A supervised multinomial classification framework for emotion recognition in textual social data

Abid Hussain Wani* and Rana Hashmy

Department of Computer Science,
School of Applied Sciences and Technology,
University of Kashmir, India
Email: abid.wani@uok.edu.in
Email: ranahashmy@kashmiruniversity.ac.in
*Corresponding author

Abstract: The task of emotion recognition from text has received much attention since the proliferation of online social networking which has woven itself into the fabric of lives of people world-over. This study is aimed at extracting the lexical and contextual information from the text and combining it with semantic information for the detection of the emotional state of a sentence. We propose a supervised framework for recognition of emotions from text in this work. Our framework utilises word embeddings from Word2Vec to extract the set of words which fall in semantic proximity of an affect-bearing word and also takes into account the context in which the words are used. We incorporate class-specific emoticon features in all our experiments as emoticons are commonly used on social media platforms. As the nature of social media text is generally very informal and has an irregular structure, our framework encompasses an appropriate mechanism to handle it. We evaluate our support vector machine-based framework on stance sentiment emotion corpus (SSEC) and Aman's dataset. The classification results achieved are better than state of art techniques currently available.

Keywords: emotion detection; emoticon mapping; supervised learning; social media analysis.

Reference to this paper should be made as follows: Wani, A.H. and Hashmy, R. (2023) 'A supervised multinomial classification framework for emotion recognition in textual social data', *Int. J. Advanced Intelligence Paradigms*, Vol. 24, Nos. 1/2, pp.173–189.

Biographical notes: Abid Hussain Wani is an Assistant Professor in the Computer Science Department at the University of Kashmir, India. His research interest lies in machine learning, natural language processing and computer networks.

Rana Hashmy is a Scientist in the Computer Science Department at the University of Kashmir, India. She completed her PhD in Computer Science in year 2007 from the Jawaharlal Nehru University, New Delhi, India. Her research interest lies in data mining, biological databases and data warehousing.

1 Introduction

The rapid growth of online social networking, as result of unprecedented advances in computing and communication technologies, has provided people a global venue for expressing and sharing their thoughts and feelings on different topics and events. Nowadays, most of the web users get themselves involved in online messaging and communicate through tweets, blogs and posts. With an unprecedented use of hand-held devices especially smartphones at an almost epidemic level, communication-over-Internet has overtaken all other forms of communication. Hence these Social Networking platforms have become a gold mine for organisations to monitor their reputation and brands by extracting and analysing content posted by the public not only about their own products and services, but also that of their competitors. Analysis of such content faces several challenges due to its typical informal, irregular structure. A good amount of research in affect and emotion recognition in text has been conducted by the people working on opinion mining and sentiment analysis, especially in the domain of product and movie reviews. Opinion mining is concerned with finding personal interpretations of observation about an object which may or may not be emotionally charged. Sentiment analysis, on the other hand, is concerned only about identifying the emotional tendencies of the content that is to classify affect according to its polarity, i.e., POSITIVE, NEGATIVE or NEUTRAL. Emotion recognition, encompasses the identification of fine-grained category of behavioural feeling which may not always be targeted towards an object.

Much of the work in emotion recognition has focused on employing supervised techniques for the identification of emotions as the results achieved are much better when compared with those from unsupervised ones. In this work we focus on the detection of six basic emotions (ANGER, DISGUST, FEAR, HAPPY, SAD and SURPRISE) in text employing support vector machine (SVM) with lexical, semantic as well as contextual features. Although a number of separate studies have been conducted on using lexical, semantic and contextual features to train supervised emotion classification algorithms, we present a supervised approach that combines lexical, semantic as well as contextual information to detect emotions.

2 Literature review

Though much of the research in affect detection has traditionally focused on opinion mining and sentiment analysis but in recent years affect recognition at fine grained categorical level has received remarkable attention (Clavel and Callejas, 2016). Aman and Szpakowicz (2007) used Naïve Bayes and SVM with lexical and non-lexical features for detection of six basic emotions in text. Khan et al. (2014) proposed an unsupervised classification framework for twitter sentiment analysis based on three-way classification algorithm. Agrawal and An (2012) employed an unsupervised context-based approach to detecting emotion from text at the sentence level based on semantic relatedness. Turney (2001) proposes an unsupervised learning algorithm for classifying reviews as recommended or not recommended based on the average semantic orientation of the adjectival and adverbial phrases. Su et al. (2006) perform automatic identification of implicit product features expressed in product reviews. Though an unsupervised classification framework does not require an annotated corpus for training but the classification accuracy achieved is generally lower than the supervised one. A number of other studies for affect detection using supervised machine learning algorithms appear in literature. Yang et al (2007) employed SVM and conditional random field (CRF) machine learning techniques to emotion classification of web blog corpora. They trained the classifiers at the sentence level taking the context of a sentence into account and applied it at the document level. In this work, we focus on the detection of universally accepted six basic emotions, i.e., ANGER, DISGUST, FEAR, HAPPY, SAD and SURPRISE from text by employing the lexical, semantic and contextual information. In our knowledge this is the first study for supervised classification of emotions in text using word embeddings from Word2Vec in conjunction with lexical features and contextual dependencies. While texting on social networking sites, many users have a tendency to express their emotions using emoticons. Boia et al. (2013), found that for more than 90% of the tweets with emoticons, emoticons indicate the correct sentiment orientation of the tweet. In this work, emoticon to emotion class mapping given by Ku and Sun (2012) used in feature building. Ku and Sun have categorised emoticons as ANGER, DISGUST, FEAR, HAPPY, SAD and SURPRISE. Six emoticon features representing the corresponding emotion classes are used in all our experiments. Keeping in view all the above studies, this study attempts at combining a number of lexical, semantic and contextual features and evaluating their relevance for supervised emotion classification of emotions in text.

3 Resources and tools

In this work we undertake a supervised learning approach for the detection of emotions in text utilising a number of language processing resources and tools. For capturing the lexical features we use three different resources: EmoLex, WordNet (WN) and WordNet-Affect (WNA). Relevant semantic information is extracted by employing vector-space model Word2Vec. We employed standard resources and tools in our work to extract feature information from the text. A brief introduction of lexical resources and semantic tools used is presented below:

a Emolex: In our study we employ the emotion associated lexicon EmoLex (also known as NRC Word-Emotion Association Lexicon) (Mohammad and Turney, 2013). For EmoLex, the annotations have been done manually by crowdsourcing using Mechanical Turk. EmoLex is a list of English words and their associations with eight emotion categories (anger, fear, anticipation, trust, surprise, sadness, joy and disgust) and two sentiments (negative and positive). In our study we take into account only those words which have the associated emotion from any of six basic emotion classes, i.e., ANGER, FEAR, SURPRISE, SAD, JOY (HAPPY in this study) and DISGUST (Ekman, 1992). This lexicon comprises of 14,182 unigrams (words).

b WordNet: WordNet (Miller, 1995) is a large lexical database just like a thesaurus wherein English words (nouns, verbs, adjectives and adverbs) are grouped into together based on their meanings. Each group is called a synsets. Synsets are interlinked not just by word forms, strings of letters, but specific senses of words. WN’s structure makes it a useful tool for computational linguistics and natural language processing and is also freely available for download. WN comprises of 117,000 synsets and each synset is linked to other synsets by means of a small number of 'conceptual relations'.

c WordNetAffect: WNA (Strapparava and Valitutti, 2004) assigns a variety of affect labels to a subset of synsets in WN. In WNA a number of WN synsets are assigned one or more affective labels (a-labels). In particular, the affective concepts representing emotional state are individuated by synsets marked with the a-label EMOTION. Besides there are also other a-labels for those concepts representing conditions stimulating moods, or emotional responses. Since this study is constrained to the six basic emotion categories we employed the extracted list from WNA for these emotions.[1] We employ only the stem words extracted from WNA and delete all duplicates.

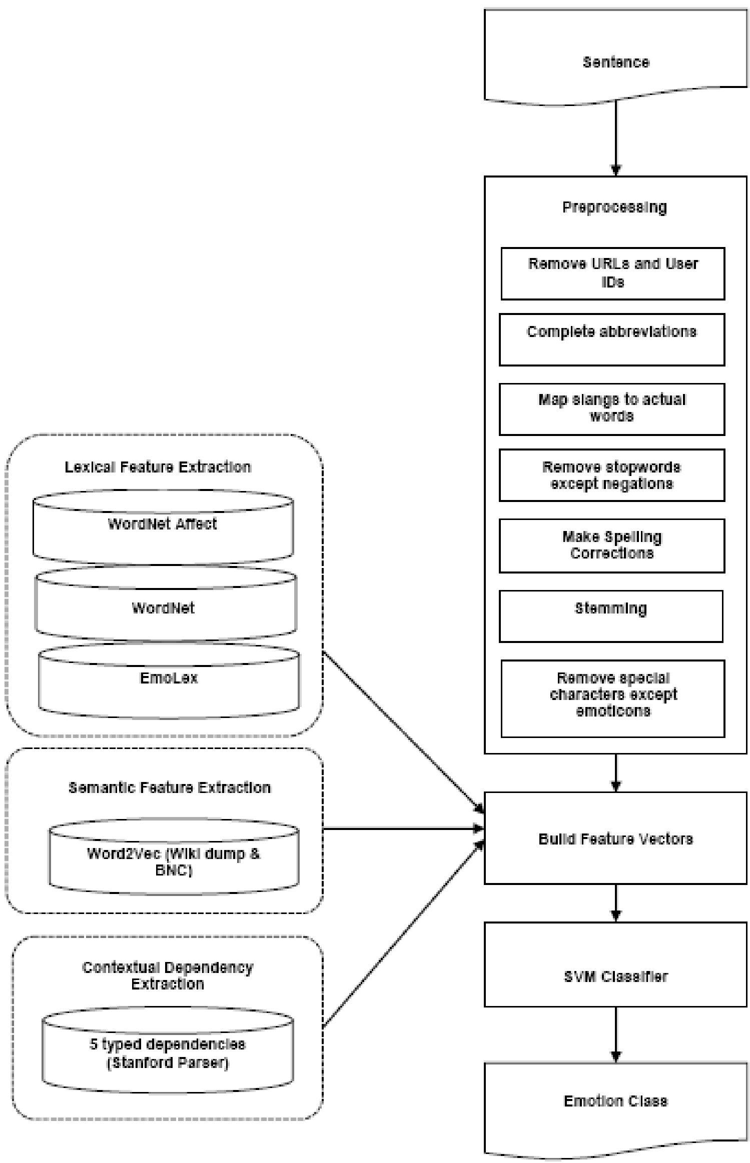
d Word2Vec: Word2vec (Mikolov et al., 2013) is a computationally-efficient predictive model for learning word embeddings from raw text. It takes a raw text corpus as input and produces a set of vectors as its output which essentially represent the feature vectors relevant to words which are there in the input corpus. The potential utility of Word2vec is not limited to just sniffing through the sentences but it can be employed to retrieve patterns from community graphs, codes, genes likes, playlists etc. The principle aim of Word2vec is to build the feature vectors for all the words in the input and cluster the vectors of related words side-by-side in the vector-space which essentially reflects the affinity among them. Word2vec produces the feature vector groupings as output that are essentially dictated by the distribution of their numerical representations. The beauty of Word2vec lies in the fact that it carries out such groupings with no help from the user. Therefore when sufficient data is provided, one can arrive at very good word semantics depending upon its previous usages (e.g., 'tree' is to 'leaf' what 'flower' is to 'petal'). In our study, we employ this Word2vec to extract and group words as per their affect sense.

4 Proposed framework

Emotion detection task is modelled as a multinomial text classification problem. Most of the work in opinion mining and sentiment analysis has focused on using supervised methods for detection of affect in text in order to achieve higher accuracy in results. We also propose to employ a improved and enriched supervised approach for recognition of emotions from text which takes into account lexical, semantic and contextual information about the textual content. Our approach provides high degree of robustness by not only taking into account the surface features of the text like in keyword spotting, but also evaluates the affective qualities of the of text using both semantic and contextual information. This enables the sensing of the emotions from the text even when affect-bearing words are absent. Further the fact that affect is at the sentence level by taking into account semantic and contextual dependencies, there are little chances of our classification framework being tricked by structural features like ambiguity and negation at the word-level. Having dwelled into the guiding idea of our approach in terms of existing approaches and theoretical considerations, the rest of this section will focus on more practical considerations, such as how lexical features are used in conjunction with the language parsing and semantic relatedness features for detecting emotions from textual data. Figure 1 depicts the proposed supervised framework for emotion recognition.

Recognition of emotions comprises of tasks like feature dictionary creation, pre-processing of input text, population of feature vector using lexical, semantic and contextual features and finally supervised classification. The output from the emotion recognition may be simply conveyed to the intended user or passed on to some other application for further processing such as product refinement, strategy shift or other decision making.

Figure 1 Proposed emotion recognition framework



The task of emotion detection from text necessitates the selection of most appropriate emotion modelling paradigm so as to handle all the relevant affect situations under consideration. In this study we employed the categorical model for emotion classification. We concentrate on the detection of set of big six emotions. The major phases of our classification framework are as follows:

a Pre-processing

The chosen datasets for the purpose of our study consists of textual deciphers on social networking sites which generally are characterised by the use of short sentences, informal wording, weak grammar, slangs etc. Thus before being subjected to further processing, the data is first pre-processed and refined in following steps:

- detect URLs and remove all the URLs from the text
- remove all the user identifiers and usernames
- replace abbreviations and slangs by their expansions
- look up for mis-spellings and replace them by the correct ones
- remove the stop words from the corpus
- apply lemmatisation
- barring emoticons, get rid of all special characters.

In order to deal with negation terms such as ‘not’, ‘n’t’, ‘none’, ‘never’, ‘neither’, ‘no’, ‘non’, ‘nothing’, we set the negation feature to 1 if they occur odd number of times, otherwise negation feature remains 0.

b Lexical feature extraction

After pre-processing the input text, we extract the relevant lexical information from WNA and WN for the six basic emotions. Table 1 shows the number of seed words extracted for each emotion class.

Table 1 Extracted words for each emotion class using WNA and WN

Emotion	*WordNet-Affect*	*WordNet*
Anger	26	237
Fear	18	206
Disgust	29	219
Happy	27	870
Sad	28	604
Surprise	15	158

The lexical information extracted is used to populate feature vectors with lexical features. We first extract affect words pertaining to each emotion class from WNA and then look for their synonyms in WN. The framework is further enriched by extracting more lexical features from a popular and large emotion lexicon, Emolex. Although Emolex contains 14,182 emotion annotated words when it works with Plutchik’s emotions, however, we work with 3,462 words relevant to Ekmans’s. Therefore, with the incorporation of Emolex the total number of unigrams goes to 5,899.

Taking a lead from the mapping of emoticons and the emotion classes as proposed by Ku and Sun in (2012), emoticon is treated as a feature in the proposed framework. Class-specific emoticons for different classes used in this study are: 12 for ANGER, 15 for FEAR, 36 for HAPPY, 54 for SAD and 11 for SURPRISE.

c Semantic feature extraction

In order to take into account the semantic features in the classification framework, we train the Word2Vec algorithm by using the Wikipedia Text Dump[2] and British National Corpus (BNC).[3] In our opinion, both these text repositories are well suited for the extraction of semantically related words as they contain data pertaining to a variety of domains and genres. Raw text from these repositories is used to train the Word2vec. With Word2Vec we have two options, the continuous bag-of-words model (CBOW) and the Skip-Gram. Though both the models focus on word embeddings, the CBOW is target-oriented while the skip-gram is source-oriented. Studies show that CBOW is better suited for not-so-bog datasets as it gets smoothed over a lot of the distributional observations. On the other hand, skip-gram is preferred for bigger datasets as it takes each context-target pair to be a fresh example. In this work, we employ the skip-gram model to extract semantic information relevant to six basic emotions. Word2Vec vectors for words in our extended emotion lexicon are looked-up and added as features for relevant emotion classes.

d Contextual information extraction

A word may convey different meaning when used under different contexts. Consider the sentence: 'His speech was incredibly disgusting'. Dropping the stop-words will leave three words 'speech', 'incredibly' and 'disgusting' in the sentence. Using lexical resource 'incredible' will fall under the emotion category HAPPY. However the emotional tendency of 'incredible' gets subtly modified by the word 'disgusting', thereby converting the emotional feel of the phrase 'incredibly disgusting' to resemble more like DISGUST than happiness. This example illustrates that a word may convey different, meaning hence the emotion, in different contexts. If a pure keyword-based method is employed, it would consider the words 'incredible' and 'disgusting' to be HAPPY and DISGUST respectively and cancel out their effect, resulting possibly in a NEUTRAL sentence. However, by taking into account the context, 'disgusting' can influence the emotion status of 'incredibly', thus resulting in the label to be DISGUST. In order to exploit this notion of influencing and dependent words in a sentence, we extract syntactic dependencies between words and employ them to capture some of the context. A syntactic dependency is represented as:

$$\alpha\left(w_1 \downarrow, w_2 \uparrow\right) \tag{1}$$

This is essentially a binary prediction, where α represents a syntactic dependency relation (grammatical relation) and the symbols ↓ and ↑ represent of modified and modifier roles of the relation respectively (Gamallo et al., 2001). Here w_1 is the dependent word while w_2 is the governor word hence the word indexed by '↓' plays the role of modified, whereas the word indexed by '↑' plays the role of modifier. We employed Stanford Parser (De Marneffe and Manning, 2012) for extracting the relevant dependencies. Though Stanford Parser (STNFP) supports about 50 grammatical relations, we focus on five types of dependencies of our interest, namely, noun-phrase-as-adverbial-modifier, negation-modifier, adverb-modifier, adjectival-modifier and adjectival-complement. Noun-phrase-as-adverbial-modifier relation where something syntactically a noun phrase is used as an adverbial modifier in a sentence. For instance in case of the sentence ‘Shares eased a fraction’; npadvmod (eased, fraction). A negation modifier is the relation between any negation word and the word modified by it. For example, in ‘I am not scared of him’; neg(scared, not). An adverb modifier of a word is an adverb or adverb-headed phrase that serves to modify the meaning of the word. For example, in the sentence ‘Genetically modified strains of bacteria’; advmod(modified, genetically). An adjectival modifier of a Noun Phrase is any adjectival phrase that serves to modify its meaning. For example, ‘John eats fried chicken’; amod(chicken, fried). An adjectival complement represents a phrase that acts as an object for it. For example, the adjectival complement dependency from ‘The food smells delicious’ is acomp(smells, delicious), where ‘delicious’ is the adjectival complement of the verb ‘smells’. We selected these five dependencies for our study as the relation objects in these dependencies can potentially modify the meaning of affect-bearing words in a sentence.

Once the sentence is parsed, we get the dependencies which are used to build the dependency features. We extended the scheme proposed by Na et al. (2012) for using contextual dependencies as features in sentiment analysis and used it to capture contextual dependencies for the set of basic emotions. The governor and dependent terms in type dependencies are converted to appropriate emotion labels using the emotion lexicon, EmoLex, so as to utilise prior scores of subjective terms. For instance, for ‘amod(car, fabulous)’, we have the following type dependency features: amod(NEUTRAL,HAPPY):1, amod(NEUTRAL,ANGER):0, amod(NEUTRAL, DISGUST):0,amod(NEUTRAL,FEAR):0, amod(NEUTRAL,SAD):0 and amod(NEUTRAL,SURPRISE): 0.

1. Word which is a noun phrase or serves as the front of an adverbial phrase which influences the semantics of a word or gets its semantics influenced. Such a dependency is ‘npadvmod’.
2. Word which changes the meaning of another word to its opposite is in ‘neg’ dependency with it.
3. Word which is non-clausal adverb or forms the front of an adverbial phrase is in ‘advmod’ dependency with its influenced word.
4. Word which is an adjective and modifies a noun or vice versa forms an ‘amod’ dependency.
5. Word which that augments the meaning of an adjective or modifies it is in ‘acomp’ dependency.

5 Evaluation and results

In this section we present the results obtained from different experiments carried on two standard datasets and evaluate the proposed classification framework. The datasets employed are the gold standard Aman Corpus and stance sentiment emotion corpus (SSEC). Aman's dataset encompasses a corpus of blog posts annotated with six basic emotions whileas SSEC is a Twitter dataset annotated with emotion labels. A number of studies have employed Aman's dataset in emotion classification evaluation studies making it a gold standard (Canales et al., 2016) but in our knowledge this is the first study to on SSEC dataset for a supervised classification utilising lexical, semantic and contextual features. Both of these datasets have been annotated according to the categorical emotion modelling paradigm using six basic emotions. We conducted several experiments to evaluate our framework for detection of six target emotions. Since emotion recognition is modelled as a text classification task in this study, we evaluate the accuracy and effectiveness of the proposed framework using confusion matrices, precision, recall and F-measure as is a norm with the other similar studies. A confusion matrix describes the performance of a classifier on a set of test data for which the true values are known. The diagonal elements in the confusion matrix represent the correctly classified data for each class whereas all other elements show incorrectly classified data. The most basic terms (which are whole numbers) used for expressing a classifier's performance are:

- True positive (tp): number of sentences correctly classified as belonging to emotion category e.
- True negative (tn): number of sentences correctly classified as not belonging to emotion category e.
- False positive (fp): number of sentences incorrectly classified as belonging to emotion category e. (Also known as a 'Type I error').
- False negative (fn): number of sentences incorrectly classified as not belonging to e. (also known as a 'Type II error').

For an emotion e, precision is obtained by taking the ration of true positives to all predicted positives, i.e., both true positives and false positives. Mathematically we have:

$$precision = \frac{tp}{tp + fp} \tag{2}$$

Recall is obtained by taking the ration between correctly classified positives by the classifier and manual classified positives (true positives + false negatives), i.e.,

$$recall = \frac{tp}{tp + fn} \tag{3}$$

F-measure is the harmonic mean of both precision and recall. Mathematically we have,

$$F = \frac{2 \times precision \times recall}{precision + recall} \tag{4}$$

Accuracy is defined as the proportion of true positive, true negatives and true NEUTRALs (true results) from all the given data.

$$accuracy = \frac{tp + tn + t_{neutrals}}{size\_of\_total\_data} \tag{5}$$

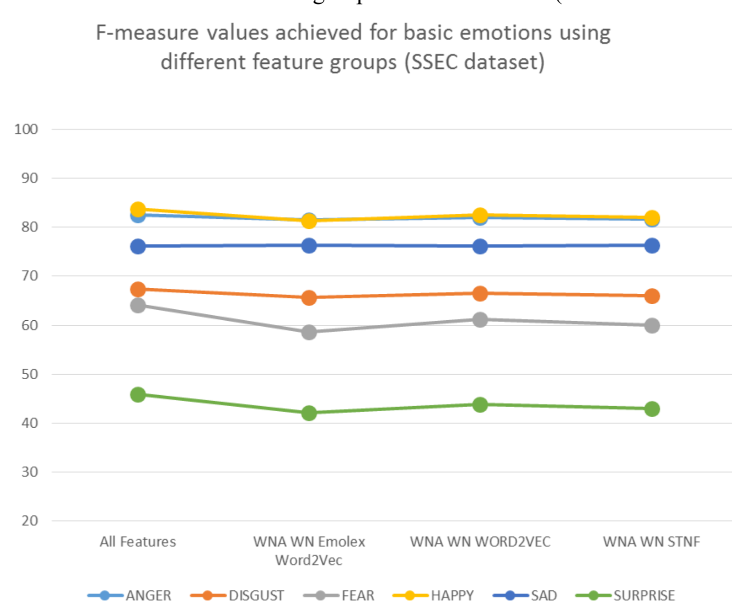
where $t_{neutrals}$ is the total number of true NEUTRALs.

a Evaluation on SSEC dataset

To train and evaluate our classifier, we employed emotion annotated SSEC (Mohammad et al., 2017), manually annotated by Schuff et al. (2017). For SSEC dataset the annotation has been carried out by six annotators. The emotion annotated corpus encompasses an aggregated annotation as well as the individual labels of each of six annotators. Though SSEC dataset provides several annotations for each tweet, we evaluate our classification framework with annotation where at least 3 out of 6 annotators agree. For tweets which are tagged with more than one emotion, we randomly choose one of the emotions as most dominant emotion. Table 2 depicts the emotion distribution of such tweets.

In our experiments we employed the SVM machine classifier for predicting the emotion category of the sentences. For computational treatment of sentences by SVM for training and classification, a sentence is represented by a vector containing values indicating the number of times each feature occurs in the sentence. We employ multi-classifier SVM with sequential minimal optimisation (Platt, 1999) using Weka (Hall et al., 2009). Table 3 shows the resulting confusion matrix and Table 4 depicts the precision, recall and F-measure results from ten-fold cross-validation experiments conducted on SSEC dataset.

Figure 2 F-measure for different feature groups for SSEC dataset (see online version for colours)



It is evident from the results that our classification framework has the best performance when all the features 'WNA+WN+EmoLex+Word2Vec+STNFP' are employed for emotion detection. F-measure achieved is highest for HAPPY emotion when 'All Features' are employed and when 'WNA+WN+EmoLex+Word2Vec' features are employed F-measure is slightly better for ANGER emotion. Figure 2 depicts the F-measure values attained using different feature groups.

Table 2 Emotion distribution in SSEC dataset

Emotion class	*Number of sentences*
Anger	1,388
Disgust	440
Fear	274
Happy	815
Sad	414
Surprise	177

Table 3 Confusion matrix for SSEC dataset

SSEC DATASET	*CONFUSION MATRICES*	*ANGER*	*DISGUST*	*FEAR*	*HAPPY*	*SAD*	*SURPRISE*
All features	ANGER	874	78	63	31	19	12
	DISGUST	86	326	17	15	16	29
	FEAR	24	21	197	13	12	16
	HAPPY	22	24	21	624	31	26
	SAD	15	17	18	16	274	11
	SURPRISE	18	13	16	42	17	85
WNA+WN +EmoLex +Word2Vec	ANGER	862	82	82	36	19	20
	DISGUST	77	320	17	24	16	33
	FEAR	22	24	186	18	15	21
	HAPPY	24	37	25	624	31	31
	SAD	13	15	19	17	286	16
	SURPRISE	17	10	19	43	16	82
WNA+WN +EmoLex +STNFP	ANGER	856	82	74	36	19	15
	DISGUST	78	305	17	19	16	32
	FEAR	26	24	182	18	19	21
	HAPPY	27	28	25	611	31	31
	SAD	13	17	18	16	267	15
	SURPRISE	17	10	19	43	16	79
WNA+WN +Word2Vec +STNFP	ANGER	745	96	74	42	26	27
	DISGUST	71	215	28	17	13	32
	FEAR	30	24	141	18	26	21
	HAPPY	25	28	25	539	43	40
	SAD	19	78	29	16	189	15
	SURPRISE	16	8	16	22	12	60

Table 4 Precision, recall and F-measure for SSEC dataset

SSEC dataset	*Emotion*	*Precision*	*Recall*	*F-measure*
All features	ANGER	84.11	81.15	82.60
	DISGUST	68.05	66.66	67.35
	FEAR	59.33	69.61	64.06
	HAPPY	84.21	83.42	83.81
	SAD	74.25	78.06	76.11
	SURPRISE	47.48	44.50	45.94
WNA+WN +EmoLex +Word2Vec	ANGER	84.92	78.29	81.47
	DISGUST	65.57	65.70	65.64
	FEAR	53.44	65.03	58.67
	HAPPY	81.88	80.82	81.35
	SAD	74.67	78.14	76.36
	SURPRISE	40.39	43.85	42.05
WNA+WN +EmoLex +STNFP	ANGER	84.16	79.11	81.56
	DISGUST	65.45	65.31	65.38
	FEAR	54.32	62.75	58.24
	HAPPY	82.23	81.14	81.68
	SAD	72.55	77.16	74.78
	SURPRISE	40.93	42.93	41.90
WNA+WN +Word2Vec +STNFP	ANGER	82.22	73.76	77.76
	DISGUST	47.88	57.18	52.12
	FEAR	45.04	54.23	49.21
	HAPPY	82.41	77.00	79.61
	SAD	61.16	54.62	57.70
	SURPRISE	30.76	44.77	36.47

b Evaluation on Aman’s dataset

We also conducted experiments on gold standard of Aman’s dataset. As suggested by Aman and Szpakowicz (2007) we employ lesser number of neutrals to avoid skewness towards emotion-free sentences. We employ multi-classifier SVM with SMO using Weka to train and validate our classifier. Table 5 shows the distribution of various emotions in the dataset.

Table 5 Emotion distribution in Aman’s dataset

Emotion class	*Number of sentences*
ANGER	179
DISGUST	172
FEAR	115
HAPPY	536
SAD	173
SURPRISE	115
NEUTRAL	600

Table 6 Confusion matrix for Aman's dataset

AMAN'S DATASET	*CONFUSION MATRICES*	*ANGER*	*DISGUST*	*FEAR*	*HAPPY*	*SAD*	*SURPRISE*	*NEUTRAL*
All features	ANGER	152	12	14	3	10	9	9
	DISGUST	12	154	8	19	6	5	16
	FEAR	1	3	97	9	5	6	4
	HAPPY	3	11	28	398	16	4	36
	SAD	2	14	5	8	146	3	17
	SURPRISE	3	26	3	11	12	101	15
	NEUTRAL	9	39	6	23	11	14	372
WNA+WN +EmoLex +Word2Vec	ANGER	148	13	16	4	13	7	9
	DISGUST	13	153	8	19	5	5	14
	FEAR	2	4	97	9	7	6	3
	HAPPY	4	12	28	392	16	4	40
	SAD	3	14	5	8	128	3	17
	SURPRISE	3	24	3	11	24	98	14
	NEUTRAL	7	40	6	32	11	18	370
WNA+WN +EmoLex +STNFP	ANGER	130	18	22	19	9	15	23
	DISGUST	13	142	8	19	5	5	23
	FEAR	2	4	80	9	7	6	9
	HAPPY	4	12	28	358	16	4	42
	SAD	8	14	16	8	120	3	16
	SURPRISE	9	24	13	14	24	86	14
	NEUTRAL	7	40	16	32	15	18	361
WNA+WN +Word2Vec +STNFP	ANGER	134	16	22	19	9	15	23
	DISGUST	11	150	8	19	5	5	23
	FEAR	2	4	82	9	7	6	9
	HAPPY	3	16	28	288	16	4	42
	SAD	18	19	16	8	112	0	16
	SURPRISE	2	24	13	14	24	86	14
	NEUTRAL	15	32	26	32	15	18	411

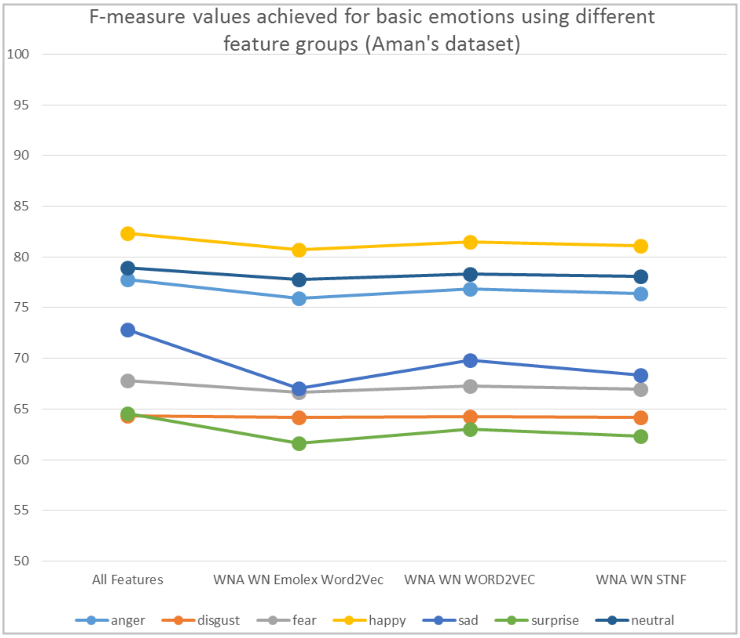
The results on Aman's dataset using all features 'WNA+WN+EmoLex+Word2Vec +STNFP' shows a remarkable improvement when compared with results obtained by Aman and Szpakowicz (2007). For instance the F-measure achieved for HAPPY emotion is 80.34 using this framework and it is 75.10 using Aman's approach. Table 6 depicts the resulting confusion matrix and Table 7 presents the precision, recall and F-measure results from ten-fold cross-validation experiments conducted on Aman's dataset.

Figure 3 shows the achieved F-measure on Aman's dataset for each of the six emotions using employing different feature sets.

Table 7 Precision, recall and F-measure for Aman's dataset

Aman's dataset	*Emotion*	*Precision*	*Recall*	*F-measure*
All features	ANGER	83.516	72.727	77.749
	DISGUST	59.459	70.000	64.301
	FEAR	60.248	77.600	67.832
	HAPPY	84.501	80.242	82.316
	SAD	70.874	74.872	72.818
	SURPRISE	71.127	59.064	64.537
	NEUTRAL	79.318	78.481	78.897
WNA+WN +EmoLex +Word2Vec	ANGER	82.222	70.476	75.897
	DISGUST	58.846	70.507	64.151
	FEAR	59.509	75.781	66.667
	HAPPY	82.526	79.032	80.742
	SAD	62.745	71.910	67.016
	SURPRISE	69.504	55.367	61.635
	NEUTRAL	79.229	76.446	77.813
WNA+WN +EmoLex+STNFP	ANGER	75.145	55.085	63.570
	DISGUST	55.906	66.047	60.554
	FEAR	43.716	68.376	53.333
	HAPPY	77.996	77.155	77.573
	SAD	61.224	64.865	62.992
	SURPRISE	62.774	46.739	53.583
	NEUTRAL	73.975	73.824	73.900
WNA+WN +Word2Vec +STNFP	ANGER	72.432	56.303	63.357
	DISGUST	57.471	67.873	62.241
	FEAR	42.051	68.908	52.229
	HAPPY	74.036	72.544	73.282
	SAD	59.574	59.259	59.416
	SURPRISE	64.179	48.588	55.305
	NEUTRAL	76.394	74.863	75.621

Figure 3 F-measure for different feature groups for Aman's dataset (see online version for colours)



6 Conclusions

This paper proposes a supervised classification framework for emotion classification in textual social data using lexical, semantic and contextual information. In this study, we employ class-specific emoticons together with lexical, semantic and contextual features to detect emotions in text. Semantic features captured by employing Word2Vec remarkably enhance the classification accuracy. For the set of six basic emotions, the classification accuracy of 75.82 is achieved on SSEC dataset using the proposed framework with all features incorporated in supervised learning. On Aman's dataset our framework achieves an accuracy of 75.13 when all features are employed. It is evident from the results that our classification framework has the best performance when all the features (WNA+WN+EmoLex+Word2Vec+STNFP) are employed for emotion detection. F-measure achieved is highest for emotions HAPPY (83.81) and ANGER (82.60) when 'All features' are employed for SSEC dataset. In case of Aman's dataset, F-measure for HAPPY emotion (82.31) the achieved F-measure is highest among all emotions. This clearly indicates that people generally employ more clear and affective words while expressing happiness.

Future research in this direction may include the use of real-world knowledge to further increase the classification accuracy. Performance evaluation of this framework on sarcastic sentences is also a future task to be undertaken.

References

Agrawal, A. and An, A. (2012) 'Unsupervised emotion detection from text using semantic and syntactic relations', in *2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology*, IEEE, pp.346–353, DOI: 10.1109/WI-IAT.2012.170.

Aman, S. and Szpakowicz, S. (2007) 'Identifying expressions of emotion in text', in *Text, Speech and Dialogue*, pp.196–205, DOI: 10.1007/978-3-540-74628-7_27.

Boia, M. et al. (2013) 'A: is worth a thousand words: how people attach sentiment to emoticons and words in tweets', in *2013 International Conference on Social Computing*, IEEE, pp.345–350, DOI: 10.1109/SocialCom.2013.54.

Canales, L. et al. (2016) 'Exploiting a bootstrapping approach for automatic annotation of emotions in texts', *Proceedings – 3rd IEEE International Conference on Data Science and Advanced Analytics (DSAA 2016)*, pp.726–734, DOI: 10.1109/DSAA.2016.78.

Clavel, C. and Callejas, Z. (2016) 'Sentiment analysis: from opinion mining to human-agent interaction', *IEEE Transactions on Affective Computing*, Vol. 7, No. 1, pp.74–93, DOI: 10.1109/TAFFC.2015.2444846.

Ekman, P. (1992) 'An argument for basic emotions', *Cognition and Emotion*, DOI: 10.1080/02699939208411068.

Gamallo, P. et al. (2001) 'Syntactic-based methods for measuring word similarity', in Matoušek V., Mautner, P., Mouček, R. and Taušer K. (Eds.): *Text, Speech and Dialogue*, TSD 2001, Lecture Notes in Computer Science, Vol. 2166, pp.116–125, Springer, Berlin, Heidelberg, DOI: 10.1007/3-540-44805-5_15.

Hall, M. et al. (2009) 'The WEKA data mining software', *ACM SIGKDD Explorations Newsletter*, ACM, Vol. 11, No. 1, p.10, DOI: 10.1145/1656274.1656278.

Khan, F.H., Bashir, S. and Qamar, U. (2014) 'TOM: Twitter opinion mining framework using hybrid classification scheme', *Decision Support Systems*, Elsevier B.V., Vol. 57, No. 1, pp.245–257, DOI: 10.1016/j.dss.2013.09.004.

Ku, L-W. and Sun, C.-W. (2012) 'Calculating emotional score of words for user emotion detection in messenger logs', in *2012 IEEE 13th International Conference on Information Reuse and Integration (IRI)*, IEEE, Las Vegas, NV, USA, pp.138–143, DOI: 10.1109/IRI.2012.6303002.

De Marneffe, M-C. and Manning, C.D. (2012) *Stanford Typed Dependencies Manual* [online] https://nlp.stanford.edu/software/dependencies_manual.pdf (accessed 30 January 2018).

Mikolov, T. et al. (2013) *Efficient Estimation of Word Representations in Vector Space*, arXiv preprint arXiv: 1301.3781 [online] https://arxiv.org/pdf/1301.3781.pdf (accessed 24 January 2018).

Miller, G. (1995) 'WordNet: a lexical database for English', *Communications of the ACM*, ACM, Vol. 38, No. 11, pp.39–41, DOI: 10.1145/219717.219748.

Mohammad, S.M., Sobhani, P. and Kiritchenko, S. (2017) 'Stance and sentiment in Tweets', *ACM Transactions on Internet Technology*, ACM, Vol. 17, No. 3, pp.1–23, DOI: 10.1145/3003433.

Mohammad, S.M. and Turney, P.D. (2013) 'Crowdsourcing a word-emotion association lexicon', in *Computational Intelligence*, pp.436–465, DOI: 10.1111/j.1467-8640.2012.00460.x.

Na, J.C. et al. (2012) 'Sentiment classification of drug reviews using a rule-based linguistic approach', in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, pp.189–198, DOI: 10.1007/978-3-642-34752-8_25, Springer, Berlin, Heidelberg.

Platt, J.C. (1999) 'Using analytic QP and sparseness to speed training of support vector machines', *Proc. Advances in Neural Information Processing Systems*, Vol. 11, No. 1, pp.557–563 [online] http://papers.nips.cc/paper/1577-using-analytic-qp-and-sparseness-to-speed-training-of-support-vector-machines.pdf (accessed 17 February 2018).

Schuff, H. et al. (2017) 'Annotation, modelling and analysis of fine-grained emotions on a stance and sentiment detection corpus', in *8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, Copenhagen, Denmark [online] http://www.romanklinger.de/publications/Schuff-etal-2017.pdf (accessed 10 February 2018).

Strapparava, C. and Valitutti, A. (2004) 'WordNet affect: an affective extension of WordNet', in *4th International Conference on Language Resources and Evaluation*, European Language Resources Association (ELRA), Lisbon [online] http://aclasb.dfki.de/nlp/bib/L04-1208 (accessed 28 January 2018).

Su, Q. et al. (2006) 'Using pointwise mutual information to identify implicit features in customer reviews', in *Lecture Notes in Computer Science*, Vol. 4285, pp.22–30, DOI: 10.1007/11940098_3, Springer, Berlin, Heidelberg.

Turney, P.D. (2001) 'Thumbs up or thumbs down?', in *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics – ACL '02*, Association for Computational Linguistics, Morristown, NJ, USA, p.417, DOI: 10.3115/1073083.1073153.

Yang, C., Lin, K.H.-Y. and Chen, H-H. (2007) 'Emotion classification using web blog corpora', in *IEEE/WIC/ACM International Conference on Web Intelligence (WI' 07)*, IEEE, pp.275–278, DOI: 10.1109/WI.2007.51.

Notes

1 http://web.eecs.umich.edu/~mihalcea/affectivetext/.

2 https://dumps.wikimedia.org/.

3 http://www.natcorp.ox.ac.uk/.