SIBERIAN STATE AEROSPACE UNIVERSITY

Krasnoyarsk, Russia




Evolutionary feature selection for emotion recognition in multilingual speech analysis

Christina Brester, Eugene Semenkin, Maxim Sidorov,
Igor Kovalev, Pavel Zelenkov

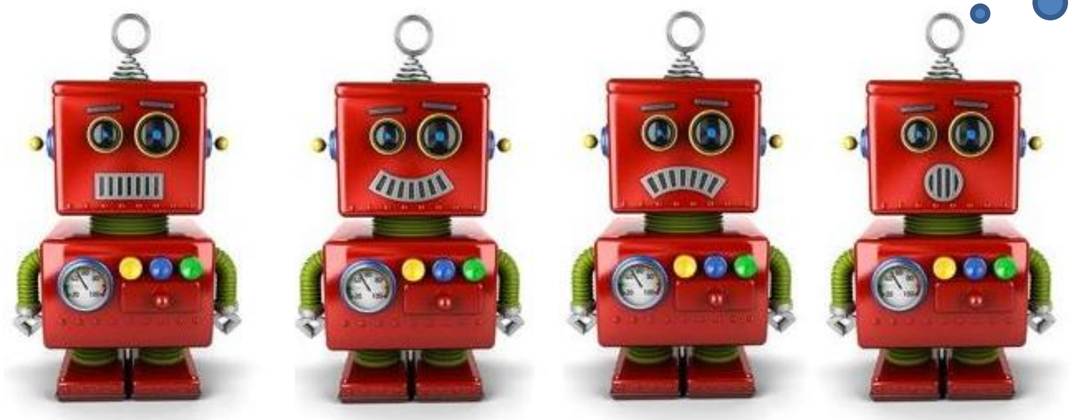



Speech-based Emotion Recognition Problem

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

… etc.

Why do we talk about *Feature Selection*?

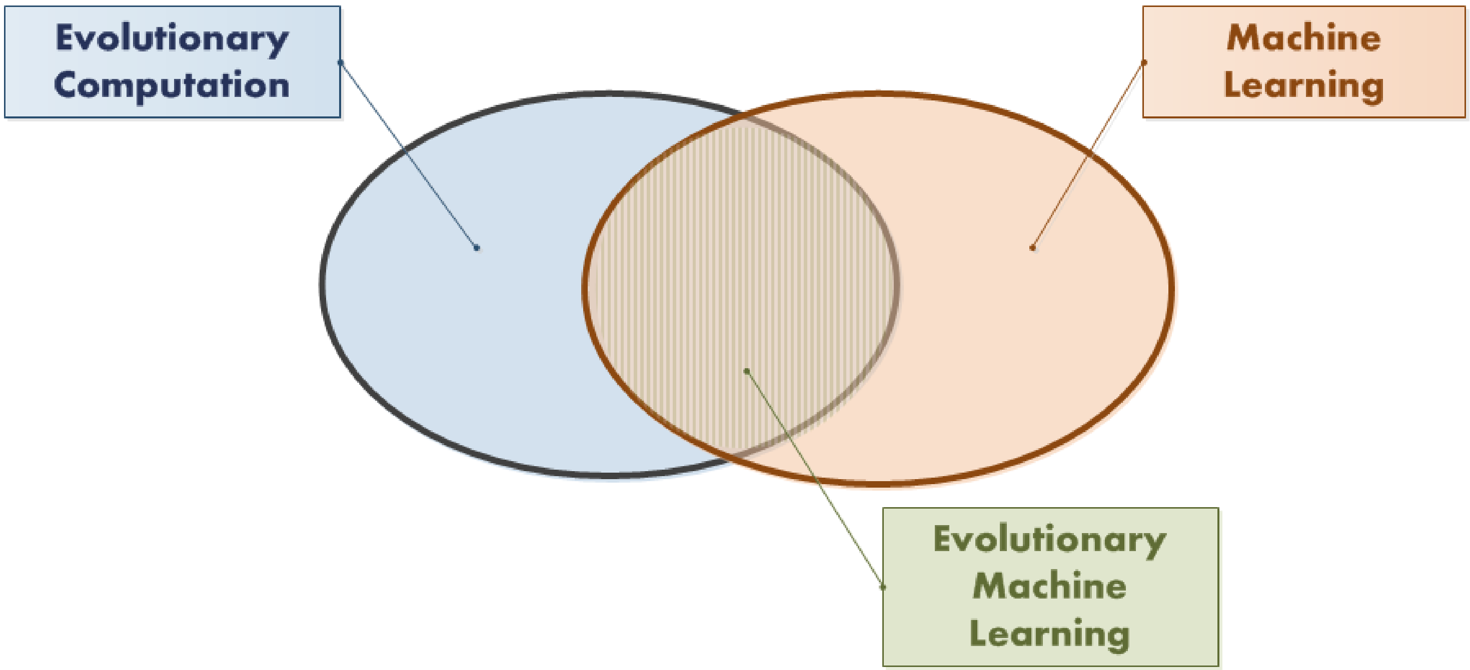


- **The number of features** extracted from the speech signal is **overwhelming**.
- **An optimal feature set** which should be used to represent the speech signals is still an open question.

Outline

- **Motivation**
  - Problem Definition
- **Background**
  - Evolutionary Computation and Machine Learning Integration
  - Feature Selection: Wrapper or Filter
- **Proposed approach**
  - Two-criteria Filter Feature Approach
  - Multi-objective Genetic Algorithms
  - The Island Model
  - Speech-based Emotion Recognition Problem
  - Corpora Description
- **Results and Discussion**
  - Baseline
  - Principle Component Analysis
  - Heuristic Feature Selection
  - Feature Selection with the Island Model and the ensemble of classifiers
- **Conclusion and Future Plans**

Integration of Evolutionary Computation and Machine Learning

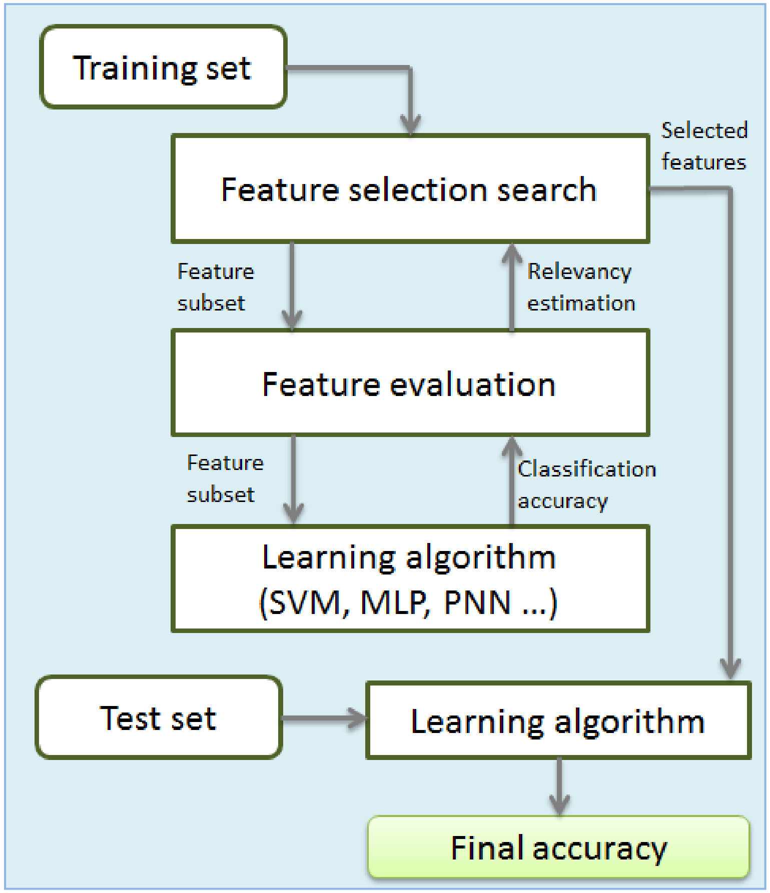


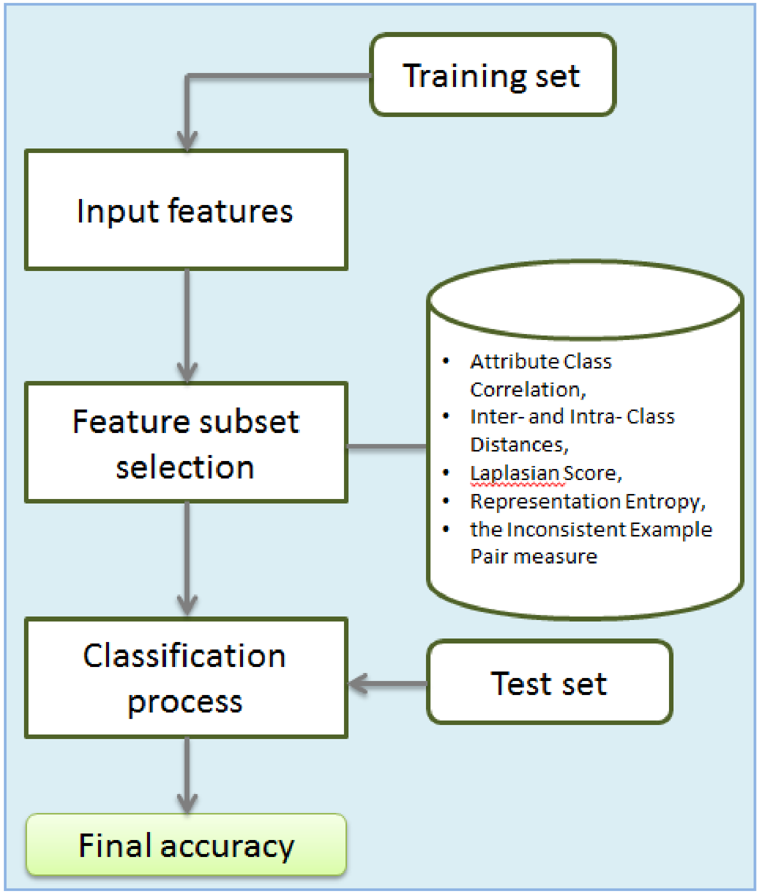
Why evolutionary?

Integration of Evolutionary Computation and Machine Learning

Pros	*Cons*
✓ The classification accuracy of the best **evolutionary** and **non-evolutionary** methods are **comparable;** ✓ Population-based search is easily **parallelized;** ✓ These methods can work in the **dynamic non-stationary environment;** ✓ **Feature selection** and **learning** in one process might be **combined;** ✓ From an optimization perspective, learning problems are typically **large, non-differentiable, noisy, deceptive, multimodal, high-dimensional, and highly constrained**. Evolutionary algorithms are an **effective tool** for such problems.	X Evolutionary methods are generally much **slower** than the non-evolutionary alternatives ➢ *solution: parallelization* X The **performance of evolutionary algorithms varies** significantly for different problems ➢ *solution: cooperative algorithms*

Two main feature selection concepts: Wrapper vs Filter





Why filter?

Two main feature selection concepts

Wrapper …

✓ involves classification models to evaluate the relevancy of each feature subset: **adjustment to** an applied **classifier**;

X requires **high computational resources.**

vs

Filter …

✓ needs significantly **fewer calculations** therefore it is rather effective in the sense of computational effort;

✓ might be effectively used **in combination with an ensemble** of diverse classifiers (MLP, SVM, Logit);

X does **not cooperate** with a **learning algorithm** and so ignores its performance entirely.

Two-criteria Filter Approach

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.

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 cooperation of evolutionary algorithms** as a strategy to avoid the of an appropriate algorithm for the problem considered.

Multi-Objective Genetic Algorithms

- *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.

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

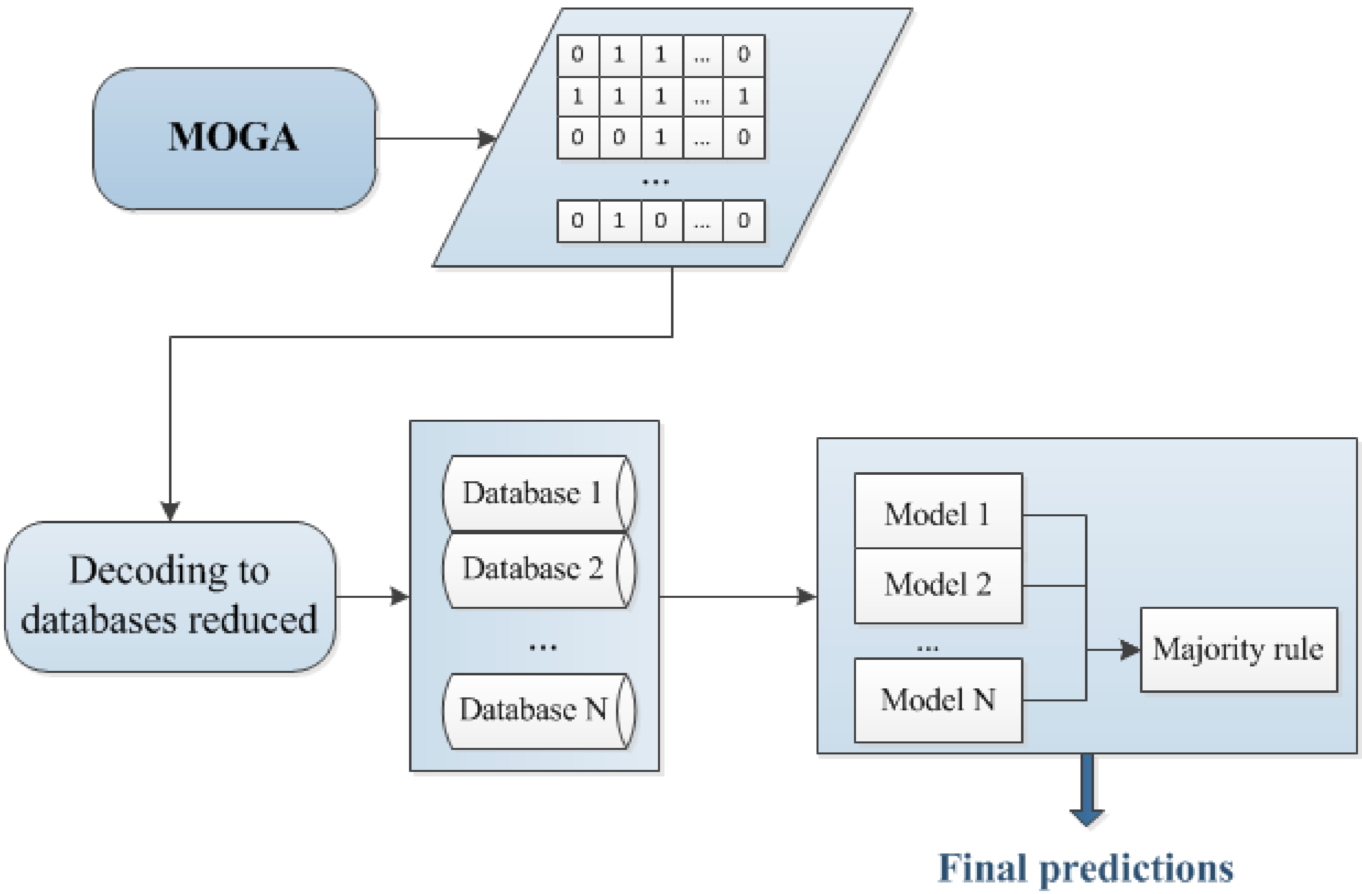
Tasks:

1. To investigate the **effectiveness of MOGAs**, which are based **on various heuristic mechanisms**, from the perspective of the feature selection procedure;
2. To implement the **cooperation of MOGAs** and observe its effectiveness.

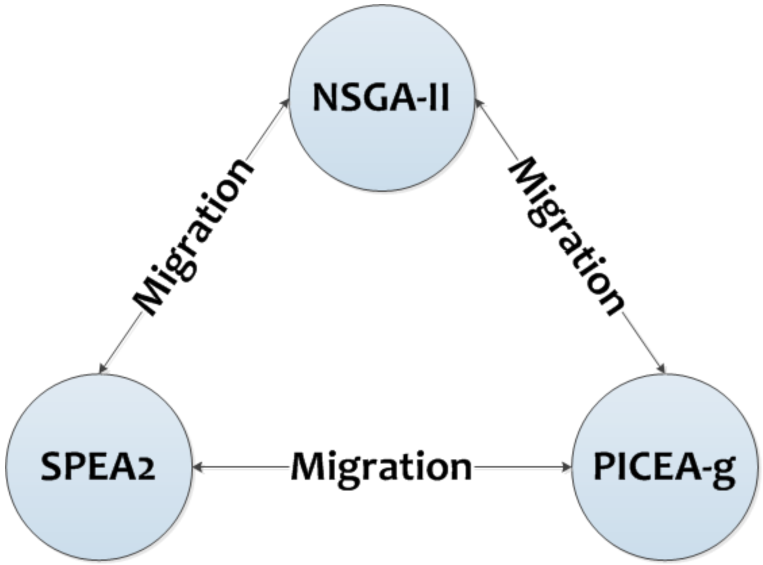
Multi-Objective Genetic Algorithms: Task 1

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

Multi-Objective Genetic Algorithms: Task 1



Multi-Objective Genetic Algorithms: Task 2



Island model …

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

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.)	
Berlin	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
- Conventional classifiers (WEKA):
  - *Support Vector Machine – SMO;*
  - *Multilayer Perceptron – MLP;*
  - *Linear Logistic Regression – Logit*.
- The *F-score* metric was evaluated.

Experiments conducted

Experiment 1:

Conventional classifiers (SMO, MLP, Logit) without Feature Selection -> **Baseline**

Experiment 2:

The same classifiers (SMO, MLP, Logit) after the application of Principal Component Analysis (the conventional attribute selection method) with the threshold values 0.75 and 0.95.

Experimental Results (1, 2)

Database	Feature Selection Method	Average Number of Selected Features	F-score Values, %		
			SMO	*MLP*	*LOGIT*
Emo-DB	Without Feature Selection	384.00	82.58	**82.98**	80.46
	PCA (0.75)	49.67	79.61	74.71	77.04
	PCA (0.95)	136.80	73.62	73.87	76.39
SAVEE	Without Feature Selection	384.00	59.31	**61.82**	60.82
	PCA (0.75)	46.67	57.86	57.46	59.86
	PCA (0.95)	130.7	46.18	50.63	51.80
LEGO	Without Feature Selection	384.00	**71.08**	64.77	70.71
	PCA (0.75)	59.83	68.05	67.19	69.03
	PCA (0.95)	162.50	70.06	66.08	70.58
UUDB	Without Feature Selection	384.00	50.44	41.94	**50.88**
	PCA (0.75)	46.67	48.48	47.53	49.61
	PCA (0.95)	156.80	49.37	47.93	49.89

Experiments conducted

Experiment 3: The two-criterion filter feature selection with MOGAs and conventional classifiers

NSGA-II, PICEA-g, and SPEA2 were used as optimizers in combination with SMO, MLP, and Logit classifiers.

- All algorithms were provided with the same amount of resources (90 generations and 150 individuals in populations).
- For each MOGA 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.

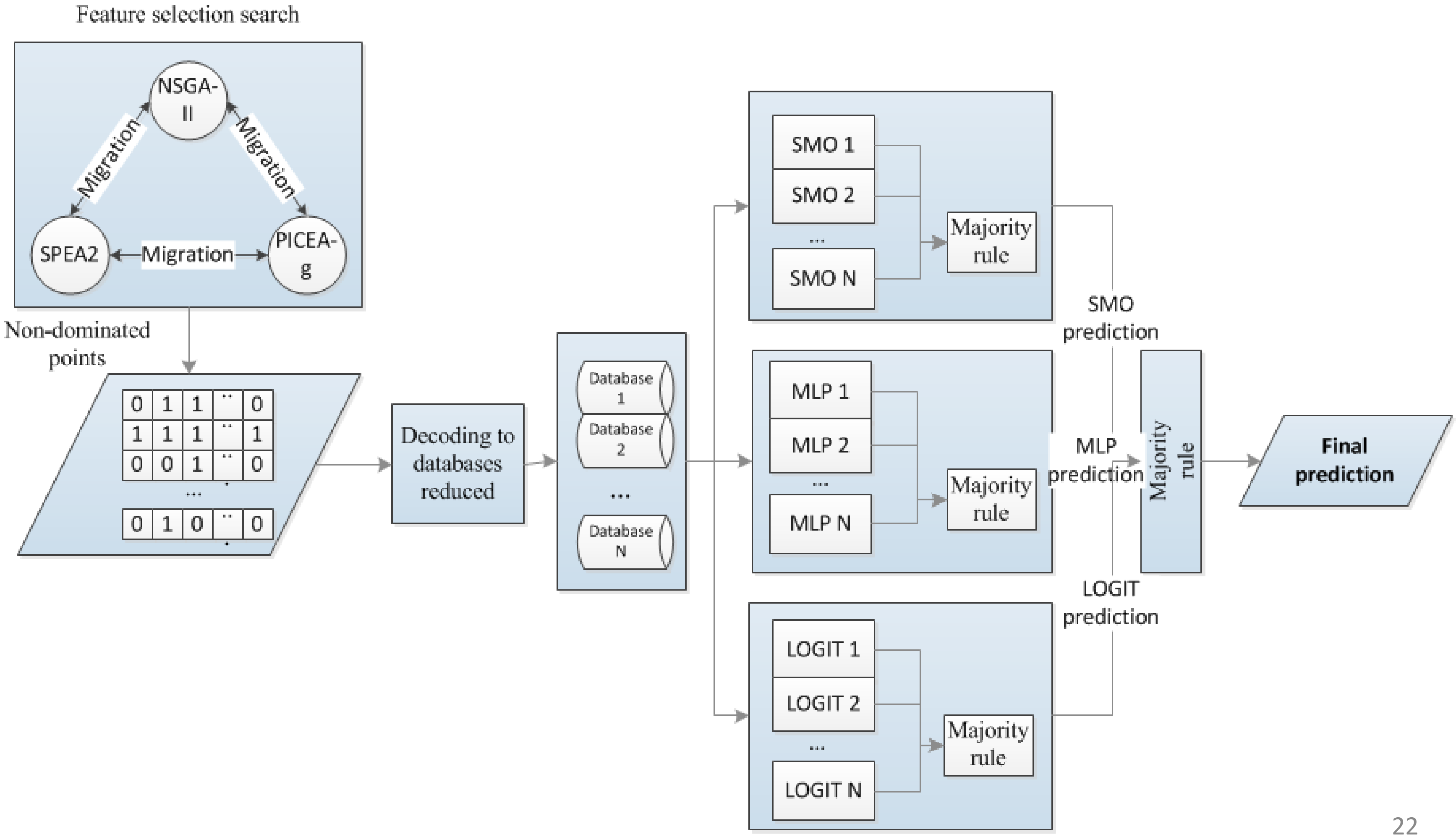
Experiments conducted

Experiment 4: The two-criterion filter feature selection with the cooperative MOGA and the ensemble of classifiers

The island model including ‘NSGA-II’, ‘PICEA-g’, and ‘SPEA2’ was applied to solve the two-criterion feature selection problem.

- All islands had an equal amount of resources (90 generations and 150/3 = 50 individuals in populations), the migration size was equal to 10 (in total each island got 20 points from two others), and the migration interval was equal to 10 generations;
- An ensemble of classifiers (SMO, MLP, and Logit) was used after the feature selection procedure.

The evolutionary feature selection scheme in combination with the ensemble of classifiers



Experimental Results (3, 4)

Database	Feature Selection Method	Average Number of Selected Features	F-score Values, %		
			SMO	*MLP*	*LOGIT*
Emo-DB	Without Feature Selection	384.00	82.58	82.98	80.46
	NSGA-II	165.16	82.77	**86.07**	82.56
	PICEA-g	180.88	83.06	84.74	84.09
	SPEA2	159.49	82.93	85.69	85.64
	Island model (NSGA-II, PICEA-g, SPEA2)	166.89	**86.26**		
SAVEE	Without Feature Selection	384.00	59.31	61.82	60.82
	NSGA-II	163.67	66.61	67.70	**69.84**
	PICEA-g	186.06	64.80	68.81	64.28
	SPEA2	166.74	64.48	67.55	66.82
	Island model (NSGA-II, PICEA-g, SPEA2)	165.94	**68.71**		

Experimental Results (3, 4)

Database	Feature Selection Method	Average Number of Selected Features	F-score Values, %		
			SMO	*MLP*	*LOGIT*
LEGO	Without Feature Selection	384.00	71.08	64.77	70.71
	NSGA-II	145.21	70.19	71.86	70.36
	PICEA-g	166.48	70.47	71.98	70.22
	SPEA2	151.18	70.61	**72.71**	70.58
	Island model (NSGA-II, PICEA-g, SPEA2)	150.68	**71.29**		
UUDB	Without Feature Selection	384.00	50.44	41.94	50.88
	NSGA-II	141.47	50.12	50.44	50.50
	PICEA-g	167.50	50.34	50.22	**50.91**
	SPEA2	145.68	50.37	50.19	50.59
	Island model (NSGA-II, PICEA-g, SPEA2)	146.85	**51.02**		

Discussion

- A **t-test** (with the significance level p=0.01) was used to compare the results:

for all of the corpora ***there was no difference*** *between* ***the best results*** *obtained in* ***Experiment 3*** *(with a classifier and a MOGA which realized a feature selection search) and the F-score values provided with* ***the island model of MOGAs*** *and* ***the ensemble of classifiers****.*

- The application of the proposed approach allowed us:

- *to* ***reduce*** *the* ***number of features*** *significantly (approximately by a factor of two);*
- *to* ***save*** *the* ***computational time*** *owing to the parallel work of the island model (roughly by a factor of 2.5 because the additional time was spent on the migration process).*

Conclusions and Future Plans

- The proposed evolutionary feature selection scheme based on an island model includes a number of algorithms and, therefore, **does not require additional experiments** to expose **the most appropriate MOGA** for the problem considered.
- It **saves computational time** due to the parallel work of islands.
- Besides, the two-criteria filter approach might be effectively used as a pre-processing stage in combination with **an ensemble of classifiers**.
- According to the results obtained, a high level of emotion recognition was achieved (up to **11.15% relative improvement** for the SAVEE database compared with the best F-score value on the full set of attributes).
- Thus, the described evolutionary feature selection technique is **an effective alternative** to conventional dimension reduction procedures such as **Principal Component Analysis**.
- Moreover, there are some other aspects related to speech-based recognition of human qualities of the user such as **gender** and **speaker identification**. Consequently, the proposed schemes might be applied to solve these problems.

Thanks a lot!