Evolutionary feature selection for emotion recognition in multilingual speech analysis

Christina Brester[1], Eugene Semenkin[2], Igor Kovalev[3], Pavel Zelenkov[4]
Institute of Computer Science and Telecommunications
Siberian State Aerospace University
Krasnoyarsk, Russia
[1]christina.bre@yandex.ru, [2]eugenesemenkin@yandex.ru, [3]kovalev.fsu@mail.ru, [4]zelenkow@rambler.ru

Maxim Sidorov
Institute of Communication Engineering
Ulm University
Ulm, Germany
maxim.sidorov@uni-ulm.de

***Abstract*—In the case when conventional feature selection methods do not demonstrate sufficient performance, alternative algorithmic schemes might be applied. In this paper we propose an evolutionary feature selection technique based on the two-criteria optimization model. To diminish the drawbacks of genetic algorithms, which are used as optimizers, we design a parallel multi-criteria heuristic procedure based on an island model. The effectiveness of the proposed approach was investigated on the Speech-based Emotion Recognition Problem, which reflects one of the crucial aspects in the sphere of human-machine communications. A number of multilingual corpora (German, English and Japanese) were engaged in the experiments. According to the results obtained, a high level of emotion recognition was achieved (up to a 11.15% relative improvement compared with the best F-score value on the full set of attributes).**

Keywords—feature selection, multi-objective genetic algorithm, island model, emotion recognition

I. INTRODUCTION

In recent years there has been a growing interest in the sphere of *Evolutionary Machine Learning*. However, some researchers highlight the negative sides of the *Evolutionary Computation* and *Machine Learning* integration. Firstly, it is always necessary to investigate a number of algorithms to define the most effective one for the problem considered because the performance of evolutionary algorithms varies significantly for different problems. Secondly, these methods require more computational resources compared with alternative non-evolutionary algorithms. Therefore, in this study we attempt to develop a feature selection technique for classification problems based on a genetic algorithm with these drawbacks removed.

Generally, the feature selection procedure can be organized as the *wrapper* approach or the *filter* one [1]. The first technique involves classification models to evaluate the relevancy of each feature subset. Although it requires high computational resources, this approach demonstrates adjustment to an applied classifier. The second technique is referred to the pre-processing stage because it extracts information from the data set and reduces the number of attributes, taking into consideration such measures as consistency, dependency, and distance. This approach needs significantly fewer calculations therefore it is rather effective in the sense of computational effort. On the one hand, the filter attribute selection procedure does not cooperate with a learning algorithm and so ignores its performance entirely. However, on the other hand, it might be effectively used in combination with an ensemble of diverse classifiers, which is quite reasonable in the case when one does not know one particular reliable and effective model. Therefore, in this paper we propose the evolutionary feature selection procedure which corresponds to the filter scheme.

We designed the attribute selection approach as a two-criteria optimization model and applied a modified multi-objective genetic algorithm to find solutions. To overcome the disadvantages of the evolutionary search, an island model is used to involve genetic algorithms which are based on different concepts. Moreover, this model allowed us to parallelize calculations and, consequently, to reduce the computational time.

The effectiveness of the proposed approach has been investigated on the *Speech-based Emotion Recognition Problem* which reflects one of the crucial questions in the sphere of human-machine communications [2]. In the experiments conducted a number of multilingual databases (English, German, and Japanese) are used.

In the previous research it was found that there was no classification model which demonstrated the highest performance for all of the corpora [3]. Therefore, we combined the developed filter technique with the ensemble of classifiers (Multilayer Perceptron, Support Vector Machine, and Linear Logistic Regression) which showed high effectiveness separately. The classification results obtained after the application of the developed pre-processing method were compared with the classification quality on the full databases (without feature selection) and after the application of Principal Component Analysis (PCA).

The rest of the paper is organized as follows: in Section II a brief description of the related work is presented, Section III contains the details of the evolutionary feature selection scheme. The speech-based emotion recognition problem and the corpora used are introduced in Section IV. The experiments

Research is performed with the financial support of the Ministry of Education and Science of the Russian Federation within the federal R&D programme (project RFMEFI57414X0037).

978-1-4799-7492-4/15/$31.00 ©2015 IEEE

conducted, the results obtained, and the main inferences are included in Section V. The conclusion and future work are presented in Section VI.

II. Related Work

Yang and Hanovar (1998) used a one-criterion genetic algorithm (GA) to determine relevant attributes in order to improve the quality of classification realized with neural networks [4]. Li Zhuo *et al.* (2008) accomplished the classification of hyperspectral images with a support vector machine; they also engaged a one-criterion GA to remove non-informative features [5]. In both cases the feature selection procedure was combined with supervised learning algorithms based on the wrapper approach scheme.

Lanzi (1997) offered to apply a heuristic method to extract attributes before executing the classification [6]. The inconsistency rate was used by a GA to assess the relevancy of reduced data sets. Due to the implementation of the filter approach it became possible not only to achieve the high performance of the C4.5 inductive algorithm but also to lower the computational cost.

Development of multi-objective optimization algorithms allowed researchers to embed these methods in the feature selection procedure to take into account several criteria. Venkatadri and Srinivasa (2010) introduced a set of measures such as *Attribute Class Correlation, Inter- and Intra-Class Distances, Laplasian Score, Representation Entropy* and *the Inconsistent Example Pair measure* to estimate the quality of reduced databases. They investigated various combinations of these criteria by means of the Non-dominated Sorting Genetic Algorithm (NSGA-II) [7]. Hamdani *et al.* (2007) also implemented NSGA-II to attain a compromise between the number of extracted attributes and the classification accuracy evaluated with the 1-NN classifier [8]. These are examples of MOGA realization in the framework of the filter and the wrapper approach respectively.

III. Evolutionary Feature Selection Scheme

A. Two-criteria Filter 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* which characterize the data set relevancy. In this 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 1.

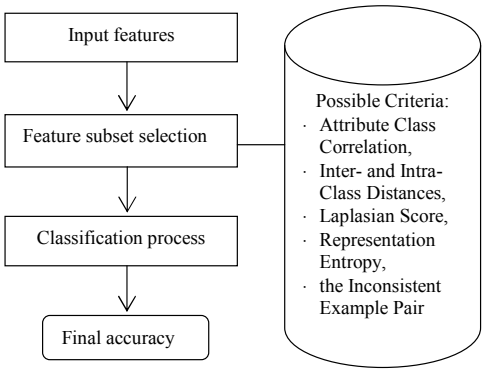

Fig. 1. General filter approach scheme.

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.

B. Multi-objective Genetic Algorithms

The common scheme of any MOGA includes the same steps as any conventional one-criterion GA:

Generate the initial population
Evaluate criteria values
While (stop-criterion!=true), do:
{Estimate fitness-values;
Choose the most appropriate individuals with the mating selection operator based on their fitness-values;
Produce new candidate solutions with recombination;
Modify the obtained individuals with mutation;
Compose the new population (environmental selection);
}

In contrast to one-criterion GAs, the outcome of MOGAs is the set of non-dominated points which form the Pareto set approximation.

Designing a MOGA, researchers are faced with some issues which are referred to fitness assignment strategies, diversity preservation techniques, and ways of elitism implementation. Therefore, in this study we investigate the effectiveness of MOGAs, which are based on various heuristic mechanisms, from the perspective of the feature selection procedure.

As we have noticed, MOGAs return the set of candidate-solutions which cannot be preferred to each other. Taking into account this fact, we have proposed a way to derive the final solution based on the set of non-dominated points. It is assumed that the outcome of the MOGA is N binary strings (the set of non-dominated solutions). Each chromosome should be decoded to the database reduced, according the rule: if a