Article

A New Feature Selection Scheme for Emotion Recognition from Text

Zafer Erenel ¹ , Oluwatayomi Rereloluwa Adegboye ¹ and Huseyin Kusetogullari ^{2,3,*}

¹ Department of Computer Engineering, European University of Lefke, Lefke, 99728 Northern Cyprus, TR-10 Mersin, Turkey; zerenel@eul.edu.tr (Z.E.); oluwatayomi.adegboye@eul.edu.tr (O.R.A.)

² Department of Computer Science, Blekinge Institute of Technology, 37141 Karlskrona, Sweden

³ School of Informatics, University of Skövde, 541 28 Skövde, Sweden

* Correspondence: huseyin.kusetogullari@bth.se; Tel.: +46-073-4223751

Received: 5 June 2020; Accepted: 1 August 2020; Published: 3 August 2020

Abstract: This paper presents a new scheme for term selection in the field of emotion recognition from text. The proposed framework is based on utilizing moderately frequent terms during term selection. More specifically, all terms are evaluated by considering their relevance scores, based on the idea that moderately frequent terms may carry valuable information for discrimination as well. The proposed feature selection scheme performs better than conventional filter-based feature selection measures Chi-Square and Gini-Text in numerous cases. The bag-of-words approach is used to construct the vectors for document representation where each selected term is assigned the weight 1 if it exists or assigned the weight 0 if it does not exist in the document. The proposed scheme includes the terms that are not selected by Chi-Square and Gini-Text. Experiments conducted on a benchmark dataset show that moderately frequent terms boost the representation power of the term subsets as noticeable improvements are observed in terms of Accuracies.

Keywords: text categorization; emotion recognition; term weighting; feature selection; machine learning

1. Introduction

Emotion recognition has become a challenging problem in the field of natural language processing since there is a large amount of data stored on the Internet. Different methods have been developed and applied to recognize emotions in various applications such as image [1], speech [2], video [3], and text [4]. Moreover, understanding emotions may increase the success of robots in the applications where there is an interaction between humans and machines [5]. However, it is computationally expensive and hard to recognize emotions in images and videos. Therefore, in this paper, we focus on emotion recognition from text as it has become significantly important while the quantity of text documents has increased dramatically on the Internet after the arrival of the new millennium.

Social networks via Internet based applications have become very popular where people convey their emotions through text [6–8]. Text classification is the task of classifying documents or data into predefined categories based on the context of content [9]. Classifiers perform the objective of classifying the document under labels close to the meaning of the content for easy recognition. They simply emulate human ability to classify a document, but it is done faster and accurately on massive information. Similarly, human emotion can be deduced from text through the application of text classification since the emotion expressed by humans can be categorized into different classes such as anger, joy, disgust, sadness, fear, and surprise [4]. This is not an exhaustive list [10]. Other emotions can be classified as secondary or tertiary forms of the six-basic forms [11]. Anger is considered as an undesirable (negative) emotion while Joy can be considered as a desirable (positive) emotion [12]. According to Lovheim [13],

there are three major monoamine systems which are serotonin, dopamine, and noradrenaline that have a great impact on emotions. Silvan Tomkins developed a comprehensive theory of fundamental emotions and classified eight different emotions [10,14,15]. Tomkins described them as basic emotions which stand for the strictly biological portion of emotion [10,14]. These basic or fundamental emotions are enjoyment/joy, interest/excitement, surprise, anger/rage, disgust, distress/anguish, fear, and shame/humiliation. Note that, several of these emotions are created based on the character or personality of the individual whereas the rest of them are formed essentially related to the social interactions. Further details of psychological effects (interior and exterior) for the individual's emotions can be found in [10,16–18]. In this work, seven basic emotion categories have been used for emotion recognition from text, which are described by the International Survey on Emotion Antecedents and Reaction [19]. These emotions are anger, disgust, fear, guilt, joy, shame, and sadness.

There are several methods used in recognizing emotion from text. One of them is the lexicon based approach [20]. In a recent study, authors implemented emotion ranking using documents where each emotion is listed in decreasing order of magnitude. Additionally, they performed word-level emotion prediction using the proposed emotion lexicon generator [20]. An alternative approach is the learning-based method. In the learning-based method, the recently trained classifier which applies a specific machine learning model recognizes emotion from a new text input [21–23].

1.1. Related Work

Before text classification takes place, the document set should go through several procedures. Firstly, the bag-of-words (BOW) representation allows each unique word in the document set to be handled as a distinct feature (or, term) [24]. Secondly, term scoring is applied in two critical areas which are term selection and term weighting. Term selection aims to select a subset of terms from the initial document collection to represent documents. Thus, it regards some terms as relevant and keeps them whereas it regards other terms as unnecessary and eliminates them by setting a threshold using numerous approaches [25,26]. The reason is twofold. Firstly, using thousands of unique terms will lead to a very high dimensional feature space [26]. This will require extra storage, memory space, and computational power [27]. Secondly, the existence of non-informative words might produce a negative impact on the decision of the automatic categorization systems. After selecting a subset of words, quantification of the relative influence of the selected words is named as term weighting.

In the field of emotion recognition from text, the BOW representation produces extremely sparse vectors due to small document lengths because text documents are made of a few number of sentences. This is named as the feature sparseness problem. The feature sparseness problem reduces the performance of the classifiers [28,29].

A previous study constructed feature space using both single words and term sets and introduced a new term weighting method to weight term sets by utilizing feature similarities to solve the feature sparseness problem [30]. In a recent study, the BOW representation has been enriched using semantic and syntactic relations between words to enhance classification performance [26]. This enrichment has improved the recall of several classifiers by reducing sparseness. The authors state that short text documents are made of a few words which turn into high dimensional sparse vectors due to the excessive number of distinct terms as a result of a high number of training samples in many domains [28]. Another related research work concluded that rare words are significant and cannot be ignored because they produce better Accuracy scores as compared to the frequent terms if used in the classifiers [31]. The authors made their experiments in the field of patent classification and technical terms that are rare improved the performance [31].

Alm et al. [32] proposed a supervised machine learning approach to classify emotions from text. The method uses a variation of Winnow update rule with different configurations of features. The authors state that it is significantly important to select the best feature set to increase the classification Accuracy score of emotion recognition. Liu et al. [33] proposed an approach using large-scale real-world knowledge for emotion classification from text. The method has been applied to classify six different

emotions which are happy, disgust, sad, fear, angry, and surprise. In [34], a hierarchical method is developed for emotion recognition from text. In order to achieve good results, authors used two different categories to classify the emotions which are positive (which represents the happiness emotion) and negative (which represents the other five emotions sadness, fear, anger, surprise, and disgust). The obtained results demonstrate that the approach performs better than the other classifiers. Moreover, in [35], another hierarchical approach is presented for emotion classification which is applied to Chinese micro-blog posts. Zhang et al. [36] developed a new feature extraction technique which is a knowledge-based topic model (KTM) to classify implicit emotion features. The SVM classifier has been applied to the extracted features to classify 19 different emotions from text. Experimental results show that authors achieved good results. Bandhakavi et al. [37] proposed a new feature extraction technique named the unigram mixture model (UMM) and compared it with the BOW. The results demonstrate that the UMM extracts features for emotion classification efficiently and outperforms the BOW. In [38], an emotion recognition model is created and applied to Indonesian text to recognize emotions. The model consists of several pre-processing stages which are normalization, stemming, term frequency-inverse document frequency weighting, and feature extraction. In the post-processing part of the model, four different classifiers which are naïve Bayes, J48, k-nearest neighbor (KNN), and SVM have been applied to the extracted features. The results show that the best result has been obtained using SVM. Moreover, an emotion recognition framework is presented for multilingual English–Hindi text. Two different classifiers which are naïve Bayes and SVM are compared and SVM performs better than the naïve Bayes classifier [39]. Other emotion recognition frameworks have been proposed and developed to classify the emotions from text using different machine learning methods such as Naïve Bayes (NB) [40], random forest (RF) [41,42], logistic regression (LR) [43] and others [21,44].

1.2. Contribution

To the best of our knowledge, the effect of moderately frequent terms in emotion recognition remains open to question. The BOW model represents the text document in the term vector where each feature is a different term. Due to this, feature selection and subsequent term weighting to compute the entries of the vectors becomes vital prior to classification. In general, feature selection methods tend to rank highly frequent terms above moderately frequent terms [45]. As a result, moderately frequent terms that may be prominent are eliminated in feature selection. In the term weighting stage, some term weighting schemes take rarity into account and give greater weights to terms that exist in a small number of documents in the training corpus [46]. In fact, the opted methods in term selection and term weighting may possibly be in conflict with each other. The proposed selection scheme is based on utilizing the moderately frequent terms using the relevance factor and the absolute value of the term occurrence probability difference to boost the representation power of the documents without increasing the dimensionality of the feature space. The scheme makes use of moderately frequent terms in the training corpus to set the entries of both training vectors and test vectors. Experiments conducted on the benchmark dataset have shown that the proposed scheme is superior in terms of Accuracy in most categories compared to the baseline methods.

2. Feature Selection and Term Weighting Schemes

This section details the well-known term selection and term weighting schemes used in the proposed framework where A is the number of documents having the term t in the positive class, B is the number of documents not having the term t in the positive class, C is the number of documents having the term t in the negative class, and D is the number of documents not having the term t in the negative class. N is the sum of all the training documents in the dataset where $N = A + B + C + D$.

2.1. Chi-Square

Chi-Square is a feature selection scheme based on filtering. It measures the dependency between a term and its class. It is used for both feature selection and term weighting. Equation (1) is used to compute the Chi-Square score of each term [47,48].

$$\text{Chi-Square}(t) = \frac{N(AD - BC)^2}{(A + C)(B + D)(A + B)(C + D)} \quad (1)$$

2.2. Gini-Text

The initial design of the Gini Index measured the impurity of term for classification. The smaller the value, lesser the impurity, the better the attribute. Gini Index is improved for feature selection and named as Gini-Text as seen in Equation (2) [49–51].

$$\text{Gini-Text}(t) = \left(\frac{A}{A+B}\right)^2 \left(\frac{A}{A+C}\right)^2 + \left(\frac{C}{C+D}\right)^2 \left(\frac{C}{A+C}\right)^2 \quad (2)$$

2.3. Relevance Frequency

The relevance frequency term weighting scheme claims that the terms with the same A/C values have the same contribution to the classification problem regardless of what the B and D values are as presented in Equation (3) [52].

$$\text{Relevance Frequency}(t) = \log\left(2 + \frac{A}{C}\right) \quad (3)$$

2.4. Binary Term Weighting

Binary term weighting scheme is a traditional strategy that is utilized for term weighting and it considers a single occurrence of terms and ignores re-occurrence of terms in a document [51]. Let us assume that D is a dataset containing various documents $D = \{d_1, \dots, d_i, \dots, d_m\}$ and m is the total number of documents. The weight of feature t_j is $W(t_j)$. $W(t_j)$ is 1 if t_j exists in and $W(t_j)$ is 0 if t_j does not exist in d_i . V is the list of features containing various unique terms where $V = \{t_1, \dots, t_j, \dots, t_n\}$ and n is the total number of unique features. For the short documents, the binary term weighting scheme is nearly as informative as other term weighting schemes which consider the re-occurrence of terms [53]. Moreover, it yields great savings in terms of computational resources [53].

3. Proposed Scheme

We propose a new scheme where the relevance frequency factor is the primary focus of interest. The authors in a previous study suggest that two terms with equal A/C factor contribute to the same extent regardless of what their B and D values are [52]. They use the relevance frequency factor for term weighting where the terms that occur mostly in the negative class receive lower weights due to high C values. The relevance frequency factor favors terms that are indicative of positive membership.

In the proposed scheme, the relevance scores of terms can be formulated as follows:

$$\begin{aligned} &\text{If } \left(\frac{A[t_i]}{A[t_i] + B[t_i]} > \frac{C[t_i]}{C[t_i] + D[t_i]}\right) \\ &\quad \text{Relevance Score}[t_i] \leftarrow \frac{A[t_i]}{A[t_i] + C[t_i]} \times \left| \frac{A[t_i]}{A[t_i] + B[t_i]} - \frac{C[t_i]}{C[t_i] + D[t_i]} \right| \\ &\quad \text{Else} \\ &\quad \text{Relevance Score}[t_i] \leftarrow \frac{C[t_i]}{A[t_i] + C[t_i]} \times \left| \frac{A[t_i]}{A[t_i] + B[t_i]} - \frac{C[t_i]}{C[t_i] + D[t_i]} \right| \end{aligned} \quad (4)$$

In the proposed formula, the relevance factor (or, the first multiplier) is either $A[t_i]/(A[t_i] + C[t_i])$ or $C[t_i]/(A[t_i] + C[t_i])$. If the A and C values of a given term t_i are equal to each other, the value of the first multiplier (or, relevance factor) is 0.5. If the term occurs only in the positive class when $C[t_i] = 0$ or in the

negative class when $A[t_i] = 0$, the value of the first multiplier is 1.0. The first multiplier does not make any discrimination against rare terms as a result of its scoring logic. If the given term t_i occurs in one document in the positive class and does not occur in the negative class, the value of the first multiplier is 1.0 for that term. Similarly, if another term t_j occurs in ten documents in the positive class and does not occur in the negative class, the value of the first multiplier will be 1.0 for that term as well. In supervised term weighting, the weight of a given term can be computed in terms of its occurrence probabilities using the training documents of positive and negative classes [53,54]. The second multiplicand in the proposed scheme is the absolute value of the term occurrence probability difference between the positive class and negative class for feature t_i . This multiplicand has been named as Δ (Delta) or ACC2 in the previous studies [51,55]. If the magnitude of the difference is high, this suggests that it is a significant term because it occurs highly in only one of the classes. Different from the weighting logic of the relevance frequency scheme, the proposed scheme favors both positive class indicative terms and negative class indicative terms in the process of selection and this is reflected in both relevance factor and Delta where the final Relevance Scores of terms are obtained by their multiplication.

The proposed scheme computes the Relevance Scores of all the unique terms in the training set and then the feature set is reduced to the topmost 1000, 900, 800, 700, 600, 500, 400, 300, 200, 150, 100, 75, and 50 terms. Similarly, Chi-Square, Gini-Text, and Delta scoring functions are used to reduce the feature set. Secondly, in a given training vector with reduced features, the entry of the selected feature is set to one if it is present and zero if it is absent in the text. After training the model, the same procedure is applied to the test vectors using the selected terms obtained from the training corpus.

3.1. Construction of Training Vectors

- Sort the terms in descending order using a scoring function by utilizing the training documents.
- Obtain binary-valued feature vectors with n terms where n is the total number of distinct terms and the first entry is the entry for the highest-ranking term.
- Set s to the selected number of features.

3.2. Construction of Test Vectors

- Use the sorted terms that are obtained using the training documents.
- Using the test vectors, obtain binary-valued feature vectors with n terms where n is the total number of distinct terms and the first entry is the entry for the highest-ranking term.
- Set s to the selected number of features.

The classification results of the proposed scheme are compared with the results of the conventional feature selection approaches.

4. Results and Discussion

In the experiments, the classifier performance is assessed using the test data to gauge Accuracy. Accuracy is the ratio of the predictions that the model made correctly. TP, FP, FN, and TN are the number of true positives, false positives, false negatives, and true negatives as presented in Equation (5).

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \quad (5)$$

4.1. Dataset

The widely used dataset ISEAR is employed for evaluating the proposed framework. The ISEAR (International Survey on Emotion Antecedents and Reaction) has seven basic emotion categories. They are anger, disgust, fear, guilt, joy, shame, and sadness. Each sentence has one category label while the quantity of sentences is 7666, annotated by 1096 people who filled the questionnaires [19,37,56].

One sample sentence from the anger category is provided in Table 1 before and after stemming. It can be said that the terms that imply the anger emotion are elusive. Moreover, they are limited in quantities.

Table 1. Sample sentence from the anger category before and after stemming.

When I was driving home after several days of hard work, there á was a motorist ahead of me who was driving at 50 km/hour and á refused, despite his low speed to let me overtake.
When wa drive home after sever dai of hard work there wa motorist ahead of me who wa drive at km hour and refus despit hi low speed to let me overtak

4.2. Experimental Setup

SVM has been used in Urdu, Hindi, English, and Chinese emotion classification systems and compared with various machine learning methods such as naïve Bayes, k-nearest neighbor, random forest classifiers [35,39,57]. The best results have been achieved using the SVM classifier. SVM has been well crafted to tackle binary classification while in emotion classification there are several categories [56]. For example, joy, anger, and shame are some of them. This circumstance can be unraveled by the one against all (OAA) approach where one category samples form the positive class and samples from all other categories form the negative class [57]. We can use the OAA approach in multi-label datasets as well because we are creating a binary SVM classifier for each emotion. As a result, a new sample can be classified into more than one emotion category depending on the decision of each classifier. In our simulations, the SVM^{light} toolbox with a linear kernel is utilized since it performs better than the nonlinear models [58].

All documents are pre-processed before training the classifiers. The Porter stemmer is applied [59]. In the experiments, four-fold cross validation is used. The dataset is split into four equal folds. Three folds are utilized for training and one fold is used for testing the model. Documents are converted into document vectors by implementing the binary weighting scheme, where the weight of the term is either one or zero depending on its presence or absence and reoccurrence of terms is not considered.

In the first set of experiments, the number of features is reduced to 1000, 900, 800, 700, 600, 500, 400, 300, 200, 150, 100, 75, and 50 terms using Chi-Square, Gini-Text, Relevance Score, and Delta filter methods to measure Accuracy due to the fact that the top 1000 features are believed to contain most of the useful terms [52,60]. The main goal of this work is to improve the classifier's performance by selecting a better subset of features.

4.3. Results

Plots report results from the experiments. Each plot in Figure 1 has four curves. They depict the experimental results obtained by the feature selection schemes.

The first set of results reflects the Accuracy performances of Chi-Square, Gini-Text, Delta feature selection schemes, and the new scheme Relevance Score in the ISEAR dataset as seen in Figure 1. Relevance Score achieved the best accuracies compared to the other schemes in three categories as depicted in Table 2. Similarly, Delta achieved the best accuracies compared to others in three categories. Chi-Square performed better than others in two categories. Gini-Text has not shown superior results in any of the categories.

Additionally, most selection schemes are producing improved results in disgust, fear, joy, and sadness when the numbers of features are increased. In contrast, Accuracy results are not increasing dramatically in anger, guilt, and joy. This suggests that there is a limited number of discriminative terms in those categories. The performance of Gini-Text decreases consistently in all categories as the number of selected terms decreases. On the contrary, Relevance Score shows consistently better performances when there are small numbers of features.

To further evaluate the results using the proposed scheme, a comparison of the best scores and second best scores obtained using the conventional feature selection approaches and the new scheme is presented in Table 2 where the bold figures and the underlined figures reflect the best and the second best Accuracy obtained in each row and the figure in brackets is the number of the selected features. The new scheme produced the best performances in three categories. Nevertheless, the number of features to obtain the best performance varies. The best performances of Chi-Square are obtained using 800 and 1000 features in joy and disgust categories. The best scores recorded in anger and shame categories are obtained using Relevance Score with only 50 and 200 features. The best performances of Delta are obtained using 1000 features in fear and guilt categories.

Figure 1. Cont.

Figure 1. Accuracy results in the International Survey on Emotion Antecedents and Reaction (ISEAR) dataset. (a) Anger, (b) disgust, (c) fear, (d) guilt, (e) joy, (f) sadness, and (g) shame.

Table 2. Best and second best accuracies and the number of terms used in ISEAR.

	Chi-Square	Gini-Text	Delta	Relevance Score
Anger	87.85(200)	87.45(50)	87.93(50)	87.99(50)
Disgust	89.72(1000)	89.25(1000)	89.38(600)	<u>89.47(700)</u>
Fear	91.32(1000)	<u>91.50(1000)</u>	91.57(1000)	91.47(600)
Guilt	88.40(200)	88.45(900)	88.50(1000)	<u>88.46(50)</u>
Joy	90.48(800)	90.00(1000)	<u>90.06(1000)</u>	90.02(400)
Sadness	<u>90.07(1000)</u>	<u>90.07(1000)</u>	90.11(500)	90.11(800)
Shame	88.73(1000)	<u>88.76(50)</u>	<u>88.76(50)</u>	88.77(200)

In order to investigate the impact of binary term weighting, the average number of terms in ISEAR dataset is computed as seen in Table 3. The first rows in the table show the average number of terms per document where the local term frequencies of existing terms are summed up. The second row in the table shows the average number of distinct terms per document where each term is counted once, and its re-occurrence is neglected. The numbers are very close. This suggests that many terms occur only one time in the documents they exist in. Binary term weighting can be considered as the most suitable choice due to the low number of re-occurrences of terms in the documents.

Table 3. Average number of terms and distinct terms per document in ISEAR.

	Anger	Disgust	Fear	Guilt	Joy	Sad	Shame
Average Number of Terms	22	20	21	21	18	18	20
Average Number of Distinct Terms	19	19	18	18	15	16	17

Moreover, Table 3 indicates that emotion documents are considerably shorter than documents used in other domains due to the fact that emotions are usually expressed in a single sentence.

In ISEAR, when the topmost fifty features are considered, Chi-Square tends to select terms with a low number of occurrences whereas Gini-Text tends to pick highly frequent terms in the dataset as shown in Figure 2 and Table 4 where $A + C$ is the number of documents that a term exists in. The proposed Relevance Score chooses features with a moderate number of occurrences. Nevertheless, these features occur mostly in the negative class due to the fact that they reside close to the vertical axis with greater C values and relatively smaller A values. Their contribution has been effective in five categories out of seven. It can be argued that the proposed selection scheme is more effective than Chi-Square because it selects frequent terms initially. More specifically, the terms that it selects are

more common than the terms that are selected by Chi-Square owing to the greater A and C values as presented in Figure 2 and Table 4. Moreover, it makes a better selection than Gini-Text since most of the terms that it selects are located nearby the C axis and they are comparatively more discriminative than the terms that are selected by Gini-Text. As a result, Relevance Score selections may create better document representations during training and testing when a low number of topmost terms are utilized by providing a better trade-off between occurrence frequencies of terms and class distributions of terms in selecting features compared to Chi-Square and Gini-Text.

Figure 2. The topmost 50 terms in Chi-Square, Relevance Score, and Gini-Text selections. (a) Anger category, (b) disgust category.

Table 4. Average (A), average (C), and average (A)/average (C) in the anger category using the topmost 50 terms.

	Gini-Text	Relevance Score	Chi-Square
Average (A)	137.28	59.18	35.54
Average (C)	718.90	343.30	132.76
Average (A)/Average (C)	0.19	0.17	0.27

The negative class is made of a large number of documents from diverse categories in the ISEAR dataset due to the nature of OAA approach. Authors in a previous study argue that negative class indicative terms can never reach the same maximum Chi-Square score as positive class indicative terms due to the nature of the class imbalance problem [61]. Their remarks agree with plots in Figure 2 and numbers in Table 4. The topmost 50 terms in the Chi-Square selection are commonly positive class indicative terms as they are closer to the horizontal axis. As a result, this will lead to a limited number of negative class indicative terms since many terms that are not selected are negative class indicative terms and this may lead to poor performance results for Chi-Square at extreme filtering levels.

In a previous study, the authors show that Gini-Text scores of rare features are low irrespective of their distribution among positive class and negative class [49]. Their observations agree with our findings. In six categories out of seven, Chi-Square performs better than Gini-Text when the topmost 50 terms are considered. Therefore, comparatively rare features that are not selected (or, excluded) by Gini-Text but distributed asymmetrically between positive class and negative class contribute to the classification performance. Moreover, when additional terms are selected, Chi-Square selects more balanced feature subsets in terms of document frequencies ($A + C$) and asymmetric class distributions compared to other schemes. In the categories disgust and joy, the Accuracy performances of Chi-Square is superior to other schemes.

To summarize, the numbers in Table 4 imply that Chi-Square is favoring positive class indicative terms because it has the highest average (A)/average (C) ratio. Gini-Text and Relevance Score are favoring negative class indicative terms to a greater extent due to lower average (A)/average (C) ratios. Average (A) and average (C) obtained using Gini-Text selection is significantly larger than average (A) and average (C) obtained using other schemes. This suggests that Gini-Text selections are flooded with common and highly repeated terms in small term sets irrespective of term class distributions in the dataset. Average (A) obtained using Relevance Score selection is slightly larger than average (A) obtained using Chi-Square selection whereas average (C) obtained using Relevance Score selection is considerably larger than average (C) obtained using Chi-Square selection. Relevance Score tends to select terms that indicate a negative class membership to a greater extent whereas Chi-Square is inclined towards terms that indicate a positive class membership at the cost of low frequency (or, occurrence).

5. Conclusions

A new feature selection scheme is proposed to investigate the effect of the selected features in emotion recognition from text. In the course of this experimentation, one emotion dataset is experimented on, using Chi-Square, Gini-Text, Delta, and Relevance Score selection schemes for feature reduction. The OAA approach is adapted for binary classification using linear SVM as the classifier. It has been shown that the selected terms by Relevance Score improved the classification performance in numerous categories. Relevance Score provides a better trade-off between occurrence frequencies of terms and class distribution of terms in selecting features.

There are several areas that need to be further investigated. In particular, another term weighting scheme can replace binary term weighting. Moreover, other feature selection methods can be further explored for improved results. Lastly, the proposed scheme uses the multiplication of the relevance factor and the Delta factor to compute the Relevance Scores of the terms. The proposed Relevance Scoring can be used in conjunction with other selection schemes to obtain improved results owing to the fact that each selection scheme has pros and cons when the number of filtered terms change. As a result, better subsets of terms that are more informative can be obtained for improved results.

Author Contributions: Data acquisition, Z.E.; conceptualization, H.K.; software, O.R.A.; experiments, O.R.A.; supervision, Z.E.; methodology, Z.E.; formal analysis, Z.E.; resources, H.K.; validation, Z.E.; review and editing; H.K. and Z.E.; writing—original draft, Z.E. All authors have read and agreed to the published version of the manuscript.

Funding: H.K. is funded by the research project scalable resource efficient systems for big data analytics by the Knowledge Foundation (Grant: 20140032) in Sweden.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Kim, J.-H.; Kim, B.-G.; Roy, P.P.; Jeong, D.-M.; Kima, B.-G. Efficient Facial Expression Recognition Algorithm Based on Hierarchical Deep Neural Network Structure. *IEEE Access* **2019**, *7*, 41273–41285. [\[CrossRef\]](#)
- Li, S.; Deng, W. Deep Facial Expression Recognition: A Survey. *IEEE Trans. Affect. Comput.* **2020**. [\[CrossRef\]](#)
- Nguyen, D.H.; Kim, S.; Lee, G.-S.; Yang, H.-J.; Na, I.-S.; Kim, S.H. Facial Expression Recognition Using a Temporal Ensemble of Multi-level Convolutional Neural Networks. *IEEE Trans. Affect. Comput.* **2019**. [\[CrossRef\]](#)
- Ekman, P. An argument for basic emotions. *Cogn. Emot.* **1992**, *6*, 169–200. [\[CrossRef\]](#)
- Cavallo, F.; Semeraro, F.; Fiorini, L.; Magyar, G.; Sincak, P.; Dario, P. Emotion Modelling for Social Robotics Applications: A Review. *J. Bionic Eng.* **2018**, *15*, 185–203. [\[CrossRef\]](#)
- De Diego, I.M.; Fernandez-Isabel, A.; Ortega, F.; Moguerza, J.M. A visual framework for dynamic emotional web analysis. *Knowl. Based Syst.* **2018**, *145*, 264–273. [\[CrossRef\]](#)
- Saldias, F.B.; Picard, R.W. Tweet Moodifier: Towards giving emotional awareness to Twitter users. In Proceedings of the 2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII), Cambridge, UK, 3–6 September 2019; pp. 1–7.

8. Franzoni, V.; Milani, A.; Gervasi, O.; Murgante, B.; Misra, S.; Rocha, A.M.A.; Torre, C.M.; Taniar, D.; Apduhan, B.O.; Stankova, E.; et al. *A Semantic Comparison of Clustering Algorithms for the Evaluation of Web-Based Similarity Measures*; Springer: Cham, Switzerland, 2016; pp. 438–452.
9. Minaee, S.; Kalchbrenner, N.; Cambria, E.; Nikzad, N.; Chenaghlu, M.; Gao, J. Deep learning based text classification: A comprehensive review. *arXiv* **2020**, arXiv:2004.03705.
10. Talanov, M.; Vallverdú, J.; Distefano, S.; Mazzara, M.; Delhibabu, R. Neuromodulating Cognitive Architecture: Towards Biomimetic Emotional AI. In Proceedings of the 2015 IEEE 29th International Conference on Advanced Information Networking and Applications, Gwangju, Korea, 24–27 March 2015; pp. 587–592.
11. Shivhare, S.N.; Khethawat, S. Emotion Detection from Text. *arXiv* **2012**, arXiv:1205.4944.
12. Ishizuka, M.; Neviarouskaya, A.; Shaikh, M.A.M. Textual Affect Sensing and Affective Communication. *Int. J. Cogn. Inform. Nat. Intell.* **2012**, *6*, 81–102. [[CrossRef](#)]
13. Lövhelm, H. A new three-dimensional model for emotions and monoamine neurotransmitters. *Med. Hypotheses* **2012**, *78*, 341–348. [[CrossRef](#)]
14. Tomkins, S. *Affect Imagery Consciousness Volume III the Negative Affects Anger and Fear*; Springer Publishing Company: New York, NY, USA, 1991.
15. Vernon, C.K. *A Primer of Affect Psychology*; The Tomkins Institute: Colchester, CT, USA, 2009.
16. Deng, J.J.; Leung, C.H.C.; Mengoni, P.; Li, Y. Emotion Recognition from Human Behaviors Using Attention Model. In Proceedings of the 2018 IEEE First International Conference on Artificial Intelligence and Knowledge Engineering (AIKE), Laguna Hills, CA, USA, 26–28 September 2018; pp. 249–253.
17. Deng, J.; Leung, C.; Li, Y. Beyond Big Data of Human Behaviors: Modeling Human Behaviors and Deep Emotions. In Proceedings of the 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), Miami, FL, USA, 10–12 April 2018; pp. 282–286.
18. Dolan, R.J. Emotion, Cognition, and Behavior. *Science* **2002**, *298*, 1191–1194. [[CrossRef](#)] [[PubMed](#)]
19. Razeq, M.A.; Frasson, C. Text-Based Intelligent Learning Emotion System. *J. Intell. Learn. Syst. Appl.* **2017**, *9*, 17–20. [[CrossRef](#)]
20. Bandhakavi, A.; Wiratunga, N.; Massie, S.; Padmanabhan, D. Lexicon Generation for Emotion Detection from Text. *IEEE Intell. Syst.* **2017**, *32*, 102–108. [[CrossRef](#)]
21. Batbaatar, E.; Li, M.; Ryu, K.H. Semantic-Emotion Neural Network for Emotion Recognition from Text. *IEEE Access* **2019**, *7*, 111866–111878. [[CrossRef](#)]
22. Ramalingam, V.V.; Pandian, A.; Jaiswal, A.; Bhatia, N. Emotion detection from text. *J. Phys. Conf. Ser.* **2018**, *1000*, 1–5. [[CrossRef](#)]
23. Hulliyah, K.; Abu Bakar, N.S.A.; Ismail, A.R. Emotion recognition and brain mapping for sentiment analysis: A review. In Proceedings of the 2017 Second International Conference on Informatics and Computing (ICIC), Jayapura, Indonesia, 1–3 November 2017; pp. 1–5.
24. Yao, J.; Zhang, M. *Feature Selection with Adjustable Criteria*; Springer: Berlin/Heidelberg, Germany, 2005; pp. 204–213.
25. Rehman, A.; Javed, K.; Babri, H. Feature selection based on a normalized difference measure for text classification. *Inf. Process. Manag.* **2017**, *53*, 473–489. [[CrossRef](#)]
26. Sebastiani, F. Machine learning in automated text categorization. *ACM Comput. Surv.* **2002**, *34*, 1–47. [[CrossRef](#)]
27. Ogura, H.; Amano, H.; Kondo, M. Feature selection with a measure of deviations from Poisson in text categorization. *Expert Syst. Appl.* **2009**, *36*, 6826–6832. [[CrossRef](#)]
28. Heap, B.; Bain, M.; Wobcke, W.; Krzywicki, A.; Schmeidl, S. Word Vector Enrichment of Low Frequency Words in the Bag-of-Words Model for Short Text Multi-class Classification Problems. *arXiv* **2017**, arXiv:1709.05778.
29. Cui, X.; Kojaku, S.; Masuda, N.; Bollegala, D. Solving Feature Sparseness in Text Classification using Core-Periphery Decomposition. In Proceedings of the Seventh Joint Conference on Lexical and Computational Semantics, New Orleans, Louisiana, 5–6 June 2018; pp. 255–264.
30. Yuan, M.; Ouyang, Y.X.; Xiong, Z. A text categorization method using extended vector space model by frequent term sets. *J. Inf. Sci. Eng.* **2013**, *29*, 99–114.
31. Khattak, A.; Heyer, G. Significance of low frequent words in patent classification using IPC Hierarchy. In Proceedings of the 11th International Conference on Innovative Internet Community Systems, Luxembourg, 19–24 June 2011; pp. 239–250.

32. Alm, C.O.; Roth, D.; Sproat, R. Emotions from text: Machine learning for text-based emotion prediction. In Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing, Stroudsburg, PA, USA, 5 October 2005; pp. 579–586.
33. Liu, H.; Lieberman, H.; Selker, T. A model of textual affect sensing using real-world knowledge. In Proceedings of the 8th International Conference on Intelligent User Interfaces, Miami, FL, USA, 12–15 January 2003; pp. 125–132.
34. Ghazi, D.; Inkpen, D.; Szpakowicz, S. Hierarchical versus flat classification of emotions in text. In Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text, Stroudsburg, PA, USA, 5–8 June 2010; pp. 140–146.
35. Xu, H.; Yang, W.; Wang, J. Hierarchical emotion classification and emotion component analysis on chinese micro-blog posts. *Expert Syst. Appl.* **2015**, *42*, 8745–8752. [[CrossRef](#)]
36. Zhang, F.; Xu, H.; Wang, J.; Sun, X.; Deng, J. Grasp the implicit features: Hierarchical emotion classification based on topic model and SVM. In Proceedings of the 2016 International Joint Conference on Neural Networks (IJCNN), Vancouver, BC, Canada, 24–29 July 2016; pp. 3592–3599.
37. Bandhakavi, A.; Wiratunga, N.; Padmanabhan, D.; Massie, S. Lexicon based feature extraction for emotion text classification. *Pattern Recognit. Lett.* **2017**, *93*, 133–142. [[CrossRef](#)]
38. Winarsih, N.A.S.; Supriyanto, C. Muljono; Winarsih, N.A.S.; Supriyanto, C. Evaluation of classification methods for Indonesian text emotion detection. In Proceedings of the 2016 International Seminar on Application for Technology of Information and Communication (ISemantic), Semarang, Indonesia, 5–6 August 2016; pp. 130–133.
39. Jain, V.K.; Kumar, S.; Fernandes, S.L. Extraction of emotions from multilingual text using intelligent text processing and computational linguistics. *J. Comput. Sci.* **2017**, *21*, 316–326. [[CrossRef](#)]
40. Mulki, H.; Ali, C.B.; Haddad, H.; Babaoğlu, I. Tw-StAR at SemEval-2018 Task 1: Preprocessing Impact on Multi-label Emotion Classification. In Proceedings of the 12th International Workshop on Semantic Evaluation, New Orleans, Louisiana, USA, 5–6 June 2018; pp. 167–171.
41. Asghar, M.Z.; Subhan, F.; Imran, M.; Kundi, F.M.; Khan, A.; Shamshirband, S.; Mosavi, A.; Koczy, A.R.V.; Csiba, P. Performance Evaluation Of Supervised Machine Learning Techniques For Efficient Detection Of Emotions From Online Content. *Comput. Mater. Contin.* **2020**, *63*, 1093–1118. [[CrossRef](#)]
42. Ghazi, D.; Inkpen, D.; Szpakowicz, S. Prior and contextual emotion of words in sentential context. *Comput. Speech Lang.* **2014**, *28*, 76–92. [[CrossRef](#)]
43. Jain, V.K.; Kumar, S.; Jain, N.; Verma, P. A Novel Approach to Track Public Emotions Related to Epidemics In Multilingual Data. In Proceedings of the 2nd International Conference and Youth School Information Technology and Nanotechnology (ITNT 2016), Samara, Russia, 17–19 May 2016; pp. 883–889.
44. Colneric, N.; Demšar, J. Emotion Recognition on Twitter: Comparative Study and Training a Unison Model. *IEEE Trans. Affect. Comput.* **2018**. [[CrossRef](#)]
45. Yang, Y.; Pedersen, O. A comparative study on feature selection in text categorization. In Proceedings of the 14th International Conference on Machine Learning, San Francisco, CA, USA, 5–8 July 1997; pp. 412–420.
46. Schönhofen, P.; Benczúr, A.A. Exploiting Extremely Rare Features in Text Categorization. In Proceedings of the 17th European Conference on Machine Learning, Berlin, Germany, 18–22 September 2006; pp. 759–766.
47. Sun, J.; Zhang, X.; Liao, D.; Chang, V. Efficient method for feature selection in text classification. In Proceedings of the 2017 International Conference on Engineering and Technology (ICET), Antalya, Turkey, 21–23 August 2017; pp. 1–6.
48. Yang, J.; Qu, Z.; Liu, Z. Improved Feature-Selection Method Considering the Imbalance Problem in Text Categorization. *Sci. World J.* **2014**, *2014*, 1–17. [[CrossRef](#)]
49. Shang, W.; Huang, H.; Zhu, H.; Lin, Y.; Qu, Y.; Wang, Z. A novel feature selection algorithm for text categorization. *Expert Syst. Appl.* **2007**, *33*, 1–5. [[CrossRef](#)]
50. Park, H.; Kwon, H.-C. Improved Gini-Index Algorithm to Correct Feature-Selection Bias in Text Classification. *IEICE Trans. Inf. Syst.* **2011**, *94*, 855–865. [[CrossRef](#)]
51. Kim, K.; Zzang, S.Y. Trigonometric comparison measure: A feature selection method for text categorization. *Data Knowl. Eng.* **2019**, *119*, 1–21. [[CrossRef](#)]
52. Calefato, F.; Lanubile, F.; Novielli, N. EmoTxt: A toolkit for emotion recognition from text. In Proceedings of the 2017 Seventh International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW), San Antonio, TX, USA, 23–26 October 2017; pp. 79–80.

53. Forman, G. An extensive empirical study of feature selection metrics for text classification. *J. Mach. Learn. Res.* **2003**, *3*, 1289–1305.
54. Erenel, Z.; Altınçay, H.; Varoğlu, E. Explicit Use of Term Occurrence Probabilities for Term Weighting in Text Categorization. *J. Inf. Sci. Eng.* **2011**, *27*, 819–834.
55. Altınçay, H.; Erenel, Z. Analytical evaluation of term weighting schemes for text categorization. *Pattern Recognit. Lett.* **2010**, *31*, 1310–1323. [[CrossRef](#)]
56. Mozafari, F.; Tahayori, H. Emotion Detection by Using Similarity Techniques. In Proceedings of the 2019 7th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS), Bojnord, Iran, 29–31 January 2019; pp. 1–5.
57. Sana, L.; Nasir, K.; Urooj, A.; Ishaq, Z.; Hameed, I.A. BERS: Bussiness-Related Emotion Recognition System in Urdu Language Using Machine Learning. In Proceedings of the 2018 5th International Conference on Behavioral, Economic, and Socio-Cultural Computing (BESC), Kaohsiung, Taiwan, 12–14 November 2018; pp. 238–242.
58. Lan, M.; Tan, C.L.; Su, J.; Lu, Y. Supervised and Traditional Term Weighting Methods for Automatic Text Categorization. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *31*, 721–735. [[CrossRef](#)] [[PubMed](#)]
59. Porter, M. An algorithm for suffix stripping. *Program* **1980**, *14*, 130–137. [[CrossRef](#)]
60. Azam, N.; Yao, Y. Comparison of term frequency and document frequency based feature selection metrics in text categorization. *Expert Syst. Appl.* **2012**, *39*, 4760–4768. [[CrossRef](#)]
61. Zheng, Z.; Wu, X.; Srihari, R. Feature selection for text categorization on imbalanced data. *ACM SIGKDD Explor. Newsl.* **2004**, *6*, 80–89. [[CrossRef](#)]

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).