

Decision Making on Customer Review using Sentiment Analysis

Santhosh Kumar K L, Jharna Majumdar

Abstract—The sentiment analysis is especially required in web based business sites, moreover profitable with person. A consistently expanding measure of results are put away in the web and also the measure of individuals would procuring things from online are expanding. Subsequently, the customers' opinions are expanding every day. The opinions toward seller provides their inclination. The opinions related to products on the Web, which are useful for both manufacturers and customers. The way toward discovering customer opinion about the item or issue is called as opinion mining. Dissecting the feelings from the separated opinions is characterized as Sentiment Analysis. The objective of opinion mining and Sentiment Analysis is to make system ready to perceive and express feeling about the given context. This work focuses on mining opinions from the sites like flipkart, which enables customer to unreservedly compose the view. It consequently extricates the opinions from the site. It likewise uses algorithms CART, Random Forest and J48 classifier to assign the reviews as negative and positive. At the end we have utilized quality parameters to gauge the level of all algorithms.

Keywords: CART, Random Forest, J48, Sentiment Analysis, Quality metrics

1. INTRODUCTION

Advances in Data Mining and Data Classification techniques has led to the re-research on Web Data Mining, enabling one to extract useful knowledge requires for our day to day needs using the world Wide Web.

Online shopping is an E-commerce application, which is a way of purchasing items (goods of regular needs) required in our day to day life or could also be purchasing services. In both the cases, however, this is done from the available information on merchandizing over the web using various browsers or search engines. Customer can look through their desired items from alternate sites by checking out on the varieties of web page of online business given by different sellers and at the same item, its accessibility and cost. Such information will possess some specific characteristics:

- (a) Variations in value, sometime small or large
- (b) Selection of the desired item
- (c) Finding accessibility from various sellers.

This method of purchasing items enables the common man knowledge of their experience towards the usage of items from E-business sites. Usually, we call them as opinion or review. It intends to decide the mind-set of the customer; which might be positive, negative or neutral towards the item. These constructive or destructive feelings

communicated by the general population are known as 'Sentiment'.

Customers can consider opinions of other purchasers while purchasing an item. This also allows him to focus on couple of suggestions and collaborate with various search engines.

Distinctive sites permit diverse review structure to be taken care. A few sites enable customer to unequivocally compose the advantageous and disadvantageous, at times alongside the rundown, in different cases there won't be any confinements on customer to compose review; with the goal that they can compose anyway they want and express their sentiments.

In the current research work, we propose the review of Laptop from Flipkart, a web based shopping site. We extract information and store them in CSV document. Using various algorithms, the reviews are identified as positive or negative. Finally the performance of each algorithm is analyzed and the results presented.

2. RELATED WORK

In this section, we will be discussing on the current research work towards the opinion mining and sentiment analysis. In [1], focuses tools and strategies utilized as a part of opinion mining. The procedure of opinion summarization outline has three primary steps, for example, "Opinion Retrieval, Opinion Classification and Opinion Summarization". User remarks are recovered from opinion sites. These remarks contain subjective data with positive or negative opinion.

In [2], centers around sentiment analysis on Amazon site. Clients of the web based shopping website Amazon are urged to post surveys of the items that they buy. Amazon utilizes a 1 to 5 scale for the items, paying little mind to their category, and it winds up testing to decide the points of interest and drawbacks to various parts of an item. In [3], concentrates just on the surveys taken from Amazon site examine the outcomes utilizing three unique methods, for example, Senti-Word-Net, Logistic-Regression and Naive-Bayes.

In [4], provides the experimental tool of Decision making towards the Sentiment analysis on Humanoid-robot with Human Robot Interaction.

In [5], centers around investigating audits from various Shopping Websites. The fundamental focal point of the framework is to investigate surveys for web based shopping

Revised Manuscript Received on April 05, 2019.

Santhosh Kumar K L, Assistant Professor, Department of CSE, Nitte Meenakshi Institute of Technology, Bengaluru, Karnataka, India

Jharna Majumdar, Dean R&D, Professor, Department of CSE, Nitte Meenakshi Institute of Technology, Bengaluru, Karnataka, India



administrations. The re-sees are grouped by positive, negative and nonpartisan. These outcomes help the user so that they can look into the specific shopping websites with appropriate positive and negative audits and rating. It utilizes "SentiWordNet lexicon" for discovering score of every word. At that point slants are delegated positive, negative and unbiased.

In [6], centers around Flipkart surveys utilizing calculations Naïve-Bayes and Decision-Tree. Utilizing clients surveys about item and audit about sellers in Flipkart as dataset and characterizes survey by negative or positive conclusion of purchaser. Such surveys are useful to some degree, promising both the customers and items creators. It introduces an observational investigation of viability of grouping item audit by semantic importance. Ordering remarks utilizes cross breed calculation joining Decision Trees and Naïve Bayes calculation.

In [7], focus on variety of Clustering methods on Amazon posts. Clients of internet shopping webpage like Amazon, indicating a huge number of client remarks sooner than buying item might be a testing task. Utilizing machine learning strategies for break down the pre-processed information from web based shopping sites to supply purchasers with improved clients information sooner than purchasing item. K means and Peak looking bunching calculations is to perform grouping of item on Amazon audits.

In [8] ROCK method is a progressive clustering system and it identifies with gathering of agglomerative various leveled grouping calculations. ROCK method arbitrarily takes test dataset recovered by database. Various leveled clustering method utilizes connections to the focuses, groups are requires inspected focuses; those focuses are used to assign left over information indicates on database legitimate clusters. In [9], overview focuses on decision tree methods for order in information mining. The most well known characterization strategy is decision tree strategies. The essential learning methodology separate and overcome procedure is utilized in choice tree. A choice trees are like tree structure, each understood hub speaks to property, each branch speaks to a yield of the test, and class mark is spoken to by each leaf hub. Methods of various kinds of trees are ID3, C4.5 and CART [10].

In [11], Classification is a procedure to mastermind the information into the ensured class-base on a couple of criteria. Grouping is called as regulated learning, in light of the occasion square measure gave with class name, a capability to unsupervised getting the hang of in the midst of which the name square measure is unidentified. In [12], the online audits and appraisals are upgraded in view of this it will be pre-carious for a few; clients will isolate the fundamental assessments which can be procured from the not significant ones.

3. THE PROPOSED METHOD

The research work proposed in this paper deals with web mining application, where customer review about a product is required. The outcome of this research will help the buyers to decide whether a product is good or bad. With the advancement of current technology, a large number of websites are available which gives customers' feedback in

terms of 'pros' and 'cons' about a product. This gives enormous ad-vantage for a customer before he/she decides the suitability of the desired product for purchase. The proposed research will overcome the drawback of the system of checking all reviews and will be able to automatically integrate the reviews of customers and give the appropriate decision on that. Current work uses classification algorithms such as CART, Random Forest and J48 along with various other mechanisms. Detailed work flow diagram of the proposed system is given below in Fig. 1.

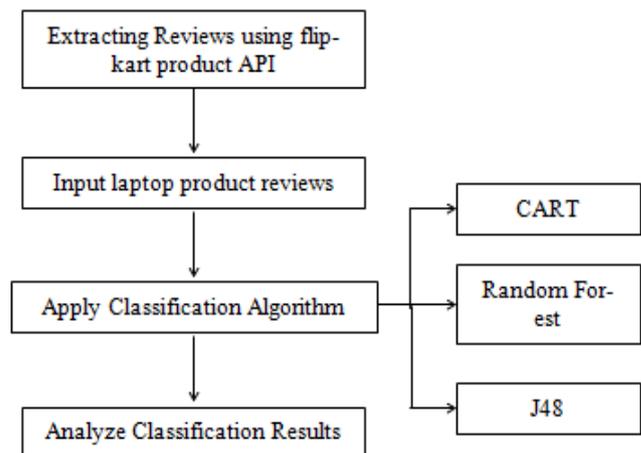


Fig. 1. Flowchart of the proposed system

4. CLASSIFICATION METHODOLOGIES FOR SENTIMENT ANALYSIS

4.1 CART

CART (Classification and Regression Tree) [10, 13] is a sort of decision tree method. In this paper, we have utilized CART grouping method is connected for arranging the audits into positive or negative words with the assistance of accessible lexicon words. The Gini Index is the parameter used to build the decision tree in CART [16].

4.2 RANDOM FOREST

Random forest is a decision tree; it is mostly used in classification and regression problems. This will prepare model which will make predictions of the value of a targeted variable through learning simple decision rules concluded by data structures. In classification problems, it comprises of an arbitrary number of random trees which can be used to find out the last outcome [14].

4.3 J48

The part criteria of J48 [15] are improved the situation both numeric and character information. The numerical esteem can part dependent on the paired split. The character information can part to double or any number of parts. The extra component for J48 is for the last esteem, choice trees prune, the starting point of guidelines, and so on.

5. RESULTS AND ANALYSIS

The performance of the classification strategies can be discovered by utilizing a portion of the accompanying parameters.

5.1 Recall

The Recall is defined as the ratio of correctly classified items over the sum of correctly classified along with missed classified items.

$$\text{Recall} = \frac{\text{correctly classified}}{\text{correctly classified} + \text{Missed classified}} \quad (1)$$

5.2 Precision

The Precision is defined as the ratio of correctly classified item over the total number of experiments conducted with correctly classified and erroneous classified items.

$$\text{Precision} = \frac{\text{correctly classified}}{\text{correctly classified} + \text{Errorly classified}} \quad (2)$$

Note: For a good classification, value of Recall should be higher than the Precision.

From Table 1 to 3, we can see the example execution measure esteems got for three distinct calculations for three unique items, for example, Dell Inspiron, HP and Lenovo. Fig. 2 demonstrates the graphical portrayal of by and large examination of execution measures got from the forbidden qualities. Among three calculations CART has relatively better outcomes over the Random Forest and J48.

Table.1. Performance Measure on – Dell Inspiron Laptop

Classifier	Recall	Precision
CART	0.841	0.357
Random Forest	0.826	0.567
J48	0.510	0.420

Table.2. Performance Measure on - HP Laptop

Classifier	Recall	Precision
CART	0.745	0.477
Random Forest	0.711	0.545
J48	0.607	0.711

Table.3. Performance Measure on - Lenovo Laptop

Classifier	Recall	Precision
CART	0.869	0.403
Random Forest	0.791	0.790
J48	0.490	0.720

6. CONCLUSION AND FUTURE WORK

As indicated in our analysis, the CART turns out to be the most efficient algorithm among the three classification algorithms for sentiment mining. The present work concentrates just on the reviews taken from ‘Flipkart site’ utilizing three different methodologies.

Considering the importance of this research in today’s world, the work can be ex-tended further considering different web sites available for the same kind of product. The other classification techniques like Artificial Neural Network, Support Vector Machine or some other well-known classification technique can be used and study their performance for this application.

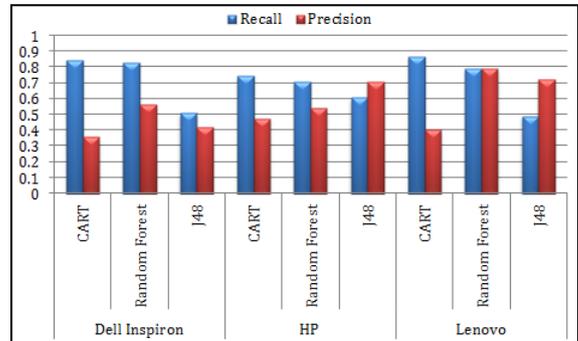


Fig. 2. Analysis of Performance measures on Sentiment classification techniques.

REFERENCES

- G.Angulakshmi, R.ManickaChezian, “An Analysis on Opinion Mining: Techniques and Tools.” International Journal of Advanced Research in Computer and Communication Engineering, Vol. 3, Issue 7, July 2014
- Callen Rain, “Sentiment Analysis in Amazon Reviews Using Probabilistic Machine Learning”, Swarthmore College Computer Society, November 2013
- Santhosh Kumar K L, Jayanti Desai, Jharna Majumdar “Opinion Mining and Sentiment Analysis on Online Customer Review”, 2016 IEEE International Conference on Computational Intelligence and Computing Research at Agni College of Technology, Chennai during December 15th to 17th 2016
- Gurunath H Naragund, Santhosh Kumar K L, Jharna Majumdar “Development of Decision Making and Analysis on Customer Reviews using Sentiment Dictionary for Human-Robot Interaction”, International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE), Volume 4, Issue 8, August 2015
- U Ravi Babu, Narsimha Reddy, “Sentiment Analysis of Reviews for E-Shopping Websites”, International Journal of Engineering And Computer Science, ISSN: 2319-7242 Volume 6 Issue 1 January. 2017
- Gurneet Kaur, Abhinash Singla, “Sentimental Analysis of Flipkart reviews using NaïveBayes and Decision Tree algorithm”, International Journal of Advanced Researching Computer Engineering & Technology (IJARCET) Volume5, January 2016
- Chantal Fry, Sukanya Manna, “Can we Group Similar Amazon Reviews: A Case Study with Different Clustering Algorithms”, IEEE Tenth International Conference on Semantic Computing 2016
- Sudipto Guha, Rajeev Rastogi, Kyuseok Shim, “ROCK: A Robust Clustering Algorithm for Categorical Attributes”, International Journal of Science, Engineering and Technology Research (IJSETR), 2015



9. Delveen Luqman Abd AL-Nabi, Shereen Shukri Ahmed, "Survey on Classification Algorithms for Data Mining: (Comparison and Evaluation)", Computer Engineering and Intelligent Systems (IISTE) Vol.4 , 2013
10. Himani Sharma, Sunil Kumar, "A Survey on Decision Tree Algorithms of Classification in Data Mining", International Journal of Science and Research (IJSR) 2015
11. Sunita Joshi, Bhuwaneshwari Pandey, Nitin Joshi "Comparative analysis of Naive Bayes and J48 Classification Algorithms", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 12, December 2015
12. Betul Dunder, Suat Ozdemir, Diyar Akay "Opinion Mining and Fuzzy Quantification in Hotel Reviews", IEEE 2016 International Symposium on Networks, Computers and Communications (ISNCC), 11-13 May 2016