World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:10, No:1, 2016

A Proposed Approach for Emotion Lexicon Enrichment

Amr Mansour Mohsen, Hesham Ahmed Hassan, Amira M. Idrees

Abstract**—**Document Analysis is an important research field that aims to gather the information by analyzing the data in documents. As one of the important targets for many fields is to understand what people actually want, sentimental analysis field has been one of the vital fields that are tightly related to the document analysis. This research focuses on analyzing text documents to classify each document according to its opinion. The aim of this research is to detect the emotions from text documents based on enriching the lexicon with adapting their content based on semantic patterns extraction. The proposed approach has been presented, and different experiments are applied by different perspectives to reveal the positive impact of the proposed approach on the classification results.

Keywords**—**Document analysis, sentimental analysis, emotion detection, WEKA tool, NRC Lexicon.

I. Introduction

NOWADAYS, Text Mining is one of the most important topics in the field of computer science area. It refers to the automated extraction of useful information from computerized huge unstructured text. Different associated fields for text mining includes text summarization, question answering systems, and sentimental analysis. According to [1], the sentimental analysis is the field of providing analysis to people's opinions, sentiments and emotions of their writings. Analyzing the person's opinion has a vital role in different analysis area such as in social networks, for example, analyzing people's opinion about a government decision or a new product. Opinions have different categories, positive, negative, and neutral. An example for positive opinion is "Alcatel is a high-quality mobile", and negative like "Samsung Mobile battery is too bad" or neutral like "I went to club yesterday" which expresses no opinion. Another level of opinion mining is emotions analysis; there are emotions categorized to be "positive" such as (Joy, surprise) or negative such as (anger, fear, sad). In this research, we focus on Ekman list of emotions [2] which are (anger, disgust, fear, joy, sadness, surprise).

The remaining of the research paper is as follows; Section II discusses different research in the field. The resources that are used such as the dataset and the lexicon are presented in Section III. Then Section IV presents the proposed architecture. Also, the experimental results and the evaluation of the proposed approach with the discussion of the results are in Section V. Finally; the conclusion is discussed in Section VI.

Amr Mansour Mohsen is with the Management Information Systems Department, Faculty of Commerce and Business Administration, Future University in Egypt, Egypt (e-mail: amr.mansour@ fue.edu.eg).

Hesham Ahmed Hassan is with the Computer Science Department, Faculty of Computers and Information, Cairo University, Egypt (e-mail: h.hassan@fci-cu.edu.eg).

Amira M. Idrees is with the Information Systems Department, Faculty of Computers and Information, Fayoum University, Egypt (e-mail: ami04@fayoum.edu.eg).

II. Related Work

Two main approaches are proposed for extracting the emotions from text documents at both sentence level and word level, they are, lexicon based techniques, and machine learning techniques. The following subsections discuss these two approaches and the work that has been performed by different researchers.

A. Lexicon Based Techniques

The main target for applying lexicon techniques is to build a dictionary containing the words with describing the emotional status for each word. In [3], a huge knowledge base has been built for concepts and their relation to each other's which is called concept net. Additionally, a direct acyclic graph is presented describing the main domain concepts in ovals and the relation between these concepts in the edges.

Many researchers have used these concepts network in detecting emotion in text such as a research proposed by [4], which built an emotion concept network according to emotion classes and with the help of the WordNet [5]. However, an extensive task is performed when there is processing to some text for identifying their emotions; the task includes processing the whole network to get all the emotions for the required documents. Another research with result accuracy of 86% by [6], which used six basic emotions (happiness, sadness, anger, fear, disgust, and surprise) with lexical chains [7], these chains are used to put the nearest words with the same meaning by a semantic relationship. In [8], the researchers aimed at building a lexicon by taking a dataset of sentences and perform a manual annotation for the words, which resulted in 59.11% accuracy. Moreover, in [9], an Arabic emotion lexicon has been built from a group of 100 documents and 2514 sentences with manual annotation of the words with their emotions of (Joy, fear, sadness, anger, disgust, surprise), the accuracy of the result has been measured by f-measure metric which revealed in 65%. Moreover, [10] presented a classification to the emotions from chatting text systems using a lexicon which provided an accuracy of 84.6%.

B. Machine Learning Techniques

Machine learning techniques are divided into two main approaches; they are supervised and unsupervised learning approaches. Supervised learning approach aims at learning from classified, labeled data about the specific domain, while the unsupervised learning approach is trying to learn from data with hidden information. A semi-supervised learning approach means that we have some indicators about the data but not the whole information. In [11], they used a supervised learning technique which is support vector machines (SVM) [12] for emotion detection in text and reached 48% accuracy. The same technique is applied in [13] with accuracy equal 71.64%. Moreover, [14] used another supervised technique which is Naïve Bayes [15] for emotion classification and reached an accuracy of 53.6%.

III. Data Resources

In this section, we describe the data resources required for this research. The resources we needed were the lexicon, the datasets, and the WEKA tool. A description of these resources is provided in the next subsections.

A. The Dataset

Two datasets are used in the experimenting step of this research; they are ISEAR and SEMEVAL. We applied the proposed approach on the two datasets and measured the results which revealed to the high accuracy of the proposed approach as will be discussed in the experiments section.

1.ISEAR Dataset

ISEAR dataset [16] stands for International Survey of Emotion Antecedents and Reactions; ISEAR was developed by a group of psychologists from different countries. 3000 students from 37 countries participated in the dataset by reporting their situations with emotions (Anger, Disgust, Fear, Joy, Sadness, Shame, and Guilt). The research group then gathered these answers. The students' data is also analyzed to determine their features such as gender, age, city, country, state when a person tells the situation and many other points. These features will be used in the classification task. The dataset is composed of 7,666 sentences that are classified with the 7 emotions and distributed as described in Table I. An example of a sentence classification from the dataset is the sentence "Misunderstand by friend" which is classified as "anger".

TABLE I
Number of Instances for Each Category in ISEAR Dataset

Emotion	Number of Instances
Anger	1,096
Disgust	1,096
Fear	1,095
Joy	1,095
Sadness	1,096
Guilt	1,093
Shame	1,096
Total	7,666

2.SEMEVAL Dataset

SEMEVAL dataset [17] uses news headlines extracted from news websites for classifying the emotions of them. The set of emotions that are used to classify the headlines contains six elements, they are (anger, disgust, fear, joy, sadness and surprise), with more classification polarity to be either positive or negative. Each sentence in the dataset is annotated with its emotion(s) with a range of each emotion from 0 to 100. Also, there is another range [-100,100] for polarity for positive and negative annotation which means that the sentence is highly positive when it takes '100', '0' when neutral (neither negative nor positive) and '-100' which will be highly negative. The dataset is composed of two sub-datasets with total 1250 sentences, the development data set which is composed of 250 headlines and the test data set which is composed of 1000 headlines. An example for "money makeovers" sentence annotation is described in Table II.

TABLE II
Cross Section from SEMEVAL Dataset

Emotion	Value
Anger	0
Disgust	0
Fear	0
Joy	31
Sadness	0
Surprise	8

B. Lexicon

Lexicon is a dictionary based technique that aims to classify the emotions of a word based on using the definition of this word in dictionaries. In this section, we will discuss the NRC emotion lexicon which is used in this research.

"NRC" [18] is a high-quality moderate sized emotion lexicon which stands for National Research Council of Canada. (NRC) is manually created using Mechanical Turk [19] and WordNetAffect [20]. The lexicon is built based on questioners that are distributed to the subscribers of Mechanical Turk [21] for annotating the specified words. Then analyzing the words is performed with providing the classification of each word to follow one of the emotions (sadness, fear, anger, trust, disgust, surprise, and anticipation, positive, negative). The lexicon includes 14182 terms that are classified to eight emotions with another level of classification to be either positive or negative. An example from the lexicon is shown in Table III. Each Word has an associated number of each emotion either 0 or 1 which reflects if the words are classified to the emotion or not as the term applicable with this emotion or not. In Table III, the word 'dislike' is classified to follow the emotions anger and disgust.

TABLE III
Example from the NRC Lexicon

Emotion	Value
Anger	1
anticipation	0
disgust	1
fear	0
joy	0
negative	1
positive	0
sadness	0
surprise	0

C. WEKA Tool

WEKA stands for Waikato Environment for Knowledge Analysis [22]; WEKA is a tool which is developed by the University of Waikato, New Zealand. It has a variety of classes for machine learning classifiers like (Support Vector Machines, Naïve Bayesian networks…). There is a graphical user interface for using the tool by providing the ability for adapting the tool with java classes and codes for more programming issues.

In this paper a set of machine algorithms from WEKA will be applied, they are, Sequential Minimal Optimization (SMO) which is an algorithm proposed [23] for training support vector machines [12]. Nearest Neighbor (IBK) [24] is a supervised technique and is a type of lazy learning [25] which is used for classification. KStar [26] or K* is a classification technique that is based on similarity with the training instances. Bagging technique [27] designed for enhancing the stability accuracy of machine learning. And Logistic model trees (LMT) [28] which is described by the classification trees, logistic regression functions and decision trees (J48) [29].

IV. Proposed Framework

In this research, a framework for documents' emotion classification is proposed; Fig. 1 provides a representation of the proposed framework while the next subsections sections discuss the framework in details. In summary, the framework applies a set of preprocessing techniques on the selected dataset and then applies the machine learning techniques for emotion classification. For more accurate results, the dictionary is enriched before being used in the learning step of the classification method.

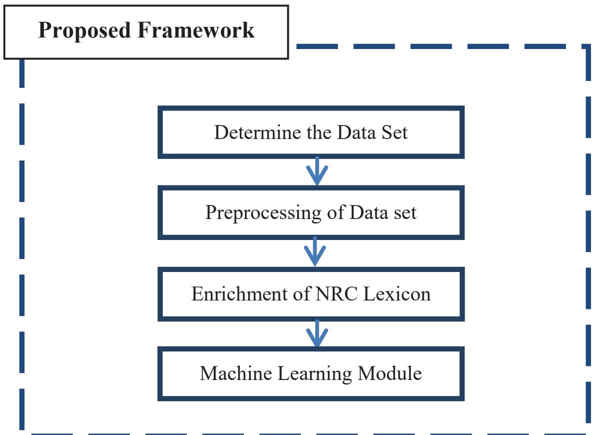

Fig. 1 The proposed framework for emotions' classification using lexicon enrichment approach

A. Determine the Dataset

This step is concerned with selecting the required dataset to be classified, as previously discussed, in the experiments applied in this research, two datasets are determined; they are ISEAR, and SEMEVAL.

B. Preprocessing of Data

Dataset preprocessing is a basic phase in any natural language processing task. We applied the preprocessing steps in the proposed system using Stanford Parser tool [30]. Preprocessing steps can be summarized as follows:

Tokenization [15], this step is performed by parsing the text and splitting it into a sequence of words or terms.

Stop words removal; this step is concerned with removing the words which will not affect the whole meaning of the text, examples of these words are (the, a, an, etc.). Moreover, removing punctuations (?;,..) and white spaces are also removed.

Sentences filtration, this step is concerned with removing the sentences that will not add value in the classification task such as "No response" or "none."

Lemmatization is the process to transform the original form of a word such as transforming "made" to "make". This step is performed by annotating the word with the part of speech (noun, verb, adjective, and adverb). While stemming means removing some ends of the same words so to achieve the same goal of Lemmatization, however, we will use Lemmatization here because it returns the word to its original form depending on the part of speech of the word

C. NRC Dictionary Enrichment

Fig. 2 shows the architecture of the lexicon enrichment module. The input to this module is the NRC lexicon and the required dataset. As previously discussed, NRC lexicon contains 14183 words which are classified to eight emotions (anger, disgust, fear, joy, sad, anticipation, trust and surprise) with polarity (positive, negative). In this component, we aim to enrich the NRC lexicon by more classified words to reach better emotions' classification for documents. Moreover, NRC lexicon defined the emotion of the word based on a specific emotion. Most of the terms are considered for specific emotion in a definite situation, therefore, in this component, we also applied modifications to the lexicon for adaptation and refinement.

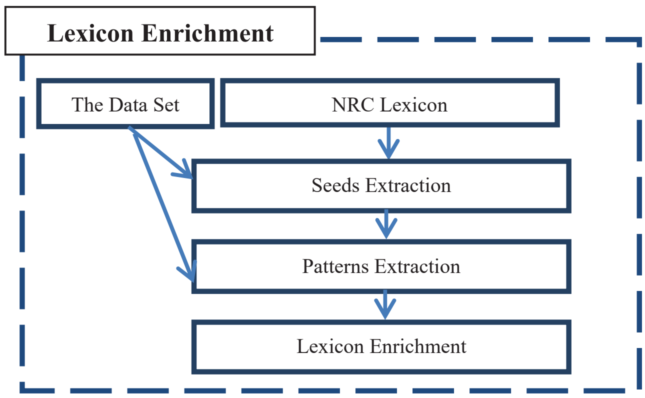

Fig. 2 Lexicon Enrichment Module

Seeds Extraction step applies matching between the words in the lexicon, and the dataset are applied in the "seeds extraction" step. For each word in the lexicon, if it is found in any sentence then it is classified as the emotion of that sentence. *Pattern Extraction* step then extracts the patterns that are related to each seed, and these patterns are then classified with the same class of the seed, we define the related pattern by the successor word and predecessor word of that seed. *Lexicon Enrichment* step uses these patterns to extract more seeds which follow the same emotion class of the pattern. The previously three steps are continuously executed until there is no seed is extracted.

D. Machine Learning Component

This component considers applying the classification algorithm over the dataset. In this research paper, we applied seven machine learning algorithms for classification and compared their results in different perspectives which revealed to a final conclusion with the best of them. In the training phase, preparing the training data is performed. For each sentence in the training data, each word is examined to find its emotion class, then weighting the features' impact is applied by determining the presence of each emotion for each sentence and then determining its frequency. This step in weighting the features has proved to increase the results' accuracy of the applied machine learning algorithms as will be shown in the experiment section.

V. Experiments and Results

This section discusses the details of the applied experiment to prove the applicability of the proposed approach. We have applied four sets of experiments in different perspectives. In all experiments, Precision, Recall, and F-Measure metrics are calculated for determining the results' accuracy. The following sub-sections demonstrate all the details for all the experiments.

A. First Set of Experiments (SEMEVAL Dataset Experiments)

In these experiments' set, we applied the proposed approach using seeds from NRC dictionary [18] with the SemEval dataset [17]. In SemEval dataset, data are classified into six emotions (Anger, Disgust, fear, joy, sadness, and surprise).

TABLE IV
Accuracy Metrics Results for 10-Fold Cross-Validation to the Proposed Approach on SEMEVAL Dataset

Emotions	Precision	Recall	F-measure
Anger	86.09%	75.00%	80.16%
Disgust	87.80%	83.72%	85.71%
Fear	83.41%	75.30%	79.15%
Joy	82.44%	80.52%	81.47%
Sadness	80.00%	72.08%	75.84%
Surprise	80.49%	65.22%	72.05%
Average	83.37%	75.31%	79.06%

In the first experiment, we applied the 10-fold cross validation over 1000 sentences of the dataset with its original emotions' classification in the dataset. Then after lexicon enrichment step, we applied the support vector machines (SMO) algorithm from WEKA for classification. Precision, Recall, and F-measure metrics are measured for each emotion to determine the accuracy of the applied algorithm for classifying each emotion. Table IV presents the results of the experiment.

In the second experiment, the following steps are performed for preparing the experiment:

- The dataset is divided into 1000 sentences as training data, and 250 sentences as testing data.
- The training data sentences were multi-classified, which means that a sentence may have been classified into more than one emotion such as "sad" and "anger".
- Following the approach proposed by [31], this classification is updated into a single-classification; each document is the classified into one emotion which has the highest classification confidence for this sentence.
- The previous step has revealed to have the distribution of 1250 headlines as follows: 198 headlines for "anger", 95 headlines for "Disgust", 321 headlines for "fear", 421 headlines for "joy", 388 headlines for "sadness", and 296 headlines for "surprise."

Then, we applied the proposed approach for lexicon enrichment as follows:

- Preprocessing and stop words steps are performed.
- In the second step, extracting seeds from the training headlines data is performed using NRC lexicon, 1,292 seeds are extracted from the training data.
- Extracting the seeds' patterns is performed; these patterns are classified with the same emotion of their seeds. These patterns are then applied on the same training data set to extract more seeds as previously discussed in the proposed approach section, and the lexicon is enriched by adding the new seeds with a classification following the same classification of the extracting pattern.
- The enrichment step revealed to 146 new words which have been added to the lexicon for enrichment. Therefore the total words that will be used for the next classification step have been raised to be 1,438 words.
- A weighting for the six emotions (anger, disgust, fear, joy, sadness, surprise) in the whole dataset is applied.
- Applying the same classification algorithm (SMO) on the testing data after lexicon enrichment has revealed to the results of accuracy that are presented in Table V. The results of the new experiments reveal that the new approach has higher average precision which means it reaches more accuracy to classify the sentence with the new emotion's class.

TABLE V
Accuracy Metrics Results of the Proposed Approach on SEMEVAL Dataset Using 1000 Sentences as Training Data

Emotions	Precision	Recall	F-measure
Anger	90.70%	59.09%	71.56%
Disgust	86.84%	63.46%	73.33%
Fear	88.33%	71.62%	79.10%
Joy	79.27%	84.42%	81.76%
Sadness	90.14%	60.95%	72.73%
Surprise	80.70%	53.49%	60.53%
Average	86.16%	65.51%	74.37%

B. Second Set of Experiments (Comparison with Other Researches)

In [31], NRC-6 lexicon is used with the six emotions as features over the SemEval dataset for measuring the accuracy of his system. Tables VI and VII present a comparison of the accuracy metrics between the work presented in [31] using NRC-6 features and the proposed architecture after applying the lexicon enrichment step. In their research [31]. SVM algorithm has been used for classification; therefore, to prove the positive impact of the lexicon enrichment step on raising the accuracy of the system, we have applied the same algorithm in our approach. Comparing the results when using NRC-6 features to our proposed approach over 1000 sentences using 10-fold cross validation shows that the overall f-measure for the 6 emotions is increased with 30 % as shown in following Table VI. Another experiment has been applied for comparing the two systems by using 1000 sentences as training data and 250 sentences as testing data is shown in Table VII which also revealed to the increase in the accuracy metrics of our proposed approach.

TABLE VI
COMPARISON OF THE ACCURACY METRICS' RESULTS BY 10-FOLD CROSS-VALIDATION EXPERIMENT ON 1000 SENTENCES

Enrichment	Precision	Recall	F-measure
[31]	24.1%	95%	38.4%
proposed approach	83.37%	75.31%	79.06%

TABLE VII
COMPARISON OF THE ACCURACY METRICS' RESULTS BY 1000 SENTENCES DATA TRAINING AND TESTING ON 250

Enrichment	Precision	Recall	F-measure
[31]	34.0	58.3	42.9
proposed approach	84.16%	65.51%	73.17%

C. Third Set of Experiments (ISEAR Dataset Experiment)

In the third set of experiments, ISEAR dataset is used for evaluation. The sentences in the ISEAR dataset is classified into five emotions; they are, (anger, disgust, fear, joy, sadness). In this experiment, 1,040 sentences are used from each emotion category so the total number of sentences that will be used in this experiment will be 5,200.

We have applied two different experiments directions, the first direction of the experiments was before applying seven machine learning classification algorithms on ISEAR dataset before enriching the lexicon, these algorithms are (Naïve Bayes, Support Vector machines (SMO), IBK, KSTAR, Bagging, LMT, J48). The second experiment by applying the same seven machine learning classification algorithms on ISEAR dataset after enriching the lexicon. In each experiment, we have performed two sets of experiments; the first set is dividing the dataset into 70% training data and 30% testing data and the second set is applying a 10-fold cross validation for the experiment to measure the accuracy. Moreover, as in ISEAR dataset, the sentences are classified into 10 emotions, we have performed the previously mentioned experiments twice, one with the complete classification of the sentences for the 10 emotions' classes, and the second with selecting the sentences that are only classified with the five emotion classes which are the focus of this research (anger, disgust, fear, joy, sadness). The details of these experiments are discussed in the next subsections.

1.First Experiment before Applying Lexicon Enrichment

In this experiment, the dataset contains 3550 sentences for training data and 1650 sentences for testing data distributed equally over all the emotions' classes. The experiments have applied the seven algorithms without applying the enrichment step on the lexicon. Table VIII presents the results of the applied experiments. As shown in the table, the best f-measure accuracy percentage for the first set were by applying the "Bagging" algorithm, and for the second set were by the IBK algorithm.

2.Second Experiment before Applying Lexicon Enrichment

In this experiment, a 10-fold cross validation technique is applied on the dataset. As in the previous experiment, the experiments have applied the seven algorithms without applying the enrichment step on the lexicon. Table IX presents the results of the applied experiments. As shown in the table, the best f-measure accuracy percentage for the first set were by applying the "LMT" algorithm, and for the second set were by the "Bagging" algorithm.

3.Third Experiment after Applying Lexicon Enrichment

In this experiment, the dataset contains 3550 sentences for training data and 1650 sentences for testing data distributed equally over all the emotions' classes. In this experiment, the lexicon enrichment step has been applied on the training dataset before applying the classification algorithms. Table XII presents the results of the applied experiments. As shown in the table, the best f-measure accuracy percentage was reached by applying the "LMT" algorithm in the two experiments. It is also noticed the accuracy has increased after lexicon enrichment for the two experiments' directions when compared with the first experiment which used the same distribution of the dataset.

4.Fourth Experiment after Applying Lexicon Enrichment

In this experiment, a 10-fold cross validation technique is applied to the dataset. In this experiment, the lexicon enrichment step has been applied on the training dataset before applying the classification algorithms. Table XIII presents the results of the applied experiments. As shown in the table, the best f-measure accuracy percentage for the first set were by applying the "LMT" algorithm, and for the second set were by the "SMO" algorithm. It is also noticed the accuracy has increased after lexicon enrichment for the two experiments' directions when compared with the first experiment which used the same distribution of the dataset.

5.Comparing of All Results Before and After Enrichment

According to the previously presented results, a final comparison has been performed to reach the final conclusion. According to these results, the lexicon enrichment step has proved its positive impact on the classification task results.

We also found that "LMT" algorithm produces the best results in most of the experiments.

It is shown that after applying the enrichment of the lexicon, F-measure is increased by 4 percentage points. Table X presents a comparison of the best results of the first and the third experiments for the six emotion classes before and after enrichment using 70% as training data and 30% as testing data. Although "Bagging" algorithm reached better results comparing with other algorithms before lexicon enrichment, however, "LMT" reached better results after lexicon enrichment.

TABLE VIII
RESULTS OF DIVIDING THE DATASET INTO 70% FOR TRAINING AND 30 % FOR TESTING BEFORE LEXICON ENRICHMENT

Machine Learning Algorithm	Emotions	Results with using all emotions' classes			Results with using the five emotions' classes		
		Precision	**Recall**	**f-measure**	**Precision**	**Recall**	**f-measure**
Naïve Bayes	*Anger*	33.75%	16.36%	22.04%	32.99%	19.39%	24.43%
	Disgust	54.70%	19.39%	28.64%	41.00%	12.42%	19.07%
	Fear	42.31%	26.67%	32.71%	65.00%	15.76%	25.37%
	Joy	24.94%	63.03%	35.74%	25.25%	77.88%	38.13%
	Sadness	25.38%	25.45%	25.42%	27.52%	21.52%	24.15%
	Average	36.22%	30.18%	28.91%	38.35%	29.39%	26.23%
IBK	*Anger*	34.62%	35.45%	35.03%	43.37%	25.76%	32.32%
	Disgust	41.10%	38.48%	39.75%	47.42%	27.88%	35.11%
	Fear	43.23%	34.85%	38.59%	51.71%	36.67%	42.91%
	Joy	45.57%	43.64%	44.58%	36.38%	65.15%	46.69%
	Sadness	31.59%	40.30%	35.42%	32.87%	43.33%	37.39%
	Average	39.22%	38.55%	38.67%	42.35%	39.76%	**38.88%**
SMO	*Anger*	40.76%	22.73%	29.18%	43.54%	19.39%	26.83%
	Disgust	46.25%	33.64%	38.95%	45.78%	31.21%	37.12%
	Fear	51.50%	36.36%	42.63%	50.62%	24.85%	33.33%
	Joy	48.76%	41.82%	45.02%	31.34%	80.91%	45.18%
	Sadness	30.85%	66.36%	42.12%	29.17%	23.33%	25.93%
	Average	43.62%	40.18%	39.58%	40.09%	35.94%	33.68%
KSTAR	*Anger*	39.66%	28.48%	33.16%	44.94%	24.24%	31.50%
	Disgust	43.06%	36.67%	39.61%	49.48%	28.79%	36.40%
	Fear	46.45%	29.70%	36.23%	51.52%	36.06%	42.42%
	Joy	45.95%	43.03%	44.44%	35.85%	64.85%	46.17%
	Sadness	31.37%	58.18%	40.76%	32.30%	44.24%	37.34%
	Average	41.30%	39.21%	38.84%	42.82%	39.64%	38.77%
Bagging	*Anger*	37.32%	31.21%	33.99%	46.67%	14.85%	22.53%
	Disgust	45.40%	43.33%	44.34%	43.19%	33.64%	37.82%
	Fear	42.97%	34.24%	38.11%	50.99%	39.09%	44.25%
	Joy	42.32%	50.91%	46.22%	36.50%	68.79%	47.69%
	Sadness	34.34%	41.52%	37.59%	32.45%	40.61%	36.07%
	Average	40.47%	40.24%	**40.05%**	41.96%	39.39%	37.67%
LMT	*Anger*	43.92%	19.70%	27.20%	43.51%	17.27%	24.73%
	Disgust	46.79%	37.58%	41.68%	45.70%	35.45%	39.93%
	Fear	49.16%	35.45%	41.20%	52.53%	31.52%	39.39%
	Joy	45.38%	49.09%	47.16%	36.07%	66.67%	46.81%
	Sadness	30.69%	59.70%	40.53%	32.53%	44.85%	37.71%
	Average	43.19%	40.30%	39.55%	42.07%	39.15%	37.71%
J48	*Anger*	35.77%	29.70%	32.45%	43.23%	20.30%	27.63%
	Disgust	47.25%	39.09%	42.79%	44.39%	28.79%	34.93%
	Fear	45.33%	41.21%	43.17%	50.67%	34.24%	40.87%
	Joy	40.77%	51.52%	45.52%	36.53%	66.97%	47.27%
	Sadness	31.09%	36.36%	33.52%	30.68%	42.12%	35.50%
	Average	40.04%	39.58%	39.49%	41.10%	38.48%	37.24%

Table XI presents a comparison of the best results of the first and the third experiments for the six emotion classes before and after enrichment using 10-fold cross validation. The results show that "LMT" reached better results before and after lexicon enrichment.

A final conclusion is reached according to the previously presented results that using "LMT" machine learning algorithm with applying the lexicon enrichment approach has revealed to the best results of all the experiments. This conclusion proves a general the positive impact of applying the lexicon enrichment step on the emotions' classification task.

VI. Conclusion and Future Work

In this research, we proposed a lexicon based classification approach for emotions classification to text documents. The proposed approach included automatic enriching to the lexicon that is used for the classification task which revealed to an enhancement in the accuracy results. The experiment is applied on two different datasets (SEMVAL, ISEAR) and with different perspectives and different methods of measures (10-fold-cross validation, and 70% training data). The proposed approach has been compared with other research and proved its competence. The results of the experiments have revealed that the best machine learning algorithm that is used was LMT with the new proposed approach.

TABLE IX
Results of 10-Folds Cross-Validation before Lexicon Enrichment

Machine Learning algorithm	Emotions	Results with using all emotions' classes			Results with using the five emotions' classes		
		Precision	**Recall**	**f-measure**	**Precision**	**Recall**	**f-measure**
Naïve Bayes	*Anger*	39.13%	36.15%	37.58%	33.85%	31.54%	32.65%
	Disgust	85.06%	21.35%	34.13%	49.16%	19.71%	28.14%
	Fear	43.90%	31.15%	36.45%	63.69%	21.92%	32.62%
	Joy	34.09%	51.83%	41.13%	26.50%	76.63%	39.39%
	Sadness	25.74%	41.06%	31.64%	37.86%	16.35%	22.83%
	Average	45.58%	36.31%	36.18%	42.21%	33.23%	31.13%
IBK	*Anger*	52.19%	53.75%	52.96%	53.53%	35.00%	42.33%
	Disgust	60.82%	50.00%	54.88%	56.00%	37.69%	45.06%
	Fear	58.23%	53.75%	55.90%	59.60%	42.69%	49.75%
	Joy	60.60%	54.13%	57.19%	38.69%	65.48%	48.64%
	Sadness	42.17%	56.15%	48.16%	37.72%	47.69%	42.12%
	Average	54.80%	53.56%	53.82%	49.11%	45.71%	45.58%
SMO	*Anger*	77.64%	47.40%	58.87%	57.22%	29.33%	38.78%
	Disgust	71.09%	58.17%	63.99%	59.44%	36.63%	45.33%
	Fear	76.90%	53.46%	63.07%	60.43%	40.67%	48.62%
	Joy	65.83%	60.38%	62.99%	35.18%	84.71%	49.72%
	Sadness	42.46%	83.17%	56.22%	42.58%	33.65%	37.59%
	Average	66.78%	60.52%	61.03%	50.97%	45.00%	44.01%
KSTAR	*Anger*	64.13%	47.79%	54.77%	51.35%	32.98%	40.16%
	Disgust	70.21%	51.44%	59.38%	58.89%	36.63%	45.17%
	Fear	69.04%	55.10%	61.28%	58.67%	41.63%	48.71%
	Joy	61.47%	57.98%	59.67%	37.90%	63.56%	47.49%
	Sadness	41.09%	73.17%	52.63%	35.28%	47.60%	40.52%
	Average	54.01%	60.19%	56.93%	48.42%	44.48%	44.41%
Bagging	*Anger*	64.59%	52.79%	58.10%	54.72%	30.67%	39.31%
	Disgust	69.81%	56.25%	62.30%	57.91%	36.25%	44.59%
	Fear	63.46%	62.79%	63.12%	58.39%	47.50%	52.39%
	Joy	57.80%	60.58%	59.15%	39.50%	70.19%	50.55%
	Sadness	43.50%	58.27%	49.82%	37.42%	45.77%	41.18%
	Average	59.83%	58.13%	58.50%	49.59%	46.08%	45.60%
LMT	*Anger*	75.83%	50.38%	60.54%	43.51%	17.27%	24.73%
	Disgust	69.68%	60.10%	64.53%	45.70%	35.45%	39.93%
	Fear	75.65%	55.58%	64.08%	52.53%	31.52%	39.39%
	Joy	62.17%	66.06%	64.06%	36.07%	66.67%	46.81%
	Sadness	45.27%	75.87%	56.70%	32.53%	44.85%	37.71%
	Average	65.72%	61.60%	61.98%	42.07%	39.15%	37.71%
J48	*Anger*	64.71%	52.88%	58.20%	54.39%	29.81%	38.51%
	Disgust	67.85%	56.83%	61.85%	57.31%	38.46%	46.03%
	Fear	63.34%	61.63%	62.48%	58.71%	43.75%	50.14%
	Joy	60.20%	58.17%	59.17%	39.35%	68.75%	50.05%
	Sadness	43.23%	60.77%	50.52%	36.79%	47.40%	41.43%
	Average	59.87%	58.06%	58.44%	49.31%	45.63%	45.23%

TABLE X
F-MEASURE RESULTS BEFORE AND AFTER LEXICON ENRICHMENT ON 3650 SENTENCES TRAINING DATA AND 1650 TESTING DATA

Enrichment		**Before**	**After**
Best Machine Learning		Bagging	LMT
F-measure results according to emotions	*Anger*	33.99%	37.04%
	Disgust	44.34%	50.29%
	Fear	38.11%	49.83%
	Joy	46.22%	46.65%
	Sadness	37.59%	38.13%
	Average	40.05%	44.39%

TABLE XI
F-MEASURE RESULTS BEFORE AND AFTER LEXICON ENRICHMENT ON 3650 SENTENCES TRAINING DATA AND 1650 TESTING DATA

Enrichment		**Before**	**After**
Best Machine Learning		LMT	LMT
F-measure results according to emotions	*Anger*	60.54%	57.43%
	Disgust	64.53%	66.73%
	Fear	64.08%	66.07%
	Joy	64.06%	63.66%
	Sadness	56.70%	58.59%
	Average	61.98%	62.50%

TABLE XII
RESULTS OF DIVIDING THE DATASET INTO 70% FOR TRAINING AND 30 % FOR TESTING AFTER LEXICON ENRICHMENT

Machine Learning algorithm	Emotions	Results with using all emotions' classes			Results with using the five emotions' classes		
		Precision	**Recall**	**f-measure**	**Precision**	**Recall**	**f-measure**
Naïve Bayes	*Anger*	*33.75%*	*16.36%*	*22.04%*	*32.99%*	*19.39%*	*24.43%*
	Disgust	*54.70%*	*19.39%*	*28.64%*	*41.00%*	*12.42%*	*19.07%*
	Fear	*42.31%*	*26.67%*	*32.71%*	*65.00%*	*15.76%*	*25.37%*
	Joy	*24.94%*	*63.03%*	*35.74%*	*25.25%*	*77.88%*	*38.13%*
	Sadness	*25.38%*	*25.45%*	*25.42%*	*27.52%*	*21.52%*	*24.15%*
	Average	*36.22%*	*30.18%*	*28.91%*	*38.35%*	*29.39%*	*26.23%*
IBK	*Anger*	*32.18%*	*36.67%*	*34.28%*	*26.54%*	*33.94%*	*29.79%*
	Disgust	*40.65%*	*38.18%*	*39.37%*	*33.96%*	*33.03%*	*33.49%*
	Fear	*46.05%*	*42.42%*	*44.16%*	*41.20%*	*37.58%*	*39.30%*
	Joy	*41.02%*	*46.36%*	*43.53%*	*39.35%*	*40.30%*	*39.82%*
	Sadness	*37.98%*	*33.03%*	*35.33%*	*40.30%*	*32.73%*	*36.12%*
	Average	*39.58%*	*39.33%*	*39.34%*	*36.27%*	*35.52%*	*35.70%*
SMO	*Anger*	*40.76%*	*22.73%*	*29.18%*	*43.54%*	*19.39%*	*26.83%*
	Disgust	*46.25%*	*33.64%*	*38.95%*	*45.78%*	*31.21%*	*37.12%*
	Fear	*51.50%*	*36.36%*	*42.63%*	*50.62%*	*24.85%*	*33.33%*
	Joy	*48.76%*	*41.82%*	*45.02%*	*31.34%*	*80.91%*	*45.18%*
	Sadness	*30.85%*	*66.36%*	*42.12%*	*29.17%*	*23.33%*	*25.93%*
	Average	*43.62%*	*40.18%*	*39.58%*	*40.09%*	*35.94%*	*33.68%*
KSTAR	*Anger*	*38.41%*	*33.64%*	*35.86%*	*34.69%*	*30.91%*	*32.69%*
	Disgust	*48.30%*	*43.03%*	*45.51%*	*42.86%*	*37.27%*	*39.87%*
	Fear	*53.60%*	*45.15%*	*49.01%*	*45.45%*	*39.39%*	*42.21%*
	Joy	*42.07%*	*57.88%*	*48.72%*	*42.65%*	*52.73%*	*47.15%*
	Sadness	*39.70%*	*40.30%*	*40.00%*	*38.40%*	*43.64%*	*40.85%*
	Average	*44.42%*	*44.00%*	*43.82%*	*40.81%*	*40.79%*	*40.56%*
Bagging	*Anger*	*39.23%*	*36.97%*	*38.07%*	*32.70%*	*31.52%*	*32.10%*
	Disgust	*44.92%*	*44.24%*	*44.58%*	*41.25%*	*40.00%*	*40.62%*
	Fear	*52.03%*	*46.67%*	*49.20%*	*46.84%*	*44.85%*	*45.82%*
	Joy	*40.51%*	*53.03%*	*45.93%*	*43.60%*	*43.33%*	*43.47%*
	Sadness	*40.91%*	*35.45%*	*37.99%*	*36.41%*	*40.61%*	*38.40%*
	Average	*43.52%*	*43.27%*	*43.15%*	*40.16%*	*40.06%*	*40.08%*
LMT	*Anger*	*37.74%*	*36.36%*	*37.04%*	*33.55%*	*30.61%*	*32.01%*
	Disgust	*48.59%*	*52.12%*	*50.29%*	*40.74%*	*43.33%*	*42.00%*
	Fear	*57.54%*	*43.94%*	*49.83%*	*49.22%*	*47.88%*	*48.54%*
	Joy	*41.83%*	*52.73%*	*46.65%*	*46.22%*	*51.82%*	*48.86%*
	Sadness	*39.35%*	*36.97%*	*38.13%*	*40.39%*	*37.58%*	*38.93%*
	Average	*45.01%*	*44.42%*	*44.39%*	*42.02%*	*42.24%*	*42.07%*
J48	*Anger*	*35.66%*	*30.91%*	*33.12%*	*32.26%*	*27.27%*	*29.56%*
	Disgust	*44.83%*	*43.33%*	*44.07%*	*39.78%*	*44.85%*	*42.17%*
	Fear	*49.00%*	*44.55%*	*46.67%*	*44.97%*	*40.61%*	*42.68%*
	Joy	*41.89%*	*57.88%*	*48.60%*	*43.05%*	*47.88%*	*45.34%*
	Sadness	*39.45%*	*34.55%*	*36.83%*	*39.22%*	*39.70%*	*39.46%*
	Average	*42.16%*	*42.24%*	*41.86%*	*39.86%*	*40.06%*	*39.84%*

TABLE XIII
RESULTS OF 10-FOLDS CROSS VALIDATION AFTER LEXICON ENRICHMENT

Machine Learning algorithm	Emotions	Results with using all emotions' classes			Results with using the five emotions' classes		
		Precision	**Recall**	**f-measure**	**Precision**	**Recall**	**f-measure**
Naïve Bayes	*Anger*	*46.80%*	*40.10%*	*43.19%*	*42.65%*	*52.73%*	*47.15%*
	Disgust	*73.90%*	*32.40%*	*45.05%*	*38.40%*	*43.64%*	*40.85%*
	Fear	*54.01%*	*32.40%*	*40.50%*	*40.52%*	*48.18%*	*44.00%*
	Joy	*31.91%*	*67.79%*	*43.40%*	*33.69%*	*76.73%*	*46.82%*
	Sadness	*35.39%*	*34.71%*	*35.05%*	*55.48%*	*33.56%*	*41.82%*
	Average	*48.40%*	*41.48%*	*41.44%*	*42.15%*	*50.97%*	*44.13%*
IBK	*Anger*	*43.08%*	*48.17%*	*45.48%*	*31.93%*	*38.65%*	*34.97%*
	Disgust	*53.12%*	*50.77%*	*51.92%*	*43.59%*	*41.83%*	*42.69%*
	Fear	*56.46%*	*53.37%*	*54.87%*	*45.68%*	*42.21%*	*43.88%*
	Joy	*54.04%*	*59.13%*	*56.47%*	*44.77%*	*44.42%*	*44.59%*
	Sadness	*52.28%*	*46.35%*	*49.13%*	*43.68%*	*39.90%*	*41.71%*
	Average	*51.80%*	*51.56%*	*51.58%*	*41.93%*	*41.40%*	*41.57%*
SMO	*Anger*	*63.84%*	*52.79%*	*57.79%*	*48.77%*	*41.92%*	*45.09%*
	Disgust	*66.67%*	*63.85%*	*65.23%*	*53.99%*	*53.94%*	*53.97%*
	Fear	*69.89%*	*61.15%*	*65.23%*	*55.04%*	*56.73%*	*55.87%*
	Joy	*61.89%*	*64.33%*	*63.08%*	*51.63%*	*59.52%*	*55.29%*
	Sadness	*50.33%*	*65.48%*	*56.92%*	*53.11%*	*50.87%*	*51.96%*
	Average	*62.52%*	*61.52%*	*61.65%*	*52.51%*	*52.60%*	*52.44%*
KSTAR	*Anger*	*57.91%*	*48.94%*	*53.05%*	*34.69%*	*30.91%*	*32.69%*
	Disgust	*63.64%*	*59.90%*	*61.71%*	*42.86%*	*37.27%*	*39.87%*
	Fear	*68.38%*	*60.29%*	*64.08%*	*45.45%*	*39.39%*	*42.21%*
	Joy	*56.64%*	*68.94%*	*62.19%*	*42.65%*	*52.73%*	*47.15%*
	Sadness	*54.01%*	*60.19%*	*56.93%*	*38.40%*	*43.64%*	*40.85%*
	Average	*54.01%*	*60.19%*	*56.93%*	*40.81%*	*40.79%*	*40.56%*
Bagging	*Anger*	*59.38%*	*51.15%*	*54.96%*	*44.92%*	*37.40%*	*40.82%*
	Disgust	*64.72%*	*60.67%*	*62.63%*	*50.71%*	*51.35%*	*51.03%*
	Fear	*64.86%*	*64.62%*	*64.74%*	*52.85%*	*52.60%*	*52.72%*
	Joy	*57.09%*	*70.87%*	*63.23%*	*50.46%*	*57.88%*	*53.92%*
	Sadness	*58.08%*	*55.96%*	*57.00%*	*48.91%*	*49.52%*	*49.21%*
	Average	*60.83%*	*60.65%*	*60.51%*	*49.57%*	*49.75%*	*49.54%*
LMT	*Anger*	*60.32%*	*54.81%*	*57.43%*	*47.69%*	*43.75%*	*45.64%*
	Disgust	*65.70%*	*67.79%*	*66.73%*	*53.02%*	*53.94%*	*53.48%*
	Fear	*72.66%*	*60.58%*	*66.07%*	*55.24%*	*56.73%*	*55.98%*
	Joy	*61.68%*	*65.77%*	*63.66%*	*53.42%*	*57.79%*	*55.52%*
	Sadness	*54.56%*	*63.27%*	*58.59%*	*52.46%*	*50.19%*	*51.30%*
	Average	*62.98%*	*62.44%*	*62.50%*	*52.37%*	*52.48%*	*52.38%*
J48	*Anger*	*53.37%*	*47.21%*	*50.10%*	*41.67%*	*34.13%*	*37.53%*
	Disgust	*61.01%*	*60.19%*	*60.60%*	*49.31%*	*51.25%*	*50.26%*
	Fear	*60.98%*	*59.81%*	*60.39%*	*50.50%*	*48.56%*	*49.51%*
	Joy	*57.05%*	*67.69%*	*61.92%*	*48.38%*	*54.42%*	*51.22%*
	Sadness	*59.08%*	*59.47%*	*59.10%*	*45.40%*	*47.88%*	*46.61%*
	Average	*58.30%*	*58.87%*	*58.42%*	*47.05%*	*47.25%*	*47.02%*

We, however, seek to increase the accuracy of the proposed approach further by finding how to detect implicit emotions in different situations such as “meeting my friend” which do not reveal if it represents joy or sad. Also using more features and different machine learning techniques for understanding will provide more enhancements for classifying the emotions. Additionally, we need to provide a more advanced step which is to classify the causality behind expressing the emotions.

REFERENCES

[1] Bing Liu, Sentiment Analysis and Opinion Mining. Chicago, USA: Morgan & Claypool Publishers, 2012.
[2] P. Ekman, Universals and Cultural Differences in Facial Expressions of Emotions. Lincioln: University of Nebraska Press, 1972.
[3] ConceptNet5, (2010). Last retrieved 29 Novemeber 2015 (Online). http://conceptnet5.media.mit.edu/
[4] R. Ezhilarasi and R. Minu, "Automatic Emotion Recognition and Classification," in International Conference on Modelling Optimization and Computing, 2012.
[5] George A. Miller, "WordNet: a lexical database for English," Communications of the ACM, vol. 38, no. 11, pp. 39-41, 1995.
[6] M. Naveen Kumar and R. Suresh, "Emotion Detection using Lexical Chains," International Journal of Computer Applications, vol. 57, no. 4, 2012.
[7] Peter D. Turney and Michael L. Littman, "Measuring Praise and Criticism: Inference of Semantic Orientation from Association," ACM

Transactions on Information Systems, vol. 21, pp. 315–346, 2003.

[8] J. Lei, Y. Raob, Q. Li, X. Quan, and Liu Wenyin, "Towards building a social emotion detection system for online news," Journal of Future Generation Computer Systems, vol. 37, pp. 438-448, 2013.

[9] Amira F. El Gohary, Torky I. Sultan, Maha A. Hana, and Mohamed M. El Dosoky, "A Computational Approach for Analyzing and Detecting Emotions in Arabic Text," International Journal of Engineering Research and Applications (IJERA), vol. 3, no. 3, pp. 100-107, May-Jun 2013. (Online). http://www.ijera.com/papers/Vol3_issue3/S33100107.pdf

[10] R. Tokuhisa, K. Inui, and Y. Matsumoto, "Emotion Classification Using Massive Examples Extracted from the Web," in the 22nd International Conference on Computational Linguistics, 2008.

[11] D. Inkpen, F. keshtkar, and D. Ghazi, "Analysis and Generation of Emotion in Texts," International Conference on Knowledge Engineering Principles and Techniques, 2009.

[12] T. Joachims, Learning to Classify Text Using Support Vector Machines: Methods, Theory, and Algorithms, 1st ed. Boston: Kluwer Academic Publishers, 2002.

[13] Chaitali G. Patil and Sandip S. Patil, "Use of Porter Stemming Algorithm and SVM for Emotion Extraction from News Headlines," International Journal of Electronics, Communication & Soft Computing Science and Engineering, vol. 2, no. 7, pp. 9-13, 2013.

[14] S. Boa et al., "Mining Social Emotions from Affective Text," IEEE Transactions on Knowledge and Data Engineering, vol. 24, no. 9, pp. 1658-1669, 2012.

[15] Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze, Introduction to Information Retrieval. New York: Cambridge University, 2008.

[16] Emotion Research. (1990s) (Online). http://emotionresearch.net/toolbox/toolboxdatabase.

[17] C. Strapparava and R. Mihalcea, "SemEval-2007 Task 14: Affective Text," in The 4th International Workshop on Semantic Evaluations, 2007.

[18] Saif Mohammad and Peter Turney, "Emotions Evoked by Common Words and Phrases: Using Mechanical Turk to Create an Emotion Lexicon," in the NAACL-HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text, California, 2010.

[19] Mechanical Turk (2007) Retrieved from: https://www.mturk.com/mturk/welcome

[20] Carlo Strapparava and Alessandro Valitutti, "WordNet-Affect: An Affective Extension of WordNet,", Lisbon, 2004, pp. 1083-1086.

[21] Peter Cohen. (2005) https://www.mturk.com/mturk. (Online). https://www.mturk.com/mturk

[22] Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten Mark Hall. (2009) the University of Waikato. (Online). http://www.cs.waikato.ac.nz/ml/weka/

[23] J. Platt, "Fast Training of Support Vector Machines using Sequential Minimal Optimization," in Advances in Kernel Methods - Support Vector Learning, B. Schoelkopf and C. Burges and A. Smola, Ed. UN: MIT Press, 1998.

[24] B. S., Landau, S., Leese, M. and Stahl, D. Everitt, Miscellaneous Clustering Methods, in Cluster Analysis, 5th ed. Chichester, UK: John Wiley & Sons, 2011.

[25] Stuart Russell and Peter Norvig, Artificial Intelligence: A Modern Approach, second edition ed. Berkeley: Prentice Hall, 2003.

[26] G. Cleary and Leonard, E. Trigg John, "K*: An Instance- based Learner Using an Entropic Distance Measure,", 1995.

[27] Amit, Anshuman Sahu, Daniel Apley, and George Runger Shinde, "Preimages for Variation Patterns from Kernel PCA and Bagging," IIE Transactions, vol. 46, no. 5, 2014.

[28] Niels Landwehr and Mark Hall and Eibe Frank, "Logistic Model Trees," Machine Learning, vol. 95, no. 1, pp. 161-205, 2005.

[29] data-mining-business-intelligence. Last retrieved: 29 November 2015 (Online). http://data-mining.business-intelligence.uoc.edu/home/j48-decision-tree

[30] Marie-Catherine de Marneffe, Bill MacCartney, and Christopher D. Manning, "Generating Typed Dependency Parses from Phrase Structure Parses," in The fifth international conference on Language Resources and Evaluation, Genoa, Italy, 2006, pp. 449-454.

[31] Saif Mohammad, "Portable Features for Classifying Emotional Text," in Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, North American, 2012, pp. 587-591. (Online). http://www.aclweb.org/anthology/N12-1071