ISSN : 2455-9679
Impact Factor : 3.340

IJRTSM

INTERNATIONAL JOURNAL OF RECENT TECHNOLOGY SCIENCE & MANAGEMENT

"AMAZON PRODUCT REVIEW ANALYSIS USING SENTIMENT BASED TEXT CLASSIFICATION TECHNIQUE: A PROPOSAL "

Richa Chunekar Kokje [1], Gajendra Singh Chaouhan [2]

[1] *M.Tech Scholar, Department of Computer Science Engineering, MIST Inodre, MP, India.*

[2] *Assistant Professor, Department of Computer Science Engineering, MIST Inodre, MP, India.*

ABSTRACT

The data mining and machine learning techniques enable us to analyze and recover target patterns using computational algorithms. These algorithms evaluate the data without human efforts. This ability of algorithm makes data mining acceptable for various applications. Among these applications a significant amount of applications are utilizing the text data, and the classical approach of text classification is not suitable for finding the emotion labels. Thus, in this paper the sentiment classification based application is proposed for analyzing the Amazon product reviews. That utilizes the NLP based text processing to accurately classify and exploring the hidden sentiments in Amazon product reviews. The proposed technique is unlike to classify the binary sentiments that classify the text into five different sentiment labels. This paper provides the basic overview of the proposed technique includes a review on recent literature and also offers a design of required sentiment based text classification system. Finally the conclusion and future work is reported in this paper.

Key Words: *Online Social Networks, supervised learning, sentiment analysis, Amazon product review, text classification, multiclass sentiment analysis*

I. INTRODUCTION

In various real world applications (i.e. Facebook, Amazon, twitter and others), a significant amount of data is generated by users. Such kind of data analysis needs significant human efforts. Additionally the existing methods of text classification are binary or less accurate. Moreover it the normal classification process is not suitable for our application requirements. Because the data classification is not subjective here, in this application we need to approximate the emotions behind the text. Basically such kind of data analysis is used for various business intelligence applications. The aim of this analysis is to extract the emotions of end consumer for a product or service [1].

Basically, the users express their emotions by using social media post or any product review. The expressed text by the consumer helps to understand the sentiments about the offered service or product [2]. The understanding about the emotions is a research domain of NLP (natural language processing) and machine learning. Thus in this work Amazon product reviews are analyzed using the NLP based text classification [3]. The proposed technique helps to analyze and recover the sentiments of consumer for the given product.

Thus, the proposed system first extracts the product reviews from the Amazon e-commerce. These reviews are in form of text thus initially text mining techniques are needed to be employ first so preprocesses of the content is opted. During the preprocessing the aim is to improve the quality of learning data. Thus noisy contents are removed from input text reviews. After preprocessing the feature extraction is applied for recovering the sentiments. Here for NLP feature extraction the NLP parser is used. During this process the keywords are selected which are having potential

ISSN : 2455-9679
Impact Factor : 3.340

sentiments. And NLP based extracted features are transformed into a fixed amount of sentiment attributes. These attributes are used for training and testing of the classifier.
This section provides the overview of the proposed work involved in the given proposal, in next section the literature review is presented for understanding of recent development on the sentiment based text classification. Further the problem domain of the work is formulated and then solution of the addressed problem is explained. Finally the conclusion and future work is provided.

II. LITERATURE REVIEW

This section provides the recent research contributions which are intended to improve or employ the sentiment classification. Therefore available research article are reported in this section.

People express their opinions about services, products, events, etc., in social media, blogs, and comments. Majority of research efforts are devoted to English data, while a great share of information is available in other languages. *Kia Dashtipour et al [4]* present a review on multilingual sentiment analysis. Authors compare own implementation on common data. Precision observed is typically lower than the one reported by the original authors. Thus, authors compare the existing works, including whether they allow for accurate implementation and for reliable reproduction of results. *Bo Zhao et al [5]* proposes a stock market prediction method exploiting sentiment analysis using financial micro-blogs. Authors analyze microblog texts to find sentiments, and then combine the sentiments and the historical data of the Shanghai Composite Index to predict the stock movements. It includes three modules: Micro-blog Filter (MF), Sentiment Analysis (SA), and Stock Prediction (SP). MF module is based on LDA to get the financial micro-blogs. The SA module first gets the sentiments of the micro-blogs. The SP module is a user-group model which adjusts the importance of different people, and combines it with stock historical data to predict the movement.

Today, we can buy everything on e-commerce websites and get it delivered. There is hardly anything that we cannot find on e-commerce websites. Judging a product just by its pictures and reviews is a difficult. To analyze these reviews and to make a decision is a tiresome task. To make the task easier *Aarti Potdar et al [6]* are proposing "Samiksha", a review bot which will generate summarization of all the user reviews. The software will produce an average numerical rating of the particular product to help the buyer. The Samiksha will prove to be convenient medium of analyzing the reviews. A huge number of videos are posted every day on social media. *Soujanya Poria et al [7]* propose a methodology for multimodal sentiment analysis, which consists in harvesting sentiments from Web videos by demonstrating a model that uses audio, visual and textual modalities. Authors used both feature- and decision-level fusion methods to merge affective information. A comparison with existing works is carried out throughout the paper. The experiments with YouTube dataset show that the proposed system achieves accuracy of nearly 80%, outperforming all state-of-the-art systems more than 20%.

One of the main characteristics of Industry is support the integration and virtualization of design and production process. *Kássio Santos et al [8]* will present a review of Industry 4.0 based on recent developments in research and practice. A multi-criteria decision making approach using the Promethee will realize an analysis of Product Development Process and Industry 4.0 concepts, considering relationship, attributing values, and some other characteristics. An overview of different opportunities for product development process in Industry 4.0 will be presented. The e-commerce is taking the ascendancy by making products available. People are relying on online products so the importance of a review is going higher. For selecting a product, a customer goes through reviews to understand a product. Going through thousands of reviews would be much easier if a model is used to polarize reviews. *T. U. Haque et al [9]* used supervised learning method on Amazon dataset to polarize it and get satisfactory accuracy.

The reviews toward sites express feeling. The process of finding user opinion about the topic or product is called opinion mining. It can also be defined as the process of automatic extraction of knowledge by opinions expressed by the user about some product. Analyzing the emotions from the extracted opinions is defined as Sentiment Analysis. *Santhosh Kumar K L et al [10]* concentrates on mining reviews from Amazon. It automatically extracts the reviews from the website. And it uses algorithm such as Naïve Bayes classifier, Logistic Regression and SentiWordNet to classify the review as positive and negative. Authors have used quality metric to measure the performance of algorithms. Reviews are helpful in attracting attention of practitioners and academics. It helps in reducing risks and uncertainty in online shopping. *MSI Malik et al [11]* examines uninvestigated variables by looking review

ISSN : 2455-9679
Impact Factor : 3.340

characteristics and important indicators. Several review content and two reviewer variables are proposed and a review helpfulness prediction model is developed. Authors derive a mechanism to extract review content variables. Six popular machine learning models and three real-life Amazon review data sets are used. The results are robust to several product categories. The results show that review content variables deliver the best performance as compared to the reviewer. This study finds that reviewer helpfulness per day and syllables in review text strongly relates to review helpfulness.

Sentiment analysis is broadly employed for extracting the polarity of text using NLP methods. *Abhilasha Singh Rathor et al [12]* focuses on examining the efficiency of three machine learning techniques Support Vector Machines (SVM), Naive Bayes (NB) and Maximum Entropy (ME) for online reviews classification. The reviews are divided as positive, neutral and negative. This is helpful for consumers who search the reviews of products to purchase. They have extracted Amazon Reviews using Amazon API. And used unigrams and weighted unigrams to train classifiers. The results have shown that algorithms work well on weighted unigrams and SVM. Customer reviews are not only helpful for customers, but it is also helpful for the manufacturers to raise their products. The reviews take the attention of the customers. Opinion Mining is playing a major role to summarize reviews and make it easy for customers. *JAWAD KHAN et al [13]* propose a supervised lazy learning model utilizing syntactic rules for the product features and opinion words in subjective review. In this algorithm, K-NN with k=3 is used for classification into two classes (subjective, objective). The experiment show that proposed method can improve the performance of existing work in terms of precision, recall and f-score.

The studied researches articles are discussed above are summarized using table 1. The table consist of author names, publication and the publication year, and finally the core contribution of research work is explained.

Table 1 the review summary

Authors	Publication/ year	Contribution
Kia Dashtipour et al [4]	2016 Springer	People express their opinions about services, products, events, using social media, blogs, and comments. Authors present a review on multilingual sentiment analysis. They compare own system on common data. Precision is typically lower than one reported other author.
Bo Zhao et al [5]	2016 IEEE	Authors propose a stock market prediction method using sentiment analysis of financial blogs. And combine the sentiments and the historical data of the Shanghai Composite Index to predict stock movements. It includes: Micro-blog Filter, Sentiment Analysis, and Stock Prediction.
Aarti Potdar et al [6]	2016 Elsevier	To analyze reviews and to make a decision is a tiresome task. Authors are proposing “Samiksha”, a review bot which will generate summarization of all the reviews. It will produce an average numerical rating for products. The Samiksha will prove to be convenient medium of analyzing the reviews.
Soujanya Poria et al [7]	2015 Elsevier	Authors propose a methodology for multimodal sentiment analysis, from Web videos by uses audio, visual and textual modalities. They used feature and decision-level fusion to merge information. The experiments with YouTube dataset show that the system achieves accuracy of nearly 80%.
Kássio Santos et al [8]	2017 Elsevier	Authors will present a review of Industry 4.0 based on recent developments. A multi-criteria decision making approach using the Promethee will realize, considering relationship, attributing values, and some other characteristics.

ISSN : 2455-9679
Impact Factor : 3.340

T. U. Haque et al [9]	2018 IEEE	For selecting a product, a customer goes through reviews to understand a product. Going through thousands of reviews would become easier if a model is used to polarize reviews. Authors used supervised learning method on Amazon dataset to polarize it.
Santhosh Kumar K L et al [10]	2016 IEEE	Author concentrates on mining reviews from Amazon. It automatically extracts the reviews and uses algorithm such as Naïve Bayes, Logistic Regression and SentiWordNet to classify the review as positive and negative. The quality metric to measure the performance of algorithms.
MSI Malik et al [11]	2017 Elsevier	Reviews help in reducing risks and uncertainty in online shopping. Authors examine review characteristics and indicators. Several content and two reviewer variables are proposed for a review helpfulness prediction model. Six popular ML models and three real-life Amazon review data sets are used. The results show that review content variables deliver the best performance as compared to the reviewer.
Abhilasha Singh Rathor et al [12]	2018 Elsevier	Author focuses on examining the efficiency of three ML techniques SVM, NB and ME for online reviews classification. This is helpful for consumers to purchase products. Using Amazon API And unigrams and weighted unigrams are used to train. The results show that algorithms work well on weighted unigrams and SVM.
JAWAD KHAN et al [13]	2016 IEEE	Author proposes a supervised lazy learning model for syntactic rules for product opinion in subjective review analysis. In this algorithm, K-NN with k=3 is used to classify data into subjective, and objective classes

III. PROPOSED WORK

The proposed work is motivated by the research article [14]. In [14] this work the sentiment based Chinese text is classified. In addition of that this method provides the prediction for the two class labels i.e. negative and positive. That is an essential, efficient and accurate approach and promising for extension. Thus the following key issues are addressed for extending the given work.

1. The proposed text processing technique is need to extend for English text classification according to the text sentiments
2. Secondly the proposed system includes 5 sentiment labels for identifying Amazon text reviews

In order to enhance the existing opinion mining model a new data model is proposed. That model offers to analyze the Amazon product reviews. The required data model for classification is demonstrated in figure 1.

In this diagram the different components of the proposed sentiment classification is associated. According to the given system the system accept the Amazon product reviews. The available on product review is extracted using the API or directly from available dataset. Both kinds of datasets are available in raw format. Therefore the preprocessing technique is applied over text data. The preprocessing is aimed to refine the valuable information and eliminate the noise from the text. During the preprocessing it is tried to enhance the quality of data. Thus first only the text component of the review text is extracted. After that the stop words and used special characters from text is removed. That process reduces the amount of data and improves the content quality.

ISSN : 2455-9679
Impact Factor : 3.340

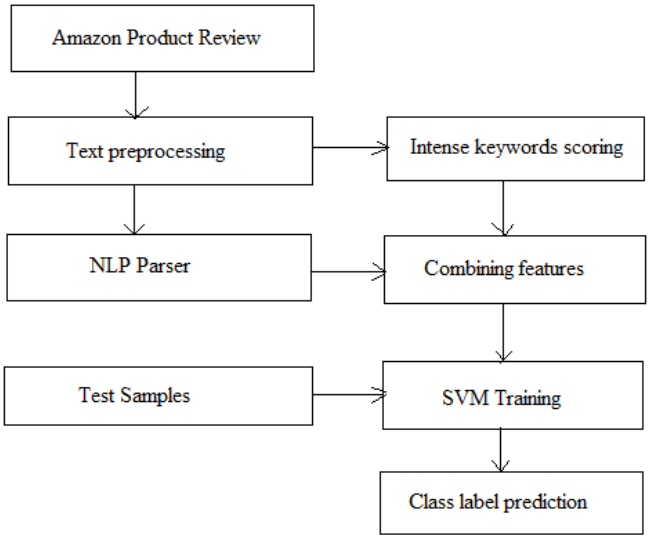

Figure 1 proposed system

After refinement of the intense keywords from the preprocessed data these keywords are preserved separately. Additionally the NLP parser is applied on the preprocessed data. The NLP parser returns the POS (part of speech) features of the given sentence. That POS are used for tagging of the text and this process is known as POS tagging. After this step the unstructured data is transformed into a 2D vector.

That 2D vector is combined with the intense keyword score. Means calculated review features are combined with two other derived features which are known as intensity score. The extended features used with the supervised learning classifier. In this presented work the SVM (support vector machine) is used in the manner of one-vs.-all. The SVM is basically a binary classifier, and for recognizing the multiple class labels additional efforts are required. The combined data is processed using the SVM classifier for training and after training test samples are classified for identifying the sentiment class labels. The classified Amazon product reviews are used for validating the implemented opinion mining technique. Thus accuracy and error rate is also computed using the classified data.

IV. CONCLUSION

The aim of the proposed work is to investigate about the text classification technique using sentiment analysis. Therefore a data mining model is presented in this paper. Before expressing the required data model a brief review on recently contributed literature is also reported. According to the obtained conclusion the sentiment analysis of a text most of the cases in reviews computed in terms of positive or negative, and very often in terms of neutral. But there are fewer efforts that are works for providing some score for the concluded sentiments. The score of an Amazon product review help to understand the satisfaction level of an end client and in similar manner dissatisfaction level. Therefore using the concluded summary a multi-class text classification system is proposed for designing and implementation. The offered approach is promising for various other application areas too.

- It is also suitable for understanding the learning process of students
- That help to understand the teachers performance and to improve the academic productivity
- To understand the banding of a product launch and similar others

The proposed model is a machine learning model which is used with learning process and helps to recognize the similar patterns. The model helps to polarize the text contents according to the sentiments of text. In near future the proposed work is demonstrating the classification model and the performance of the system proposed.

ISSN : 2455-9679
Impact Factor : 3.340

REFERENCES

[1] S. Pouyanfar, Y. Yang, S. C. Chen, M. L. Shyu, S. S. Iyengar, *"Multimedia Big Data Analytics: A Survey"*, ACM Computing Surveys, Vol. 51, No. 1, Article 10. Publication date: January 2018.

[2] S. Shayaa, S. Ainin, N. I. Jaafar, S. B. Zakaria, S. W. Phoong, W. C. Yeong, M. Ali A. Garadi, A. Muhammad, A. Z. Piprani, *"Linking consumer confidence index and social media sentiment analysis",* Cogent Business & Management (2018), 5: 1509424.

[3] W. Medhat, A. Hassan, H. Korashy, *"Sentiment analysis algorithms and applications: A survey"*, Ain Shams Engineering Journal 2014, 5, 1093-1113.

[4] K. Dashtipour, S. Poria, A. Hussain, E. Cambria, A. Y. A. Hawalah, A. Gelbukh, Q. Zhou, *"Multilingual Sentiment Analysis: State of the Art and Independent Comparison of Techniques"*, Cogn Comput (2016) 8:757–771 DOI 10.1007/s12559-016-9415-7.

[5] B. Zhao, Y. He, C. Yuan, Y. Huang, *"Stock Market Prediction Exploiting Microblog Sentiment Analysis"*, 978-1-5090-0620-5/16/$31.00 c 2016 IEEE.

[6] A. Potdar, P. Patil, R. Bagla, R. Pandey, Prof. N. Jadhav, *"SAMIKSHA - Sentiment Based Product Review Analysis System",* Procedia Computer Science 78 (2016) 513 – 520, 2016 The Authors. Published by Elsevier B.V.

[7] S. Poria, E. Cambria, N. Howard, G. B. Huang, A. Hussain, *"Fusing audio, visual and textual clues for sentiment analysis from multimodal content"*, Neurocomputing 174 (2016) 50–59, & 2015 Elsevier B.V. All rights reserved.

[8] K. Santos, E. Loures, F. Piechnicki, O. Canciglieri, *"Opportunities Assessment of Product Development Process in Industry 4.0"*, Procedia Manufacturing 11 (2017) 1358 – 1365, © 2017 Elsevier B.V.

[9] T. U. Haque, N. N. Saber, F. M. Shah, *"Sentiment Analysis on Large Scale Amazon Product Reviews"*, 2018 IEEE International Conference on Innovative Research and Development (ICIRD), 978-1-5386-5283-1/18/$31.00 ©2018 IEEE.

[10] S. Kumar K L, J. Desai, J. Majumdar, *"Opinion Mining and Sentiment Analysis on Online Customer Review"*, 978-1-5090-0612-0/16/$31.00 ©2016 IEEE.

[11] M. Malik, A. Hussain, *"An analysis of review content and reviewer variables that contribute to review helpfulness"*, Information Processing and Management xxx (2017) xxx-xxx.

[12] A. S. Rathor, A. Agarwal, P. Dimri, *"Comparative Study of Machine Learning Approaches for Amazon Reviews"*, procedia Computer Science 132 (2018) 1552–1561.

[13] J. Khan, B. S. Jeong, *"Summarizing Customer Review Based On Product Features and Opinion"*, Proceedings of the 2016 International Conference on Machine Learning and Cybernetics, Jeju, South Korea, 10-13 July, 978-1-5090-0390-7/16/$31.00 ©2016 IEEE.

[14] Y. Fang, H. Tan, J. Zhang, *"Multi-Strategy Sentiment Analysis of Consumer Reviews Based on Semantic Fuzziness"*, 2169-3536 2018 IEEE.