Recommender System for Sentiment Analysis using Machine Learning Models

A. Naresh[a], P. Venkata Krishna[b]

[a] Department of Computer Science, Bharathiar University, Tamilnadu, India
[b] Department of Computer Science, Sri Padmavati Mahila Visvavidyalayam, Tirupati, India
[a] pandu5188@gmail.com, [b] pvk@spmvv.ac.ins

Article History: Received: 10 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 28 April 2021

Abstract: In recent years, with the rapid growth of Internet technology, online shopping has become a rapid way for users to purchase and consume desired products. Large volume of user-generated content on social media sites like twitter has resulted in tweet sentiment analysis. Sentiment analysis supports as base for decision support systems and recommendation systems and it becomes an essential tool on online platforms to extract the information on user emotional state to improve user satisfaction. This paper proposes an effective sentiment analysis recommender system framework using machine learning models.

Keywords: Machine Learning, Sentiment Analysis, Recommender System, Twee

1. Introduction

Recommender Systems are software tools and practices that deliver recommendations for various to a user [1-2]. The suggestions provided by recommender systems will support users to take decisions regarding the products or items. Recommender systems are used to build recommendations by processing information from actively gathered varied kind of data. The data that is used for processing information depended upon the type of recommender systems [3-4]. These recommender structures play a major role in offering recommendations about the products or items, where we can utilize the widely available information about the products or items or anything through Online Social Networks (OSN). Automatic Recommender system on the cloud can recommend products or items by providing the opinions or recommendations about the product based on the queries of the user; this is provided through the cloud platform. The cloud platform is a very good stage for people to convey their views and feelings towards the products and items. The Online Social Network such as Twitter, Facebook will act as a good cloud platform for the recommender system.

Sentiment analysis encourages the customers to find perspectives on others and furthermore helps in choosing the buy with the opinion of different customers. Highlight improvement in products is additionally made possible by assessment mining to satisfy the necessities of advertising knowledge and product benchmarking in the industrial business. The Figure 1 shows the sentiment analysis process.

At first step the product reviews are are cleaned for eliminating the stop word and uncommon character. Then, sentiments are recognized for choosing the essential features which helps in classifying the positive and negative estimations. Finally, with respect to opinion extremity as positive or negative, better decision making was acted in the buying of products.

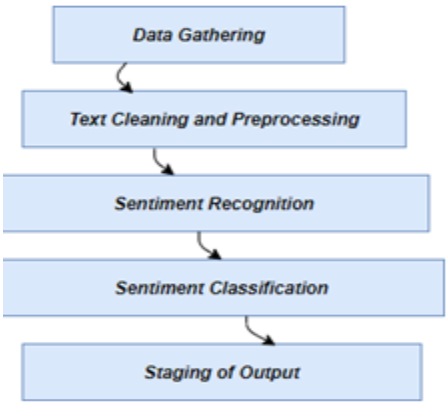

Fig. 1. Sentiment Analysis process

Sentiment analysis stages:

A. Data collection and pre-processing

In data collection and preprocessing stage, the acquired content is inspected for recognizing the opinions. It is significantly dependent on the individual method utilized, to the words which express sentiments are taken out. At that point, preprocessing is performed for eliminating the irrelevant opinions. It is significant for mining keywords from the content which offered an accurate classification.

B. Classification

In classification stage, identification of content polarity is performed. Generally, three classes are utilized for order, for example, positive, negative, and neutral. Furthermore, Classification calculations are additionally used in sentiment analysis, in view of the administered or unaided strategies with pre checked models. It is basic for preparing the model utilized for classification with domain-specific data.

C. Aggregation and presentation of results

At aggregation and presentation stage, the gained assessment classification measure result is gone through aggregation measure after specific calculations for assessing the basic assessment of the researched text. Subsequently, introduction is completed straightforwardly for speaking to the sentiment textual words.

The machine learning models are very useful to classify the collected information and predict the polarity of sentiment as either positive or negative or neutral. In this article we are giving an overview of the framework that is used for the recommender system based on machine learning models for sentiment analysis. This article also provides an overview of different machine learning models chosen for sentiment classification based on supervised and unsupervised learning techniques.

In this article, section 2 revels the related work, section 3 discussed the architecture of proposed framework, section 4 discussed various machine learning models for sentiment analysis, section 5 revels expected results and finally conclusion ends this paper.

2. Related Work

Preethi et al in 2017 [5] proposed a recommender approach on the cloud by applying deep learning (DL) concept to sentiment analysis called RNN-based Deep-learning Sentiment Analysis (RDSA) framework which was built on ‘Recommend Me’ App. This proposed framework was built based on the Learning Automata based Sentiment Analysis (LASA) framework.

In [6], proposed a Bayesian-Inference founded recommendation structure for online social groups. This recommendation provides better results than trust load recommendation and collaborative filtering recommendation. This recommender system uses conditional probability for calculating rating similarity, where content ratings are shared with friends. The main contribution of the proposed approach describes a Bayesian-inference founded recommendation structure with an accurate and personalized recommendation by leveraging implanted social composition inside a social network.

In [7], proposed a sentiment analysis-based recommender system that provides product rating using sentiment analysis. This system provides more accurate results by comparing with various classification methods such as Naïve Bayes and

SVM on extracted tweets from twitter. The main contribution of this approach includes the adoption of dual prediction technique and SVM text classifier. The experimental evaluation was performed on twitter data provides performance accuracy up to 82 percent.

In [8], proposed a sentiment analysis-based recommender system that provides improved recommendation using faith and sentiment implication from online social networks. This paper proposes a framework of implicit social trust and sentiment-based recommender system. This framework was verified by experimenting with real-time social data from Twitter. This recommender system uses machine learning methods such as naïve bayes, logistic regression, and decision trees.

In [8], proposed a recommender system that uses opinions or sentiment of users regarding the services provided by E-shopping website. For classifying the sentiments provided by the user they used stochastic learning algorithm. This recommender system will find out and does not allow fake reviews by using the MAC address of the system. The main contribution of this framework includes the implementation of stochastic learning and context-based engine to overcome the drawbacks of traditional recommender system and for improved performance.

The literature Survey on sentiment analysis- based recommender systems concludes, most of the frameworks that were evaluated on movie domain provide better performance results compared to another domain. The RDSA framework [5] shows the highest performance accuracy among the studied sentiment analysis based recommender systems. The experimental assessment of the recommended framework was done by applying the recursive neural network for sentiment categorization in movie and restaurant reviews. The evaluation of the framework was not observed in real-time data and short text or micro-blog textual data.

Use of machine learning (ML) models to sentiment examination has been famous these days. The motivation to pick ML models, as it gives improved execution and exact outcomes over learning assignments. ML strategies will perform both component extraction and order for record and short content.

In [9], developed a new model, using lexicon- based techniques and ML strategies together. As a naive strategy for opinion examination, the

Foundation of each word in a sentence takes a score from a word reference and the last extremity score of the significant sentence is determined by utilizing added substance score-based models. ML models are prepared to perform precise assessment explanations by utilizing highlights dependent on extremity scores of writings. The final supervised model can accomplish sentiment comments of new Turkish writings inside a 73% achievement rate with no human intercession. The dataset was built by utilizing online business site (www.hepsiburada.com).

In [10], the literary cry surveys of organizations are analyzed to allocate a probability for the review as having positive or negative estimation. The information considered for the estimation investigation are the surveys on eateries about food, administration, cost, and feel. ML approaches in the nltk library of python can end up being valuable in any such examination on Natural Language Processing and the library has been utilized broadly in this work. Every calculation utilized has been compared based on their efficiency.

In [11], proposes utilization of supposition examination characterization as a powerful strategy for analyzing literary information originating from assets of assets on web. Sentiment analysis is a strategy for information mining that assesses literary information expending AI strategies. Because of tremendous scope of assessments of clients, their surveys, feedback, and recommendations accessible over the web assets, it is such a great amount of irreplaceable to find, analyze and combine their perspectives for improved decision making. Sentiment analysis presents a compelling and proficient assessment of customers continuously which can enormously influence the dynamic cycle for business area.

3. Proposed Work

Sentiment analysis technique serves as a decision support system for the recommendation system. The goal of combining recommender system and sentiment analysis provides the most accurate results to the users. Sentiment analysis uses different levels and types of techniques used for feature extraction and methods that are used for classification. These techniques and methods provide the resultant user query with the negative or positive, sometimes positive or negative or neutral sentiment so that the recommender system can make decision-based on the sentiment analysis procedure response.

Recommender systems support the users to make decisions regarding their preferences about the product or item. Sentiment analysis works an essential role in building a recommender scheme. The fundamental idea behind this framework is to collect information from nodes. Apply sentiment analysis technique on the collected information using different machine learning styles such as support vector machine, k-nearest neighbor, decision tree etc. will works well to classify the information in recommender system. The proposed framework that is used for recommender structure built on machine learning models for sentiment analysis is shown in figure 2.

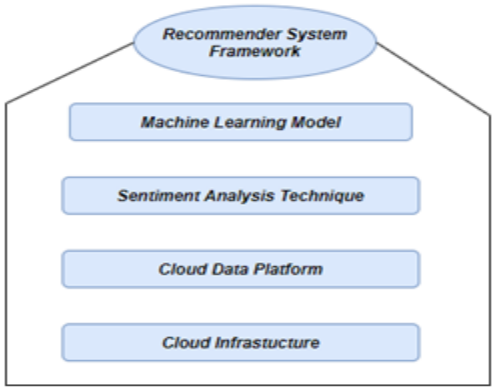

Fig 2: Proposed Recommender System Framework

4. Machine Learning Models Machine Learning Includes The Principles Of

Statistics, probability, optimization, and control theory. Large numbers of machine learning algorithms are deployed in practice that allows computers to advance actions based on practical data [12-13]. Accuracy, training time, number of parameters and features are critical while choosing a machine learning algorithm. The implicit knowledge acquired by the training data of machine knowledge algorithms can be effectively used for the assessment of the unseen data. Performance of the model is evaluated by various measures such as accuracy, precision, and recall. In the broad spectrum of data analysis, a snapshot of various machine learning algorithms is discussed here.

ML based models are widely applied in the application areas such as vision processing,

forecasting, pattern recognition, robotics, and expert system. Machine learning can be categorized in to supervised and unsupervised learning methods. Supervised learning utilizes a known dataset called „training dataset" to make forecasts. The training dataset contains input data and class label to construct a model that can make forecasts of the answer values for a new dataset. A test dataset is often utilized to authenticate the model. Classification (discrete labels) and regression (real values) are the two categories of supervised learning.

This section gives the outline of the supervised machine learning technique chosen for the sentiment classification using different machine learning models for recommender systems.

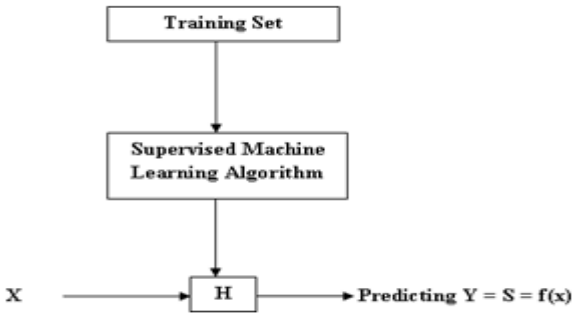

Fig. 3. Supervised Learning

Using the inputs to predict the values of output is called supervised learning. The sentiment classification technique used in this chapter was a supervised machine learning technique. The approach to perform a supervised machine learning technique was followed as shown in figure 4. For illustrations consider a training data set as the set of documents with provided sentiments as follow:

Table 1: Training Data Set

Documents D	Sentiment S {0, 1}
Document D1	0
Document D2	1
:	:
Document Dn	1

Table 2 Test Data Set

Documents D X	Sentiment S Y- {0, 1}
Document D1 Document D2 : Document Dn	? ? : ?

The training set undergoes processing centered on the supervised machine learning process, it provides features or attributes of the document Di. The aim of the supervised learning is, for the offered training set to learn a function H: X Y. Here X is an input document based on the query and

Y is to predict the sentiment of the provided input documents where H is the hypothesis. The supervised learning algorithm is used for regression and classification.

The supervised learning technique predicts the polarity Y also represented as sentiment variable S of the target data or test data based on the training dataset X with provided hypothesis H of a finite set of classes such as negative or positive. Based on the query of the input data X, the polarity of the provided input was described as either negative or positive with the support of training records. The terminology used in the supervised learning approach is described as follows:

X – Input data based on the query H – Hypothesis

Y= S = f(X) = {0, 1}

To achieve the hypothesis of the predicted sentiment of the given input data, this chapter presents an outline of the functioning of machine learning models.

5. Result Discussion

Data is gathered from the twitter data source by utilizing application programming interface (API). It is the user interface among both clients and sources. Here the source means a twitter site, which contains a collection of client tweets. For implementation, an airline dataset is taken. The observations contain multiple attributes like tweet- id, sentiment, remarks, creation time, area, and so on. The airline dataset was taken as input. 1200 tweets are considered in this work.

The gathered tweets are preprocessed and arranged into three classes as positive, negative and neutral. Three basic machine learning models K- nearest neighbor (KNN), support vector machine (SVM) and decision tree (DT). The performance of these models is compared using various measures like accuracy, precision, recall and f-measure. Table 3 shows the classifies performance on airline dataset and table 4 refers classification of tweets. Figure 4 shows the graphical representation of tweets classification.

Table 3. ML classifies performance on airline dataset

Classifiers	Accuracy (%)	Precision	Recall	F-Measure
KNN	67	70.5	69.3	67.9
SVM	68	69	68.1	68.7
DT	80	81.4	81.4	80.95

Table 4: Classification of tweets

Sentiment class	Total tweets	Classified
Positive	1200	560
Negative	1200	362
Neutral	1200	278

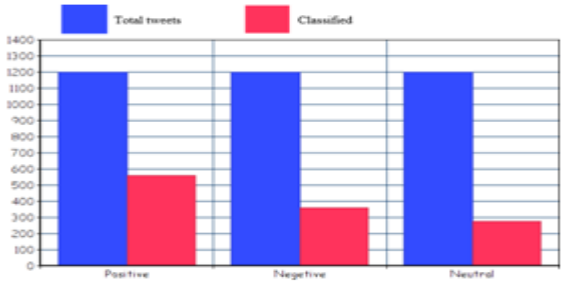

Fig 4. Graphical representation of tweets classification

6. Conclusion

In this article machine learning based recommender system is proposed for sentiment analysis. An airline tweets data set is considered here, and three basic supervised ML techniques are applied and finally the performance of same is assessed in terms of accuracy, precision, recall and f-measure. Based on the chosen machine learning models, the framework was designed based on supervised learning technique for sentiment analysis with recommender systems.

References

1. R Francesco, L Rokach, and B Shapira, “Introduction to recommender systems handbook”. Springer US, 2011, pp 1-35.
2. K. McCarthy, M. Salam´o, L. Coyle, L. McGinty, B. Smyth and P. Nixon, “Group recommender systems: a critiquing based approach". IUI '06: Proceedings of the 11th international conference on intelligent user interfaces, New York, NY, USA, ACM, 2006, pp. 267– 269.
3. M.C. Puerta Melguizo, L. Boves, A. Deshpande and
4. O.M. Ramos, “A proactive recommendation system for writing: helping without disrupting”, Proceedings of the 14th European conference on Cognitive ergonomics, 2007, pp. 89–95.
5. R. Burke, “Hybrid web recommender systems”, The Adaptive Web, Springer Berlin Heidelberg, 2007, pp. 377–408.
6. G. Preethi, P.V. Krishna, S. Obaidat Mohammad, V. Saritha and Y. Sumanth, “ Application of Deep Learning to Sentiment Analysis for recommender system on cloud”, International Conference on Computer, Information and Telecommunication Systems (CITS) , 2017, pp. 93-97.
7. X. Yang, Y. Guo and Y. Liu, "Bayesian-Inference-Based Recommendation in Online Social Networks, " IEEE Transactions on Parallel and Distributed Systems, vol. 24, no. 4, 2013, pp. 642-651.
8. M. P. Anto, M. Antony, K. M. Muhsina, N. Johny, V. James and A. Wilson, "Product rating using sentiment analysis," International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), 2016, pp. 3458-3462.

A. Dimah and Z. Xiao-Jun, “Improving Recommendation Using Trust and Sentiment Inference from OSNs”, International Journal of Knowledge Engineering-IACSIT, 2015, pp. 9-17.

9. Rumelli, Merve, et al. "Sentiment Analysis in Turkish Text with Machine Learning Algorithms." 2019 Innovations in Intelligent Systems and Applications Conference (ASYU). IEEE.
10. Hemalatha, S., and Ramathmika Ramathmika. "Sentiment Analysis of Yelp Reviews by Machine Learning." 2019 International Conference on Intelligent Computing and Control Systems (ICCS). IEEE, 2019.
11. Chaturvedi, Saumya, Vimal Mishra, and Nitin Mishra. "Sentiment analysis using machine learning for business intelligence." 2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI). IEEE, 2017.
12. Jason, B. "A tour of machine learning algorithms." Journal of Intelligent Learning Systems and Applications 9.4 (2013): 50-62.
13. Jonnavithula, S. K., Jha, A. K., Kavitha, M., & Srinivasulu, S. (2020, October). Role of machine learning algorithms over heart diseases prediction. In AIP Conference Proceedings (Vol. 2292, No. 1, p. 040013). AIP Publishing LLC.