ICINCO 2014
11th INTERNATIONAL CONFERENCE ON INFORMATICS IN CONTROL, AUTOMATION AND ROBOTICS
VIENNA, AUSTRIA 1 - 3 SEPTEMBER, 2014

Acoustic Emotion Recognition: Two Ways of Feature Selection Based on Self-Adaptive Multi-Objective Genetic Algorithm

Christina Brester (*abahachy@mail.ru*),
Eugene Semenkin (*eugenesemenkin@yandex.ru*)
Institute of Computer Sciences and Telecommunication,
Siberian State Aerospace University, Krasnoyarsk, Russia

Maxim Sidorov (*maxim.sidorov@uni-ulm.de*)
Institute of Communications Engineering,
University of Ulm, Germany

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

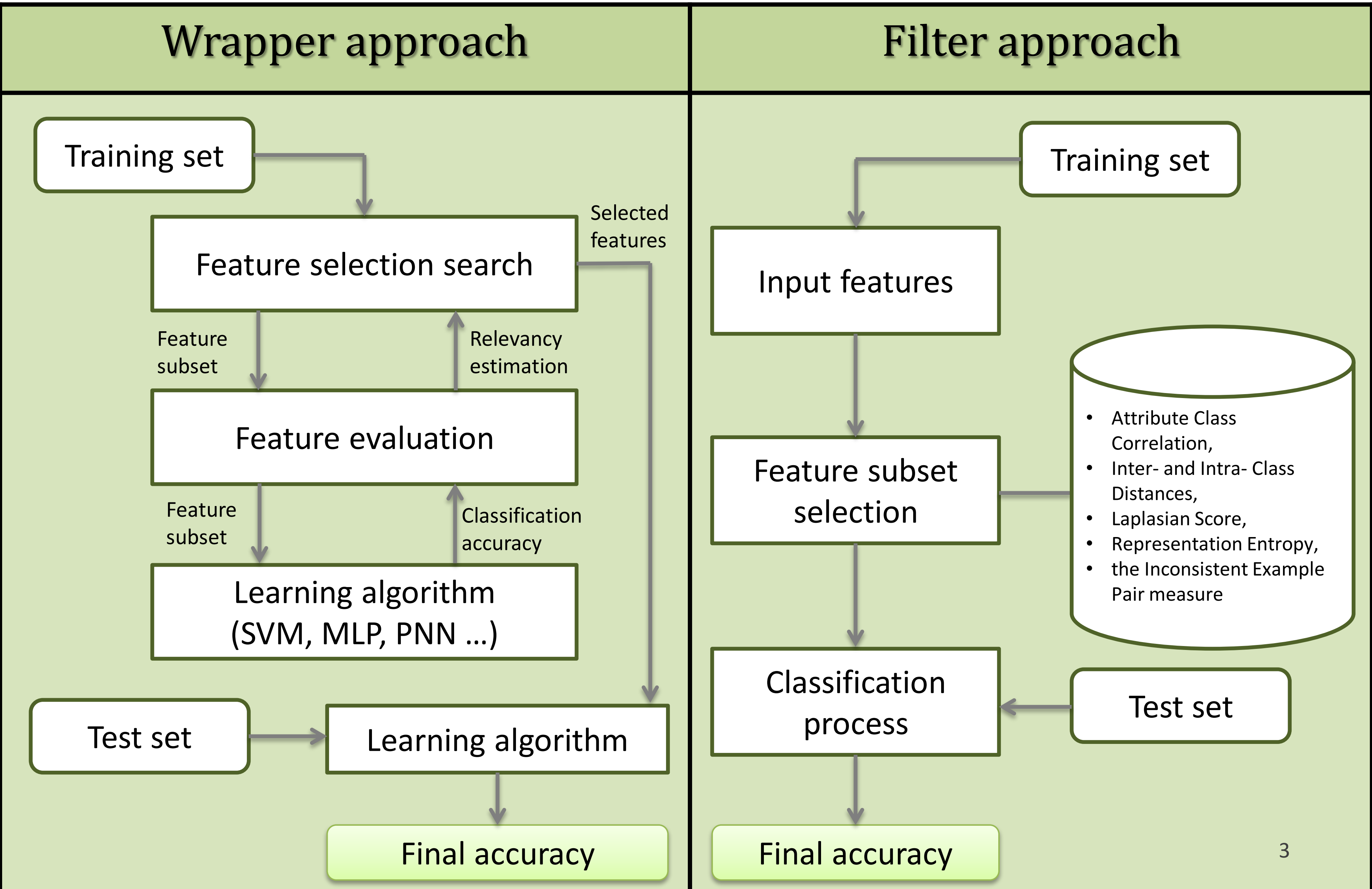
$\bar{x}_i$ – independent variable,
y_i – dependent variable, $i = \overline{1,n}$,
$y_i \in C$, where $C = \{c_1, c_2, \ldots, c_r\}$ – finite set,
r – the number of classes.

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).

Feature selection concepts: *formal models*

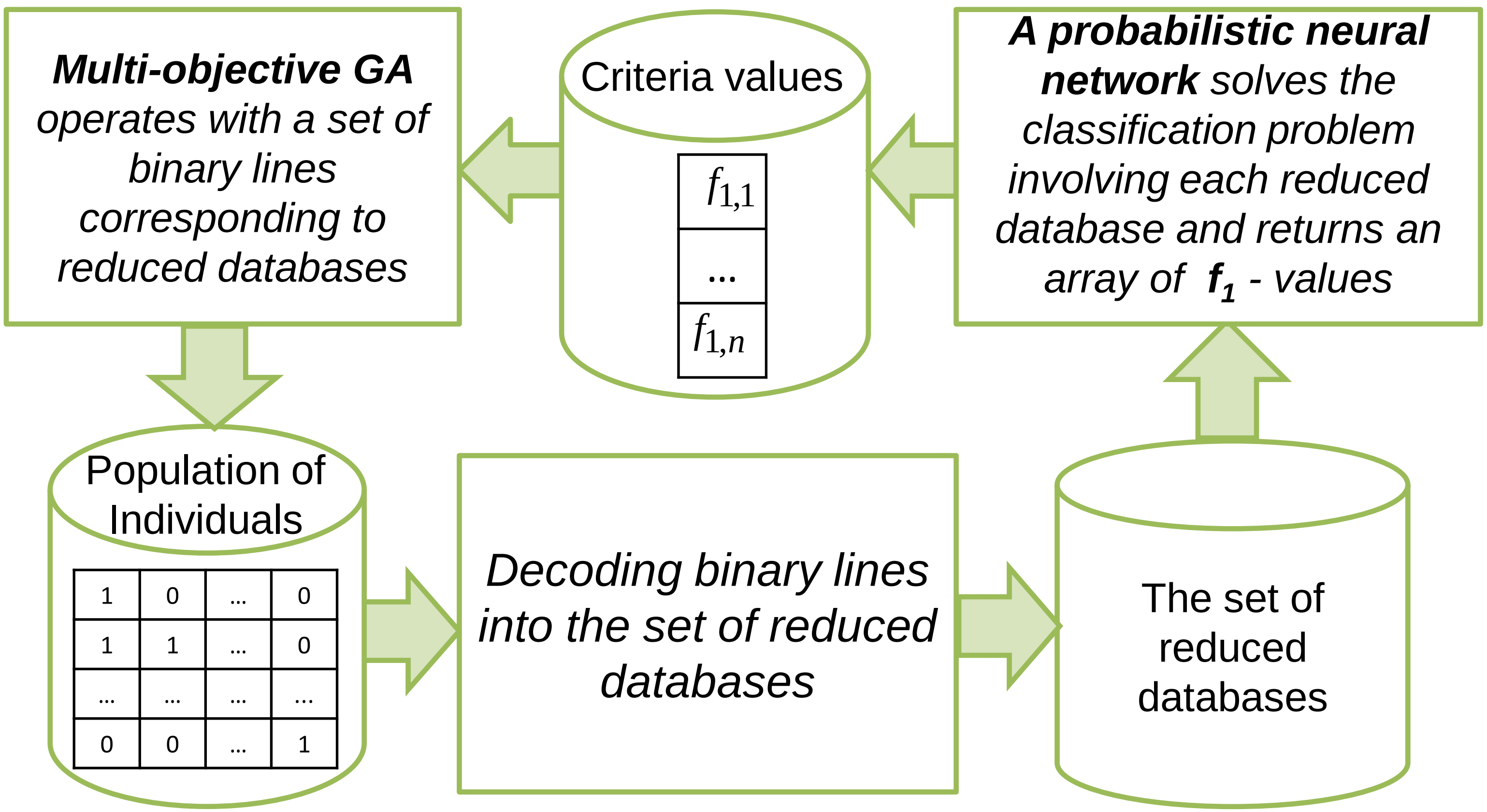


Wrapper approach: *the actual model*

f1 - the relative classification error,

f2 - the number of selected features,

$$f1 \to min,\ f2 \to min$$



Christina Brester (*abahachy@mail.ru*), Eugene Semenkin (*eugenesemenkin@yandex.ru*),
Institute of Computer Sciences and Telecommunication, Siberian State Aerospace University, Krasnoyarsk, Russia,
Maxim Sidorov (*maxim.sidorov@uni-ulm.de*), *Institute of Communications Engineering, University of Ulm, Germany*

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(…,…) denotes the Euclidian distance,
p_r and n_r represent the central example and the number of examples in the r-th class.

Christina Brester (*abahachy@mail.ru*), Eugene Semenkin (*eugenesemenkin@yandex.ru*),
Institute of Computer Sciences and Telecommunication, Siberian State Aerospace University, Krasnoyarsk, Russia,
Maxim Sidorov (*maxim.sidorov@uni-ulm.de*), *Institute of Communications Engineering, University of Ulm, Germany*

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.

Christina Brester (*abahachy@mail.ru*), Eugene Semenkin (*eugenesemenkin@yandex.ru*),
Institute of Computer Sciences and Telecommunication, Siberian State Aerospace University, Krasnoyarsk, Russia,
Maxim Sidorov (*maxim.sidorov@uni-ulm.de*), *Institute of Communications Engineering, University of Ulm, Germany*

Self-adaptation concept

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.

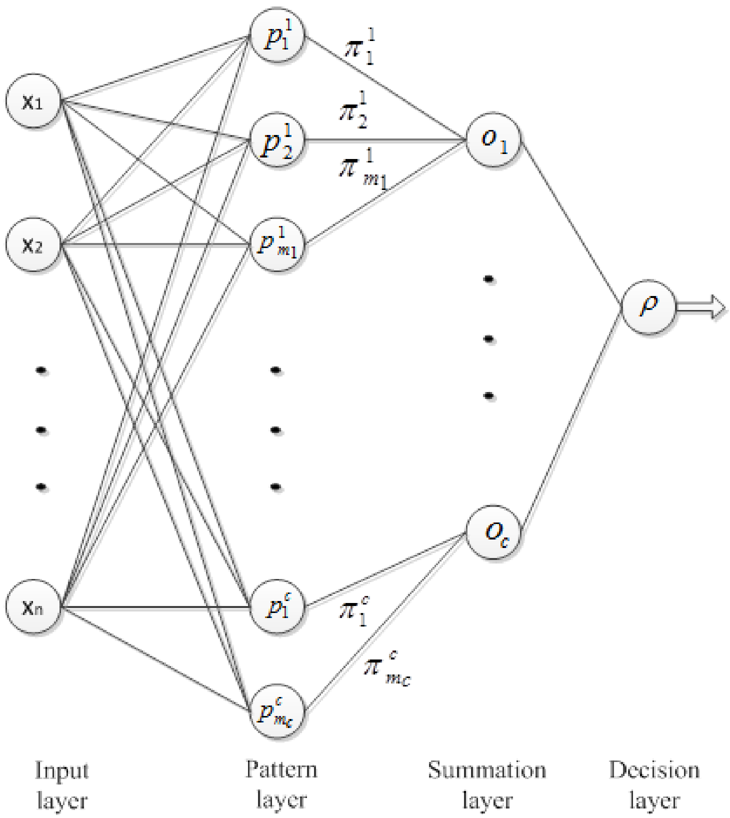
Probabilistic Neural Network Architecture

- The number of input neurons is equal to the amount of features;
- The quantity of elements in the pattern layer is equal to the training sample size;
- The amount of elements in the summation layer is equal to the number of classes.

Activation functions in the second layer are:

$$f_j(\bar{x}) = \exp\left(\frac{-\sum_i (w_{i,j} - x_i)^2}{\sigma^2}\right)$$

– Gaussian kernels formed using training data w_j as centers.



Christina Brester (*abahachy@mail.ru*), Eugene Semenkin (*eugenesemenkin@yandex.ru*),
Institute of Computer Sciences and Telecommunication, Siberian State Aerospace University, Krasnoyarsk, Russia,
Maxim Sidorov (*maxim.sidorov@uni-ulm.de*), *Institute of Communications Engineering, University of Ulm, Germany*

Outline

- Motivation
- Problem understanding
- Theoretical aspects:
  - Two schemes of Feature Selection: wrapper and filter approaches
  - Self-adaptive Multi-Objective algorithm
  - Probabilistic Neural Networks
- Corpora description
- Experiment conditions
- Experimental results
- Conclusion

Christina Brester (*abahachy@mail.ru*), Eugene Semenkin (*eugenesemenkin@yandex.ru*),
Institute of Computer Sciences and Telecommunication, Siberian State Aerospace University, Krasnoyarsk, Russia,
Maxim Sidorov (*maxim.sidorov@uni-ulm.de*), *Institute of Communications Engineering, University of Ulm, Germany*

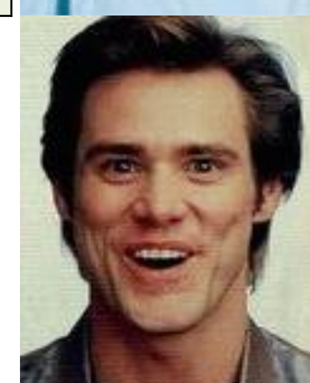
Corpora description

Database	Language	Full length (min.)	Number of emotions	File level Duration		Emotion level Duration		Notes
				Mean (sec.)	Std. (sec.)	Mean (sec.)	Std. (sec.)	
Berlin	German	24,7	7	2,7	1,02	212,4	64,8	Acted
SAVEE	English	30,7	7	3,8	1,07	263,2	76,3	Acted
VAM-Audio	English	47,8	4	3,02	2,1	717,1	726,3	Non-acted


Berlin	SAVEE	VAM
Neutral, anger, fear, joy, sadness, boredom, disgust	Sad-bored, angry-anxious, relaxed-serene, happy-exciting	Anger, disgust, fear, happiness, sadness, surprise, neutral







Experiment conditions

Feature selection search	The Strength Pareto Evolutionary Algorithm (E. Zitzler, L. Thiele, 1999)
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).

Christina Brester (*abahachy@mail.ru*), Eugene Semenkin (*eugenesemenkin@yandex.ru*),
Institute of Computer Sciences and Telecommunication, Siberian State Aerospace University, Krasnoyarsk, Russia,
Maxim Sidorov (*maxim.sidorov@uni-ulm.de*), *Institute of Communications Engineering, University of Ulm, Germany*

Results: analysis and inferences

	Relative classification accuracy, %		
	Berlin	SAVEE	VAM
PNN	58.90 (384)	47.32 (384)	67.07 (384)
PCA+PNN	43.7 (129.3)	26.5 (123.6)	59.4 (148.6)
SPEA_wrapper+PNN	71.5 (68.4)	48.4 (84.1)	70.6 (64.8)
SPEA_filter+PNN	76.2 (138.6)	60.8 (142.0)	73.2 (152.8)

Christina Brester (*abahachy@mail.ru*), Eugene Semenkin (*eugenesemenkin@yandex.ru*),
Institute of Computer Sciences and Telecommunication, Siberian State Aerospace University, Krasnoyarsk, Russia,
Maxim Sidorov (*maxim.sidorov@uni-ulm.de*), *Institute of Communications Engineering, University of Ulm, Germany*

Conclusion

- 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% for some of the corpora.

- According to obtained results, the heuristic search for feature selection in the ER problem is much more effective than application of the PCA-based technique that leads to decreasing the classification accuracy.

- The usage of more accurate classifiers and more effective MOEA might improve the performance of the system.

Christina Brester (*abahachy@mail.ru*), Eugene Semenkin (*eugenesemenkin@yandex.ru*),
Institute of Computer Sciences and Telecommunication, Siberian State Aerospace University, Krasnoyarsk, Russia,
Maxim Sidorov (*maxim.sidorov@uni-ulm.de*), *Institute of Communications Engineering, University of Ulm, Germany*