Self-adaptive multi-objective genetic algorithms for feature selection

Christina Brester,
Eugene Semenkin

Siberian State Aerospace University,
Krasnoyarsk, Russia

Maxim Sidorov,
Wolfgang Minker

University of Ulm,
Ulm, Germany

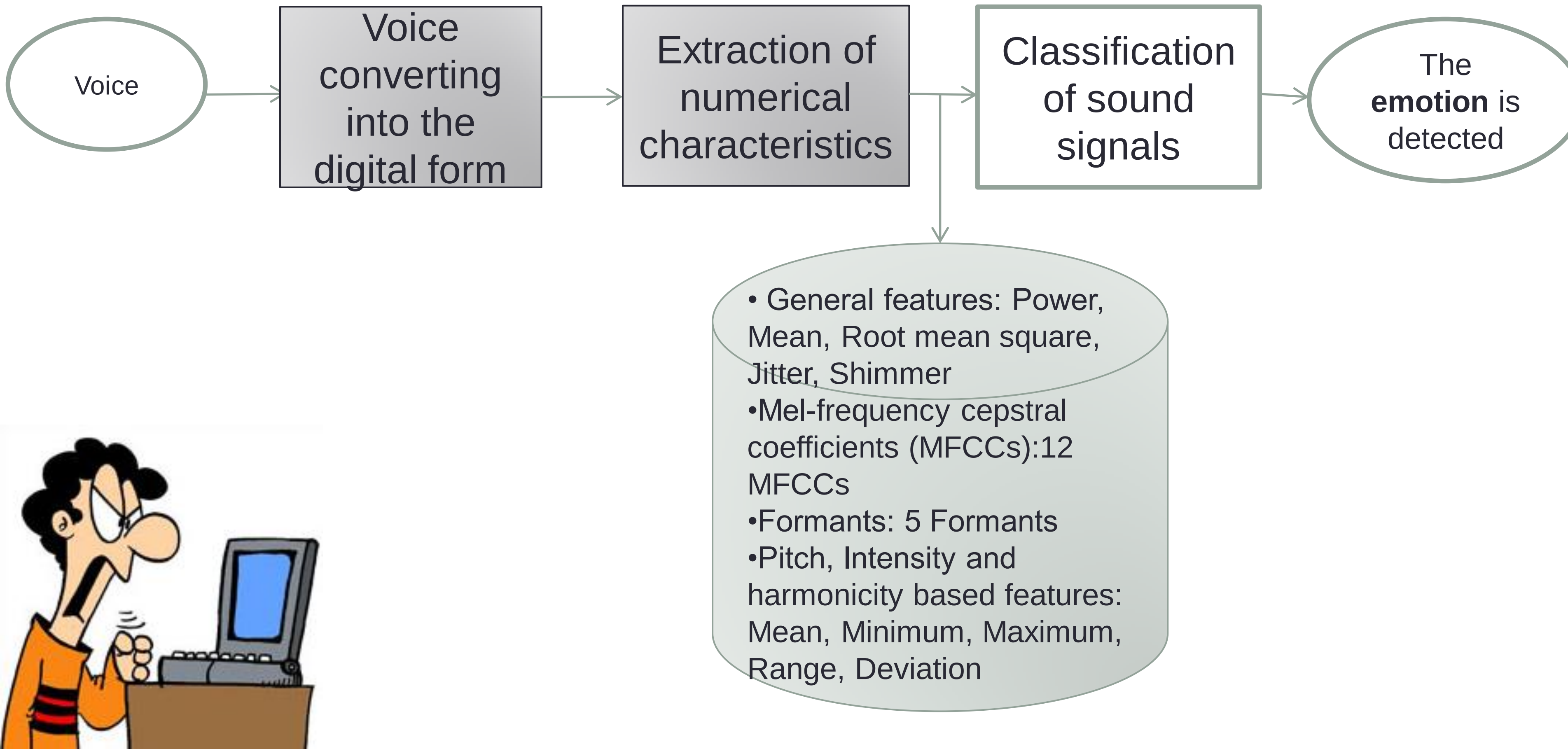
Motivation

- An opportunity to recognize human emotions might be useful in various applications:

- call centers quality monitoring;
- improvement of spoken dialogue systems.

- **An optimal feature set** which should be used to represent the speech signals is still an open question.

Speech-based Emotion Recognition Problem



Problem definition

Sample

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

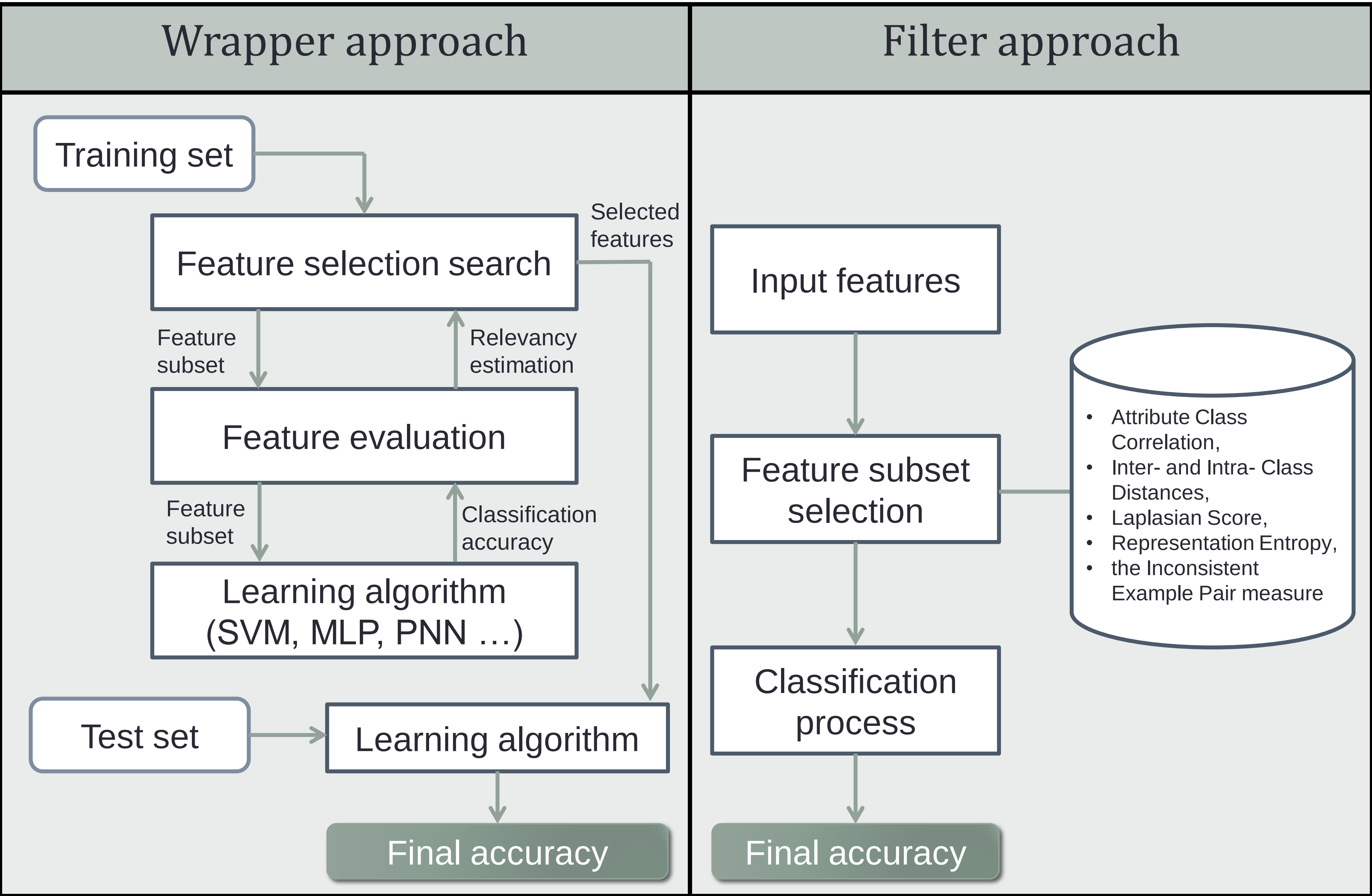
$\bar{x}_i$ – independent variable,
y_i – dependent variable, $i = \overline{1,n}$,
$y_i \epsilon C$, where $C = \{c_1, c_2, \dots, c_r\}$ – finite set,
r – the number of classes.

New examples

$x_{1,1}$	$x_{1,2}$	$\dots$	$x_{1,m}$	?
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
$x_{l,1}$	$x_{l,2}$	$\dots$	$x_{l,m}$	?

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

Feature selection concepts: *formal models*



Feature selection search

Main concepts:

- An optimization model with **binary representation**:

1	0	0	…	1

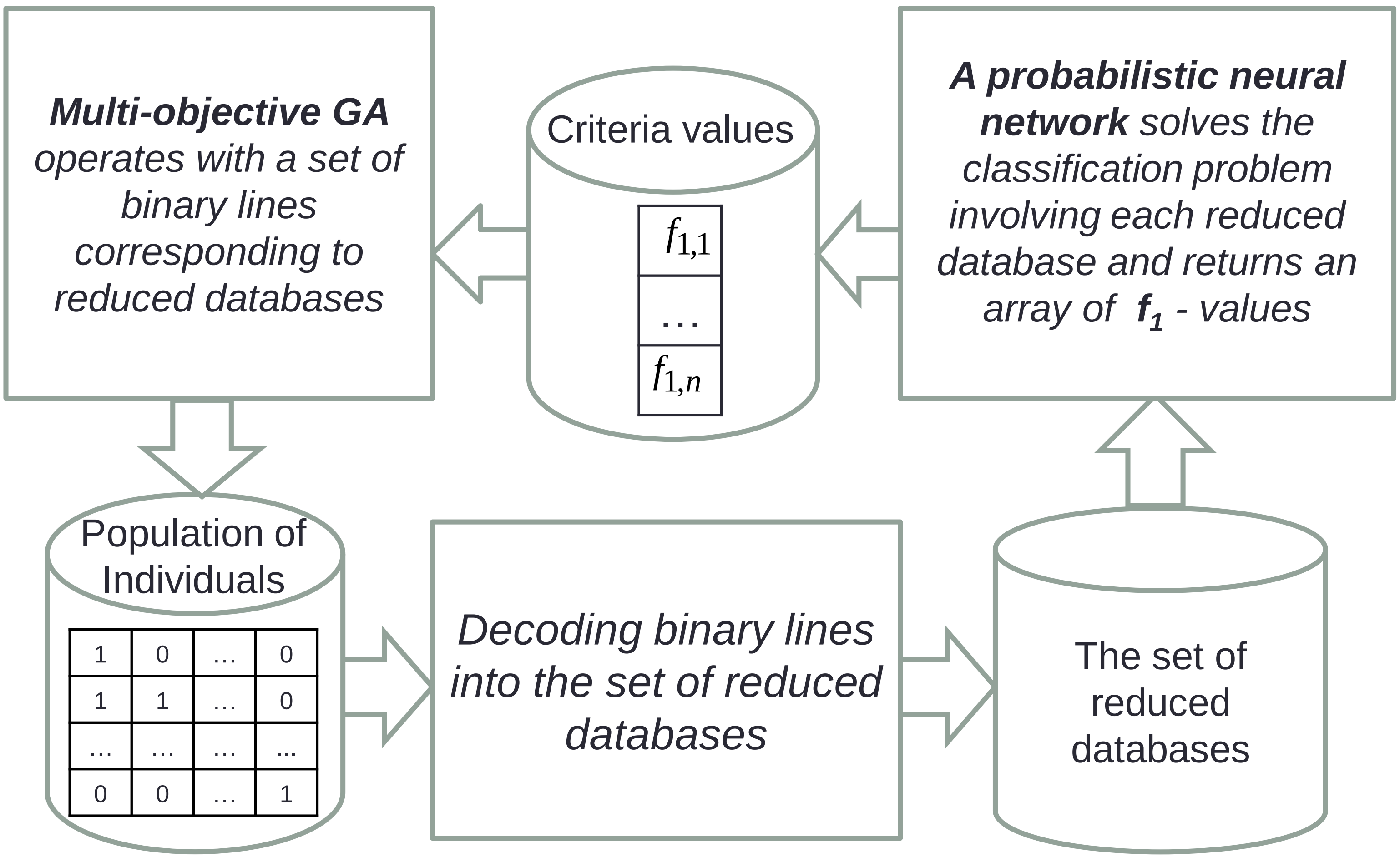
unit corresponds to the relevant attribute;
zero denotes the irrelevant attribute.

- **Evolutionary (genetic) algorithms** as a technique for optimizing both discrete and continuous criteria.

- **The self-adaptation idea** as a strategy to organize the automatic choice of algorithm settings.

Wrapper approach: *the actual model*

f1 - the relative classification error,
f2 - the number of selected features,
$$f1 \to \min, \; f2 \to \min$$



Filter approach: *the actual model*

f1 – the Intra-Class Distance (IA),
f2 - the Inter-Class Distance (IE),
f1 → min, f2 → max

- Attribute Class Correlation,
- ***Inter- and Intra- Class Distances,***
- Laplasian Score,
- Representation Entropy,
- the Inconsistent Example Pair measure

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

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

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

Multi-objective evolutionary algorithms

Method	Basic concepts
Preference-inspired co-evolutionary algorithm using goal vectors (PICEA-g) [Wang, 2013]	Pareto-dominance idea; Elitism; Incorporating decision maker preferences.
Multi-objective evolutionary algorithm based on decomposition (MOEA/D-DRA) [Zhang *et al*.,2009] (the leader of CEC 2009 MOEA competition)	Decomposition; Dynamic resource allocation.
The Strength Pareto Evolutionary Algorithm (SPEA) [E. Zitzler, L. Thiele, 1999]	Pareto-dominance idea; Elitism.
Genetic algorithm with the rank aggregating fitness function (GA-RAFF) [P. Bentley, J. Wakefield, 1997]	Aggregating criteria, Average ranking.

Self-adaptation concept 1

Genetic operators	**Preference-inspired co-evolutionary algorithm using goal vectors (PICEA-g)** **[Wang, 2013]**
Selection	Conventional tournament
Crossover	Conventional uniform
Mutation	The scheme proposed by Daridi *et al.* (2004) was engaged. This heuristics is equal to: $p_m = 1 / 240 + 0.11375 / 2^t$, where p_m is the mutation probability, t is the current generation number.

Self-adaptation concept 2

<table>
<tr><th>Genetic operators</th><th>Multi-objective evolutionary algorithm based on decomposition (MOEA/D-DRA)
[Zhang et al.,2009]
(the leader of CEC 2009 MOEA competition),

Genetic algorithm with the rank aggregating fitness function (GA-RAFF)
[P. Bentley, J. Wakefield, 1997]</th></tr>
<tr><td>Selection</td><td rowspan="3">Application probabilities q_i^k for each i-th variant of the k-th operator were introduced.
After the adaptation interval (that was the certain number of objective function evaluations) the probabilities are recalculated taking into account fitness of individuals generated by the given operator:

$$q_i^k = 0.2 / \mathrm{n}^k + 0.8 \cdot ratio_i^k / scale^k,$$
where $scale_k = \sum_i ratio_i^k$.

The first summand does not allow any probability to be equal to zero (that makes all variants of operators available throughout the algorithm execution).</td></tr>
<tr><td>Crossover</td></tr>
<tr><td>Mutation</td></tr>
</table>

Self-adaptation concept 3

Genetic operators	The Strength Pareto Evolutionary Algorithm (SPEA) [E. Zitzler, L. Thiele, 1999]
Selection	Conventional tournament
Crossover	The self-configurable recombination operator is based on the *co-evolution* idea: the population is divided into groups and each group is generated with a particular type of recombination (it may be *one-point*, *two-point* or *uniform* crossover). The efficiency of operators is compared in pairs in every *T*-th generation to reallocate resources on the basis of the fitness values. «Fitness» is proportional to the number of non-dominated individuals generated with a certain type of crossover and stored in the outer set.
Mutation	The scheme proposed by Daridi *et al.* (2004) was engaged. This heuristics is equal to: $p_m = 1 / 240 + 0.11375 / 2^t$, where p_m is the mutation probability, t is the current generation number.

Database description

Database	Language	Notes	Emotions
Berlin	German	Acted emotions	Neutral, anger, fear, joy, sadness, boredom, disgust

Full length (min.)	File level Duration		Sample size	The number of features	
	Mean (sec.)	Std. (sec.)		Baseline	Extended
24,7	2,7	1,02	535	37	384

Baseline results

	Relative classification accuracy, %	The number of selected features
PNN (baseline)	56.68	37
PNN (extended data set)	58.90	384
PCA+PNN	43.70	129.3

Experiment conditions

Learning algorithm	The probabilistic neural network (PNN) [D.F. Specht, 1990]
Experiment conditions	25 runs; random division in proportion 70-30%; stratification
Computational resources	100 individuals, 100 generations
Final solution	The candidate-solution that provides the minimum of the classification error on the validation data set (20% of the training data).

Experiment results

	Wrapper			Filter		
	Classification accuracy, %	Average number of features	Gain , %	Classification accuracy, %	Average number of features	Gain , %
PICEA-g	**73.05**	**85.5**	**28.83**	**75.37**	**128.0**	**32.97**
MOEA/ D-DRA	69.73	160.1	23.02	73.63	126.4	29.90
GA-RAFF	**73.02**	**101.5**	**28.83**	71.78	134.8	26.64
SPEA	71.46	68.4	26.08	**76.20**	**138.6**	**34.44**
GA	70.70	155.1	24.74	-	-	-

Using Wilcoxon nonparametric criteria (with significance level α = 0.05) it might be found that the GA-PNN system which is not oriented to the feature reduction does not outperform any approaches taking into consideration two criteria.

Conclusion

- We revealed advantages of using MOEAs in the feature selection procedure to solve the speech-based emotion recognition problem.

- Obtained results reflect superiority of the developed approach in contrast to application the PCA-hybrid system.

- An application of the PNN-MOEA hybrid system for selecting the most representative features and maximizing the accuracy of the supervised learning algorithm could decrease the number of features from 384 to 64.8 and increase the ER accuracy up to 34.44%.

- Future lines of this study lie in the following directions:
  - there is an opportunity for the most effective MOEAs to co-operate with each other to achieve better results,
  - this approach should be investigated on the set of other classification problems (speaker or gender identification).