International Journal of Engineering Research and Technology.
ISSN 0974-3154 Volume 7, Number 2 (2014), pp. 153-159
© International Research Publication House
http://www.irphouse.com

Text Emotion Detection using Neural Network

Rupinder Singh, Veenu Mangat and Mandeep Kaur

University Institute of Engineering and Technology
Panjab University Chandigarh
rupinder0590@gmail.com, veenumangat@yahoo.com, mandeep@pu.ac.in

Abstract

Text emotion detection refers to identifying the type of emotion getting used by the text. The process involves two process training and testing. The training section involves training the classifier with the text and the testing section involves the identification of the type of text used. After this accuracy of the classifier is checked by measuring how many correct labels of the text does the classifier identifies. This paper focuses on the enhancement of the text emotion detection using back propagation neural network. The classification results have improved by 5 to 10 percent

KEYWORDS: Text emotion, Feature Extraction, Classification, Neural Networks

INTRODUCTION

Detecting emotional state of a person by analyzing a text document written by him/her appear challenging but also essential many times due to the fact that most of the times textual expressions are not only direct using emotion words but also result from the interpretation of the meaning of concepts and interaction of concepts which are described in the text document. Recognizing the emotion of the text plays a key role in the human-computer interaction. Emotions may be expressed by a person's speech, face expression and written text known as speech, facial and text based emotion respectively. Sufficient amount of work has been done regarding to speech and facial emotion recognition but text based emotion recognition system still needs attraction of researchers[1]. In computational linguistics, the detection of human emotions in text is becoming increasingly important from an applicative point of view.

Emotion is expressed as joy, sadness, anger, surprise, hate, fear and so on. Since there is not any standard emotion word hierarchy, focus is on the related research about emotion in cognitive psychology domain. "Emotions In Social Psychology", in which he explained the emotion system and formally classified the human emotions through an emotion hierarchy in six classes at primary level which are Love, Joy, Anger, Sadness, Fear and Surprise. Certain other words also fall in secondary and tertiary levels. [2]

Keyword Spotting Technique

The keyword pattern matching problem can be described as the problem of finding occurrences of keywords from a given set as substrings in a given string [4]. This problem has been studied in the past and algorithms have been suggested for solving it. In the context of emotion detection this method is based on certain predefined keywords. These words are classified into categories such as disgusted, sad, happy, angry, fearful, surprised etc. Process of Keyword spotting method is shown in the figure 1.

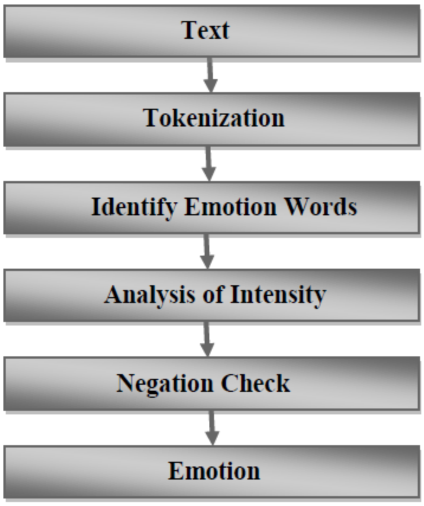


Figure 1. Keyword Spotting Technique

Keyword spotting technique for emotion recognition consists of five steps shown in fig.1 where a text document is taken as input and output is generated as an emotion class. At the very first step text data is converted into tokens, from these tokens emotion words are identified and detected. Initially this technique takes some text as input and in next step we perform tokenization to the input text. Words related to emotions will be identified in the next step afterwards analysis of the intensity of emotion words will be performed. Sentence is checked whether negation is involved in it or not then finally an emotion class will be found as the required output.

Design of Neural Networks for Emotion Recognition

General description

Now with all the background knowledge, we can start the design of neural networks for emotion recognition. The 12 features data obtained from the face tracker are used as the input of the 12 input nodes in a neural network. The output layer contains 2-7 nodes that represent the emotion categories, depending on different networks. There are 1 or 2 hidden layers and the number of hidden nodes ranges from 1 to 29x29. The learning rate, momentum number, and the parameter of the sigmoid activation function are automatically adjusted during the training procedure. In some networks, the Powell's method is considered, while in others, a set of empirical ways are combined, i.e. take the peak frames of the emotion data sequence, sort the training set, delete some of the emotions, normalize the output, set threshold to the weights, etc. The test results are based on Cohn-Kanade database and on the authentic database separately. The activation function we used is the sigmoid function. In the following sections, we present a description of all parameters, and their combinations' results in experiments.[3]

Weights

Back-propagation is a gradient descent search, so it is easy to stop at a local minimum, while randomly selected weights help to avoid this. If the weights are too large, the networks tend to get saturated. The solution is to ensure that, after weight initialization and before learning, the output of all the neurons is small value between [-0.5, 0.5]. We initialize the weights by a random function and ignore those weights that are larger than a specific threshold which can also be adjusted as one of the parameter of the network. One question is if during the training procedure, should we constrain the weights as well? This partially depends on how large the input is, because the sigmoid function is very close to one when the input is greater than 10. Since our feature data for input is very small, usually smaller than 2, we set the threshold of the weights, ignoring those weights that exceed this constraint during training. It came out that this did not make much difference at improving the hit rate. On the other hand, when we tried a very strict threshold in a 2-hidden-layer neural network during the training procedure, sometimes it led awful performance of the 2-layer network. This is because the parameters of the activation we set in a 2-hidden-layer network were not proper for that threshold, and caused the saturation of the neuron. Hence, we gave a very large threshold after initialization to the weights to avoid similar problems.

Another thing we should take care of is the starting point, which can also affect the search direction to find a good local minimum(figure 2)If we start at point A, we obtain the global minimum, while from C, we get the local minimum. So we should try different staring points by initializing the weights with different random values. We tested this in some of the networks and found that the accuracy curve fluctuates, but not too much.

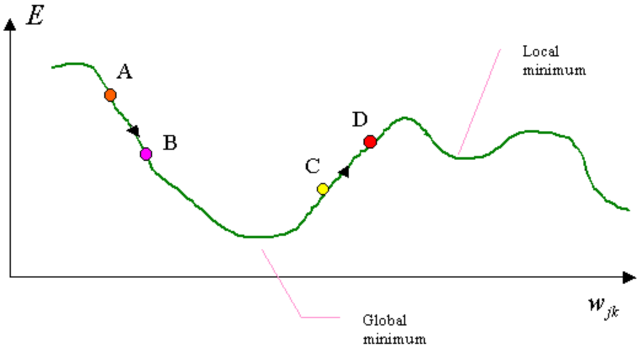


Figure: Local minimum and global minimum

Key parameters and their combination

In design of neural networks, there are some critical parameters that need to be set, i.e. the learning rate α, the momentum number λ, and the activation function parameter σ.

The speed of the learning is governed by the learning rate α. The momentum number carries along the weight change, thus it tends to smooth the error-weight space. An improper value of σ will cause the neuron saturated. In general, the performance of neural networks will be very awful, if these values are not chosen correctly. Unfortunately, there are no precise rules or mathematics definition upon when one should choose what particular value of these parameters. Normally, the setting of the parameters is done empirically. Does it help in finding better combinations if we let the computer do part of the job? We tried this in the following way.

First, we defined three different categories. The increase or decrease step size of α, λ, and σ is given by input or macro definition. This depends on how frequently these categories should be changed during training procedure, e.g., for those categories that need little interference, we give a macro definition for training efficiency. When better accuracy occurs, the rates together with the parameters which lead to this accuracy are recorded in a file. We repeat the training until the accuracy stops improving for some turns. When testing, we construct a neural network by reading these parameters from the file. Since the training with the complete combinations of these parameters costs quite a long time, we only tried a part of these three parameters’ combinations. Therefore, it is possible to miss some better combinations of them.

Feature Extraction

To achieve a successful classification, it is extremely important to extract the relevant features from the processed speech data. The most important features for emotions classification are summarized as follows as

Pitch

Pitch is the most distinctive difference between male and female. A person’s pitch originates in the vocal cords/folds, and the rate at which the vocal folds vibrate is the frequency of the pitch. Various Pitch Detection Algorithms (PDAs) have been developed: Autocorrelation method, Harmonic Product Spectrum (HPS), Robust Algorithm for Difference Function (AMDF) method, Cepstrum Pitch Determination (CPD), Simplified Inverse Filtering Tracking (SIFT). and Direct Time Domain Fundamental Frequency Estimation (DFE). Most of them have a very high accuracy for voiced pitch estimation, but the error rate considering voicing decision is still quite high. Moreover, the PDAs performance degrades significantly as the signal conditions deteriorate. The automatic glottal inverse filtering method and iterative adaptive inverse filtering (IAIF) was used as a computational tool for getting an accurate estimation for pitch

Formants

The formants are one of the quantitative characteristics of the vocal tract. In the frequency domain, the location of vocal tract resonances depends upon the shape and the physical dimensions of the vocal tract. Each formant is characterized by its center frequency and its bandwidth as in [4],. The formants can be used to discriminate the improved articulated speech from the slackened one. The formant bandwidth during slackened articulated speech is gradual, whereas the formant bandwidth during improved articulated speech is narrow with steep flanks. A simple method to estimate formant frequencies and formant bandwidths relies on linear prediction analysis.

Energy

Energy is one of the most important features that give good information about the emotion. The long term definition of signal energy is defined as in (1):

$$Energy=\sum(x_{normalised})^2$$

There is little or no utility of this definition for time-varying signals, speech. So the short term energy contour is evaluated because it is related to the arousal level of emotions as in (2):

$$Energy_n=\sum_{m=n-N+1}^{n}[x(m)w(n-m)]^2$$

where *w(n-m)* is the window, n is the sample that the analysis window is centered on, and N is the window size.

ALGORITHM

1. Upload Files for all categories (SAD HAPPY ANGRY)
2. Store values in in db
3. Target(1:4- for every category)
4. Y=Net.train(Uploaded_Vaues_files, targets,epochs)
5. Upload a value for testing
6. Test Sample=Input Sample
7. G=Newff(Y,Uploaded Set,10) where 10 is the number of neurons
8. If G<=1
9. CATEGORY-SAD

```
10. Else if 1<G<2
11. CATEGORY HAPPY
12. Else if2< G <3
13. CATEGORY ANGRY
```

RESULT OF CLASSIFICATION

The results are categorized as below

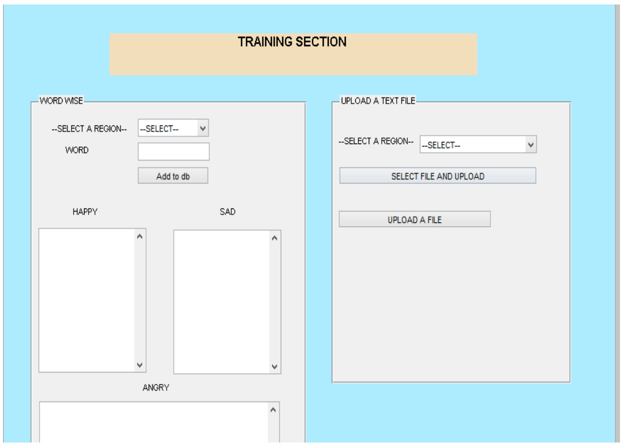
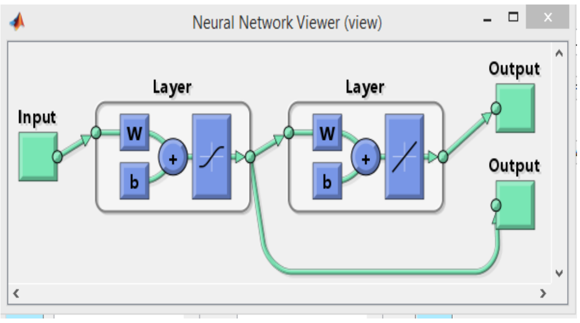


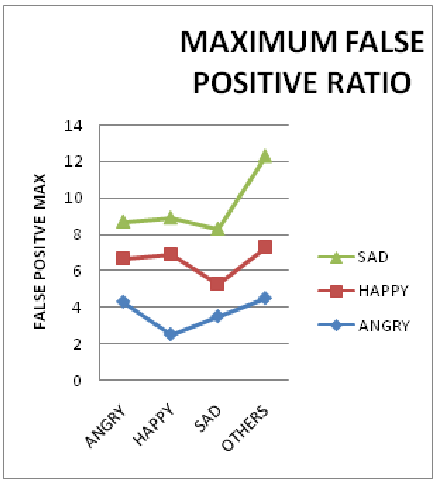
Figure represents the main working window where training and testing can be performed. The left side contains the training part where as the right side contains the testing part.

The training section involves the value of the uploading of the files and the testing section involves the classification.



The above figure[] represents the configuration of the classification using neural network in which back propagation neural network has been called. The input contains two layers. The first layer is the input set of data which the user has uploaded to be tested and the second input is the stored database values. The output would be computed according to the estimation values as mentioned in the algorithm. The results can be classified according to the following matrix

FILE CONTENT LENGTH	CATEGORY SAD	CATEGORY HAPPY	CATEGORY ANGRY
100 words	85 %	92 %	93 %
200 words	89 %	94%	95 %
250 words	89.2 %	93.7%	96.3%



The above figure represents the false positive ratio of the contents when plotted into the real time scenario in which Sad has the maximum occurance followed by happy and then angry.

The current research work opens a lot of doors for the future research workers. The current system does not signify any mixed emotion data and also the classifiers can be upgraded to BFO.

REFERENCES

[1] Nicu Sebe, Michael S. Lew, Ira Cohen, Ashutosh Garg, Thomas S. Huang *Emotion Recognition Using a Cauchy Naive Bayes Classifier* ICPR, 2002

[2] G. Little Wort, I. Fasel, M. Stewart Bartlett, J. Movellan *Fully automatic codeing of basic expressions from video*, University of California, San Diego

[3] C. Maaoui, A. Pruski, and F. Abdat, “Emotion recognition for human machine communication”, Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 08), IEEE Computer Society, Sep. 2008, pp. 1210-1215, doi: 10.1109/IROS.2008.4650870

[4] Chun-Chieh Liu, Ting-Hao Yang, Chang-Tai Hsieh, Von-Wun Soo, ”Towards Text-based Emotion Detection: A Survey and Possible Improvements ”,in International Conference on Information Management and Engineering,2009.

[5] N. Fragopanagos, J.G. Taylor, “Emotion recognition in human–computer interaction”, Department of Mathematics, King‟s College, Strand, London WC2 R2LS, UK Neural Networks 18 (2005) 389–405 march 2005.