JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Feature wise Sentiment based E-Commerce Review system using Machine Learning

[4]Chitra Rathod, [2]Prof. Deep Joshi, [3]Prof. Anjuman Ranavadiya , [4]Zala Krishna, [5]Patel Deesha
[1] ME Student, [2] Assistant Professor, [3]Assistant Professor,[4] ME Student, [5]Assistant Professor
Computer Ennginеering Department
Grow More Group of Institutions ,,Himmatnagar, Dist. S. K.,Gujarat – 383001

ABSTRACT

With the ease of E-Commerce and Mobile Commerce, a new domain for research on reviews by customer has grown. So many researchers have worked toward the same topic and also got significant results. Sentiment analysis over short text and particularly over customer reviews is not as straight forward work as other sentiment analysis work due to variety of products, product features and fake review issues.

Topic modeling technique with sentiment analysis can help for better classification of customer reviews. Various machine learning techniques can be applied to dig the real sentiments from the given reviews. In our research work we are going to apply improvement on Graph based multiclass classification for customer reviews from Amazon.

Keywords: Sentiment Analysis, DNN, LDA, MDK-LDA, SVM

INTRODUCTION

In General Term Machine Learning is a technique in which Machine is trained to learn based on Historical Data and Events. In machine learning there are two main phase: Training Phase and Testing Phase. In Training Phase a model is trained over historical data and various events from this data. In Testing Phase model is tested over pre classified data and various parameters are measured.

1.1 Applications of Machine Learning

Mostly we can classify application of Machine Learning in two categories.

A) Classification:

"The Classification of a dataset is such type of technique that is mostly used to detect the category or classes of new data from the sample data. It detects based on earlier provided data that is known as training data. In Classification technique, a model (or algorithms) learns from the input informational dataset samples (Training Data) or samples and then classifies new sample into a number of classes or categories"[12].

There are main three types of learning used for Classification

i) Supervised Learning
ii) Unsupervised Learning
iii) Reinforcement Learning

B) Prediction:

"Prediction of a model in machine learning refers to the outcome of any algorithm after training on a historical dataset and tested to new testing data with aim of forecasting. Example of Prediction in machine learning are like Weather data Prediction, Customer data Purchase Prediction, Query Prediction etc. Mostly it works over probability criteria. Prediction has no general world meaning in machine learning. Its purely based on Mathematical Model"[11].

1.2 NLP

Natural Language Processing (NLP) is generally used for computers to understand the human languages. In background, NLP analyzes the grammatical structure of sentences and the individual meaning of words, then uses algorithms to extract meaning and deliver outputs.[17] we can also say, it makes 'sense of human language' so that it can automatically perform different tasks. Best examples of NLP in action are various virtual assistants, like Google Assist, Siri, and Alexa. NLP understands written as well as spoken text like "Hey Alexa, Which College is Best of Master Degree ?" and transforms it into a proper way, making it easy for machines to understand. Another well-known application of NLP is chatbots. They help support systems solve issues by understanding common language requests for the support and responding automatically by their system based on user question. There are lots of everyday apps you use, where we probably encountered NLP without even noticing the use of NLP. Text recommendations when writing an email in many email clients, offering to translate a Facebook post written in a different language, or filtering unwanted promotional emails into your spam folder. In short, the goal of Natural Language Processing is to make human language – which is complex, ambiguous, and extremely diverse – easy for machines to understand.

1.3 Common NLP Techniques

Natural Language Processing (NLP) applies main two techniques to understand text: syntactic analysis and semantic analysis [5].

Syntactic Analysis:

Syntactic analysis – or parsing – analyzes text using basic grammar rules to identify sentence structure, it recognize how words are organized, and how words are related to each other in a sentence.
Some of its main sub-tasks may include:

Tokenization:

Tokenization consists of breaking up a text into smaller parts that is called tokens to make text easier to handle. For processing it is done.

Part of speech tagging (PoS tagging):

Part of speech tagging (PoS tagging) labels tokens (extracted words) as verb, adverb, adjective, noun, etc. POS helps infer the meaning of a word. Place of any word can have different meaning. Ex. 'book a ticket' , 'buy the book'

Lemmatization & stemming:

Lemmatization & stemming consist of reducing inflected words to their base form to make them easier to analyze. Best example of stemming is 'Swimming to 'swim'.

Stop-word removal removes frequently occurring words (Mostly the words like articles etc) that don't add any semantic value, such as I, they, have, like, yours, etc.

Semantic Analysis

Semantic analysis mainly focus on capturing the meaning of content. In first stage, it studies the meaning of every word (lexical semantics). Then, it checks for the combination of words and what they mean in context. The main sub-tasks of semantic analysis are:
Word sense disambiguation tries to identify in which sense a word is being used in a given context.
Relationship extraction attempts to understand how entities (places, persons, organizations, etc) relate to each other in a text.

1. RELATED WORKS

For our research work we have reviewed many research papers those have worked with Sentiment Analysis using NLP in recent time. Research [1] proposed a sentiment multi classification method based on a directed weighted model. The model represents the sentiment entity vocabulary as the sentiment nodes and represents the relation between nodes as the directed weighted link. They have applied MDK-LDA Algorithm and used Amazon Dataset with 10,000 Reviews. They have achieved 74.70% accuracy; and it is better compare to BERT that is 73.36 %. Apart from accuracy improvement they also reduced CPU Time compare to BERT. In their research work no Feature wise Sentiment categorized and single source of review has been used. So it may lead to false result. Also they have not applied any fake review detection technique.

Figure 1 Multiclassification of Sentiment[1]

In research work[2] the researchers proposed a new sentiment analysis model-SLCABG, which is based on the sentiment lexicon and combines Convolutional Neural Network (CNN) and attention-based Bidirectional Gated Recurrent Unit

(BiGRU). They have achieved 93.50 % accuracy in their work over large dataset. They have performed their research over book reviews and to fastern the process they have trimmed the reviews. Following table 1 shows the result achived by the research work[2].

Model	Accuracy	Precision	Recall	F1
Naive Bayes	57.9%	55.6%	79.2%	65.3%
SVM	67.7%	93.8%	38.4%	54.5%
CNN	90.9%	91%	90.2%	90.6%
CNN+	91.4%	90.8%	91.6%	91.2%
SLCBAG	93.5%	93%	93.6%	93.3%

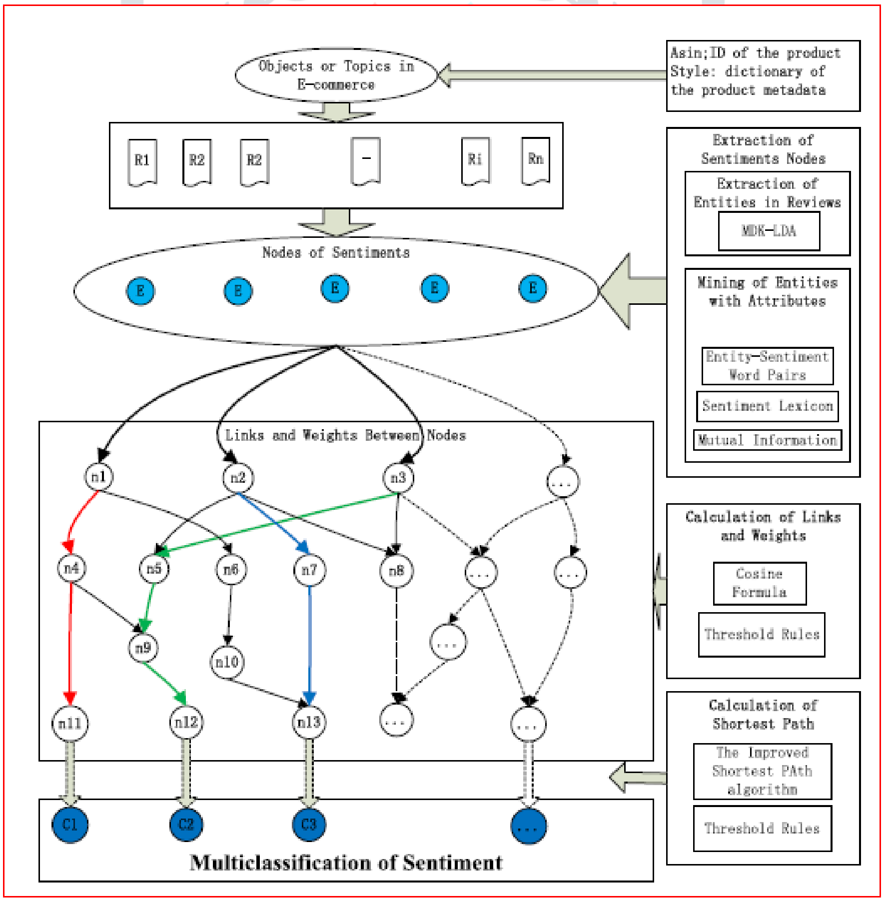
Table 1 Result comparison for research [2]

In research work [3], a novel sentiment analysis model named MBGCV is designed to improve the accuracy, MBGCV employs a multichannel paradigm and integrates Bidirectional Gated Recurrent Unit (BiGRU), Convolutional Neural Network (CNN) and Variation Information Bottleneck (VIB). They have used Chinese product reviews and the Chinese hotel reviews dataset with size of 21,107 records. Their model is better for imbalanced dataset. They have achieved 91% accuracy.

Researchers in work [4] explored diverse feature sets and classifiers for sentiment quantification. In addition, empirical performance analysis of conventional machine learning techniques, ensemble-based methods and state-of the-art deep learning algorithms on basis of features set, is performed. They have applied Expectation Maximization (EM) algorithm with CNN. 1.6 million Tweets dataset is used and achieved Absolute Error 0.26.

In research [5], researchers designed two equations to compute review helpfulness and review time scores, and they fine-tune BERT model to predict the review sentiment orientation probability. They also designed a formula to assign a numerical score to each review. Helpfulness score, time score and sentiment orientation score are calculated. Movie Review Dataset with 50,000 Reviews is tested. For their dataset they invited 32 users and 3 experts to assign a score to get more effective data.

In research [6], they proposed a fake review detection model by using Text Classification and



techniques related to Machine Learning. They applied Support Vector Machine, K-Nearest Neighbour, and logistic regression (SKL), using a bigram model that detects fraudulent reviews based on the number of pronouns, verbs, and sentiments. They used Yelp and TripAdvisor dataset and achieved 95% and 89.03% accuracy for their work.

Research work [7] proposes the construction of a directed weighted graph (ADG structure) based on some yielded information from FP-Growth frequent pattern identification algorithm on their corpus of Persian sentences. ELDA, SAM, an MMI-based and an LRT-based algorithms indicates the robustness of their approach. They achieved 91% accuracy in their work.

Research [8] consumer sentiment analysis is carried out using CNN. It is made automatically utilizing artificial intelligence approaches. According to the findings of a study on sentiment analysis on an E-Commerce-based web store for women, the apparels review dataset using the CNN. CNN method with the word vector generator and TF-IDF can produce a higher accuracy of 94%.

Paper	Advantages	Limitations
[1]	Achieved 74.70% Accuracy; better compare to BERT that is 73.36 % Also reduced CPU Time compare to BERT	No Feature wise Sentiment categorized. Single Source of Review may lead to false result. No Fake Review detection applied.
[2]	Achieved 93.50 % Accuracy; better compare to BiGRU that is 93.1 % Large Dataset is processed for Training	Only Book Reviews are processed. No Multiclass factor checked. Trimmed Reviewed to fasten the process.
[3]	Better model for Imbalanced Data. Achieved 91% Accuracy	No Feature based multiclass classification done. No Fake Reviews are measured.
[4]	Expectation Maximization (EM) algorithm is used.. CNN are trained for experimentation purposes. 1.6 million tweets dataset is used.	More training time. High Resource utilization
[5]	32 users and 3 experts were invited to assign a score to get more effective data. Highest average satisfaction scores given by both users and experts	Detection of Fake Reviews are missing. No Features or Multiclass approach applied
[6]	Classifiers: Support Vector Machine, K-Nearest Neighbour, and logistic regression (SKL), using a bigram model that detects fraudulent reviews based on the number of pronouns, verbs, and sentiments. Dataset: yelp , TripAdvisor	Result may be reduced in absence of Proper Pre-Labeled Training Data as all are supervised methods
[7]	Used FP-Growth Frequent Pattern for Graph Generation Achieved 91% accuracy	Polarity Detection and Sentiment based on Polarity can be applied
[8]	Higher accuracy of 94% Less Complexity	Require Manual Labelling for training dataset

Table 2 Related Works

2. METHODOLOGY

We are applying Fake Review Detection and apply the Score of Fake Probability to Weight for the Links. We also extracted Features of the Product those are mentioned in the review and analyze the Final Result based on Featured also.
Steps for Proposed Methodology:

Step 1: Read Amazon Review from Dataset
Step 2: Preprocess the Reviews (Remove Duplicate etc)
Step 3: Check Fake Review Score
Step 4: Extract Product Features mentioned in Review
Step 5: Extract Sentiment Nodes (Entity using MDK-LDA)
Step 6: Assign Attributes for Nodes
Step 7: Calculate Links between Nodes and Assign Weight:

- Cosine Value
- Threshold values
- Fake Review Score

Step 8: Calculate Shortest Path
Step 9: Multiclass Sentiment
Step 10: Feature Based Analysis

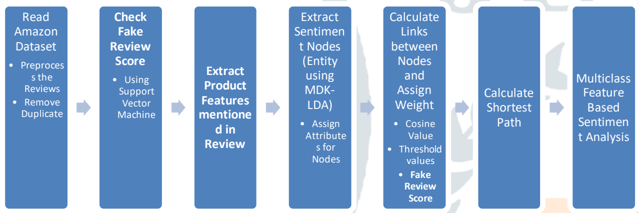


Figure 2 Proposed Steps

For our research work we have used amazon Phone Accessories dataset (2018). This dataset contains various features like :id, name, asins, brand, categories, keys, manufacturer, reviews.date, reviews.dateAdded, reviews.dateSeen, reviews.didPurchase, reviews.doRecommend, reviews.id, reviews.numHelpful, reviews.rating, reviews.sourceURLs, reviews.text, reviews.title, reviews.userCity, reviews.userProvince ,reviews.username.
This dataset contains 34660 reviews for Phone and Accessories.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 34660 entries, 0 to 34659
Data columns (total 21 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   id                    34660 non-null  object
 1   name                  27900 non-null  object
 2   asins                 34658 non-null  object
 3   brand                 34660 non-null  object
 4   categories            34660 non-null  object
 5   keys                  34660 non-null  object
 6   manufacturer          34660 non-null  object
 7   reviews.date          34621 non-null  object
 8   reviews.dateAdded     24039 non-null  object
 9   reviews.dateSeen      34660 non-null  object
 10  reviews.didPurchase   1 non-null      object
 11  reviews.doRecommend   34066 non-null  object
 12  reviews.id            1 non-null      float64
 13  reviews.numHelpful    34131 non-null  float64
 14  reviews.rating        34627 non-null  float64
 15  reviews.sourceURLs    34660 non-null  object
 16  reviews.text          34659 non-null  object
 17  reviews.title         34655 non-null  object
 18  reviews.userCity      0 non-null      float64
 19  reviews.userProvince  0 non-null      float64
 20  reviews.username      34658 non-null  object
dtypes: float64(5), object(16)
memory usage: 5.6+ MB
Number of Unique ASINs: 42
```

Figure 3 Dataset Details

3. RESULTS AND DISCUSSION

With our proposed methodology we have achieved 94% accuracy with Amazon. We also analyzed reviews based on various features of the Phone. That includes 'Battery','RAM','Display','Speed' etc.

.

	precision	recall	f1-score	support
	0.00	0.00	0.00	5
Negative	0.61	0.26	0.37	156
Neutral	0.44	0.12	0.19	292
Positive	0.95	0.99	0.97	6473
accuracy			0.94	6926
macro avg	0.50	0.34	0.38	6926
weighted avg	0.92	0.94	0.92	6926

Accuracy: 0.9393589373375686

4. CONCLUSION

After reviewing some of the research work for Sentimental Analysis of E-Commerce, we have founded that most of researchers have worked toward entire review sentiment, feature wise review is missing and also fake probability is missing. We have implemented the system that covers these points and it has improved the result in terms of Accuracy. Also fake review score generation helped for better understanding of the sentiments.

As a future work we can apply more features to detect Artificial Intelligence based fake review and sentiment based on various user classes.

REFERENCES

[1] S. Zhang, D. Zhang, H. Zhong and G. Wang, "A Multiclassification Model of Sentiment for E-Commerce Reviews," in IEEE Access, vol. 8, pp. 189513-189526, 2020, doi: 10.1109/ACCESS.2020.3031588.

[2] L. Yang, Y. Li, J. Wang and R. S. Sherratt, "Sentiment Analysis for E-Commerce Product Reviews in Chinese Based on Sentiment Lexicon and Deep Learning," in IEEE Access, vol. 8, pp. 23522-23530, 2020, doi: 10.1109/ACCESS.2020.2969854.

[3] T. Gu, G. Xu and J. Luo, "Sentiment Analysis via Deep Multichannel Neural Networks With Variational Information Bottleneck," in IEEE Access, vol. 8, pp. 121014-121021, 2020, doi: 10.1109/ACCESS.2020.3006569.

[4] K. Ayyub, S. Iqbal, E. U. Munir, M. W. Nisar and M. Abbasi, "Exploring Diverse Features for Sentiment Quantification Using Machine Learning Algorithms," in IEEE Access, vol. 8, pp. 142819-142831, 2020, doi: 10.1109/ACCESS.2020.3011202.

[5] A. Benlahbib and E. H. Nfaoui, "Aggregating Customer Review Attributes for Online Reputation Generation," in IEEE Access, vol. 8, pp. 96550-96564, 2020, doi: 10.1109/ACCESS.2020.2996805.

[6] H. Tufail, M. U. Ashraf, K. Alsubhi and H. M. Aljahdali, "The Effect of Fake Reviews on e-Commerce During and After Covid-19 Pandemic: SKL-Based Fake Reviews Detection," in IEEE Access, vol. 10, pp. 25555-25564, 2022, doi: 10.1109/ACCESS.2022.3152806.

[7] S. Jamshidi Nejad, F. Ahmadi-Abkenari and P. Bayat, "A Combination of Frequent Pattern Mining and Graph Traversal Approaches for Aspect Elicitation in Customer Reviews," in IEEE Access, vol. 8, pp. 151908-151925, 2020, doi: 10.1109/ACCESS.2020.3017486.

[8] Mehta, K., & Panda, S. P. (2022). Sentiment Analysis on E-Commerce Apparels using Convolutional Neural Network. International Journal of Computing, 21(2), 234-241. https://doi.org/10.47839/ijc.21.2.2592

[9] Deniz, E.; Erbay, H.; Coşar, M. Multi-Label Classification of E-Commerce Customer Reviews via Machine Learning. Axioms 2022, 11, 436. https://doi.org/10.3390/axioms11090436

[10] Wassan S, Shen T, Xi C, Gulati K, Vasan D, Suhail B. Customer Experience towards the Product during a Coronavirus Outbreak. Behav Neurol. 2022 Feb 2;2022:4279346. doi: 10.1155/2022/4279346. PMID: 35287287; PMCID: PMC8917442.

[11] M. Bouazizi and T. Ohtsuki, "Multi-class sentiment analysis on twitter: Classification performance and challenges," in Big Data Mining and Analytics, vol. 2, no. 3, pp. 181-194, September 2019, doi: 10.26599/BDMA.2019.9020002.

[12] “Dataset” https://www.kaggle.com/datasets/grikomsn/amazon-cell-phones-reviews