Grey Relational Classification of Consumers' Textual Evaluations in E-Commerce

Hüseyin Fidan

Mehmet Akif Ersoy University, Industrial Engineering Department, Burdur, Turkey, hfidan@mehmetakif.edu.tr

Abstract

Companies have gained important advantages by the development of electronic commerce. Consumer evaluations in electronic environment offer great possibilities for analysis. The fact that the consumer opinions are comprised of textual data, analyzes have complicated and challenging process. In recent years, it is seen that text mining methods are used in analyzes in the literature. However, the evaluations of consumers which are formed by short texts make it necessary to realize the analysis with insufficient data. The weighting methods such as Term Frequency and Term Frequency-Inverse Document Frequency as well as common used classification algorithms such as Naïve Bayes and Support Vector Machine have some inadequacies in short text analysis. In this study, a grey relational classification model based on Vector Space Model and Bag of Words has been developed. The model was first applied to the positive-negative categorization of the evaluations, then, applied to the classification of negative evaluations. It was determined that the accuracy level of the model is higher than the classification algorithms commonly used in short text. According to the results of the research, 9637 negative evaluations in 24479 consumer opinion were determined, and 50.4% of the negative evaluations were found to have the most problems related to product.

Keywords: Consumer relationships management, Consumer evaluation analysis, Text mining, Grey system theory, Grey relational classification

1   Introduction

Electronic commerce (e-commerce) offers important advantages to consumers and firms. Especially the Web 2.0 technologies have bring a new transaction environment to the companies. So, business processes have gained new dimensions, improvments and possibilities. Data in e-commerce offer great opportunities that cannot be ignored for analysis of customer understanding, identifying needs, enhancing customer satisfaction and building loyalty. In this respect, it is important to analyze the customer evaluations for survival and profitability of firms in the new competition environment which shaped with internet technologies [70].

In the studies on consumer evaluations, the data which consist of products, consumer satisfactions and company ratings are acquired from survey methods or e-commerce sites. Although, these studies have major contributions to the literature, they are inadequate in presenting realistic results because they are not based on the textual views of the consumers [75]. In a few studies based on textual evaluations, semi-structured data that e-commerce sites supply have been analyzed by text mining [97]. However, the data provided by the e-commerce site will not produce healthy results, because it is possible to remove negative opinions by the company.

Text classification problem is one of the main topics in text mining. The automatic classification of consumer opinions is a subject of interest in text mining literature. Especially, classification of textual data consisting of a few words or limited characters, such as twitter messages, news head-lines, consumer opinion, etc. is a challenging process. Satisfactory results cannot be achieved in text mining because evaluations of consumers are composed of short texts [66]. Short texts mean less data to be used for analysis. So, text mining methods that still have some deficiencies such as less accuracy and long process time in case of insufficient data which called sparsity problem. The aim of the study is to improve a classification method that has higher accuracy in short text classifications. Grey Relational Analysis (GRA) is one of the common methods referred in incomplete information. In literature, machine learning based algorithms are widely applied in short text analysis, but GRA is rarely used. It will be an appropriate choice for analyzing of short text data by GRA in case of incomplete information.

In this study, the textual evaluations of consumers have been obtained via the developed web mining application and analyzed by text mining. The novelty of the study is to develop a new grey-based classification model that can be used instead of the algorithms which have some weakness in short text analysis. In this context, the training dataset was formed with 500 evaluations randomly selected from the non-structural data obtained by consumer evaluations and categorized. Through the training process, bag of words (BoW) were obtained and GRA was performed by creating Vector Space Model (VSM) in which each consumer evaluation in the training dataset was used as a reference. It was obtained that the accuracy of the improved model is determined to have high accuracy values according to the test result. The comparison results show that the accuracy of the improved model is higher than the Naïve Bayes (NB) and Support Vector Machine (SVM) algorithms. Finally, the GRA classification model was applied to classify 24,479 consumer evaluations and it was determined that the most problematic issue among the evaluations is product related. The results of the study show that the GRA should be used to classify short-text consumer evaluations for higher accuracy rates.

Following the introduction of the study, the literature review was given dealing with the subject. Grey System Theory and GRA were given in the third section. The materials and methods of the research carried out in the fourth part, the findings of the analysis and the comparison of the results was shown in the fifth section. Finally, conclusions and recommendations were given in the last section.

2   Literature Research

The section of literature research discusses the customer relationships management and text mining process in context of the short text classification studies.

2.1   Customer Relationships Management

Customer relationships management (CRM) is one of the most important firm strategy that aims both at gaining new customers and the keeping old customers to build profitable processes in long run. [70]. The CRM processes that add businesses value to the firms aim to develop policies to discover customers, their preferences, needs and expectations [28]. Companies that increase their customer satisfaction and achieve their loyalty gain significant competitive advantages [70]. For this reason, CRM is a firm strategy that enables firms to stay in the market [54].

Although it is claimed that CRM-related processes were based on the commercial transactions of Mesopotamian farmers [9], initial systematic researches on CRM have begun in 1980s by Robert D. Kestnbaum's studies. In Kestnbaum method which is called as *Direct Marketing* or *Database Marketing*, database was created with customer information, data were analyzed with statistical methods and the obtained results are used to sale products to the other consumers by email [14]. The approach, which includes data collection, analysis and implementation processes, is also the basis of today's CRM systems. With the widespread use of the internet, increased data

volume has made it possible to establish meaningful relationships and obtain meaningful information from intense data. The inadequacy of traditional analysis combined with the increase in data volume has led to the use of data mining methods in CRM systems. Thus, data acquisition methods of the companies such as questionnaire and complaint boxes have changed with the possibilities of the Web 2.0 [73]. Internet applications such as Facebook, Twitter, forums, blogs, compliant systems and e-mails provide important possibilities during the data acquisition process and offer to realizing more effective analysis in CRM.

2.2 Studies on Customer Evaluations

Process and product evaluations of customers which are among the main topics in CRM have crucial importance for companies to develop the right strategies. In this respect, customer feedbacks are the major topics of interest for the organizations as source of information [32], [63], [85]. Especially depends on the increasing e-commerce transactions after 2000's analysis of customer evaluations data in digital environment have become an interesting field and attracted the companies' attention. The first studies on consumer complaints on web pages have been carried out in these years. Generally, CRM applications to achieve customer loyalty and satisfaction are divided into two groups, operational and analytical [7], [33]. The operational CRM aims to polish up the automation of company processes and the analytical CRM seeks the customer characteristics and behaviors [59].

Data mining, which is defined as revealing hidden knowledge from huge amount of data, is the preferred method in analytical CRM applications nowadays [4], [7], [38], [73]. Most of the studies carried out using methods such as classification, clustering, prediction and associations are related to customer satisfaction scoring about products and services such as hospitals [1], hotels [6], [47], telecommunication [81], mobile phones [18], cars [26], auctions [13] etc. Instead of scoring systems that do not fully reflect consumer opinions, analyzing of consumer complaints and assessments will be a more realistic approach. However, such analyses are rare in the literature, because consumer evaluations are in the text type [59]. Inadequacies of data mining methods in the analysis of non-structural data make it difficult to analyze the text data [83].

Analysis of non-structural data have begun to be more common since 2000s, and analysis of customer text evaluations have also increased. In the study of Li and Yamanishi [49], which is one of the first studies on analyzing customer comments, opinions and requests with textual data, applied rule-based text mining to consumer evaluations that obtained by open-ended questionnaire. The study emphasized that the method can be used effectively in firm decision processes. Nasukawa and Nagana [58] analyzed opinions of the customer by obtaining data from the consumer support center instead of the questionnaire method. They determined the count of the customer problems by calculating the word frequencies in the text. Hu and Liu [36] have demonstrated that customer problems can be clustered by using the consumer evaluation data. They have obtained the data from an e-commerce site and have realized the clustering as positive or negative in a text mining based categorization study on the qualification of electronic products. Park and Lee [67] have classified the opinions and complaints about consumer profiles and products by using decision tree algorithm. They have determined the group of the products according to the characteristics of a new product by finding associative words. Gamon et al. [26] performed clustering of consumer evaluations with non-structural data from a readily available database of consumer opinions on cars. In the research, the consumers' evaluations are divided into three categories as positive, negative and other by weighting the terms with TF-IDF. According to Onan et al. [59], the term weighting methods such as TF, TF-IDF are insufficient in short text analysis and they have suggested the ensemble learning methods such as bagging.

Recent studies on CRM and text mining focus on the realizing automated classification systems. Coussement and Poel [15] have designed a system that automatically classifies customer e-mails by using linguistic style features for loyalty estimation. Zhang, Narayanan and Choudhary [97], who provided consumer opinions from Amazon.com, analyzed camera and television evaluations and developed an automated evaluation system by product ranking scores. Lee and Bradlow [46] have claimed that market characteristics of the products can be obtained by consumer product evaluations instead of product scores. They have implemented an application that automatically determines the market characteristics of the products by using text mining to identify the words used in consumer evaluations.

When the consumer evaluations analysis literature is examined, it is seen that the researches are carried out in different branches of science and analyzes are generally realized using the consumer's evaluation scores. Datasets are usually compiled by researchers from data sources such as company web pages, forums, complaint sites, social media and e-mail. Compilation of data by researchers manually facilitates the analysis, but leads to limited analyzes as few data are used on the other side. In the other hand, consumer evaluations that contain a few words lead to the sparsity problem in analyzes. The sparsity problem, which is one of the main problems in short text analysis, leads to unhealthy results or analyzes cannot be concluded.

2.3 Text Mining

The data resources such as chat rooms, forums, blogs, social media, complaint sites contain non-structural data dealing with social life such as text, pictures, graphics, video etc. According to Tuzhilin [88], the analysis of the non-structural data will be inevitably a very interesting field in the next decade. The analyzing unstructured data is a very important issue for business community since 85% of the data produced in commercial processes is composed of

textual data [86]. Unstructured data analysis requires the textual operations that appropriate to spoken language structure due to the insufficiency of data mining approaches. One of the first study on textual data analysis was carried out by Hebrail and Mersais in 1992. The researchers attempting to extract key words based on Natural Language Processing (NLP) from project narratives recorded in the database in text form have performed clustering analysis of documents by using key words [34]. However, the process of converting textual data into structural data was first described by article of Feldman and Dagan [22]. In the study, after the textual data in the documents were converted into structural data, meaningful concepts were determined by using the text categorization approach of Hebrail and Marsais [22].

There are two approaches to text mining as contextual and lexical analysis [29]. In the contextual analysis, it is aimed to determine word or word associations according to the frequency values of the words in the text. In this approach, similarities between texts are determined using n-gram, word frequency, and TF-IDF statistical methods [5]. Lexical analyzes are carried out in the context of the NLP in order to examine texts meaningfully [72]. Within the context of lexical analysis, applications such as sentiment mining, opinion mining, thought mining are being carried out and used to extract political, commercial and cultural meanings [56]. This study focuses on the analysis of textual data by using the contextual approach.

The classification carried out with supervised machine learning algorithms aims to determine which group belongs to a given data after the training process. Textual data must be represented numerically for classifying. This approach, called Bag of Words (BoW), is IR based and assumes that texts can be represented by word numbers in the text [77]. According to Sebastiani [79], in the case of multiple classifications, the entire text can be represented with VSM which based on BoW. Decision trees, Naive-Bayes (NB), Classification and regression trees (CART), Support vector machines (SVM) are the algorithms widely used for classifications [2].

Text mining is a complex process consisting of four steps, in which textual data is converted into structural data and analyzed by data mining approaches. The representation of the relational process steps is given in Figure 1.
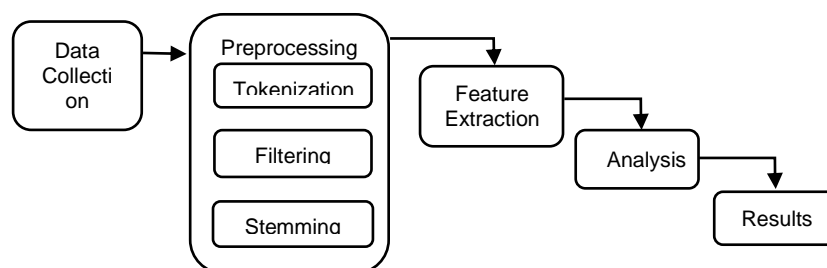

Figure 1: Text mining process

As seen in Figure 1, text mining process has four sequential steps. First step is to collecting raw data for analysis. Second step is to applying NLP methods for preparing data. The feature extraction is to converting text data into structural form in the third step, and the last step is analysis. Explanations of the steps are given below in detail.

2.3.1 Data Collection

Data sources such as documents, web pages, e-mails, patient records are used in the process of collecting non-structural data. Usually, data collection process can be realized manually by the researchers or can be obtained by the support applications provided by the data source such as RSS, API and Web service [97]. In this study, an application was developed by C# programming language to obtain data from a web site that do not provide support services and the obtained non-structural data was transferred to the database.

2.3.2 Data Preprocessing

The data preprocessing that is the most important step in the text mining, influences the success of the analysis result directly. According to Korde and Mahender [43], pre-processing of data preparation is important in classification of textual data as well as algorithms. This process involves tokenization, filtering and stemming. Tokenization is the process of removing the conjunctions and marks (comma, point, exclamation, etc.) from the text. In this context, the conjunction words and marks are placed in a reference series and extraction is applied according to this series. Filtering is NLP-based and is the process of extracting the prefixes and suffixes in the text that are specific to the language. The determining of the root of the word which is called stemming is the most difficult process in data preprocessing [82]. Finding the root, which varies according to the grammatical structure of the language, is also based on the NLP. Finding roots is not applied in some applications due to it takes a long time in analysis [82]. In this case, the text mining is performed with raw data.

51

2.3.3 Feature Extraction and Analysis

Feature extraction is a transformation process of textual data into structural form so that data mining approaches can be applied. In general, counts of the words in the text are extracted using BoW and VSM methods and these values are regarded as the attributes of the document. In other words, the conversion of textual data into structural form is a property extraction process of the text data. If more than one document is compared, the specified document properties should be weighted according to other documents [2], [5]. TF and TF-IDF that are normalization methods, are used for weighting the terms, and similarity values are determined by similarity measures such as Cosine, Manhattan, Gini etc. [78], [87].

One of the most common used feature extraction approaches is Bag of Words (BoW) method which is defined as the series that created by determining the word frequency numbers in a document [2], [29]. The basic idea of the BoW is the most repeated words represent the document [29]. Thus, the document is expressed in numerical form. Another approach for the feature extraction is Vector Space Model (VSM). VSM, which is a method to convert the text data into numerical vectors, is the most common way of representing the text. In this method, the text is represented by a vector formed by the frequency values of each word [2]. In document comparisons, the weight values should be determined according to the other documents. Usually the TF-IDF method is used for calculating the weight values.

TF-IDF method has simple calculation and allows the multiple document comparison [29], however, it has some deficiencies [75]. Imperfect information, called data sparsity in data mining and text mining, is seen as one of the most crucial problem in analyzing textual data [66]. The VSM using TF-IDF, which yields healthy results in the analysis of long texts, is insufficient in the analysis of short text data [75]. For this reason, TF-IDF method will not be the most proper choice for analyzing consumer evaluation that is short text.

2.3.4 Text Classification Algorithms

Classification is one of the most widely used technique in text mining. Machine learning algorithms have been commonly applied for the classifications such as NB, SVM, KNN and RF. As mentioned above, it is agreed in literature that NB and SVM are more successful algorithms in short text classification.

The NB algorithm, which is based on Bayes theorem, is a probabilistic approach for classification problem. It is one of the most commonly used algorithms because it has simple implementation and high classification performance [57], [76]. The NB algorithm which is intensively used in text classification processes has two application methods as multivariate Bernoulli (MvB) and Multinominal (Mn) algorithms [2]. In the MvB method, vectors are created based on the presence or absence of the words in the document. In other words, MvB is a binary property extraction method. In the Mn approach based on BoW, it is assumed that word frequencies will represent the document. In the Mn approach, the similarity is determined by the sum of the probabilities of the terms in the text.

The SVM used in the classification of linear and nonlinear data is one of the widely used algorithm in text classification. SVM, which is a statistical learning model, uses a function called hyperplane to divide data into two parts. The functions that called support vectors show the distance of the data to the hyperplane. Thus, distances between the data are determined [89]. In other words, the hyperplane that determines the distance between the support vectors is used to realize the category separations of data. The SVM is commented as an appropriate choice for text classifications because of the fact that the text data has the huge feature vector, contains sparse data and unrelated features [29].

Classification performance and process time are problematic issues which are extensively investigated in the text classification [37]. The low performance of the algorithms used in text classification led researchers to improve these algorithms, to apply hybrid methods, and to combine various algorithms. However, these methods cause huge increases in processing times [8], [62]. In addition, high probability of missing data in text affects the performance of analysis negatively.

2.4 Related Works on Classification of Short Text

Having insufficient information in the dataset is one of the main problems in text mining process [21], [68], [76]. Two statistical approaches are used to solve the incomplete information problem that reduces efficiency of analysis. The first situation that creates the insufficient information is inadequate number of samples. Some statistical methods such as K-folds cross validation and boosting are used to increase the number of samples for solution. The missing data in the samples is the second situation. Data removing, value estimation and averaging techniques are applied to such samples in case of missing information. Applying the techniques of both increasing of samples and eliminating of missing data mean the changes in values of dataset that called data manipulation. Data manipulation can lead to the divergence of analytical results from actual results, in other words it can cause the deviations in results [90]. The manipulation process can affect the results with higher deviations in text mining due to the text data does not have a certain regular structure.

Various studies that do not contain manipulation in dataset apply several algorithms for analysis. In these studies, it is aimed to determine which algorithm has more effective results or to determine the effects of the algorithm

improvements. Sahami [76], who stated that limited learning caused by insufficient data is one of the main factors affecting the classification, investigated the performances of classifying numerical and textual data by NB algorithm. According to the study, it has been found that the accuracy rate is 94-95% for numerical data classification, while accuracy rates fall in the range of 86-87% in classification of textual data. Sahami's work has revealed that textual data is more influenced by the data sparsity problem and the classification performance is low. Faguo et al. [21] emphasized that the accuracy rates in the classification of short texts are low and generally NB, K-Nearest Neighbor (KNN) and SVM algorithms are used in text classifications. According to researchers, accuracy rates of the algorithms range from 60% to 70%. Pawar and Gawande [68] stated that some studies recommend the SVM algorithm, some studies NB. There is no consensus in the literature on algorithms and that new models should be developed to increase accuracy in short text analysis [68]. Chavan et al. [12] have emphasized that NB and SVM algorithms have been used in the analysis of short texts usually. In recently study, classification of short texts when the insufficient data increases, Wärnling and Bissmark [90] have found that the highest performance decrease was observed in SVM and J48 algorithms, and the least performance decrease was observed in NB. The study emphasizes the use of NB in the classification of short texts because the probability of data sparsity is very high.

Due to the low accuracy rates in the results of short text analysis, the algorithms seem to be inadequate. For the solutions of the problem of low accuracy, some studies suggest hybrid approaches in which several algorithms are combined while some of them suggest the improvements of algorithms by using different techniques. Narayanan, Arora and Bhatia [57], who emphasized that the performance of NB classification can be improved by property selection before analysis, have found that the accuracy rate of 73.7% of the NB classification is increased to 88.8%. In another study, Dino et al. [17] have applied NB and SVM algorithms together in the classification of short texts, have determined that the NB accuracy rate of 81.25% increased to 85.42%. Prabowo and Thelwall [69] showed that ensemble methods which combine several algorithms enhance the classification performances in text categorization. Onan et al. [62] used AdaBoost methods with NB, SVM, LR and Random Forest (RF) algorithms and found that the Bagging-RF algorithm gave the most effective results in keywords extraction performance. When the obtained results of the study are examined, it is seen that the RF accuracy ratio, which is 76.29%, increases to 79.64% with Bagging. This means the increase of about 3.4% in accuracy performance.

Fuzzy Rules (FR) is provided as a different approach in classification of texts. Romero et al. [74] have classified text data by using FR and found that the accuracy of the FR-based classification model was 81%. Kumbhar, Mali and Atique [45] have indicated that FR classification accuracy of 80.99% is increased to 87.8% by supporting genetic algorithms. Besides, they found that this increase was directly related to the number of fuzzy rules. According to the authors, the accuracy rate of the classification has increased by the rise of number of rules. However, the depending on the rise of rules, also the increasing of processing time for the classifications has been observed. In other words, FR can be used in the classification of textual data, but the FR approach is insufficient in the classification of short texts [61]. Khoury, Karray and Kamel [42] stated that meaningful fuzzy rules could not be produced for short texts that contain a few word, and they found that fuzzy approaches did not give meaningful results in short text analysis. In addition, the authors stated that the small number of training data have caused the sparsity problem and large number of training data have caused to produce too many rules which led to long training period [42]. On the other hand, it has been emphasized in recent years that Deep Learning (DL) methods have impressive results in machine learning applications. Especially DL methods have remarkable results in large amount of voice and image data [44], [95]. Jiang et al. [39] stated that DL methods can be applied in text classification problem. The researchers have obtained that DL techniques can be applied in feature learning to eliminate the problems of the high dimensionality [39]. However, DL is major topic in machine learning approach, it has some limitations. Krizhevsky et al. [44] have emphasized that performance of the DL analysis depends on the data as much as possible. It means that since large amount of data has been required for DL analysis, DL is not suitable for analysis that contains limited data. Besides, DL needs some improvements because of long computational time, sophisticate process steps, difficult parameter configurations and not interpretable results [55].

A different approach is noteworthy in studies related to insufficient data. Grey Relational Analysis (GRA) is one of the most common Grey System component used in analysis in case of insufficient information. Most of the studies emphasized the reason of applying GRA is inadequacy of traditional methods due to missing information. So, GRA is seen as a powerful alternative in uncertainty analysis due to incomplete information [19], [27], [71]. In this context, several studies have been performed by using GRA such as performance and efficiency analysis [64], [80], [93], decision making problem [31], [51], [91], [96], selection problem [3], [10], [48], [65], [71], [98]. Surprisingly, The GRA, which is a powerful tool in the analysis of insufficient data, is not seen in short text analysis. Although there are some studies dealing with clustering analysis that used numerical data [11], [24], [40], [64], [92], [99], the GRA-related short text analysis cannot found in classification literature.

The consumer opinions in e-commerce contain inadequate information because it is short text. More consumers express their evaluations in just a few words that means consumer evaluations have higher probability for insufficient information in e-commerce. In this case, data manipulation such as increasing the sample size or removing samples will lead to deviations in the results, as it will affect the meaning of consumer evaluations. For this reason, the classification of short texts should be realized without data manipulations. It is argued that the NB and SVM algorithms are a viable option for short text classification because the accuracy rates are generally higher than the other algorithms [12], [62], [90]. However, the increasing of inadequate data reduces the efficiencies of the

commonly used classification algorithms. This problem that requires more study stands as the main issue to be solved in short text classification.

3 Grey Relational Analysis

The Grey System Theory (GST) was proposed for solving problems with small sample and missing information by Ju Long Deng in 1982 article [16]. This study was accepted as a new approach to producing results in case of uncertainty and incomplete information [53]. The precise unknown information in GST is represented by black, the precise known information by white, and the partial known information between these two end points is represented by grey [52]. The fact that the GST gives effective results at limited information has attracted the interest of the researchers, and many studies have been carried out using the theory. The GST related studies have been published in various disciplines such as agriculture, economy, meteorology, management, medicine, computer science, etc. [41], and generally focused on optimization, prediction and decision making. The performing better results in case of incomplete information are shown as the reason for the using GST in all studies [50], [91], [92], [96]. The GST contains five major parts such as grey relational analysis (GRA), grey decision, grey prediction, grey control and grey programming [91].

The GRA is a relational ranking, classification and decision making method developed by using the GST [24], [25], [52], [53]. The method aims to determine the levels of relation between the reference and each alternatives in a grey system. The relationship levels between the alternatives are called the Grey Relational Degrees [94]. The GRA is used to make the most appropriate selection according to the various criteria among the alternatives. In this regard, the GRA provides the best alternative choice in a multi-criteria decision-making process [35]. So, GRA can be considered as an optimization method.

Although GRA is intensively used in many researches, there are a few studies the using GRA in data mining and text mining. In data mining, using the large amount of data may have caused to researchers not concentrated too much to the missing data. Han et al. [30] emphasized the importance of missing information in data mining for performance of the analysis and they have predicted the missing values by using GRA. Fidan [25] stated that data mining techniques fail in case of limited data and suggested the GRA to resolve the limitations. However, since it removes some of the limitations of the central clustering algorithms, it is possible to attain some studies that use the GRA-based clustering method [24], [40], [92]. Grey Relational Clustering (GRC) has eliminated the center determination and recursive process problems in data mining clustering algorithms such K-means, K-medoids and Fuzzy C-Means. [11], [24], [99].

Text mining is one of the study fields where the missing data affects the analysis negatively [21], [68], [76], [90]. Although some studies suggest some approaches mentioned above to solve the missing data problem in text mining, there is no agreement on certain solution in the literature. In particular, short text data limits the clustering and classification process and reduce performance of analysis. In this context, the use of GRA in text mining studies will provide positive contributions to the literature.

3.1 GRA Calculation Steps

Calculations are performed in six steps to benchmark and rank alternatives in GRA. These steps are; creation of a comparison matrix, creation of a normalization matrix, creation of an absolute value table, creation of a grey relational coefficient matrix, calculation of grey relational degrees [94].

Step 1: Firstly, decision matrix is created with alternatives a and criterions c. The count of alternatives in the decision matrix is denoted by m and the count of criterions is denoted by n. As it is shown in equation (1) and (2), X_a represents each alternatives and X_{ac} represents the value of cth criterion of ath alternative in decision matrix X. For example, the first customer evaluation is denoted as X_1 and the first word in the first customer evaluation is denoted as X_{11}.

$$a = 1,2,3,\dots.m \quad c = 1,2,3,\dots.n \tag{1}$$

$$X = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1n} \\ X_{21} & X_{22} & \cdots & X_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ X_{m1} & X_{m2} & \cdots & X_{mn} \end{bmatrix} \tag{2}$$

Step 2: Based on the values in the decision matrix, reference series is added and comparison matrix is obtained in second step. The values of the reference series to be added can be selected according to the maximum or minimum values of the decision criteria in the alternatives as well as the optimum criterion value that an ideal alternative can obtain. The comparison matrix X_0 is shown in equation (3).

$$X_0 = \begin{bmatrix} X_{01} & X_{02} & \dots & X_{0n} \\ X_{11} & X_{12} & \cdots & X_{1n} \\ X_{21} & X_{22} & \cdots & X_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ X_{m1} & X_{m2} & \cdots & X_{mn} \end{bmatrix} \tag{3}$$

Step 3: The process of reducing data to small intervals is called normalization [84]. After the normalization process, it becomes possible to compare the series with each other, thus it is avoided that discrete values affect the analysis adversely. The normalization method in GRA varies according to whether the criterion values are maximum or minimum. If there is a positive contribution of a desired value of the series (optimal-based), the normalization is performed using the equation (6) and the normalization matrix in the equation (7) is obtained.

$$X_{ac}^* = \frac{X_{ac} - minX_{ac}}{\text{maxX}\,ac - minX_{ac}} \tag{4}$$

$$X_{ac}^* = \frac{maxX_{ac} - X_{ac}}{\text{maxX}\,ac - minX_{ac}} \tag{5}$$

$$X_{ac}^* = \frac{|X_{ac} - X_{0c}|}{\text{maxX}\,ac - X_{0c}} \tag{6}$$

$$X^* = \begin{bmatrix} X_{01}^* & X_{02}^* & \cdots & X_{0n}^* \\ X_{11}^* & X_{12}^* & \cdots & X_{1n}^* \\ X_{21}^* & X_{22}^* & \cdots & X_{2n}^* \\ \vdots & \vdots & \cdots & \vdots \\ X_{m1}^* & X_{m2}^* & \cdots & X_{mn}^* \end{bmatrix} \tag{7}$$

Step 4: The absolute value table shown in equation (9) is generated by using the equation (8) where the absolute difference of each alternative series with the reference series. For example, Δ_{11} value in equation (9) is calculated by absolute subtraction of X_{01}^* and X_{11}^* values in equation (7).

$$\Delta_{ac} = |X_{0c}^* - X_{ac}^*| \tag{8}$$

$$\Delta = \begin{bmatrix} \Delta_{01} & \Delta_{02} & \dots & \Delta_{0n} \\ \Delta_{11} & \Delta_{12} & \cdots & \Delta_{1n} \\ \Delta_{21} & \Delta_{22} & \cdots & \Delta_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ \Delta_{m1} & \Delta_{m2} & \cdots & \Delta_{mn} \end{bmatrix} \tag{9}$$

Step 5: Grey relational coefficients that demonstrate the degree of closeness of the decision matrix to the comparison matrix are calculated by using absolute difference values in equation (9). Equation (12) shows the calculation method of the grey relational coefficients.

$$\Delta_{max} = max_a max_c \Delta_{ac} \tag{10}$$

$$\Delta_{min} = min_a min_c \Delta_{ac} \tag{11}$$

$$\gamma_{ac} = \frac{\Delta_{min} + \rho \Delta_{max}}{\Delta_{ac} + \rho \Delta_{max}} \tag{12}$$

ρ value in equation (12) is called distinguishing coefficient and $\rho \in [0,1]$. It is used for narrow and broaden grey coefficients data intervals [35]. In cases where there is large difference between the data values, p value is chosen close to 0 [94]. In the literature it is emphasized that p value does not affect the ranking of the alternatives and is generally taken as $p = 0.5$ [20].

Step 6: The grey relational degree, which indicates the relationship level between the reference series and the alternatives, can be calculated in two ways. Equation (13) is used to calculate the degrees if the decision criteria have equal importance. In the other case, equation (14) is used [94]. w_{ac} value in equation (14) indicates the weight of the decision criterion. δ_a value calculated by Eq. (13) or (14) is close to 1 indicates that the relationship between the reference series and the alternative is high [53].

$$\delta_a = \frac{1}{n} \sum_{c=1}^{n} \gamma_{ac} \tag{13}$$

55

$$\delta_a = \sum_{j=1}^{n} w_{ac}\gamma_{ac} \qquad (14)$$

4 Materials and Methods

This section contains the aim of the study, research model, brief explanations of the research model and evaluation metrics of classification analysis.

4.1 Purpose and Scope of the Research

The purpose of the research is to analyzing consumer evaluations on e-commerce transactions with text mining and to establishing a GRA based novel approach to classification problem. In this context, the scope of the research is to analyzing the consumer opinions on e-commerce transactional real data which obtained by web mining techniques.

4.2 Research Model

The method of the research performed using GRA in text mining is presented in Figure 2.
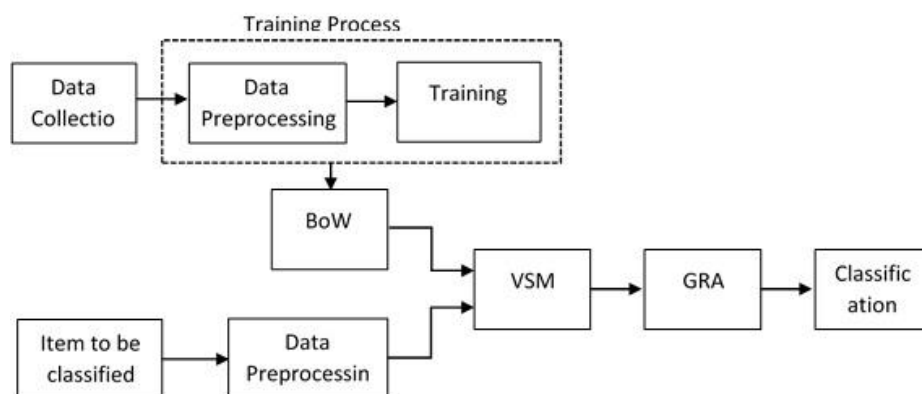

Figure 2: Model of the research

4.3 Dataset

The akakce.com site, which has been operating as an e-commerce price search engine since 2000, uses a forum system in which consumer opinions are included to increase service qualifications. From 2014 onwards, it presents consumer opinions and firm responses to social use. There is no service rating in the forum system and consumer opinions are textual data. The permission of the company regarding the use of the data has been obtained and the dataset has been created with the data by web mining technique from *akakce.com/forum* page. The data retrieved from the forum system in 03.03.2018 by developed C# application was saved in the SqlServer 2008 database. A total of 30625 records were identified including 26321 consumer evaluations and 4304 store responses. 1842 records was removed from the analysis because of having no text data entered by the consumers. Thus, the dataset is created with 24479 consumer evaluation records for classification.

The structure of the data is given in Figure 3. It should be noted that the data used in the research is in Turkish. This is not a limitation for applying the model to the different languages. In other words, the model can be used with the other languages easily. Since data pre-processing involves NLP-based approaches, analysis should be performed by the considering the language of the text. This is not a requirement of the model, but a necessity of text mining.

4.4 Data Preprocessing and Training

Each recorded consumer evaluation includes vendor, consumer, date, transaction, and consumer opinion data. In Figure 3, where the recording structure is given, it can be seen that no sign or delimiter is used to separate the data from each other. Since the analysis will only be carried out within the context of consumer opinions, it has been seen an appropriate approach to splitting according to the space character for obtaining the evaluation data. After deleting the records that would not be included in the analysis, randomly selected 500 consumer evaluations were categorized as positive or negative. 113 consumer opinions identified as negative and pointing to a problem were identified and the problem categories were identified and categorized. The first 10 records from the consumer issues whose problem categories are identified are presented in Table 1.

	id	bilgi
1	1	Satıcı: Hedef VitaminÜlkü Değirmenci02.03.2018Sipariş tarihi: 22.02.20181-29Sipariş No: AA0513313072071445 siparişim ne zaman gelecek
2	2	Mağaza Cevabı 03.03.20182-30Sayın Ülkü Değirmenci, siparişiniz RM 115188 takip numarasıyla aynı gün kargoya verilmiş olup, alıcısı bulunamadı notuyla tekrar bize iade faturası kesilmiş
3	3	Satıcı: Kırmızı Beyaz Eczaneİnci Biber03.03.2018Siparişim zamanında ulaştı.Eksiksiz teslim edildi.Sipariş tarihi: 08.02.20183-31Sipariş No: AA0505503070109777
4	4	Satıcı: BeyazAVM.comMehmet Can03.03.2018Siparişim zamanında ulaştı. Eksiksiz teslim edildi.4-32sıkıntısız bir alıveriş oldu. hem uygun fiyat hem de sorduğum sorulara zamanında cevap

Figure 3: Raw data structure

Table 1: Categories and classifications of the consumer evaluations

Consumer	Evaluations	Considered category	Considered class
1	siparişim ne zaman gelecek *(In English: What time will my order arrive)*	Negative	Delivery
2	bir alıveriş oldu. hem uygun fiyat hem de sorduğum sorulara zamanında cevap verildi *(In English: ? shopping. Both it is affordable and responded to my questions timely)*	Positive	-
3	biri hemen gelmiş olmasına rağmen diğer 2 hafta olmuş olmasına rağmen hala gelmedi. Ürünüm geldiği takdirde geri dönüşte bulunacağım *(In English: Although one of them had arrived immediately, the other still did not come, although it has been 2 weeks. I will be back when the product comes)*	Negative	Delivery
4	kargoya verileli kaç gün oldu fakat elime çok geç ulaştı gidip şubeden aldım yalan söylüyorlar aradım kimse yoktu diyor gelip şubeden alın dediler *(In English: how many days have been courier, but it arrived very late, I went to the cargo branch and they lie, They said there was nobody there)*	Negative	Delivery
5	telefon ile senkronize olmadı İçerisinden garanti belgesi ve kullanam kılavuzuda çıkmadı Kargo dahil satın aldığım halde iade işleminde 20 TL kargo ücreti istendi *(In English: it was not synchronized with the phone. It did not appear in the guarantee document and instruction manual. I have requested a 20 TL shipping fee for the refund process Although I have purchased including the cargo)*	Negative	Product
6	Hiçbir sorun çıkmadı ucuza aldım *(In English: I did not have any problems bought cheap)*	Positive	-
7	baby olarak siparis verdigim halde farklı ürün geldi gelen urun normal sterimar Memnun kalmadım *(In English: I ordered different products as baby the product is normal sterimar Dissatisfied)*	Negative	Product
8	güzel ve ucuz bir alışverişti teşekkürlerrrrrr *(In English: nice and cheap shopping thank you)*	Positive	-
9	Mail mesaj wattsapp yoluyla istedim İletilmedi Anket için bana ulaştılar O zaman istedim Ancak o zaman ilettiler faturayı Bunun dışında bir aksilik yaşamadım *(In English: I wanted invoice via mail message and WhatsApp. I could not have. They contacted me to for the survey. Only then they transmitted the receipt. Other than that, I have not experienced a negation)*	Negative	Support
10	siparişim 4 iş günü olmasına rağmen halen daha tedarik sürecinde gözüküyor *(In English: Although the order is 4 workdays ago, it still seems to be in the process of supply)*	Negative	Delivery

The filtering and the stemming have been applied to consumer evaluations that categorized and classified for training dataset. Zemberek 0.11.1 Turkish NLP library was used to find roots via developed an application in Java and then the word roots were saved in the database. Dataset preprocessing has been completed with the determination of word roots for each evaluation, thus the dataset has been prepared for training process and feature extraction.

4.5   Data Structuring and Classification

The classification is the machine learning process that has been realized by using structural data. For this reason, the text data which is unstructured can be classified by converting into numerical data. BoW and VSM methods are applied for the conversion by using the numbers of word frequency. The classifications can be done after conversion process. However, VSM tables to be created contain many zeros because the word in an item may not be in another item. In other words, unique words in items cause the many zeros in the VSM of the short texts. In the case of inadequate data which is seen in especially limited training data in the analysis of short texts, the classification methods are inadequate, the classification performance decreases and even the analysis cannot be performed.

In the model, the text data was converted to the numerical format and vectored with VSM table using the BoW created after the training process. VSM was created with the words in the item to be classified and the number of the words that are in BoW of the each category. For the solution of the classification problem due to insufficient data, the classification process was carried out by the determining the closest item to the reference.

4.6 Evaluation Methods

Evaluation methods are statistical measurement tools which applied to determine the accuracy of the models [62]. There are two kinds of evaluation methods such as external and internal. If data to be analyzed were labelled previously then external methods are used, otherwise internal methods [60]. Since the classification models have training process in which require labelled data, external evaluation methods are used such Precision (P), Recall (R), F measure (F) and Accuracy-error ratio (Acc).

In the equations below, tp represents the true positives, tn represents the true negatives, fp represents the false positives, fn represents the false negatives and N represents the total number of samples. Precision refers to ratio of correctly classified positives and is calculated as in equation (15). Recall which is shown in equation (16) refers to ratio of actual positives, which is correctly classified, in identified positives. F measure is explained as harmonic mean of P and R values [59] and given in equation (17). Accuracy-error ratio represents the ratio of true classified items in all items and is given in equation (18).

$$P = \frac{tp}{tp + fp} \tag{15}$$

$$R = \frac{tp}{tp + fn} \tag{16}$$

$$F = \frac{2.P.R}{P + R} \tag{17}$$

$$Acc = \frac{tp + tn}{N} \tag{18}$$

5 Experimental Results

Experimental results are presented under three different subtitle as training process findings, vector space model findings and GRA classification findings. The evaluation of the model is presented in the subsection of test process and model accuracy. Finally, comparing classification results of the algorithms are given in this section.

5.1 Training Process Findings

500 randomly selected records from the dataset to be used for the training process were evaluated and it was determined that 113 consumer evaluations contain problems. Thus, the training dataset was formed with 113 negative and 387 positive consumer evaluations. Four class were specified like product, delivery, support and return-exchange related issues in accordance with the literature while the training dataset was prepared [23]. After two learning steps with the training dataset, consumer evaluations were categorized as positive or negative firstly, then in second step negative evaluations were classified. It has been observed that there are 46 complaints regarding product, 51 regarding delivery, 9 regarding support and 7 regarding return-exchange in the negative evaluations. Since the research will only be carried out on negative evaluations, the BoW table of the classified negative evaluations is generated. The BoW table has been shown in Table 2.

Table 2: BoW for training dataset

Classification	Positive Evaluations		Negative Evaluations	
	Number of evaluations	Number of words	Number of evaluations	Number of words
Product	-	-	44	581
Delivery	-	-	47	650
Support	-	-	12	149
Return-exchange	-	-	10	200
Total	387	5644	113	2180

5.2 Vector Space Model Findings

The VSM table was prepared for each negative evaluations by using BoW values of each class. Each column of the VSM belongs to each word in the evaluation. Reference values show the count of the words in the evaluation. The example VSM table prepared for the consumer1 evaluation is given in Table 3.

Table 3: VSM table for customer1 evaluation

	Words of the Evaluation			
	sipariş *(In English: order)*	ne *(In English: what)*	zaman *(In English: time*	gelecek *(In English: arrive*
Reference	1	1	1	1
Product	3	0	2	1
Delivery	19	1	5	3
Support	8	1	3	1
Return exchange	4	1	1	0

Reference values of the customer1 evaluation consist of the words as 1 *sipariş*, 1 *ne*, 1 *zaman* and 1 *gelecek*. The count of the words according to the classes in customer1 evaluation is also shown in the VSM table and the values obtained from BoW dealing with each class. For example, *sipariş* word was mentioned in the BoW 19 times in delivery class and 8 times in support class.

5.3   Grey Relational Classification Findings

The GRA was performed based on the VSM table. The class of the evaluation that constitutes the reference series will be determined by the number of words in the class with the greatest number of words. In other words, the category with the highest number of words in the reference evaluation will be the classifier. The decision matrix added to the reference series is presented in Table 4.

Table 4: Reference series and decision matrix

	sipariş	ne	zaman	gelecek
	max	max	max	max
Reference	19	1	5	3
Product	3	0	2	1
Delivery	19	1	5	3
Support	8	1	3	1
Return-Exchange	4	1	1	0

Normalization was applied to decision matrix in order not to adversely affect the analysis of the value differences between the alternatives. Normalization matrix is shown in Table 5 and table of absolute differences is given in Table 6.

Table 5: Normalization matrix

	sipariş	ne	zaman	gelecek
Reference	1	1	1	1
Product	0	0	0,25	0,3333
Delivery	1	1	1	1
Support	0,3125	1	0,5	0,3333
Return-Exchange	0,0625	1	0	0

Table 6: Absolute differences table

	sipariş	ne	zaman	gelecek
Reference	0	0	0	0
Product	1	1	0,75	0,6667
Delivery	0	0	0	0
Support	0,6875	0	0,5	0,6667
Return-Exchange	0,9375	0	1	1

The grey relational coefficients that determine the level of similarity between the reference series and the classes are given in Table 7. In the calculation of the coefficients, p value is taken as 0.5 according to the literature. The grey relational coefficients are calculated by assuming that the each words in the customer evaluation have equal importance.

Table 7: Grey relational coefficients and Grey relational degrees

	sipariş	ne	zaman	gelecek	Grey Relational Degrees
Reference	0	0	0	0	0
Product	0,3333	0,3333	0,4	0,4286	0,3738
Delivery	1	1	1	1	**1**
Support	0,4211	1	0,5	0,4286	0,5874
Return-Exchange	0,3478	1	0,3333	0,3333	0,5036

59

Table 7 shows that the class with the highest similarity to the reference series is the delivery. The classifications of the first 10 negative evaluations are given in Table 8.

Table 8: Classified consumer evaluations

Customer	Evaluations	Considered class	Determined class
1	siparişim ne zaman gelecek *(In English: What time will my order arrive)*	Delivery	Delivery
2	biri hemen gelmiş olmasına rağmen diğer 2 hafta olmuş olmasına rağmen hala gelmedi. Ürünüm geldiği takdirde geri dönüşte bulunacağım *(In English: Although one of them had arrived immediately, the other still did not come, although it has been 2 weeks. I will be back when the product comes)*	Delivery	Delivery
3	kargoya verileli kaç gün oldu fakat elime çok geç ulaştı gidip şubeden aldım yalan söylüyorlar aradım kimse yoktu diyor gelip şubeden alın dediler *(In English: how many days have been courier, but it arrived very late, I went to the cargo branch and they lie, They said there was nobody there)*	Delivery	Delivery
4	telefon ile senkronize olmadı İçerisinden garanti belgesi ve kullanam kılavuzuda çıkmadı Kargo dahil satın aldığım halde iade işleminde 20 TL kargo ücreti istendi *(In English: it was not synchronized with the phone. It did not appear in the guarantee document and instruction manual. I have requested a 20 TL shipping fee for the refund process Although I have purchased including the cargo)*	Product	Product
5	baby olarak siparis verdigim halde farklı ürün geldi gelen urun normal sterimar Memnun kalmadım *(In English: I ordered different products as baby the product is normal sterimar Dissatisfied)*	Product	Product
6	Mail mesaj wattsapp yoluyla istedim İletilmedi Anket için bana ulaştılar O zaman istedim Ancak o zaman ilettiler faturayı Bunun dışında bir aksilik yaşamadım *(In English: I wanted invoice via mail message and WhatsApp. I could not have. They contacted me to for the survey. Only then they transmitted the receipt. Other than that, I have not experienced a negation)*	Support	Support
7	siparişim 4 iş günü olmasına rağmen halen daha tedarik sürecinde gözüküyor *(In English: Although the order is 4 business days ago, it still seems to be in the process of supply)*	Delivery	Delivery
8	bir hafta beklediler sonra ürün kalmadı yerine başka ürün verelim dediler Derhal paramın iadesini talep ettim 12 gün geçti hâlâ paramı iade etmediler *(In English: I waited a week and then they gave me another product I demanded my money immediately no pay 12 days passed)*	Return-Exchange	Return-Exchange
9	sipariş verdim ve hala bana ulasmadi *(In English: I had ordered and it has not still reached me)*	Delivery	Delivery
10	kullanım tarihi geçmiş ürün gönderdiler *(In English: they have sent expired product)*	Product	Product

Table 8 shows the evaluations will be in which class according to the grey relational degrees. It is seen that the considered classes before training process and determined classes after analysis are the same.

5.4 Test Process and Model Accuracy

The test dataset was prepared before performing model accuracy analysis. Since 113 negative evaluations were used in the training dataset, 113 negative evaluations were randomly selected for the test dataset. The selected negative evaluations for the test process was labeled as product, delivery, support, and return exchange for the considered classification. Thus, the test dataset to be used in the accuracy analysis was created. After applying preprocessing to the test dataset, each evaluation in test dataset and the BoW values obtained from the training dataset were placed in the VSM table, and class is specified by GRA. These operations were realized for the all evaluations in the test dataset respectively and classifications of evaluations in the test dataset have been determined.

Accuracy of the model was performed by the F measure and accuracy-error ratio. High values of both F and Acc indicate the higher accuracy of the classification [62]. The accuracy analysis of the model was realized with using test dataset results are given in Table 9.

Table 9: Model accuracy and error rates

Class	tp	tn	fp	fn	P	R	F	Acc
Product	44	64	1	4	0,97	0,91	0,94	0,95
Delivery	45	62	2	4	0,95	0,91	0,93	0,94
Support	6	99	6	2	0,5	0,75	0,6	0,93
Return-Exchange	6	104	3	0	0,66	1	0,8	0,97

Table 10: Classified evaluations

Class	Number of positive evaluations	Number of negative evaluations	Ratio in negative evaluations (%)
Product	-	4859	50,4
Delivery	-	3673	38,1
Support	-	691	7,2
Return-Exchange	-	414	4,3
Total	14.842	9637	

5.5   Comparing Results

In the literature on the classification of short texts, it is stated that NB and SVM algorithms give successful results. It is seen that these algorithms have been applied in the most of the studies dealing with text analysis. For this reason, the comparison of the results of the developed GRA classification model has been performed with the results of the NB and the SVM algorithms. The NB and the SVM algorithms were applied to training dataset and test dataset which had been used for GRA classification and WEKA 3.8.2 that is a data mining application was used for NB classification and SVM classification. The accuracy ratios obtained from the analysis are given in Table 11.

Table 11: Accuracy rates of algorithms

Class	NB classification						SVM classification						GRA classification					
	tp	tn	fp	fn	F	Acc	tp	tn	fp	fn	F	Acc	tp	tn	fp	fn	F	Acc
Product	29	67	16	1	0,77	0,85	20	62	25	6	0,56	0,72	44	64	1	4	0,94	0,95
Delivery	35	64	12	2	0,83	0,87	23	56	24	10	0,57	0,70	45	62	2	4	0,93	0,94
Support	10	90	2	11	0,6	0,88	3	73	9	28	0,14	0,67	6	99	6	2	0,6	0,93
Ret. Exc.	9	88	0	16	0,53	0,86	5	76	4	28	0,23	0,71	6	104	3	0	0,8	0,97
Average					0,68	0,86					0,37	0,70					0,81	0,94

The results show that the model is a successful approach for short text classification. It is seen in Table 11 that the accuracy rates of the developed GRA classification are higher than the NB and the SVM classification algorithms. It is observed that NB classification performance has higher accuracy rates than the SVM. This determination supports the studies which claim the NB has better results in short text classification such as Onan et al. [62], Warling and Bissmark [90] and Chavan et al. [12]. In addition, the NB accuracy rate is determined 86% in Acc average, which is similar to the findings of Sahami [76].

As seen in Table 11, F measure has lower values then Acc values. Since F measure method represents the true classifications, as well as the false classifications in the calculation, it reveals more realistic results in unbalanced data. Although accuracy values of F measure are less than Acc values, F values have proved that the model has higher accuracy and the GRA classification is a viable choice for short text classification.

On the other hand, it is observed in literature that the performance of short text classifications increase between 3% and 5% by the improvements in algorithms or by the ensemble methods. GRA classification has 94% accuracy for Acc and 81% accuracy for F measure in average and it means that the model has 8-13% performance increase over the NB classification and 24-44% over the SVM. The accuracy rates in Table 11 prove that the GRA classification model has better performance than other algorithms in short text classification.

6   Conclusion

Evaluations of consumers about products, transactional process and services in e-commerce are one of the most important tool in decision-making of the firms. For this reason, classification and interpretation of consumer evaluations will contribute to the development of policies that solve problems in this direction. In this study, it is revealed that consumer evaluations obtained by web mining technique from forum pages of a website can be classified by grey relational analysis based on text mining methods. In this context, consumer evaluations were divided into two groups as positive and negative opinions, then classifications of the consumer problems were realized among negative evaluations.

According to the classification findings of the test dataset, the accuracy rate of the developed model is determined 93-97% due to the Acc and 60-94% due to the F measure. These values show that grey relational classification has a high accuracy level. After the test process, the research dataset that has 14842 positive and 9637 negative evaluations was analyzed. In the classification of negative evaluations, it has been determined that the most problematic issue is related to the product. The product problem was determined in 4859 consumer evaluations. The ratio of the product problem in negative evaluations is 50.4%. In other words, half of the problems experienced in shopping in e-commerce are due to product related issues such as different, damaged, missing part products.

The GRA classification accuracy ratios were compared with the widely used NB and SVM algorithms for short text classification. It was observed that the model has higher performance than the accuracies of other algorithms both F measure and Acc. According to the Acc values it has been found that the GRA classification model has 10% more accuracy than the NB algorithm and 35% more accuracy than the SVM. Besides, according to the F measure values, the proposed model has 23% more accuracy than the NB and 44% more accuracy than the SVM. Both F measure and Acc analysis results show that the GRA classification model is more suitable and viable option for short text classifications.

There are two limitations in the research, namely data collection and data preprocessing processes. After examining the data, it was seen that the consumer opinions were combined with the data such as order number, rating, which was added automatically by the system. This situation that occurs in some consumer evaluations, has led to loss the first words in the evaluations. However, model accuracy ratios of the training dataset shows that the losing of the words does not affect the analysis. Another limitation is insufficiency of NLP library. Zemberek 0.11.1 which is used as a Turkish NLP library, has some deficiencies in the point of finding incorrect word root. However, since the same library is used in both the training dataset used in the learning process and the research dataset root finding process, it was assumed that these deficiencies will not affect the analysis. In this context, Turkish NLP libraries such as Zemberek, Tspell, TurkishNlp, which are still in development process, have crucial importance to give better results in text mining research.

The data used for the analysis is in Turkish. For this reason, a Turkish NLP library for language processing was chosen. However, the fact that the data used in the analysis is in Turkish does not mean that the model cannot be applied to other languages. This is not a restriction but an advantage for languages that have advanced NLP library such as in English. Because, English NLP libraries are more powerful than Turkish NLP libraries. For this reason, the implementation of the study with the dataset in English can reveal clearer results, due to the NLP-based limitations are reduced. Also, the model will not have any problem the analyzing folky terms of the language if the terms can be separable. However, it will be insufficient in finding semantic responses of the folky terms.

In the study, a novel classification model has been improved based on GRA and applied to the real transactional data in ecommerce. It has been determined that the developed model has better result over the NB and the SVM algorithms. The grey relational degrees were used for the classification of consumer evaluations and the classifications were realized by grey relational degrees that have the highest relation with the reference evaluation words. However, some consumer opinions may have content that concerns more than one class. Since grey relational degrees show similarity levels to the reference, the evaluations that can be represented by more than one class can also be analyzed using this method. In this context, the study can be expanded so that more than one class can be generated. Moreover, since the processing time of grey relational degrees is very short and calculation process is not complicated compared to the other approaches, real-time classification systems can be realized easily by using the developed model. However, since the number of columns in the VSM table increases as the text data size increases, it is estimated that the GRA processing time will be long. The prolongation of waiting time is not desirable in real-time systems. So, the model is recommended for short text data classification in real-time applications.

References

[1] F. Alemi, L. Clementz and D. C. Aron, Feasibility of real-time satisfaction surveys through automated analysis of patients unstructured comments and sentiments, Manage Health Care, vol. 21, no. 1, pp. 1-11, 2012.
[2] M. Allahyari, S. Pouriyeh, M. Assefi, S. Safaei, E. D. Trippe, J. B. Gutierrez, and K. Kochut, A brief survey of text mining: Classification, clustering and extraction techniques, in Proceedings in KDD Bigdas, Canada, 2017, pp. 1-13.
[3] V. M. Athawale and S. Chakraborty, A comparative study on the ranking performance of some multi-criteria decision-making methods for industrial robot selection, International Journal of Industrial Engineering Computations, vol. 2, no. 4, pp. 819-830, 2011.
[4] B. Baesens, S. Viaene, D. V. D. Poel, J. Vanthienen, and G. Dedene, Bayesian neural network learning for repeat purchase modelling in direct marketing, European Journal of Operational Research, vol. 138, no. 1, pp. 191-211, 2002.
[5] S. Beliga, A. Mestrovic and M. S. Ipsic, An overview of graph based keyword extraction methods and approaches, Journal of Information and Organizational Science, vol. 39, no. 1, pp. 1-20, 2015.
[6] K. Berezina, A. Bilgihan, C. Cobanoglu, and F. Okumus, Understanding satisfied and dissatisfied hotel customers: Text mining of online hotel reviews, Journal of Hospitality Marketing & Management, vol. 25, no. 1, pp. 1-24, 2015.
[7] A. Berson, S. Smith and K. Thearling, Building Data Mining Applications for CRM. NewYork: McGraw-Hill, 2000.
[8] L. Breiman, Random forests, Machine Learning, vol. 45, no. 1, pp. 5-32, 2001.
[9] S. Chakravorty, Customer relationship management: A content analysis of issues and best practices, Ph. D. dissertation, Florida International University, Florida, 2006.
[10] J. W. K. Chan and T. K. L. Tong, Multi-criteria material selections and end-of-life product strategy: Grey relational analysis approach, Materials & Design, vol. 28, no. 5, pp. 1539-1546, 2007.

[11] K. C. Chang and M. F. Yeh, Grey relational analysis based approach for data clustering, IEEE Proc.-Vis. Image Signal Process, vol. 152, no. 2, pp. 165-172, 2005.

[12] G. S. Chavan, S. Manjare, P. Hegde, and A. Sankhe, A survey of various machine learning techniques for text classification, International Journal of Engineering Trends and Technology (IJETT), vol. 15, no. 6, pp. 288-292, 2014.

[13] K. C. Chen, Text mining e-complaints data from e-auction store with implications for internet, Marketing Research. Journal of Business & Economics Research, vol. 7, no. 5, pp. 15-24, 2009.

[14] R. J. Courtheoux and L. G. Sterk. (2003). In memoriam: Robert D. Kestnbaum. Maa-Online. [Online]. Available: http://maa-online.com/Articles/Kestnbaum.pdf.

[15] K. Coussement and D. V. Poel, Improving customer complaint management by automatic email classification using linguistic style features as predictors, Decision Support Systems, vol. 44, pp. 870-882, 2008.

[16] J. L. Deng, Control problems of grey systems, System and Control Letters, vol. 1, no. 5, pp. 288-294, 1982.

[17] L. H. Dino, V. P. Kallimani and R. R. Kumar, Text documents preprocessing with the Bayes formula for classification using the Support vector machine, IEEE Transactions of Knowledge and Data Engineering, vol. 20, no. 9, pp. 1264-1272, 2008.

[18] S. Dixit and S. Kr, Collaborative analysis of customer feedbacks using Rapid Miner, International Journal of Computer Applications, vol. 142, no. 2, pp. 29-36, 2016.

[19] Y. Dou, Q. Zhu and J. Sarkis, Evaluating green supplier development programs with grey-analytical network process-based methodology, European Journal of Operational Research, vol. 233, pp. 420-431, 2014.

[20] I. Ertugrul, T. Oztas, A. Ozcil and G. Z. Oztas, Grey relational analysis approach in academic performance comparison of university: A case study of Turkish universities, European Scientific Journal, June 2016 special edition, pp. 128-139, 2016.

[21] Z. Faguo, Z. Fan and Y. Bingru, Research on short text classification algorithm based on statistics and rules, in Proceedings Third International Symposium on Electronic Commerce and Security, Guangzhou, 2010, pp. 3-7.

[22] R. Feldman and I. Dagan, Knowledge discovery in textual databases (KDT), in Proceeding of KDD-95, Montreal, 1995, pp. 112-117.

[23] H. Fidan, Asimetrik bilginin elektronik ticaret üzerindeki etkileri: Tüketici güveni üzerine bir uygulama, Ph. D. disertation, Süleyman Demirel University, Institute of Social Science, Isparta, 2014.

[24] H. Fidan, Grey relational clustering analysis of e-commerce customer loyalty, AJIT-e: Online Academic Journal of Information Technology, vol. 9, no. 32, pp. 163-182, 2018.

[25] H. Fidan, Association rules-based gray relational approach for e-commerce recommender system, in Machine Learning Techniques for Improved Business Analytics (G. Dileep Kumar, Ed.). Hershey: IGI Global, 2018, pp. 64-93.

[26] M. Gamon, A. Aue, S. Corston-Oliver and E. Ringger, Pulse: Mining customer opinions from free text, in Proceedings of the International Symposium on Intelligent Data Analysis (IDA), Madrid, 2005, pp. 121-132.

[27] D. Golmohammadi and M. Mellat-Parast, Developing a grey-based decision making model for supplier selection, International Journal of Production Economics, Vol. 137, pp. 191-200, 2012.

[28] D. L. Goodhue, B. H. Wixom and H. J. Watson, Realizing business benefits through CRM: Hitting the right target in the right way, MIS Quarterly Executive, vol. 1, no. 2, pp. 79-94, 2002.

[29] J. Han, M. Kamber and J. Pei, Data Mining Concepts and Techniques. USA: Morgan Kaufmann Publications, 2012.

[30] B. Han, S. Xiao, L. Liu, and Z. Wu, A new methods for filling missing values by grey relational analysis, in Proceedings 2nd International Conference on Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC), Dengleng, 2011, pp. 2721-2724.

[31] X. Hao, An empirical analysis on Guangdong industrial energy intensity based on panel data, in The 19th International Conference on Industrial Engineering and Engineering Management (E. Qi, J. Shen and R. Dou, Eds.). New York: Springer Science & Business Media, 2013, pp. 109-114.

[32] W. He, S. Zhao and L. Li, Social media competitive analysis and text mining: A case study in the pizza industry, International Journal of Information Management, vol. 33, pp. 464-472, 2013.

[33] Z. He, X. Xu, J. Z. Huang, and S. Deng, Mining class outliers: Concepts, algorithms and applications in CRM, Expert Systems with Applications, vol. 27, pp. 681-697, 2004.

[34] G. Hebrail and J. Marsais, Experiments of textual data analysis at electricité de France, in New Approaches in Classification and Data Analysis, Studies in Classification, Data Analysis, and Knowledge Organization (E. Diday, Y. Lechevallier, M. Schader, P. Bertrand and B. Burtschy, Eds.). Berlin, Heidelberg: Springer, 1994, pp. 569-576.

[35] A. Hinduja and M. Pandey, Multicriteria recommender system for life insurance plans based on utility theory, Indian Journal of Science and Technology, vol. 10, no. 14, pp. 1-8, 2017.

[36] M. Hu and B. Liu, Mining and summarizing customer reviews, in Proceedings 10th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), Seattle, 2004, pp. 168-177.

[37] S. Jiang, G. Pang, M. Wu, and L. Kuang, An improved K-nearest-neighbor algorithm for text categorization, Expert Systems with Applications, vol. 39, no. 1, pp. 1503-1509, 2012.

[38] T. Jiang and A. Tuzhilin, Segmenting customers from population to Individuals: Does 1-to-1 keep your customers elers forever, IEEE Transactions on Knowledge and Data Engineering, vol. 18, pp. 1297-1311, 2006.

[39] M. Jiang, Y. Liang, X. Feng, X. Fan, Z. Pei, Y. Xue, and R. Guan, Text classification based on deep belief network and softmax regression, Neural Computing & Application, vol. 29, no. 1, pp. 61-70, 2018.

[40] X. Jin, Grey relational clustering method and its application, The Journal of Grey System, vol. 5, no. 3, pp. 181-188, 1993.

[41] D. Julong, Introduction to grey system theory, The Journal of Grey System, vol. 1, no. 1, pp. 1-24, 1989.
[42] R. Khoury, F. Karray, M. Kamel, A fuzzy classifier for natural language text using automatically-learned fuzzy rules, in Proceedings 2nd International Conference on Artificial & Computational Intelligence for Decision, Control and Automation - International Conference on Machine Intelligence (ACIDCA-ICMI'2005), Tozeur, 2005, pp. 5-7.
[43] V. Korde and C. N. Mahender, Text classification and classifiers: A survey, International Journal of Artificial Intelligence & Applications (IJAIA), vol. 3, no. 2, pp. 85-99, 2012.
[44] A. Krizhevsky, I. Sutskever and G. Hinton, ImageNet classification with deep convolutional neural networks, Neural Information Processing Systems, vol. 25, no. 2, pp. 1106-1114, 2012.
[45] P. Kumbhar, M. Mali, M. Atique, A genetic-fuzzy approach for automatic text categorization, in Proceedings IEEE 7th International Advance Computing Conference, Hyderabad, 2017, pp. 572-578.
[46] T. Y. Lee and E. T. Bradlow, Automated marketing research using online customer reviews, Journal of Marketing Research, vol. 48, no. 5, pp. 881-894, 2011.
[47] S. E. Levy, W. Duan and S. Boo, An analysis of one-star online reviews and responses in the Washington, D.C. lodging market, Cornell Hospitality Quarterly, vol. 54, no. 1, pp. 49-63, 2011.
[48] G. D. Li, D. Yamaguchi and M. Nagai, A grey-based decision-making approach to the supplier selection problem, Mathematical and Computer Modelling, vol. 46, no. 3-4, pp. 573-581, 2007.
[49] H. Li and K. Yamanishi, Mining from open answers in questionnaire data, in Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, 2001, pp. 443-449.
[50] S. Liu, The current developing status on grey system theory, The Journal of Grey System, vol. 19, no. 2, pp. 111-123, 2007.
[51] H. W. Liu, and G. J. Wang, Multi-criteria decision-making methods based on intuitionistic fuzzy sets, European Journal of Operational Research, vol. 179, no. 1, pp. 220-233, 2007.
[52] S. Liu, J. Forrest and Y. Yang, A brief introduction to grey systems theory, Grey Systems: Theory and Application, vol. 2, no. 2, pp. 89-104, 2012.
[53] S. Liu and Y. Lin, Grey Information Theory and Practical Applications. New York: Springer Science+Business Media, 2006.
[54] B. Loh, K. Koo, K. Ho, and R. Idrus. (2011) A review of customer relationship management system benefits and implementation in small and medium enterprises. Semantic Scholar. [Online]. Available: https://pdfs.semanticscholar.org/f0df/d0f7bac330f95d1dd3d87998ab2eed0f8301.pdf
[55] L. Miralles-Pechuan, D. Rosso, F. Jimenez, and J. M. Garcia, A methodology based on deep learning for advert value calculation in CPM, CPC and CPA networks, Soft Computing, vol. 21, no. 3, pp. 651-665, 2017.
[56] A. Montoyo, P. M. Barco and A. Balahur, Subjectivity and sentiment analysis: An overview of the current state of the area and envisaged developments, Decesion Support System, vol. 53, pp. 675-679, 2012.
[57] V. Narayanan, I. Arora and A. Bhatia, Fast and accurate sentiment classification using an enhanced Naive Bayes model, in Proceedings 14th International Conference on Intelligent Data Engineering and Automated Learning - IDEAL 2013 Lecture Notes in Computer Science, Hefei, 2013, pp. 194-201.
[58] T. Nasukawa, and T. Nagano, Text analysis and knowledge mining system, IBM Systems Journal, vol. 40, no. 4, pp. 967-984, 2001.
[59] E. W. Ngai, L. Xiu and D. C. K. Chau, Application of data mining techniques in customer relationship management: A literature review and classification, Expert Systems with Applications, vol. 36, pp. 2592-2602, 2009.
[60] H. Nizam and S. S. Akın, Sosyal medyada makine öğrenmesi ile duygu analizinde dengeli ve dengesiz veri setlerinin performanslarının karşılaştırılması, in Proceedings XIX. Türkiye'de İnternet Konferansı, İzmir, 2014, pp. 129-136.
[61] T. M. Nogueira, S. O. Rezende and H. A. Camargo, On the use of fuzzy rules to text document classification, in Proceedings 10th International Conference on Hybrid Intelligent Systems, Atlanta, 2010, pp. 19-24.
[62] A. Onan, S. Korukoglu, and H. Bulut, Ensemble of keyword extraction methods and classifiers in text classification, Expert Systems With Applications, vol. 57, pp. 232-247, 2016.
[63] E. K. Özyirmidokuz and M. H. Özyirmidokuz, Analyzing customer complaints : A web text mining application, in Proceedings of INTCESS14- International Conference on Education and Social Sciences, Istanbul, 2014, pp. 734-743.
[64] M. Padmaja, and D. Haritha, Software effort estimation using grey relational analysis with K-Means clustering, in Information Systems Design and Intelligent Applications Advances in Intelligent Systems and Computing, vol. 672 (V. Bhateja, B. Nguyen, N. Nguyen, S. Satapathy, and DN Le, Eds.). Singapore: Springer, 2018, pp 924-933.
[65] M. S. Pakkar An integrated approach to grey relational analysis, analytic hierarchy process and data envelopment analysis, Journal of Centrum Cathedra: The Business and Economics Research Journal, vol. 9, no. 1, pp. 71-86, 2016.
[66] W. Pan, E. Zhong and Q. Yang, Transfer learning for text mining, in Mining Text Data (C.C. Aggarwal and C.X. Zhai, Eds.). Boston: Springer, 2012, pp. 223-257.
[67] Y. Park and S. Lee, How to design and utilize online customer center to support new product concept generation, Expert Systems with Applications, vol. 38, no.8, pp. 10638-10647, 2011.
[68] P. Y. Pawar and S. H. Gawande, A comparative study on different types of approaches to text categorization, International Journal of Machine Learning and Computing, vol. 2, no. 4, pp. 423-426, 2012.

[69] R. Prabowo and M. Thelwall, Sentiment analysis: A combined approach, Journal of Informetrics, vol. 3, no. 2, pp. 143-157, 2009.

[70] K. Rababah, H. Mohd and H. Ibrahim, A study of the perception of the benefits of customer relationship management (CRM) system in Malaysian private hospitals, in Proceedings International Conference on Business and Economics Research IACSIT, Malaysia, 2011, pp. 327-330.

[71] R. Rajesh, and V. Ravi, Supplier selection in resilient supply chains: A grey relational analysis approach, Journal of Cleaner Production, vol. 86, pp. 343-359, 2015.

[72] K. Ravi and R. Ravi, A survey on opinion mining and sentiment analysis: Tasks, approaches and applications, Knowledge-Based Systems, vol. 89, pp. 14-46, 2015.

[73] O. Reinhold and R. Alt, Analytical social CRM: Concept and tool support, in Proceedings 24th Bled eConference eFuture: Creating Solutions for the Individual, Organisations and Society, Slovenia, 2011, pp. 226-241.

[74] F. P. Romero, P. J. Iranzo, A. Soto, M. F. Satler, and J. G. Casero, Classifying unlabeled short texts using a fuzzy declarative approach, Language Resources and Evaluation, vol. 47, no. 1, pp 151-178, 2013.

[75] J. Ropero, A. Gomez, A. Carrasco, C. Leon, and J. Lugue, Term weighting for information retrieval, in Fuzzy Logic - Algorithms, Techniques and Implementations (P. D. Elmer, Ed.). London: InTech, 2012, pp. 173-192.

[76] M. Sahami, Learning limited dependence bayesian classifiers, In KDD, vol. 96, pp. 335-338, 1996.

[77] G. Salton, A vector space model for automatic indexing, Information Retrieval and Language Processing, vol. 18, no. 11, pp. 613-620, 1975.

[78] G. Salton and C. Buckley, Term-weighting approaches in automatic text retrieval, Information Processing & Management, vol. 24, no. 5, pp. 513-523, 1988.

[79] F. Sebastiani, Machine learning in automated text categorization, ACM Computing Surveys, vol. 34, no. 1, pp. 1-47, 2002.

[80] Q. Song, M. Shepperd, and C. Mair, Using grey relational analysis to predict software effort with small data set, in Proceedings IEEE International Software Metrics Symposium, Italy, 2005, pp. 10-35.

[81] E. A. Stoica and E. K. Özyirmidokuz, Mining customer feedback documents, International Journal of Knowledge Engineering, vol. 1, no. 1, pp. 68-71, 2015.

[82] F. Tala, A study of stemming effects on information retrieval in Bahasa Indonesia, M.S. thesis, Institute for Logic, Language and Computation, University of Amsterdam, 2003.

[83] C. Taylor. (2018, March). Structured vs. unstructured data. Datamation. [Online]. Available: https://www.datamation.com/big-data/structured-vs-unstructured-data.html.

[84] C. H. Tsai, C. L. Chang and L. Chen, Applying grey relational analysis to the vendor evaluation model, International Journal of the Computer the Internet and Management, vol. 11, no. 3, pp. 45-53, 2003.

[85] D. K. Tse and P. C. Wilton, Models of consumer satisfaction formulation: An extension, Journal of Marketing Research, vol. 24, pp. 204-212, 1988.

[86] V. Tunalı and T. T. Bilgin, Examining the impact of stemming on clustering Turkish texts, in Proceedings Innovations in Intelligent Systems and Applications (INISTA) International Symposium, Trabzon, 2012, pp. 1-4.

[87] P. D. Turney and P. Pantel, From frequency to meaning: Vector space models of semantics, Journal of Artificial Intelligence Research, vol. 37, pp. 141-188, 2010.

[88] A. Tuzhilin, Customer relationship management and web mining: The next frontier, Data Mining and Knowledge Discovery, vol. 24, no. 3, pp. 584-612, 2012.

[89] V. Vapnik, The Nature of Statistical Learning Theory. New York: Springer-Verlag, 1995.

[90] O. Warnling and J. Bissmark. (2017, June). The sparse data problem within classification algorithms. Diva. [Online]. Available: http://www.diva-portal.se/smash/get/diva2:1111045/FULLTEXT01.pdf

[91] G. Wei, Gray relational analysis method for intuitionistic fuzzy multiple attribute decision making, Expert Systems with Applications, Vol. 38, pp. 11671-11677, 2011.

[92] W. Wu, C. Lin, K. Peng, and C. Huang, Applying hierarchical grey relation clustering analysis to geographical information systems-A case study of the hospitals in Taipei city, Expert Systems with Application, Vol. 39, pp. 7247-7254, 2012.

[93] Y. Xiong, and Z. Wang, Study on efficiency evaluation methods of forestry construction project, In Environment, Energy and Sustainable Development (W-P. Sung, J. Kao and R Chen, Eds.). Florida, USA: CRC Press, 2014, pp. 7-13.

[94] B. F. Yıldırım, Gri ilişkisel analiz, in Çok Kriterli Karar Verme Yöntemleri (B. F. Yıldırım and E. Önder, Eds.). Bursa: Dora Basım Yayın, 2015, pp. 229-236.

[95] D. Yu, S. Wang, Z. Karam, and L. Deng, Language recognition using deep-structured conditional random fields, Acoust Speech Signal Process, vol. 41, no. 3, pp. 5030-5033, 2010.

[96] X. Yuan, Grey relational evaluation of financial situation of listed company, Journal of Modern Accounting and Auditing, vol. 3, pp. 41-44, 2007.

[97] K. Zhang, R. Narayanan and A. Choudhary, Voice of the customers: Mining online customer reviews for product feature-based ranking, in Proceedings 3rd Wonference on Online Social Networks (WOSN'10), Boston, 2010, pp. 11-11.

[98] S. Zhang and S. Liu, A GRA-based intuitionistic fuzzy multi-criteria group decision making method for personnel selection, Expert Systems with Applications, vol. 38, no. 9, pp. 11401-11405, 2011.

[99] Q. Zhiming and W. Xiaoli, Application of grey relational clustering and data mining in information extraction, in Proceedings IEEE International Seminar on Business and Information Management (ISBIM) - Wuhan, Hubei, China , 2008, pp. 3-6.