Multicriteria Neural Network Design in the Speech-based Emotion Recognition Problem

Christina Brester[1], Eugene Semenkin[2], Maxim Sidorov[3], Olga Semenkina[4]
[1,2,4] *Institute of Computer Sciences and Telecommunication, Siberian State Aerospace University, Krasnoyarsk, Russia*
[3] *Institute of Communications Engineering, University of Ulm, Germany*
[1]*christina.brester@gmail.com,* [2]*eugenesemenkin@yandex.ru,* [3]*maxim.sidorov@uni-ulm.de,* [4]*semenkina.olga@mail.ru*

Keywords: Neural network, multicriteria design, cooperative genetic algorithm, speech-based emotion recognition, feature selection.

Abstract: In this paper we introduce the two-criterion optimization model to design multilayer perceptrons taking into account two objectives, which are the classification accuracy and computational complexity. Using this technique, it is possible to simplify the structure of neural network classifiers and at the same time to keep high classification accuracy. The main benefits of the approach proposed are related to the automatic choice of activation functions, the possibility of generating the ensemble of classifiers, and the embedded feature selection procedure. The cooperative multi-objective genetic algorithm is used as an optimizer to determine the Pareto set approximation in the two-criterion problem. The effectiveness of this approach is investigated on the speech-based emotion recognition problem. According to the results obtained, the usage of the proposed technique might lead to the generation of classifiers comprised by fewer neurons in the input and hidden layers, in contrast to conventional models, and to an increase in the emotion recognition accuracy by up to a 4.25% relative improvement due to the application of the ensemble of classifiers.

1 INTRODUCTION

The sphere of human-machine interactions is closely related to *affective computing* (Picard, 1995), which is the interdisciplinary domain including algorithms, systems and devices aimed at recognizing, processing, and simulating human emotions. In most cases all these techniques engage video or audio data to analyse users' emotions. Also there are multi-modal systems which fuse visual information and acoustic characteristics extracted from speech signals. However, in this paper we consider the human emotion recognition problem in the framework of intellectual spoken dialogue systems and, therefore, we apply only audio data.

Previously it was found that compared with various classification models neural networks showed rather high effectiveness for the speech-based emotion recognition problem (Brester *et al.*, 2014). In the experiments conducted a multilayer perceptron (MLP) with one hidden layer trained by the error backpropagation algorithm (BP) was used.

Conventionally, the number of neurons in the hidden layer is proportional to the amount of classes in the sample and the dimensionality of the feature vector. In the case of the emotion recognition problem the quantity of input attributes is very large: generally, we extract 384 acoustic characteristics from the speech signal. As a result, the MLP structure is exaggerated and contains too many neurons in its hidden layer. Moreover, while designing MLPs, researchers have to choose the activation function for each neuron, which is not a trivial task. By default a sigmoid is widely used, despite the fact that there are a lot of other activation functions which are easier in the sense of computational complexity and at the same time might be effectively applied without detriment to the recognition accuracy.

Taking into account these points, we decided to improve the MLP performance by optimizing its structure. In this study we propose a two-objective optimization model which allows us to generate appropriate MLPs based on two criteria: the *classification accuracy* and *computational complexity*. Using this strategy, it is possible to design the MLP whose performance is comparable with the accuracy of the conventional model and whose structure is optimal in the sense of computational complexity. The main advantages of the approach proposed also include the automatic choice of activation functions, the embedded feature selection procedure, and the option of generating the ensemble of classifiers.

The rest of the paper is organized as follows: in Section II a description of the two-objective model for the neural network design and the cooperative genetic algorithm, which is applied to optimize the criteria introduced, are presented. In Section III there is a definition of the speech-based emotion recognition problem and the corpora used. The experiments conducted, the results obtained, and the main inferences are included in Section IV. The conclusions and future work are presented in Section V.

2 PROPOSED APPROACH

2.1 Multicriteria Optimization Model for Neural Network Design

In this study we propose the two-criterion optimization model for neural network design, specifically, for the automatic generation of MLPs with one hidden layer. By taking into account two objectives, it is possible to attain a trade-off between the classification accuracy and computational complexity.

Criterion 1. The relative classification error:

$$minimize{:}\ K1 = E = \frac{N_{incorrectly}}{N_{all}}, \quad (1)$$

where $N_{incorrectly}$ is the number of instances classified incorrectly, N_{all} is the common number of instances.

Criterion 2. Computational complexity:

$$minimize{:}\ K2 = N_{weights} + \sum_{j=1}^{N_{neurons}} K_j(i), \quad (2)$$

where $K_j(i) = \frac{T_i^{act}}{T^{weight}}$ is the coefficient reflecting the relative computational complexity of evaluating the *i*-th activation function of the *j*-th neuron; *i* is the identification number of the activation function in the finite set comprised by alternative variants of activation functions; T_i^{act} is the time spent on evaluating the *i*-th activation function; T^{weight} is the time required to process one connection; $N_{weights}$ is the number of connections in the MLP; $N_{neurons}$ is the number of neurons in the MLP. T_i^{act} and T^{weight} are assessed empirically. It is essential to note that K_i is independent of the software used because T_i^{act} is normalized by T^{weight}.

To solve this two-criterion problem, we suggest applying a multi-objective genetic algorithm (MOGA), which operates with binary strings coding diverse MLP structures. Each candidate solution, called a *chromosome*, contains identification numbers of all neurons from the hidden layer (Figure 1). Zero corresponds to the absence of neurons. Input parameters include the set of activation functions with their ID-numbers and the maximum number of neurons in the hidden layer.

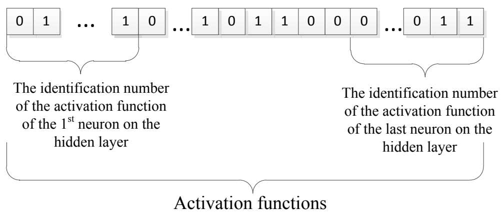

Figure 1: The presentation of the MLP structure as a binary string.

The backpropagation algorithm is applied to train MLPs with different numbers of neurons in the hidden layer and estimate the criterion $K1$.

Moreover, we propose using the cooperative multicriteria genetic algorithm as a multi-objective optimization procedure to diminish the drawbacks of the evolutionary search (Brester *et al.*, 2015a). The next section contains a concise description of this heuristic multi-agent procedure and its advantages.

2.2 Cooperative Multi-objective Heuristic Procedure

While designing a MOGA, researchers are faced with some issues which are related to fitness assignment strategies, diversity preservation techniques, and ways of elitism implementation (Zitzler *et al.*, 2004). To eliminate a number of problems which arise while designing multicriteria evolutionary methods, in this study we use a cooperation of several genetic algorithms (GA) based on various heuristic mechanisms. An island model is applied to involve a few GAs which realize different concepts. Moreover, this model allows us to parallelize calculations and, consequently, to reduce computational time.