CLASSIFICATION OF EMOTIONS FROM TEXT USING SVM BASED OPINION MINING

Sandip S. Patil  
Assistant Professor  
SSBT’s COET, Bambhori, Jalgaon, 425001, (India)  
Email: san_78004@yahoo.co.in

Asha P. Chaudhari  
Research Scholar  
S.S. B. T. COET, Bambhori, Jalgaon 425001, (India)  
E-mail: ashachaudhari@gmail.com

Abstract  
The news sentences often yields multiple emotions. In this paper, we are presenting a novel method for classifying news sentences into six categories of emotions. The corpus consists of 1000 news sentences and the emotion tag considered was anger, disgust, fear, happiness, sadness and surprise. The dictionary of WordnetAffect is used for word presence and classification of news sentences. The dataset of Affective Text is used for training and testing. In this paper we are proposing the SVM classification using Quadratic programming for automatic emotion classification and analysis.

1. Introduction  
Social computing is a new age research area inspired by the human-human interactions in society. Human behavior, reflected through their expressions, gesture or spoken and written language, is profoundly influenced by their interactions with the society from birth to death. Consequently, the behavior of a person largely depends on the behavior of the other persons in the society. Sometimes the decision making process considers cues from the social contexts. For example, one tends to inclined to a particular product provided that other social actors are affined to the same. The new age digital systems are trying to exploit this role of social context in a number of applications like recommender system, intelligent tutoring system and many others.

The above discussion imposes that there is a need to develop an effective method for classifying the text according to emotions. The main objective of this paper is stated as follows.
“Provided a set of predefined emotion labels (e.g. joy, fear, surprise), classify the titles with the appropriate emotion label and with a valence indication using SVM classification.”

The stated objective is further divided into following sub problems:
• Prepare the dataset of news headlines
Annotate the dataset with predefined emotions as defined in Affective Text [2]
Using NLP prepare the database matrix of test emotions and training emotions
classify the training set with a support vector machine using quadratic
programming algorithm (active constraints method)
computes the prediction of a support vector machine using the kernel function
and its parameter for classification
Compute the accuracy of the classification.

The rest of this paper is organized as follows. In Section 2, we introduce some related
works. Section 3 describes the corpus and the emotion dictionary used in our
experiment. In section 4, we explain the process of building our emotion classification
system, and present the results in section 5, finally we provide the conclusion and
future works in section 6.

2. Related Work

Research on the sentiment classification is challenging, and more research have
been done on this topic. Plaban et al [1] introduce a multi-label emotion classification
model based on ADTboost MH (a boosting algorithm called ADTboost MH). Word
present in the sentences and the polarity of the subject, object and verb are used as
features. The classifier performs better than that on singular words only. The average
precision was 79.5%. Hu et al [2], implemented a Bayes text classification. Their results
show that the Naive Bayes classification method achieves a high performance on text
classification. However, this method just categorizes texts into only two classes:
positive and negative, which excludes reader’s emotions such as angry, happy etc.
Zornitsa et al presented an approach of headline emotion classification based on
frequency and co-occurrence information collected from the World Wide Web [3].
Kevin Hsin-Yih Lin et al classified news into emotions using various combinations of
feature sets and identifying the emotional influences of news articles on readers [4].
XU Lin hong et al computed semantic similarity of the vocabulary and tagged
vocabulary in HowNet (http://www.keenage.com/), adopted the derogatory or
commendatory terms as features of classification, utilized Support Vector Machine
classifier to identify the text orientation, and dealt with the negative sentence via
matching negative rules [5]. Chinese character bi-gram, words, metadata, affix
similarity, word emotion and emotion categories are used as features. They got the
highest prediction accuracy 76.88% for bored and 89.66% for useful. Class useful is
some kind of vague on emotion expression [6].

Yu Zhang et al explored how to incorporate emotional aspects of dialog into
existing dialog processing techniques and worked on making a Chinese emotion
classification model which is used to recognize the main affective attribute from a
sentence or a text [7]. Prem Melville et al developed a unified framework in which one
can use background lexical information in terms of word-class associations, and refine
this information for specific domains using any available training examples [8].
Plaban Kumar Bhowmick et al present a novel method for classifying news sentences into multiple emotion categories using an ensemble based multi-label classification technique called RAKEL [9].

In short, some sentiment classifications have better performance over others and some experiments focus on the headline data or news sentences. We use effective feature selection method and implement a sentiment classifier with better performance.

### 3. Support Vector Machine

Support Vector Machines (SVM) is a machine learning model proposed by V. N. Vapnik [10]. The basic idea of SVM is to find an optimal hyperplane to separate two classes with the largest margin from pre-classified data. After this hyperplane is determined, it is used for classifying data into two classes based on which side they are located. By applying appropriate transformations to the data space before computing the separating hyperplane, SVM can be extended to cases where the margin between two classes is non-linear.

#### 3.1 Linearly Separable Case

If the training data are linearly separable, then there exists a pair \( (w, b) \) such that

\[
W^T X_i + b \geq 1, \quad \text{for all } X_i \in P
\]

\[
W^T X_i + b \leq -1, \quad \text{for all } X_i \in N
\]

The decision function is of the form

\[
f(x) = \text{sign}(w^T x + b)
\]

Figure 1. Optimal separating hyperplane for Binary classification problem.

\( W \) is termed the weight vector and \( b \) the bias (or \( b^- \) is termed the threshold). The inequality constraints (1) can be combined to give

\[
\gamma_i(W^T X_i + b) \geq 1, \quad \text{for all } X_i \in P \cup N
\]
3.2 Nonlinearly Separable Case

Given a training set of instance-label pairs \((x_i, y_i)\), \(i = 1, 2, 3, \ldots \ell\) where \(x_i \in \mathbb{R}^n\), the class label of the \(i^{\text{th}}\) pattern is denoted by \(y_i \in \{1, -1\}\). Nonlinearly separable problems are often solved by mapping the input data samples \(x_i\) to a higher dimensional feature space \(\phi(x_i)\). The classical maximum margin SVM classifier aims to find a hyperplane of the form \(w^T \phi(x) + b = 0\), which separates patterns of the two classes. So far we have restricted ourselves to the case where the two classes are noise-free. In the case of noisy data, forcing zero training error will lead to poor generalization. To take account of the fact that some data points misclassified, we introduce a vector of slack variables \(\Xi = (\xi_1, \ldots, \xi_\ell)^T\) that measure the amount of violation of the constraints (3). The problem can then be written as

\[
\begin{align*}
\text{Minimize} & \quad \frac{1}{2} w^T w + C \sum_{i=1}^\ell \xi_i \\
\text{subject to} & \quad y_i (w^T \phi(x_i) + b) \geq 1 - \xi_i, \\
& \quad \xi_i = 0, \ i = 1, 2, 3, \ldots \ell
\end{align*}
\]

The solution to (4)-(5) yields the soft margin classifier, is termed because the distance or margin between the separating hyperplane \(w^T \phi(x) + b = 0\) is usually determined by considering the dual problem, which is given by

\[
L(w, b, \alpha, \Xi, \Gamma) = \frac{1}{2} \|w\|^2 + \sum_{i=1}^\ell \alpha_i [y_i (w^T \phi(x_i) + b) - 1 + \xi_i] - \sum_{i=1}^\ell y_i \xi_i + C \sum_{i=1}^\ell \xi_i
\]

where \(\Lambda = (\alpha_1, \ldots, \alpha_\ell)^T\) and \(\Gamma = (\gamma_1, \ldots, \gamma_\ell)^T\), are the Lagrange multipliers corresponding to the positivity of the slack variables. The solution of this problem is the saddle point of the Lagrangian given by minimizing \(L\) with respect to \(w, \Xi\) and \(b\), and maximizing with respect to \(\Lambda \geq 0\) and \(\Gamma \geq 0\). Differentiating with respect to \(w, b\) and \(\Xi\) and setting the results equal to zero,

It is obtained

\[
\begin{align*}
\frac{\partial L(w, b, \alpha, \Xi, \Gamma)}{\partial w} &= -\sum_{i=1}^\ell \alpha_i y_i \phi(x_i) = 0, \\
\frac{\partial L(w, b, \alpha, \Xi, \Gamma)}{\partial b} &= -\sum_{i=1}^\ell \alpha_i y_i = 0, \\
\text{and} \\
\frac{\partial L(w, b, \alpha, \Xi, \Gamma)}{\partial \xi_i} &= c - \alpha_i - \gamma_i = 0;
\end{align*}
\]

\[
\text{Minimize} \quad \sum_{i=1}^\ell \sum_{j=1}^\ell y_i y_j a_i a_j k(x_i, x_j) - \sum_{i=1}^\ell \alpha_i
\]

Subject to
\[ \ell \sum_{i=1}^{\ell} \alpha_i y_i = 0, \text{ and } 0 \leq \alpha_i \leq C, \ i=1,2,3 \ldots \ell \]  

(7)

Here, \( \alpha_i, i=1,2,3 \ldots \ell \) denotes the Lagrange multipliers and the matrix \( K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j) \) are termed as Kernel matrix [11][12]. Training vector \( x_i \) is mapped into a higher dimensional feature space and construct an optimal hyperplane. SVM also restricts the choice of Kernel that the Quadratic programming is a convex problem. Therefore, it guarantees that global optimization with corresponding Kernel. SVM uses training data as Support Vectors and uses Lagrange multipliers to represent the Support Vectors. The classifier can be constructed using the decision function in the form

\[ y(x) = \text{sgn}[\sum C_i y_k (x, x_k) + b)] \]

4. A Framework of Emotion Classification using SVM

We proposed to focus on the emotion classification of text document. The text documents typically consist of a few words and are often written by creative people with the intention to provoke emotions, and consequently to attract the readers' attention i.e. news or news headlines, reviews of movie or product or any type of services etc. These type of data may be collected from web or it may be written by any customers or writer. These characteristics make these text document particularly suitable for use in an automatic emotion recognition setting, as the affective/emotional features (if present) are guaranteed to appear in these short sentences.

Figure 2 shows the general framework for emotion classification of news sentences using SVM. The overall process is divided into two steps:

Step 1: In this step we train the SVM classifier with train data and compute the parameters which are used for classification of test data. Figure 1 shows the process of training the SVM for classification.

Step 2: In this step we validate the SVM by the test data using the parameters computed from train data. The process is shown in figure 2.

Working
1. Preparation of the database matrix

   a. Compute the matrix TrainDatabaseIn from the 250 training news sentences. The matrix is of size 250 X7. The first column deals with the occurrence of the news sentence. The subsequent columns deals with the 6 emotions in the dictionary. The entry will correspond to ‘1’ if the news belong to particular emotion else ‘0’.

   b. Compute the matrix TrainDatabaseOut from the annotations given in the database related to each of 250 news headlines.

   c. Compute the matrix TestDatabaseIn from the 1000 test news sentences. The matrix is of size 1000 X7. The first column deals with the occurrence of the news sentence. The subsequent columns deals with the 6 emotions in the dictionary. The entry will correspond to ‘1’ if the news belong to particular emotion else ‘0’.

   d. Compute the matrix TestDatabaseOut from the annotations given in the database related to each of 1000 news headlines.

   e. Prepare the Database matrix of TrainDatabaseIn, TrainDatabaseOut, TestDatabaseIn, TestDatabaseOut

2. Classify the training set with a support vector machine using quadratic programming algorithm (active constraints method). Compute the parameters of classification from the training set.

3. Validate the SVM for the test set using the parameters computed from the training data.

4. Compute the precision as the fraction of number of true positives to the sum of false positive and true positives. Compute the recall as the fraction of number of true positives to the sum of true positives and false negatives. Compute F1 measure as accuracy of classification for each emotion class.

5. Training and Testing with the specified classifier using cross validation technique.

5. Experimental Results and Analysis

5.1 Training and Testing Data

   As mentioned before, emotions are a matter of subjective feelings and as such they can be hardly described and objectively measured. And this is a problem. When we are to deal with training an emotion classification learning model, an absolute necessity is to have quality training data. A classification model can have only such good accuracy as the training data had.

   Two data sets were used: a development data set consisting of 250 annotated headlines, and a test data set with 1,000 annotated headlines.
5.2 Experimental Results

The research intends to compare the efficiency of SVM. Detection and identification of emotional and non-emotional words generalized as the following: True positive (TP): the number of emotional words detected when it is actually emotional words. True negative (TN): the number of non-emotional words detected when it is actually non-emotional words. Classifiers have long been evaluated on their accuracy only. An often-used measure in the information retrieval and natural language processing communities is Overall Accuracy. An often-used another measure in the information retrieval and natural language processing communities is the F1-measure. According to Yang and Liu [15], this measure was first introduced by C. J. Van Rijsbergen [16]. They state, the F1 measure combines recall (R) and precision (P) with an equal weight in the following form:

\[ \text{F1} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \]

where \( R = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100 \% \) and \( P = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100 \% \). TP is the number of true positives, i.e., the number of non-emotional words predicted correctly. TN is the number of true negatives, i.e., the number of words correctly predicted as non-emotional. FP is the number of false positives, i.e., the number of words incorrectly predicted as non-emotional words. FN is the number of false negatives, i.e., the number of words incorrectly predicted as emotional.

Table 1. Experimental results (in terms of F1-Measure) for emotions trained from WordNet Affect and tested on SemEval Test set

<table>
<thead>
<tr>
<th>Training Set</th>
<th>Test Set</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SemEval Affective Text Dataset</td>
</tr>
<tr>
<td>WordNet Affect</td>
<td></td>
</tr>
<tr>
<td>Anger</td>
<td></td>
</tr>
<tr>
<td>Disgust</td>
<td></td>
</tr>
<tr>
<td>Fear</td>
<td></td>
</tr>
<tr>
<td>Joy</td>
<td></td>
</tr>
<tr>
<td>Sadness</td>
<td></td>
</tr>
<tr>
<td>Surprise</td>
<td></td>
</tr>
</tbody>
</table>

In this experiment, we are considering the effect of adding emotional words from WPARD and WorNet-Affect and ISEAR into our training set. Result shows that classification performance increased for SVM. SVM gives drastic and improved difference compared to other classifiers.
6. Conclusion

Emotion classification of text is very important in applications like emotional text-to-speech (TTS) synthesis, human computer interaction, etc. Past studies on emotion classification focus on the writer’s emotional state conveyed through the text. This paper addresses the reader’s emotions provoked by the text. In this paper we explore sentence level emotion classification. Firstly, we extract news headlines and related reader emotion information from the web. Then we classify news headlines into reader emotion categories using support vector machine (SVM), and examine classification performance under different feature settings. Experiments show that certain feature combinations achieve good results.

References