ICIC Express Letters
Volume **15**, Number **9**, September **2021**

ICIC International ©2021 ISSN 1881-803X
pp. 951–960

TWO PHASE WEIGHTED CORRELATION FEATURE SELECTION APPROACH FOR TEXT CLASSIFICATION USING MULTILABEL CLASSIFIERS

Bhavana Raghunath Bhamare[1,2,*] and Prabhu Jeyanthi[3]

[1]Department of Computer Science and Engineering
[3]Department of Information Technology
Sathyabama Institute of Science and Technology
Jeppiaar Nagar, Rajiv Gandhi Salai, Chennai, Tamilnadu 600119, India
[*]Corresponding author: kanawadebr@gmail.com; bhavanak@isquareit.edu.in

[2]Department of Information Technology
International Institute of Information Technology (I[2]IT)
Hinjawadi, Pune 411057, India

Received December 2020; accepted March 2021

Abstract. *In text categorization, the classification performance depends on the features used for the training of the classifier and the classifiers used. Many existing approaches consider text classification as a single label classification problem and use binary or multiclass classifiers for it. However, a review instance may belong to multiple topics simultaneously. Therefore, the objectives of this research are to propose a feature selection strategy and to use a multilabel classification approach to evaluate its performance. This study proposes a feature selection approach named weighted correlation feature selection (WCFS), in line with the first objective, which emphasizes on feature relevance and avoids redundancy among them. In this supervised approach, relevant features are selected using the weight of each feature and the redundancy among features is avoided by calculating the correlation between features. The second objective is to evaluate the system's performance using multilabel classifier. The proposed method has gained significant results compared to other methods with 0.779 Micro F1 score using binary relevance – support vector machine (BR-SVM) classifier.*

Keywords: Information gain (IG), Gini index (GI), Chi-square, Multilabel classification, Correlation

1. **Introduction.** As a substantial amount of data is available on the Internet, there is a need for an automated system to categorize such unstructured text. This electronic text is available in the form of blog contents, news feeds, review contents, and Twitter or Facebook posts, etc. To process such data, there is a need for automated text categorization techniques. Text categorization has many applications like sentiment analysis, spam detection, document summarization, and context detection. It also has many applications in the medical field. For textual data, the features are the words in a sentence. In this, the accuracy of the classification model depends on the feature extraction, selection strategies and the classifiers used. Different feature extraction strategies are used in earlier works like N-grams, parts of speech (POS) tag based features, grammatical rule-based features and context-based features [1].

For selecting the relevant features, feature selection strategies are used. The term frequency (TF) and the term frequency-inverse document frequency (TF-IDF) are the common methods used in many research works. The techniques of feature selection (FS) can be widely divided into supervised and unsupervised. Some of the supervised FS approaches are Filter and Wrapper techniques. Wrapper models form subsets of features

DOI: 10.24507/icicel.15.09.951

from the whole collection of features. They choose features heuristically by calculating the accuracy of the machine learning algorithm trained on the chosen subset of features. However, if there are a large number of features, wrapper approaches are computationally inefficient and infeasible. Methods for wrappers may attempt to over fit small training data. On the other hand, filter methods use a mathematical formula to pick features. Correlation, Chi-square, Gini index, and information gain are some of the filter methods. Filter strategies are independent of the machine learning algorithm. In this article, for aspect extraction, we rely on filter approach.

Feature selection techniques:

Chi-square:

The higher Chi-square value shows that the feature is more dependent on the class, so such features can be selected for training classifiers [1]. Chi-square statistic is defined in Equation (1):

$$x^2(f,c) = \left[\frac{R \times (AD - CB)^2}{(A+C)(B+D)(A+B)(C+D)}\right] \tag{1}$$

A is the co-occurrence count of class c and feature f. B is the number of times f occurs without c. C is the number of times c occurs without f. D is the number of times neither f nor c occurs. R is the number of records.

Information gain (IG):

IG is a strategy which is useful to decide how much the feature is informative about the class [2]. IG is computed as in Equation (2). $P(C_j)$ is the probability of class C_j, k is the number of classes, $P(C_j|f)$ is the conditional probability of C_j in presence of f and $P\left(C_j|\bar{f}\right)$ is the probability of C_j in absence of f.

$$\begin{aligned} InfoGain(f) = &-\sum_{j=1}^{k} P(C_j) \log P(C_j) + P(f) \sum_{j=1}^{k} P(C_j|f) \log P(C_j|f) \\ &+ P\left(\bar{f}\right) \sum_{j=1}^{k} P\left(C_j|\bar{f}\right) \log P\left(C_j|\bar{f}\right) \end{aligned} \tag{2}$$

Classification:

The classification problems fall into two categories like one-label and multi-label classification problems. Under one label, it includes two class and multiclass problems that predict one class label only. However, a single instance in a training dataset may contain multiple labels. To handle such classification problems, the multilabel classification approach is used. In this, the classifier predicts multiple labels if existing for a single instance. Multilabel classification problem can be handled by applying the problem transformation strategies, adapted algorithm, and ensemble strategies [3].

The problem transformation strategy transforms the multilabel problem into a single label problem. In the adapted algorithms, the single label classifiers are adapted for multilabel classification like multi-label – k nearest neighbor (ML-KNN). In ensemble approach, a set of multiclass classifiers or a set of multilabel classifiers can be used to create a multilabel ensemble classifier. In this research, we applied the transformation strategy. The classifiers used for classification along with transformation strategy are support vector machine (SVM), Naïve Bayes (NB), and random forest (RF). The problem transformation can be done by applying binary relevance (BR), classifier chain (CC), and label powerset (LP) [4]. The different problem transformation approaches are briefed below.

Binary relevance (BR): In this approach, an ensemble of binary classifiers is trained, one for each label. Each classifier predicts class belongingness as yes or no. The union of all predicted classes is considered as a label of the given test instance. In this, the dependency between labels is ignored.

Classifier chains (CC): In CC, a chain of classifiers is generated where the output of previous classifiers is given as input features to the next one. For the training example (X, T), the CC train classifiers in the form of chains as $(X, t_1), ([X, t_1], t_2), ([X, t_1, t_2], t_3)$. This approach considers the label dependencies.

Label powerset (LP): This method considers the correlations between labels L and creates a powerset of labels. The number of classifiers is $2^{|L|}$. The powerset generated for a dataset with 3 labels is $(\{000\}, \{100\}, \{010\}, \{001\}, \{110\}, \{011\}, \{101\}, \{111\})$.

The multilabel training data is represented as in Equation (3). Here, t_1 to t_k are labels and x_1 to x_N are features.

$$\textit{Training dataset representation} = \begin{bmatrix} t_1^1 & t_2^1 & \dots & t_k^1 & x_1^1 & x_2^1 & \dots & x_N^1 \\ t_1^2 & t_2^2 & \dots & t_k^2 & x_1^2 & x_2^2 & \dots & x_N^2 \\ t_1^n & t_2^n & \dots & t_k^n & x_1^n & x_2^n & \dots & x_N^n \end{bmatrix} \quad (3)$$

The two objectives of this research are 1) to propose a feature selection approach; 2) to analyze the performance of aspect prediction using multilabel classification. The dataset used for this research is SemEval 2014 restaurant review dataset. A review instance contains one or more aspect categories like food, service, ambience, and price. The focus of the proposed system is to identify the aspects of a given review sentence. In this system, from each review sentence, features are extracted using term frequency (TF). Here, the value of TF is 3. The features are selected using a two-phase process. In the first phase of feature selection, a weight is calculated for each feature and the feature having the maximum weight is selected. In the second phase, the remaining features in each aspect category are selected by considering the correlation among features. The weight supports to select relevant features and correlation avoids redundancy. Many earlier works used single label classifiers like multi-class classifiers to predict the aspect label for a review instance [5]. This problem is overcome in the proposed system by applying a multilabel classifier for aspect prediction.

The paper is organized as follows. Section 1 presents an introduction. Section 2 reviews related work. The proposed system is elaborated in Section 3. Experimental results are discussed in Section 4 that is followed by conclusion.

2. Related Work.

Feature selection plays a vital role in applications like sentiment analysis, text classification, aspect identification, recommender system, and medical field. It helps to reduce the feature search space and feature dimensionality, which in turn improves classification speed and accuracy. Cai et al. [6] published a review on various supervised, semi-supervised, and unsupervised FS methods. A two-step feature selection (FS) strategy of univariate feature selection followed by feature clustering was proposed in [7]. The univariate FS technique was used here to decrease the search space and then clustering was applied to picking typically independent sets of features. With extensive review and analysis of more than 13 datasets, the authors demonstrated the efficiency of the suggested methodology. It shows that the performance gain obtained using Naive Bayes is significant. Deng et al. [8], Sarkar and Goswami [9] analyzed and assessed the efficiency of certain FS techniques. Sarkar and Goswami [9] analyzed the performance of the system on Chi-square, IG, mutual information (MI), symmetrical uncertainty methods using classifiers such as NB, SVM, and decision tree (DT). On various text-based datasets, the authors presented the findings of FS techniques.

A feature ranking metric known as the relative discrimination criterion was proposed in [10]. By considering the difference between the document frequency of a feature in positive and negative documents along with the area under curve (AUC) value, the relevance of a feature was determined. Ren et al. [11] proposed a framework to address some problems known as a robust learning framework. A robust learning framework (RLF) unifies the area under curve maximization, outlier detection, and FS.

Bahassine et al. [12] suggested an enhanced classification system for the Arabic text that uses the enhanced Chi-square FS to boost the classification performance. The combination of the ImpCHI technique and the SVM classifier works best in terms of F-measure, precision, and recall according to experimental research. A strategy for solving dimensionality problems was proposed in [13] by integrating clustering with a correlation metric to create a subset of good features. Initially, by using the k-means clustering process, unrelated features were terminated and then selected the appropriate features by using correlation measures from each cluster. Hu et al. [14] implemented a new FS approach known as dynamic relevance and joint mutual information maximization (DRJMIM) to fix two concerns of previous systems, such as determining the relevance of the feature without taking consideration of the relevance of the candidate feature and the chosen feature, eliminating any interdependent features as redundant features. In this research, the redundant features are paired with a dynamic weight to eliminate the probability of true redundant features to be picked. Hancer et al. [15] proposed new criteria known as a filter criterion inspired from ReliefF, mutual information, and Fisher Score concepts. This method selected highest-ranked features based on Fisher Score and ReliefF. This approach performed better than mutual information feature selection approach in single-objective and multi-objective differential evolution frameworks.

Lee et al. [16] implemented the new S-C4.5-SMOTE (synthetic minority over-sampling technique) bagging algorithm, which uses wrapper FS to support clinical decision-making. This approach solves the issue of data distortion by keeping the dataset balanced. Liu et al. [17] suggested an FS strategy to classify the text based on an independent feature search space. The proposed technique was known as relative document – term frequency difference (RDTFD), which aims to isolate the features in all content archives into two feature sets. The proposed method focuses on improving the high-class association of the features and diminishing the relationship between the features and reduces the search space. A new wrapper FS algorithm for classification problems, such as a hybrid genetic algorithm (GA) and extreme learning machine-based feature selection algorithm, was introduced in [18]. In this approach, the GA was used to pick the feature subsets from the feature space. The set of these subsets was further used to train the ensemble classifiers to predict accuracy.

Madasu and Elango [19] focused on the performance of various FS methods for sentiment analysis (SA). The TF-IDF was used as the feature extraction technique to generate the feature vocabulary. To pick the best feature collection from the feature vocabulary, the authors did experiments on different FS techniques. Selected features were used to train numerous machine learning classifiers and ensemble classifiers. This analysis proves that ensemble classifiers outperform the other classifiers and neural network approaches. Kermani et al. [20] suggested a new hybrid FS approach based on global PMI (pointwise mutual information). This hybrid approach was the combination of filter and wrapper approaches using IG in the first phase as a filter approach and PMI based FS as a wrapper approach in the second phase. Schouten et al. [21] used an ontology-based approach for aspect extraction and sentiment classification on SemEval 2015 restaurant review dataset. In this method, the multiclass classifier was used for aspect category prediction and for each aspect category one SVM classifier was trained. The features considered for training classifier were presence and absence of lemmatized word, ontology concept and the Word-Net synsets.

The study in [23] considers the label oriented score of a text with the assigned label. The text in the training set was divided as text assigned to a label and text unassigned to a label. Here, the frequency of text assigned to a label was compared with the text which was unassigned to a label. If this score is more, then the text belongs to a subset of text assigned to the label.

Many earlier methodologies as discussed in this section focused on the filter, wrapper and hybrid approaches of feature selection along with multiclass classifiers. Very few approaches considered redundancy among features. The proposed system handles the problem of redundant features and uses multilabel classifier for classification to avoid the drawback of using multiclass classifier.

3. **Proposed System.** Figure 1 shows the proposed system architecture. This system is tested for aspect prediction function on the restaurant review dataset. The dataset used is the SemEval restaurant review dataset [22]. Each review in the dataset is labelled with one or more aspect categories like food, service, ambience, price and miscellaneous. The dataset contains 3044 training sentences and 13% of the sentences have more than 2 aspect categories. Earlier approaches considered it as a single label problem and used multiclass classifiers to predict a label. This drawback is avoided in the proposed system. The proposed methodology uses multi-label classifiers to predict class labels of a review. The performance of multi-label classifiers is boosted by applying a two-phase feature selection approach.

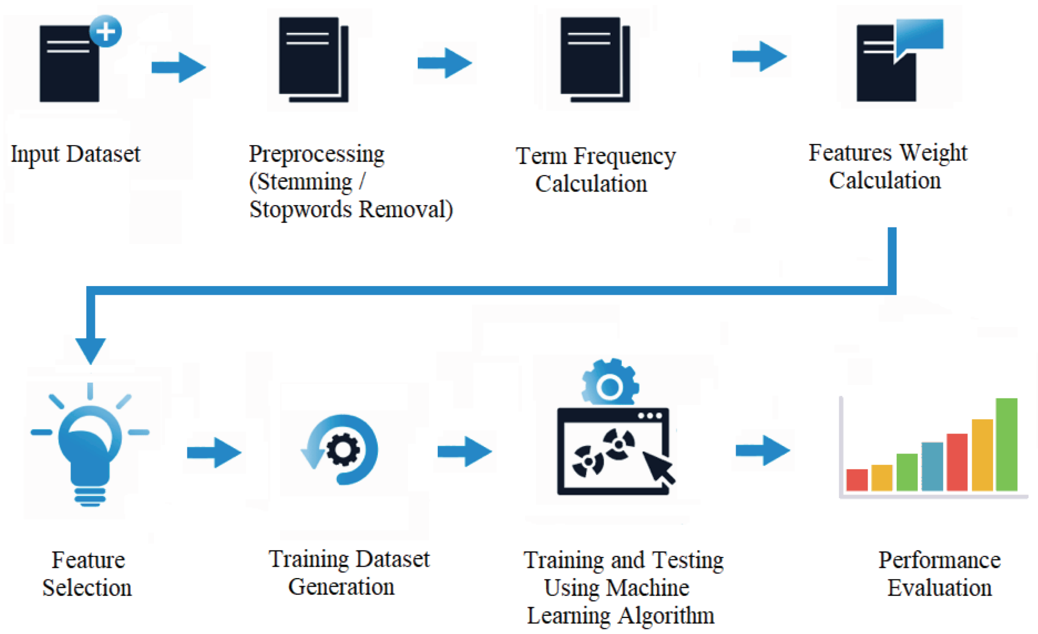

Figure 1. The proposed system architecture

The working flow of the system is as follows.

Step 1: Stemming is applied after pre-processing for each review sentence. Further, stop words are removed.

Step 2: For each feature, the term frequency is calculated in an aspect category and in a dataset. Features having a term count greater than 3 are extracted.

Step 3: Weight of each feature f in each aspect category c is calculated using Equation (4).

$$W_{fc} = \frac{Numebr\ of\ times\ f\ \ occurring\ in\ category\ c}{occurance\ count\ of\ f\ \ in\ a\ dataset} \tag{4}$$

After Step 3, the features in all categories are sorted in decreasing order of weight.

Step 4: Feature selection method.

Feature selection:

Feature selection is done by different approaches like information gain, Chi-square statistic, and the proposed two-phase weighted correlation feature selection (WCFS) approach. The performance of these feature selection strategies is compared using multilabel classifiers like BR, CC, and LP.

Two phase weighted correlation feature selection (WCFS):

This is a two-phase feature selection approach. WCFS is used to select features from each category. The proportion of how many features to select from each category is decided from Equations (5)-(7). In the first phase, in each category, the feature with maximum weight is selected. To select the next term, the correlation of each non-selected term is computed with the selected term. Furthermore, the difference between weight and correlation is computed for each non-selected term. The term having maximum difference value is selected. It helps to select the terms with maximum weight and low correlation value. The maximum weight helps to select the relevant terms and low correlation helps to avoid redundancy. This way, the remaining features/terms are selected from each category as per the proportion count obtained in Equation (7).

$$totalCount = \sum_{j=i}^{c} number\ of\ features\ in\ aspect\ category\ i \tag{5}$$

$$getPercent = \frac{k}{totalCount} \tag{6}$$

$$select\_no\_of\_features_c = no\_of\_features_c \times getPercent \tag{7}$$

k is the number of features to select from the feature space. $no\_of\_features_c$ are the total number of features extracted from aspect category c. $select\_no\_of\_features_c$ are the fraction of features selected from aspect category c.

Step 5: After feature selection, training file is created.

Step 6: Generated training file is passed to multilabel classifier. 10-fold cross validation technique is used for training and testing.

4. **Experimental Results and Discussion.** In this section, the experimental results and the evaluation metrics used for multilabel classification are explained.

Evaluation Metrics:

Accuracy:

Accuracy for an instance is the proportion of correctly predicted labels to the total number of labels for that instance. Total accuracy is the average over all instances.

Micro average F1 score and macro average F1 score (label-based measures):

For micro averaging, TP (true positive), FP (false positive), and FN (false negative) of all classes are used to calculate micro average precision and micro average recall [5]. Later, the harmonic mean of micro averaged precision and recall is micro-average F1 score. In macro average, precision is the average of the precisions of all classes. In a similar fashion, the macro average recall is computed. Macro average F1 is the harmonic mean of the average precision and average recall. Equations (8)-(11) depict the calculation of micro precision, micro recall, macro precision and macro recall respectively.

$$micro\_precision = \frac{\sum_{k_i \in K} TP_{k_i}}{\sum_{k_i \in K} TP_{k_i} + FP_{k_i}} \tag{8}$$

$$micro\_recall = \frac{\sum_{k_i \in K} TP_{k_i}}{\sum_{k_i \in K} TP_{k_i} + FN_{k_i}} \tag{9}$$

$$macro\_precision = \frac{\sum_{k_i \in K} Precision_{k_i}}{|K|} \tag{10}$$

$$macro\_recall = \frac{\sum_{k_i \in K} Recall_{k_i}}{|K|} \tag{11}$$

TP_{k_i} is true positive value of class k_i. FP_{k_i} is false positive value of class k_i. FN_{k_i} is false negative value of class k_i. $Precision_{k_i}$ is precision of class k_i. $Recall_{k_i}$ is recall of class k_i.

Hamming loss:

It is the fraction of incorrectly predicted labels. The lower value of Hamming loss indicates better prediction performance of the classifier. In Equation (12), T indicates number of labels, t_i and a_i are predicted and actual labels for instance i.

$$Hamming\ Loss = \frac{1}{|X|}\sum_{i=1}^{X}\frac{XOR(t_i, a_i)}{|T|} \tag{12}$$

The evaluation measures used in this work are accuracy, hamming loss, macro F1 score and micro F1 score. The performance of the proposed feature selection strategy is compared with IG and Chi-square statistics. To test the performance, multilabel classifiers are used. Very few existing approaches used multilabel classifiers for aspect prediction. A review instance may contain more than one aspect category. However, multiclass classifiers predict only one category and multilabel classifiers can predict multiple categories if existing in a review.

Table 1 shows the comparative performance of the CHI square, IG, and the proposed system (WCFS) respectively on 500, 700, 1000, and 1500 features using BR, CC and LP problem transformation strategies on SVM, NB, and RF classifiers. The results of WCFS are comparable and improved with an increased number of features. The performance of CHI square is observed better than IG. The experimental results show that the results of the proposed system are comparable with CHI square for 1000 features. The proposed approach obtained significantly improved results for 1500 features in comparison to the other strategies using BR-NB, BR-SVM, CC-NB, CC-SVM, CC-RF, LP-NB, LP-SVM, and LP-RF. The Micro-F1 value of WCFS and CHI square are same using LP-NB classifier for 1500 features. CHI square and IG are the filter approaches of feature selection. These methods are suitable to select relevant features but do not address redundancy among features. In this experimentation, the redundancy is handled using correlation between features. This experimentation demonstrates that with increased number of features the performance of the proposed system is improved. It is also noticed that the proposed system shows better results on SVM classifier for BR, CC, and LP strategies and the maximum Micro-F1 score obtained is 0.779 using BR-SVM.

Schouten et al. [21] used ontology driven approach with multiclass classifier. The F1 score gained using this approach was 0.628. The proposed system has gained a maximum 0.779 Micro F1 score using BR-SVM (multilabel) classifier for SemEval 2014 restaurant review dataset. The results are taken using 10-fold cross-validation on MEKA 1.9.0 tool. As a future scope, this system can be extended by testing it on different larger datasets for the various number of features. The proposed methodology gives better results with an increased number of features.

Experimental setup:

The configuration of the system used to carry out this experimentation is as mentioned below:

Windows 10 OS equipped with a Core i5 processor, 8 GB DDR RAM with a GeForce MX150 NVIDIA graphics card, and 1 TB HDD. A Netbeans 8.0.2 IDE is used to execute the program and the system is designed using java (version JDK 1.8).

5. **Conclusion.** The proposed work is focusing on two parameters: one is to propose a two phase feature selection approach and the second is to test the performance of the system using multilabel classifiers. Many existing systems used multiclass classifiers. Multiclass classifiers can predict only one label from an instance of a dataset. However, a test instance may contain more than one class labels. In the proposed methodology multi-label classifier is used for aspect prediction. In this approach, the search space is minimized using weight-based feature selection method that ensures the relevancy of features. The redundancy is handled by using correlation. The features for which the

TABLE 1. Accuracy (Acc), Hamming Loss (HL), Macro F1 (M_F1), and Micro F1 values of CHI, IG and WCFS methods using 500, 700, 1000, and 1500 features on BR, CC, and LP (Bayes/SVM/RF) classifiers

BR		500				700				1000				1500			
		Acc	HL	M_F1	Micro F1	Acc	HL	M_F1	Micro F1	Acc	HL	M_F1	Micro F1	Acc	HL	M_F1	Micro F1
Bayes	**CHI**	**0.603**	0.162	0.619	**0.668**	**0.605**	0.151	0.641	**0.682**	**0.601**	0.151	0.641	**0.682**	0.588	0.167	0.607	0.66
	IG	0.595	0.169	0.594	0.654	0.595	0.169	0.595	0.654	0.594	0.168	0.593	0.656	0.581	0.168	0.599	0.656
	WCFS	0.52	0.155	0.621	0.65	0.554	0.155	0.632	0.667	0.566	0.152	0.638	0.674	**0.594**	0.156	0.638	**0.678**
SVM	**CHI**	0.712	0.104	0.743	**0.773**	0.72	0.103	0.749	**0.776**	0.708	0.106	0.75	0.77	0.699	0.128	0.698	0.731
	IG	0.69	0.112	0.711	0.751	0.703	0.109	0.714	0.76	0.696	0.115	0.709	0.75	0.698	0.111	0.729	0.76
	WCFS	0.665	0.116	0.727	0.743	0.705	0.106	0.752	0.771	0.712	0.103	0.761	**0.778**	0.714	0.103	0.761	**0.779**
RF	**CHI**	0.708	0.114	0.725	**0.756**	0.708	0.114	0.725	0.755	0.714	0.111	0.732	0.761	0.699	0.111	0.708	0.757
	IG	0.698	0.126	0.693	0.742	0.706	0.123	0.703	0.748	0.724	0.113	0.728	**0.768**	0.717	0.116	0.718	**0.762**
	WCFS	0.683	0.122	0.718	0.739	0.7	0.113	0.725	**0.756**	0.706	0.108	0.727	0.764	0.724	0.112	0.771	0.722
CC																	
Bayes	**CHI**	0.617	0.167	0.599	0.661	0.629	0.16	0.636	0.675	0.632	0.159	0.642	0.678	0.632	0.161	0.642	0.678
	IG	0.606	0.172	0.593	0.652	0.606	0.172	0.594	0.653	0.61	0.17	0.599	0.655	0.612	0.172	0.607	0.66
	WCFS	0.625	0.162	0.628	**0.669**	0.632	0.16	0.64	**0.678**	0.637	0.158	0.646	**0.683**	0.634	0.133	0.676	**0.712**
SVM	**CHI**	0.748	0.107	0.748	**0.773**	0.749	0.106	0.75	**0.774**	0.747	0.107	0.753	0.773	0.733	0.113	0.732	0.761
	IG	0.719	0.119	0.705	0.743	0.723	0.118	0.707	0.747	0.725	0.117	0.711	0.75	0.715	0.122	0.71	0.743
	WCFS	0.714	0.12	0.73	0.745	0.742	0.109	0.751	0.77	0.75	0.106	0.763	**0.778**	0.748	0.106	0.76	**0.777**
RF	**CHI**	0.727	0.116	0.719	**0.753**	0.722	0.118	0.709	**0.748**	0.732	0.113	0.714	**0.757**	0.72	0.117	0.688	0.746
	IG	0.709	0.124	0.68	0.733	0.709	0.123	0.672	0.733	0.72	0.119	0.671	0.742	0.714	0.119	0.652	0.738
	WCFS	0.701	0.125	0.706	0.732	0.722	0.117	0.707	0.748	0.731	0.113	0.712	0.756	0.724	0.115	0.694	**0.75**
LP																	
Bayes	**CHI**	0.642	0.157	0.636	**0.677**	0.642	0.156	0.636	**0.678**	0.64	0.158	0.633	**0.675**	0.629	0.166	0.624	**0.665**
	IG	0.621	0.166	0.603	0.66	0.616	0.169	0.596	0.653	0.617	0.168	0.596	0.656	0.598	0.185	0.582	0.633
	WCFS	0.624	0.168	0.615	0.659	0.635	0.163	0.628	0.67	0.633	0.165	0.624	0.668	0.629	0.166	0.624	**0.665**
SVM	**CHI**	0.728	0.115	0.723	**0.752**	0.731	0.114	0.722	**0.754**	0.732	0.115	0.729	0.754	0.72	0.119	0.744	0.712
	IG	0.716	0.121	0.69	0.737	0.721	0.119	0.701	0.742	0.712	0.123	0.692	0.734	0.719	0.12	0.712	0.743
	WCFS	0.697	0.128	0.702	0.725	0.722	0.119	0.718	0.746	0.735	0.113	0.733	**0.758**	0.734	0.114	0.736	**0.758**
RF	**CHI**	0.695	0.13	0.662	**0.717**	0.691	0.132	0.652	**0.712**	0.693	0.132	0.654	**0.713**	0.679	0.137	0.62	0.698
	IG	0.672	0.139	0.62	0.696	0.676	0.138	0.615	0.697	0.688	0.134	0.63	0.707	0.673	0.14	0.605	0.691
	WCFS	0.668	0.141	0.646	0.694	0.67	0.14	0.628	0.693	0.679	0.136	0.629	0.701	0.707	0.124	0.706	**0.736**

difference between weight and correlation is more are selected. The performance of the classifier depends on the quality of the features. WCFS helps to select relevant features by avoiding redundancy among them. The performance of the system is tested using multilabel classifiers. The maximum micro F1 score gained in this experimentation is 0.779 using BR-SVM multilabel classifier. This system can be explored in the future by applying it on different larger size datasets having review instances with multiple classes.

REFERENCES

[1] I. S. Thaseen and C. A. Kumar, Intrusion detection model using fusion of chi-square feature selection and multi class SVM, *Journal of King Saud University – Computer and Information Sciences*, vol.29, no.4, pp.462-472, 2017.

[2] M. Labani, P. Moradi, F. Ahmadizar and M. Jalili, A novel multivariate filter method for feature selection in text classification problems, *Engineering Applications of Artificial Intelligence*, vol.70, pp.25-37, 2018.

[3] J. A. Kumar, S. Abirami and T. E. Trueman, Multilabel aspect-based sentiment classification for abilify drug user review, *IEEE 2019 11th International Conference on Advanced Computing (ICoAC)*, pp.376-380, 2019.

[4] M. Afzaal, M. Usman, A. C. Fong and S. Fong, Multiaspect-based opinion classification model for tourist reviews, *Expert Systems*, vol.36, no.2, 2019.

[5] J. Tao and X. Fang, Toward multi-label sentiment analysis: A transfer learning based approach, *Journal of Big Data*, vol.7, no.1, pp.1-26, 2020.

[6] J. Cai, J. Luo, S. Wang and S. Yang, Feature selection in machine learning: A new perspective, *Neurocomputing*, vol.300, pp.70-79, 2018.

[7] S. D. Sarkar, S. Goswami, A. Agarwal and J. Aktar, A novel feature selection technique for text classification using Naive Bayes, *International Scholarly Research Notices*, 2014.

[8] X. Deng, Y. Li, J. Weng and J. Zhang, Feature selection for text classification: A review, *Multimedia Tools and Applications*, vol.78, no.3, pp.3797-3816, 2019.

[9] S. D. Sarkar and S. Goswami, Empirical study on filter based feature selection methods for text classification, *International Journal of Computer Applications*, vol.81, no.6, 2013.

[10] A. Rehman, K. Javed, H. A. Babri and M. Saeed, Relative discrimination criterion – A novel feature ranking method for text data, *Expert Systems with Applications*, vol.42, no.7, pp.3670-3681, 2015.

[11] K. Ren, H. Yang, Y. Zhao, W. Chen, M. Xue, H. Miao, S. Huang and J. Liu, A robust AUC maximization framework with simultaneous outlier detection and feature selection for positive-unlabeled classification, *IEEE Trans. Neural Networks and Learning Systems*, vol.30, no.10, pp.3072-3083, 2018.

[12] S. Bahassine, A. Madani, M. Al-Sarem and M. Kissi, Feature selection using an improved chi-square for Arabic text classification, *Journal of King Saud University – Computer and Information Sciences*, vol.32, no.2, pp.225-231, 2020.

[13] S. Chormunge and S. Jena, Correlation based feature selection with clustering for high dimensional data, *Journal of Electrical Systems and Information Technology*, vol.5, no.3, pp.542-549, 2018.

[14] L. Hu, W. Gao, K. Zhao, P. Zhang and F. Wang, Feature selection considering two types of feature relevancy and feature interdependency, *Expert Systems with Applications*, vol.93, pp.423-434, 2018.

[15] E. Hancer, B. Xue and M. Zhang, Differential evolution for filter feature selection based on information theory and feature ranking, *Knowledge-Based Systems*, vol.140, pp.103-119, 2018.

[16] S. J. Lee, Z. Xu, T. Li and Y. Yang, A novel bagging C4.5 algorithm based on wrapper feature selection for supporting wise clinical decision making, *Journal of Biomedical Informatics*, vol.78, pp.144-155, 2018.

[17] Y. Liu, S. Ju, J. Wang and C. Su, A new feature selection method for text classification based on independent feature space search, *Mathematical Problems in Engineering*, 2020.

[18] X. Xue, M. Yao and Z. Wu, A novel ensemble-based wrapper method for feature selection using extreme learning machine and genetic algorithm, *Knowledge and Information Systems*, vol.57, no.2, pp.389-412, 2018.

[19] A. Madasu and S. Elango, Efficient feature selection techniques for sentiment analysis, *Multimedia Tools and Applications*, vol.79, no.9, pp.6313-6335, 2020.

[20] F. Z. Kermani, E. Eslami and F. Sadeghi, Global filter-wrapper method based on class-dependent correlation for text classification, *Engineering Applications of Artificial Intelligence*, vol.85, pp.619-633, 2019.

[21] K. Schouten, F. Frasincar and F. de Jong, Ontology-enhanced aspect-based sentiment analysis, in *Web Engineering. ICWE 2017. Lecture Notes in Computer Science*, J. Cabot, R. De Virgilio and R. Torlone (eds.), Cham, Springer, 2017.

[22] *http://metashare.ilsp.gr:8080/repository/browse/semeval-2014-absa-train-data-v20-annotation-guidelines/683b709298b811e3a0e2842b2b6a04d7c7a19307f18a4940beef6a6143f937f0/*, 2014.

[23] M. H. Nguyen, A label-oriented approach for text classification, *International Journal of Innovative Computing, Information and Control*, vol.16, no.5, pp.1593-1609, 2020.