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.

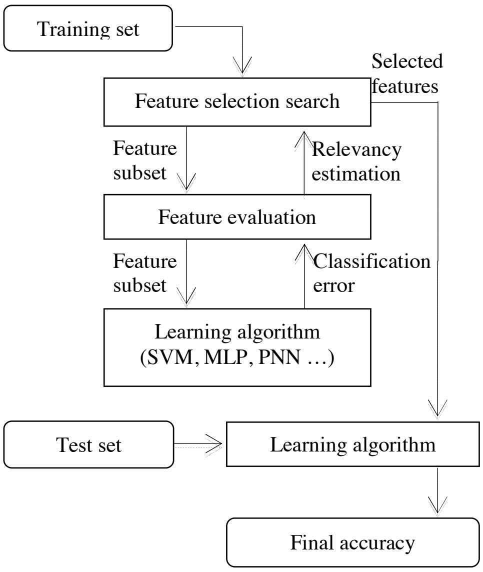

Figure 1: The wrapper approach.

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:

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

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

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.

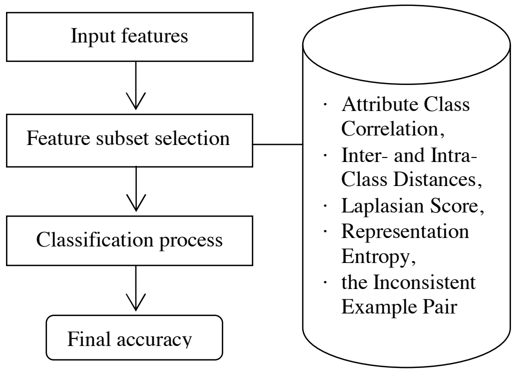

Figure 2: The filter approach.

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» $\}$;