Sentiment Analysis of Product-Based Reviews Using Machine Learning Techniques

JitendraSoni*, Dr.KirtiMathur**,AmanPathan*,Pravar More*, VaishaliBaghel*, Toshi Anand*, TushitaThusu*

*Institute of Engineering & Technology, Devi AhilyaVishwaVidyalaya, Indore
**International Institute of Professional Studies, Devi AhilyaVishwavidyalaya, Indore
morepravar@gmail.com

ABSTRACT

Sentiment Analysis can be explained as sorting and dividing sentiments on the basis of emotions, mainly: good, bad or neutral. Opinion Mining is the main job done using NLP (natural language processing). Opinion Mining can also be known as Sentiment Analysis. Despite the research on this topic having gained a lot of attention in the recent years, some issues still remain. In this paper, our attempt is to resolve these issues. Not being able to sort Sentiments on the basis of contradiction. Studying and understanding the process, a simple way for Sentiment classification is proposed. We gathered data from Amazon.com for this study. We have performed categorization based on factors: sentence level and review level and have obtained promising outcomes. Also, other than all of this, we are also giving an insight into our mood analysis study in the following paper.

I.   INTRODUCTION

Sentiment refers to an attitude, belief, or conclusion brought on by an emotion. In general terms it is specified as opinion mining. It depicts and analysis how people feel or think about different products. It has become habitual for us to look for reviews and ratings before purchasing something, even before going to a restaurant or on a vacation. Nowadays there are companies and social media sites that analyze their user information on the basis of the data they collect and then present the desirable results to the customer. This helps the company in self evaluation and improve its services. Sentiment analysis of product reviews work in a similar way. Here we are analyzing each review with the help of natural language processing toolkit, the sentence is broken into tokens and then each word is analyzed. We will be using various techniques for this purpose which are properly mentioned in further sections. Sentiment analysis as a technology has been mainly useful for companies to improve their performance among people, to analyze their own products and take necessary measures for better market performance. In this study, we are considering mobile phone reviews from Amazon.in and then analyzing it using the below mentioned ways. Section 2 gives information about the prior literature review done in this field, section 3 provides us the methodology used in analyzing and collecting the data, section 4 summarizes the whole study and gives a proper conclusion.

II.   Research Design and Methodology

Data Collection

Data collection for our research involved the acquisition of over 5 million product reviews from Amazon.com between the months of January and March 2021.

These reviews are submitted by more than 3 million customers for over 20,000 products, which were categorized into four groups: Beauty and wellness, Books, Electrical devices and Home. Each review contained essential information such as the ID of the reviewer, ID of the product, rating of the product, time at which review was done, Usefulness, and Text Check.

(a)

(b)
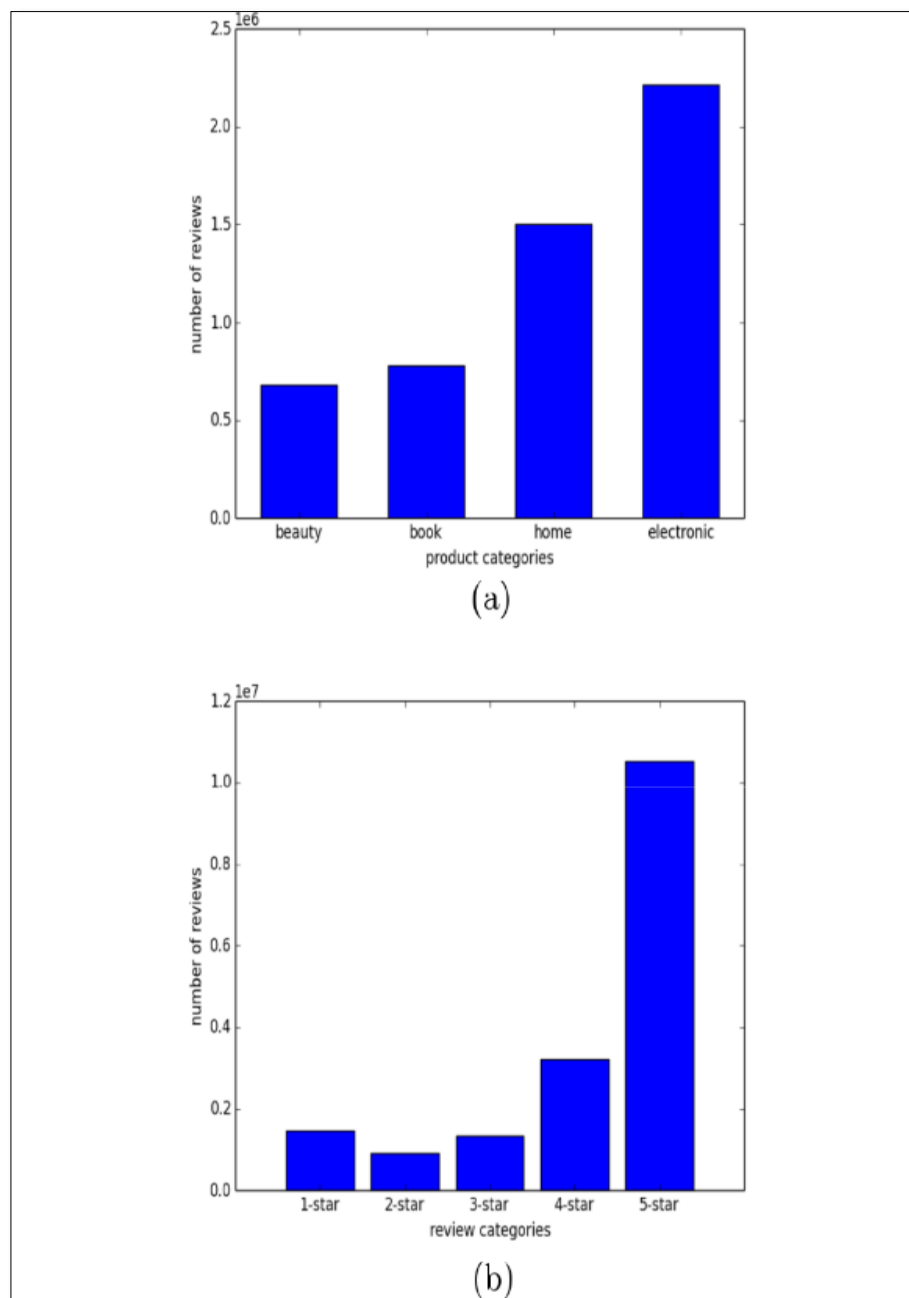
Ratings were provided on a scale of 1-5 stars, without any half-star rating options(above fig(b)).

Sentiment sentences extraction and Part-of-speech tagging

To analyze the subjective content of the reviews, we extracted all sentences that contained positive or negative words, while ignoring the objective content. Syntactic analysis of the sentences was done using part-of-speech tagging, where each word in the sentence was identified based on its syntactic role.

Nouns, pronouns, verbs, prepositions, adjectives, adverbs, conjunctions, and interjections are eight parts of speech in English language. Our POS taggers proved essential in filtering out words that do not express emotions such as nouns and pronouns. We were also able to distinguish between words that may be employed in different parts of speech thanks to POS taggers.

Sentiment score computation for sentiment token

The following is the formula for t's sentimental score (SS):

$$SS(t) = \frac{\sum_{i=1}^{5} i \times \gamma_{5,i} \times Occurrence_i(t)}{5 \times \sum_{i=1}^{5} \gamma_{5,i} \times Occurrence_i(t)}$$

Occurrence $i(t)$ represents the number of times that t has appeared in reviews with an i-star rating where i ranges from 1 to 5. The figure above indicates that our dataset is uneven and that there are various number of reviews for each star rating. Accordingly, we introduce the ratio,$\gamma_{5,i}$, which is given as

$$\gamma_{5,i} = \frac{|5-star|}{|i-star|}$$

where $i=1,...,5$

There are two processes for the classification of emotional polarity:

Sentence-level classification and Review-level classification. We want to categorise a sentence as positive or not positive(negative) based on the emotion it conveys. It is impossible to manually categorise the vast amount of data we have as positive or negative. So, to do so, we need an algorithm. The programme counts both positive and negative inputs in each and every sentence. If the positive words are more than negative than the sentence is marked as positive and vice-versa. For review-level classification there is star-scale ratings.

Feature Vector Formation

The information that we obtained from the original dataset is sentiment tokens and sentiment scores. These are the characteristics or traits that are used for categorization of emotions. To tailor a classifier, each entry in the training data must be changed into a vector consisting these features known as feature vector.The challenges that we face while creating this vector are:

-large number of features or feature values should not be present in vectors

- to suit the classifier, vector must have the same amount of dimensions

We use 2 binary strings "1" and "0" to represent the occurrence of each token. For storing the triggered strings we use the hashing function in Python.

III. Outcome and discussion Methods of evaluation

Based on each classification model's averaged FI-score, performance is calculated as:

$$F1_{avg} = \frac{\sum_{i=1}^{n} \frac{2 \times P_i \times R_i}{P_i + R_i}}{n}$$

Where n is the number of classes, Pi is the precision of the ith class, Ri is the recall of the ith class, Pi and Ri are calculated using 10-fold cross validation.

The dataset is partitioned into 10 equal subsets for 10-fold cross validation, with each subset comprising an equal number of positive and negative class vectors. One of the 10 subsets is set aside for testing, and the other nine are used for training. We carry out this procedure ten times, saving a different tenth for testing each time. Then we take average of these 10 result to get a single result.

Overall, this methodology allowed us to collect and analyze a vast amount of data effectively, providing valuable insights into customer sentiment towards different product categories on Amazon.com .

IV. Conclusion

The objective of this research is to classify emotional polarity as positive, negative or neutral. As we have seen that sentiment analysis deals with different emotions, notions or human feelings. Also, after understanding the research for sentiment analysis using product review major issue we have dealt with is emotional polarity. We have taken data from Amazon.com for conducting different operations in this research paper. We have used NLP and ML techniques to analyse the data. A proper classification of sentence-level and review-level analysis is conducted in this paper. In addition, we have selected a proper solution for overcoming the problem of emotional polarity here.

References

1. S. ChandraKala1 and C. Sindhu2, "OPINION MINING AND SENTIMENT CLASSIFICATION: A SURVEY,".Vol .3(1),Oct 2012,420-427G.Angulakshmi , Dr.R.ManickaChezian ,"An Analysis on Opinion Mining: Techniques and Tools". Vol 3(7), 2014 www.iarcce.com.

2. Callen Rain,"Sentiment Analysis in Amazon Reviews Using Probabilistic Machine Learning" Swarthmore College, Department of Computer Science.

3. Padmani P .Tribhuvan,S.G. Bhirud,Amrapali P. Tribhuvan," A Peer Review of Feature Based Opinion Mining and Summarization"(IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (1), 2014, 247-250 ,www.ijcsit.com.

4. Carenini, G., Ng, R. And Zwart, E. Extracting Knowledge from Evaluative Text. Proceedings of the Third International Conference on Knowledge Capture (K-CAP'05), 2005.

5. Dave, D., Lawrence, A., and Pennock, D. Mining the Peanut Gallery: Opinion Extraction And Semantic Classification of Product Reviews. Proceedings of International World Wide Web Conference (WWW'03), 2003.

6. Zhu, Jingbo, et al. "Aspect-based opinion polling from customer reviews." IEEE Transactions on Affective Computing, Volume 2.1,pp.37-49, 2011.

7. Na, Jin-Cheon, Haiyang Sui, Christopher Khoo, Syin Chan, and Yunyun Zhou. "Effectiveness of simple linguistic processing in automatic sentiment classification of Product reviews." Advances in Knowledge Organization Volume9, pp. 49-54, 2004.

8. Nasukawa, Tetsuya, and Jeonghee Yi. "Sentiment analysis: Capturing favorability Using natural language processing." In Proceedings of the 2nd international conference On Knowledge capture, ACM, pp. 70-77, 2003.

9. Li, Shoushan, Zhongqing Wang, Sophia Yat Mei Lee, and Chu-Ren Huang. "Sentiment Classification with Polarity Shifting Detection." In Asian Language Processing (IALP), 2013 International Conference on, pp. 129-132. IEEE, 2013.

10. W. Zheng, H. Jin, Y. Zhang, X. Fu and X. Tao, "Aspect-Level Sentiment Classification Based on Auto-Adaptive Model Transfer," in IEEE Access, vol. 11, pp. 34990-34998, 2023, doi: 10.1109/ACCESS.2023.3265473.

11. J. Khan, N. Ahmad, S. Khalid, F. Ali and Y. Lee, "Sentiment and Context-Aware Hybrid DNN With Attention for Text Sentiment Classification," in IEEE Access, vol. 11, pp. 28162-28179, 2023, doi: 10.1109/ACCESS.2023.3259107.