Feeler: Emotion Classification of Text Using Vector Space Model

Taner Danisman[1] and **Adil Alpkocak**[1]

Abstract. Over the last quarter-century, there is increasing body of research on understanding the human emotions. In this study, automatic classification of anger, disgust, fear, joy and sad emotions in text have been studied on the ISEAR (International Survey on Emotion Antecedents and Reactions) dataset. For the classification we have used Vector Space Model with a total of 801 news headlines provided by "Affective Task" in SemEval 2007 workshop which focuses on classification of emotions and valences in text. We have compared our results with ConceptNet and powerful text based classifiers including Naive Bayes and Support Vector Machines. Our experiments showed that VSM classification gives better performance than ConceptNet, Naive Bayes and SVM based classifiers for emotion detection in sentences. We achieved an overall F-measure value of 32.22% and kappa value of 0.18 for five class emotional text classification on SemEval dataset which is better than Navie Bayes (28.52%), SVM (28.6%). We have tested and discussed the results of classification using cross-validation technique for emotion classification and sentiment analyses on both the ISEAR and SemEval datasets. In addition to the classification experiments we have developed an emotion enabled video player which automatically detects the emotion from subtitle text of video and displays corresponding emoticon.

Keywords: Emotion Detection in Text, Vector Space Model, Emotion Perception, Human Computer Interaction.

1 INTRODUCTION

Current state-of-art in computer human interaction largely ignores emotion whereas it has a biasing role in human-to-human communication in our everyday life. In the mean time, a successful computer human interaction system should be able to recognize, interpret and process human emotions. Affective computing could offer benefits in an almost limitless range of applications. However, the first step is Human Emotion Recognition (HER), and it is getting more attention recently. In HER, the data gathered to recognize human emotion is often analogous to the cues that humans use to perceive emotions in others. Hence, human emotion recognition is multimodal in nature, and includes textual, visual and acoustic features. Text seems to be the most studied modality since the text is relatively easier to process than others.

HER from text can be simply envisioned to be a classification problem of a given text according to predefined emotional classes. In this case, it first requires a preparation of proper training set for each emotional class and selection of good features. One of the solutions for this issue is Bag of Word (BoW). It's very similar to keyword spotting [5] and lexical affinity [6]. BoW approach that is widely used in information retrieval, and tries to generate a good lexicon for each emotional class and feature extraction. However, creation of emotional lexicon is both time consuming and labor-intensive task since usually requires manual annotations. On the other hand, the number of words in lexicons is very limited, and it is not desired for most classifiers using the BoW approach. Moreover, user's vocabulary may differ from the document vocabulary. In literature, an alternate approach for this issue has been proposed by [1]. They use blog based emotion datasets, where every blog document is already labeled by authors. It seems that it is good for generating large scale lexicon for a better representation for a given language. Blogs have more than 200-300 words per document on average. However, assigning a single emotional label to a document having many words is not very meaningful. Therefore, a better training set for each emotional class must consider sentences and words, not paragraphs. After preparing a proper training set and selecting good features, the next task is to classify a given text.

To date, many approaches have been proposed for HER from text. These approaches can be grouped into three main groups: keyword spotting, statistical NLP, and ontology based approaches. Each approach has its own advantages and disadvantages. In addition, there is no rigid line between these approaches. Keyword spotting is easy to implement, and based on predetermined set of terms to classify the text into emotion categories. Despite its simplicity, creation of an effective lexicon is difficult too since only 4% of words used in texts have emotional value [14]. For these reasons it is not suitable for wide range of domains. The second group is based on statistical NLP approaches. This approach is similar to lexical affinity where affinities of words are still used but as a feed for a machine learning algorithm. In case of lexical affinity, words have some probabilistic value representing the affinity for a particular emotion class. However, it requires high quality, large-scale training dataset for a better classification. The third groups is based on ontologies, heavily uses semantic networks like WordNet-Affect [4] and ConceptNet [15] are linguistic resources for lexical representation of affective information using commonsense knowledge. ConceptNet is an integrated commonsense knowledgebase with a natural language processing toolkit MontyLingua which supports many practical textual reasoning tasks over real world documents without additional statistical training.

In this paper, we propose a VSM approach for HER from text.

[1] Computer Engineering Department, Dokuz Eylul University, Tinaztepe Campus, 35160 Izmir/TURKEY.
Email: {**taner,alpkocak}@cs.deu.edu.tr**

- We have used sentences from ISEAR [2] dataset, emotional words from Wordnet-Affect and polarity of words from WPARD datasets.
- Our approach uses Vector Space Model for HER.
- We measured the effect of stemming and emotional intensity on emotion classification in text.
- Third, we have developed an emotion enabled video player which automatically detects the emotions from subtitle text and displays emotions as emoticons during video play.

The rest of the paper is organized as follows. In section two, we have explained the emotion classification problem in text. In section three, Vector Space Model and methodology is presented. Section four shows the experimental results performed on the SemEval test set. Finally, section five concludes the study and provides a general discussion.

2 PROBLEM DEFINITION and RELATED WORK

One side of the problem is the selection of a qualified dataset for machine learning methods. In order to cover most of the words in a given language, a large-scale dataset is needed. In addition this dataset should have variation of emotional content, independent emotional responses from different cultures to eliminate cultural affects of emotion.

Manual creation of large-scale datasets is difficult and time consuming task. Blog based datasets provides large-scale lexicons as presented in [1]. They worked on large collection of blog posts (122,624 distinct web pages) for classifying blog text according to the mood reported by its author during the writing. According to their results, increasing the amount of training data leads an additional increase in classification performance. On the other hand, the quality of the dataset is important for better classification.

All these requirements lead us to use ISEAR (International Survey on Emotion Antecedents and Reactions) dataset in our experiments. ISEAR consists of 7,666 sentences and snippets in which 1096 participants from fields of psychology, social sciences, languages, fine arts, law, natural sciences, engineering and medical in 16 countries across five continents completed a questionnaire about the experiences and reactions to seven emotions in everyday life including joy, fear, anger, sadness, disgust, shame, and guilt. Surprisingly, ISEAR dataset is not studied yet for text based emotion classification. Previous studies using the ISEAR dataset try to find relationships among emotions and different cultures, genders, ages, and religions. On the other hand this corpus is well suited to use for emotional text classification purposes. Table 1 shows samples from this dataset for the anger emotion.

"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".
"A colleague asked me to study with her. I could not explain things as perfectly as she had expected. So she reacted in an aggressive manner." ....

Table 1. ISEAR anger samples

2.1. Related Work

Achievements in this domain can be used in next generation intelligent robotics, artificial intelligence, psychology, blogs, product reviews, and finally development of emotion-ware applications such as emotion-ware Text to Speech (TTS) engines for emotional reading of text. CRM and service oriented companies like Right Now Technologies and NICE Systems produces customer service software SmartSense™ and NICE Perform™ respectively which recognizes customer emotions using keyword spotting technique and prosodic features of speech then performs flagging, prioritizing and routing inquiries and customers based on emotional content.

[7] developed a new aggregator to fetch news from different news resources and categorize the themes of the news into eight emotion types using semantic parsers and SenseNet [8]. [9] studied the natural language and affective information using cognitive structure of affective information. They developed ALICE chat-bot based on Artificial Intelligence markup language (AIML) script to improve interaction in a text based instant messaging system that uses emoticons or avatar that represents the sensed emotion to express the emotional state.

According to [10] emotion annotation for text is a hard problem and inter-annotator agreement value k=.24-.51. [11] employed a commonsense knowledgebase OMCS (Open Mind Common Sense) having 400,000 facts about everyday world to classify sentences into basic emotions (happy, sad, angry, fearful, disgusted, and surprised) categories. [5] developed an emotion extraction engine that can analyze the input text in a chat dialogue, extract the emotion and displays the expressive image on the communicating users display. Their parser only considers sentences in present continuous tense, sentences without starting auxiliary verbs (No question sentences allowed), positive sentences, etc. [12] considered the emotional expressions for text-to-speech engines and emotional reading. They partitioned the text into nouns adjectives and adverbs and used the frequency of words to determine the emotional class. [13] tried to detect emotion from both speech and textual data. They manually defined the emotional keywords and emotion modification words. They have used "very" and "not" as a modification word where the only difference between "very happy", "happy", and "not happy" is the emotional intensity. As they are using keyword-spotting technique (they have 500 words labeled as emotion words), they reported that textual recognition rate is lower than speech based recognition. According to their work, emotion recognition performance of multimodal system is better than performance of individual modalities.

3 VECTOR SPACE MODEL

Vector Space Model (VSM) is widely used in information retrieval where each document is represented as a vector, and each dimension corresponds to a separate term. If a term occurs in the document then its value in the vector is non-zero. Let us assume that we have *n* distinct terms in our lexicon. Then, lexicon, ℓ, is represented as a set of ordered terms, and more formally, it is defined as follows:

$$\ell = \{t_1, t_2, t_3, \ldots, t_n\}$$

Then, an arbitrary document vector, $\vec{d}_i$, is defined as follows:

$$\vec{d}_i = \langle w_{1i}, w_{2i}, ..., w_{ni} \rangle$$

where w_{ki} represents the weight of k^{th} term in document i. In literature, there several different ways of computing these weight values have been developed. One of the best known schemes is *tf-idf* weighting. In this scheme, an arbitrary normalized w_{ki} is defined as follows;

$$w_{ki} = c(t_k, d_i) = \frac{tf_{ik} \log(N/n_k)}{\sqrt{\sum_{k=1}^{n} (tf_{ik})^2 [\log(N/n_k)]^2}} \text{ where;}$$

$t_k = k^{\text{th}}$ term in document d_i

tf_{ik} = frequency of the word t_k in document d_i

$idf_k = \log\left(\frac{N}{n_k}\right)$ inverse document frequency of word t_k in entire dataset

n_k = number of documents containing the word t_k,
N = total number of document in the dataset.

Each emotion class, M_j, is represented by a set of documents, $M_j=\{d_1,d_2,\ldots,d_c\}$. Then, we have created a model vector for an arbitrary emotion, $\vec{E}_j$, by taking the mean of $\vec{d}_j$ vectors for an arbitrary emotion class. More formally, each $\vec{E}_j$ is computed as follows:

$$\vec{E}_j = \frac{1}{|M_j|} \sum_{\vec{d}_i \in M_j}^{|M_j|} \vec{d}_i$$

where $|M_j|$ represents the number of documents in M_j. After preparing model vectors for each emotion class, the whole system is represented with a set of model vectors, $D=\{E_1,E_2,...,E_s\}$ where s represents the number of distinct emotional classes to be recognized.

In VSM, documents and queries are represented as vectors, and cosine angle between the two vectors used as similarity of them. Then normalized similarity between a given query text, Q, and emotional class, E_j, is defined as follows:

$$\text{sim}(Q, E_j) = \sum_{k=1}^{n} w_{kq} * E_{kj}$$

In order to measure the similarity between a query text and the D matrix of size $s \times n$, first we convert the query text into another matrix $n \times 1$ similar to D where n is the size of the lexicon and s is the number of emotions. Then for each emotion (each row of D matrix), we make multiplication between the query matrix Q and one row of D matrix. After these multiplications we have m scalar values representing the cosine similarity. The index of the maximum of these values is selected as the final emotional class. More formally:

The classification result is then,

$$\text{VSM}(Q) = \arg\max_j (\text{sim}(Q, E_j))$$

The basic hypothesis in using the VSM for classification is the contiguity hypothesis where documents in the same class form a contiguous region, and regions of different classes do not overlap.

4 EXPERIMENTATION

Before starting on a research on emotion classification, the first question is "Which emotions should be addressed?" There are many different emotion sets exists in the literature including basic emotions, universal emotions, primary and secondary emotions, neutral vs. emotional, and for some cases the problem is reduced to a two class classification problem (Sentiment Analysis) using the Positive and Negative values as class labels. Simple classification sets give better performance than expanded sets of emotions which require cognitive information and deeper understanding of the subject. In our research study, we have used five emotion classes (anger, disgust, fear, sad, and joy) that form the intersection between the ISEAR dataset and the SemEval test set. Therefore, the number of emotion classes s=5.

For the classification, we have used Naïve Bayes, Support Vector machines and Vector Space Model classifiers. We have considered the effect of the stemming, negation and intensity of emotions on classification performance. We have used WEKA tool [16] for the Naïve Bayes and SVM classification. In order to compare the performance of VSM and other classifiers, we have considered the mean F1-measure value and the kappa statistics which considers the inter-class agreements.

First, we have used set theory, which deals with collections of abstract objects to find the intersections and set differences of objects in a given set. For the graphical simplicity, we only show three emotional classes (anger, disgust, and fear) with a few words in Figure 1 where each circle represents an emotional class and entries represent the words.

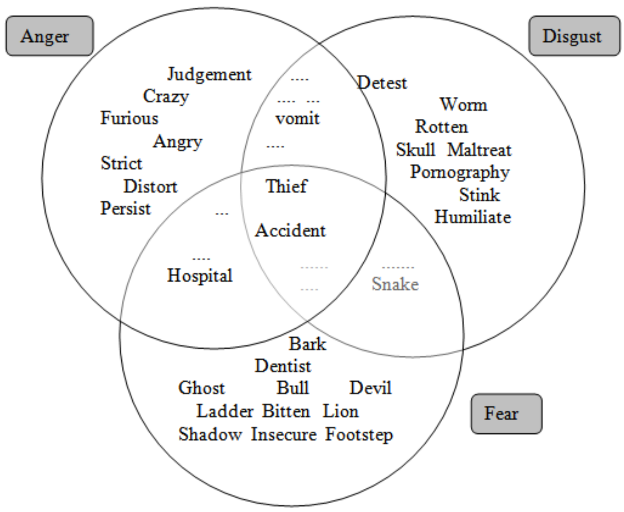

Figure 1. Words belonging to specific emotions in ISEAR dataset after finding set differences

Using the set difference words, the word "ghost, bark, dentist and lion" is appeared only in sentences representing the fear emotion whereas "successful, rejoice, sunshine, tremendous, accredit, and carnival" appeared only in joy sentences. The following table represents the automatically extracted emotion words, which are non-intersected words in ISEAR database.

Anger	Disgust	Fear	Joy	Sad	Shame	Guilt
rules	detest	ghost	happiest	weep	aloud	pinch
strict	worm	bark	joyous	sadden	zip	decline
disregard	rotten	dentist	envelop	coma	infidelity	feign
judgment	skull	devil	success	farewell	underwear	counselor
crazy	hypocrite	ladder	rejoice	drought	smear	harsh
hatred	stink	bitten	ecstatic	tragic	dialect	caretaker
persist	humiliate	lion	sunshine	saddest	spell	insinuate
…	…	…	…	…	…	…

Table 2. Emotion related words from ISEAR dataset using set difference for each emotion set.

These results give hope us to use tf-idf (Term Frequency-Inverse Document Frequency) values for emotion classification because these non-intersected words have very low term frequency values (among 1-10) but have high inverse document frequency values. In addition, by using the tf-idf values we are also able to use the words in intersected areas. Therefore, we selected to use VSM (Vector Space Model) for emotion classification in text.

4.1. Stop Word removal strategy

This study also showed that some words are only appeared in specific sentences belonging to a single emotion, so stop-word removal based on minimum term frequencies is not suitable for emotion detection. Stop words are usually the most frequent words including articles (a, an, the), auxiliary verbs (be, am, is, are), prepositions (in, on, of, at), conjunctions (and, or, nor, when, while) that do not provide additional improvement for search engines but increase the computational complexity by increasing the size of the dictionary. The important aspect of stop-word removal in emotion detection is the words, not their frequencies. There are several publically available stop-word lists available where these lists consist of approximately 400-500 most frequent words in a given language. However, public stop-word lists consider the information retrieval and they do not consider words carrying emotional content. Therefore we first need to remove some of the emotional words from the stop-word list including negative verbs (not, is not, does not, do not, should not, etc.). In addition, we replaced the word "very" with blank and the word "blank not blank" is replaced by "blank not". In addition, Words in Table 2 are removed from the stop-word list to improve the classification rate. We ignored the part of speech tagging on input text because of its effect of reducing the classification accuracy as described in [17].

Since non-alpha tokens are automatically removed by TMG [18], the exclamation marks and question marks are replaced by descriptive new words "XXEXCLMARK" and "XXQUESMARK" respectively. Negative short forms are also replaced by negative long forms such that "doesn't" is replaced by "does not". After these replacements, the following sentences are changed as follows:

"I don't love you!" => "I do not love you XXEXCLMARK" => "I do NOTlove you XXEXCLMARK"

"I am not very happy." => "I am not happy." => "I am NOThappy."

As seen in the above examples, the word "happy" and "love" is used to create new words "NOTlove" and "NOThappy". In this way, we can discriminate the word "love" having positive meaning and "NOTlove". In the same way, the new word "NOThappy" has a negative meaning.

Initially we have used stemming for finding morphological root of a given word. Stemmers in linguistic are widely used in search engines and query based systems to improve the efficiency of these systems. For emotion classification, stemming also removes the emotional meaning from the words. We found that tense information also affects the emotional meaning of the words. For example the words "marry" and "love" is frequently shown in joy sentences while the words "married" and "loved" are appeared in sad sentences.

4.2. Training and Test sets

For training, we have used combination of ISEAR, Wordnet-Affect and WPARD datasets. Testing is performed on SemEval Task 14 "Affective Text" test set.

Our main training dataset, ISEAR, is further expanded by adding emotional words from Wordnet-Affect [4] and WPARD (Wisconsin Perceptual Attribute Rating Database) [3] to improve the emotional classification of sentences. Each word in Wordnet-Affect and WPARD is replicated up to average number of terms per document which is 16 (as seen on Table 4) in our experiment to make ISEAR like sentences. In this case, the sentences are constructed using the same words.

WPARD is like a polarity dataset were collected from 342 undergraduate students using online form to rate how negative or positive were the emotions they associated with each word, using a scale from -6 (very negative feeling) to +6 (very positive feeling), with 0 being a neutral feeling. Table 3 shows samples from this dataset.

Word	Value	Word	Value
rape	-5.60	hope	+4.43
killer	-5.55	honeymoon	+4.48
funeral	-5.47	home	+4.50
slavery	-5.41	sunset	+4.53
cancer	-5.38	beach	+4.58
corpse	-4.95	family	+4.58
slave	-4.84	friend	+4.60
war	-4.78	peace	+4.62
coffin	-4.73	kiss	+4.64
morgue	-4.72	holiday	+4.73
cigarette	-4.49	fun	+4.91

Table 3. Sample cross-section from WPARD [3] dataset

Before extracting the features, we have preprocessed the ISEAR dataset and manually eliminated some of the inconsistent and incomplete entries (such as "[No response]" lines). Normalization is performed using the TMG toolbox [18] and get the following distribution as seen in Table 4.

Emotion	Number of sentences	# of words before stop word removal	Average # of terms before normalization	Average # of terms after normalization
Angry	1,072	26,3	24.8	17.7
Disgust	1,066	22,8	21.6	15.8
Fear	1,080	25,6	23.9	17.1
Joy	1,077	21,1	19.8	14.2
Sad	1,067	21,3	20.2	14.6
Shame	1,052	24,9	23.9	16.9
Surprise	1,053	23,5	22.6	15.9
Average	1,066	23,7	22.4	16.0

Table 4. Number of sentences per emotion in ISEAR Dataset

SemEval Task 14 “Affective text” test set is used for testing. Table 5 shows the sample cross-section in XML format and Table 6 shows corresponding ground truth for this test set.

```
<corpus task=”affective text”>
<instance id="500">Test to predict breast
cancer relapse is approved</instance>
<instance id="501">Two Hussein allies are
hanged, Iraqi official says</instance>
<instance id="502">Sights and sounds from
CES</instance>
<instance id="503">Schuey sees Ferrari unveil
new car</instance>
…
```

Table 5. Sample cross-section from SemEval test set

Instance Id	Anger	Disgust	Fear	Joy	Sadness	Surprise
500	0	0	15	38	9	11
501	24	26	16	13	38	5
502	0	0	0	17	0	4
503	0	0	0	46	0	31
…	…	…	…	…	…	…

Table 6. Corresponding ground truth data for SemEval test set

4.3. Experiments

In order to build up the D matrix, first we made normalizations including limited stop-word elimination, term-length thresholds, which is 3 in our case. We did not consider global and local thresholds. Average number of terms per document before the normalization is 22.43 and after the normalization number of index terms per document is 16 and the dictionary size is 5,966 terms. This result leads us to a D matrix of size 7,466×5,966. As the size of average number of index term elements per document is 16, the D matrix is very sparse. After computing E_j vectors, the new size is 5×5,966.

After the normalization step, we have computed the term frequency and inverse document frequency (tf-idf) values that provide a level of information about the importance of words within the documents. The tf–idf weight is a weight often used in information retrieval and text mining. This weight is a statistical measure used to evaluate how important a word is to a document and how important the word is to all the documents in the collection.

Experiment 1: We studied the effect of emotional intensity to classification performance on the SemEval test set. In our experiment we have selected emotions having either positive or negative valence value greater than a threshold *T* where *T* is between 0-70. According to Figure 2, F1-Measure value increases proportionally with the *T* when *T* is between 30 to 70. It shows that increased emotional intensity also increases the classification performance.

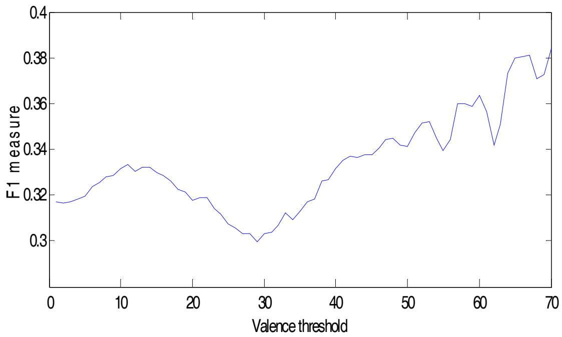

Figure 2. Valence threshold versus F1-measure on VSM classifier

Experiment 2: We have studied the effect of stemmers on emotion classification in text using 10 fold cross-validation on the ISEAR dataset and on unseen test set SemEval. A stemmer is an easy to implement algorithm which determines a stem (or morphological root) form of a given inflected (or, sometimes, derived) word form. In some cases it is known as suffix remover. We found that some words having the same morphological root can have different emotional meaning. For example, the word “marry” classified as joy and “married” is classified as sad emotions. In spite of this, those samples are very limited. Our experiments showed that use of the stemming algorithms still gives additional increase in classification accuracy as seen in Table 7, Table 8, and Table 9 for Naïve Bayes, SVM and VSM classifiers. Bold values represent the best scores considering three classifiers.

Training Set	Stemming	Naïve Bayes Classifier 5 Class Emotional Classification					
		10 Fold cross validation on the ISEAR dataset			Test on SemEval test set		
		Kappa	Mean F1	Accuracy	Kappa	Mean F1	Accuracy
ISEAR	Yes	.59	67.0	67.2	.14	27.1	31.3
	No	.59	67.2	**67.4**	.09	23.3	26.8
ISEAR+WPARD+ WORDNET_AFFECT	Yes	.51	61.2	60.8	.16	**29.0**	35.0
	No	.46	57.8	57.0	.12	25.3	30.8

Table 7. Naive Bayes results for five class emotion classification

		Support Vector Machine 5 Class Emotional Classification					
		10 Fold cross validation on the ISEAR dataset			Test on SemEval test set		
Training Set	**Stemming**	**Kappa**	**Mean F1**	**Accuracy**	**Kappa**	**Mean F1**	**Accuracy**
ISEAR	Yes	.59	67.5	**67.4**	.11	24.5	27.2
	No	.58	67.0	66.9	.09	23.4	26.4
ISEAR+WPARD+ WORDNET_AFFECT	Yes	**.61**	**68.3**	70.2	.12	24.9	27.0
	No	.56	65.0	67.1	.09	23.7	28.0

Table 8. Support Vector Machine results for five class emotion classification

		Vector Space Model Classifier		
		SemEval test set		
Training Set	**Stemming**	**Kappa**	**Mean F1**	**Accuracy**
ISEAR	Yes	0.16	28.7	**36.0**
	No	0.11	26.1	32.0
ISEAR+WPARD+ WORDNET_AFFECT	Yes	**0.17**	28.5	34.8
	No	0.11	25.5	32.0

Table 9. Vector Space Model results for five class emotion classification

In addition to stemming experiment, we have considered the effect of adding emotional words from Wordnet-Affect and WPARD dataset into our training set. Results showed that, classification performance increased for Naïve bayes and SVM classifiers but in case of VSM the performance is reduced and there is only a small increase in kappa. This is because; we only added the word itself not sentences in our training set. Therefore during the normalization step, words come from Wordnet-Affect and WPARD behaved like a document which results a decrease in accuracy as seen in Table 8.

Experiment 3: In this experiment, we only considered positive and negative classes. Therefore, we combined the anger, disgust, fear, and sad emotions in Negative class while joy is the only member of the Positive class. Table 10 shows the results of this classification for different classifiers where the best performance for cross-validation comes from SVM classifier with 79.5% F-Measure value and 59.2% with VSM classifier.

Previous studies achieve up to 42.4% F1-measuse using coarse-grained evaluation for polarity detection on this dataset as reported in [19] while VSM approach achieves 59.2% F1-measure.

For emotion classification, previous studies on this dataset achieves up to 30.3% F1-measure for single class and 11% on average for six-class emotion classification using coarse-grained evaluation. Evaluation criteria of these studies can be found in [19].

Our results achieve up to 49.6% F1-measure for single classes and 32.2% on average for five-class emotion classification as seen on Table 11.

Classifier	**Test method/set**	**Positive**	**Negative**	**Overall F1**
Naïve Bayes	10Fold Cross Validation /ISEAR	64.1	89.9	74.8
Naïve Bayes	SemEval	55.3	60.6	57.8
libSVM	10Fold Cross Validation /ISEAR	**69.0**	**93.8**	**79.5**
libSVM	SemEval	49.9	**66.3**	56.9
VSM	SemEval	**59.1**	59.4	**59.2**

Table 10. Experimental results for polarity in terms of F-Measure using cross-validation on the ISEAR dataset

For the stop word experiment , “English.stop” file from Porter stemmer and “common_words” file from TMG are used. As seen on Table 11, almost all best F-Measure (mean of precision and recall) scores come from our classifier with 32.22% value.

In case of ConceptNet, we have used XML-RPC based client to communicate with ConceptNet server. For the evaluation, ConceptNet outputs a prediction vector $P(S) = \langle p_1(S), p_2(S), \ldots, p_m(S) \rangle$ of size m where S represents a sentence or a snippet, $p_i(S)$ represents prediction value of i^{th} emotion class for the sentence S. Final classification result selects the maximum of $p_i(U)$ and assigns the corresponding class label using

$$P(S) = \arg\max_i \; p_i(S)$$

Classifier	**Stop Word**	**Anger**	**Disgust**	**Fear**	**Joy**	**Sad**	**Overall F1**
Naïve Bayes	Porter	20.2	5.2	41.9	39.6	32.6	27.9
Naïve Bayes	Tmg	21.5	5.4	**42.7**	40.5	32.5	28.5
libSVM	Porter	17.7	**9.5**	39.0	42.7	34.1	28.6
libSVM	Tmg	14.5	8.8	40.0	42.0	33.9	27.8
VSM	Porter	22.1	9.1	40.1	49.2	**37.1**	**31.5**
VSM	Tmg	**24.2**	9.3	41.1	**49.6**	36.7	**32.2**
ConceptNet	N/A	7.8	**9.8**	16.8	**49.6**	26.3	22.1

Table 11. Experimental results (in terms of F1-Measure) for emotions trained from ISEAR and tested on SemEval Test set

We have also create a video player which detects the emotion of subtitle texts and speech signal using the VSM and SVM classifiers trained on the ISEAR and displays the corresponding emoticon as seen in Figure 3. Emotion detection in speech signal is performed using ensemble of support vector machines.

Figure 3. Emotion-ware video player screenshot from Finding Nemo[2]

5 CONCLUSION and FUTURE WORK

In this paper, we proposed a VSM approach for HER from text. We measured the effect of stemming and emotional intensity on emotion classification in text. We showed that Vector Space model based classification on short sentences can be as good as other well-known classifiers including Naïve Bayes and SVM and ConceptNet.

We also studied the effect of stemming to emotion classification problem. According to our experiments, use of stemming removes and decreases the emotional meaning from words. But these examples are very rare in our test set therefore use of stemming still increases the classification performance for all classifiers.

Finally, we have developed an emotion enabled video player, which shows video, emotional states and valence information at the same time.

As future work, we are planning to combine multiple modalities in video (audio, visual and text) to improve the classification performance.

Acknowledgments. This study was supported by the Scientific and Technological Research Council of Turkey (TUBITAK) Project No: 107E002.

6 REFERENCES

[1] G. Mishne, Experiments with mood classification in blog posts. In: *Style2005 – 1st Workshop on Stylistic Analysis of Text for Information Access*, at SIGIR (2005).

[2] K. R. Scherer and H.G. Wallbott, Evidence for universality and cultural variation of differential emotion response patterning. *Journal of Personality and Social Psychology*, 66:310-328 (1994).

[3] D. A. Medler, A. Arnoldussen, J.R. Binder, and M.S. Seidenberg, The Wisconsin Perceptual Attribute Ratings Database. http://www.neuro.mcw.edu/ratings/ (2005).

[4] C. Strapparava and A. Valitutti., WordNet-Affect: an affective extension of WordNet. In: *Proceedings of the 4th International Conference on Language Resources and Evaluation* (LREC 2004), Lisbon, pp. 1083-1086, May (2004).

[5] A. C. Boucouvalas and X. Zhe, Text-to-Emotion. Engine for Real Time Internet Communication. In: *Proceedings of the 3rd International Symposium on CSNDSP*, Staffordshire University, UK, pp. 164-168, (2002).

[6] A. Valitutti, C. Strapparava, and O. Stock, Developing affective lexical resources. *PsychNology Journal*, 2(1):61-83 (2004).

[7] M. Shaikh, H. Prendinger, and M. Ishizuka, Emotion sensitive news agent: An approach towards user centric emotion sensing from the news. *The 2007 IEEE/WIC/ACM International Conference on Web Intelligence* (WI-07), Silicon Valley, USA, pp. 614-620 Nov. (2007).

[8] M. Shaikh, H. Prendinger, and M. Ishizuka, SenseNet: A linguistic tool to visualize numerical-valence based sentiment of textual data (Poster). *Proceedings 5th International Conference on Natural Language Processing* (ICON-07), Hyderabad, India, pp. 147-152, (2007).

[9] M. Shaikh, H. Prendinger, and M. Ishizuka, A cognitively based approach to affect sensing from text, In: *Proceedings of 10th Int'l Conf. on Intelligent User Interface* (IUI 2006), Sydney, Australia, pp.349-351 (2006).

[10] C. O. Alm, D. Roth, and R. Sproat, Emotions from text: Machine learning for text-based emotion prediction. In: *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 579-586, Vancouver, British Columbia, Canada, (2005).

[11] H. Liu, H. Lieberman, and T. Selker, A model of textual affect sensing using real-world Knowledge. In: *Proceedings of International Conference on Intelligent User Interfaces* (IUI-03) pp. 125–132, (2003).

[12] F. Sugimoto and M. Yoneyama, A Method for Classifying Emotion of Text based on Emotional Dictionaries for Emotional Reading, In: *Proceedings of the IASTED International Conference on Artificial Intelligence and Applications*, Austria, pp. 91-96 (2006).

[13] Z.J. Chuang and H. Wu, Multi-Modal Emotion Recognition from Speech and Text. *International Journal of Computational Linguistics and Chinese Language Processing*, 9(2):1-18 (2004).

[14] J. W. Pennebaker, M. E. Francis, and R. J. Booth, Linguistic Inquiry and Word Count: LIWC 2001. Mahwah, NJ: Lawrence Erlbaum, (2001).

[15] H. Liu and P. Singh, ConceptNet: A Practical Commonsense Reasoning Toolkit. *BT Technology Journal* 22(4):211-226. Kluwer Academic Publishers, (2004).

[16] I. H. Witten and E. Frank (2005) *Data Mining: Practical machine learning tools and techniques*, 2nd Edition, Morgan Kaufmann, San Francisco, (2005).

[17] E. Boiy, P. Hens, K. Deschacht, and M.-F. Moens, (2007) Automatic Sentiment Analysis in On-line Text, ELPUB2007. Openness in Digital Publishing: Awareness, Discovery and Access In : *Proceedings of the 11th International Conference on Electronic Publishing*,Vienna, Austria 13-15, June (2007).

[18] D. Zeimpekis and E. Gallopoulos, Tmg: *A matlab toolbox for generating term-document matrices from text collections.* Technical report, University of Patras, Greece, (2005).

[19] C. Strappava and R. Mihalcea, SemEval-2007 Task 14: Affective Text, In: *Proceedings of the 4th International Workshop on Semantic Evaluations* (SemEval-2007), pages 70–74, Prague, June (2007).

[2] © Finding Nemo is a Copyright owned by Walt Disney and Pixar Entertainment