Collaborative Filtering Recommendation System through Sentiment Analysis

Fatemeh Abbasi[a] , Ameneh Khadivar[b*]

[a]Postdoctoral researcher, Department of Management, Faculty of Social Sciences and Economics, Alzahra University, Tehran, Iran, f.abbasi@alzahra.ac.ir
[b]Associate Prof., Department of Management, Faculty of Social Sciences and Economics, Alzahra University, Tehran, Iran, a.khadivar@alzahra.ac.ir*(corresponding author)

Article History: Received: 14 July 2020; Accepted: 2 January 2021; Published online: 5 February 2021

Abstract: Purpose: In recent years, the advent of social networking sites has attracted more attention to review-based recommender systems. The purpose of developing such systems is to use the valuable information, which can be obtained from users' textual reviews. This paper presents a collaborative filtering recommender system using sentiment analysis.

Design / Methodology: For this purpose, a sample of 7210 comments about 221 books from Amazon website are used to sentiment analysis. We used ensemble models to extract users' opinions. Weighted vote-based classifier ensemble technique is used for ensemble modeling. The required data were collected from Amazon.com through Web Crawlers in Java. The data were limited to Amazon users' comments to specific book topics such as Business Intelligence. We applied different methods including text normalization and ensemble methods for doing the sentiment analysis.

Finding: The results showed that sentiment analysis of user reviews has a positive effect on recommending popular goods by users and also on the performance of recommender systems.

Practical Implication: These results show that with understanding the effect of sentimentanalysis for analysis unstructured data , online retailers could use it for policy making and recommend new suggestion to their customers. Also this system helps consumers to make informed decisions.

Originality/value: This study combines sentiment analysis and recommender systems and shows remarkable improvement in the performance of recommender system.

Keywords: Recommender system, Collaborative filtering, Sentiment analysis, Ensemble models, Weight-based voting .

---

1. Introduction

Remarkable increase of competition in business has led to increasing the importance of creation, preservation and development of relationships with customers. On the other hand, the widespread use of information technology in businesses has provided new ways and methods of communication between customers. World Wide Web (WWW) was created to develop information sharing such that daily use of several websites and social networks has now become a part of everyone's life all over the world (Yu, Duan & Cao, 2013). Today, customers, who have access to the Web, can buy the required goods and products at any time or place. For example, book is a product that people are more likely to purchase online via such websites as Amazon. Furthermore, they can express their opinions about a given product in the form of text, and these reviews can help the potential customers to make a purchase.

Today, data collection is no longer considered as a challenge; the real challenge is to extract, convert and analyze large volumes of structured and unstructured data. As it is easy to produce and disseminate text, the Web has a text format and the information in it consists of facts and opinions. Weblogs, social networks and e-commerce sites all provide a ground to express opinions, which can be used to get the users' opinions about social events, political movements and marketing campaigns (Yu, Duan, & Cao, 2013). Large amounts of information in websites are in the form of text and semi-structured data; the analysis of which is particularly important for their hidden knowledge. First, information retrieval techniques such as text indexing were developed to handle the textual data, though they are not efficient for large volumes of ever growing data. Without knowing the content of documents, it will be difficult to formulate appropriate questions to extract useful information from such a huge amount of data. Users need some tools to compare different documents, organize them based on their relevance, and discover the patterns(Liu B., 2012). To fulfill the current needs via the use of large amounts of information available in the Web, which can be found in either structured or unstructured form, we methods that allow access to the data and extract information from them. To extract knowledge from facts and opinions, different methods like text mining and sentiment analysis are used. Text mining is used for automated knowledge extraction from unstructured texts, which includes methods and algorithms to discover useful patterns, as well as classification and extraction of information (Radovanović & Ivanović, 2008). Sentiment analysis, which refers to the study of views, opinions, emotions and attitudes toward an entity expressed in texts, is used to explore opinions from unstructured web texts (Yan, He, Shen & Tang, 2014). Sentiment analysis, also known as opinion mining, deals with identification, extraction, and classification of opinions, emotions and attitudes about various issues. It can provide valuable information to place advertisements in web pages and appropriately solve the problem of automated classification and organization of data (Williams, Bannister, Arribas-Ayllon, Preece & Spasic, 2015).

Today, many e-commerce sites recommend products to users. Product recommendation might be based on best-selling items, demographic features of users, analysis of the users' previous behaviors, and prediction of their future purchasing behavior (Schafer, Konstan & Riedl, 1999). Recommender systems suggest users' popular items through extracting their tastes and preferences. These systems enhance the sale of websites in three ways:

1) They convert website visitors to buyers. Visitors of an e-commerce website often visit various pages and products and leave the website without making a purchase. In this case, recommender systems help them to find the required product.

2) The systems lead to cross-selling. They increase the cross-selling and the average order size by providing suggestions.

3) Recommender systems increase the trust level by creating value added resulting from communicating with customers (Aggarwal C., 2016). Recommender systems actually aim to rank items of the system based on their similarity to the users' interests to recommend higher-ranked items.

In this paper, we present a collaborative recommender system for the users' textual reviews. To do this, we used ensemble method through sentiment analysis. Due to the fact that the output accuracy of the analysis on unstructured data is highly valid, it is necessary to pay due attention to this issue. The remainder of the article is structured as follows: Section 2 reviews the studies conducted in the area of the present study. In Section 3, methodology and general process of the study are presented. Section 4 outlines the findings, and finally, Section 5 includes the conclusions.

2. Literature review

Sentiment analysis is a research field that analyzes the opinions, emotions, evaluation and attitudes of people toward the entities such as products, services, organizations, individuals, issues, events, topics and attributes. Sentiment analysis is also known as opinion mining, opinion extraction, sentiment mining, subjectivity analysis, affect analysis, motion analysis and review mining; however, in industry and university, it is mostly referred to as motion analysis and opinion mining. Sentiment analysis, generally, focuses on the positive and negative feelings of opinions (Liu B., 2012).

Singh et al. (2011) proposed a new approach for ensemble recommender systems. To improve the research results, they combined content-based approach with sentiment analysis. In the end, items with similar content and those labeled as positive in classification of opinions were recommended to users. The proposed approach was implemented on comments about movies (Singh, Mukherjee & Mehta, 2011). Kumar et al. improved movie recommender systems bycombining collaborative filtering and sentiment analysis. They explained in detail various types of recommender systems, taught classifiers and how to analyze emotions by Naïve Bayes classifier. The results indicated that combining sentiment analysis and collaborative filtering increased precision (Kumar Singh, Mukherjee & Kumar Mehta, 2011). Koukourikos et al.'s (2012), in "Techniques of sentiment analysis of user opinions", introduced a resource of recommendation to other users. Sentiment analysis of users' opinions was conducted using a dictionary-based approach on teaching sources and the primary results showed that such opinions are valuable information that can be used as perceived rate of a particular user in a recommender system (Koukourikos, Stoitsis & Karampiperis, 2012). Gurini et al. (2013) proposed a recommender system with a new weighting function called as Sentiment Volume Objectivityfunction where both the interests and feelings of users are considered. Such an approach is helpful in providing more useful suggestions. Primary results of the proposed approach revealed that it works better than other methods examined in the literature review (Gurini, Gasparetti, Micarelli & Tre, 2013). Chen et al. (2015), in their "Recommender systems based on user reviews", reviewed how to use the opinions of users to develop content-based and collaborative recommender systems. This system focuses on removing cold-start[1] and sparsity[2] problems. The study included two sections: creating opinion-based user profile, and product profile. User profile consists of another subclass where opinions are used to create profile based on words and improve the ranks. Under the product profile, opinions are enriched by comparison to enhance the quality of evaluation (Chen, Chen & Wang, 2015).

[1] Cold-start: It concerns the issue that a system cannot recommend an item to anyone unless a large number of users buy significant number of that item. Cold start describes a user who has just joined the recommender system and no sufficient information is available about him.
[2]This appears when despite significant number of users, there is not enough rating or purchase of items and this leads to user matrices or sparse matrix.

Alhamdi and Zeng (2015) proposed a new framework for recommender systems called as "Implicit Social Trust and Sentiment", which extracts user preferences by examining Online Social Networks (OSNs). The research provides a framework to use a new source of data in order to deliver personalized recommendations via search in text posts that friends share through micro-blogging. The proposed framework provides opinions in numerical scales and includes three parts: 1) Measurement of implicit trust between friends based on the intercommunication acts, 2) Measurement of the sentiment rating to reflect the knowledge behind the posts, and 3) Identification of the degree of impact of trust level between friends and sentiments resulting from reviews on recommendations by using machine learning regression algorithms such as linear regression, random forest, and support vector regression (SVR). The framework indicates a semantic relationship between the rating categories. Empirical results, using data from Twitter on the proposed framework, verified the effectiveness of implicit social trust and sentiment (Alahmadi & Zeng, 2015). Jayashree and Kulkarni (2017), in "Recommendation system with sentiment analysis as feedback", proposed an intelligent system to suggest a variety of choices to facilitate the decision making. They developed a Hotel Recommendation System that combines the collaborative filtering with sentiment classification to provide suggestions; to improve the recommendation results, sentiment classification is included as the feedback. There is also performance comparison between the two classifiers of "Naïve Bayesian" and "K-Nearest Neighbor" to classify the sentiments. The proposed hybrid approach helps provide suggestions using item-related reviews in the case where the items have no particular ratings. The results showed that the proposed approach is more accurate comparing to the case where only classification is used (Jayashree & Kulkarni, 2017).

Review of the literature shows that two classes of articles can be identified: some have merely looked at sentiment analysis while others have studied the combination of recommender systems with sentiment analysis. Articles on sentiment analysis primarily look at machine learning and dictionary-based approaches. Concerning the monitored approach, most articles deal with monitored learning and compare the results of two or three algorithms, and based on the results of more accurate algorithm, they model a sentiment analysis. Analysis of the articles reveals that ensemble approach to sentiment analysis has not been taken into account. As the ensemble approach to sentiment analysis can produce more accurate results, more attention should be paid to this method. Also regarding the recommender systems, ensemble approach to sentiment analysis has not been investigated that can be the topic for future research works.

3. Research Method

Here, we are interested to improve the currently available recommender systems based on sentiment analysis. For this purpose, we tried to analysis on Amazon book review using ensemble model. The results of sentiment analysis can be used as feedback for recommending book. We achieved our goals through eight steps, which will be explained in detail separately. Figure 1 shows the research framework. The study contains data collecting, pre-processing, modeling, sentiment analysis and recommendations, which will explained in the process of the research.

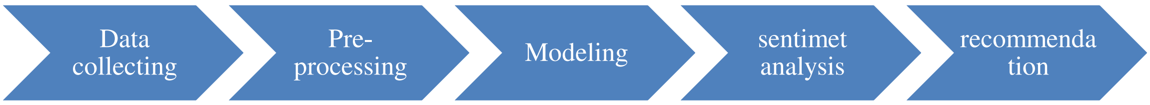

Figure 1: Research framework

Data collecting

In this research, a web crawler is used to collect the required data through the Amazon website by inserting initial keywords such as “Business Intelligence”. To design the crawler, Java programming language is used. This software is specifically designed to survey the Amazon website so that it can find the relevant books from the website among all books based on a searched topic. During the search process, more than 221 books, 7210 user comments and other information about books with the subject of Business Intelligence were extracted. The dataset contains title of book, title of comment, user rating, name of the commenter and comment text. All data were saved in the form of JSON file.

Data pre- processing

Data pre-processing is the most important step in text mining, Natural Language Processing (NLP) and Information Retrieval (IR). As texts often include particular formats such as numbers, dates and words that do not help text exploration and can be eliminated, unnecessary formats have to be removed before any data analysis is done (Gurusamy & Kannan, 2014). Online texts often contain noise and uninformative parts such as HTML tag, script and advertisements. Furthermore, some words in the text have no impact on the general orientation of it, and keeping those parts or words might be problematic for processing and accurate analysis. As a result, pre-processing of data is really necessary to improve the quality and accuracy of analysis (Haddi, Liu & Shi, 2013).

In this study, techniques used for data pre-processing include: case folding, tokenization, replacement of noisy data, n-grams, Part of Speech (POS) tagging, removal of stop-words, and stemming.

Case folding: In this step, all characters are converted into uppercase or lowercase so that if a word is repeated several times but in different cases, it is included once in the modeling process (Chibelushi & Thelwall, 2009). In this research, all uppercase letters have been converted to lowercase ones.

Tokenization: Tokenization or breaking up of a text is to segment it into smaller units such as words, terms and symbols, commonly known as "tokens" (Katariya & Chaudhari, 2015). In the present study, all comments were segmented into smaller units, and finally, the words obtained from the texts were considered as independent variables. Moreover, words with less than four letters or more than 25 letters were ignored to avoid the inclusion of ineffective words as variables in the model.

Replacement of noisy data: Noisy data refers to abbreviations and acronyms. For example, "TY" is the acronym for "Thank You" (Stavrianou, Andritsos, & Nicoloyannis, 2007). Such words often have emotional weight, so they have to be replaced by the original words. To replace such meaningless words in this research, a list of acronyms was prepared.

N-grams: N-gram model is widely used in text mining and NLP. An n-gram is a continuous sequence of n items of a text or speech. An n-gram of size 1 is called "unigram", size 2 is a "bigram", and size 3 is a "trigram". For example, in "This is a sentence", if n=2 (bigram), then n-grams are: This is, is a, a sentence. One application of n-gram is for developing features for supervised machine learning models. It is also used for sentiment classification to select features. This method refers to breaking up of a text into different parts where n represents the number of words in each part. In a systematic survey, Cheng et al. (2007) investigated the relationship between the number of ns and the information obtained from the text, the results of which are illustrated in Figure 2. Their findings indicated that as the number of n increases, the amount of information obtained from the text decreases, and the maximum amount of information is obtained when n=2 (Cheng, Yan, Han, & Hsu, 2007). Accordingly, we used bigram.

Part of Speech (POS) tagging: It is a linguistic method used since 1960, which has recently attracted the researchers in NLP to extract features of their products. In POS tagging, each word is assigned a tag as verb, adjective or adverb based on its grammatical function in the text(Zubair Asghar, Khan, Ahmad & Kundi, 2014). It is very effective for feature extraction in terms of accuracy (Zubair Asghar, Khan, Ahmad, & Kundi, 2014). As the words in this research that function as verb, adjective and adverb carry the emotional weight of the text, only the words with these functions were selected for analysis in the modeling phase.

Removal of stop-words: Words that, despite their high frequency in the text, have less semantic value. They include pronouns, prepositions and conjunctions that do not carry any information (Ramasubramanian & Ramaya, 2013).

Stemming: This method is used to find the stems. In this phase, all words are reduced to their original stem. Stemming algorithms remove affixes according to the grammatical rules. In the present research, the Snowball Stemmer Library, which is the most popular and standard approach, was used (Krouska, Troussas, & Virvou, 2016). Table 1 shows steps of pre-processing from one of the examined articles.

Table 1: Pre-processing steps

Pre-processing steps	Before processing	After processing
Noisy data	I recommend this book .to anyone .TMI	I recommend this book to anyone .. Too Much Information
Tokenization	. I recommend this book to anyone .Too Much Information	'I', 'recommend', 'this', 'book', 'to', 'anyone', '.', 'Too', 'Much', 'Information'
Stop-word	'I', 'recommend', 'this', 'book', 'to', 'anyone', '.', 'Too', 'Much', 'Information'	'I', 'recommend', 'book', 'anyone', '.', 'Too', 'Much', 'Information’
Case Fold	'I', 'recommend', 'book', 'anyone', '.', 'Too', 'Much', 'Information’	‘i', 'recommend', 'book', 'anyone', '.', 'too', 'much', 'information'
Stemming	‘i', 'recommend', 'book', 'anyone', '.', 'too', 'much', 'information'	I, recommend, book, anyon, too, much, inform
Part of Speech	. I recommend this book to	('I', 'PRP'), ('recommend', 'VBP'),

tagging (POS tagging)	anyone .Too Much Information	('this', 'DT'), ('book', 'NN'), ('to', 'TO'), ('anyone', 'NN'), ('.', '.'), ('Too', 'VB'), ('Much', 'JJ'), ('Information', 'NN')
Bi-gram	. I recommend this book to anyone .Too Much Information	('I', 'recommend'), ('recommend', 'this'), ('this', 'book'), ('book', 'to'), ('to', 'anyone'), ('anyone', '.'), ('.', 'Too'), ('Too', 'Much'), ('Much', 'Information')

Modeling

In the second step, sentiment analysis model is made by using Ensemble Methods. Ensemble method is used for modeling based on some classifier algorithms, and these algorithms work together to classify the comments. In this research, seven classifiers of Naïve Bayes, Linear Support Vector Classification (Linear SVC), Multinominal Naïve Bayes (MNB), Multi-lay Perceptron (MLP), Benoulli Naïve Bayes (Benoulli NB), Logestic Regression and Nu-Vector Classification (Nu-SVC) from Python NLTK Library were used for modeling. These algorithms are used in supervised learning method to make predictions. Supervised machine learning approach is a kind of learning with specific input and output and a supervisor that gives some information to the learner. In this way, the system tries to learn a function that maps an input into an output. In a machine learning-based classification, there are two sets of texts: training set, and test set. Training set is used by an automatic classifier to learn distinctive features from the texts, and test set is used for evaluating the classifier (Medhat, Hassan & Korashy, 2014). At the end, the final result of each comment is determined by using weighted approach.

Ensemble method is based on combining a set of classifiers, which can lead to higher performance accuracy than using them independently. In weighted vote-based classifier ensemble technique, different classifiers have different voting weights and classifiers with better performance may get higher weights. Therefore, it is very important to select appropriate weights of votes (Zhang, Zhang, Cai & Yang, 2014). In this research, different weights were assigned based on the accuracy of the above mentioned seven classifiers. In assigning weights, more accurate classifiers were assigned higher weights, so the weights number 1, 2 and 3 were used for estimation. It is to be noted that various cases of weighting have been evaluated among which the above mentioned weights showed better performance.

Evaluating the model for sentiment analysis

After implementing the model using ensemble approach, now the precision, recall and F measures have to be estimated. Precision and recall are measures applied in information retrieval for determining the proportion of documents retrieved by the system according to the users' needs (V. Raghavan A. & S. Jung, 1989).

Measures for evaluating negative classes:

(1) $$\text{Recall} = \frac{TN}{FP+TN}$$

(2) $$\text{Precision} = \frac{TN}{TN+FN}$$

(3) $$\text{F measure} = \frac{2*Recall*Precision}{Recall+Precision}$$

Measures for evaluating positive classes:

(4) $$\text{Recall} = \frac{TP}{FN+TP}$$

(5) $$\text{Precision} = \frac{TP}{TP+FP}$$

(6) $$\text{F measure} = \frac{2*Recall*Precision}{Recall+Precision}$$

Measures for evaluating classifier algorithms:

(7) $$\text{Accuracy rate} = \frac{TN+TP}{TN+FN+TP+FP}$$

(8) $$\text{Error rate} = \text{accuracy rate} - 1 = \frac{FN+FP}{TN+FN+TP+FP}$$

One method for evaluating classifier algorithms is K-fold cross validation. To reduce the error in modeling, this method has been used to separate train and test data. To do so, the whole data is divided into K equal parts. From this K subset, each time, one fold is used as testing set and the rest (K-1) as training test. This process is repeated K times so that, each time, one different fold is evaluated and one particular precision is estimated for the proposed model. At the end, the average result of these K times is used for final estimation. The value most commonly used for K in technical texts is 10 (Kohavi, 1995)(P. Bradford & E. Brodley, 2001). Hence, we also selected the K-value as 10. The model was implemented 10 times and the average results were, finally, calculated.

Sentiment analysis

After evaluating the model, sentiment analysis was performed on 7210 reviews obtained from Amazon.com and the users' positive or negative reviews about books were collected.

Item similarity

Based on the scores obtained from sentiment analysis (from -1 to +1) and the ratings assigned by the users to products (numeric ratings of 1, 2, 3, 4 and 5), the user-item-matrix was formed. This matrix is used to determine the users' preference and detect the nearest neighbors of users and items. An essential part in collaborative filtering-based recommender systems is to calculate item similarity and choose the most similar items and goods. The primary idea behind computing the similarity between i and j items is to isolate the users, who have rated both of these items, and then compute the similarity between them. There are various methods to compute item similarity in recommender systems based on collaborative filtering, the most popular of which is correlation-based similarity. In this approach, similarity between the two items of i and j was estimated using Pearson’s correlation and Equation 10 (Sarwar, Karypis, Konstan & Riedl, 2001):

(9) $$\text{Sim (i, j)} = \frac{\sum_{u \in U}(R_{u,i} - \bar{R}_i)(R_{u,j} - \bar{R}_j)}{\sqrt{\sum_{u \in U}(R_{ui} - \bar{R}_i)^2}\sqrt{\sum_{u \in U}(R_{uj} - \bar{R}_j)^2}}$$

Where, $R_{u,i}$ is the user u rating for item i, $\bar{R}_i$ is the average scores of item i, and u denotes the user, who rated items i and j. In this research, item similarity was computed using Pearson’s Correlation, and based on user-item ratings, the results of which (collaborative filtering-based similarity) were applied for predicting user preferences.

Recommendation

When similar items are identified, new goods and products are recommended to a target user based on the similarity of an item to the ones that he/she has rated. In this research, weighted sum method was used to provide recommendation. By calculating the total weight of scores assigned by the users, the target users’ preferences are predicted and recommended according to Equation 10 (Sarwar, Karypis, Konstan & Riedl, 2001):

(10) $$P_{u,i} = \frac{\sum_{allsimilaritems,N}(S_{i,N} * R_{u,N})}{\sum_{allsimilaritems,N}(|S_{i,N}|)}$$

Where, $P_{u,i}$denotes item *i* recommendation to user *u*, $S_{i,N}$ denotes the similarity between item *N* (previously rated by the user) and item *i*, and $R_{u,N}$ denotes the score given by user *u* to the item *N*.

Recommender system evaluation

One important step to make sure that the recommender system can be extended well is to evaluate its performance. There are some well-known measures to examine the precision or performance of recommender systems, including accuracy, coverage, and F-measure (to balance the two measures) that can reflect the recommender performance well (Hassan & Hamada, 2017).

Accuracy: It reflects how many percent of the overall recommendations is correct; in other words, it measures the truth of recommendations. Favorite items are the set of goods that should be suggested to the users and recommended items are the set of goods recommended by the system to the users:

(11) $$\text{Accuracy} = \frac{\text{Recommended Items} \cap \text{Favorite Items}}{\text{Recommended Items}}$$

Coverage: It indicates how many percent of user's favorite items have been recommended to him/her; then:

(12) $$\text{Coverage} = \frac{\text{Recommended Items} \cap \text{Favorite Items}}{\text{Favorite Items}}$$

(13) $$\text{F-measure} = \frac{\text{precision} \times \text{coverage} \times 2}{\text{precision} + \text{coverage}}$$

4. Results

After transforming the document data into structure data using TF-IDF to extract terms and key phrases, we fed the data to k-means algorithms. After transforming the text data to structured data in pre-processing step to extract terms and phrases, we fed the data to the classifiers for modeling. As mentioned before, seven classifiers, including Naïve Bayes, Linear Support Vector Classification (Linear SVC), Multi-nominal Naïve Bayes (MNB), Multi-layer Perceptron (MLP), Bernoulli Naïve Bayes (Bernoulli NB) and Logestic Regression algorithms were used for modeling. To evaluate each of the seven algorithms, the measures of precision, recall, F-measure, precision rate, and error rate were used. Table 2 presents the precision related to each algorithm. These results were obtained by repeating the model for each classifier for 10 times. It was revealed that Multinominal Naïve Bayes with the precision rate of 0.7911 had higher precision than the other algorithms for modeling.

Table 2: Evaluation of classifier algorithms

Classifier	Precision	Recall	F-measure	Accuracy rate	Error rate
Naïve Bayes	0.69	0.675	0.678	0.704	0.295
Linear Support Vector Classification (Linear SVC)	0.744	0.743	0.743	0.7555	0.243
Multi-nomial Naive Bayes (MNB)	0.783	0.774	0.777	0.7911	0.207
Multi-layer perceptron (MLP)	0.753	0.748	0.747	0.7599	0.239
Bernoulli Naive Bayes (Bernoulli NB)	0.668	0.647	0.650	0.6836	0.315
Logistic Regression	0.783	0.774	0.777	0.7907	0.208
Nu-Vector Classification (Nu-SVC)	0.761	0.752	0.753	0.768	0.231

When the results of each algorithm was determined, weighted vote-based classifier ensemble technique was used. Table 3 shows the evaluation results of the ensemble approach. As you can see, precision and performance of the ensemble approach are better than those of the individual classifiers. This causes more accurate prediction in areas where classifiers fail to have an appropriate performance. The results indicated that the ensemble approach made up the errors of classifiers and developed a model with higher precision.

Table 3: Evaluating the weighted vote-based ensemble approach

	Precision	Recall	F- measure	Accuracy rate	Error rate
Weighted vote-based classifier ensemble technique	0.7958	0.7804	0.7854	0.8003	0.1989

In the next step, the proposed model was implemented on 7210 comments about 221 books on business intelligence extracted from Amazon.com, and then the positive or negative reviews of users for each book were determined. Table 4 presents a sample of scores obtained from sentiment analysis (i.e. positive and negative reviews about each book as well as the related scores assigned by the users). In this table, the review score denotes the positive or negative rating for each book, e.g. -0.61538 means that 61% of the classifiers have recognized a review as negative.

Table 4: Positive and negative reviews

User comments	Sentiment score)positive / negative(	User rate
This is an excellent and thorough overview of the dynamics of emotional maturity and responsibility. I can't wait to finish the next book, "Working with Emotional Intelligence." Thank you Dr. Goleman and your staff for all the research and hard work. This work will continue to impact our world in the coming decade.	+1	5

User comments	Sentiment score)positive / negative(	User rate
The reviews that degrade Goleman's work are rather petty. This book was not written to debunk the importance of IQ, but rather to explore the other side of humanity that we often overlook-emotionality.	-0.61538	4

Therefore, a user-item matrix was created to analyze the user reviews. In this stage, similarity between items is calculated based on three kinds of data: user rating of books (from 1 to 5), a score obtained from the sentiment analysis of user reviews (a digit between +1 and -1) and the combination of user rating and sentiment analysis score. When similarity between the items is estimated, the items similar to the ones rated previously by the same user, is recommended to him/her. To combine the user rating and the score obtained from sentiment analysis, sentiment score was normalized to place between 0-5. The latter was normalized to be placed between 0-50. To combine these two scores, the average score was found.

In this step, recommendation score is calculated using the weight sum method. For example, if a user rates a book as 5 and the similarity between this book and the one he/she has not rated is 0.9, a new book with the score of 4.5 will be recommended to him/her. In this method, books, which have both high scores and have received high rating by the user, would be recommended to him/her. Table 5 presents a sample of recommendations to users.

Table 5: A sample of recommendations to users

Commenter's Name	Books rated by users	Recommended books	Score of recommended book
Wanderley Oliveira Mendes	Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die.	Data Smart: Using Data Science to Transform Information into Insight	+4.745678
		Super-intelligence: Paths, Dangers and Strategies	+3.269767
		Emotional Intelligence 2.0	+2.543639
		HBR's 10 Must Reads on Emotional Intelligence by Harvard Business Review (2016-08-09).	+1.853950
		Emotional Intelligence: Why It Can Matter More Than IQ?	-1.017456

In this system, negative score means that the item is not favored by the user; therefore, it is removed from the list of items to be recommended. Moreover, among the items with positive score and defined threshold set, only those scored higher than 3.5 will be recommended to the user so that they may be favored more.

To evaluate the system, books scored higher than 4 were selected for evaluation, and the recommended scores were compared with the user's real rating to assess the degree of precision. In this stage, precision of the recommender system results was compared with various scores obtained using the measures of accuracy, coverage and F-measure.

Table 6. Evaluating the recommended scores based on three types of data

CriteriaEvaluation / Data	Accuracy	Coverage	F-measure
User rate	0.67782	0.83672	0.7489

Sentiment score	0.58190	0.93124	0.7162
Combined user rate and sentiment score	0.5544	0.92148	0.6923

The data given in Table 6 indicate that recommender systems based on sentiment analysis cover higher percentage of user's favorite items and have positive effect on the systems' performance. In the final step and to prepare the system, Python Tkinter package was used to design the user interface.

Figure 2: User interface of the recommender system

5. Conclusion

This research looked at proposing a recommender system to purchase goods using the sentiment analysis of buyers. Sentiment analysis plays an important role in various areas of e-commerce. Here, the recommendations were provided based on three types of data: a) user rating to each book in quantitative form according to the pre-determined values of 1, 2, 3, 4 and 5; b) estimated scores based on the sentiment analysis of users' text reviews on each book; and c) the average scores given by the users and the scores obtained from sentiment analysis.

The research results showed that unstructured text data analysis can positively affect the percentage of favorite recommendations for users and performance of the recommender system. It can also be argued that applying user reviews can have positive effect on providing recommendations in line with user interests. Users' reviews actually reflect their feelings and opinions about products and can be used as a more precise basis to provide better recommendations to the users.

Most studies conducted so far have taken into account a single classification method to propose a model for sentiment analysis, while the present research has used ensemble approaches including the seven classifiers of Naïve Bayes, Linear Support Vector Classification (Linear SVC), Multi-nominal Naïve Bayes (MNB), Multi-layer Perceptron (MLP), Bernoulli Naïve Bayes (Bernoulli NB), Logestic Regression, and Nu-Vector Classification (Nu - SVC). The final result of each review has been determined via majority vote and weighted voting approached, which have been unprecedented. The model proposed in this research can be applied in other areas, as well as for the sentiment analysis of user reviews with the least changes required. In this research, user information items such as demographic information and user purchase were not available. Furthermore, the present system can merely provide recommendations to those users, who have previously used a product, i.e., they are current customers of the product. If more information is included about the items and users, the extended sample can provide recommendations to new users as well. To generalize the results, it is suggested that the recommended method be implemented using data from other industries. Also, in order to provide more precise recommendations, it is recommended to include the time factor in future system designs.

References

Liu, B. (2015). *Sentiment analysis: Mining opinions, sentiments,and emotions.* Cambridge University Press.

Radovanović, M., & Ivanović, M. (2008). TEXT MINING: APPROACHES AND APPLICATIONS. *Novi Sad J. Math, 38*(3), 227-234.

Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2001). Item-Based Collaborative Filtering Recommendation Algorithms. *WWW '01 Proceedings of the 10th international conference on World Wide Web* (pp. 285-295). Hong Kong: ACM .

Aggarwal, C. (2016). *Recommender Systems: The Textbook.* NY: Springer.

Alahmadi, D., & Zeng, X.-J. (2015). ISTS: Implicit social trust and sentiment based approach to recommender systems. *Expert Systems With Applications*, 8840-8849.

Chen, L., Chen , G., & Wang, F. (2015). Recommender systems based on user reviews: the state of the art. *User Model User-Adap Inter*, 99–154.

Cheng, H., Yan, X., Han, J., & Hsu, C.-W. (2007). Discriminative Frequent Pattern Analysis for Effective Classification. *2007 IEEE 23rd International Conference on Data Engineering.* Istanbul: IEEE.

Chibelushi, C., & Thelwall , M. (2009). Text Mining for Meeting Transcript Analysis to Extract Key Decision Elements. *International MultiConference of Engineers and Computer Scientists*, 710-715.

Gurini, D., Gasparetti , F., Micarelli , R., & Tre, R. (2013). A Sentiment-Based Approach to Twitter User Recommendation. *Rec Sys 2013.* Hong Kong: RSWEB.

Gurusamy, V., & Kannan, S. (2014). Preprocessing Techniques for Text Mining. *RTRICS.* Podi.

Haddi, E., Liu, X., & Shi, Y. (2013). The Role of Text Pre-processing in Sentiment Analysis. *Procedia Computer Science*, 26-32.

Hassan, M., & Hamada, M. (2017). A Neural Networks Approach for Improving the Accuracy of Multi-Criteria Recommender Systems. Applied Sciences, 7(9), 868

Jayashree, R., & Kulkarni, D. (2017). Recommendation System with Sentiment Analysis as Feedback Component. *Proceedings of Sixth International Conference on Soft Computing for Problem Solving* (pp. 359-367). Singapore: Springer.

Karampiperis, P., Koukourikos, A., & Stoitsis, G. (2014). Collaborative Filtering Recommendation of Educational Content in Social Environments Utilizing Sentiment Analysis Techniques. *Recommender Systems for Technology Enhanced Learning* , 3-23.

Katariya, N. P., & Chaudhari, M. S. (2015). TEXT PREPROCESSING FOR TEXTMINING USING SIDE INFORMATION. *International Journal of Computer Science and Mobile Applications*, 1-5.

Kohavi, R. (1995). A Study of CrossValidation for Accuracy Estimation and Model Selection. *Appears in the International Joint Conference on Articial Intelligence (IJCAI)*.

Koukourikos, A., Stoitsis, G., & Karampiperis, P. (2012). Sentiment Analysis: A tool for Rating Attribution to Content in Recommender Systems. *2nd Workshop on Recommender Systems for Technology Enhanced Learning* (pp. 61-70). Saarbrucken: Software and Knowledge Engineering Laboratory (SKEL).

Krouska, A., Troussas, C., & Virvou, M. (2016). The effect of preprocessing techniques on Twitter. *Chalkidiki.* Greece: IEEE.

Kumar Singh, V., Mukherjee, M., & Kumar Mehta, G. (2011). Combining Collaborative Filtering and Sentiment Classification for Improved Movie Recommendations. *International Workshop on Multi-disciplinary Trends in Artificial Intelligence* (pp. 38-50). Berlin: Pringer, Heidelberg.

Leimstoll, U., & Stormer, H. (2007). Collaborative recommender systems for online shops. *Americas Conference on Information Systems(AMCIS).* Retrieved from Available in: http://aisel.aisnet.org/cgi/viewcontent.cgi?article=1666&context=amcis2007.

Liu, B. (2010). Sentiment Analysis and Subjectivity. In *Invited Chapter for the Handbook of Natural Language Processing* (pp. 1-38). Boca: Taylor and Francis Group. Retrieved from http://gnode1.mib.man.ac.uk/tutorials/NLPhandbook-NLPhandbook-

Liu, B. (2012). *Sentiment Analysis and Opinion Mining (Synthesis Lectures on Human Language Technologies).* Williston: Morgan & Claypool Publishers.

Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications:A survey. Ain Shams Engineering Journal, 5(4), 1093-1113.

P. Badfordb, J., & E. Brodley, C. (2001). The Effect of Instance-Space Partition on Significance. *Machine Learning*, 269-286.

Ramasubramanian, C., & Ramya, R. (2013). Effective Pre-Processing Activities in Text Mining using Improved Porter's Stemming Algorithm. *International Journal of Advanced Research in Computer and Communication Engineering*, 4536-4538.

Schafer, J., Konstan, J., & Riedl, J. (1999). Recommender systems in e-commerce. *Proceedings of the 1st ACM Conference on Electronic Commerce*, 158-166.

Singh, V., Mukherjee, M., & Mehta, G. (2011). Combining a Content Filtering Heuristic and Sentiment Analysis for Movie Recommendations. *Springer-Verlag Berlin Heidelberg*, 659-664.

Stavrianou, A., Andritsos, P., & Nicoloyannis, N. (2007). Overview and semantic issues of text mining. *ACM SIGMOD Record*, 23-34.

Raghavan, V., Bollmann, P., & Jung, G. S. (1989). A critical investigation of recall and precision as measures of retrieval system performance. ACM Transactions on Information Systems (TOIS), 7(3), 205-229.

Williams, L., Bannister, C., Arribas-Ayllon, M., Preece, A., & Spasic, I. (2015). The role of idioms in sentiment analysis. *Expert Systems with Applications*, 7375–7385.

Yan, G., He, W., Shen, J., & Tang, C. (2014). A bilingual approach for conducting Chinese and English social media sentiment analysis. *Computer Networks*, 491–503.

Yu, Y., Duan, W., & Cao, Q. (2013). The impact of social and conventional media on firm equity value:A sentiment analysis approach. Decision Support Systems. *Decision Support Systems*, 916-926.

Zhang, Y., Zhang, H., Cai, J., & Yang, B. (2014). A Weighted Voting Classifier Based on Differential Evolution. *Abstract and Applied Analysis*. Retrieved from http://dx.doi.org/10.1155/2014/376950

Zubair Asghar, M., Khan, A., Ahmad, S., & Kundi, F. (2014). A Review of Feature Extraction in Sentiment Analysis. *Journal of Basic and Applied Scientific Research*, 181-186.