Journal of Advances in Computer Research
Quarterly pISSN: 2345-606x eISSN: 2345-6078
Sari Branch, Islamic Azad University, Sari, I.R.Iran
(Vol. 7, No. 2, May 2016), Pages: 85-99
www.jacr.iausari.ac.ir

Implicit Emotion Detection from Text with Information Fusion

Nooshin Riahi✉**, Pegah Safari**

Computer Engineering Department, Alzahra University, Tehran, Iran

nriahi@alzahra.ac.ir; phsafari@yahoo.com

Received: 2015/04/12; Accepted: 2015/12/05

Abstract

In this paper we have proposed an approach for emotion detection in implicit texts. We introduced a combinational system based on three subsystems. Each one analyzes input data from a different aspect and produces an emotion label as output. The first subsystem is a machine learning method. The second one is a statistical approach based on vector space model (VSM) and the last one is a keyword-based subsystem with an information fusion component to aggregate the final output of main system. We analyzed the performance of our proposed system on ISEAR dataset with seven emotions: anger, joy, sad, shame, fear, disgust and guilt. The results show that our combinational system outperforms each subsystem with overall f-measure of 0.68 and at least up to 0.71 in terms of F_1 in emotion level except for anger. The overall performance of the main system is 9.13% better than machine learning module, 16.6% better than VSM and 23% better than keyword-based.

Keywords: *Implicit Emotion Detection, Combinational System, Information Fusion, Machine Learning, Vector Space Model, Keyword Based.*

1. Introduction

Emotions have been widely studied in psychology and behavior science, as they are an important element in human nature and play key role in their communications. In computational linguistic, the automatic detection of emotions from texts becomes more important especially from an applicative point of view. For instance in sentiment analysis which decides just the valence of the text, emotion detection as a fine-grained technique can reflect the feelings of costumers about different aspects of a specific product and reveal the imperfections or favorite features in new merchandise and enhance business plans [1].

In human interfaces using an emotion detection system can bring more intelligence to the human-computer interactions and change the way people communicate with computers. Even studying mass social blogs or post of people on personal pages can reveal the social movements or trends toward a specific political party or shows the abnormalities in society and even alarms potential suicide attempts [2].

With this wide variety of applications, recently emotion detection received a special attention which results in growth of different detection methods. These methods can be categorized into three main approaches: machine learning, keyword-based and hybrid technique which is the combination of the first two methods. Each approach faces some specific limitations and pitfalls. For example machine learning techniques are limited to few labeled corpora and may result in domain specific classifiers. Also it's difficult to specify emotional features to train the classifier. Keyword-based methods have some shortcomings as well which is the restriction to the fixed emotion related words and their imperfection in detecting emotion of sentences without any specified emotional words or ambiguous keywords. In addition to these shortages, there are some substantial challenges related to the nature of emotion detection problem. Some of these important challenges are as follows [3]:

- There is still no standard emotion categorization model accepted by all of the researchers.
- Some texts have different emotions dissolved into text and into each other which makes the task so difficult.
- Sometimes emotion is embedded through semantic structure of sentence using implicit expressions such as metaphors, proverbs or even sarcasms.
- According to the appraisal theory the emotion perception of people from a specific situation may vary based on their experiences, goals or evaluations. So there is no guarantee to achieve the same emotion perception of someone who has described a situation from his own point of view [4].

The last two challenges address the implicit emotion detection system which is the ideal detection that can confront everyday life situations and bring the most intelligence into human-computer interactions. Some researches have been done in this field but their results sound unreasonable and it seems that there is still a long way to go to evolve these kinds of emotion detection systems. So in this study we have focused on implicit emotion detection and have proposed a combinational system consisted of three main subsystems and an information fusion component to aggregate the results. With this combinational system we can analyze our input data from different aspects and use them in a complementary way.

In the following we have organized the structure of our paper as follows. In section 2 we present a literature survey of textual emotion detection. In Section 3 through each subsection, we elaborate the architecture of subsystems in addition to the information fusion component. The evaluation results are discussed in section 4 and finally section 5 concludes study and provides general discussion.

2. Related Works

Fields such as computational linguistics, Natural Language Processing (NLP) and Affective computing interest in large collections of documents provided through internet or out of many textual analysis tasks and as a result there is a growing area of interest in the automatic detection techniques. This attention formed different methods in emotion detection with three main approaches as keyword based technique, machine learning and hybrid approach which is the combination of machine learning and keyword methods. Also some of the information retrieval techniques are attended and highly used such as statistical models and specifically VSM (vector space model) which is most of the times combined with dimensionality reduction methods to achieve better performance. [5] is one of these efforts used VSM as a supervised method with ISEAR dataset as training and testing the results on 801 news headlines provided by "Affective task" in SemEval 2007. It just considered 5 emotions: anger, disgust, fear, joy and sad with F_1=0.32 tested on SemEval and F_1=0.28 tested on ISEAR dataset.

[6] proposed an unsupervised approach which is based on PMI(Point wise Mutual Information) scores to measure the semantic relatedness of words with each emotion category and compute an emotion vector for them. Also some syntactical dependencies have been considered to support semantic aspects such as negation or complementary adjectives. This research is based on the presence of keywords related to the seed words of each emotion class. So without these words system cannot detect the sense of the text and also the selection of seed words for emotion classes will affect the overall performance. The results are evaluated on Alm and ISEAR datasets with 4 emotions: joy, anger combined with disgust, fear and sad results F1=0.52 on ISEAR and F1=0.57 on Alm dataset.

In [7] an emotion detection system for online news has been built which is composed of a document selection module, POS (Part Of Speech) tagging and social emotion lexicon generation part which results in averaged Pearson's correlation coefficient=0.52 tested on dataset consisted of 40897 Chinese news articles. [8] is one of the successful efforts in implicit emotion detection which is based on ISEAR dataset which results in building EmotiNet as a knowledge base for representing and storing affective reactions to real-life contexts and in [9] the authors have used EmotiNet to explore its limitations and compare its performance in an implicit detection system with other common methods. The results were tested on 1081 ISEAR examples filtered out 175 samples used to build the core of EmotiNet. Although this method involves doing some complex preparations such as extracting action chains and finding the tuple of actor, action type and patient for each action but the final results are in the same range of the other common methods with averaged-F_1=0.45.

In [10] the authors exploited Extreme Learning Machine (ELM) to develop a cognitive model for emotion categorization which can represent any concept in an affective space with four dimensions: pleasantness, attention, sensitivity and aptitude. The proposed system called ELM-based model was evaluated with 6813 common sense concept and on patientOpinion dataset consist of 2000 patient opinions. The major effort of this paper was on building a cognitive model and used it in Opinion Mining to detect just positive and negative opinions upon 6 categories of clinical service, communication, food, parking, staff and timeliness which results in averaged-F_1=0.81.

[11] automatically generated an emotion related dataset from 2.5 million tweets of Tweeter by gathering related hash tags and used them in two machine learning technique (MNB and LIBLINEAR), considered 7 emotions: joy, sadness, anger, love, fear, thankfulness and surprise results in F_1=0.53. Although a lot of these efforts have been done in emotion detection field but most of them focus on explicit expressions and implicit efforts are still imperfect and have a long way to go.

3. Methodology

In this study we proposed a combinational system to detect implicit emotion expressions which is composed of four main subsystems in addition to an information fusion component to aggregate their results and decide the final emotion label of documents. Each subsystem implements one of the main approaches in emotion detection and analyzes data from a different aspect. The first one is a machine learning system which is able to capture semantic and syntactic information of text with feature extraction and training a SVM classifier. As the second module we have used Vector Space Model (VSM) categorized as statistical approaches with ability to extract the word level information and map our input data into the affective space which will result in a complementary aspect of data analysis. The last subsystem is a keyword-based method to detect some situation descriptions with an explicit emotion expression at the first or the last of text. Finally the information fusion component gathers the results of modules and annotates the test document if and only if the emotion labels of all three subsystems are the same, else it's left abandoned.

3.1 Overview of the Proposed System

Figure 1 shows the overall structure of our main system which includes three subsystems with an information fusion component to integrate the result of subsystems and produce the final output as emotion label of test documents. Information fusion encompassed theories, techniques or tools in order to aggregate information acquired from different sources and its goal is to achieve a decision or action which is better qualitatively or quantitatively than the result of each individual sources[12]. In the following sections we will introduce each part of our system elaborately.

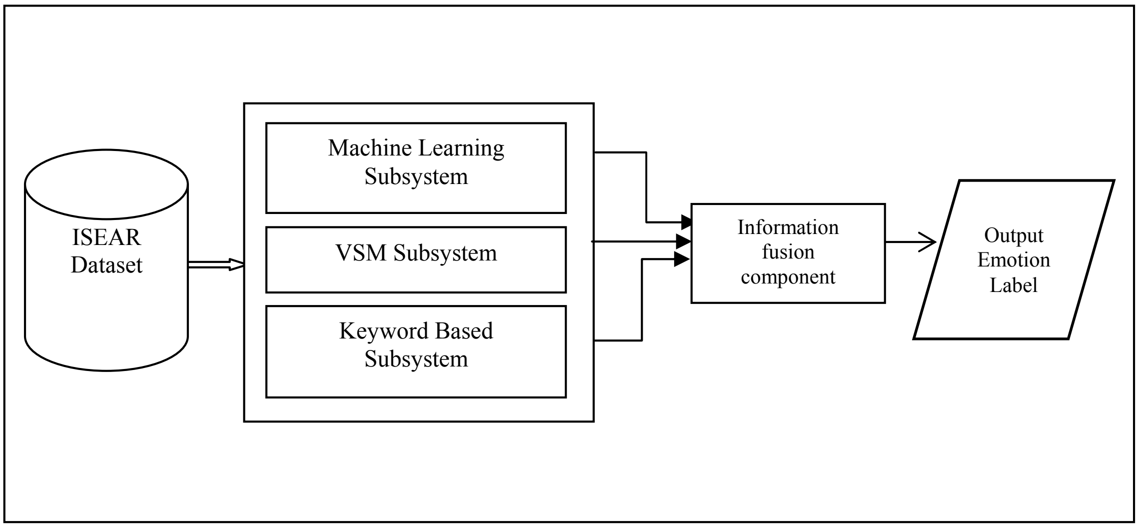

Figure 1: Overview of the emotion detection system.

3.2 Machine Learning Subsystem

Figure 2 shows the structure of this subsystem.

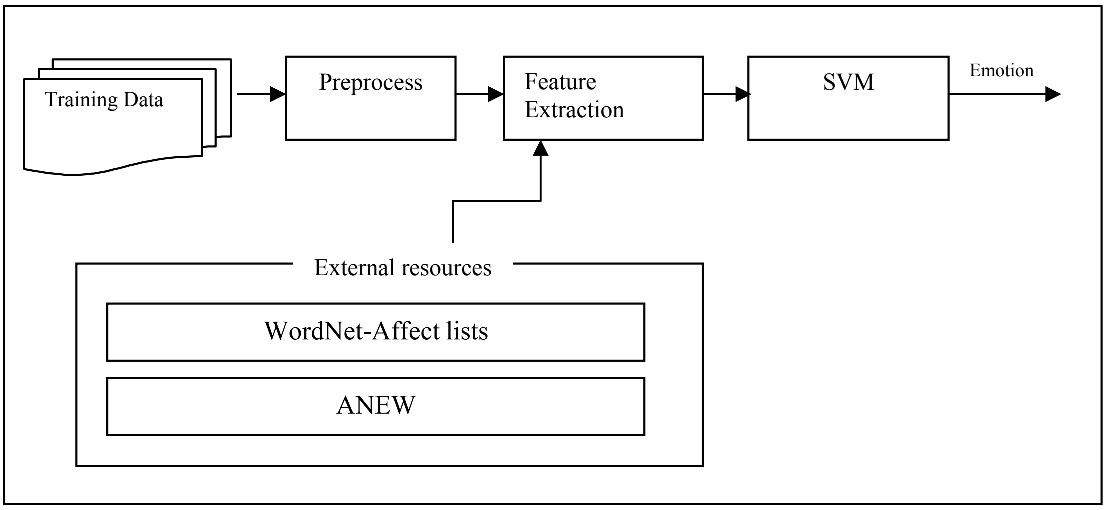

Figure 2: structure of machine learning subsystem

As the first step in the preprocess, we have filtered out all of the documents with [NO RESPONSE] content that shows the participants' inability to describe a suitable situation with the specified emotion. Then the short negative form of expressions such as "shouldn't" is transformed into the long form: "should not". In the next step, we tokenized our input data, split them into sentences and used morphological analyzer to achieve the root of each token. As punctuations and numbers are useless in emotion contents, they are removed except exclamation ("!") and question marks ("?") which are thought to bear some emotional context. Then we removed tokens with less than 3 characters because we have considered 3 as a threshold for meaningful words which results in eliminations of some common stop words such as "a" and "an" or some abbreviated names and titles (Mr and Ms) and even metrics such as "km".

After preprocess, we can extract different features to train our SVM classifier. These features are as follows:

Unigram: this feature gives us lexical information of documents with extracting root of tokens and weighting them according to the tf-idf scheme. Using tf-idf is the best choice because it considers both word frequency (tf) and at the same time using idf it gives higher scores to the rare and uncommon words.

Bigram: using bigram help us to consider negation in syntactical level of documents and also have some level of semantic information. Again it uses the root of tokens with tf-idf weighting system.

POS tags: distribution of POS tags in subjective and objective documents are different even in negative and positive texts [13]. So counting the number of each part of speech in training documents as ngram of POS tags, can help us to capture useful syntactical information.

Frequency of emotion words in WordNet-Affect lists: WordNet-Affect lists are the list of words found in WordNet synsets organized as separated lists according to the 6 emotion classes of Ekman model. For each synset in WordNet They contain synset's number, its POS tag and the words related to the specified emotion [14]. As the intersection of the seven emotions in ISEAR dataset with Ekman model results in five emotions, we have used 5 lists including anger, joy, sad, disgust and fear. So in the training documents, we counted the number of words found in WordNet-Affect lists if they matched the POS tags of the listed words and created a 5 dimensional vector feature.

ANEW score: ANEW (Affective Norm of English Words) is a set of normative emotional ratings for a collection of English words (N=1035) where after reading the words, subjects reported their emotions in a three dimensional representation. This collection provides the related values for valence, arousal and dominance using Self-Assessment Manikin (SAM). So for each word w, the normative database provides coordinates $\bar{w}$ in an affective space as [15]:

$$\bar{w} = (valence,\ arousal,\ dominance) \tag{1}$$

As a feature we find the occurrences of ANEW words in each document and calculate the average of scores as follows:

$$\text{Avg} = \frac{\sum_{i=1}^{n} \bar{w}}{n} \tag{2}$$

Where n is the number of all ANEW words found in a document.

The last two features where used in combination of commonly features such as ngram to capture more semantic information of input data. For example ANEW scores are conventionally used in dimensional emotion classification methods but we have used them as a feature in categorical classification.

After extracting these features in the feature extraction phase of our machine learning subsystem, they are fed into SVM classifier to build the model with which the test documents are evaluated and labeled.

3.3 VSM Subsystem

Figure 3 shows the overall structure of this subsystem.

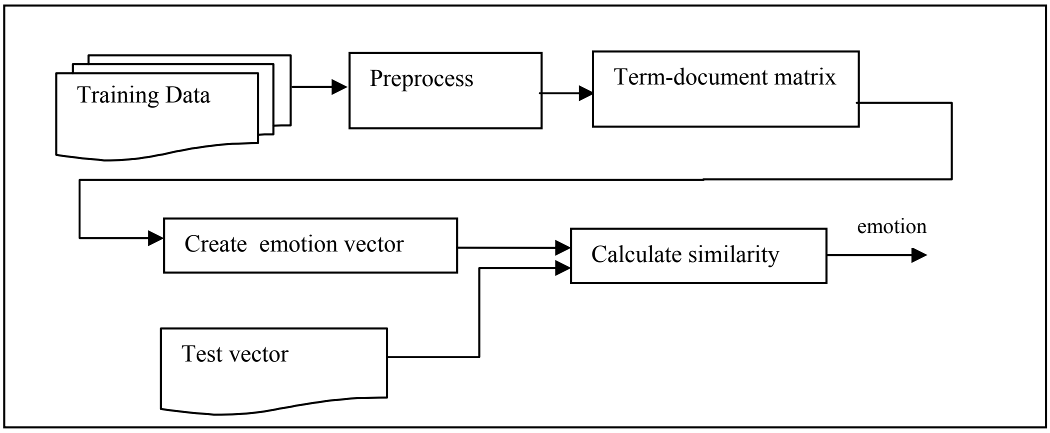

Figure 3: Structure of VSM subsystem

All of the preprocess steps are the same as machine learning subsystem but here the stop words are removed as well. This removal eliminates noisy and unimportant information and reduces the size of term-document matrix which is built with the root of tokens and as a result it counts all different forms of a word as its base form.

If we denote the document vector of an arbitrary sample, $\bar{d}_i$ as follows:

$$\bar{d}_i = < w_{i1}, w_{i2}, w_{i3}, \ldots, w_{in} > \tag{3}$$

Where w_{ik} is the weight of word k in document i and is calculated according to the tf-idf schema, then for each emotion class C_j, the set of documents labeled as C_j are aggregated to form its base vector as follows:

$$E_j = \frac{\sum_{d_i \in C_j}^{|C_j|} d_i}{|C_j|} \tag{4}$$

Where $|C_j|$ is the number of documents in C_j.

In order to label the test document we calculated its similarity with the base vector of each emotion class E_j which is cosine angle between two vectors and defines as:

$$\text{sim}\left(d_t, E_j\right) = \frac{d_t . E_j}{|d_t||E_j|} \tag{5}$$

Where d_t is the vector of test document and $|d_t|$ is the size of d_t. $|E_j|$ is the size of E_j as well. Finally the label of test document is selected according to the label of the base vector with maximum similarity as follows:

$$VSM\left(d_t\right) = argmaxj\left(sim\left(d_t, E_j\right)\right) \tag{6}$$

3.4 Keyword Based Subsystem

Figure 4 shows the structure of this subsystem.

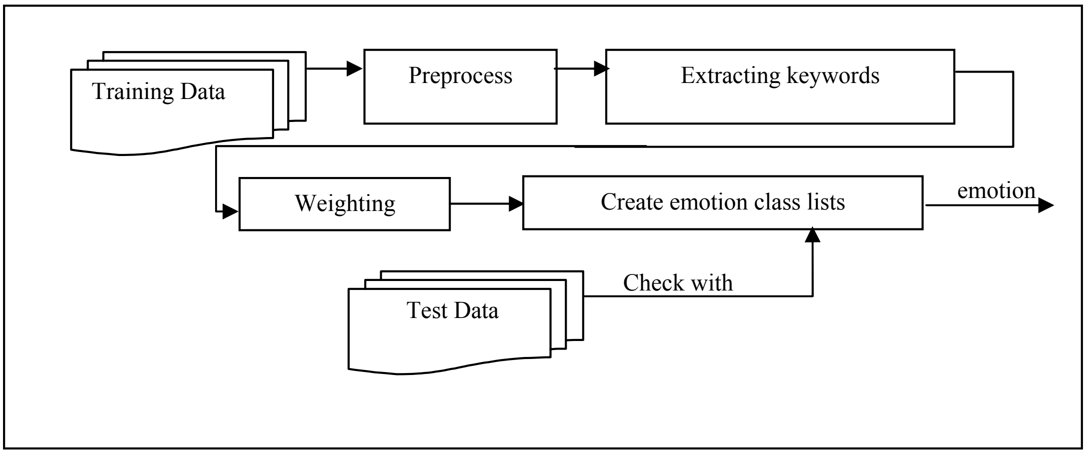

Figure 4: Structure of keyword based subsystem

In this module again all of the preprocess steps are the same as machine learning and VSM subsystem. In order to extract keywords we removed stop words and as a result key phrases are left which are weighted as follows:

$$\text{WLLS}\left(w_i, c_j\right) = P(w_i \mid c_j) \log \frac{P(w_i \mid c_j)}{P(w_i \mid \sim c_j)} \tag{7}$$

Where $P(w_i|c_j)$ is the ratio of frequencies of w_i in emotion class c_j to the count of all words in c_j and $P(w_i|\sim c_j)$ is the ratio of frequencies of w_i in all emotion classes except c_j to the count of all words in these classes. This schema weights words specific to an emotion class c_j higher than the words with balanced distribution in all of emotion classes. After this computation, there will be seven lists for each class with its collected keywords and their correspondent weights; So as to label the new samples, after stop word removal and extracting key phrases, we sum up the weight of each word for each emotion class. The label of the class with maximum value will be selected as the final emotion label of the test document. Score of class e is computed as follows:

$$\text{score}(e) = \sum_{i=1}^{n} \text{weight}_e(k_i) \tag{8}$$

Where $\text{weight}_e(k_i)$ is the weight of keyword k_i according to the emotion class e and n is the total number of keywords found in the document from the emotion list e. final label is selected as follows:

$$\text{EmotionLabel}(d_t) = \text{label of argmax}(\text{score}(e)) \tag{9}$$

This subsystem can efficiently detect explicit expressions in some implicit situations and reduce the complexity of our overall system while increasing the accuracy when facing these kinds of examples.

3.5 Information Fusion Component

This component receives the results of subsystems as input and decides the final output of the whole system. In this level, we have examined two of the simplest ways of information fusion to avoid a large amount of computation. In the first option we can annotate documents if at least two of the three subsystems agree on the same emotion or as an alternative way we can restrict our system outputs just to the documents with emotion labels which are the same in all of the three subsystems. As these modules work complementary and analyze data from different aspects, choosing the second way of aggregation sounds more reasonable and the results in the evaluation section confirms this theory. In a formal way we can say:

$$\text{emotion}(d_t) = \begin{cases} e & \text{if } e_{\text{machine leaning}} = e_{\text{VSM}} = e_{\text{keyword}} \\ \text{null} & \text{otherwise} \end{cases} \tag{10}$$

It means that the test document d_t will receive emotion label e if and only if all of the three subsystems label it as e else it’s left without label and deserted.

4. Analysis and Results

In this study we have used ISEAR (International Survey on Emotion Antecedents and Reactions) dataset. ISEAR consists of 7666 sentences in which 1096 participants from fields of psychology, social sciences, languages, fine arts, law, natural sciences, engineering and medical in 16 countries described an experience or reaction of their everyday life to seven emotions including joy, fear, anger, sad, shame, guilt and disgust [16].

ISEAR is one of the most challenging datasets in emotion detection because in comparison to the other datasets such as Alm or SemEval, the emotions are expressed more implicit and this is one of the main reasons why this dataset is much more used in the new methodologies with implicit emotion detection approaches. But the remaining challenges of ISEAR relates to the style of respondents' description which is shown in Table 1 for the anger. Some participants describe the emotion elicited situation in multi sentences while the other just mention with clauses that make the semantic parsing of documents more unviable. Also according to the appraisal theory, in a specific situation, experiences and goals of each person will result in a different emotional experience. All of these challenges make ISEAR examples closer to the real life emotional situations and make our system more realistic.

Table 1: ISEAR anger samples

"A close person lied to me"
"A colleague asked me for some advice and as he did not have enough confidence in me he asked a third person."
"When I was child and did not achieve my goals."

We have implemented all components of our proposed system with GATE (General Architecture of Engineering) except the VSM subsystem which was developed in Matlab. GATE is a powerful open source tool in natural language processing and is developed by Sheffield University under GNU license [17]. The tokenization, sentence splitting, POS tagging and morphological analysis were all done by ANNIE which is an information extraction system used in GATE as the main processing component.

With filtering of ISEAR and elimination of [NO RESPONSE] documents, 87 samples were removed from the total number of 7666 documents in this dataset which results in removing 9 anger, 13 sad, 24 shame, 5 sad, 4 joy, 17 guilt and 15 disgust documents. This elimination causes the imbalanced distribution of emotion classes especially the shame category which lost most of its samples. Therefore we selected 1075 documents from each emotion class in order to have the same share of data in the training phase. Then the total number of 7525 documents were split into training set and test set with the ratio of 80:20 meaning 6020 documents in the training set and the remaining 1505 documents as the test set. Table 2 shows the results of the SVM classifier which was implemented by SVMLib as the machine learning subsystem.

Table 2: The result of machine learning subsystem with SVM classifier

	F_1	Precision	Recall
Anger	0.503	0.482	0.525
Disgust	0.556	0.603	0.516
Fear	0.757	0.738	0.776
Guilt	0.525	0.574	0.483
Joy	0.703	0.679	0.730
Sad	0.584	0.619	0.553
Shame	0.525	0.482	0.576

The overall performance of the subsystem in the term of averaged-F_1 is 0.5947 with the same precision and recall= 0.5947.

In the next experiment we have investigated the effect of extracted features. First of all unigram and bigram features have been considered as the base set and then each feature was added to these two elements to show its effect on the overall performance of the classifier. Unigram and bigram can be considered as the base feature set because they can capture the essential lexical information and even syntactical level of data with considering negation in bigram.

The results are shown in table 3 where precision and recall values have been calculated according to the predefined formula in GATE software and results in the same value as F_1. In this table, Feature set 1 consists of unigram and bigram of tokens as the base set and in the feature set 2 in addition we have added the frequencies of affective words found in WordNet-Affect lists which shows 1% improvement in the results. Feature set 3 integrated ANEW scores with feature set 1 while in the feature number 4, POS tags are added to the base set. Finally set 5 is the combinations of all features consisted of unigram, bigram, ngram of POS tags, frequencies of affective words and ANEW scores.

Table 3: the effect of each feature set in SVM classifier

	F_1
Feature set 1	**0.5867**
Feature set 2	**0.5947**
Feature set 3	**0.5887**
Feature set 4	**0.5894**
Feature set 5	**0.5967**

As the results show, the effect of POS tags in feature set 4 and ANEW scores in feature set 3 are trivial and they don't effect on the overall performance of the system. The deficiency of POS tags can be interpreted as the result of the short length of documents in ISEAR dataset. Most of these samples have at most 3 sentences and even some situations has been described with clauses which underestimate the effect of POS tags and show us that this feature would be more effective in long poses and bulky texts. Also distribution of POS tags will be more meaningful in explicit examples where the count of some tags such as adjectives in a specific emotion class would be different from another one. But in our case most of the examples are in the same style expressing emotion in an implicit way through describing a situation and the distribution of POS tags would be the same without any affective information. Also the low effect of ANEW scores can be the result of limited words in this lexical dataset.

In table 4 we investigated the effect of stop word removal and as the results show, the elimination of these words deteriorated the output of machine learning subsystem. It shows that their presence is important in some features such as bigram. In the other hand some of the ANEW words are removed during this process and some of POS tags such as pronouns are eliminated as well. Although the importance of these two features are insignificant against the overall performance but their proportion of deterioration in F-measure should be considered.

Table 4: the effect of stop words removal in the result of SVM classifier

	F_1	Precision	Recall
Without stop word removal	0.5967	0.5967	**0.5967**
With stop word removal	0.0769	0.0769	**0.0769**

The output of VSM subsystem is shown in table 5 which results in average F_1=0.522, Precision= 0.559 and Recall=0.522.

Table 5: the result of VSM (Vector Space Model) subsystem

	F_1	Precision	Recall
Anger	0.439	0.412	0.469
disgust	0.455	0.565	0.381
Fear	0.676	0.754	0.614
Guilt	0.510	0.457	0.576
Joy	0.590	0.527	0.669
Sad	0.493	0.772	0.362
shame	0.494	0.427	0.586

Table 6 shows the output of keyword based subsystem and in figure 5 we can compare the result of subsystems with each other. In figure 5 we can see that machine learning has the best performance compared to VSM and keyword methods because in this subsystem the negation in syntactical level is captured while the other two modules cannot consider it. Again in comparison of VSM with keyword subsystem, VSM outperformed which was predictable because ISEAR documents with explicit emotion expressions are rare and most of the samples describe emotion through semantic layers which is out of reach of keyword-based methods.

Table 6: The results of keyword based subsystem

	F_1	Precision	Recall
anger	0.140	0.143	0.140
disgust	0.466	0.550	0.404
fear	0.654	0.741	0.586
guilt	0.474	0.392	0.6
Joy	0.549	0.455	0.693
Sad	0.476	0.410	0.567
shame	0.450	0.441	0.460

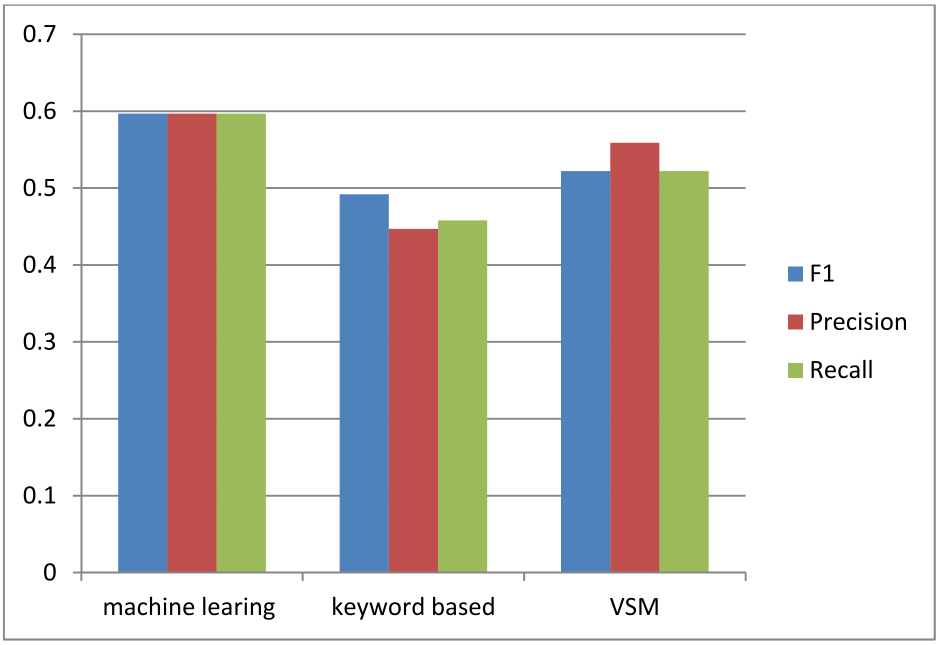

Figure 5: Comparing the performance of subsystems with each other

As shown in table 2, 5 and 6 in all of these subsystems, fear and joy emotions have higher F_1 than the other classes. It shows that people are more aware about these two feelings and they can describe them clearly with more explicit expressions or through more specified key phrases. However anger is the emotion with the least F_1 in all subsystems and it can show that expressing this feeling is much harder for people. Perhaps some times they are unaware of annoying factors in environment to describe or it might be the result of each person's perception of anger which is different from the each other.

Finally we have investigated the ways of aggregating the output of our overall system. Table 7 shows the result of aggregation in majority voting when just two subsystems out of the three agree on the same emotion and table 8 shows the result in the situation when all of the subsystems should agree on the same tag.

In the first situation 1347 documents from the total number of 1505 samples are labeled which means 89% of documents but the overall F_1=0.58 is lower than F_1=0.59 in the machine learning subsystem. In the second situation although 47% of data received emotion tag which means 721 documents, but the performance of the overall system with averaged-F_1=0.68 is 9.13% better than machine learning subsystem, 16.6% better than VSM and shows 23% improvement in keyword based subsystem. Figure 6 shows this improvement in the graphical way.

Table 8 shows a huge progress in the term of F_1 for each emotion class compared to the single subsystems except anger emotion with a dramatical decrease due to the poor results of the keyword-based module. As mentioned anger was the hardest emotion to detect in all of the subsystems and the keyword module had the lowest F-scores therefore it declines the performance of the whole system for anger emotion. So if we put anger aside and just consider the remaining 6 emotion classes, the averaged-F_1=0.76 will be resulted which means 17%, 24% and 31% improvement against machine learning, VSM and keyword-based modules respectively. Again table 8 shows 10% better performance than the majority voting which implies that the results of subsystems are complementary and all of them should be considered in the voting phase but in fact we believed that not all of them have the same saliency and priority. We can give them suitable weights which is considered as the future work and discussed in the next section.

Table 7: The results of combinational system with majority voting

	F_1	Precision	Recall
anger	0.470	0.551	0.410
disgust	0.549	0.613	0.497
fear	0.722	0.774	0.678
guilt	0.577	0.540	0.619
joy	0.635	0.545	0.762
sad	0.640	0.75	0.558
shame	0.533	0.477	0.605

Table 8: The results of combinational system with agreement of all three subsystems

	F_1	Precision	Recall
anger	0.22	0.25	0.18
disgust	0.734	0.730	0.738
fear	0.854	0.823	0.886
guilt	0.728	0.744	0.712
joy	0.805	0.744	0.876
sad	0.756	0.839	0.688
shame	0.719	0.673	0.772

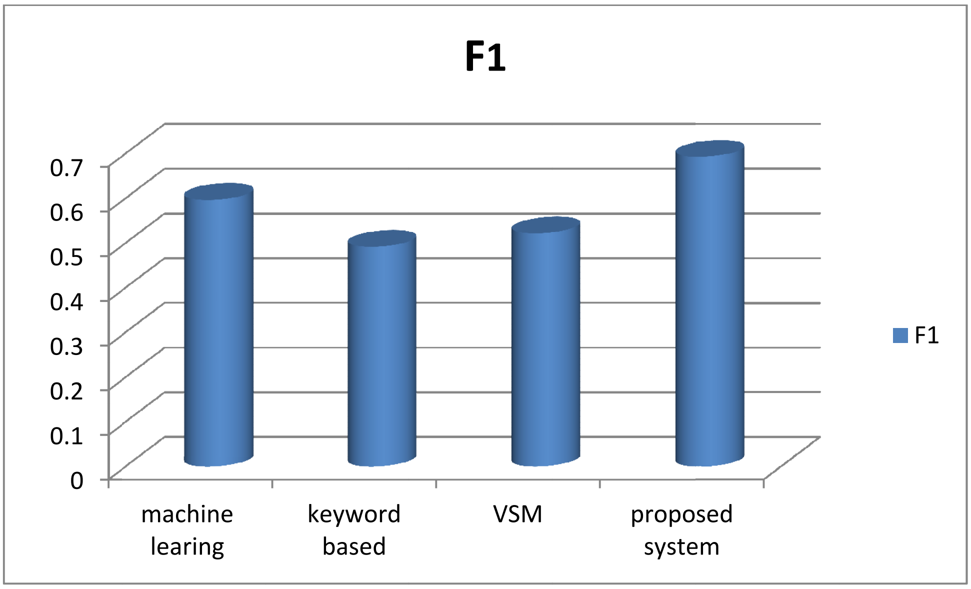

Figure 6: The performance of proposed system in comparison to each subsystem

Although we can use the results of each component as the baseline and interpret the superiority of our combinational system over individual modules as its novelty and superiority against commonly used methods in emotion detection field but in a convenient way and in order to have a better vision of our model's performance we have compared it with some of the novel or common methods almost trained with ISEAR dataset.

Table 9: Comparing the result of combinational system with the other common methods

Method	Dataset	Emotions	Overall F_1
Keyword Baseline	ISEAR	4 classes: joy, sad, anger-disgust, fear	0.32
SVM	ISEAR	5 classes: anger, disgust, fear, joy, sad	0.68
LIBLINEAR	Twitter Dataset	7 classes: Joy, sad, anger, love, fear, thankfulness, surprise	0.53
EmotiNet	ISEAR	7 classes: Joy, sad, anger, disgust, fear, guilt, shame	0.45
Our SVM module	ISEAR	7 classes: Joy, sad, anger, disgust, fear, guilt, shame	0.59
Our VSM module	ISEAR	7 classes: Joy, sad, anger, disgust, fear, guilt, shame	0.52
Our Keyword-based module	ISEAR	7 classes: Joy, sad, anger, disgust, fear, guilt, shame	0.45
Our Combinational System	ISEAR	7 classes: Joy, sad, anger, disgust, fear, guilt, shame	**0.68**
Our Combinational System	ISEAR	6 classes: Joy, sad, disgust, fear, guilt, shame	**0.76**

Although our model annotates 47% of documents but its performance in the term of F_1 is much higher than the other methods especially when anger is eliminated from the set of emotional tags which results are shown in the last column.

5. Conclusion and Future Work

In this paper, we proposed an emotion detection system which is composed of three subsystems: machine learning, VSM and keyword based module. We showed that each subsystem can analyze data from a different aspect. Their results are aggregated and used to annotate the test document if and only if all of the three subsystems agree on the same emotion class otherwise it is left abandoned. The performance of the proposed system is 9.13% better than machine learning subsystem, 16.6% better than VSM and 23% better than keyword based method.

In the future work, we will weight the output of each subsystem according to its performance and the vote of subsystem with higher F_1 in a specific emotion class will be prioritized the other two subsystems.

In addition we can use dimension reduction methods in our VSM subsystem to reduce the dimension of term-document matrix and perform semantic analysis with the help of LSA or PLSA method.

References

[1] Pang, Bo, and Lillian Lee. "Opinion mining and sentiment analysis."Foundations and trends in information retrieval 2.1-2 (2008): 1-135.

[2] Zhe, Xu, and A. C. Boucouvalas. "Text-to-emotion engine for real time internet communication." Proceedings of International Symposium on Communication Systems, Networks and DSPs. 2002.

[3] C Jain, Mukesh, and V. Y Kulkarni. "TexEmo: Conveying Emotion from Text-The Study." International Journal of Computer Applications 86.4 (2014): 43-49.

[4] Roseman, Ira J., and Craig A. Smith. "Appraisal theory: Overview, assumptions, varieties, controversies." (2001).

[5] Danisman, Taner, and Adil Alpkocak. "Feeler: Emotion classification of text using vector space model." AISB 2008 Convention Communication, Interaction and Social Intelligence. Vol. 1. 2008.

[6] Agrawal, Ameeta, and Aijun An. "Unsupervised emotion detection from text using semantic and syntactic relations." Web Intelligence and Intelligent Agent Technology (WI-IAT), 2012 IEEE/WIC/ACM International Conferences on. Vol. 1. IEEE, 2012.

[7] Lei, Jingsheng, et al. "Towards building a social emotion detection system for online news." Future Generation Computer Systems 37 (2014): 438-448.

[8] Balahur, Alexandra, et al. "Emotinet: A knowledge base for emotion detection in text built on the appraisal theories." Natural Language Processing and Information Systems. Springer Berlin Heidelberg, 2011. 27-39.

[9] Balahur, Alexandra, Jesús M. Hermida, and Andrés Montoyo. "Detecting implicit expressions of emotion in text: A comparative analysis." Decision Support Systems 53.4 (2012): 742-753.

[10] Cambria, Erik, et al. "An ELM-based model for affective analogical reasoning."Neurocomputing 149 (2015): 443-455.

[11] Wang, Wenbo, et al. "Harnessing twitter" big data" for automatic emotion identification." Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom). IEEE, 2012.

[12] Waltz, Edward, and James Llinas. Multisensor data fusion. Vol. 685. Boston: Artech house, 1990.

[13] Pak, Alexander, and Patrick Paroubek. "Twitter as a Corpus for Sentiment Analysis and Opinion Mining." LREC. Vol. 10. 2010.

[14] Strapparava, Carlo, and Alessandro Valitutti. "WordNet Affect: an Affective Extension of WordNet." LREC. Vol. 4. 2004.

[15] Bradley, Margaret M., and Peter J. Lang. Affective norms for English words (ANEW): Instruction manual and affective ratings. Technical Report C-1, The Center for Research in Psychophysiology, University of Florida, 1999.

[16] Fischer, Agneta H., et al. "Gender and culture differences in emotion." Emotion4.1 (2004): 87.

[17] Cunningham, Hamish. "GATE, a general architecture for text engineering."Computers and the Humanities 36.2 (2002): 223-254.