



Sentiment Analysis of Seller Reviews

Vidya Zope¹, Samarth Ladia², Vishal H Talreja³, Chirag Kripalani⁴, Sunny Ramrakhiani⁵
Mentor¹, Student^{2,3,4,5}
VESIT, Mumbai, India

Abstract:

Online transactions have seen a rapid growth in the last few years and consumer reviews have become an important part of this growing business. Seller Trust Profile can be computed using various models. These models make use of the feedback reviews which consist of ratings as well as comments given by the user. Models that compute seller trust score based on user feedback overlook the important details in the text and give inaccurate trust scores based on user ratings. We propose a novel model for computing seller trust profile based on the user reviews and we would recommend this trust profile to the user.

I. INTRODUCTION

The Internet offers a global platform for Ecommerce, communication, and opinion sharing. It has many blogs dedicated to topics like finance, travel, sports, entertainment, history, environment, business etc. on which people express their opinions in terms of feedbacks and comments. Mining through these terabytes of user reviews is a demanding task. Opinion mining has many useful applications. Generating user-query specific summaries is also an interesting application of opinion mining. The main focus is efficient feature extraction, sentiment polarity classification, and summary generation of online reviews and estimating a seller trust profile out of those reviews.. Nowadays, several websites are available on which several products are advertised and sold. Before making a purchase an online user typically browses through many other products of different brands and different sellers before reaching a final decision. This apparently simple information retrieval task actually requires a lot of comparison, since all sellers sell similar features and prices for most products. Most online e-commerce sites allow users to post feedbacks of products purchased. There are also devoted sites that post reviews by experts and end users. There are also online shopping sites which already have separated seller and product reviews. These seller reviews if appropriately classified and summarized can play an instrumental role in influencing a buyer's decision. The main difficulty in analyzing these online users' reviews is that they are in the form of a language that has developed naturally in use. Natural language processing is difficult; analyzing online unstructured textual reviews is even harder. Some of the major problems with processing unstructured text are dealing with spelling mistakes, incorrect punctuation, and irrelevant words, use of non-dictionary words or slang terms, and undefined abbreviations. So, the task of summarizing noisy, unstructured online reviews demands extensive Preprocessing. And as in this paper we will be analysing the reviews using sentiment analysis algorithm which will be using Naive bayes and NLP processors to get the true sentiment of the review and generate the seller trustscore which will allow users to judge the product on product reviews irrespective of seller's performance and can choose the best seller for the product order using our seller trust score.

II. REQUIREMENTS

For defining the requirements of the project capabilities, features or attributes of the project are required. Requirements are then decided to determine which requirements will be included and excluded from the project. Depending on that requirements are classified into two categories Functional and non-functional requirements.

A. Functional Requirements

A functional requirement document defines the functionality of a system or one of its subsystems. It also depends upon the, expected users, the type of software and type of system where the software is used. The functional requirements of our project are as follows:

i. Login

User should be able to login into his/her E-commerce account

ii. Network Connection

The application should automatically connect to network.

iii. Retrieve Reviews

The application should retrieve reviews from website. A set list of keywords will be made.

iv. Computation

Computing seller trust profile based on fine grained analysis of feedback comments.

v. Comparison

Unique methodology to compare reviews and draw conclusions. The computed trust scores for users are allotted based on trust and performance.

vi. Ratings

Based on the score of the user, the reviews are rated. The reviews will be scanned and based on the number of matches with the list of key words, they will be rated accordingly.

vii. Recommendation

The higher rated reviews from users with higher scores will feature above the others, thereby providing other users with trusted and reliable help.

B. Non-Functional Requirements

A non-Functional requirement is a software testing technique that verifies the attributes of the system such as memory leaks, performance or robustness of the system.

i. Well defined Classification

Distinct well defined sentiment classification.

ii. Word Extraction

Opinionated word extraction from reviews should be retrieved as quickly as possible.

iii. Security

The trust score should be secure.

iv. GUI

Good user interface for enhanced user experience. System should be as automated as possible.

III. CLASSIFICATION

Classification is a function that assigns items in a collection to target categories or classes. The goal of classification is to accurately predict the target class for each case in the data. Our application aims at Multi-labelled classification of keywords using supervised machine learning algorithms. Keywords are classified into predefined classes: Positive, Negative and Neutral.

A. Positive Reviews

Positive reviews refer to the ones with positive keywords. These are the desirable contents that help rate the seller higher. These are further classified into three types:

i. Single Pure Positive

These include single word positive keywords like ‘good’, ‘impressive’, ‘nice’, etc. which are basically positive adjectives.

ii. Multiple Pure Positive

These include multiple words asserting a positive tone. It includes, but is not restricted to, the use of adverbs and adjectives. For example, ‘very good’, ‘extremely nice’, ‘super awesome’, etc.

iii. Double Negative

These include phrases with multiple negative keywords used together to imply a positive a tone. Eg. ‘Not bad’

B. Negative Reviews

Negative reviews refer to the ones with negative keywords. These are non-desirable contents that help rate the seller lower.. These are further classified into two types:

i. Single Pure Negative

These include single word negative keywords like ‘pathetic’, ‘bad’, ‘useless’, etc. which are basically positive adjectives.

ii. Multiple Pure Negative

These include multiple words asserting a positive tone. It includes, but is not restricted to, the use of adverbs and adjectives. For example, ‘very bad’, ‘extremely horrible’, etc. Neutral reviews include the irrelevant reviews or simply the ones that are not applicable to our system. It has the following types:

C. Neutral Reviews

Neutral reviews include the irrelevant reviews or simply the ones that are not applicable to our system. It has the following types:

i. Non Informative

Comments with content that is not relevant to a particular seller. Eg. If a review speaks about mobile phones in a review of a home appliances seller.

ii. Garbage and Spam

Seller reviews which contain non-comprehensible content, and garbage values. Spam refers to a variety of prohibited behaviours that violate the Website rules. Spam can be generally described as unsolicited actions that negatively impact other users. This includes many forms of automated account interactions as well as attempts to mislead users.

iii. Non Biased

Relevant reviews with comprehensible content but which does not help in the classification.

IV. COMPARISON OF ALGORITHMS

The comparisons of various algorithms that we will use in the system are as follows:

Table.1. Comparison of Machine learning classification algorithms

Parameter	NB	SVM	KNN
Complexity	Very Simple	Complex	Moderately Complex
Memory	Minimum	Memory Intensive	Memory Intensive
Features	Independent feature Improves performance	Can perform better with dependent features	Can perform good with both
Decision Boundary	Linear/parabolic/elliptic	Any	Any
Prediction Speed	Fast	Moderate	Slow
Training Speed	Fast	Moderate	Slow

V. TOOLS AND TECHNOLOGIES USED

A. Technologies

The various technologies that will be requiring in the system are as follows:

i. R language:

R is a language and environment for statistical computing and graphics. R is an integrated suite of software facilities for data manipulation, calculation and graphical display

ii. Shiny:

Shiny is an open source R package that provides an elegant and powerful web framework for building web applications using R.

B. Tools

The various tools that will be requiring in the system are as follows:

i. R studio:

R Studio is a free and open source integrated development environment (IDE) for R.

ii. Shiny Dashboard:

R is one of the most popular languages and environments for working with data, and Shiny enables users to build fairly complex web applications using only the R language without knowing any CSS, HTML, or JavaScript. A dashboard built with R/Shiny requires its specification to be contained in two R scripts: ui.R and server.R. Ui.R is where you will specify what the dashboard looks like and provide placeholders for content (e.g. plots, tables, dynamic UI). In server.R, you can refer to user's inputs to generate dynamic content.

iii. Sublime:

Sublime Text is a cross platform source editor with a Python API (API). It natively supports many programming languages and markup languages, and its functionality can be extended by users with plugins.

iv. Tableau:

Tableau Software helps to produces a family of interactive data visualization products focused on business intelligence.

v. E-Commerce Websites API

Ecommerce websites allows you to interact with its data i.e reviews & several attributes about reviews using E-commerce API's.

vi. UML:

The Unified Modeling Language (UML) is a general-purpose, developmental, modeling language in the field of software engineering that is intended to provide a standard way to visualize the design of a system.

VI. ARCHITECTURAL DIAGRAM

The Architectural diagram of our system is as follows:

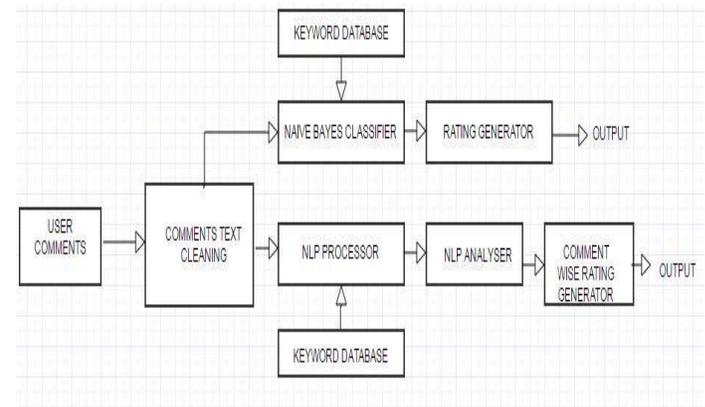


Figure.1. Architectural Diagram

The reviews are fetched from e-commerce website which undergoes text cleaning to extract the keywords from the review. These keywords are then used by Naive Bayes classifier and NLP processor which analyse the keywords using keyword database and generates the rating of the review using sentiment analysis algorithm. After keyword analysis the seller trust scores are generated.

VII. LITERATURE SURVEY

A. Cross-Domain Sentiment Analysis of Product Reviews by Combining Lexicon-based and Learn-based Techniques

Product reviews can direct consumers' purchasing behaviours and sellers' marketing strategies. Therefore, in this paper, they have proposed a novel sentiment analysis of Chinese reviews and have proposed CLL method. They have the lexical analysis based on three domains viz books, hotels and electronics. Furthermore, they use four categories of features (including 16 features in total) to build six classifiers. They conduct the proposed model CLL. The experimental results show that domain lexicons outperform the basic lexicon no matter in which domain. The proposed method performs better than state-of-the-art methods in domains of books and hotels, and is slightly inferior in the domain of electronics.

B. A New way for Semi Supervised Learning Based on Data Mining for Product Reviews

Swift increase in online users along with growing power of review sites has given birth to Sentiment analysis or Opinion mining, which aims at guessing what other people think and comment. Nowadays, several websites are there on which a variety of products are sold. Before making a purchase an online user typically goes through many similar products of different brands before reaching a decision. This apparently simple information retrieval task actually requires a lot of comparison, since all sellers sell similar features and prices for most products. The system gives sentiment analysis of product reviews and not only positive and negative reviews but also gives neutral reviews.

C. Comment Based Seller Trust Model for E-Commerce

Online transactions have seen a rapid growth in the last few years and consumer reviews have become an important part of this growing business. Seller Trust Profile can be computed using various models. These models make use of the feedback reviews which consist of ratings as well as comments given by the user. Models that compute seller trust score based on user feedback overlook the important details in the text and give inaccurate trust scores based on user ratings. We propose a novel model for computing seller trust profile based on the user reviews and we would recommend this trust profile to the user. In this, they have proposed a methodology to compare and verify results. A few common advertisers on Amazon and Flipkart have been identified and trust profiles have been made for them. The trust scores for the sellers on Amazon are lesser than those shown on the website and those for Flipkart are nearly equal to those mentioned on its website. The relative error that is computed is higher on Amazon as compared to Flipkart.

D. A Survey on Creating Trustworthy Seller Profile Using Multidimensional Trust Model

E-commerce uses Reputation based trust models to greater extent. Reputation trust score for seller is obtained by gathering feedback ratings. The “all good reputation” problem however is frequent in current trust systems in most of the websites trust scores are invariably high for sellers and it is difficult for potential shoppers to select appropriate sellers and make truthful decision. In this paper, based on the observation that shoppers often express perspectives openly in free text feedback comments and by mining feedback comments we propose Comment Trust Evaluation. They have proposed a model which is based on multidimensional aspects for computing reputation trust scores from user feedback comments. Feedback comments are mined for dimension ratings and weights and an algorithm is proposed and uniting techniques of natural language processing, opinion mining and topic modelling. .

VIII. REFERENCES

- [1]. Kushagra Bhargava, Tanvi Gujral, Mihir Chawla, Tanya Gujral,”Comment Based Seller Trust Model for E-Commerce,”2016 International Conference on Computational Techniques in Information and Communication Technologies (ICCTICT).
- [2]. Mukta Patkar,Pooja Pawar, Mony Singh,Ashwini Save,” A New way for Semi Supervised Learning Based on Data Mining for Product Reviews,”2nd IEEE International Conference on Engineering and Technology (ICETECH), 17th & 18th March 2016, Coimbatore, TN,India.
- [3]. Neha Khandave, Shubhangi Wakhare, Kanchan Tope, Sagar Vetal,” A Survey on Creating Trustworthy Seller Profile Using Multidimensional Trust Model,”DOI:10.1 5680/IJIRSET.2 015. 04100 32, Vol. 4, Issue 12, December 2015.
- [4]. Kaili Mao, Jianwei Niu, Xuejiao Wang, Lei Wang and Meikang Qiu,”Cross-Domain Sentiment Analysis of Product Reviews by Combining Lexicon-Based and Learn-Based

Techniques,”2015 IEEE 17th International Conference on High Performance Computing and Communications (HPCC), 2015 IEEE 7th International Symposium on Cyberspace Safety and Security (CSS), and 2015 IEEE 12th International Conf on Embedded Software and Systems (ICESSE).