Acoustic Emotion Recognition: Two Ways of Features Selection Based on Self-Adaptive Multi-Objective Genetic Algorithm

Christina Brester[1], Maxim Sidorov[2], Eugene Semenkin[3]
[1;3] *Institute of Computer Sciences and Telecommunication, Siberian State Aerospace University, Krasnoyarsk, Russia*
[2] *Institute of Communications Engineering, University of Ulm, Germany*
[1]*abahachy@mail.ru,* [2]*maxim.sidorov@uni-ulm.de,* [3]*eugenesemenkin@yandex.ru*

Keywords: Heuristic feature selection, multi-objective genetic algorithm, self-adaptation, probabilistic neural network, speech-based emotion recognition.

Abstract: In this paper the efficiency of feature selection techniques based on the evolutionary multi-objective optimization algorithm is investigated on the set of speech-based emotion recognition problems (English, German languages). Benefits of developed algorithmic schemes are demonstrated compared with Principal Component Analysis for the involved databases. Presented approaches allow not only to reduce the amount of features used by a classifier but also to improve its performance. According to the obtained results, the usage of proposed techniques might lead to increasing the emotion recognition accuracy by up to 29.37% relative improvement and reducing the number of features from 384 to 64.8 for some of the corpora.

1 INTRODUCTION

While solving classification problems it is reasonable to perform data preprocessing procedures to expose irrelevant attributes. Features might have a low variation level, correlate with each other or be measured with mistakes that lead to a deterioration in the performance of the learning algorithm.

If standard techniques (such as Principal Component Analysis (PCA)) do not demonstrate sufficient effectiveness, alternative algorithmic schemes based on heuristic optimization might be applied.

In this paper we consider two approaches for feature selection: according to the first one, the relevancy of extracted attributes is evaluated with a classifier; the second one is referred to the data preprocessing stage and engages various statistical metrics which require fewer computational resources to be assessed. In both cases Probabilistic Neural Network (PNN) is used as a supervised learning algorithm (Specht, 1990).

We investigate the efficiency of the introduced algorithmic schemes on the set of emotion recognition problems which reflect one of the crucial questions in the sphere of human-machine communications. Nowadays program systems processing voice records and extracting acoustic characteristics are becoming more widespread (Boersma, 2002), (Eyben *et al*., 2010). However, the number of features obtained from the speech signal might be overwhelming and due to the reasons mentioned above it is not rational to involve all of this data in the classification process. Therefore it is vitally important to determine the optimal feature set used by a learning algorithm to recognize human emotions.

2 MODELS FOR FEATURE SELECTION

2.1 Wrapper and filter approaches

In (Kohavi, 1997) basic algorithmic schemes for feature selection are presented.

The *wrapper* approach is a combination of an optimization algorithm and a classifier that is used to estimate the quality of the selected feature set. In this study we propose a multi-objective optimization procedure operating with two criteria which are the relative classification error (assessed on the set of validation examples) and the number of selected features; both criteria should be minimized. The usage of these criteria allows not only to improve the performance of involved classifiers but also to reduce the amount of data required for training. The scheme for this approach is shown in Figure 1.

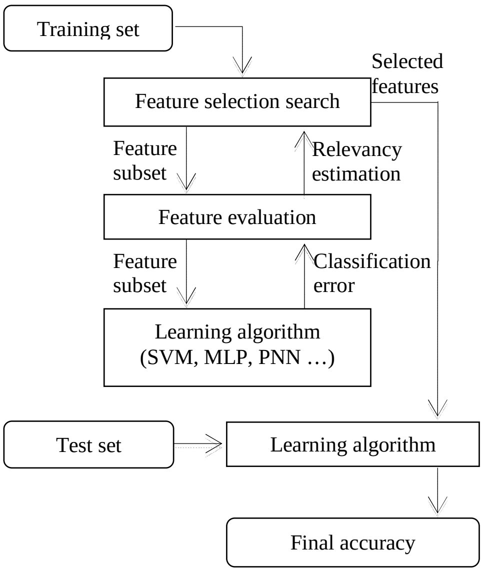

Figure 1: The wrapper approach.

Feature selection with the *filter* approach is based on estimating statistical metrics such as *Attribute Class Correlation, Inter- and Intra- Class Distances*, *Laplasian Score*, *Representation Entropy and the Inconsistent Example Pair measure* (Venkatadri and Srinivasa, 2010) which characterize the data set quality. In that case we also introduce the two-criteria model, specifically, the Intra-class distance (IA) and the Inter-class distance (IE) are used as optimized criteria:

$$IA = \frac{1}{n}\sum_{r=1}^{k}\sum_{j=1}^{n_r} d(p_j^r, p_r) \to min, \quad (1)$$

$$IE = \frac{1}{n}\sum_{r=1}^{k} n_r d(p_r, p) \to max, \quad (2)$$

where p_j^r is the *j*-th example from the *r*-th class, p is the central example of the data set, $d(...,...)$ denotes the Euclidian distance, p_r and n_r represent the central example and the number of examples in the *r*-th class.

The scheme of the filter method is shown in Figure 2.

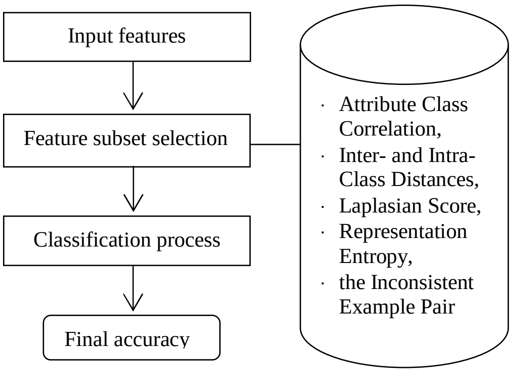
Input features
Feature subset selection
Classification process
Final accuracy
· Attribute Class Correlation,
· Inter- and Intra-Class Distances,
· Laplasian Score,
· Representation Entropy,
· the Inconsistent Example Pair

Figure 2: The filter approach.

As a feature selection technique we use a multi-objective genetic algorithm (MOGA) operating with binary strings, where *unit* and *zero* correspond to a relative attribute and an irrelative one respectively. Moreover, to avoid choosing the algorithm settings it is reasonable to apply the self-adaptive modification of MOGA (Eiben *et al.*, 1999).

2.2 Self-adaptive Strength Pareto Evolutionary Algorithm

The search for the optimal feature set from the database was realized through involving a multi-criteria evolution procedure. We modified the Strength Pareto Evolutionary Algorithm (SPEA) (Zitzler and Thiele, 1999) using the self-adaptation idea. The proposed approach works as follows:

Inputs:

- N : the population size;
- $\overline{N}$: the maximum number of non-dominated points stored in the outer set;
- M : the maximum number of generations.

Parameters of the self-adaptive crossover operator:

- *«penalty»*: a fee size for recombination types defeated in paired comparisons;
- *«time of adaptation»* T : the number of generations fulfilled before every reallocation of resources among recombination types;
- *«social card»*: the minimum allowable size of the subpopulation generated with a crossover operator type;
- available recombination types: $J = \{$ 0 | «single-point crossover»; 1 | «two-point crossover»; 2 | «uniform crossover» $\}$;

- n_j, $j \in J$: the amount of individuals in the current population generated by the *j*-th type of crossover.

Outputs:

- $PS = \overline{P} = \{ \overline{x}_i \}, 1 \le i \le \overline{N}$: the approximation of the Pareto set;
- PF : the approximation of the Pareto front.

Step 1. Initialization

Generate an initial population $P_t = \{ x_i \},\ t = 0,$ $i = \overline{1, N}$, uniformly in the binary search space: probabilities of boolean *true* and *false* assignments are equal. Define initial values $n_j = \dfrac{N}{|J|}$.

Step 2. Evaluation of criteria values

For each individual from P_t, do:

2.1. Compile the feature subsystem from the database corresponding to the current binary string.

2.2. Estimate criteria values for all individuals from the current population.

Step 3. Composing the outer set

3.1. Copy the individuals non-dominated over P_t into the intermediate outer set $\overline{P}'$.

3.2. Delete the individuals dominated over $\overline{P}'$ from the intermediate outer set.

3.3. If the capacity of the set $\overline{P}'$ is more than the fixed limit $\overline{N}$, apply the clustering algorithm (hierarchical agglomerative clustering).

3.4. Compile the outer set $\overline{P}_{t+1}$ with the individuals from $\overline{P}'$.

Step 4. Fitness-values determination

Calculate fitness-values for individuals both from the current population and from the outer set.

Step 5. Generation of new solutions

Set $j = 0$. For each $j-$realized recombination type, $j \in J$, do:

1) Set $k = 0$ and repeat:
2) Select two individuals from the united set $\widehat{P} = \overline{P}_{t+1} \cup P_t$ by 2-tournament selection.
3) Apply the current type of recombination to individuals chosen in step (2).
4) Perform a mutation operator: the probability p_m is determined according to the rule (Daridi et al., 2004):

$$p_m = \frac{1}{240} + \frac{0.11375}{2^t}, \tag{3}$$

where t is the current generation number.

5) If $k = n_j$, then $j = j + 1$, otherwise $k = k + 1$.

Step 6. Stopping Criterion

If $t = M$, then stop with the outcome $PS = \overline{P}_{t+1}$, otherwise $t = t + 1$.

Step 7. Resources reallocation

If t is multiple to T, do:

7.1. Determine «*fitness*»-values q_j for all $j \in J$:

$$q_j = \sum_{l=0}^{T-1} \frac{T - l}{l + 1} \cdot b_j, \tag{4}$$

where $l = 0$ corresponds to the latest generation in the adaptation interval, $l = 1$ corresponds to the previous generation, etc. b_j is defined as following:

$$b_j = \frac{p_j}{|\overline{P}|} \cdot \frac{N}{n_j}, \tag{5}$$

where p_j is the amount of individuals in the current outer set generated by the *j*-th type of recombination operator, $|\overline{P}|$ is the outer set size.

7.2. Compare all crossover operator types in pairs based on their «*fitness*»-values. Determine s_j to be the size of a resource given by the *j*-th recombination type to those which won:

$$s_j = \begin{cases} 0, if\ n_j \le social\_card \\ int(\dfrac{n_j - social\_card}{n_j}), if \begin{array}{l} (n_j - h_j \cdot penalty) \le \\ \le social\_card, \end{array} \\ penalty, otherwise, \end{cases} \tag{6}$$

where h_j is the number of losses of the *j*-th operator in paired comparisons.

Table 1: Databases description

Database	Language	Full length (min.)	Number of emotions	File level duration		Emotion level duration		Notes
				Mean (sec.)	Std. (sec.)	Mean (sec.)	Std. (sec.)	
Berlin	German	24.7	7	2.7	1.02	212.4	64.8	Acted
SAVEE	English	30.7	7	3.8	1.07	263.2	76.3	Acted
VAM	German	47.8	4	3.02	2.1	717.1	726.3	Non-acted

7.3. Redistribute resources n_j based on s_j values, $j \in J$.

Go to **Step 2**.

In Steps 2 and 5 standard SPEA schemes of the fitness assignment and selection are used.

3 PERFORMANCE ASSESSMENT

3.1 Corpora description

In the study a number of speech databases have been used and this section provides their brief description.

Berlin The Berlin emotional database (Burkhardt et al., 2005) was recorded at the Technical University of Berlin and consists of labeled emotional German utterances which were spoken by 10 actors (5 female). Each utterance has one of the following emotional labels: neutral, anger, fear, joy, sadness, boredom or disgust.

SAVEE The SAVEE (Surrey Audio-Visual Expressed Emotion) corpus (Haq and Jackson, 2010) was recorded as a part of an investigation into audio-visual emotion classification, from four native English male speakers. The emotional label for each utterance is one of the standard set of emotions (anger, disgust, fear, happiness, sadness, surprise and neutral).

VAM The VAM database (Grimm et al., 2008) was created at the Karlsruhe University and consists of utterances extracted from the popular German talk-show ”Vera am Mittag” (Vera in the afternoon). The emotional labels of the first part of the corpus (speakers 1-19) were given by 17 human evaluators and the rest of the utterances (speakers 20-47) were labeled by 6 annotators on a 3-dimensional emotional basis (valence, activation and dominance). To produce the labels for the classification task we have used just a valence (or evaluation) and an arousal axis. The corresponding quadrant (counterclockwise, starting in the positive quadrant, and assuming arousal as abscissa) can also be assigned emotional labels: happy-exciting, angry-anxious, sad-bored and relaxed-serene.

Two corpora (Berlin, SAVEE) consist of acted emotions, whereas VAM database comprises real ones. Acted and non-acted emotions have been considered for the German language, but there are only non-acted emotions in English utterances.

In comparison with Berlin and SAVEE corpora, the VAM database is highly unbalanced (see Emotion level duration columns in Table 1).

Emotions themselves and their evaluations have a subjective nature. That is why it is important to have at least several evaluators of emotional labels. Even for humans it is not always obvious which decision to make about an emotional label. Each study, which proposes an emotional database, also provides an evaluators’ confusion matrix and a statistical description of their decisions.

There is a statistical description of the used corpora in Table 1.

3.2 Experiments and results

To assess the efficiency of the developed approaches we compared the PNN-classifier performance on the extended data sets comprised of 384-dimensional feature vectors, the PCA-PNN and the MOGA-PNN system execution on the reduced set of attributes (Table 2).

For every experiment the classification procedure was run 20 times. The data set was randomly divided into training and test samples in a proportion of 70-30%. In all experiments MOGAs were provided with an equal amount of resources (for each run 10100 candidate solutions were examined in the search space). The final solution was determined from the set of non-dominated candidates as the point with the lowest error on the

Table 2: Experimental results

Method	Relative classification accuracy, %		
	Berlin	SAVEE	VAM
PNN	58.9 (384)	47.3 (384)	67.1 (384)
PCA-PNN	43.7 (129.3)	26.5 (123.6)	59.4 (148.6)
SPEA_wrapper-PNN	71.5 (68.4)	48.4 (84.1)	70.6 (64.8)
SPEA_filter-PNN	76.2 (138.6)	60.8 (142.0)	73.2 (152.8)

validation sample (20% of the training data set).

Table 2 contains the relative classification accuracy for the presented corpora. In parentheses there is the average number of selected features.

4 CONCLUSION

This study reveals advantages of using the MOGA-PNN combination in feature selection by the example of the speech-based emotion recognition problem. The developed algorithmic schemes allow not only a reduction in the number of features used by the PNN-classifier (from 384 up to 64.8) but also an improvement in its performance up to 29.4% (from 58.9% on the full data set to 76.2% on the set of selected features).

It was found that the filter approach permits the achievement of better results in the sense of the classification accuracy, whereas the usage of the wrapper approach decreases the number of features significantly.

Moreover, we compared these heuristic procedures with conventional PCA with the 0.95 variance threshold. According to the obtained results, the heuristic search for feature selection in the emotion recognition problem is much more effective than the application of the PCA-based technique that leads to a decrease in the classification accuracy.

Although the PNN accuracy is rather high, there is an opportunity to investigate the self-adaptive multi-objective genetic algorithm hybridized with more accurate classifiers. Besides, other effective MOGAs might be used to improve the performance of the proposed technique.

REFERENCES

Boersma, P., 2002. Praat, a system for doing phonetics by computer. *Glot international*, 5(9/10): pp. 341–345.

Burkhardt, F., Paeschke, A., Rolfes M., Sendlmeier, W. F., Weiss, B., 2005. A database of german emotional speech. *In Interspeech*, pp. 1517–1520.

Daridi, F., Kharma N., Salik, J., 2004. Parameterless genetic algorithms: review and innovation. *IEEE Canadian Review*, (47): pp. 19–23.

Eiben A.E., Hinterding R., and Michalewicz Z., 1999. Parameter control in evolutionary algorithms. *IEEE Transactions on Evolutionary Computation*, 3(2): pp. 124–141.

Eyben, F., Wöllmer, M., and Schuller, B., 2010. Opensmile: the Munich versatile and fast opensource audio feature extractor. In *Proceedings of the international conference on Multimedia*, pp. 1459–1462. ACM.

Grimm, M., Kroschel, K., Narayanan, S., 2008. The vera am mittag german audio-visual emotional speech database. In Multimedia and Expo, 2008 *IEEE International Conference on*, 865–868. IEEE.

Haq, S., Jackson, P., 2010. *Machine Audition: Principles, Algorithms and Systems*, chapter Multimodal Emotion Recognition, 398–423. IGI Global, Hershey PA.

Kohavi, R., John, G.H., 1997. Wrappers for feature subset selection. *Artificial Intelligence*, 97: pp. 273-324.

Specht, D.F., 1990. Probabilistic neural networks. *Neural networks, 3(1):* pp. 109–118.

Venkatadri, M., Srinivasa Rao, K., 2010. A multiobjective genetic algorithm for feature selection in data mining. *International Journal of Computer Science and Information Technologies, vol. 1, no. 5*, pp. 443–448.

Zitzler, E., Thiele, L., 1999. Multiobjective evolutionary algorithms: A comparative case study and the strength pareto approach, Evolutionary Computation, IEEE Transactions on, vol. 3, no. 4, pp. 257–271.