International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:7, No:3, 2013

Emotion Recognition Using Neural Network: A Comparative Study

Nermine Ahmed Hendy and Hania Farag

***Abstract*—**Emotion recognition is an important research field that finds lots of applications nowadays. This work emphasizes on recognizing different emotions from speech signal. The extracted features are related to statistics of pitch, formants, and energy contours, as well as spectral, perceptual and temporal features, jitter, and shimmer. The Artificial Neural Networks (ANN) was chosen as the classifier. Working on finding a robust and fast ANN classifier suitable for different real life application is our concern. Several experiments were carried out on different ANN to investigate the different factors that impact the classification success rate. Using a database containing 7 different emotions, it will be shown that with a proper and careful adjustment of features format, training data sorting, number of features selected and even the ANN type and architecture used, a success rate of 85% or even more can be achieved without increasing the system complicity and the computation time.

***Keywords*—**Classification, emotion recognition, features extraction, feature selection, neural network

I. Introduction

IN past years, many researchers have paid attention to the recognition of nonverbal information, and have especially focused on emotion recognition. Many kinds of physiological characteristics are used to extract emotions, such as voice, facial expressions, hand gestures, body movements, heartbeat and blood pressure. Among these modalities, facial expressions and speech are known to be more effective for expressing the emotions. From the literature, it is observed that human perception of the emotions is about 55% from facial expressions, 38% from the speech and 7% from the text as in [1].

Speech Emotion Recognition (SER) can find several applications such as call centers management, commercial products, life-support systems, virtual guides, customer service, lie detectors, conference room research, emotional speech synthesis, art, entertainment and others.

Two different approaches dominate the research on SER: feature-based or classifier-based as in [2]. Feature-based approach aims to extract emotion-related features from the human speech. Some researchers worked on extracting as many speech features as possible for emotion recognition as in [3], [4]. Others emphasized on getting a limited but efficient set of features to improve the classification as in [5], [6]. In addition, working on finding new speech emotional database for certain unfamiliar languages was considered in [7]. Feature selection and standardization was investigated in [2]-[4], [8].

On the other hand, the classification-based approaches focus on designing a classifier to determine distinctive boundaries between emotions. Many researches concentrate on classification using one type of Artificial Neural Network (ANN) as in [9]-[11]. Comparison between Multi-Layer Perceptron ANN and Generalized Feed Forward ANN was considered in [12]. Razak et al. compared the performance of ANN against Fuzzy classifiers as in [13]. Alternative classifiers were also considered such as Hidden Markov Model (HMM) as in [14], k-Nearest Neighbor (k-NN) as in [15], Gaussian Mixture Model (GMM) as in [16], Support Vector Machines (SVM) as in [17], and Decision Tree Algorithms (DTA) as in [18]. Various studies showed that choosing the appropriate classifier can significantly enhance the overall performance of the system as in [19]. References [20]-[22] show that classifier performance was improved by dividing speech data in two groups according to gender.

This work compares the performance of 5 different topologies of ANN while considering the effect of other important factors such as the data format and the number of features used. The effect of training data sorting on the classifier performance was also considered for the first time. Several tests were carried out to elaborate the idea using a free online database as in [23] containing 7 different emotions. The classification was based on the most common features used in earlier works. The results will show what system design aspects are significant and how system can be tuned to achieve the maximum classification success rate.

The paper is organized as follows. In section II the SER system is briefly described. In Section III, the database and test preparation are discussed. Section IV is dedicated for the experimental results. Finally, conclusions are drawn in Section V.

II. Speech Emotion Recognition System

Like any typical recognition system, the speech emotion recognition system has the emotional speech as an input and the classified emotion as an output. The system contains four main stages, preprocessing, feature extraction, feature preparation and finally the classifier.

Fig. 1 shows a flowchart for a typical emotional recognition system. The input of the system is the speech files, which first undergoes some preprocessing. The next step is to extract the main features of the input speech that will differentiate between the different emotions. Then the feature selection,

Nermine Hendy is with the Alexandria Institute of Engineering and Technology, Alexandria, Egypt (e-mail: nermine_ah@yahoo.com).
Hania Farag., is with Alexandria University, Faculty of Engineering, Alexandria, Egypt (e-mail: hania11@yahoo.com).

removal and standardization algorithms are applied to get the optimum feature vectors. The vector is then presented to the classifier in training and testing scheme. The final output is the classified emotion according to the input speech.

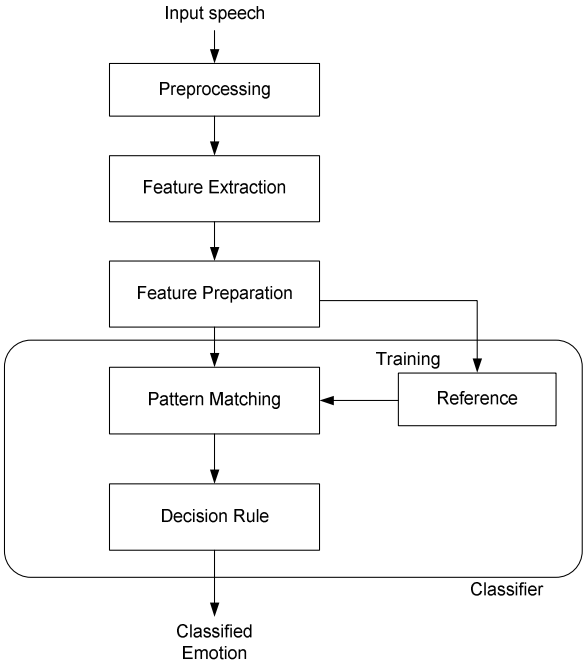

Fig. 1 Flowchart for the Emotional Recognition System

A. Preprocessing

Signal pre-processing is an important stage of the system as it has a big impact on the performance of the classifier. It is important to feed the neural network with suitable inputs for robust classification. It covers Analogue to Digital (A/D) conversion, digital filtering, normalization, signal segmentation, windowing and removal of non speech signal. The digitized speech samples are first normalized, and then the d.c. component is removed. Next, the samples are segmented into 30msec frames with a 10msec overlap using a hamming window. Finally, silence and unvoiced frames are detected and removed, as in [19].

B. 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 in [3], [4], [19], [24]:

1. 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 as in [25], Harmonic Product Spectrum (HPS) as in [26], Robust Algorithm for Pitch Tracking (RAPT) as in [27], Average Magnitude Difference Function (AMDF) method as in [28], Cepstrum Pitch Determination (CPD) as in [29], Simplified Inverse Filtering Tracking (SIFT) as in [30] and Direct Time Domain Fundamental Frequency Estimation (DFE) as in [31]. 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, which is used in [32], [33]. The pitch related features extracted and used are further described in section III.

2. 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], [19]. 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.

3. 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):

$$Energy = \sum(x_{normalized})^2 \quad (1)$$

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 [19] as in (2):

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

where *w(n-m)* is the window, n is the sample that the analysis window is centered on, and N is the window size. The energy related features extracted and used are further described in section III.

4. Spectral Energy

The fourth feature is the energy of certain frequency bands as in [4]. There are many contradictions in identifying the best frequency band of the power spectrum in order to classify emotions. Many investigators put high significance on the low frequency bands, such as the 0–1.5 kHz band whereas others suggest the opposite as in [34].

5. Duration and Voicing

A set of timing features, which display prosodic characteristics of the utterance was extracted. The duration for

voiced and un-voiced segment is extracted as in [4], [19], [35]. The number of voiced segments and the speech rate is calculated as the inverse duration of the voiced part of speech determined by the presence of pitch pulses or it can be found by the rate of syllabic units.

6. Zero Crossing Rates

The zero crossing rate contours, the number of time-domain zero-crossings in each frame of the signal, is calculated for each signal. Similarly, the short-time zero crossing rate is defined as the weighted average of the number of times the speech signal changes sign within the time window. The zero-crossing rate was calculated for each 20 ms frame of a sample's data as in [4]. Representing this operation in (3) and (4) as:

$$z_{\hat{n}} = \sum_{m=-\infty}^{\infty} 0.5|sgn\{x[m]\} - sgn\{x[m-1]\}|w[\hat{n}-m] \quad (3)$$

where

$$sgn\{x\} = \begin{cases} 1 & x \geq 0 \\ -1 & x < 0 \end{cases} \quad (4)$$

7. Jitter and Shimmer

In speech and voice science jitter and shimmer are used to describe the short-term – cycle-to-cycle – variations in fundamental frequency and amplitude, respectively. The jitter and shimmer related features, perturbation factor (PF) and perturbation quotient (PQ), are measured for the voiced segments of a signal.

8. Other Spectral Features

The Mel-frequency Cepstral Coefficients (MFCCs) provide a better representation of the signal than the frequency bands since they additionally exploit the human auditory frequency response as in [35].

Linear Predictive Coding (LPC) is a popular signal transforms methods that were used for feature extraction in speech recognition works. Basically, the LPC algorithm produces a vector of coefficients that represent a smooth spectral envelope of the Direct Fourier Transform (DFT) magnitude of a temporal input signal.

C. Feature Preparation

The extracted features vectors need some preparation before entering the last step of classification. Preparation consists of features removal, standardization and selection. Certain features have to be removed, because of many missing values observed as in [4], [36] When more than 2% of the total number of feature values is missing, the corresponding feature is discarded.

In some papers features were normalized before selection stage [4], [14], [37]. While others did not mention this step as in [12], [13], [20], [21] and [38].

Then the features are fed as input to the feature selection stage. Most classifiers are negatively influenced by redundant, correlated or irrelevant features. Thus, in order to reduce the dimensionality of the input data, a feature selection algorithm is implemented to choose the most significant features of the training data for the given task.

Feature selection presents several advantages. To begin with, small feature subsets require less memory and computations, whereas they also allow for a more accurate statistical modeling, thus improving performance. On the contrary, large feature sets may yield a prohibitive computational time for classifier training. Additionally, feature selection determines which features are the most beneficial and eliminates irrelevant features, leading to a reduction in the cost of acquisition of the data. Furthermore, if all the features are employed, there is the curse of dimensionality as well as the risk for over fitting. In addition, feature selection can boost performance when the number of training utterances is not sufficient or when a real-time problem needs to be handled.

Different feature selection algorithms were introduced like Forward Feature Selection (FFS), Principal Components Analysis (PCA), Genetic Algorithm (GA) and Sequential Forward Floating Search (SFFS) which can be used to encode the main information of the feature space more compactly as in [3], [8], [36], [39], [40].

D. The Classifier

Classification is the final stage of the SER system. Choosing the classifier type also has a great effect on the classification accuracy. Artificial Neural Network (ANN) is an efficient pattern recognition mechanism which simulates the neural information processing of human brain. The ANN processes information in parallel with a large number of processing elements called neurons and uses large interconnected networks of simple and non linear units. The computational intelligence of neural networks is made up of their processing units, characteristics and ability to learn. During learning the system parameters of NN vary over time and are characterized by their ability of local and parallel computation, simplicity and regularity.

Different ANN was used for emotion classification with different classification accuracy as in [7], [11], [20], [21]. The mentioned different aspects will be tested for different ANN types, topologies and parameters. Comparing results for different types of ANN will show the effect of features format, features selection and data sorting on each.

III. Database and Test Preparation

E. Database Used in the Experiment

The database used in this paper is called Berlin emotional speech database. This database is a simulated speech database, developed by the Technical University, Institute for Speech and Communication, Department of Communication Science, Berlin as in [23]. This database consists of seven basic emotions: anger, boredom, disgust, fear, happiness, sadness and neutral simulated using 535 speech data, categorized as in Fig. 2.

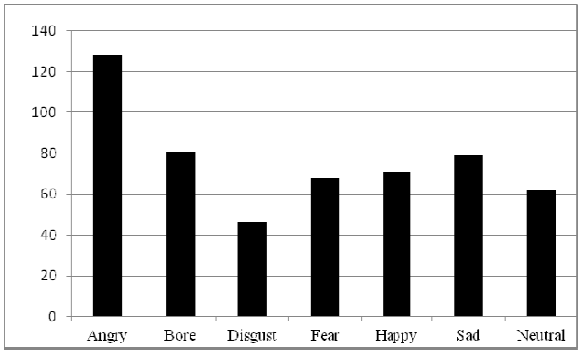

Fig. 2 Speech Samples Count by Emotion

F. Features Extraction

Several features were used in literatures to identify the different emotions using ANN as in [10]. In this context, 175 features were extracted from the different speech data and summarized in Table I to Table VI.

TABLE I
PITCH CONTOUR RELATED FEATURES

Indices	Features
1-6	Maximum, minimum, mean, range, standard deviation and interquartile range of pitch values.
7-11	Maximum, minimum, mean, range and standard deviation of first derivative pitch contour values
12-15	75 and 90th percentile of pitch and first derivative of pitch contour.
122-125	Maximum, mean, median, interquartile range of durations for the plateaux at minima
126-127	Mean and interquartile range of pitch values for the plateaux at minima
128-131	Maximum, mean, median, interquartile range of durations of the rising slopes of pitch contours
132-133	Mean and interquartile range of pitch values within the rising slopes of pitch contours
134-137	Maximum, mean, median, interquartile range of durations of the falling slopes of pitch contours
138-139	Mean and interquartile range of pitch values within the falling slopes of pitch contours

TABLE II
FORMANT CONTOUR RELATED FEATURES

Indices	Features
16-18	Mean of the 1st, 2nd and 3rd formant.
19-21	Max of the 1st, 2nd and 3rd formant.
22-24	Min of the 1st, 2nd and 3rd formant.
25-27	Median of the 1st, 2nd and 3rd formant.
28-30	Standard deviation of the 1st, 2nd and 3rd formant.
31-33	Interquartile range of the 1st, 2nd and 3rd formant.
34-36	BW of the 1st, 2nd and 3rd formant.

TABLE III
ENERGY CONTOUR RELATED FEATURES

Indices	Features
37-42	Mean, minimum, maximum, standard deviation, range and interquartile range of Short-Time energy contour.
43-45	30, 50 and 90th percentile of STenergy contour.
46-51	Mean, minimum, maximum, standard deviation, range and interquartile range of first derivative of Short-Time energy contour.
52-53	30 and 50th percentile of first derivative of STenergy contour.
54-59	Mean, minimum, maximum, standard deviation, range and interquartile range of db-energy contour.
60-62	30, 50 and 90th percentile of db-energy contour.
63-68	Mean, minimum, maximum, standard deviation, range and interquartile range of first derivative of db-energy contour.
69-71	30, 50 and 90th percentile of first derivative of db-energy contour.
152-155	Maximum, mean, median, interquartile range of durations of the rising slopes of db-energy contours
156-157	Mean and interquartile range of db-energy values within the rising slopes of db-energy contours
158-161	Maximum, mean, median, interquartile range of durations of the falling slopes of db-energy contours
162-163	Mean and interquartile range of first derivative db-energy values within the falling slopes of first derivative of db-energy contours
164-167	Maximum, mean, median, interquartile range of durations of the rising slopes of first derivative of db-energy contours
168-169	Mean and interquartile range of first derivative of db-energy values within the rising slopes of first derivative of db-energy contours
170-173	Maximum, mean, median, interquartile range of durations of the falling slopes of first derivative of db-energy contours
174-175	Mean and interquartile range of first derivative of db-energy values within the falling slopes of first derivative of db-energy contours

TABLE IV
SPECTRAL ENERGY RELATED FEATURES

Indices	Features
72-75	Spectral energy below 250,600, 1000, 1500Hz.
76-79	Spectral energy between the frequency ranges 250-600, 600-1000, 1000-1500, 250-1000Hz.

TABLE V
DURATION AND VOICING RELATED FEATURES

Indices	Features
80-84	Minimum, mean, range, median and standard deviation of voiced duration of speech.
85-89	Minimum, mean, range, median and standard deviation of unvoiced duration of speech.
90	Speaking rate

TABLE VI
EXTRA EXTRACTED FEATURES

Indices	Features
91-95	Mean, minimum, maximum, standard deviation and range of zero crossing rate.
96	Number of voiced segment
97-100	PF,PQ of jitter and shimmer
101-112	12 MFCC coefficient
113-121	9 LPC coefficient

G. Feature Selection

According to comparison done in [41] the Forward Feature Selection (FFS) algorithm was chosen for the selection stage since the classification test is done on a medium scale data. In addition, FFS algorithm is a simple, fast, effective and widely accepted technique as in [42]. The algorithm selected an optimum feature set of 129 in length.

H. Training Data Arrangement

The speech data was arranged in two groups: training group and test group. In the training group several speech data was replicated to ensure having equal number of data sets for each emotion resulting a final number of 480 speech files for training purpose and 121 files for testing.

The training data was further ordered in several arrangements according to speaker gender and emotion to explore the effect of data arrangement on the classification accuracy. The different data arrangements used are shown in Table VII.

TABLE VII
DIFFERENT DATA SORTING

Dataset name	Description
Data_1	Male-Female Emotions Disordered.
Data_2	Male-Female, Ordered by Emotion.
Data_3	Male Only, Disordered Emotion.
Data_4	Male Only, Ordered by Emotion.
Data_5	Female Only, Emotion Disordered.
Data_6	Female Only, Ordered by Emotion.
Data_7	Male-Female, Ordered by 2 Groups of Emotion..
Data_8	Male-Female, Ordered by Emotions and Gender.

I. ANN Used

The ANN architectures used for comparison purpose are:

1. Back Propagation Neural Networks (BPNN).
2. Learning Vector Quantization Neural Networks (LVQNN).
3. Probabilistic Neural Network (PNN).

Several types of BPNN were used, namely:

1. Feed Forward Back Propagation Neural Networks (FFBPNN).
2. Cascade Forward Back Propagation Neural Networks (CFBPNN).
3. Fit Back Propagation Neural Networks (FITBPNN).

The target from using different types was to explore the impact of the ANN type on the classification accuracy. Furthermore, different architectures were used for some of these types to compare the performance. In the next section, different experiments implemented on emotional recognition are described in details.

IV. EXPERIMENTAL RESULT AND DISCUSSIONS

After carrying numerous experiments on emotional recognition, it was found that the most significant factors that impact the classification success are:

1. Features format.
2. Number of features selected.
3. Training data sorting.
4. ANN type and architecture.

In this section, the impact of each of these factors is tested using Matlab and the resulting data is summarized for further analysis.

J. Impact of Features Format

In this test all five ANN types were used in the classification of the database in two different formats. The first, features were used in their row format while in the second; features were standardized around its mean and standard deviation using (5):

$$\hat{x}_{mnv} = \frac{x-\mu}{\sigma} \quad (5)$$

Feature standardization preserves all original feature relationships and does not introduce any bias in the features. These two types of formats are used to train a FFBPNN with input layer of 30 neurons, one hidden layer of 20 neurons, tansig function at input and hidden layer and purelin activation function at output layer. The FITBPNN trained with 40 neurons on the input layer, one hidden layer of 20 neurons and with default training functions and algorithms. The CFBPNN trained with 40 neurons with default training functions and algorithms. The LVQNN was trained using 40 hidden neurons. PNN trained with same spread for certain data groups of same dimension. Results are summarized in Table VIII, Table IX and Fig. 3.

TABLE VIII
AVERAGE SUCCESS RATE USING ROW FEATURES

ANN	Angry	Bore	Disgust	Fear	Happy	Sad	Neut.	Avg.
FFBP	30%	15%	24%	35%	29%	26%	26%	26%
FIT	13%	23%	19%	20%	28%	25%	29%	22%
CF	18%	41%	25%	17%	16%	15%	26%	23%
LVQ	33%	23%	26%	24%	20%	23%	22%	24%
PNN	74%	74%	33%	43%	49%	25%	48%	49%

TABLE IX
AVERAGE SUCCESS RATE WHEN FEATURES ARE STANDARDIZED

ANN	Angry	Bore	Disgust	Fear	Happy	Sad	Neut.	Avg.
FFBP	74%	67%	67%	77%	84%	69%	78%	74%
FIT	71%	75%	54%	62%	68%	70%	68%	67%
CF	69%	72%	66%	70%	70%	64%	67%	68%
LVQ	58%	78%	61%	70%	63%	80%	85%	71%
PNN	96%	83%	73%	76%	88%	79%	91%	84%

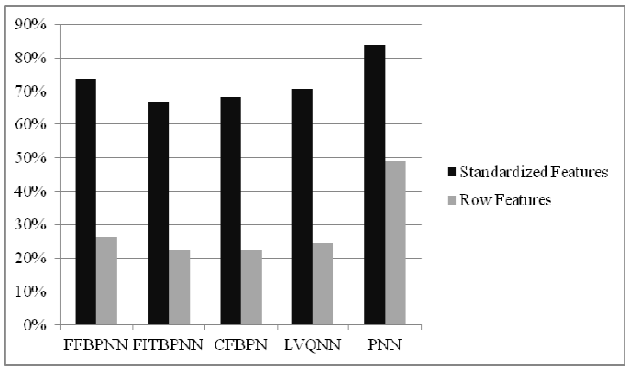

Fig. 3 Impact of Features Format for Different ANN Types

The results show that using row features degrade the classifier performance for all ANN used. Feature standardization improves the features generalization ability and guarantees that all the features obtain the same scale in order to ensure an equal contribution of each feature to the feature selection algorithm. Based on the above, it was decided to rely on standardized features for all the following tests.

K. Impact of Features Selection

Feature selection was implemented using FFS and the optimum number of features identified by the algorithm 129 features. Different tests were applied to compare the classification success rate when different numbers of features,

including the whole set, are used in ANN training. For each test, the rules of thumb as described in [43] was used to select the number of input and hidden neurons, if applicable, based on the training data dimension. This test was repeated for 3 different types of ANN and results are shown in Fig. 4.

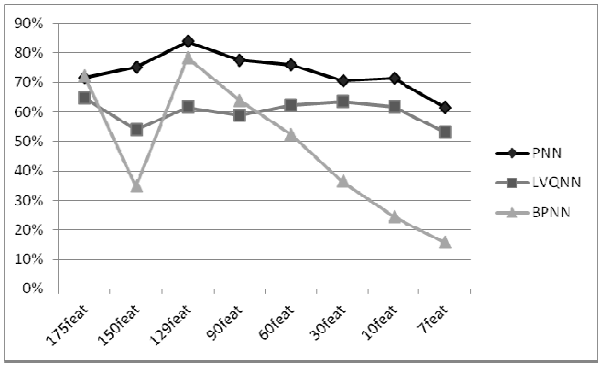

Fig. 4 Impact of Features Selection for Different ANN Types

Results shows that, the BPNN reached its highest success rate at 129 features matching the optimum predicted by the FFS algorithm used. However, the success rate dramatically degrades with the decrease of number of features. Hence, BPNN is not recommended when a reduced number of features is a must to reduce the processing time in online applications. LVQNN, on the other hand, offers high robustness with varying the number of features, although the maximum performance is still achieved at the full set of features. It is also clear that the PNN offers superior performance for all different sets of features, although the maximum is still achieved at the optimum point predicted by the FFS algorithm. When a reduced number of features is required, PNN is definitely the best choice. Choosing the best feature set with suitable spread setting enhance the performance accuracy and makes PNN a good online classifier.

To conclude with, the FFS algorithm was able to get the optimum number of features for both PNN and BPNN while for LVQ the highest success rate was achieved with the complete features set. Results also show that when a reduced set of features has to be used, PNN and LVQNN are preferred over the BPNN.

L. Impact of Data Sorting

The different data sets described in Table VII were used to train a FFBPNN with one hidden layer of 20 neurons, tansig function at input and hidden layer and purelin activation function at output layer. The LVQNN was trained using 40 hidden neurons. PNN trained with certain spread for certain data groups of same dimension. Results are shown in Fig. 5.

Results show that the training data sorting can significantly impact the recognition average success rate. It affects performance of BPNN and LVQ network. Hence, for any specific application, the researcher should try several combinations of training data sorting in order to achieve the maximum success rate. While for PNN, it can be noticed that data sorting doesn't affect the classification accuracy as long as the data size is the same.

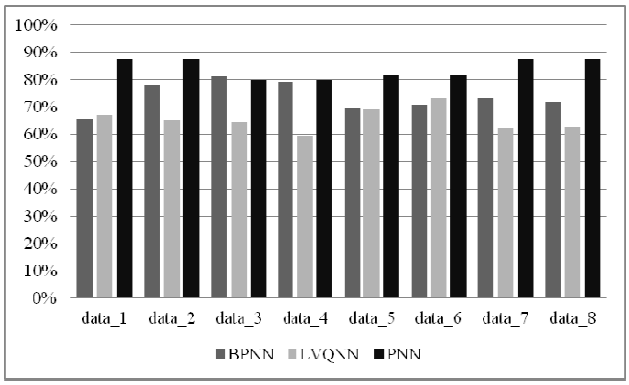

Fig. 5 Impact of Data Sorting in Different ANN Types

Last but not least, ANN topology and design have a significant effect on the classification accuracy. For BPNN and LVQNN choosing the number of hidden layers, number of neurons used for each, the activation function training and learning algorithm and rate used are important factors to consider. For PNN choosing the suitable spread values is equally important.

Tests revealed that PNN can be considered as a robust and fast ANN. In addition PNN offers the highest results repeatability when compared with BPNN and LVQNN.

V. Conclusion

In literatures, lots of efforts were spent to extract more features from speech in order to enhance the classification accuracy. In contrast, this work emphasizes on other factors that are proven to be equally important to achieve good classification results. It was shown that it is possible with a reduced number of features to get a good classification when ANN is used. Using a reduced number of features can help to considerably decrease the computational effort and time delay especially when dealing with online systems. In addition, using reduced number of features can considerably simplify the hardware implementation of the classifier if needed.

VI. References

[1] C. Busso, Z. Deng, S. Yildirim, M. Bulut, C. Lee, A. Kazemzadeh, et al., "Analysis of Emotion Recognition Using Facial Expressions, Speech and Multimodal Information.," in *International Commission on Mathematical Instruction*, 2004, pp. 205-211.

[2] S. Ramakrishnan, "Recognition of Emotion from Speech: A Review, Speech Enhancement, Modeling and Recognition- Algorithms and Applications," *ISBN 978-953-51-0291-5, Hard cover, 138 pages, Publisher: InTech, Chapters published March 14, 2012* under CC BY 3.0 license DOI: 10.5772/2391, pp. 66-80.

[3] Zhongzhe Xiao, "Features Extraction and Selection for Emotional Speech Classification," in *Advanced Video and Signal Based Surveillance, 2005. AVSS ,IEEE Conference*, Ecully, France, Sept. 2005, pp. 411- 416.

[4] Margarita Kotti, Fabio Paternò, "Speaker-Independent Emotion Recognition Exploiting a Psychologically-Inspired Binary Cascade

Classification Schema," in *International Journal of Speech Technology*, vol. 15, no. 2, June 2012, pp. 131-150.
[5] K. J. Patil,P. H. Zope,S. R. Suralkar, "Emotion Detection From Speech Using Mfcc & Gmm," *International Journal of Engineering Research & Technology (IJERT)*, vol. 1, no. 9, November- 2012.
[6] Björn Schuller, Gerhard Rigoll, Manfred Lang, "Speech Emotion Recognition Combining Acoustic Features And Linguistic Information In A Hybrid Support Vector Machine - Belief Network Architecture," in *Acoustics, Speech, and Signal Processing. (ICASSP '04). IEEE International Conference on*, 2004, pp. I- 577-80.
[7] Mina Hamidi, Muharram Mansoorizade, "Emotion Recognition From Persian Speech With Neural Network," *International Journal of Artificial Intelligence & Applications (IJAIA)*, vol. 3, no. 5, September 2012,p. 107.
[8] Constantine Kotropoulos, Dimitrios Ververidis, "Sequential Forward Feature Selection With Low Computational Cost," in *European Signal processing conference*, Turkey, 2005.
[9] Keshi Dai , Harriet J. Fell, Joel MacAuslan, "Recognizing Emotion In Speech Using Neural Networks," in *Telehealth/AT '08 Proceedings of the IASTED International Conference on Telehealth/Assistive Technologies*, ACTA Press Anaheim, CA, USA, 2008, pp. 31-36.
[10] Mehmet S. Unluturk, Kaya Oguz, Coskun Atay, "Emotion Recognition Using Neural Networks," in *10th WSEAS (World Scientific and Engineering Academy and Society) international conference on Neural networks*, USA, 2009, pp. 82-85.
[11] Yafei Sun, "Neural Networks for Emotion Classification," 2003.
[12] K B Khanchandani, Moiz A Hussain, , "Emotion Recognition Using Multilayer Perceptron and Generalized Feed Forward Neural Network," *Journal of Scientific and Industrial Research (JSIR)*, vol. 68, no. 05, May 2009, pp. 367-371.
[13] A.A. Razak, R. Komiya, M. Izani, Z. Abidin, "Comparison Between Fuzzy and NN Method for Speech Emotion Recognition," in *Information Technology and Applications, 2005. ICITA*, 2005, vol 1, pp. 297- 302.
[14] Björn Schuller, Gerhard Rigoll, Manfred Lang, "Hidden Markov Model-Based Speech Emotion Recognition," in *Multimedia and Expo, 2003. ICME '03. Proceedings*, 2003, vol.1, pp. I-401-4.
[15] M. Murugappan, "Human Emotion Classification Using Wavelet Transform and KNN," in *Pattern Analysis and Intelligent Robotics (ICPAIR),* 2011, pp. 148-153.
[16] Dimitrios Ververidis, Constantine Kotropoulos, "Emotional speech Classification Using Gaussian Mixture Models," in *Circuits and Systems. ISCAS 2005. IEEE International Symposium*, 2005, pp. 2871-2874.
[17] Chung-hsien Wu , Ze-jing Chuang, "Emotion Recognition Using IG-based Feature Compensation and Continuous Support Vector Machines," in *International Journal of Computational Linguistics and Chinese Language Processing*, vol.12, no. 1, 2007, pp.65-78.
[18] Jarosław Cichosz, Krzysztof Ślot, "Emotion Recognition in Speech Signal Using Emotion-Extracting Binary Decision Tree," in *Polish State Fund for Research*, 2008.
[19] Dimitrios Ververidis, Constantine Kotropoulos, "Emotional Speech Recognition: Resources, Features, and Methods" in *Speech Communication*, April 2006.
[20] A. Firoz Shah., A. Raji Sukumar, P. Babu Anto., "Discrete Wavelet Transforms and Artificial Neural Networks for Speech Emotion Recognition" in *International Journal of Computer Theory and Engineering*, vol. 2, no. 3, June 2010.
[21] A. Firoz Shah.,A. Raji Sukumar.,P. Babu Anto., "Automatic Emotion Recognition from Speech using Artificial Neural Networks with Gender-Dependent Databases," in *International Conference on Advances in Computing, Control, and Telecommunication Technologies (IEEE)*, 2009.
[22] Dimitrios Ververidis, Constantine Kotropoulos, "Automatic Speech Classification to Five Emotional States Based on Gender Information," in *European Signal Processing Conference (EUSIPCO)*, Austrilia, 2004, pp. 341-344.
[23] Institute for Speech and Communication Technical University. Berlin Database of Emotional Speech. [Online]. http://pascal.kgw.tu-berlin.de/emodb/index-1280.html
[24] Tomi Kinnunen, Haizhou Li, "An Overview of Text-Independent Speaker Recognition: from Features to Supervectors," in *Speech Communication*, vol. 52, no. 1, January 2010, pp. 12-40.
[25] L. Rabiner, "On the Use of Autocorrelation Analysis for Pitch Detection" in *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 25, no. 1, Feb 1977, pp. 24-33.
[26] W. J. Hess, *Pitch Determination of Speech Signals: Algorithms and Devices*.: Springer Series in Information Sciences, 1983 pp.415-460.
[27] D. Talkin, "A Robust Algorithm for Pitch Tracking (RAPT)," in *Speech Coding and Synthesis, Elseiver Science*, Amsterdam, 1995, pp. 495-518.
[28] M. J. Ross, H. L. Shaffer, A. Cohen, R. Freudberg, H. J. Manley, "Average Magnitude Difference Function Pitch Extractor," in *Acoustics, Speech and Signal Processing, IEEE Transactions on* , vol 22, no. 5, Oct.1974., pp. 353–362.
[29] A. M. Noll, "Cepstrum Pitch Determination," in *Journal of the Acoustical Society of America*, vol. 41, no. 2, 1967, pp. 293-309.
[30] L. Rabiner, M. J. Cheng, A. E. Rosenberg, C. A. McGonegal, "A Comparative Performance Study of Several Pitch Detection Algorithms," in *IEEE Transactions on ASSP*, vol. 24, 1976, pp. 399-417.
[31] H. Bořil, P. Pollák, "Direct Time Domain Fundamental Frequency Estimation of Speech in Noisy Conditions," in *Proceeding of EUSIPCO2004, Wien,* vol. 1, Austria, 2004, pp. 1003-1006.
[32] Tuomo Raitio, Antti Suni, Junichi Yamagishi, Hannu Pulakka, Jani Nurminen, Martti Vainio, Paavo Alku, "HMM-Based Speech Synthesis Utilizing Glottal Inverse Filtering," in *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 19, no. 1, January 2001, pp. 153-165.
[33] Alku, Paavo, "Glottal Wave Analysis with Pitch Synchronous Iterative Adaptive Inverse Filtering," in *Speech Communication - Eurospeech '91*, vol. 11, no. 2-3, June 1992, pp. 109-118.
[34] Johannes Pittermann, Angela Pittermann, Wolfgang Minker, *Handling Emotions in Human-Computer Dialogues*.: Springer, 2009.
[35] Jia Rong, "Acoustic Features Extraction for Emotion Recognition," in *Computer and Information Science, 6th IEEE/ACIS International Conference*, Melbourne , July 2007, pp. 419- 424.
[36] N. Kwak, "Input Feature Selection for Classification Problems" in *Neural Networks, IEEE Transactions*, vol. 13, no. 1, Jan 2002, pp. 143-159.
[37] Wauter Bosma, Elisabeth Andr E, "Exploiting Emotions to Disambiguate Dialogue Acts," in *Intelligent User Interfaces - IUI ,* March 2004, pp. 85-92.
[38] Hao Tang ;Chu, S.M. ; Hasegawa-Johnson, M. ; Huang, T.S. , "Emotion Recognition from Speech VIA Boosted Gaussian Mixture Models," in *Multimedia and Expo, 2009. ICME. IEEE* , 2009, pp. 294-297.
[39] Jianping Huaa,Waibhav D.Tembeb, EdwardR.Dougherty, "Performance of Feature-Selection Methods in the Classification of High Dimension Data,"in *Elsevier*, vol. 42, no. 3, 2009, pp. 409-424.
[40] M. Soryani , N. Rafat, "Application of Genetic Algorithms to Feature Subset Selection in a Farsi OCR," in *Proceedings of World Academy of Science, Engineering and Technology*, 2006, pp. 113-116.
[41] Mineichi Kudo , Jack Sklansky, "Comparison of Algorithms that Select Features for Pattern Classifiers," *Pattern Recognition-PR*, vol. 33, no. 1, 2000, pp. 25-41.
[42] Delft Pattern Recognition Research, Faculty EWI - ICT, Delft University of Technology. PRTools, A Matlab toolbox for pattern recognition. [Online]. http://prtools.org/
[43] Kevin L. Priddy, Paul E. Keller, *Artificial Neural Networks: An Introduction*, 1st ed.: SPIE Press, 2005,pp. 107-116.