A NOVEL MACHINE LEARNING SYSTEM FOR SENTIMENT ANALYSIS AND EXTRACTION

Osama Mohammad Rababah[1], and Nour Alokaily[2]

[1]Information Technology Department, The University of Jordan, Amman, Jordan
[2]School of archaeology and Tourism, The University of Jordan, Amman, Jordan

ABSTRACT

The huge volume of online reviews makes it difficult for a human to process and extract all significant information to make decisions. As a result, there has been a trend to develop systems that can automatically summarize opinions from a set of reviews. In this respect, the automatic classification and information extraction from users' comments, also known as sentiment analysis (SA) becomes vital to offer users the best responses to users' queries, based on their preferences. In this paper, a novel system hat offers personalized user experiences and solves the semantic-pragmatic gap was presented. Having a system for forecasting sentiments might allow us, to extract opinions from the internet and predict online user's favorites, which could determine valuable for commercial or marketing research. The data used belongs to the tagged corpus positive and negative processed movie reviews introduced by Pang and Lee[1]. The results show that even when a small sample is used, sentiment analysis can be done with high accuracy if appropriate natural language processing algorithms applied.

KEYWORDS

Machine Learning, Big Data, Natural Language Processing, Sentiment Analysis

1. INTRODUCTION

Sentiment analysis consists of the usage of language processing, text analysis, and computational linguistics to identify subjective opinion. Usually, the new data entries are compared to already classified samples, which belong to the same category.SA is the procedure of determining the polarity or intention of a written text [2].

According to[3], SA includes five steps to analyze sentiment data. The first step begins with data collection which consists of collecting data from user-generated content contained in blogs, forums, and social media networks. The collected data can be messy and expressed by different methods or by using different words, slangs, and context of writing. Manual analysis of such an enormous amount of data is not possible and exhausting. As a result, text analytics and natural language processing are used to mine and classify the data. Secondly, is the text preparation step, that is consists of cleaning the extracted data before analysis. Non-textual contents and contents that are unsuitable for the study are documented and detached. The third step is emotion detection, in which the extracted sentences of the reviews and opinions are scrutinized; sentences with individual expressions are retained, and sentences with objective communication are

discarded. The fourth step is sentiment classification where personal sentences are classified in positive, negative, good, bad, like, dislike, but classification can be made by using multiple points. Finally, it is the presentation of the output step where the key objective of sentiment analysis here is to transform unstructured text into meaningful information. At the end of the study, the test results are displayed on graphs. Also, time can be analyzed and can be graphically displayed a sentiment timeline with the chosen values of frequency, percentages, and averages over time [2].

The efficient auto-summarization of texts is a separate field of study in the computational linguistics community. One of its main goals is to offer users a way to access the content of their interest in a quicker and more efficient way [3]. SA, on the other hand, aims to be able to divide correctly text data into categories based on the opinions the authors expressed about particular issues, using natural language. To be able to offer personalized user experiences, these two fields can be analyzed holistically [4]. The novel system proposed in this article does that by merging an auto-summarization algorithm with a sentiment analysis algorithm and examining the results using the relevant metrics.

Accessing and searching reviews is frustrating when users have an imprecise idea of the product or its features and they need a recommendation or a close match. Keyword-based search does not usually provide good results, as the same keywords can appear in both good and bad reviews[5]. Another challenge in understanding studies is that a reviewer's general rating might be dedicated to the product features in which might not be of interest to the user searching. Additional challenges include having the sentiment word with an opposite meaning in a particular domain. Sarcastic sentences may violate the meaning of sentences; therefore, close attention to the words used in such sentences is needed. Other issues include when people write a word in different means which may not give us an indication that it is the same word. People's methods of expression can be inconsistent while most of the traditional text processing methods hang on the fact that a minor variance between two pieces of text doesn't alter the meaning.

This paper is prearranged as follows. Section 2 is a literature review of relevant work. Section 3 gives an overview of the methodology we adopt for this research article. Sections 4 and 5 present test cases and results obtained by running the novel system. Finally, Section 6demonstrates the conclusion.

2. STATE OF ART

The term SA first appeared in [5]. However, the research on sentiments appeared earlier [6]-[10]. The literature on SA focused on diverse fields, from computer science to management sciences, social sciences and business due to its importance to various tasks such as subjective expressions [11], sentiments of words [12], subjective sentences [13], and topics [5] [13] and [14].

SA can be approached in different manners, either by categorizing data into two groups: positive or negative[15] or by using numerous intermediary classes, such as the multiple stars reviews [1]. The sentiment classification approaches can be classified into machine learning, lexicon-based and hybrid approach [16].

The increase in new categories of online information also changes the type of summarization that is of interest. Summarization has newly been combined with work on SA [17]-[19]. Given the

numerous different reviews that one can find on the web, the problem is to identify common opinions. Some of the approaches that have been tried so far include: determining semantic properties of an object, defining the intensity of an opinion, and determining whether opinion is important. In this paper, we present a novel system that was uniting an auto-summarization algorithm with a SA algorithm to increase personalized user experience.

The auto-summarization of texts was done using the tools offered by the NLTK toolkit (NLTK.org) [20], which provide the opportunity to tag sentences syntactically and calculate word frequencies and perform stop word elimination, by using the pre-defined English corpora.

3. A NOVEL SYSTEM ARCHITECTURE

The figure below shows the system flowchart. The auto summarization consists of 6 steps which start from the original text document that is given as an argument (step 1) and generate a summary of that text by selecting the n most relevant sentences (where n is a user-defined variable) (step 6). Steps (2) to (5) encompass of sentence tagging and word frequency and relevance calculations.

The tagged texts are handled and then handed to a naïve Bayesian classifier, along with their tags as training data. Once the classifier has been trained, new observations are given for classification. Figure 1 shows the steps of auto-summarization and Sentiment analysis.

The measures used to scale the performance of the sentiment analysis were as follow:

The sensitivity or the true positive rate and sometimes called recall, measures the proportion of positives that are correctly identified. Recall is calculated as below:

True Positives / TruePositives + FalseNegatives.

The precision or the positive predictive value is the fraction of relevant retrieved instances such as the percentage of negative restaurant or movie reviews that are truly negative. Precision is calculated as below:

True Negatives / TruePositives + FalsePositives.

Accuracy is defined as the closeness of agreement between the result of a measurement and a true value [21]. Accuracy is calculated as below:

True Positives + True Negatives / TruePositives + FalseNegatives+

TrueNegatives + FalsePositives
True positives: positive comments correctly identified as positive.
True negatives: negative comments correctly identified as negative.
False positives: negative comments incorrectly identified as positive.
False negatives: positive comments incorrectly identified as negative.

The values that were found for each of these meters are shown and discussed in the results section of this paper.
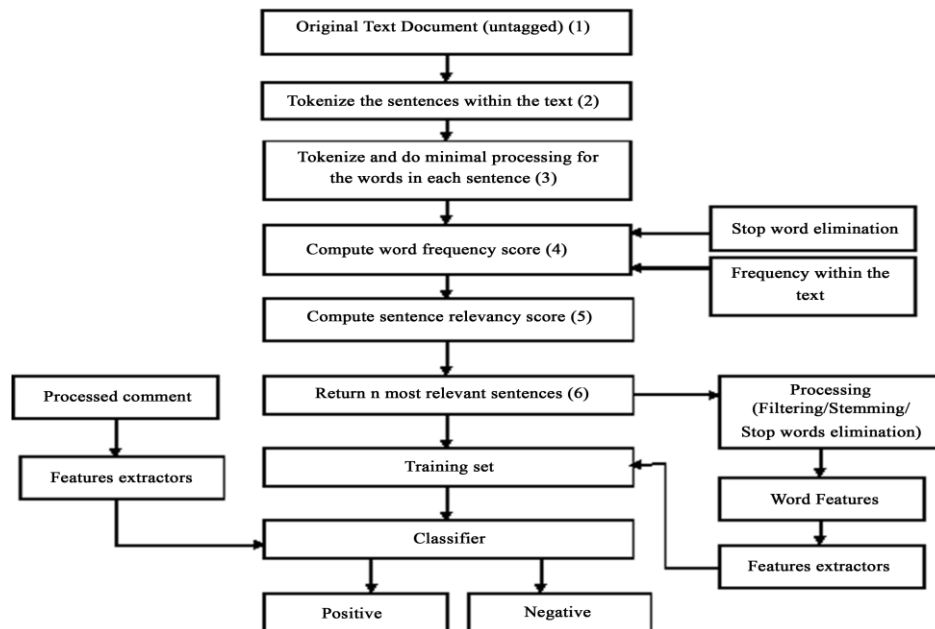

Figure 1.Systemflowchart.

4. SYSTEM TESTING

Five different forms of test cases were used:

- *No Proc process* which uses the original texts of the comments for both training and classification, with the dataset divided 20%/80%.
- *Min Proc process* only eliminates punctuation and uppercase letters, still uses the original complete textual comments for classification.
- *Sum on Sum* where all comments are summarized first and then they are used for training the Bayesian network and testing (again the 80%/20% ratio was used for the classification/testing).
- *Sum on full* where the Bayesian network is trained with the full text of the comments and the summaries are given as new items to be classified.
- *Full on Sum* where the Bayesian network is trained with the text of the summaries and the full textual comments texts are used for classification.

The 20%/80% ratio for training vs. classification was respected for all test cases, and no text was used for both training and testing (a summary of a text is considered the equivalent of the original text in this regard).

5. RESULTS

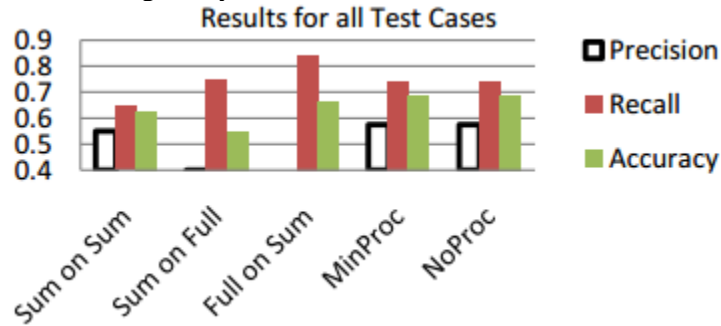The results found from running the system for each of the test cases shown below.

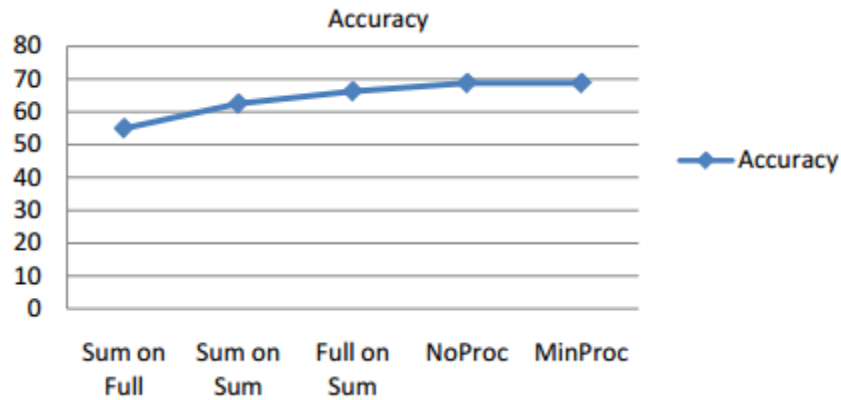Figure 2.Complete value set for all test cases.

Figure 3. Accuracy depending on the test case

Table 1.Numeric values for all the test cases.

	Precision	Recall	Accuracy
Sum on Full	0.15	0.75	0.55
Sum on Sum	0.55	0.65	0.63
Full on Sum	0.4	0.84	0.66
NoProc	0.58	0.74	0.69

As shown in Table 1 and Figure 2 the best metrics were the Sum on Sum, Min- Proc, and NoProc, the top accuracy were NoProc and MinProc. Also, it's clear from Figure 3, the best accuracy was the NoProc and MinProc. Moreover, in contrast with the other metrics, accuracy has the minimum variation.

6. CONCLUSIONS

After running the system, results shows accuracy is improved when texts of the same type are used for training and testing. However, the accuracy of the system does not vary greatly between test cases—as opposed to the other metrics. Furthermore, there was no change in the results found for texts that were not processed and the ones that had undergone minimal processing; showing polarity were not influenced by the usage of upper case vs. lower case or punctuation signs.

The precision drops radically when the system is trained with different types of texts than the trained with (Sum on Full and Full on Sum test cases), the explanation being the same as the one for the accuracy drop. As further developments for the proposed system, the following directions could be investigated: 1) Variation of the number of sentences in the summaries depending on the length of the original text—assuring that the length of the original text does not affect the training algorithm; 2) Extra processing methods could be added to the algorithm, such as stemming and stop word elimination, and the results should be reexamined to determine if the performance metrics improve for the mix test cases—Sum on Full and Full on Sum.

ACKNOWLEDGEMENTS

REFERENCES

[1]   Pang, B. and Lee, L. (2004) A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts. Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics. http://dx.doi.org/10.3115/1218955.1218990

[2]   Gupta, C., Jain, A., & Joshi, N. (2019). A Novel Approach to feature hierarchy in Aspect Based Sentiment Analysis using OWA operator. In Proceedings of 2nd International Conference on Communication, Computing and Networking (pp. 661-667). Springer, Singapore.

[3]   D'Andrea, A., Ferri, F., Grifoni, P. and Guzzo. T. (2015) Approaches, Tools and Applications for Sentiment Analysis Implementation. International Journal of Computer Applications, 125,26-33.

[4]   Abdulla, N., Ahmed, N., Shehab, M., AlAyyoub, M., Al-Kabi, M. and Al-Rifai, S. (2014) Towards Improving the Lexicon-Based Approach for Arabic Sentiment Analysis. International Journal of Information Technology and Web Engineering (IJITWE), 9, 55-71. http://dx.doi.org/10.4018/ijitwe.2014070104

[5]   Sindhu, R., Jamail, R. and Kumar, R. (2014) A Novel Approach for Sentiment Analysis and Opinion Mining. International Journal of Emerging Technology and Advanced Engineering, 4,522-527.

[6]   Nasukawa, T. and Yi, J. (2003) Sentiment Analysis: Capturing Favorability Using Natural Language Processing. Pro- ceedings of the 2nd International Conference on Knowledge Capture, Florida, 23-25 October 2003, 70-77. http://dx.doi.org/10.1145/945645.945658

[7]   Morinaga, S., Yamanishi, K., Tateishi, K. and Fukushima, T. (2002) Mining Product Reputations on the Web. Pro- ceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 341-349. http://dx.doi.org/10.1145/775047.775098

[8] Pang, B., Lee, L. and Vaithyanathan, S. (2002) Thumbs up? Sentiment Classification Using Machine Learning Tech- niques. Proceedings of the 7th Conference on Empirical Methods in Natural Language Processing,79-86.

[9] Alaei, A. R., Becken, S., &Stantic, B. (2019). Sentiment analysis in tourism: capitalizing on big data. Journal of Travel Research, 58(2), 175-191.

[10] Turney, P. (2002) Thumbs up or Thumbs down? Semantic Orientation Applied to Unsupervised Classification of Re- views. Proceedings of the 40th ACL,417-424.

[11] Wiebe, J. (2000) Learning Subjective Adjectives from Corpora. Proceedings of National Conference on Artificial Intelligence.

[12] Wilson, T., Wiebe, J. and Hoffmann, P. (2009) Recognizing Contextual Polarity: An Exploration of Features for Phrase-Level Sentiment Analysis. Computational Linguistics, 35, 399-433. http://dx.doi.org/10.1162/coli.08-012-R1-06-90

[13] Peacock, D. C., & Khan, H. U. (2019). Effectiveness of Social Media Sentiment Analysis Tools with the Support of Emoticon/Emoji. In 16th International Conference on Information Technology-New Generations (ITNG 2019) (pp. 491-494). Springer, Cham.

[14] Yi, J., Nasukawa, T., Niblack, W. and Bunescu, R. (2003) Sentiment Analyzer: Extracting Sentiments about a Given Topic Using Natural Language Processing Techniques. Proceedings of the 3rd IEEE International Conference on Data Mining (ICDM 2003), Florida, 19-22 November 2003, 427-434. http://dx.doi.org/10.1109/icdm.2003.1250949

[15] Hiroshi,K.,Tetsuya,N.andHideo,W.(2004)Deeper Sentiment Analysis Using Machine Translation Technology.Proceedings of the 20th International Conference on Computational Linguistics (COWLING 2004), Geneva, 23-27 August 2004, 494-500. http://dx.doi.org/10.3115/1220355.1220426

[16] Govindarajan, M. (2013) Sentiment Analysis of Movie Reviews Using Hybrid Method of Naive Bayes and Genetic Algorithm. International Journal of Advanced Computer Research, 3,139-145.

[17] Albanese, M. (2013) A Multimedia Recommender System. ACM Transactions on the Internet Technology (TOIT), 13, 1-32.http://dx.doi.org/10.1145/2532640

[18] Carenini, G. and Cheung, J. (2008) Extractive vs. NLG-Based Abstractive Summarization of Evaluative Text: The Ef- fect of Corpus Controversiality. Proceedings of the International Natural Language Generation Conference, 33-41. http://dx.doi.org/10.3115/1708322.1708330

[19] Lerman, K., Blair-Goldensohn, S. and McDonald, R. (2009) Sentiment Summarization: Evaluating and Learning User Preferences. Proceedings of the Conference of the European Chapter of the Association for Computational Linguistics, 514-522. http://dx.doi.org/10.3115/1609067.1609124

[20] NLTK.org. (n.d.). Retrieved 01 28, 2016. http://www.nltk.org/index.html

[21] Jagdale, R. S., Shirsat, V. S., &Deshmukh, S. N. (2019). Sentiment analysis on product reviews using machine learning techniques. In Cognitive Informatics and Soft Computing (pp. 639-647). Springer, Singapore.