Article

JIS

Journal of Information Science
1–14

DOI: 10.1177/0165551515613226
jis.sagepub.com
SAGE

A feature selection model based on genetic rank aggregation for text sentiment classification

Aytuğ Onan
Celal Bayar University, Turkey

Serdar Korukoğlu
Ege University, Turkey

Abstract

Sentiment analysis is an important research direction of natural language processing, text mining and web mining which aims to extract subjective information in source materials. The main challenge encountered in machine learning method-based sentiment classification is the abundant amount of data available. This amount makes it difficult to train the learning algorithms in a feasible time and degrades the classification accuracy of the built model. Hence, feature selection becomes an essential task in developing robust and efficient classification models whilst reducing the training time. In text mining applications, individual filter-based feature selection methods have been widely utilized owing to their simplicity and relatively high performance. This paper presents an ensemble approach for feature selection, which aggregates the several individual feature lists obtained by the different feature selection methods so that a more robust and efficient feature subset can be obtained. In order to aggregate the individual feature lists, a genetic algorithm has been utilized. Experimental evaluations indicated that the proposed aggregation model is an efficient method and it outperforms individual filter-based feature selection methods on sentiment classification.

Keywords

Feature selection; rank aggregation; sentiment classification

1. Introduction

Sentiment analysis, also called opinion mining, is the process of extracting subjective information, such as opinions, sentiments and attitudes in the source materials towards an entity. It is an interdisciplinary research field that combines tools and techniques from natural language processing, text mining and computational linguistics [1]. Opinions are important factors and influencers of the decision-making process. The determination of people's opinions toward a particular event can be extremely important in several fields, such as management sciences, political science, economics and other disciplines of social sciences [2]. The Web provides a rich and progressively expanding source of information to reach opinions/sentiments regarding a particular topic, product, event or individual.

Sentiment analysis process can be modelled as a classification problem. Sentiment analysis can be conducted at different levels of detail. Based on the levels, sentiment analysis can be broadly divided into three main levels: document-level, sentence-level and aspect-level sentiment analysis [3]. Document-level sentiment classification aims to determine the overall sentiment orientation of an entire document, such as a review text, assuming that each document contains information regarding a single entity. Sentence-level sentiment analysis aims to identify subjective and objective sentences. In sentence-level sentiment analysis, the sentiment orientation of subjective sentences is also identified. The classification of review documents at document or sentence level of granularity does not fully reveal the opinions regarding

Corresponding author:
Aytuğ Onan, Department of Computer Engineering, Celal Bayar University, 45140, Manisa, Turkey.
Email: aytug.onan@cbu.edu.tr

different features of a particular entity. Hence, aspect-level sentiment analysis concerns the classification of sentiments by focusing on the particular features/aspects of entities [3].

The approaches used in sentiment classification can be divided into two distinct groups: machine learning-based and lexicon-based methods. In text and web mining applications, decision tree classifiers, rule-based classifiers, probabilistic classifiers and Naïve Bayes classifiers, linear classifiers, such as Support Vector Machines and neural networks have been widely employed [4]. In order to process the data set of text documents, text features are extracted and selected. Term presence and term frequency, part of speech, opinion words, phrases and negations are typical text-based methods used in sentiment classification for representing the data set [3]. Then, the machine learning algorithm is trained with a labelled data set and a classification model is constructed in a supervised manner. In contrast, lexicon-based sentiment analysis methods determine the sentiment orientation of an entity based on a sentiment lexicon, which consists of sentiment terms collected in advance [3].

The identification of an appropriate representation and the application of an efficient feature selection method are two essential tasks for building a robust model with high classification performance. Text mining is a domain that suffers from the high dimensionality and irrelevancy of text features [4]. Owing to its simplicity and relatively high performance, filter-based feature selection methods, such as information gain, chi-square statistics and mutual information, have been successfully employed in the applications of text mining [3, 4]. In order to represent the text data set, a bag-of-words (BOW) framework has been widely employed. In this framework, a document is represented as a bag of its words with their corresponding frequencies in the document without taking grammar or word order into account. Since a bag-of-words framework has a simple structure, the sentiment data sets used in the experimental evaluations are represented via this framework.

As mentioned above, sentiment analysis is a subfield of text mining and the characteristics of the sentiment analysis data exhibit similar problems to those encountered in text mining data sets. Hence, the feature selection becomes an important research direction for sentiment analysis. Taking these issues into account, this paper presents an ensemble feature selection method for classifying text sentiment data. The proposed method utilizes widely employed individual filter-based feature selection methods, such as information gain (IG), chi-square (CHI), gain ratio (GR), symmetrical uncertainty (SU), Pearson correlation coefficient (PCC), ReliefF algorithm (RL) and probabilistic significance measure (PS) as the base methods. The ranking lists of the individual base methods are then converted into a single ranking list based on a rank aggregation approach with genetic algorithms.

The rest of this paper is organized as follows. Section 2 briefly reviews the existing work on sentiment analysis and rank aggregation-based feature selection. Section 3 presents feature selection methods. Section 4 briefly describes the classification algorithms used in the experimental evaluations. Section 5 presents the proposed genetic rank aggregation-based feature selection model. Section 6 presents the experimental results and Section 7 presents the concluding remarks.

2. Related work

This section briefly reviews the existing literature on feature selection methods in the machine learning-based sentiment/text classification and rank aggregation-based feature selection approaches. Tan and Zhang [5] empirically evaluated the effectiveness of four feature selection methods (document frequency, chi-square statistics, mutual information and information gain) on five learning algorithms (centroid classifier, *K*-nearest neighbour algorithm, winnow classifier, Naïve Bayes classifier and Support Vector Machines) on a Chinese sentiment corpus.

Chen et al. [6] presented two feature selection metrics (multi-class odds ratio and class discriminating measure) for multi-class text classification with Naïve Bayes classifier. The multi-class counterpart of odds ratio extends the measure from binary class domains to the multi-class problems.

Wang et al. [7] presented an improved Fisher's discriminant ratio-based feature selection method for text sentiment classification. The performance of the proposed feature selection method is compared with different feature selection methods with Support Vector Machine classifier.

Mesleh [8] examined the effectiveness of feature selection methods on Arabic text classification. Several feature selection methods, such as chi-square statistics, information gain, mutual information and odds ratio have been evaluated on Support Vector Machines. The experimental results indicate that chi-square statistics often yield better performance for classifying Arabic text documents.

Duric and Song [9] presented a feature selection method for sentiment classification, which concerns subjective expressions using a content and syntax model. The features obtained with the proposed feature selection method yield competitive results with the conventional feature selection methods when the maximum entropy algorithm is utilized as a classifier.

Uysal and Gunal [10] developed a probabilistic feature selection method. The method aims to assign a high score to the distinctive terms, which frequently occur in a particular class and do not occur in the other classes. Similarly, the method aims to assign a low score to the irrelevant terms, which rarely occur in a single class. The terms with high frequency in all classes are assigned a low score, whereas the terms with high frequency in some of the classes are assigned a relatively high score. Based on these requirements, a probabilistic scoring framework is utilized to obtain a ranking of features.

Gunal [11] presented a hybrid feature selection scheme for text classification, which consists of a filter-based and wrapper-based feature selection. The feature selection scheme has been examined in terms of different feature set sizes, data set characteristics, classification algorithms and evaluation metrics. The F_1 measure was utilized in the experimental evaluations. The experimental results demonstrated that the combination of several feature selection methods can obtain a better feature subset than the individual feature selection methods.

Yang and Yu [12] presented a feature selection and sentiment similarity measure for Chinese sentiment data sets. In order to analyse similar micro-blogging accounts, a sentiment similarity measure based on the Karhunen–Loeve transform has been presented. First, information gain-based feature selection was applied to the feature set. To evaluate the different feature sets, C4.5, Support Vector Machine and Naïve Bayes classifiers are utilized.

Sheydai et al. [13] presented an association rule mining-based scheme for text classification. In this scheme, a feature selection was utilized to reduce the dimensionality of text documents. Clustering was utilized to cluster informative features based on their class labels. The performance of the classification scheme was compared with *K*-nearest neighbour, Naïve Bayes and Support Vector Machine classifier. The experimental results yield better performance while using the association rule mining-based classification algorithm for text classification.

Javed et al. [14] presented a two-stage feature selection scheme for text classification. In this scheme, a feature ranking approach, such as information gain or bi-normal separation measure, is applied first. Then, a feature subset selection method, such as the Markov blanket filter, is utilized. Multiple feature rankings obtained by different feature selection methods can be combined by rank aggregation to obtain a more robust feature set.

Jong et al. [15] presented an ensemble feature ranking method that combines the feature rankings obtained by different runs of an evolutionary feature selection approach referred as ROGER (ROC-based genetic learner). The ROGER algorithm obtains a linear combination of features based on the area under the ROC curve.

Prati [16] examined the performance of four rank aggregation methods, namely Borda, Condorcet, Schulze and Markov Chain. In order to obtain feature rankings, information gain, gain ratio, symmetric uncertainty, chi-square, OneR and ReliefF feature selection methods were utilized. Experimental results indicated that Schulze rank aggregation method yields better performance.

Dittman et al. [17] evaluated the performance of rank aggregation methods for ensemble gene selection. As feature ranking methods, 25 feature ranking methods, such as area under the ROC curve, deviance, *F*-measure, geometric mean and Gini index, have been utilized. In order to analyse the effect of different rank aggregation methods, nine rank aggregation methods have been implemented. These methods are enhanced Borda, exponential weighting, highest rank, lowest rank, mean, median, robust rank, Round Robin and stability selection rank aggregation. According to the experimental evaluations, rank aggregation methods generally yield the best performance, whereas the differences among the different rank aggregation methods are not statistically significant.

Bouaguel et al. [18] presented an ensemble feature selection method that combines the Relief algorithm, correlation-based feature selection and information gain measure. In order to obtain a consensus ranked list from the ranked lists of individual feature selection algorithms, a genetic algorithm has been utilized.

Wald et al. [19] examined whether ensemble methods can improve the performance of individual feature selection methods in gene selection. As individual feature selection methods, area under the ROC curve, probability ratio, fold change ratio, signal-to-noise ratio and information gain have been selected. In order to combine these individual methods, a mean rank aggregation method has been utilized. The experimental results indicated that the performance of information gain and fold change ratio methods have been generally enhanced with the inclusion of mean rank aggregation.

Bouaguel et al. [20] presented a rank aggregation-based feature selection method for credit scoring domain. Relief, Pearson correlation coefficient and mutual information methods have been used as the individual feature selection methods. These methods are combined via two different aggregation techniques (majority voting and mean aggregation). The experimental results indicated that aggregation techniques generally yield better performance for credit scoring domain.

Sarkar et al. [21] presented a feature selection method that combines information gain, chi-square and symmetrical uncertainty-based feature selection methods with Borda rank aggregation.

In ensemble feature selection, the way of combining multiple individual feature selection methods is a critical decision, but the determination of efficient components to be included in the feature selection model is also another important issue. In this regard, the proposed ensemble feature selection method has been formulated as a result of extensive

empirical analysis of individual feature selection methods with different combinations of aggregation methods. The contribution of this paper is twofold. In order to enhance the performance of sentiment text classification, the number of works dedicated to the statistical feature selection methods is very limited. Hence, this paper presents an extensive empirical analysis of individual feature selection methods. To our knowledge, this is the first study that combines individual feature selection methods with rank aggregation for the text classification domain.

3. Feature selection methods

Feature selection is the process of obtaining an appropriate feature subset from the data set so that the classification algorithms can deal efficiently with high-dimensional feature spaces. Feature selection methods aim to eliminate irrelevant or redundant features and to reduce the training time required to build a classification model [22]. Feature selection methods can be broadly divided into two groups: filter-based and wrapper-based feature selection methods [23]. Filter-based methods evaluate the merit/usefulness of features based on heuristics/evaluation metrics, whereas wrapper-based methods select the features based on the performance of a machine learning algorithm to optimize the predictive performance. Filter-based feature selection methods are divided into two groups: individual feature measures and group feature measures [24]. Individual feature measures evaluate the merit of features based on a particular evaluation metric. Based on the value of this metric, a ranking of the features is obtained. Group feature measures evaluate the merit of feature subsets. Compared with the group-based measures, individual feature measures are more efficient in terms of running time. This section briefly introduces the individual filter-based measures utilized in the framework.

3.1. Information gain

Information gain-based feature ranking is a filter-based feature selection method, which is widely utilized in text mining domain. For a particular class C, the entropy value of an attribute A is calculated as given by equations (1) and (2) [25]:

$$H(C) = -\sum_{c \in C} p(c)\log_2 p(c) \tag{1}$$

$$H(C|A) = -\sum_{a \in A} p(a) \sum_{c \in C} p(c|a)\log_2 p(c|a) \tag{2}$$

Information gain is used to measure additional information obtained with the existence of a particular attribute A. For each attribute A_i, the information gain between the attribute and the class is determined as given by equation (3) [25]:

$$IG_i = H(C) - H(C|A_i) \tag{3}$$

3.2. Chi-square

Chi-squared feature ranking evaluates the merit of each feature individually with the chi-squared statistical measure. In this measure, the absence of independence between a term t and a category c is examined [26]. The measure is calculated as given by equations (4) and (5) [26]:

$$\chi^2(c, t) = \frac{N \times (AD - BC)}{(A + C)(B + C)(A + B)(C + D)} \tag{4}$$

$$\chi^2_{\max}(t) = \max_i(\chi^2(t, c_i)) \tag{5}$$

where A is the number of times t belongs to c, N is the total number of documents, B is the number of times t occurs without belonging to category c, C is the number of times c occurs without t, D is the number of documents that do not contain both t and c [26]. The measure is calculated for each category and term.

3.3. Gain ratio

Information gain is a biased measure towards features with high values. In order to get rid of this bias, gain ratio evaluates the features by dividing the information gain of the predicted attribute to the entropy of the observed attribute as given by equation (6) [27]:

$$GR = \frac{IG}{H(X)} \tag{6}$$

Gain ratio takes values in [0–1]. The value of one for gain ratio indicates that attribute X can definitely predict the other attribute Y, whereas the value for zero indicates that there is not a meaningful relation between the two attributes.

3.4. Symmetrical uncertainty coefficient

Similar to gain ratio, the symmetrical uncertainty coefficient aims to eliminate the bias of information gain towards the attributes with higher values. Symmetrical uncertainty coefficient is calculated by dividing the information gain by the entropy of the observed and predicted attributes as given by equation (7) [27]:

$$SU = 2 \times \left[\frac{IG}{H(Y) + H(X)} \right] \tag{7}$$

The symmetrical uncertainty coefficient takes values in range of [0–1]. The value of 1 for the gain ratio indicates that attribute X can definitely predict the other attribute Y, whereas the value of 0 indicates that there is not a meaningful relation between the two attributes.

3.5. Pearson correlation coefficient

Pearson correlation coefficient is used to measure the correlation between two attributes as given by equation (8) [28]:

$$R(i) = \left[\frac{\sum_{k=1}^{m} (x_{k,i} - \overline{x_i})(y_k - \overline{y})}{\sqrt{\sum_{k=1}^{m} (x_{k,i} - \overline{x_i})^2 \sum_{k=1}^{m} (y_k - \overline{y})^2}} \right] \tag{8}$$

3.6. ReliefF algorithm

Relief algorithm [29] is a filter-based method that uses a feature relevance criterion to rank the features. Unlike the statistical measures used for ranking the quality of attributes, the Relief algorithm takes contextual information into account. Hence, it can handle properly with the attributes when there is a strong dependency between them [30]. However, the Relief algorithm can only deal with two-class problems. Hence, the ReliefF algorithm [31] was introduced. It is an extension of Relief algorithm to deal with incomplete and noisy data with multi-class problems. The ReliefF algorithm evaluates the merit of an attribute by repeatedly sampling an instance. It can operate on both discrete and continuous data. The algorithm randomly selects an instance R_i. Then it searches for k of its nearest neighbours from the same class (nearest hits H_j) and k of its nearest neighbours from the other classes (misses M_j (C)). For each attribute, the quality estimation is determined based on the nearest hits and misses.

3.7. Probabilistic significance measure

The probabilistic significance measure is a filter-based feature ranking method [32]. For two instances with different class labels, informative features of these two instances should take different values. Based on this assumption, the probabilistic significance measure ranks the attributes based on a two-way function. For each attribute A_i, attribute to class relation (AE) is given by equation (9) and the class to attribute relation (CE) is given by equation (10) [32]:

$$AE(A_i) = \left(1/k \sum_{r=1,2,\ldots,k} \upsilon_i^r \right) - 1.0 \tag{9}$$

$$CE + (A_i) = (1/m) \times \left(\sum_{j=1,2,\ldots,m} \wedge_i^j \right) - 1.0 \tag{10}$$

where k denotes number of different values for a particular attribute, m denotes the total number of classes, j denotes the examined class, ${v_i}^r$ denotes the discrimination power of an attribute and ${\wedge_i}^j$ denotes the discrimination power of a class. For an informative attribute, it is expected to have higher values in terms of both attribute to class and class to attribute relation [32].

4. Classification algorithms

In text classification, decision tree classifiers (such as ID3, C4.5 and C5 algorithms), neural networks, Support Vector Machines, instance-based classifiers (such as *K*-nearest neighbour algorithm) and Bayesian classifiers (such as Naïve Bayes) have been widely employed [4]. These machine learning methods can yield promising results to classify text documents. As emphasized in advance, datasets of text/web mining domain have high-dimensional feature space. Hence, some machine learning methods, such as Support Vector Machines and neural networks, may be costly to train a large data set [33, 34]. Though they are speedy and have a simple structure, Naïve Bayes and *K*-nearest neighbour algorithms can obtain very promising results with appropriate parameter settings [35]. Taking these issues into account, this study takes the Naïve Bayes classifier and *K*-nearest neighbour algorithm as the base classifiers. This section briefly describes these classification algorithms used in the experimental study.

4.1. Naïve Bayes classifier

Naïve Bayes classifier is a statistical classification algorithm that is based on Bayes's theorem. This classifier is based on the assumption of class conditional independence that simplifies the required calculations. Owing to its simple structure, computational efficiency and high predictive performance, the Naïve Bayes algorithm has been successfully employed for text classification tasks. Though it has a simple structure, the algorithm can yield high predictive results, as decision tree classifiers and artificial neural networks [36].

4.2. K-Nearest neighbour algorithm

The *K*-nearest neighbour algorithm is an instance-based classification algorithm. In the algorithm, the classification model is built based on the similarity between the *K* closest training instances of a particular instance. The training instances are represented by *n*-dimensional features and each instance corresponds to a single point in *n*-dimensional space. Each training instance is kept in an *n*-dimensional instance space and the class label for a new instance is determined based on the majority voting of class labels of its *K*-nearest neighbours [37].

5. Proposed genetic rank aggregation-based feature selection model

Filter-based feature selection methods can obtain different rankings for the same data set. The aggregation of several different feature rankings may be beneficial to obtain an enhanced ranking of the features [38]. The proposed method models the feature selection as a rank aggregation problem and the feature rankings obtained by different filter-based methods are combined. Given a set of rankings of the same candidate sets, the rank aggregation seeks to find a single better ranking from these multiple rankings [21]. The rank aggregation problem can be modelled as an optimization problem where the objective is to obtain a final ranking that is the closest to the all individual ranking lists, as given by equation (11) [39]:

$$\Phi(\delta) = \sum_{i=1}^{m} w_i d(\delta, L_i) \tag{11}$$

where δ denotes a list of length k=[L_i], L_i, ith ranked list, w_i denotes the weight associated with the list and d denotes the distance function among the examined lists. The optimization problem seeks to find an optimal list (δ*) that minimizes the total distance to the existing lists according to equation (12) [39]:

$$\delta^* = \arg\min \sum_{i=1}^{m} w_i d(\delta, L_i) \tag{12}$$

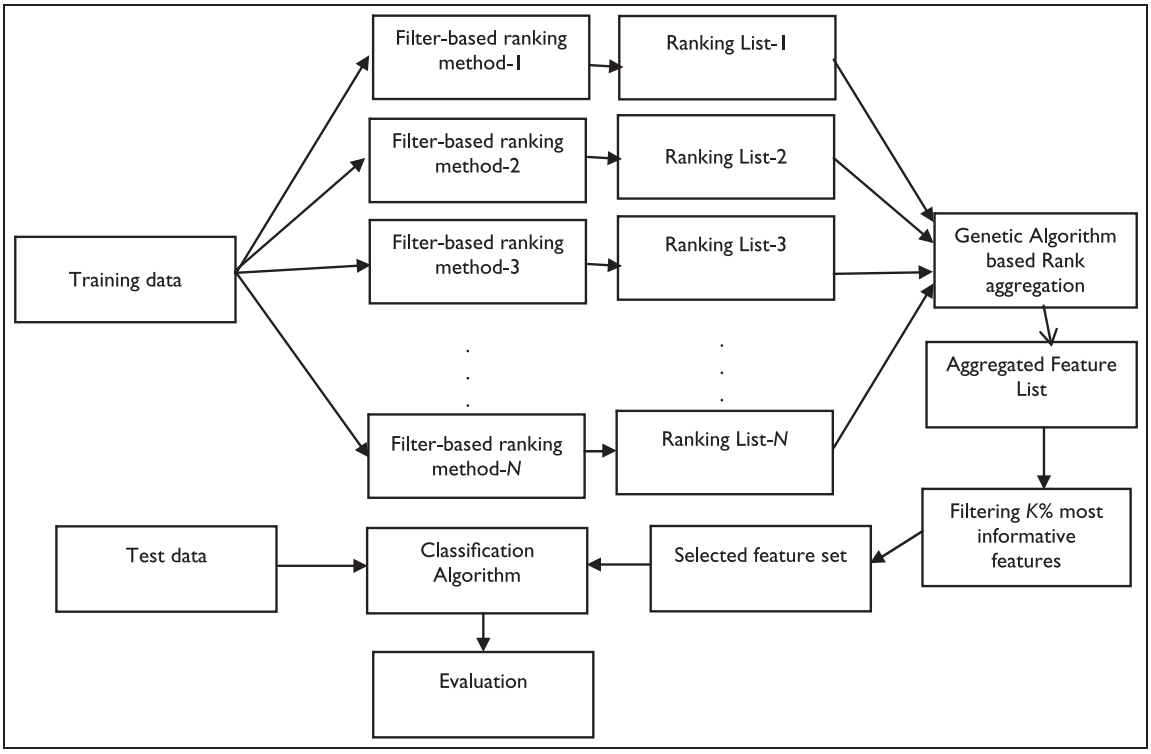

Figure 1. The feature selection framework (GA).

In this minimization procedure, several different distance functions, such as Spearman foot rule distance and Kendall's tau distance, may be utilized. The rank aggregation framework presented here utilizes the Spearman foot rule distance, which is calculated as follows [39]:

$$S(\delta, L_i) = \sum_{t \in L_i \cup \delta} \left| r^{\delta}(t) - r^{L_i}(t) \right| \tag{13}$$

where $r^{\delta}(t)$ and $r^{L}{}_{i}$ (t) denotes the ranks of a particular candidate t in the two examined lists.

The general framework of the feature selection model is presented in Figure 1. First, feature ranking lists are obtained based on seven different filter-based feature selection methods, namely information gain, chi-square, gain ratio, symmetrical uncertainty coefficient, Pearson's correlation coefficient, ReliefF algorithm and probabilistic significance measure. For the features with the same values of calculated metrics, the same rank position is given. For the last features with the lowest metric values, the rank positions are taken as the last position. In other words, let *A, B* and *C* be the three features with information gain value of zero and let there be 10 features in total. Then, the rank of these three attributes are regarded as 10 (the last position) since each shares the same value for the metric. The brute force solutions to the rank aggregation problem may be computationally infeasible for the lists with even a small number of elements. Hence, it is appropriate to use a metaheuristic algorithm to solve the problem [40]. The framework presented here utilizes the genetic algorithm in conjunction with the feature selection methods. In order to represent the chromosomes in the population, a path representation method is utilized. In this representation, each chromosome is represented by a permutation of length n, such that the position of a particular element represents the ranking of the element. The genetic algorithm utilizes Spearman foot rule distance as a fitness function, as given by equation (13). Population size parameter is determined based on the values outlined in Aledo et al. [40]. Tournament selection (k=2) is utilized as the selection mechanism. In this selection, k individuals are randomly selected from the population and the selected individuals are evaluated based on their fitness values. Among the individuals, the one with the highest fitness value is selected. In each generation, $k.n$ chromosome pairs are randomly selected and the best individuals (with the lowest fitness value) are selected. The crossover and mutation mechanisms are utilized to obtain offspring from the individuals. Since the individuals are represented by the path representation, the classical crossover and mutation operators cannot be utilized [41]. In the path representation, several different crossover mechanisms, such as partially mapped crossover, order crossover, order-based

Table 1. Descriptive information for the data sets [44, 45]

Dataset	Positive class label	Negative class label	Number of features
Camera	250	248	1352
Camp	402	402	2045
Doctor	739	739	1578
Drug	401	401	1438
Laptop	88	88	2010
Lawyer	110	110	2474
Music	291	291	1398
Radio	502	502	1923
TV	235	235	2834

crossover and position-based crossover, can be utilized [41]. A recent empirical analysis of the utilization of genetic algorithms [40] in rank aggregation indicates that the most efficient crossover and mutation methods for the problem are position-based crossover [42] and insertion mutation operator [43]. Hence, these two operators are utilized in the genetic algorithm.

6. Experimental design and results

6.1. Datasets

In order to evaluate the effectiveness of the proposed feature selection framework, we have used nine well-known sentiment analysis data sets from different domains, namely Camera, Camp, Doctor, Drug, Laptop, Music, Lawyer, Radio and TV data sets [44]. The Camera data set contains digital camera evaluations extracted from Amazon.com; the Camp data set contains evaluations for summer camps extracted from CampRatingz.com; the Doctor data set contains doctor evaluations retrieved from RateMDs.com; the Drug data set is obtained from DrugrRantingz.com; the Laptop and Music data sets are also obtained from Amazon.com; the Lawyer data set is obtained from LawyerRatingz.com; the Radio data set is obtained from RadioRatingz.com; and the TV data set is obtained from TVRatingz.com. Except for the Camera data set, sentiment analysis data sets have equal numbers of instances with positive and negative labels. The descriptive information regarding the data sets is summarized in Table 1, where the number of features indicates the feature numbers with unigram data representation. Since the BOW framework has a simple structure and a common use for sentiment analysis, we have adopted this scheme for feature representation [46]. In addition, the features are represented via a term-frequency model with unigram features owing to the efficiency of this configuration in the sentiment classification [47].

6.2. Evaluation measures

To evaluate the predictive performance of feature selection methods, classification accuracy (ACC) and F-measure values are utilized. Classification accuracy (ACC) is the proportion of true positives and true negatives obtained by the classification algorithm over the total number of instances as given by the following equation (14):

$$ACC = \frac{TN + TP}{TP + FP + FN + TN} \tag{14}$$

where TN, TP, FP and FN represent the number of true negatives, the number of true positives, the number of false positives and the number of false negatives, respectively. Precision (PRE) is the proportion of the true positives against the true positives and false positives as given by equation (15):

$$PRE = \frac{TP}{TP + FP} \tag{15}$$

Recall (REC) is the proportion of the true positives against the true positives and false negatives as given by equation (16):

$$REC = \frac{TP}{TP + FN} \tag{16}$$

The F-measure takes values between 0 and 1. It is the harmonic mean of precision and recall as determined by equation (17):

$$F-\text{measure} = \frac{2 * PRE * REC}{PRE + REC} \tag{17}$$

6.3. Experimental procedure

In the experiments, a 10-fold cross-validation procedure is utilized. In this way, the original data set is randomly divided into 10 equal-sized subsamples. For each time, a single subsample is used for validation, whereas the rest are kept for training. The process is repeated 10 times and average results are reported. In the experimental results, filter-based feature selection and classification algorithms are performed with the machine learning toolkit WEKA (Waikato Environment for Knowledge Analysis) version 3.7.11, which is an open-source platform that contains many machine learning methods implemented in Java. The proposed feature selection framework is also implemented in Java. For the classifiers, the default parameters of WEKA are utilized. In order to evaluate the effectiveness of proposed feature selection scheme, the performances of seven different filter-based methods (IG, CHI, GR, SU, PCC, RL and PS) are taken into account. In order to obtain a single ranking from different ranked lists, a number of rank aggregation methods have been presented in the literature. Among the earlier works presented in Section 2, the most extensive analysis of rank aggregation techniques is given in Dittman et al. [17]. Hence, we have adopted this comparison framework to evaluate our genetic rank aggregation model. In this framework, mean (MA), median (MEA), highest rank (HRA), lowest rank (LRA), stability selection (SSA), exponential weighting (EWA), enhanced Borda (EBA), Round Robin (RRA) and robust rank aggregation (RORA) methods are utilized. Mean aggregation computes the mean value of the feature's rank among the combined ranked lists and the mean value is used as a value for the feature. Similarly, median aggregation computes the median value of the feature and assigns it as a value for the feature. In the highest rank method, the best (the smallest) rank of the feature among the combined ranked lists is selected, whereas the worst (the largest) value is assigned in the lowest rank aggregation [17]. In stability selection, a threshold value is used to determine the value of each feature. For each list, whether the examined feature exceeds the threshold value or not is determined and based on this judgement, a position is assigned to the feature [48]. The exponential weighting method extends the stability selection by assigning points based on $e^{-r/s}$, where r denotes the rank of feature and s denotes the threshold value. The enhanced Borda rank aggregation method obtains a ranking value for each attribute by multiplying each attribute's Borda count by its stability selection score [48]. The Round Robin rank aggregation method starts with assigning random order values to the different lists. The features of the lists are combined into a single list by selecting the elements of lists based on order values of lists and the rank of attributes in each list [17]. The Robust rank aggregation method obtains a probabilistic model for features that assigns significance scores to keep only statistically relevant features in the combined list. In this model, features that are ranked better compared with the null hypothesis of uncorrelated inputs are given a significance score [49]. In addition to the individual feature selection methods, the results obtained by rank aggregation methods are also presented. In feature ranking methods, the number of attributes to be kept in the data set is an essential parameter of determination. In this experiment, we adopted the idea of changing the number of features proportional to the total number of features in each data set. Hence, different proportions ranging from 10 to 90% are taken into account in the experimental evaluation. The predictive performance and usefulness of each feature subset are evaluated by Naïve Bayes and K-nearest neighbour algorithms in terms of classification accuracy and F-measure.

6.4. Results and discussion

The experimental procedure is repeated on the nine public sentiment analysis data sets mentioned above. In Tables 2 and 3, average (mean) classification accuracies for different proportions of features (10–90%) for different filter-based feature selection methods, namely IG, CHI, GR, SU, PCC, RL and PS and different rank aggregation methods, have been presented with Naïve Bayes and K-nearest neighbour algorithms, respectively. In the tables, the best results achieved for each classification algorithm are indicated using bold type. As it can be observed from Tables 2 and 3, the proposed genetic algorithm-based framework, which obtains an aggregated feature list based on the individual filter-based feature selection methods, outperforms the base filter-based feature selection methods. The highest predictive performance on the Naïve Bayes classifier is 94.71%, which is achieved by genetic algorithm-based feature aggregation framework (GA)

Table 2. Classification accuracies with Naïve Bayes algorithm for the data sets

		10%	20%	30%	40%	50%	60%	70%	80%	90%
Individual methods	IG	86.92	88.33	87.00	87.25	84.82	85.03	84.47	83.63	82.69
	CHI	86.65	87.81	87.25	85.97	84.86	85.13	84.56	83.72	82.66
	PCC	86.90	89.72	90.79	91.44	89.82	90.90	90.10	88.71	86.53
	Gain	83.45	86.49	87.24	87.12	84.79	85.06	84.61	83.73	82.70
	RL	64.57	68.77	72.52	75.59	76.80	79.87	80.58	80.92	81.51
	PS	81.80	85.62	87.14	85.95	84.86	85.07	84.58	83.74	82.83
	SU	85.95	87.54	87.25	86.63	85.66	84.96	83.88	83.54	82.69
Aggregation methods	MA	88.42	89.03	84.85	82.95	84.82	89.99	88.98	89.94	89.39
	MEA	87.48	88.89	88.41	83.00	84.83	88.98	89.79	88.99	89.41
	HRA	89.48	89.88	87.23	87.49	89.84	89.83	87.81	89.84	89.88
	LRA	88.38	88.10	84.98	83.98	88.39	89.90	89.71	89.00	89.44
	SSA	89.13	88.94	73.94	80.78	83.43	88.73	88.94	88.90	87.50
	EWA	82.03	88.38	84.75	83.04	84.41	89.93	89.99	88.95	89.45
	EBA	89.99	89.48	84.94	82.43	83.43	89.79	89.09	89.91	89.39
	RRA	89.09	87.41	89.48	89.29	88.14	88.48	88.97	87.00	88.48
	RORA	89.49	89.93	87.23	87.60	87.03	87.48	89.75	89.00	87.26
	GA	90.17	92.72	93.50	94.50	94.15	**94.71**	94.45	93.80	92.94

Table 3. Classification accuracies with *K*-nearest neighbour algorithm for the data sets

		10%	20%	30%	40%	50%	60%	70%	80%	90%
Individual methods	IG	66.66	64.71	61.82	64.38	60.41	61.93	61.85	61.70	60.35
	CHI	67.14	64.37	61.73	60.54	60.94	61.67	61.90	61.68	60.54
	PCC	65.77	64.30	63.85	61.96	61.47	62.21	61.95	61.88	60.60
	Gain	71.80	68.22	62.69	61.28	60.63	61.60	61.79	61.41	60.12
	RL	56.33	57.51	56.21	56.69	55.91	59.04	59.12	59.56	59.34
	PS	71.56	67.62	62.48	60.58	60.72	61.55	61.80	61.34	60.52
	SU	66.06	63.39	62.28	61.48	60.24	61.77	61.87	61.71	60.35
Aggregation methods	MA	63.32	79.45	77.83	62.80	76.07	77.72	76.68	77.72	77.80
	MEA	67.87	79.76	77.71	77.38	76.78	76.95	76.72	77.90	77.17
	HRA	69.07	62.91	61.42	63.03	60.04	61.27	79.65	77.72	79.77
	LRA	79.08	67.77	78.50	74.81	76.27	76.90	76.40	77.72	77.07
	SSA	76.82	77.78	77.44	77.07	76.17	77.63	77.48	77.72	74.63
	EWA	78.99	66.71	78.07	77.31	76.27	76.94	76.41	77.78	76.02
	EBA	66.34	78.32	62.26	76.21	77.89	76.27	76.77	77.74	77.80
	RRA	87.23	87.84	84.07	87.69	87.44	86.17	86.28	86.39	84.82
	RORA	87.34	87.06	86.38	83.69	86.11	86.40	87.84	84.94	86.49
	GA	90.87	**92.02**	90.94	90.48	90.26	90.96	90.91	90.95	90.54

when 60% of the most informative features in the aggregated feature list are selected. The highest predictive performance on the *K*-nearest algorithm is 92.02%, which is achieved by GA-based feature aggregation framework when 20% of the most informative features in the aggregated feature list are selected. Amongst the base filter-based feature selection methods, the best (the highest) predictive performance on Naïve Bayes classifier has been achieved by PCC-based feature selection for 40% of the most informative features in the measure-based ranking. The classification accuracy for this scheme is 91.44%. The results of PCC-based feature selection for other numbers of features often outperform the other base filter-based feature selection methods for the Naïve Bayes classifier. For the *K*-nearest neighbour algorithm, the best (the highest) predictive performance amongst the base filter-based feature selection methods has been achieved by GR-based feature selection when 10% of the most informative features are selected. The classification accuracy obtained by this combination is 71.80%. Another point of consideration is the performance of rank aggregation methods over individual filter-based feature selection algorithms. As it can be observed from Tables 2 and 3, aggregation methods generally yield better results compared with the individual filter-based methods. There are several configurations where rank aggregation methods outperform the individual methods. For the Naïve Bayes algorithm, the best rank aggregation performance is achieved by the proposed genetic algorithm-based approach. This is followed by highest rank

Table 4. *F*-Measure values with Naïve Bayes algorithm for the data sets

		10%	20%	30%	40%	50%	60%	70%	80%	90%
Individual methods	IG	0.87	0.88	0.87	0.87	0.85	0.85	0.85	0.84	0.82
	CHI	0.87	0.88	0.88	0.86	0.86	0.86	0.85	0.84	0.82
	PCC	0.87	0.90	0.91	0.91	0.90	0.91	0.90	0.89	0.86
	Gain	0.84	0.87	0.87	0.87	0.86	0.86	0.85	0.84	0.83
	RL	0.66	0.69	0.73	0.77	0.79	0.81	0.81	0.81	0.82
	PS	0.83	0.87	0.87	0.86	0.86	0.86	0.85	0.84	0.83
	SU	0.86	0.88	0.87	0.87	0.86	0.85	0.84	0.84	0.82
Aggregation methods	MA	0.87	0.88	0.84	0.82	0.85	0.90	0.89	0.90	0.90
	MEA	0.87	0.88	0.88	0.82	0.84	0.89	0.89	0.89	0.90
	HRA	0.87	0.90	0.87	0.87	0.90	0.89	0.87	0.90	0.90
	LRA	0.84	0.87	0.85	0.84	0.88	0.89	0.88	0.89	0.89
	SSA	0.93	0.88	0.75	0.81	0.83	0.88	0.89	0.89	0.87
	EWA	0.83	0.87	0.85	0.83	0.84	0.89	0.90	0.89	0.89
	EBA	0.86	0.88	0.85	0.82	0.83	0.89	0.90	0.89	0.89
	RRA	0.89	0.87	0.89	0.89	0.88	0.88	0.89	0.87	0.88
	RORA	0.89	0.88	0.87	0.88	0.87	0.87	0.90	0.89	0.87
	GA	0.90	0.93	0.93	0.94	0.94	**0.95**	0.94	0.94	0.93

Table 5. *F*-Measure values with *K*-nearest neighbour algorithm for the data sets

		10%	20%	30%	40%	50%	60%	70%	80%	90%
Individual methods	IG	0.66	0.65	0.60	0.64	0.63	0.64	0.61	0.59	0.57
	CHI	0.66	0.65	0.60	0.62	0.64	0.63	0.61	0.59	0.57
	PCC	0.63	0.61	0.61	0.59	0.61	0.59	0.59	0.60	0.57
	Gain	0.69	0.69	0.61	0.61	0.63	0.64	0.61	0.59	0.56
	RL	0.55	0.52	0.54	0.58	0.61	0.61	0.58	0.56	0.55
	PS	0.68	0.68	0.61	0.61	0.63	0.63	0.61	0.59	0.57
	SU	0.65	0.63	0.60	0.62	0.62	0.63	0.61	0.59	0.57
Aggregation methods	MA	0.63	0.79	0.78	0.63	0.76	0.77	0.77	0.78	0.78
	MEA	0.68	0.80	0.78	0.77	0.77	0.77	0.77	0.78	0.77
	HRA	0.69	0.63	0.61	0.63	0.60	0.61	0.80	0.78	0.80
	LRA	0.79	0.68	0.78	0.75	0.76	0.77	0.76	0.78	0.77
	SSA	0.77	0.78	0.77	0.77	0.76	0.78	0.77	0.78	0.75
	EWA	0.79	0.67	0.78	0.77	0.76	0.77	0.76	0.78	0.76
	EBA	0.66	0.78	0.62	0.76	0.78	0.76	0.77	0.78	0.78
	RRA	0.87	0.88	0.84	0.88	0.87	0.86	0.86	0.86	0.85
	RORA	0.87	0.87	0.86	0.84	0.86	0.86	0.88	0.85	0.86
	GA	0.91	**0.92**	0.91	0.91	0.90	0.91	0.91	0.91	0.91

aggregation, Round Robin rank aggregation and robust rank aggregation in respective order. For the *K*-nearest neighbour algorithm, the best rank aggregation performance is again achieved by the proposed genetic algorithm-based approach. The second and third best performance results are achieved by Round Robin rank aggregation and robust rank aggregation, respectively.

In Tables 4 and 5, average (mean) *F*-measure values for different proportions of features (10–90%) for different filter-based feature selection methods, namely IG, CHI, GR, SU, PCC, RL and PS and different rank aggregation methods have been presented with Naïve Bayes and *K*-nearest neighbour algorithms, respectively. Similar to the classification accuracies listed in Tables 2 and 3, the proposed GA-based feature aggregation framework outperforms the base filter-based feature selection (ranking) methods. The highest *F*-measure results are also achieved with the same configurations of GA-based feature aggregation and Naïve Bayes with 60% of the most informative features and GA-based feature selection-based aggregation and *K*-nearest neighbour algorithm with 20% of the most informative features.

In Figures 2 and 3, average accuracy rate comparisons for the individual feature selection methods and the rank aggregation methods with different numbers of features on Naïve Bayes and *K*-nearest neighbour algorithms are presented, respectively. In the figures, the *x*- and *y*-axes are, respectively, the proportions of features kept in the data set and classification accuracy rates. The figures clearly depict that the predictive performance of the proposed GA-based feature

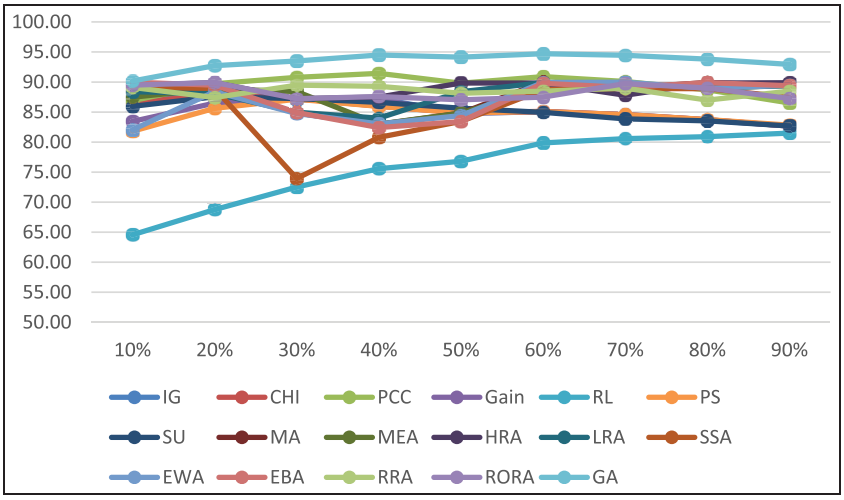

Figure 2. Average accuracy rate comparisons for feature selection algorithms with Naïve Bayes classifier.

aggregation framework yields better predictive performance on all data set configurations for Naïve Bayes and *K*-nearest neighbour algorithms. The individual (base) filter-based methods, such as RL, exhibit highly changing patterns based on the different numbers of features selected, whereas the proposed genetic algorithm-based feature selection method exhibits an approximately flat pattern. This indicates the robustness and efficiency of the proposed genetic algorithm-based feature aggregation method over the individual filter-based approaches.

As it can be observed from the results presented in Tables 2 and 3, the Naïve Bayes algorithm generally outperforms the classification performance of the *K*-nearest neighbour algorithm for text sentiment classification. A recent paper on sentiment classification [46] examines the classification performance of different classification algorithms and data representations for text sentiment classification. The results indicates that the Naïve Bayes and radial basis function networks yield better performance compared with Support Vector Machines, C4.5 and *K*-nearest neighbour algorithm for text sentiment classification. This may be explained by the intuitive structure of the Naïve Bayes algorithm with a limited number of parameters to tune. In contrast to Naïve Bayes, the *K*-nearest neighbour algorithm involves several parameters, such as the number of nearest neighbours, distance functions and weighting function. These parameters may greatly affect the performance of the algorithm, which is kept constant during the experimental analysis [50].

7. Conclusion

Sentiment analysis is an important research direction of web mining and text mining. Social data mining is a domain where the processing of huge amounts of data in an efficient manner is a substantial research direction. Feature selection is the process of selecting an appropriate feature subset in classification model construction so that readability of the classification model can be enhanced, the training time required to train the learning algorithm can be reduced and the generalization ability can be enhanced whilst eliminating overfitting. In addition, the selection of an appropriate feature subset in classification model construction can yield better predictive performance. Owing to its speed and high predictive performance, individual filter-based feature methods have been widely employed for text and web mining. Hence, this paper presents an ensemble approach that integrates the individual feature ranking lists obtained by different filter-based feature selection methods. The presented framework obtains individual feature rankings with information gain, chi-square, gain ratio, symmetrical uncertainty, Pearson correlation coefficient, ReliefF algorithm and probabilistic significance measure-based feature selection methods. Then, these individual lists are amalgamated into a single ranking list via a genetic algorithm. The experimental results are evaluated on nine public sentiment analysis data sets from various domains. In the evaluation, Naïve Bayes and *K*-nearest neighbour algorithms are used as the learning algorithms. Classification accuracy and *F*-measure are utilized as the evaluation metrics. The experimental results indicate that the proposed genetic algorithm-based feature selection scheme is an efficient method for sentiment analysis, since it can effectively process feature space while achieving high classification accuracy. In the experimental evaluation, the highest (best) average classification accuracy is 94.71% which is obtained when Naïve Bayes classifier is utilized in conjunction with the proposed genetic algorithm-based feature selection scheme and 60% of the most informative features in the aggregated list are selected.

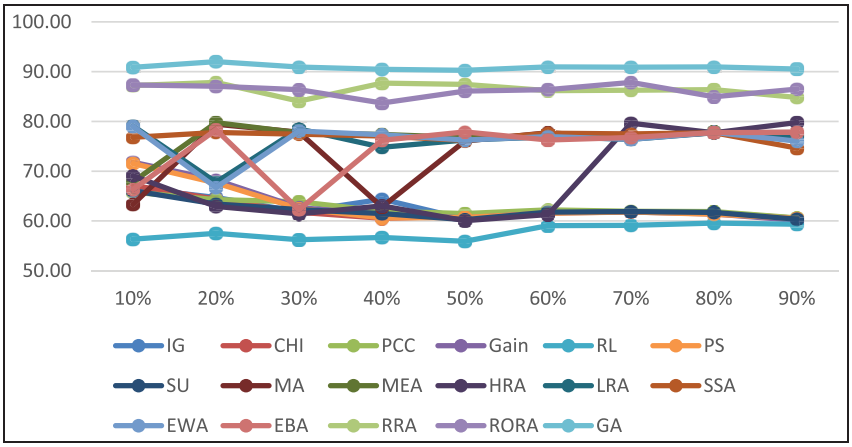

Figure 3. Average accuracy rate comparisons for feature selection algorithms with *K*-nearest neighbour algorithm.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

[1] Wikipedia. Sentiment Analysis, http://en.wikipedia.org/wiki/Sentiment_analysis (accessed 23 July 2015).
[2] Liu B. *Sentiment analysis and opinion mining*, 1st edn. California: Morgan & Claypool, May 2012, p. 167.
[3] Medhata W, Hassan A and Korashy H. Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal* 2014; 5(4): 1093–1113.
[4] Aggarwal CC and Zhai CX. A survey of text classification algorithms. In: Aggarwal CC and Zhai CX (eds) *Mining text data*, 1st edn. Berlin: Springer, 2012, pp. 77–128.
[5] Tan S and Zhang J. An empirical study of sentiment analysis for Chinese documents. *Expert Systems with Applications* 2008; 34: 2622–2629.
[6] Chen J, Huang H, Tian S and Qu Y. Feature selection for text classification with Naïve Bayes. *Expert Systems with Applications* 2009; 36: 5432–5435.
[7] Wang S, Li D, Song X, Wei Y and Li H. A feature selection method based on improved fisher's discriminant ratio for text sentiment classification. *Expert Systems with Applications* 2011; 38: 8696–8702.
[8] Mesleh AM. Feature sub-set selection metrics for Arabic text classification. *Pattern Recognition Letters* 2011; 32(14): 1922–1929.
[9] Duric A and Song F. Feature selection for sentiment analysis based on content and syntax models. *Decision Support Systems* 2012; 53(4): 704–711.
[10] Uysal AK and Gunal S. A novel probabilistic feature selection method for text classification. *Knowledge-Based Systems* 2012; 36: 226–235.
[11] Gunal S. Hybrid feature selection for text classification. *Turkish Journal of Electrical Engineering and Computer Science* 2012; 20(S2): 1296–1311.
[12] Yang DH and Yu G. A method of feature selection and sentiment similarity for Chinese micro-blogs. *Journal of Information Science* 2013; 39(4): 429–441.
[13] Sheydai N, Saraee M and Shahgholian A. A novel feature selection method for text classification using association rules and clustering. *Journal of Information Science* 2015; 41(1): 3–15.
[14] Javed K, Maruf S and Babri HA. A two-stage Markov blanket based feature selection algorithm for text classification. *Neurocomputing* 2015; 157: 91–104.
[15] Jong K, Mary J, Cornuejols A, Marchiori E and Sebag M. *Ensemble feature ranking*. Lecture Notes in Computer Science, Vol. 3202. Berlin: Springer, 2004, pp. 267–278.
[16] Prati RC. Combining feature ranking algorithms through rank aggregation. In: *Proceedings of international joint conference on neural networks*, 2012, pp. 1–8.
[17] Dittman DJ, Khoshgoftaar TM, Wald R and Napolitano A. Classification performance of rank aggregation techniques for ensemble gene selection. In: *Proceedings of the twenty-sixth international FLAIRS conference*, 2012, pp. 420–425.
[18] Bouaguel W, Brahim AB and Limam M. Feature selection by rank aggregation and genetic algorithms. In: *Proceedings of the international conference on knowledge discovery and information retrieval and the international conference on knowledge management and information sharing*, 2013, pp. 74–81.

[19] Wald R, Khoshgoftaar TM and Wald R. Ensemble gene selection versus single gene selection: Which is better? In: *Proceedings of the twenty-sixth international FLAIRS conference*, 2012, pp. 350–355.
[20] Bouaguel W, Mufti GB and Limam M. *Rank aggregation for filter feature selection in credit scoring*. Lecture Notes in Computer Science, Vol. 8284. Berlin: Springer, 2013, pp. 7–15.
[21] Sarkar C, Cooley S and Srivastava J. Robust feature selection technique using rank aggregation. *Applied Artificial Intelligence* 2014; 28(3): 243–257.
[22] Dash M and Liu H. Feature selection for classification. *Intelligent Data Analysis* 1997; 1: 131–156.
[23] Chandrashekar G and Sahin F. A survey on feature selection methods. *Computers and Electrical Engineering* 2014; 40: 16–28.
[24] Diao R. Feature selection with harmony search and its applications. PhD Thesis, Aberystwyth University, 2014.
[25] Hall MA and Holmes G. Benchmarking attribute selection techniques for discrete class data mining. *IEEE Transactions on Knowledge and Data Engineering* 2003; 15(6):1437–1447.
[26] Yang Y and Pedersen JO. A comparative study on feature selection in text categorization. In: *Proceedings of the fourteenth international conference on machine learning*, 1997, pp. 412–420.
[27] Hall MA and Smith LA. Feature selection for machine learning: Comparing a correlation-based filter approach to the wrapper. In: *Proceedings of the twelfth international florida artificial intelligence research society conference*, 1999, pp. 235–239.
[28] Guyon I and Elisseeff A. An introduction to variable and feature selection. *Journal of Machine Learning Research* 2003; 3: 1157–1182.
[29] Kira K and Rendell LA. A practical approach to feature selection. In: *Proceedings of international conference on machine learning*. San Francisco, CA: Morgan Kaufmann, 1992, pp. 249–256.
[30] Robnik-Sikonja M and Kononenko I. Theoretical and empirical analysis of ReliefF and RReliefF. *Machine Learning* 2003; 53: 23–69.
[31] Konenenko I. Estimating attriutes: Analysis and extension of Relief. In: *Proceedings of European conference on machine learning*. Berlin: Springer, 1994, pp. 171–182.
[32] Ahmad A and Dey L. A feature selection technique for classificatory analysis. *Pattern Recognition Letters* 2005; 26: 43–56.
[33] Colas F and Brazdil P. Comparison of SVM and some older classification algorithms in text classification tasks. In: Bramer M (ed.) *Artificial intelligence in theory and practice*. Berlin: Springer, 2006, pp. 169–178.
[34] Zhan W and Yifan H. A Comparison among three neural networks for text classification. In: *Proceedings of the 8th international conference on signal processing*, Beijing, 16–20 November 2006, pp. 1–4. New York: IEEE.
[35] Colas FPR. Data mining scenarios for the discovery of subtypes and the comparison of algorithms. PhD Thesis, Leiden University, 2009.
[36] Han J and Kamber M. *Data mining: Concepts and techniques*, 2nd edn. San Francisco, CA: Morgan Kaufmann, March 2006, p. 800.
[37] Aha DW, Kitler D and Albert MK. Instance-based learning algorithm. *Machine Learning* 1991; 6: 37–66.
[38] Dwork C, Kumar R, Naor M and Sivakumar D. Rank aggregation methods for the Web. In: *Proceedings of the 10th international World Wide Web conference*. New York: ACM, 2001, pp. 613–622.
[39] Pihur V, Datta S and Datta S. Weighted rank aggregation of cluster validation measures: A Monte Carlo cross-entropy approach. *Bioinformatics* 2007; 23(13): 1607–1615.
[40] Aledo JA, Gamez JA and Molina D. Tackling the rank aggregation problem with evolutionary algorithms. *Applied Mathematics and Computation* 2013; 222: 632–644.
[41] Larranaga P, Kuijpers CMH, Murga RH, Inza I and Dizdarevic S. Genetic algorithms for the travelling salesman problem: A review of representations and operators. *Artificial Intelligence Review* 1999; 13:129–170.
[42] Syswerda G. Schedule optimization using genetic algorithms. In: Davis L (ed.) *Handbook of genetic algorithms*. New York: Van Nostrand Reinhold, 1991, pp. 332–349.
[43] Fogel D. An evolutionary approach to the traveling salesman problem. *Biological Cybernetics* 1988; 60(2):139–144.
[44] Whitehead M and Yaeger L. Building a general purpose cross-domain sentiment mining model. In: *Proceedings of the world congress on computer science and information engineering*. New York: IEEE, 2009, pp. 472–476.
[45] Wang G, Sun J, Ma J, Xu K and Gu J. Sentiment classification: The contribution of ensemble learning. *Decision Support Systems* 2014; 57: 77–93.
[46] Onan A and Korukoğlu S. Ensemble methods for opinion mining. In: *Proceedings of the 23th signal processing and communications applications conference*. New York: IEEE, 2015, pp. 212–215.
[47] Pang B and Lee L. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval* 2008; 2(1–2): 1–135.
[48] Haury AC, Gestraud P and Vert JP. The influence of feature selection methods on accuracy, stability and interpretability of molecular signatures. *Plos One* 2011; 6(12): e28210.
[49] Kolde R, Laur S, Adler P and Vilo J. Robust rank aggregation for gene list integration and meta-analysis. *Bioinformatics* 2012; 28(4): 573–580.
[50] Batista GEAPA and Silva DF. How *k*-nearest neighbor parameters affect its performance. In: *Proceedings of the Argentine symposium on artificial intelligence*. 2009, pp. 1–12.