MULTI-OBJECTIVE HEURISTIC FEATURE SELECTION FOR SPEECH-BASED MULTILINGUAL EMOTION RECOGNITION

Christina Brester[1], Eugene Semenkin[1], and Maxim Sidorov[2]

[1]*Institute of Computer Science and Telecommunications,*
Reshetnev Siberian State Aerospace University,
Krasnoyarsky rabochy Av. 31, 660037, Krasnoyarsk, Russian Federation

[2]*Institute of Communications Engineering,*
Ulm University, Albert Einstein-Allee 43, 89081, Ulm, Germany

Abstract

If conventional feature selection methods do not show sufficient effectiveness, alternative algorithmic schemes might be used. In this paper we propose an evolutionary feature selection technique based on the two-criterion optimization model. To diminish the drawbacks of genetic algorithms, which are applied as optimizers, we design a parallel multi-criteria heuristic procedure based on an island model. The performance of the proposed approach was investigated on the Speech-based Emotion Recognition Problem, which reflects one of the most essential points in the sphere of human-machine communications. A number of multilingual corpora (German, English and Japanese) were involved in the experiments. According to the results obtained, a high level of emotion recognition was achieved (up to a 12.97% relative improvement compared with the best F-score value on the full set of attributes).

Keywords: multi-objective optimization, feature selection, speech-based emotion recognition

1 Introduction

While solving classification problems it is reasonable to accomplish pre-processing of data used by a learning algorithm due to several reasons: features might have a low variation level, correlate with each other or be measured with mistakes.

In [1] it is demonstrated that if there are irrelevant attributes in the sample, then the performance of decision trees, trained by ID3, 4.5, CART, decreases significantly, which is also valid for the k-nearest neighbors algorithm. Although the Nave Bayes classifier is robust in the sense of irrelevant features, one may observe the tremendous deterioration of its performance if there are correlating attributes.

For example, on the MONK1 problem [2] (this database contains characteristics of a robot) the 4.5 algorithm generates a tree with 15 nodes: 5 nodes are related to checking non-informative features (there are three irrelevant attributes in the sample). In that case the relative classification error equals 24.3%. However, if irrelevant features are eliminated, then the error decreases up to 11.1%. The same behavior of the 4.5 algorithm is showed on the credit scoring and diabetes diagnostics problems [3].

Generally, the feature selection procedure can be organized as the *wrapper* approach or the *filter* one [1]. The first technique involves classification models to evaluate the relevancy of each feature subset. Although it requires high computational resources, this approach demonstrates adjustment to an applied classifier. The second technique is referred to the pre-processing stage because it extracts information from the data set and reduces the number of attributes, taking into consideration such measures as consistency, dependency, and distance. This approach needs significantly fewer calculations therefore it is rather effective in the sense of computational effort. On the one hand, the filter attribute selection procedure does not cooperate with a learning algorithm and so ignores its performance entirely. However, on the other hand, it might be effectively used in combination with an ensemble of diverse classifiers, which is quite reasonable in the case when one does not know one particular reliable and effective model. Therefore, in this paper we propose the feature selection procedure which corresponds to the filter scheme.

Moreover, in recent years there has been a growing interest in the sphere of *Evolutionary Machine Learning*. However, some researchers highlight the negative sides of the *Evolutionary Computation* and *Machine Learning* integration. Firstly, it is always necessary to investigate a number of algorithms to define the most effective one for the problem considered because the performance of evolutionary algorithms varies significantly for different problems. Secondly, these methods require more computational resources compared with alternative non-evolutionary algorithms. Therefore, in this study we attempt to develop a feature selection technique for classification problems based on a genetic algorithm with these drawbacks removed.

The effectiveness of the proposed approach has been investigated on the *Speech-based Emotion Recognition Problem* which reflects one of the crucial questions in the sphere of human-machine communications [4]. In the experiments conducted a number of multilingual databases (English, German, and Japanese) are used.

The rest of the paper is organized as follows: in Section 2 a brief description of the proposed approach is presented. It includes the details of the evolutionary feature selection scheme and some key points of a multi-objective genetic algorithm (MOGA) and its modification. The speech-based emotion recognition problem and the corpora used are introduced in Section 3. The experiments conducted, the results obtained, and the main inferences are included in Section 4. The conclusion and future work are presented in Section 5.

2 Evolutionary Feature Selection Scheme

We designed the attribute selection approach as a two-criterion optimization model and applied a modified multi-objective genetic algorithm to find solutions. To overcome the disadvantages of the evolutionary search, an island model was used to involve genetic algorithms which were based on different concepts. Moreover, this model allowed us to parallelize calculations and, consequently, to reduce the computational time.

In this Section there is a description of the two-criterion model used and some details about multi-objective genetic algorithms are applied.

2.1 Two-criterion Filter Approach

Feature selection with the filter approach is based on estimating statistical metrics such as *Attribute Class Correlation, Inter- and Intra- Class Distances, Laplasian Score, Representation Entropy* and *the Inconsistent Example Pair measure* which characterize the data set relevancy [5].

In this case we introduce the two-criterion model: *the Intra-class distance (IA)* and *the Inter-class distance (IE)* are used as optimized criteria:

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

$$IE = \frac{1}{n} \sum_{r=1}^{k} n_r d(p_r, p) \rightarrow \max, \qquad (2)$$

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.

As a feature selection technique we use a MOGA operating with binary strings, where *unit*

and *zero* correspond to a relevant attribute and an irrelevant one respectively.

2.2 MOGA and its Cooperative Modification for Feature Selection

The common scheme of any MOGA includes the same steps as any conventional one-criterion genetic algorithm (GA), but in contrast to it, the outcome of a multi-objective algorithm is the set of non-dominated points which form the Pareto set approximation.

Designing a MOGA, researchers are faced with some issues which are referred to fitness assignment strategies, diversity preservation techniques, and ways of elitism implementation. Therefore, at first, in this study we investigated the effectiveness of MOGAs, which were based on various heuristic mechanisms. *Non-Sorting Genetic Algorithm II (NSGA-II)* [6], *Preference-Inspired Co-Evolutionary Algorithm with goal vectors (PICEA-g)* [7], and *Strength Pareto Evolutionary Algorithm 2 (SPEA2)* [8] were used as tools to optimize the introduced criteria (1), (2).

Secondly, we designed a cooperative modification of a MOGA to combine different heuristics based on an island model and, consequently, to avoid the choice of the most effective MOGA. Moreover, at different stages of its execution various heuristics might be advantageous. An island model [9] of a GA implies the parallel work of several algorithms. *The initial number of individuals M is spread across L subpopulations: Mi=M/L, i=1,...,L. At each T-th* generation algorithms exchange the best solutions (migration). There are two parameters: migration size, the number of candidates for migration, and migration interval, the number of generations between migrations. It is also necessary to define the island model topology, in other words, the scheme of migration. We use the fully connected topology that means each algorithm shares its best solutions with all other algorithms included in the island model. NSGA-II, PICEA-g, and SPEA2 are involved as parallel working islands. In the study [10] there is a detailed description of this cooperative multi-objective algorithm.

3 Speech-based Emotion Recognition Problem

One of the obvious ways to improve the intellectual abilities of spoken dialogue systems is that related to their personalization. While communicating, machines should perceive the qualities of the user (as people usually do) such as age, gender and emotions to adapt its answers for the particular speaker.

In this paper we consider one particular aspect of the personalization process that is speech-based emotion recognition. This problem and methods used to solve it lie in the sphere of *affective computing* [11].

Generally, any approach applied to this recognition problem consists of three main stages.

At first, it is necessary to extract acoustic characteristics from the collected utterances. At the INTERSPEECH 2009 Emotion Challenge an appropriate set of acoustic characteristics representing any speech signal was introduced. This set of features comprised attributes such as power, mean, root mean square, jitter, shimmer, 12 MFCCs and 5 formants. The mean, minimum, maximum, range and deviation of the following features have also been used: pitch, intensity and harmonicity. The number of characteristics is 384. To get the conventional feature set introduced at INTERSPEECH 2009, the Praat [12] or OpenSMILE [13] systems might be used. Secondly, all extracted attributes or the most relevant of them should be involved in the supervised learning process to adjust a classifier. At the final stage, the signal that has to be analysed is transformed into an unlabelled feature vector (also with the usage of the Praat or OpenSMILE systems) and then the trained classification model receives it as the input data to make a prediction.

3.1 Corpora Description

In the study a number of speech databases have been used and this Section provides their concise description.

The Emo-DB emotional database (German) [14] was recorded at the Technical University of Berlin and consists of labelled emotional German utterances which were spoken by 10 actors (5 female). Each utterance has one of the following

emotional labels: neutral, anger, fear, joy, sadness, boredom or disgust.

The SAVEE (Surrey Audio-Visual Expressed Emotion) corpus (English) [15] was recorded as a part of an investigation into audio-visual emotion classification from four native English male speakers. The emotional label for each utterance is one of the standard set of emotions (anger, disgust, fear, happiness, sadness, surprise and neutral).

The LEGO emotion database (English) [16] comprises non-acted American English utterances extracted from an automated bus information system of the Carnegie Mellon University at Pittsburgh, USA. The utterances are requests to the Interactive Voice Response system spoken by real users with real concerns. Each utterance is annotated with one of the following emotional labels: angry, slightly angry, very angry, neutral, friendly, and non-speech (critical noisy recordings or just silence). In this study different ranges of anger have been merged into a single class and friendly utterances have been deleted. This pre-processing results in a 3-class emotion classification task.

The UUDB (The Utsunomiya University Spoken Dialogue Database for Paralinguistic Information Studies) database (Japanese) [17] consists of spontaneous Japanese human-human speech. The task-oriented dialogue produced by seven pairs of speakers (12 female) resulted in 4,737 utterances in total. Emotional labels for each utterance were created by three annotators on a five-dimensional emotional basis (interest, credibility, dominance, arousal, and pleasantness). For this work, only the pleasantness and arousal axes are used. The corresponding quadrant (anticlockwise, starting in the positive quadrant, and assuming arousal as abscissa) can also be assigned emotional labels: happy-exciting, angry-anxious, sad-bored and relaxed-serene.

4 Performance Assessment

In the previous research it was found that there was no classification model which demonstrated the highest performance for all of the corpora [18]. Therefore, we decided to combine the developed filter technique with a number of classifiers which showed high effectiveness separately [19]:

- **Support Vector Machine (SMO).** To design a hyperplane separating sets of examples Sequential Minimal Optimization (SMO) is used for solving the large scale quadratic programming problem.

- **Multilayer Perceptron (MLP).** A feedforward neural network with one hidden layer is trained with the error backpropagation algorithm (BP).

- **Linear Logistic Regression (Logit).** This linear model describes the relationship between labels and independent variables using probability scores.

In all experiments the F-score metric [20] was assessed to compare the quality of classification (the more effective the classifier used, the higher F-score value obtained). To derive more statistically significant results, the 6-fold cross-validation procedure was implemented for each database.

Firstly, we obtained the classification results on the full databases (without any feature selection method at all). Table 1 contains the averaged F-score values. It might be noticed that on different corpora various classifiers show the best results. Therefore, it is reasonable to involve several models in the ensemble of classifiers.

Table 1. The classification results on full databases (Experiment 1).

Database	F-score, %		
	MLP	SVM	LOGIT
Emo-DB	80.83	**81.71**	80.46
SAVEE	59.55	59.31	**60.82**
LEGO	68.19	**71.08**	70.71
UUDB	49.34	50.44	**50.88**

Secondly, the effectiveness of the conventional Principal Component Analysis (PCA) as a feature selection procedure was investigated in combination with MLP, SMO, LOGIT. The classification results are presented in Table 2: for each database the first and second rows contain F-score values after the application of PCA with the 0.75 and 0.95 thresholds correspondingly.

Table 2. The effectiveness of PCA (Experiment 2).

Database	The average number of features	F-score, %		
		MLP	SVM	LOGIT
Emo-DB	49.7	74.71	**79.61**	77.04
	136.8	73.87	73.62	76.39
SAVEE	46.7	57.46	57.86	**59.86**
	130.7	50.63	46.18	51.80
LEGO	59.8	67.19	68.05	69.03
	162.5	66.08	70.06	**70.58**
UUDB	46.7	47.53	48.48	49.61
	156.8	47.93	49.37	**49.89**
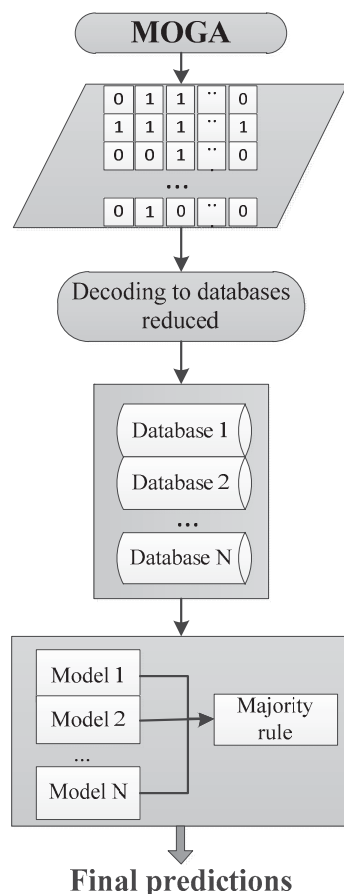
Figure 1. The scheme of the approach proposed (Experiment 3).

While analysing the results in Table 2, we may conclude that for all of the databases the application of PCA leads to the deterioration of the classifier performance. Moreover, for the Emo-DB corpus the highest F-score value after feature selection is lower than even the worst result on the full dataset. These inferences illustrate the necessity of alternative feature selection procedures.

The next experiment was based on the two-criterion model (1), (2). NSGA-II, PICEA-g, and SPEA2 were used as optimizers in combination with MLP, SMO, and Logit classifiers. Generally, for each database this experiment was conducted 9 times (3 MOGAs and 3 classifiers).

As we have noticed, MOGAs return the set of candidate-solutions which cannot be preferred to each other. Taking into account this fact, we have proposed a way to derive the final solution based on the set of non-dominated points. It is assumed that the outcome of the MOGA is N binary strings (the set of non-dominated solutions). Each chromosome should be decoded to the database reduced, according the rule: if a gene is equal to '0' then eliminate the corresponding attribute, and if a gene is equal to '1' then include the respective feature in the database reduced. In short, we obtain N different sets of features and train N various classifiers based on these data. For each test example the engaged models vote for different classes according to their own predictions. The final decision is defined as a collective choice based on the majority rule (Figure 1).

Taking into consideration predictions of several classifiers is a good alternative to choosing one particular solution from the set of non-dominated points. In fact, candidates, which demonstrate high effectiveness on the training data, might often be the worst on the test data. Therefore, to avoid such cases, we use the scheme described.

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

Table 3. The classification results after heuristic feature selection (Experiment 3).

Database	MOGA	The average number of features	F-score, %		
			MLP	SVM	LOGIT
Emo-DB	NSGA-II	165.2	**86.07**	82.77	82.56
	PICEA-g	180.9	84.74	83.06	84.09
	SPEA2	159.5	85.69	82.93	85.64
SAVEE	NSGA-II	163.7	67.70	66.61	**69.84**
	PICEA-g	186.1	68.81	64.80	64.28
	SPEA2	166.7	67.55	64.48	66.82
LEGO	NSGA-II	145.2	71.91	70.45	70.47
	PICEA-g	166.5	71.98	70.47	70.22
	SPEA2	151.2	**72.71**	70.61	70.58
UUDB	NSGA-II	141.5	50.44	50.12	50.50
	PICEA-g	167.5	50.22	50.34	**50.91**
	SPEA2	145.7	50.19	50.37	50.59

selection, uniform recombination and the mutation probability $p_m=1/n$, where n is the length of the chromosome.

In this experiment it has been exposed that (Table 3):

– in most cases the usage of a MOGA for selecting informative features leads to the improvement of F-score values compared with results on full datasets;

– for some classifiers this improvement is significant: from 80.83% to 86.07% for MLP on Emo-DB and from 60.82% to 69.84% for LOGIT on SAVEE;

– minor decreasing of F-score values is not statistically significant (t-test with the significance level $p=0.05$).

However, we may observe that various MOGAs provide the highest F-score values for different classifiers and different databases.

Moreover, for Emo-DB and LEGO the classification models demonstrating the best results on the full dataset (SMO) were outperformed by other models (MLP) after the feature selection procedure. Therefore, it is unlikely to predict the best combination of a classifier and a MOGA.

Consequently, instead of one particular classification model it is reasonable to use the ensemble of MLP, SVM, and LOGIT and the filter approach

allows us to do it. Besides, the usage of the cooperative MOGA based on the island model may eliminate the choice of the effective optimizer.

The scheme of this general approach, which includes the cooperative MOGA and the ensemble of classifiers, is presented in Figure 2.

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. The genetic operators were the same (as in Experiment 3).

Table 4. The cooperative MOGA with an ensemble of classifiers (Experiment 4).

| Database	The average number of features	F-score, %
Emo-DB	166.9	**86.26**(↑ 5.57%)
SAVEE	165.9	**68.71**(↑ 12.97%)
LEGO	150.7	**71.29**(↑ 0.30%)
UUDB	146.9	**51.02**(↑ 0.28%)

The results obtained are presented in Table 4. In addition to the F-score value and the average number of extracted features, for each corpus there is also relative improvement of the F-score metric in comparison with its highest value on the full dataset (percentage in parentheses).
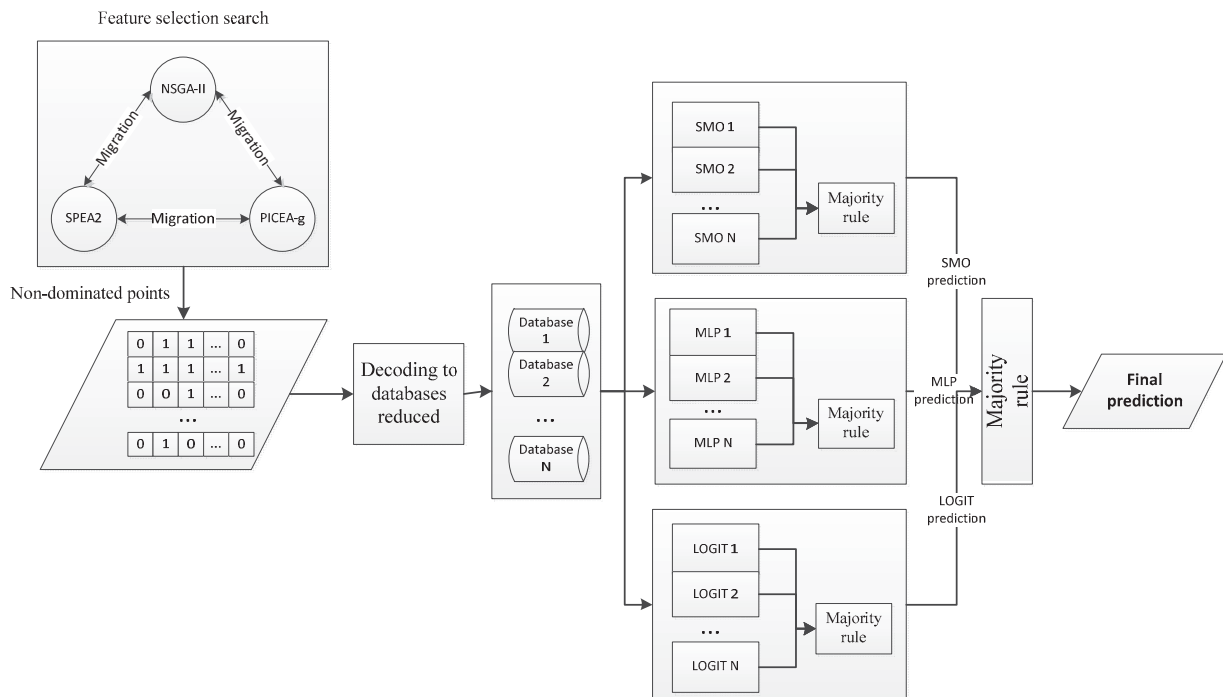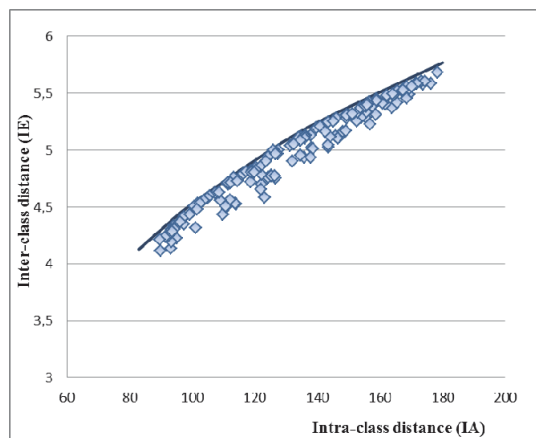
Figure 2. The scheme of the approach based on the cooperative MOGA and the ensemble of classifiers (Experiment 4).
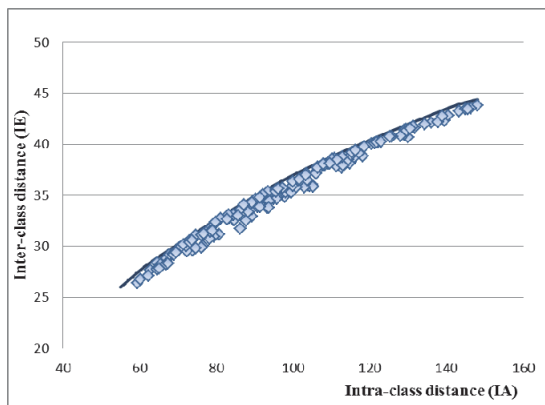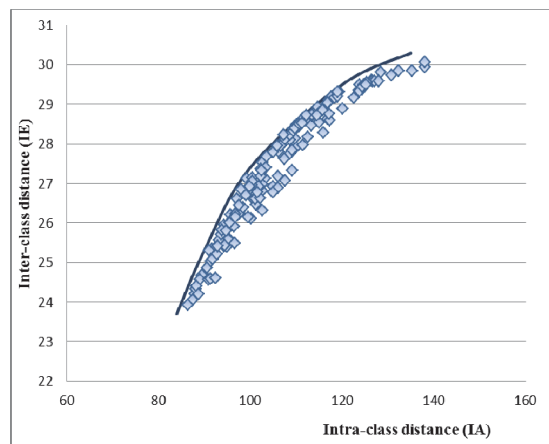

Figure 3. Individuals of the final population in the criterion space (the approximation of the Pareto front) for Emo-DB.
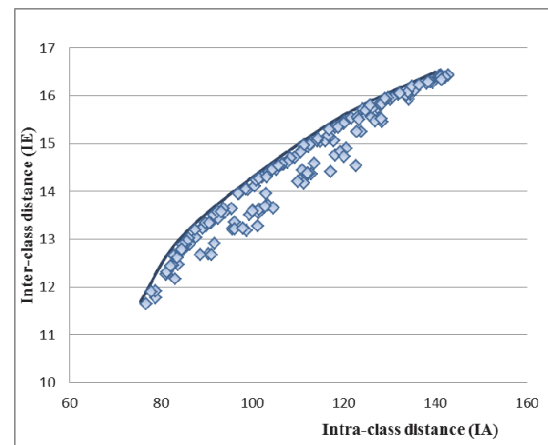
Figure 4. Individuals of the final population in the criterion space (the approximation of the Pareto front) for SAVEE.

Figure 5. Individuals of the final population in the criterion space (the approximation of the Pareto front) for LEGO.

Figure 6. Individuals of the final population in the criterion space (the approximation of the Pareto front) for UUDB.

The results provided with the combination of different MOGAs and the ensemble of classification models were compared with the highest F-score values obtained with one of the MOGAs and one of the classifiers separately (these values are highlighted in Table 3).

A t-test (with the significance level *p=0.01*) was used for comparison. As a result, for all of the corpora there was no difference between the best results obtained in the previous experiment (Table 3) and the F-score values provided with the cooperative MOGA and the ensemble of classifiers.

The application of the proposed approach allowed us not only to achieve the highest F-score values but also to reduce the number of features significantly (approximately by a factor or two).

In Figures 3-6 for each database the individuals of the final population, which is obtained after one of runs, are depicted. In most of the cases we may observe that the approximation of the Pareto front is almost linear.

The usage of the cooperative MOGA allows us not only to avoid the choice of the most effective algorithm for the problem considered, but also to reduce the time spent on its work (the time required for feature selection). On average, for all of the corpora due to the parallel execution of three island model components it became possible to decrease time costs roughly by a factor of 2.55 (the additional time was spent on the migration process), which was especially important for LEGO and UUDB because of their high dimensionality (the huge amount of instances in the dataset). In Table 5 there is the average time spent on optimizing criteria (1), (2) by every MOGA.

**Table 5**. The average time of MOGA executing, sec.

Database	MOGA			
	SPEA2	NSGA-II	PICEA-g	Cooperative MOGA
Emo-DB	121.02	99.02	113.95	44.41
SAVEE	119.42	91.75	105.73	39.61
LEGO	948.57	778.99	988.79	358.00
UUDB	983.70	798.99	982.20	368.03

However, there is one more possibility to reduce the time required for feature selection. In addition to the population, also the sample might be spread across the islands. The training dataset is divided into L parts randomly, where L is the number of components in the island model. Each part is used by one particular algorithm in the cooperative MOGA. During the migration process, islands exchange not only the best individuals, but also parts of the sample: training examples are sent from one island to another. So, at every migration stage all islands should be provided with new training instances (Figure 7). At each L-th migration process the training dataset is divided into L new parts again.

Table 6. Experimental results with the exchange of training examples (Experiment 5).

Database	The average number of features	F-score, %				Time spent on feature selection, sec.
		MLP	SMO	LOGIT	Ensemble	
Emo-DB	162.58	83.26	81.12	82.47	**84.15**	19.50
SAVEE	165.36	63.92	62.79	64.04	**64.84**	20.14
LEGO	141.33	**71.84**	70.44	70.82	71.59	111.14
UUDB	136.48	50.14	50.05	50.69	**50.71**	114.75
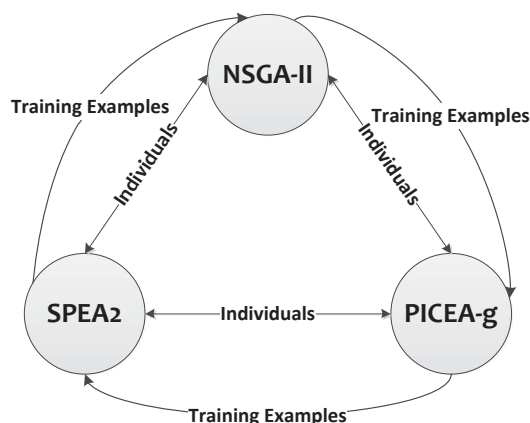
Figure 7. The migration process with the exchange of training examples.

Therefore, in the next experiment we incorporated the subsample exchange technique into the feature selection procedure. At each generation of the cooperative MOGA execution, we had three parts of the training dataset, which were sent from one island to another during migration (new parts were generated during each third migration process).

The cooperative MOGA was provided with the same amount of resources as it was in the previous experiment (Experiment 4). This heuristic feature selection approach was investigated in combination with a number of classifiers (MLP, SMO, LOGIT) and with the ensemble of these models. The results obtained are presented in Table 6.

In this experiment for most of the corpora the highest F-score value was achieved with the ensemble of classifiers (compared with results obtained by these models separately).

A t-test (with the significance level *p=0.01*) was used to compare these results with the F-score values obtained with the ensemble of classifiers in Experiment 4 (without any subsample exchange tech-

nique). As it was revealed, there was no difference in the analysed values.

However, in this experiment we reached the essential decreasing of time costs spent on feature selection. Due to the usage of the subsample exchange technique, we managed to reduce the time spent by a factor of 2.27 for Emo-DB, 1.97 for SAVEE, 3.22 for LEGO and 3.21 for UUDB (in comparison with the time required by the cooperative MOGA without any subsample exchange technique). Therefore, in the framework of the proposed two-criterion feature selection model the parallel work of islands in the cooperative MOGA with the subsample exchange technique might be absolutely beneficial in terms of time costs.

5 Conclusion

In this paper we introduced the evolutionary feature selection technique based on the two-criterion optimization model. In the approach proposed the cooperative MOGA is used as an optimizer. This algorithm includes several heuristics and, therefore, does not require additional experiments to expose the most effective MOGA for the problem considered. Also it saves computational time due to the parallel work of components and the possibility to incorporate the subsample exchange technique. Besides, this approach is effectively used as a pre-processing stage in combination with an ensemble of classifiers.

The thorough investigation of the proposal was accomplished on the set of emotional databases. The experimental results proved that the described evolutionary feature selection technique might be an effective alternative to conventional dimension reduction procedures such as Principal Component Analysis. Generally, a high level of emotion recog-

nition was achieved (up to 12.97% relative improvement for the SAVEE database compared with the best F-score value on the full set of attributes).

Moreover, the promising results prove that it might be reasonable to apply the proposed algorithmic schemes to solve some other problems related to speech-based recognition of human qualities such as gender or speaker identification.

References

[1] R. Kohavi, G.H. John, Wrappers for feature subset selection. Artificial Intelligence, 97, pp. 273-324, 1997.

[2] S.B. Thrun, The Monk's problems: a performance comparison of different learning algorithms, Tech. Rept. CMU-CS-91-197, Carnegie Mellon University, Pittsburgh, PA, 1991.

[3] G.H. John, Enhancements to the data mining process. Ph.D. Thesis, Computer Science Department, Stanford University, CA, 1997.

[4] M. Venkatadri, K. Srinivasa Rao, A multiobjective genetic algorithm for feature selection in data mining, International Journal of Computer Science and Information Technologies, vol. 1, no. 5, 2010, pp. 443–448.

[5] Ch. Brester, M. Sidorov, E. Semenkin, Acoustic Emotion Recognition: Two Ways of Feature Selection Based on Self-Adaptive Multi-Objective Genetic Algorithm, Proceedings of the International Conference on Informatics in Control, Automation and Robotics (ICINCO), 2014, pp. 851-855.

[6] K. Deb, A. Pratap, S. Agarwal, T. Meyarivan, A fast and elitist multiobjective genetic algorithm: NSGA-II, IEEE Transactions on Evolutionary Computation 6 (2), 2002, pp. 182-197.

[7] R. Wang, Preference-Inspired Co-evolutionary Algorithms, A thesis submitted in partial fulfillment for the degree of the Doctor of Philosophy, University of Sheffield, 2013, p. 231.

[8] E. Zitzler, M. Laumanns, L. Thiele, SPEA2: Improving the Strength Pareto Evolutionary Algorithm for Multiobjective Optimization, Evolutionary Methods for Design Optimisation and Control with Application to Industrial Problems EUROGEN 2001 3242 (103), 2002, pp. 95-100.

[9] D. Whitley, S. Rana, and R. Heckendorn, Island model genetic algorithms and linearly separable problems, Proceedings of AISB Workshop on Evolutionary Computation, Manchester, UK. Springer, volume 1305 of LNCS, 1997, pp. 109-125.

[10] Ch. Brester, E. Semenkin, Cooperative Multi-objective Genetic Algorithm with Parallel Implementation // Advances in Swarm and Computational Intelligence, LNCS 9140, 2015, pp. 471–478.

[11] R.W. Picard, Affective computing. Tech. Rep. Perceptual Computing Section Technical Report No. 321, MIT Media Laboratory, 20 Ames St., Cambridge, MA 02139, 1995.

[12] P. Boersma, Praat, a system for doing phonetics by computer, Glot international, vol. 5, no. 9/10, 2002, pp. 341–345.

[13] F. Eyben, M. Wllmer, and B. Schuller, Opensmile: the Munich versatile and fast opensource audio feature extractor, Proceedings of the international conference on Multimedia, 2010. ACM, pp. 1459–1462.

[14] F. Burkhardt, A. Paeschke, M. Rolfes, W. F. Sendlmeier, and B. Weiss, A database of german emotional speech, In Interspeech, 2005, pp. 1517–1520.

[15] S. Haq, P. Jackson, Machine Audition: Principles, Algorithms and Systems, chapter Multimodal Emotion Recognition, IGI Global, Hershey PA, Aug. 2010, pp. 398–423.

[16] A. Schmitt, S. Ultes, and W. Minker, A parameterized and annotated corpus of the cmu let's go bus information system, Proceedings of International Conference on Language Resources and Evaluation (LREC), 2012.

[17] H. Mori, T. Satake, M. Nakamura, and H. Kasuya, Constructing a spoken dialogue corpus for studying paralinguistic information in expressive conversation and analyzing its statistical/acoustic characteristics, Speech Communication, 53, 2011.

[18] Ch. Brester, M. Sidorov, E. Semenkin, Speech-based emotion recognition: Application of collective decision making concepts, Proceedings of the 2nd International Conference on Computer Science and Artificial Intelligence (ICCSAI2014), 2014, pp. 216–220.

[19] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I. H. Witten, The WEKA Data Mining Software: An Update, SIGKDD Explorations, Vol. 11, Issue 1, 2009.

[20] C. Goutte, E. Gaussier, A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. ECIR'05 Proceedings of the 27th European conference on Advances in Information Retrieval Research, 2005, pp. 345–359.

Christina Brester completed her Bachelor's (2012) and Master's degree (2014) in systems analysis and control at the Reshetnev Siberian State Aerospace University (SibSAU) in Krasnoyarsk, Russia. Her Master's thesis was on Adaptive Multi-objective Genetic Algorithms. Currently, she is a PhD student in Computer Science and Engineering at SibSAU. Her research interests include evolutionary computation, neuro-evolutionary algorithms, machine learning and speech analysis.

Eugene Semenkin received his Master in Applied Mathematics degree from Kemerovo State University (Kemerovo, USSR) in 1982, his PhD in Computer Science from Leningrad State University (Leningrad, USSR) in 1989 and his DSc in Engineering and Habilitation from the Siberian State Aerospace University (Krasnoyarsk, Russia) in 1997. Since 1997, he has been a professor of sys-tems analysis at the Institute of Computer Science and Telecommunications of the Siberian State Aerospace University. His areas of research include the modelling and optimization of complex systems, computational intelligence and data mining. He has been awarded the Tsiolkovsky Badge of Honour by the Russian Federal Space Agency and the Reshetnev medal by the Russian Federation of Cosmonautics.

Maxim Sidorov has been a research assistant and PhD student at the University of Ulm (Germany) since 2012. In 2010 he received the Bachelor of Science degree in applied mathematics and computer science from the Siberian Federal University, Krasnoyarsk, Russia. In 2012 he graduated and obtained the Master of Science degree in systems analysis and control from the Reshetnev Siberian State Aerospace University, Krasnoyarsk, Russia. His areas of scientific interest include intellectual information technologies, data mining, machine learning and its applications, and dialogue technology.