commercial and non-commercial reasons. Social media monitoring is the process of identifying and determining what is being said about a brand, individual or product through different social and online channels.

In recent years, online social networking has revolutionized interpersonal communication. The newer research on language analysis in social media has been increasingly focusing on the latter's impact on our daily lives, both on a personal and a professional level. Natural language processing (NLP) is one of the most promising avenues for social media data processing. It is a scientific challenge to develop powerful methods and algorithms which extract relevant information from a large volume of data coming from multiple sources in various formats. Most of the challenges of social media texts are analyzing in contrast with traditional documents. Research methods in information extraction, automatic categorization and clustering, automatic text summarization and indexing, and machine translation need to be adapted to a new kind of data. The current research on NLP tools and methods for processing the non-traditional information from social media data that is available in large quantities (big data), and it demonstrates how creative NLP approaches can integrate appropriate linguistic information in a variety of fields, including social media monitoring, healthcare, business intelligence, industry, marketing, security, and defence. We examine the assessment measures currently in use for NLP and social media applications, as well as the most recent evaluation campaigns and collaborative assignments on fresh social media datasets. These tasks are coordinated by the Association for Computational Linguistics or the National Institute of Standards and Technology through the Text Analysis Conference and the Text REtrieval Conference. Lastly we go over the significance of this dynamic field and how NLP can greatly benefit from it in the upcoming ten years as mobile technology, cloud computing, virtual reality, and social networking develop.

Keywords: Odia, Language, Sentiment Analysis, Tagging, Morphology

Introduction
Social media has been widely used and become an important communication tool since the age of Internet. It is an effective way to spread out information and express opinions. Since many people use social media every day, a large number of reviews, feedbacks, article have been created. Many organizations use social media to reach out their customers. It is important for organizations to automatically identify each customer review whether it is positive or negative; this is called "Sentiment Analysis".

Sentiment analysis, also known as opinion mining, will continue to play a significant role, allowing businesses to monitor social media and gain real-time insights into how customers feel about their brand or products. Using natural language processing (NLP) tools to assess brand sentiment can assist businesses in identifying areas for improvement, detecting negative comments on the fly (and responding proactively), and gaining a competitive advantage. Analysis of the impact of marketing efforts and evaluating how consumers react to events such as a new product introduction are two more intriguing use cases for sentiment analysis in social media monitoring.

Due to the fast growth of World Wide Web the online communication has increased. In recent times the communication focus has shifted to social networking. In order to enhance the text methods of communication such as tweets, blogs and chats, it is necessary to examine the emotion of user by studying the input text. Online reviews are posted by customers for the products and services on offer at a website portal. This has provided impetus to substantial growth of online purchasing making opinion analysis a vital factor for business development. To analyze such text and reviews sentiment analysis is used. Sentiment analysis is a sub domain of Natural Language Processing which acquires writer's feelings about several products which are placed on the internet through various comments or posts. It is used to find the opinion or response of the user. Opinion may be positive, negative or neutral. In this paper a review on sentiment analysis is done and the challenges and issues involved in the process are discussed. The approaches to sentiment analysis using dictionaries such as SenticNet, SentiFul, SentiWordNet, and WordNet are studied. Dictionary based approaches are efficient over a domain of study. Although a generalized dictionary like WordNet may be used, the accuracy of the classifier get affected due to issues like negation, synonyms, sarcasm, etc.

Twitter have been created since 2006 and gained popularity nowadays. A sentiment analysis on Twitter was introduced in 2009. The limitation of 140 characters per tweet makes Twitter easier to classify the sentiment. However, commonly used classical techniques, e.g., Multinomial Naïve Bayes (MNB), Support Vector Machine (SVM) and Maximum Entropy (MaxEnt), are based on bag-of-words model which the sequence of words is ignored. This results in inefficient sentiment analysis because the sequence of words can affect the emotion. For example, "bad" and "not" are both negative, but the phrase "not bad" which is composed of these two words has positive meaning.

Human decision making or thinking is always affected by others thinking, ideas and opinions. The growth of social web gives a huge amount of user generated data such as comments, opinions and reviews about products, services and events. This data will be useful for consumers as well as manufacturer. While buying any product online consumers usually check comments or opinion of others about the product. Manufacturer can understand the response of that product and get insight into its products strength and weaknesses based on the sentiment of the customers. These opinions are helpful for both business. organizations and individuals but the huge amount of such opinionated text data becomes burden to users. To analyze and summarize the opinions expressed in these enormous opinionated text data is a very interesting domain for researchers. This new research domain is typically called Sentiment Analysis or Opinion Mining. Sentiment analysis is used to automatically mine the opinions and emotions from text, speech, and database sources with the help of Natural Language Processing (NLP). Sentiment analysis does the classification of opinions in the text into categories like "positive" or "negative" or "neutral". It's often referred to as subjectivity analysis, opinion mining and appraisal extraction. For buying any product customer wants to see the opinion of other about the product. Customer reviews are very important for business process since to make future decision business organizations should know what customers are saying about their product or service that an organization is providing. It will provide important functionality for voice of customer and brand reputation management. Thus it is helpful in business process and also for the customers Major areas of research in Sentiment analysis are Subjectivity Detection, Sentiment Prediction, Aspect Based Sentiment Summarization, and Text Summarization for Opinions, Contrastive Viewpoint Summarization, Product Feature Extraction, and Detecting Opinion Spam. Subjectivity Detection is a task of finding whether text is opinionated or not. Sentiment Prediction is about predicting the polarity

of text whether it is positive or negative. Aspect Based Sentiment Summarization generates sentiment summary in the form of star ratings or scores of features of the product. Text Summarization generates a few sentences that summarize the reviews of a product. Contrastive Viewpoint Summarization puts an emphasis on contradicting opinions. Product Feature Extraction is a work that extract product feature from its review. Detecting Opinion Spam is concern with identifying fake or bogus opinion from reviews.

Methodology and Result Analysis

A systematic literature review methodology has been adopted by this article (Gupta et al., 2018; Chakraborty and Kar, 2017). We selected the Scopus database to analyze text mining literature which has been published in service management because Scopus indexes a very wide coverage of engineering and management literature in terms of journals, conferences and book chapters (Grover et al., 2018; Singh et al., 2020; Agarwal et al., 2017). Further, Scopus provides various options in which users can search for the relevant literature through advanced search. Therefore, for this study, we extracted the research papers from the Scopus database. Initially, we downloaded the research papers from the Scopus database in September 2020.

For downloading the research papers, we decided on the search terms "Text Mining", "NLP", "Semantic Analysis", "Topic Modeling", "Service", and "Management". We used the Boolean "AND" and "OR" operator in the search field to combine the search term. Therefore, the structure of the search query was "Text Mining" OR "NLP" OR "Semantic Analysis" OR "Topic Modeling" AND "Service" OR "Management". It ensures that search terms will come in extracted research papers. We applied a search on "Article Title" and "Article Keywords" for the initial downloading of published research literature. In the first phase, we fetched 6273 research papers from the query result. In the second phase, we restricted our analysis only to journals published literature, then the number of research papers reduced to 3893. In the third phase, we focused only on the business and management area, then the number of research papers reduced to 327. In the fourth phase, we restricted our search only to FT (Financial Times, British) publications; A*, A, and B ranked journals in the Australian Business Dean's Council journal ranking schema and 4*, 3 and 2 in the Chartered Association of Business School's journal ranking schema, to restrict to relatively higher quality articles, then the paper count reduced to 155. These three-journal ranking schema were used as collectively they are more inclusive and yet fairly popular globally as a standard for assessing quality of publications. In the fifth phase, we screened titles, keywords, and abstracts of the selected papers so that we can exclude inappropriate papers. We excluded those papers that were not relevant to the objective of the systematic review process of text mining applications in service and management. Finally, we got 125 re- search papers that were entirely appropriate for studying text mining applications in the service and management field.

Conclusion and Future work

From the pioneering contributions to the domain of NLP, IR and in specific to sentiment analysis it is observed that sentiment analysis can play a very important role in the development of successful business and research. Sentiment analysis can be carried out in different languages and may extend to other areas such as image processing, data aggregation, etc. There are several dictionaries available for sentiment analysis, of which Senti WordNet is used more often. This paper reviews approaches, issues and challenges involved in sentiment analysis and classification. Three types of approaches are described with their relative merits and limitations. It is found that sentiment analysis using dictionary-based approach is swift compared to machine learning-based approach since it require no prior training. A framework of sentiment classification is explained describing the main steps in sentiment analysis using dictionary-based approach.

References

[1] Varghese,R., Jayasree, M., "A Survey on Sentiment Analysis and Opinion Mining", IJRET: International Journal of Research in Engineering and Technology, Volume: 02 Issue: 11, October 2013.

[2] Patil, S. and T. Patil, "Automatic Generation of Emotions for Social Networking Websites using Text Mining", 4th International Conference on Computing, Communication And Networking Technologies (ICCCNT) IEEE, pp4-6, July 2013.

[3] Pimpalkar, A., Wandhe, T., Swati Rao M., Kene M., "Review of Online Product using Rule Based and Fuzzy Logic with Smileys", IJCAT – International Journal of Computing and Technology, Volume 1, Issue 1, February 2014.

[4] Kaushik, C.and Mishra A., "A Scalable, Lexicon Based Technique for Sentiment Analysis" International Journal in Foundations of Computer Science and Technology (IJFCST), Vol.4, No.5, September 2014.

[5] Vohra, S. M. and Teraiya, J. B., "A Comparative Study of Sentiment Analysis Techniques" Journal of Information, Knowledge and research in Computer Engineering Volume – 02, Issue – 02, October 2013.

[6] Baccianella, S., Esuli, A. and F. Sebastiani, "SENTIWORDNET 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining", The International Conference on Language Resources and Evaluation (LREC). Vol. 10, 2010.

[7] Nakagawa, T., S. Kurohashi, "Dependency Tree based Sentiment Classification using CRFs with Hidden Variables", Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics. Association for Computational Linguistics, 2010.

[8] A. Go, R. Bhayani, and L. Huang, "Twitter sentiment classification using distant supervision", CS224N Project Report, Stanford, Vol. 1, pp. 12, 2009.

[9] A. Bermingham and A. F. Smeaton, "Classifying sentiment in microblogs: is brevity an advantage" in Proceedings of the 19th ACM international conference on Information and knowledge management, 2010, pp. 1833-1836.

[10] X. Wang, Y. Liu, C. Sun, B. Wang, and X. Wang, "Predicting Polarities of Tweets by Composing Word Embeddings with Long Short-Term Memory," in Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, 2015, vol. 1, pp. 1343–1353.

[11] N. Kalchbrenner, E. Grefenstette, and P. Blunsom, "A convolutional neural network for modelling sentences", Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, Volume 1, pp 655-665, 2014.