SIBERIAN STATE AEROSPACE UNIVERSITY
Krasnoyarsk. Russia




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

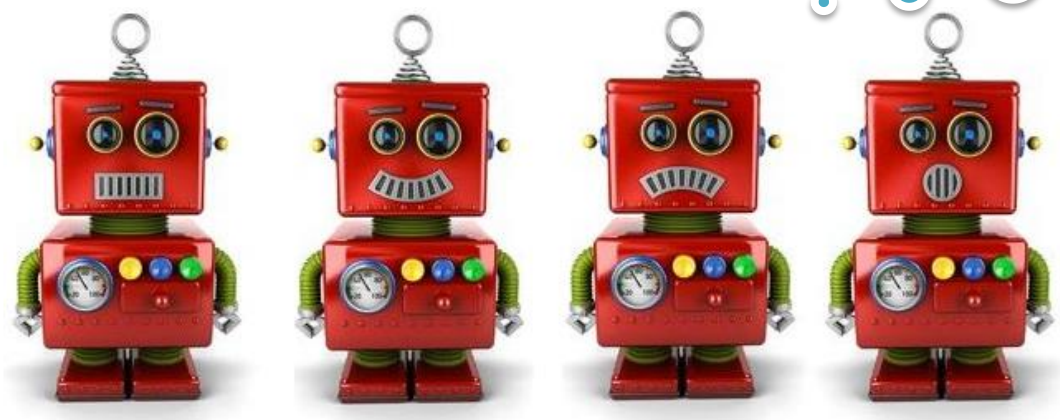
Christina Brester, Eugene Semenkin, Maxim Sidorov, and Olga Semenkina




Speech-based Emotion Recognition Problem

- Spoken Dialogue Systems Improvement
- Robotics
- Call-centers quality monitoring

... etc.



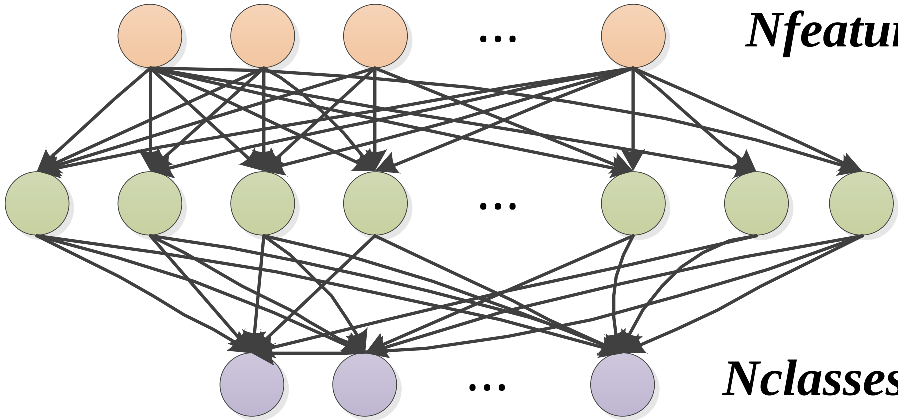
Why do we talk about *Neural Network Design*?

The most effective classifiers are:

- Multilayer Perceptron (MLP)
- Support Vector Machine (SVM)
- Logistic Regression

Conventional MLP

- Structure:

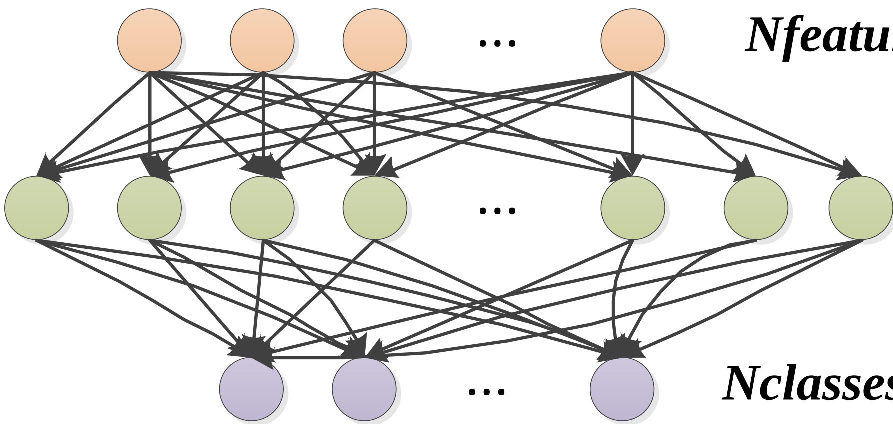


- Activation function:

$$sigmoid\quad f(x) = \frac{1}{1+e^{-x}}$$

Conventional MLP

- Structure:



Nfeatures = 384

Nneurons=(Nfeatures+Nclasses)/2 +1

Nclasses

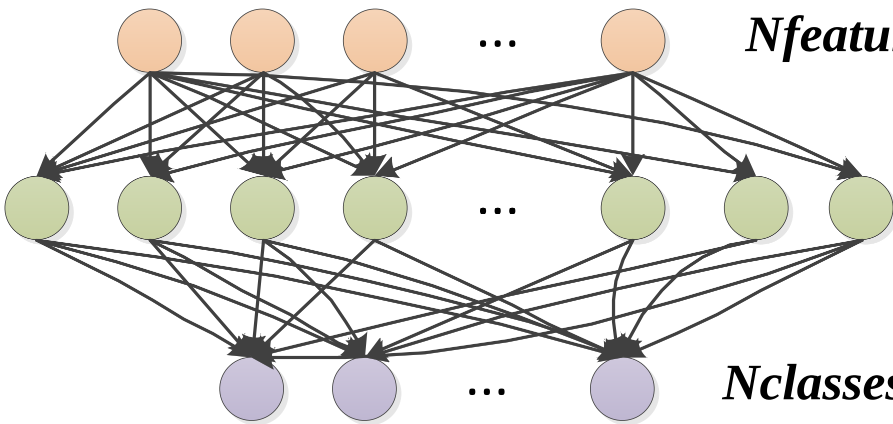
! This structure is exaggerated

- Activation function:

$$sigmoid \quad f(x) = \frac{1}{1+e^{-x}}$$

Conventional MLP

- Structure:



Nfeatures = 384

Nneurons=(Nfeatures+Nclasses)/2 +1

Nclasses

! This structure is exaggerated

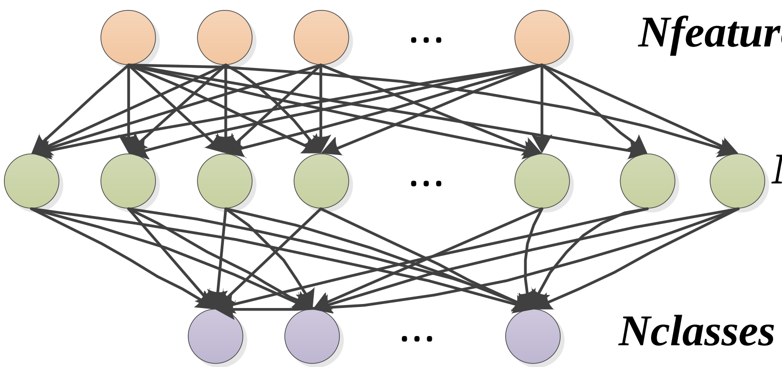
- Activation function:

$$sigmoid \quad f(x) = \frac{1}{1+e^{-x}}$$

! There are a lot of other activation functions which are easier in the sense of computational complexity

Conventional MLP

- Structure:



- Activation function:

$$sigmoid \quad f(x) = \frac{1}{1+e^{-x}}$$

! Our proposal:
A two-objective optimization model for generating MLPs based on two criteria: the ***classification accuracy*** and ***computational complexity***.

Advantages:
- the automatic choice of activation functions,
- the embedded feature selection procedure,
- the option of generating the ensemble of classifiers.

Outline

- **Motivation**
  - Speech-based Emotion Recognition Problem
  - Conventional MLP (cons)
- **Proposed Approach**
  - Two-criterion Optimization Model for Neural Network Design
  - Cooperative Multi-objective Heuristic Procedure
- **Problem Definition**
  - Speech-based Emotion Recognition Problem (in detail)
  - Corpora Description
- **Results and Discussion**
  - Conventional MLPs
  - Automatically designed MLPs and their Ensembles
  - Embedded Feature Selection Procedure
- **Conclusion and Future Plans**

Two-criterion Optimization Model for Neural Network Design

$$\begin{cases} \text{Criterion 1. } \textbf{The relative classification error} \rightarrow \mathit{min} \\ \text{Criterion 2. } \textbf{Computational complexity} \rightarrow \mathit{min} \end{cases}$$

Criterion 1. **The relative classification error**:

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

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

Two-criterion Optimization Model for Neural Network Design

$$\begin{cases} \text{Criterion 1. } \textbf{The relative classification error} \rightarrow min \\ \text{Criterion 2. } \textbf{Computational complexity} \rightarrow min \end{cases}$$

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.

Two-criterion Optimization Model for Neural Network Design

$$\begin{cases} \text{Criterion 1. } \textbf{The relative classification error} \rightarrow min \\ \text{Criterion 2. } \textbf{Computational complexity} \rightarrow min \end{cases}$$

Criterion 2. **Computational complexity:**

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

where $K_j(i) = \dfrac{T_i^{act}}{T^{weight}}$ 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.

Two-criterion Optimization Model for Neural Network Design

- **A multi-objective genetic algorithm (MOGA)** is to solve the two-criterion problem;
- **The backpropagation algorithm** is applied to train MLPs with different numbers of neurons in the hidden layer and estimate the Criterion 1.
- **Input parameters** include:
  - the set of activation functions with their ID-numbers;
  - the maximum number of neurons in the hidden layer.

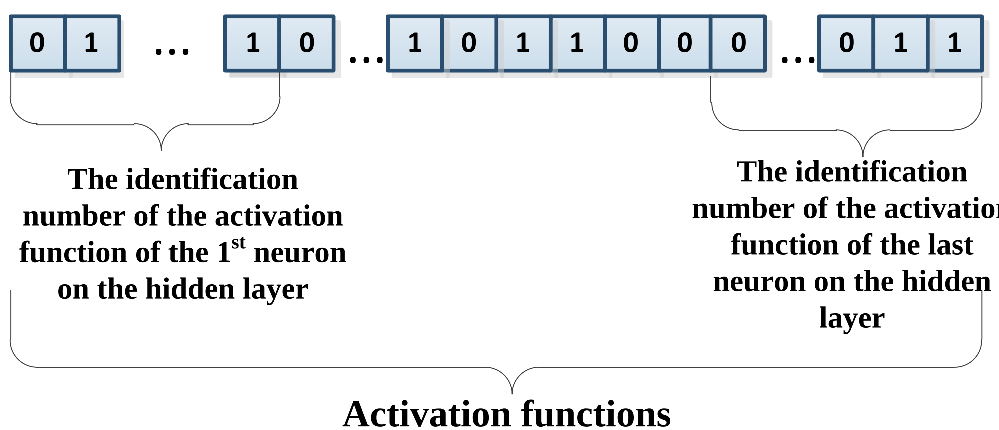


Figure 1. The presentation of the MLP structure as a binary string

Multi-Objective Genetic Algorithms (MOGAs)

- *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**);*

}

Multi-Objective Genetic Algorithms

Designing a MOGA, researchers are faced with some issues:

- fitness assignment strategies,
- diversity preservation techniques,
- ways of elitism implementation.

➢ Possible solution: Cooperation of genetic algorithms which are based on different concepts

Basic features of the MOGA used

MOGA	Fitness Assignment	Diversity Preservation	Elitism
NSGA-II	Pareto-dominance (niching mechanism) and diversity estimation (crowding distance)	Crowding distance	Combination of the previous population and the offspring
PICEA-g	Pareto-dominance (with generating goal vectors)	Nearest neighbour technique	The archive set and combination of the previous population and the offspring
SPEA2	Pareto-dominance (niching mechanism) and density estimation (the distance to the k-th nearest neighbour in the objective space)	Nearest neighbour technique	The archive set

Cooperative Multi-objective Heuristic Procedure

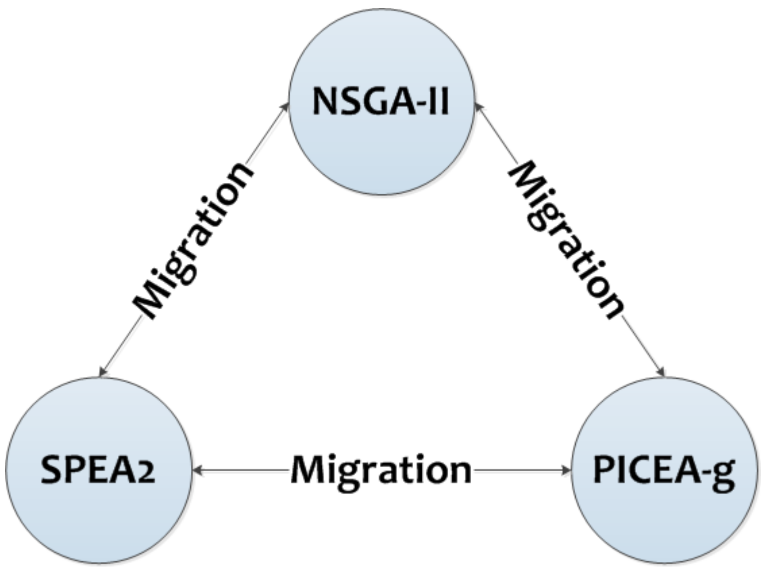


Figure 2. The island model

Island model …

- ✓ is based on parallel work of islands;
- ✓ has an ability to preserve genetic diversity;
- ✓ could be applied to separable problems.

At each *T*-th generation algorithms exchange the best solutions (***migration***).

There are two parameters:

migration size, the number of candidates for migration;

migration interval, the number of generations between migrations.

Speech-based Emotion Recognition Problem

List of extracted features

- General features: Power, Mean, Root mean square, Jitter, Shimmer
- Mel-frequency cepstral coefficients (MFCCs):12 MFCCs
- Formants: 5 Formants
- Pitch, Intensity and harmonicity based features: Mean, Minimum, Maximum, Range, Deviation
- Etc.

Voice

Voice conversion into the digital form

Extraction of numerical characteristics

Classification of sound signals

The **emotion** is detected

Sample

$x_{1,1}$	$x_{1,2}$	…	$x_{1,m}$	y_1
$x_{2,1}$	$x_{2,2}$	…	$x_{2,m}$	y_2
$x_{3,1}$	$x_{3,2}$	…	$x_{3,m}$	y_3
…	…	…	…	…
$x_{n,1}$	$x_{n,2}$	…	$x_{n,m}$	y_n

$\bar{x}_i$ – independent variable,
y_i – dependent variable. $i=\overline{1,n}$,
$y_i \in C$, where $C=\{c_1, c_2, \ldots, c_r\}$ – finite set,
r – the number of classes.

New examples

$x_{1,1}$	$x_{1,2}$	…	$x_{1,m}$	?
…	…	…	…	…
$x_{l,1}$	$x_{l,2}$	…	$x_{l,m}$	?

Goal:
To classify new objects based on the sample (supervised learning).

Corpora description

Database	Language	Full length (min.)	Number of emotions	File level duration		Notes
				Mean (sec.)	Std. (sec.)	
Emo-DB	German	24.7	7	2.7	1.02	Acted
SAVEE	English	30.7	7	3.8	1.07	Acted
LEGO	English	118.2	3	1.6	1.4	Non-acted
UUDB	Japanese	113.4	4	1.4	1.7	Non-acted

Experiments conducted

Common for all experiments:

- 6-fold cross-validation procedure
- The ***F-score (%)*** metric was evaluated.

Experiment 1 :

The conventional MLP classifier trained with the BP algorithm was applied (the program system *WEKA*).

Experiment 2 :

The proposed **two-criterion optimization model** was used to design MLP classifiers automatically.

Experiments conducted

Experiment 2:

For each component of the MOGA (NSGA-II. PICEA-g. and SPEA2) the following settings were defined:

- binary tournament selection;
- uniform recombination;
- the mutation probability $p_m=1/n$, where n is the length of the chromosome.

All islands had an equal **amount of resources**:

- **20 generations** and **30/3 = 10 individuals** in populations;
- **the migration size** was equal to 3 (in total each island got 6 points from two others);
- **the migration interval** was equal to 5 generations.

! As a result:

One model or **an ensemble of classifiers** (non-dominated candidates)

Activation functions used

№	Activation function	K_i value
1	$f(x)=\begin{cases}-1, & x<-1\\ x, & -1\le x\le 1\\ 1, & x>1\end{cases}$	6.46
2	$f(x)=1$	2.69
3	$f(x)=e^{-\frac{x^2}{2}}$	22.48
4	$f(x)=tanh(x)$	23.14
5	$f(x)=\frac{1}{1+e^{-x}}$	22.20
6	$f(x)=1-e^{-\frac{x^2}{2}}$	20.55
7	$f(x)=atan(x)$	27.01

$$* \; K_j(i)=\frac{T_i^{act}}{T^{weight}}$$

Experiments conducted

Experiment 3:

The feature selection procedure is embedded in the model design process.

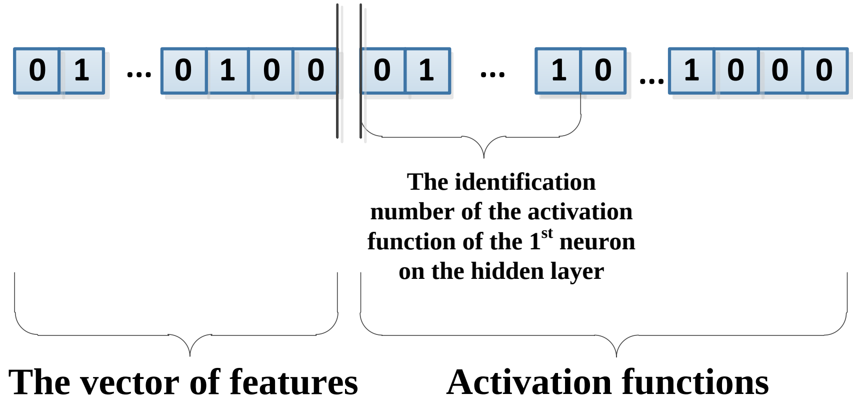


Figure 3. The presentation of the MLP structure with the feature selection procedure

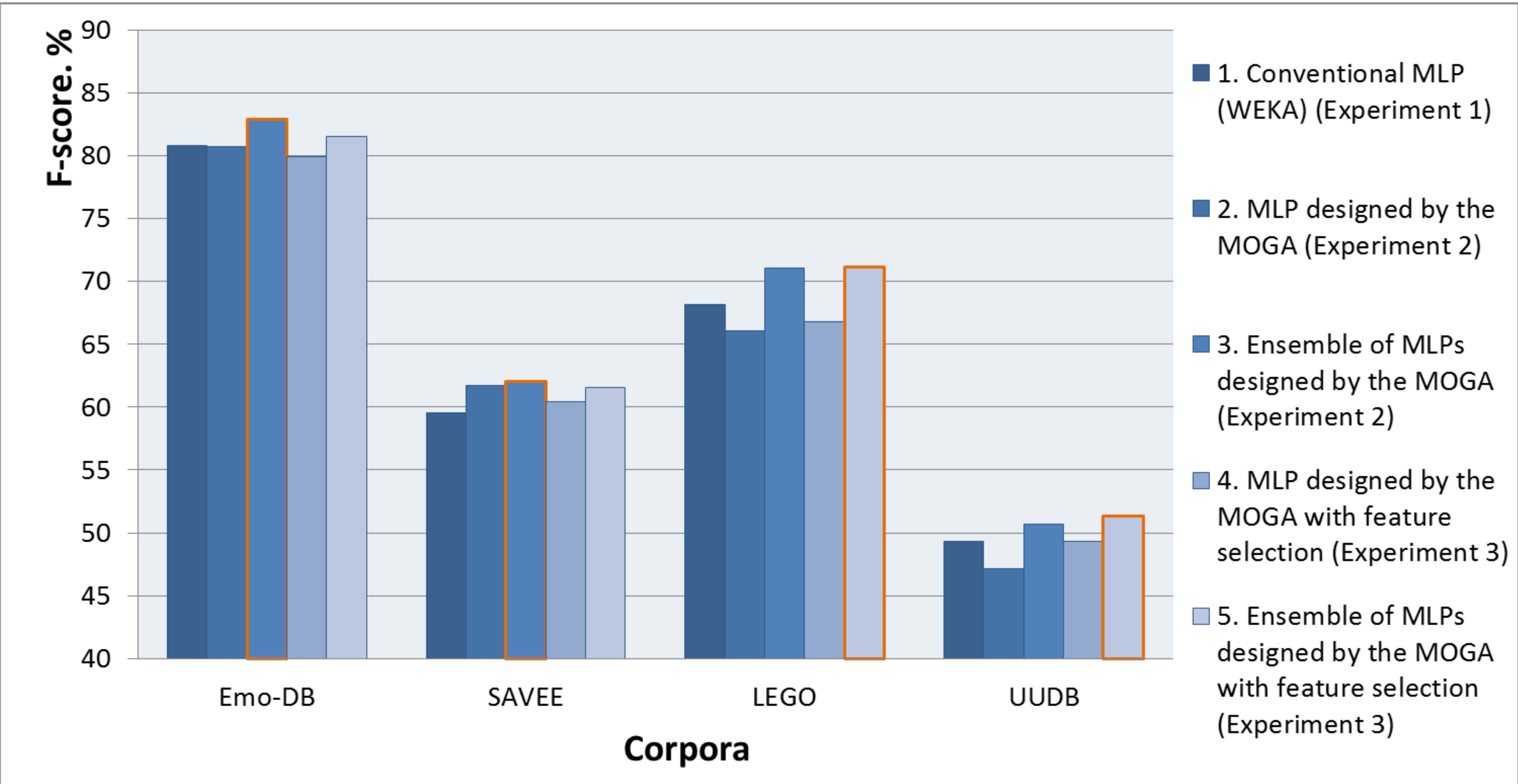
All islands had an equal **amount of resources**:

- **30 generations** and **60/3 = 120 individuals** in populations;
- **the migration size** was equal to 3 (in total each island got 6 points from two others);
- **the migration interval** was equal to 5 generations.

! As a result:

One model or **an ensemble of classifiers** (in our experiment there were **15 models** in the ensemble)

Experimental Results



Experimental Results

The average amount of neurons in layers

	Emo-DB		SAVEE		LEGO		UUDB	
	Input	Hidden	Input	Hidden	Input	Hidden	Input	Hidden
1. Conventional MLP (WEKA) (Experiment 1)	384.0	197.0	384.0	197.0	384.0	195.0	384.0	195.0
2. MLP designed by the MOGA (Experiment 2)	384.0	152.2	384.0	151.5	384.0	154.3	384.0	154.0
3. Ensemble of MLPs designed by the MOGA (Experiment 2)	384.0	151.6	384.0	152.7	384.0	153.1	384.0	154.7
4. MLP designed by the MOGA with feature selection (Experiment 3)	265.0	139.0	269.2	155.7	273.0	151.7	272.0	154.0
5. Ensemble of MLPs designed by the MOGA with feature selection (Experiment 3)	268.1	142.0	261.5	149.6	271.7	149.9	269.2	148.8

Conclusion and Future plans

✓ **The main benefit** of this approach is the opportunity to generate effective MLP structures taking into consideration two objectives *'classification performance'* and *'computational complexity'*.

Other advantages:

✓ In the experiments conducted it was revealed that this technique allowed us to design MLP classifiers with simpler structures, whose accuracy was comparable with (or even higher than) the performance of conventional MLPs containing more neurons in the hidden layer.

✓ In the framework of this technique, it is also possible to design ensembles of classifiers; their application leads to the essential improvement of the classification quality.

✓ The binary representation of MLP structures permitted us to embed the feature selection procedure and additionally to simplify classifiers.

Future plans:

Finally, there are some other questions related to the human-machine communication sphere. The proposed scheme might be applied without any changes to **the speech-based speaker identification problem** as well as to **speaker gender** or **age recognition**.

Thanks a lot!