NEURAL NETWORK APPROACHES FOR CHILDREN'S EMOTION RECOGNITION IN INTELLIGENT LEARNING APPLICATIONS

Felix Albu[1], Daniela Hagiescu[2], Liviu Vladutu[3], Mihaela-Alexandra Puica[2]

[1] *Valahia University of Targoviste (ROMANIA)*
[2] *Danimated SRL (ROMANIA)*
[3] *Polytechnic University of Bucharest (ROMANIA)*

Abstract

In this paper the children's emotion recognition performance of several neural networks approaches is described. The Radial Basis Function (RBF), Probabilistic Neural Networks (PNN), Extreme Learning Machines (ELM) and Support Vector Machines (SVM) variants were tested on recorded speech signals and face detected images.

For the speech signal the Mel Frequency Cepstral Coefficients (MFCC) and other parameters were computed together with their mean and standard deviation in order to obtain the feature vector for the neural network input.

For images, input parameters for emotion detection consisted in several distances computed between certain facial features using space coordinates for eyes, eyebrows and lips.

In case of RBF networks and speech signals we investigated the influence of the number of centres chosen by the k-means algorithm on the recognition performance on both training and test databases. The FAU Aibo Emotion corpus database was used because it has recordings from 51 children aged 10 to 13 years while interacting with a Sony Aibo robot. It is shown that there is a limitation of performance over a certain number of centers for the chosen identified emotions.

Another promising technique for classification of speech feature vectors is the use of ELM. They are Single-hidden Layer Feedforward Neural (SFLN) networks. In this case, random values are allocated to the weights of the hidden layer and the output weights are found by matrix operations.

Our simulations have shown a similar behavior of ELM networks with the RBF networks. There is a limitation of ELM performance after an increase of the number of hidden neurons over a specific number. Also, it is shown that the variant called Online Sequential ELM (OS-ELM) obtains very close classification performance to that of ELM.

For facial emotion recognition a subset of 20 subjects ages 6 to 9 (10 boys and 10 girls) from The Dartmouth Database of Children's Faces was used. Different types of RBF networks (Classic RBF, Multi-Stage RBF, and Probabilistic Neural Networks) with variable number of hidden neurons were trained and tested.

SVMs are a new type of supervised nonlinear learning paradigms which were used in the last decades both for classification and regression analysis. They have shown similar performance to RBF networks in our emotion detection simulations.

The results prove the effectiveness of several neural networks techniques in estimating the children affective state that can have important implications on technology-enhanced learning and intelligent software applications for children. It is shown that child affective modeling it is as important as their cognitive modeling when it comes to deciding the next tutoring step and how it should be delivered.

Keywords: Intelligent tutor, neural networks, emotion recognition.

1 INTRODUCTION

The building of an intelligent tutor that evaluates the school-aged children's handwritten symbols and takes into account the emotions is a difficult problem. Usually, the teacher is assessing child's facial expression and voice pattern and adapt the pedagogical approach. The teacher also takes into account the child personality traits. Therefore, an intelligent tutor should mimic the behavior of a teacher.

There are multiple methods to assess the child interest starting from recorded voice and face detected images [5]. We propose to identify three types of emotions - positive, negative and neutral. Although

most papers (e.g. [6]-[8] etc.) identify more emotions, we found that the mentioned ones are useful for our application. Depending on the identified emotion and long term child's attitude, a recommendation for a specific strategy is made.

The paper is organized as follows. Several neural network structures are succinctly presented in Section 2. In section 3 several details about the speech emotion detection method are provided. Section 4 describes the expression classification technique using face landmarks such us space coordinates for eyes, eyebrows and lips. Section 5 presents the implications of technology-enhanced learning. Finally, the conclusions regarding the emotion based strategy and ideas for further improvements of the intelligent tutor are proposed.

2   NEURAL NETWORK STRUCTURES

Many machine learning methods have been used for emotion detection from both speech and images, each one having advantages and disadvantages. These systems typically use high-dimensional feature vectors in order to attain a high emotion recognition accuracy. In this paper we investigated the use of Radial Basis Functions (RBF) [9], the Extreme Learning Machines (ELM) [10], the Probabilistic Neural Networks (PNN) [11] and the Support Vector Machines (SVM) [12-13]. Other methods that can be used are the Gaussian Mixture Models (GMM) [14], k-NN [15] or other advanced recurrent neural networks (RNN) based solution [16].

2.1   The radial basis functions

The radial basis function (RBF) have been introduced in [9]. We use the implementation of [17] that has three main steps of the training process [18]:

1. Prototype selection through k-means clustering;

2. Evaluation of the beta coefficient for each RBF neuron;

3. Training of output weights for each category solving the normal equations. This approach is simpler, faster and guaranteed to yield the optimum weight values [17].

The Gaussian similarity function has been chosen, $\varphi(\mathbf{x}) = e^{-\beta \|\mathbf{x}-\mu\|^2}$ where μ is the mean of the distribution and β is a parameter that controls the width of the RBF neuron activation function. We set $\sigma = \frac{1}{m}\sum_{i=1}^{m}\|x_i - \mu\|$ and $\beta = 1/(2\sigma^2)$ where m is the number of training samples of the cluster and x_i is the ith training sample in the cluster [17].

2.2   The extreme learning machines

A promising technique are the Extreme Learning Machines (ELM). They are Single-hidden Layer Feedforward Neural (SLFN) networks. In this case random weights are allocated to the hidden neurons and never updated. The output weights of the SFLN are learned in a single step using matrix operations. It was shown that ELM is a very fast neural network with very good generalization properties [10]. However, the ELM performance is very sensitive to the number of hidden neurons.

In case of a network with one hidden layer and L hidden nodes the output is $f(\boldsymbol{\beta}) = \sum_{i=1}^{L} \beta_i h_i(\mathbf{x}) = \mathbf{h}(\mathbf{x})$ where $\mathbf{h}(\mathbf{x}) = [h_1(\mathbf{x}),...,h_L(\mathbf{x})]$ is the output vector of the hidden layer with respect to the input $\mathbf{x}$ and $\boldsymbol{\beta} = [\beta_1,...,\beta_L]^T$ is the vector of output weights between the hidden neurons and the output nodes [19].

The matrix $\mathbf{H} = [\mathbf{h}(\mathbf{x}_1),...,\mathbf{h}(\mathbf{x}_N)]^T$ is defined or a training dataset S with N samples. This leads to $\boldsymbol{\beta} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{T}$, where $\mathbf{T}$ is the vector of real target t_i with respect to an input. Therefore the ELM network needs to calculate the pseudoinverse of $\mathbf{H}$ [20].

In order to learn the model online, the online sequential ELM (OS-ELM) was proposed in [21]. In this case we have $\beta_0 = P_0 H_0 T_0$ and $P_0 = \left(H_0^T H_0\right)^{-1}$, for an initial training dataset S_0 and N_0 training samples. When new training samples are available the output weights are updated by

$$\beta^{(k+1)} = \beta^{(k)} + P_{k+1} H_{k+1}^T \left(T_{k+1} - H_{k+1}\beta^{(k)}\right) \text{ and } P_{k+1} = P_k - P_k H_{k+1}^T \left(I + H_{k+1} P_k H_{k+1}^T\right)^{-1} H_{k+1} P_k, \text{ where}$$

H_{k+1} is the hidden layer output for the (k+1)th arriving training data [19].

2.3   Probabilistic neural networks

A probabilistic neural network (PNN) is an implementation of the kernel discriminant analysis algorithm organized as a multilayered feedforward network with four layers: input layer, pattern layer, summation layer and output layer.

Our current implementation uses 11 distances obtained from the children facial landmarks (as described in [18]) as the inputs for the PNN, which combines the theory of Bayesian classification and the estimation of probability density functions (PDFs) for the 11 above-mentioned inputs. Currently, all the computed distances are given as input to the network used for pattern recognition of emotion, unlike work done in [11] where relevant inputs are computed as prerequisite. An offline analysis of this type is in progress using RapidMiner [26], a software platform which provides an integrated environment for machine learning, data mining, text mining, predictive analytics and business analytics. Knowing which are the most relevant inputs (like in [27]) we can weight them in a mixed classifier (which is to combine the deterministic and probabilistic information).

However, even so, the classification results obtained with PNN were the highest, along with those obtained with multi-stage RBF neural networks, as can be seen in table 1 from Section 4.2.

2.4   Support vector machines

The Support Vector Machines (SVMs) were first introduced in 1992 by Boser, Guyon and V. Vapnik [28] and they had an astounding evolution since then. Moreover, they lead the way to the Kernel Machines, a larger class of learning algorithms of which SVMs are a particular instance [29].

Over the last 15 years SVM have proven to be successful for applications in many fields (bioinformatics, natural language processing, handwriting recognition, speech [13] or emotion recognition).

The main reason that made us chose SVM for our application is that this class of kernel methods implicitly defines the class of possible patterns by introducing a notion of similarity between data, i.e. in this case different children that share the same emotional status are sharing the same change of the biometric features (e.g. for surprise, inner and outer eyebrows raised, jaw drop and upper lid raised or for sadness, the lip corners depressed and chin raised). These are summarized in [30] and we have also shown an example in Fig. 6.

As for any other kernel-based machines, the learning algorithms for SVMs are composed of two main parts:

- a general purpose learning machine;
- a problem specific kernel function.

From the kernels that we used in our classification problem (Radial Basis Function kernels, polynomial or linear kernels) the first class (RBF-based) were the most successful, as the results show in Table 1. For the implementation, we have chosen libSVM [31], a library which in the last 15 years has been used in several applications. A typical use of libSVM involves two steps: first, training a data set to obtain a model and second, using the model to predict the accuracy for a testing data set. The (ANSI) C implementation is one of the fastest version of the libSVM library and therefore this version was used for training, testing and scaling our data.

Currently, there is work in progress regarding optimization of the hyperparameters specific to a SVM (like the one controlling the slack-variables penalty weight, C, and gamma). This is because the Gaussian kernel has the widest spread in SVM applications, whose bandwidth parameters are γ or σ (where σ is the spread).

The training of an SVM consists in the following optimization problem:

$$min_{w,\xi} = 1/2\|w^2\| + C\sum_i(\xi_i)$$

where w represents the hyperplane normal vector and C is the weight of the penalty function which is formulated by the sum of all ξ slack variables [33].

A recent model for the algorithm optimization, which has proven to be one of the most successful, is presented in [34]. Also LibSVM offers scripts for performing grid search and Python scripts for some parameters optimal selection.

3 SPEECH EMOTION DETECTION TOOL

Speech emotion recognition is a tough problem because it is difficult to find effective features. Important nonverbal cues such as speech rate, pitch can help increase the recognition rate.

The detected emotions are the following: positive, negative and neutral. A feature vector is computed by using speech samples. These feature vectors are presented as an input to a neural network structure. A training is performed on shuffled feature vectors on a training database. Two thirds are used for the training database, while the last third are used for the test. The recognition results are averaged over 100 trials.

3.1 Feature selection choice

The feature vector is formed with the following parameters (the short term energy, the zero-crossing rate, the spectral roll-off, spectral centroid and 13 mel frequency cepstral coefficients (MFCC).

The first element is the short term energy. This parameter is computed for each frame of 10 ms of speech. It is widely used for silence periods detection and audio classes discrimination. The energy of voiced frames is higher than that of unvoiced frames [7].

The zero-crossing rate is the rate of sign changes for the speech signal [7]. It is a rough indicator of the frequency content of the speech signal and it is used together with the short term energy for silence period detection. It is known that unvoiced or noisy frames have a higher zero-crossing rate than voice frames [7].

The spectral roll-off is a measure of the spectral shape and it is defined as the frequency below which 80% of the magnitude distribution is concentrated [8]

The spectral centroid feature is the center of mass of the spectrum and it is a measure of the spectral position [8].

Mel-frequency coefficients (MFCC) are perceptually motivated features that are also based on the STFT. The design of Mel-frequency filters models the human auditory system and takes into account the psychoacoustics.

The MFCC are standard features for many speech applications [7]. The MFCC are perceptually motivated and takes into account the psychoacoustics by modeling the human auditory system. The details about computing these MFCC vectors can be found in [7]. We compute for each frame 13 MFCC.

These vectors are computed for each frame and the final feature vector is formed by their mean and standard deviation for each speech sample leading to 34-length vector (Fig.1).

Mean zero crossing rate	Mean spectral roll off	Mean spectral centroid	Mean short time energy	Std zero crossing rate	Std spectral roll off	Std spectral centroid	Std short time energy	13 Mean MFCC	13 Std MFCC

Fig. 1: a) 34-element feature vector for a speech file.

An example of such feature vector for a negative emotion is shown in Fig. 2.
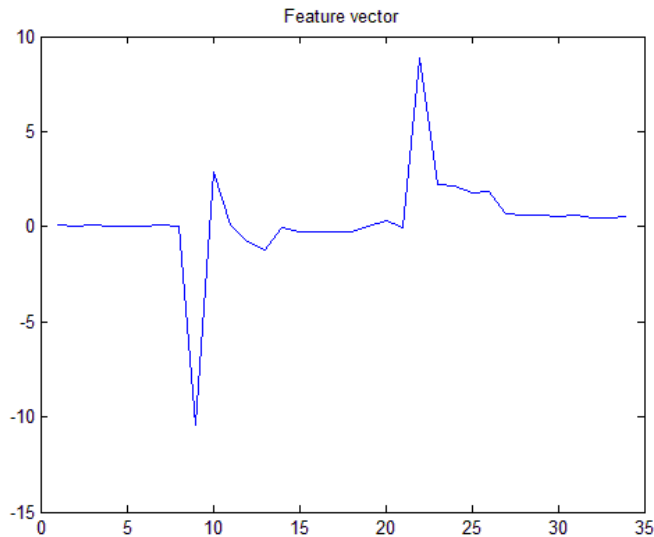
Fig. 2: An example of a feature vector for a "negative" emotion

The FAU-AIBO speech samples were used [22], [23]. It was mentioned in [22-23] that the human performance is about 80% for two classes and about 65% for four classes, therefore we can expect a performance around 75% for three classes. The FAU Aibo Emotion Corpus database includes the recordings of 51 children aged 10 to 13 years while interacting with a Sony Aibo robot. The recordings were segmented in small wav files on 16 bits and sampled at 16 kHz. The emotions annotated by people were recorded and an overall evaluation for full sentence is made together with information about the sex and age of the children. Three emotions were considered (positive, negative and neutral). We tested the RBF, ELM and OS-ELM networks for clean speech signals and white noise corrupted files with SNR = 10 dB.

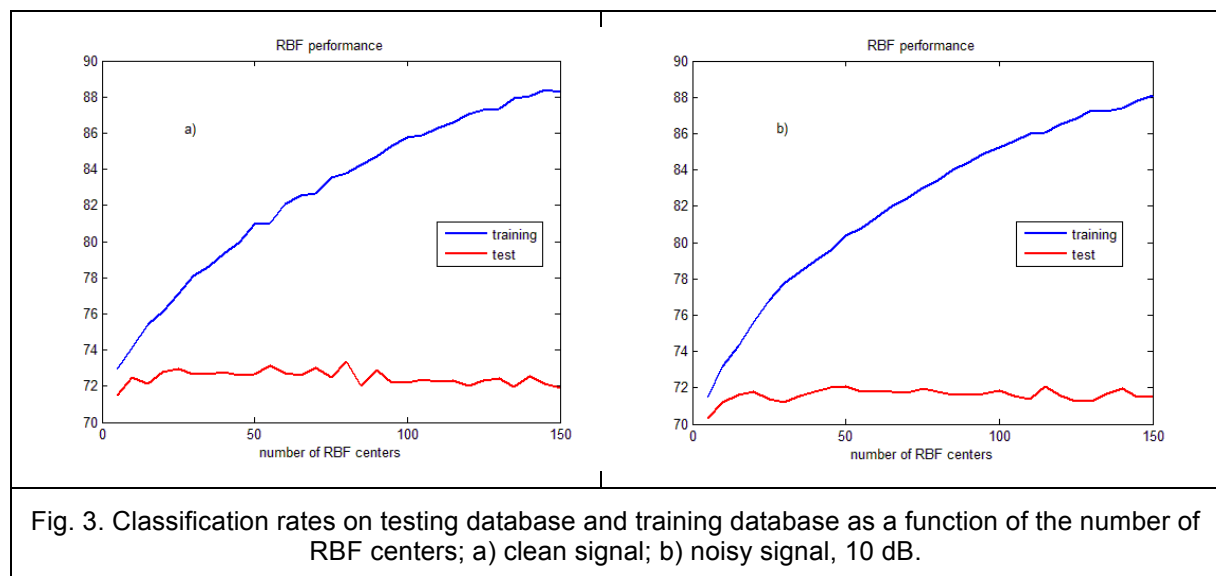In Fig. 3 the results on the test database are shown for the RBF network.

Fig. 3. Classification rates on testing database and training database as a function of the number of RBF centers; a) clean signal; b) noisy signal, 10 dB.

The maximum average RBF performance on the testing database is about 73.5% and the optimum number of centers is 80 for the noise-less case. It can be noticed from Fig. 3a that the testing classification rate is flattening after about 50 chosen RBF centers. The 10 dB additive noise leads to a reduction with about 1.5% of the recognition error. The maximum performance is about 72% and is obtained for 115 RBF centers. It can be said that the RBF emotion recognition performance is robust to noise and a similar behavior can be noticed regarding the flatness after about 50 chosen RBF centers and a trend towards over-fitting after about 100 RBF centers.

The ELM simulations shows a different behavior for a small number of hidden neurons and confirms the ELM sensitivity for this case observed in other applications [10]. The maximum average classification rate is about 72.1% for the noise-less case and 71.3% for the noisy speech samples. In both cases the ELM network needs 125 hidden neurons. It can be noticed from Fig. 4 that the classification rate is low for a number of hidden neurons smaller than 45.
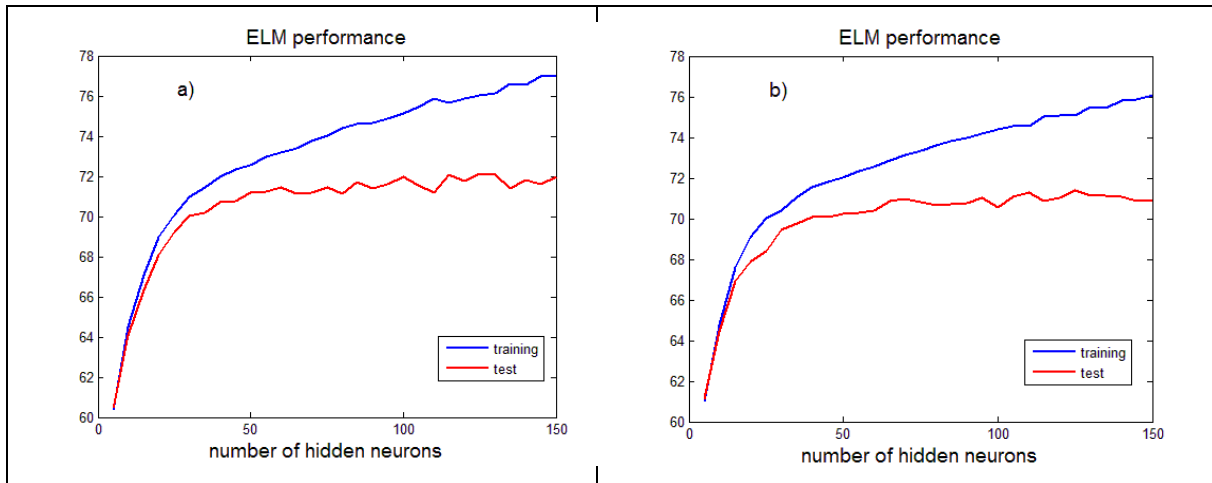

Fig. 4. Classification rates on testing database and training database as a function of the number of the hidden neurons with sigmoidal activation of the ELM; a) clean signal; b) noisy signal, 10 dB.

A better suited variant for a changing acoustic environment is the online sequential ELM (OS-ELM) [21]. In Fig. 5 the results of ELM, OS-ELM for both clean and noisy signals are shown.
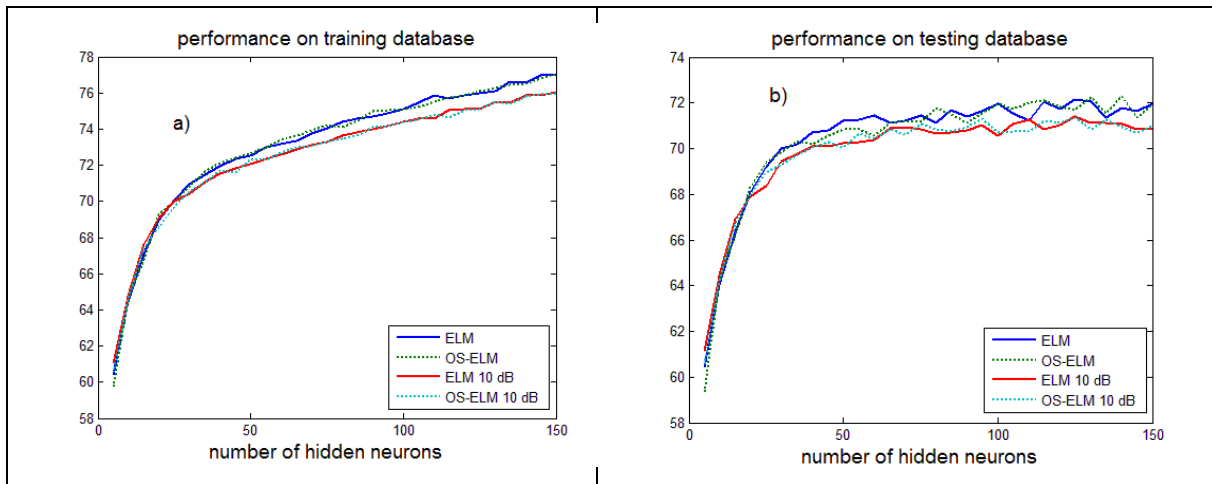

Fig. 5. Classification rate as a function of the number of the hidden neurons with sigmoidal activation for ELM and OS-ELM a) on training database; b) on testing database.

The activation function is the sigmoidal function. It can be noticed that the OS-ELM network achieves slightly better recognition only for some hidden neuron numbers. The overall conclusion is that the classification results of RBF is better than that of the ELM/OS-ELM networks. However, the speed of the ELM variants is higher than that of the RBF networks even if they need a higher number of hidden neurons in order to achieve the same recognition rate. It can be said that the RBF provides a better complexity/performance compromise than ELM. The results are very similar with those obtained by using a much larger vector length from [18] (i.e. more than three times). This lead to an important complexity reduction of the overall computation times.
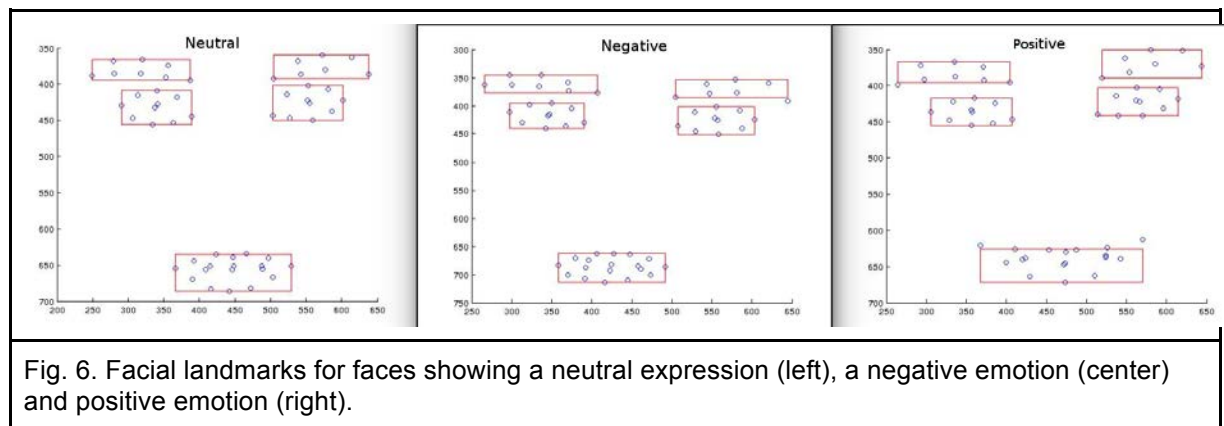
4   FACE DETECTED EMOTIONS

In human-human interaction, facial expressions are the first cues to indicate one's affective state. Even if the person tries to hide it, the face muscles contract to produce specific expressions that betray the emotion felt. For this reason, automatic emotion recognition from facial expressions is a well-studied field of research. Nevertheless, most of the studies concern emotion detection using adult subjects. In our work, we focus on children ages 5 to 9 years old, since they are the target for the intelligent tutor developed [18].

4.1   Data selection, preprocessing and transformation

We use the Dartmouth Database of Children Faces [32], which contains thousands of images taken for 100 children (50 boys and 50 girls) ages 5 to 15 years old showing 8 facial expressions: anger, disgust, fear, happy, neutral, pleased, sad and surprise. We extract from this database only the frontal images leaving us with 1276 images and we cluster them in three classes: positive (happy, pleased, surprised), negative (disgust, sad, angry) and neutral. We leave out fear since it is less likely to be triggered in an educational context. We also separate a set of 242 images that correspond to 20 subjects (10 boys and 10 girls) which are part of our target age group (5 to 9 years old) and we focus the testing process on these samples.

The first step is to detect the landmarks on children's faces, thus distinguishing the elements of interest on the face: eyes, eyebrows and mouth. By muscle contraction, shape of these elements and distance between them changes for different emotional states. In Fig. 6 the points that track the above mentioned elements and the rectangles that enclose them are plotted for one subject showing the three facial expressions considered: neutral, negative and positive. Even though the differences are small, there can be seen that a negative expression has a smaller mouth and closer eyebrows and a positive expression has a wider mouth and eyebrows are spread apart (in comparison with the neutral).

Fig. 6. Facial landmarks for faces showing a neutral expression (left), a negative emotion (center) and positive emotion (right).

We thus compute 11 distances that are to define these differences: two distances between centre of the left/right eye and center of the left/right eyebrow, two distances between corner of the left/right eye and corner of the left/right eyebrow, four distances measuring width and height of the eyes, distance between corner of the eyebrows and width and height of the mouth. These 11 distances form the input for the tested classifiers.

4.2 Classification results

The tested classifiers are Multi-stage Radial Basis Function Neural Network (RBF-NN), Probabilistic Neural Network (PNN) and the multi-class classification Support Vector Machines (SVM).

For each classifier we varied the parameters and the input and test data. Thus, we has three test cases:

- training set was composed of all the 1276 images and test set was composed of the selected 242 images; in this case, all the subjects with all the facial expressions from the test data set are included in the training set and therefore we expect to obtain the highest accuracy;

- training set was composed of 1040 images remaining after excluding the 242 selected images from the 1276 total images; in this case, no subject is included in the training set; moreover, the tested subjects are aged 5 to 9 years old and training is done on elder subjects; the expected accuracy is thus the smallest;

- training set was composed of 1040 images plus a variable number of images from the set of 242 selected images; in this case, the facial expressions of a given subject are found both in training and test set, but not for the same emotion.

For the RBF neural network we varied the number of centers, obtaining the best accuracy for 25 centers (in the cases where the test set contains new subjects or new emotions) and 40 centers (in the case where the test set is included in the training set).

For the PNN we varied the spread of the radial basis functions in the networks, obtaining the best accuracy for spread=0.1 (in the cases where subjects in the test set are included in the training set) and for spread = 02 (in the case where the test set contains only new subjects).

For the SVM we varied the kernel function and its parameters. The best results were obtained for radial basis kernel function with gamma=9 and C=10 (for the case where test data is included in training data) and, gamma=10 and C=10 (for the case where subjects in test data are included in training data) and gamma=1 and C=4 (for the case where test data contains only new subjects).

Table 1 resumes the best classification results obtained for the three test cases presented above.

Table 1. Best classification results for tested classifiers for face-emotions.

Classifier	Best accuracy on test data included in the training data	Best accuracy on test data with subjects included in the training data	Best accuracy on test data including only new subjects
RBFNN	91.32%	91.80%	83.06%
PNN	97.25%	91.27%	80.58%
SVM	98.35%	90.16%	84.71%

The classification results presented in Table 1 confirm our expectations for the three test cases considered. The accuracy has slightly decreased from the values obtained in previous work (94%) [18], but this is not a real concern since the current training set is larger and also the test set contains images not present in the training set.

5 AFFECT-AWARE LEARNING APPLICATIONS

Human emotion, as one of the 12 major challenges in the field of cognitive neuroscience [35], is completely intertwined with cognition in guiding children in the process of acquiring the necessary skills. Our contribution seeks to extend prior work in the field of intelligent tutors (e.g. [36]) and show that not only the understanding of negative feelings (like frustration or boredom) are important in the learning process but also the positive ones (like pleasure or happiness) because they can help design a better tutor - a positive emotional state of the student suggests that the pedagogical strategy applied up to the given moment is successful and the tutor should, in this case, minimize its intervention.

In developing the tutor we consider that the detection of a negative user affective state should trigger motivation-increase mechanisms (like telling a story that would increase student curiosity in the

current topic). Also, the successful completion of a learning cycle should trigger a reward mechanism. This concept of applying game mechanics to non-game contexts (in this case, an educational context) is called gamification, since the reward mechanisms is always present in the game design through passing to the next level of the game, or earning extra-points or lives. That is also motivated from a psychological and neurological point of view, since "Motivation is one emotion strongly linked to learning and has been defined as an inner drive that causes a person to act with direction and persistence" [37]. One possible decision that the tutor can take when the sentiment-analysis algorithm detects a negative feeling (like boredom, sadness or frustration) is to offer the child some kind of hints using easy, cheerful, multimedia-based multiple-choice quizzes.

6   CONCLUSIONS

In this paper we evaluate several network structures for an emotion recognition application. The emotion recognition is based on recorded speech signals and face detected images. We investigated the robustness of the classification rate to noise on both speech and image signals. Our future work will be focused on correlating the results using images and speech and improve the results by selecting the most suitable features for the chosen network architecture. Also, we take into consideration improving our classification tools, as specified in Sections 2.3 and 2.4. Moreover, a more complex approach would imply aggregating the output from multiple kernels like in the recent work of prof. ErkkiOja et al. [38].

ACKNOWLEDGMENTS

REFERENCES

[1]   Haq, S. and Jackson, P.J.B. (2010). Multimodal Emotion Recognition, In W. Wang (ed), Machine Audition: Principles, Algorithms and Systems, IGI Global Press, ISBN 978-1615209194, chapter 17, pp. 398-423.

[2]   Slot, K., Cichosz J., Bronakowski L. (2009) Application of Voiced-Speech Variability Descriptors to Emotion Recognition. CISDA 2009: 1-5.

[3]   Espinosa H. P., García C. A. R., Pineda L. V. (2011). EmoWisconsin: An Emotional Children Speech Database in Mexican Spanish. ACII (2) 2011: 62-71.

[4]   Krishna K.K.V., Satish P.K. (2013). Emotion Recognition in Speech Using MFCC and Wavelet Features. 3rd IEEE International Advance Computing Conference.

[5]   Albu F., Hagiescu D., and Puica M. (2014). Quality evaluation approaches of the first grade children's handwriting. The 10th International Scientific Conference eLearning and software for Education Bucharest, 17-23.

[6]   Kruskall, J. and M. Liberman, (1983) The Symmetric Time Warping Problem: From Continuous to Discrete", In Time Warps, String Edits and Macromolecules: The Theory and Practice of Sequence Comparison, pp. 125-161, Addison-Wesley Publishing Co., Reading, Massachusetts.

[7]   Deller J.R., Proakis J. G. and Hansen J. H.L., (1993) Discrete Time Processing of Speech Signals, Macmillan, New York.

[8]   Giannakopoulos, T., &Pikrakis, A. (2014). Introduction to Audio Analysis: A MATLAB® Approach. Academic Press.

[9]   Broomhead, D., & Lowe, D. (1988). Multivariable functional interpolation and adaptive networks. Complex Systems, 2, 321-355

[10]   Huang G., Huang G.-B., Song S., and You K. (2015). Trends in Extreme Learning Machines: A Review,"Neural Networks, vol. 61, no. 1, 32-48.

[11]   D.F. Specht, (1998). Probabilistic Neural Networks for Classification, Mapping, or Associative Memory. IEEE International Conference on Neural Networks, vol. I, 525-532.

[12] Vapnik V. (1995). The Nature of Statistical Learning Theory ed.; SpringerVerlag, New York.

[13] Yu C., Tian Q., Cheng F. and Zhang S. (2011). Speech Emotion Recognition Using Support Vector Machines. Advanced Research on Computer Science and Information Engineering. vol. 152, G. Shen and X. Huang, Eds., ed: Springer Berlin Heidelberg, 215–220.

[14] Hao T., Chu S.M., Hasegawa-Johnson M. and Huang T.S. (2009) Emotion recognition from speech VIA boosted Gaussian mixture models. IEEE ICME 2009, 294–297.

[15] Feraru M. and Zbancioc M. (2013) Speech emotion recognition for SROL database using weighted KNN algorithm. Electronics, Computers and Artificial Intelligence (ECAI), 1–4.

[16] Haykin, S., Neural Networks and Learning Machines (3rd Edition), 2008.

[17] https://chrisjmccormick.wordpress.com/2013/08/15/radial-basis-function-network-rbfn-tutorial/

[18] Albu F., Hagiescu D., Puica M., Vladutu L. (2015). Intelligent tutor for first grade children's handwriting application. 9th International Technology, Education and Development Conference, Madrid, Spain, 2- 4 March 2015, 3708 – 3717

[19] Wong P. K., Vong C. M., Gao X. H., and Wong K. I., (2014). Adaptive control using fully online sequential-extreme learning machine and a case study on engine air-fuel ratio regulation. Mathematical Problems in Engineering, vol. 2014, Article ID 246964, 11 pages.

[20] Han K., Yu D. and Tashev I. (2014) Speech Emotion Recognition Using Deep Neural Network and Extreme Learning Machine. Interspeech 2014, 223–227.

[21] Liang N.-Y., Huang G.-B., Saratchandran P., and Sundararajan N. (2006). A Fast and Accurate On-line Sequential Learning Algorithm for Feedforward Networks. IEEE Transactions on Neural Networks, vol. 17, no. 6, 1411-1423

[22] Steidl S. (2009). Automatic Classification of Emotion-Related User States in Spontaneous Children's Speech, Logos-Verlag.

[23] Batliner A., Steidl S., Hacker C., and Noth E. (2008). Private Emotions vs. Social Interaction – a Data-driven Approach towards Analysing Emotion in Speech. User Modeling and User-Updated Interaction, Vol. 18, No. 1-2, 175-206.

[24] Firoz S.A., Raj S.A. and Babu A.P. (2009). Automatic Emotion Recognition from Speech Using Artificial Neural Networks with Gender-Dependent Databases. Advances in Computing, Control and Telecommunication Technologies, ACT '09, 162–164.

[25] Chu H.C., Tsai W., Liao M.J., Cheng W.K., Chen Y.M. (2014). An E-Learning Model Featuring Facial Emotion Recognition and Regulation: Mathematic Learning of High-Functioning Autism Student as an Example, EDULEARN 2014.

[26] RapidMiner: Data Mining Use Cases and Business Analytics Applications (Chapman & Hall/CRC Data Mining and Knowledge Discovery Series) – October 25, 2013, by Markus Hofmann (Editor), Ralf Klinkenberg (Editor).

[27] Malik, H., Mishra, S. (2014). Feature selection using RapidMiner and classification through probabilistic neural network for fault diagnostics of power transformer. The 2014 Annual IEEE India Conference (INDICON), 1-6.

[28] Boser B. E., Guyon I. M. and Vapnik V. N., A training algorithm for optimal margin classifiers. Proceedings of the fifth annual workshop on Computational learning theory COLT '92, 144-152.

[29] http://www.kernel-machines.org/

[30] Barroso E., Santos G. and Proenca H. (2013). Facial Expressions: Discriminability of Facial Regions and Relationship to Biometrics Recognition".IEEE Workshop on Computational Intelligence in Biometrics and Identity Management (CIBIM), 77-80.

[31] Chang C.-C. and Lin C.-J. (2011). LIBSVM: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2 (3).

[32] Dalrymple K.A., Gomez J, and Duchaine B. (2013). The Dartmouth Database of Children's Faces: Acquisition and Validation of a New Face Stimulus Set", PLoS ONE 8(11).

[33] Cortes C. and Vapnik V. (1995). Support-vector networks. Machine Learning, 20(3), 273–297.

[34]   Hutter F., Hoos H. H. and Leyton-Brown K. (2011). Sequential Model-Based Optimization for General Algorithm Configuration. Learning and Intelligent Optimization, Lecture Notes in Computer Science Volume 6683, Springer, 507-523.

[35]   Norman, D.A. (1981). Twelve issues for cognitive science. Perspectives on Cognitive Science, Hillsdale, NJ: Erlbaum, 265–295.

[36]   McQuiggan, S. and Lester, J. (2006). Diagnosing self-efficacy in intelligent tutoring systems: an empirical study. M. Ikeda, K. Ashley and T.W. Chan (Eds.) Eighth International Conference on Intelligent Tutoring Systems, 565-574.

[37]   Woolf B., Burleson W., Arroyo I., Dragon T., Cooper D. and Picard R. (2009). Affect-aware tutors: recognising and responding to student affect. Int. J. Learning Technology, Vol. 4, No. 3/4, 129-164.

[38]   Zhang H., Gonen M., Yang Z., Oja E. (2015) Understanding Emotional Impact of Images Using Bayesian Multiple Kernel Learning. Neurocomputing, in press.