Self-Configuring Ensemble of Neural Network Classifiers for Emotion Recognition in the Intelligent Human-Machine Interaction

Evgenii Sopov[1] and Ilia Ivanov[2]
Department of Systems Analysis and Operations Research
Siberian State Aerospace University
Krasnoyarsk, Russia
evgenysopov@gmail.com[1], ilyaiv92@gmail.com[2]

***Abstract*—Reducing the dimensionality of datasets and configuring learning algorithms for solving particular practical tasks are the main problems in machine learning. In this work we propose multi-objective optimization approach to feature selection and base learners hyper-parameter optimization. The effectiveness of the proposed multi-objective approach is compared to the single-objective approach. We have chosen emotion recognition problem by audio-visual data as a benchmark for comparing the two mentioned approaches. We have chosen neural network as a base learning algorithm for testing the proposed approach to parameter optimization. As a result of multi-objective optimization applied to parameter configuration we get the Pareto set of neural networks with optimal parameter values. In order to get the single output, the Pareto optimal neural networks were combined into an ensemble. We have examined several ensemble model fusion techniques including voting, average class probabilities and meta-classification. According to results, multi-objective optimization approach to feature selection provides an average 2.8% better emotion classification rate on the given datasets than single-objective approach. Multi-objective approach is 5.4% more effective compared to principal components analysis, and 13.9% more effective compared to not using any dimensionality reduction at all. Multi-objective approach applied to neural networks parameter optimization provided on average 7.1% better classification rate than single-objective approach. The results suggest that the proposed multi-objective optimization approach is more effective at solving considered emotion recognition problem.**

I. INTRODUCTION

Today machine learning and knowledge discovery in data lie at the core of intelligent systems design. Choosing the proper machine learning algorithm for particular problem is crucial for building an effective system. Another key point is configuring the chosen algorithm's parameters to achieve the best performance.

There are many ways to configure the learning algorithm parameters. The simplest one is manual configuring, but it's reasonable only when there are a few parameters and there is useful information presented about problem features. For majority of algorithms, this is not the case, that is why manual configuring becomes unsuitable, and other techniques come in handy.

The traditional approaches to algorithm parameter optimization include grid search, bayesian optimization, random search and gradient optimization. The more sophisticated approaches involve using of optimization algorithms for choosing the optimal sets of parameter values.

In this work, we propose the multi-objective optimization approach to learning algorithms parameter optimization, using multi-objective genetic algorithms for designing neural networks. We compare the proposed approach with single-objective optimization approach on the emotion recognition benchmark problem.

In single-objective problem statement, the parameters of neural network, such as number of neurons and number of learning iterations, are used as input parameters, and emotion classification rate is an objective to be optimized. The result of the optimization is the optimal set of neural network parameter values. In multi-objective problem statement, number of neurons is introduced as a second objective that is minimized. As a result of optimization a Pareto set of optimal sets of parameter values is obtained. The neural networks with these effective parameter values are combined into an ensemble. Several schemes of ensemble classifiers output fusion are examined: voting, average class probabilities, meta-classification by support vector machine (SVM) classifier.

Another key aspect of solving any machine learning problem is feature selection and dimensionality reduction. The models that are based on less amount of data features are generally more simple and robust, provide better generalization and require less computational resources. There are two main ways of dimensionality reduction in machine learning called feature transformation and feature selection.

Feature transformation methods, as the name states, transform the set of available features to the set of other

978-1-4799-7560-0/15 $31.00 © 2015 IEEE
DOI 10.1109/SSCI.2015.252

features. The number of the transformed features must be smaller to ensure the reduction of feature space dimensionality. The popular method that uses this technique is principal components analysis (PCA).

Feature selection methods select the subset of the most important features out of the initial set of features. The more information the feature contains for predicting the dependent variable, the more its importance. The methods of this class usually require an optimization procedure to find an optimal subset of features. Researchers take the model effectiveness (usually classification rate) as an optimization criterion. Such approach to feature selection uses single-objective optimization. In this work we propose a multi-objective optimization approach to feature selection. We add the second optimization criterion - the number of selected features that is to be minimized. Adding the second criterion strengthens the effect of feature space reduction. We compare single-objective and multi-objective optimization approaches to feature selection to each other, to PCA method and to the variant of not using dimensionality reduction at all.

Emotion recognition problem is being used in this work as a benchmark for comparing the mentioned approaches. Emotion recognition is the relevant part of computer-machine interaction (HMI). Today the major companies that provide technological devices have to think more about the interface between the machine and the user. In order to build a successful interface we need to teach machines responding to human actions in an intelligent way. In order to do so, machines need to collect as much information about the human user as possible. The collected information includes gender, age, emotional state etc. Much research has been done on building intelligent dialogue systems (DS) that are able to collect this kind of information.

Emotion recognition is one of the most challenging parts of the global task of building an effective DS. The challenges occur because human emotions are usually hidden. Emotions also tend to change fast. Nevertheless, in this work we make an attempt of automatically classifying person's emotional state based on the voice and facial video image. We chose neural network as a classification algorithm, because neural networks proved to be especially effective in image classification.

The rest of the paper is organized as follows. Section 2 provides insight into significant related work on state-of-the-art multi-objective optimization and its application to base learner algorithms parameter optimization. The adopted methodology can be found in Section 3. Section 4 describes the emotion recognition problem. Experimental results are presented and discussed in Section 5. Finally, conclusion and further work can be found in Section 6.

II. Significant Related Work

The approach proposed in this work uses a multi-objective optimization procedure. We will give insight into the state-of-the-art multi-objective optimization algorithms. The described algorithms proved their effectiveness over time and remain a strong point of reference in many scientific publications.

Strength Pareto Evolutionary algorithm (SPEA) is one of the state-of-the-art multi-objective optimization (MO) algorithms proposed by Zitzler [1]. It directly employs the idea of Pareto dominance in order to find the population of non-dominated points that would approximate Pareto front. An internal clustering procedure is applied to encourage population points diversity. Non-dominated Sorting Genetic algorithm-2 (NSGA-2) is another efficient MO algorithm developed by Deb [2]. This is an upgrade of his previous proposed algorithm, NSGA. At the core of this algorithm lies the idea of sorting the population of points according to their mutual non-dominance, first the non-dominated points in the population are found, their ranks are assigned to one. These points are removed from the population, the next subset of non-dominated points is found, their ranks are assigned to two, those points are removed from the population. This procedure continues until all population members are ranked, after that go the regular genetic operators - selection, crossover, mutation, elitism. Also, the classic Vector Evaluated Genetic algorithm (VEGA) proposed by Schaffer [3] was included in the experiments framework, as this MO algorithm provides good baseline accuracy when compared to other state-of-the-art MO algorithms on benchmark problems.

The idea of using optimization algorithms for tuning the learning algorithm parameters has already been investigated by a number of authors. The classic approaches use gradient based optimization algorithms. But when we have no information about the objective, classic methods demonstrate low performance. In this case the only option is to use the search algorithms that do not require the information about the objective.

The greatest number of research was conducted in the field of neural networks parameter optimization. Bergstra et. al. [4] in their work showed that random search is statistically better for neural network and deep belief network hyper-parameter optimization than grid search and manual search. They claim that for most datasets that they used only a few hyper-parameters really mattered to the resulting accuracy of the algorithm, and for different datasets those hyper-parameters also differed. This phenomenon makes grid search a poor choice for configuring algorithms for new datasets.

Larochelle et. al. [5] used greedy layer-wise procedure to train deep multi-layer neural networks. The authors split the process of network parameter tuning into two phases. In the first phase the network parameter subsets corresponding to distinct network layers are trained using an unsupervised learning criterion. In the second phase, all network parameters are tuned using back propagation and gradient descent on a global supervised cost function. The network parameters are initialized with values obtained in the first phase.

In [6] the authors used a genetic algorithm to search for an accurate and diverse set of trained networks. First, they create and train the initial population of networks, then use genetic operators to create new networks. The diversity of each network with respect to the current population is measured according to the dispersion of each distinct network output and the output of current population

ensemble. The fitness of each network is calculated as the weighted sum of accuracy and diversity estimates. The final population of networks is combined into an ensemble, using the weighted sum output fusion, where each network's weight is proportional to its accuracy.

Smith et. al. in [7] proposed a hybrid multi-objective evolutionary algorithm for optimizing the structure of recurrent neural networks for time series prediction. They use several methods for selecting individuals from the obtained Pareto set. The first method selects all individuals below a threshold in the Pareto front, the second one is based on the training error. Individuals near the knee point of the Pareto front are also selected, and finally the individuals are selected based on the diversity of individual predictors. The authors claim that such hybrid approach to selecting a subset of optimal neural networks outperforms the first and the second approaches when used separately.

Feature selection is the preprocessing step in machine learning. Feature selection is mostly spread in machine learning problems with a very high number of attributes. It is performed for several reasons. Model simplification, less amount of computational resources and increased model generalization by reducing model variance. For complex machine learning problems feature selection becomes a computationally expensive task. For that reason researchers use optimization algorithms for finding the globally optimal subset of features. Feature selection methods that use optimization procedures are called metaheuristic methods. Metaheuristic methods are divided into three classes based on how they combine the selection algorithm and the model construction: filter methods, wrapper methods and embedded methods. Filter methods select features regardless of the model. They take into account only general notions like the correlation of the attribute and the dependent variable. Wrapper methods evaluate feature subsets [16]. This makes possible to detect interactions between variables, but increases the computation time. Finally, in embedded systems the learning algorithm includes its own variable selection algorithm [17]. This reduces computational cost, but in this case the learning algorithm need to know what a good selection is beforehand.

Emotion recognition problem has been researched by a number of authors. Here we give an insight into some of them. The paper by Rashid et al. [11] explores the problem of human emotion recognition and proposes the solution of combining audio and visual features. First, the audio stream is separated from the video stream. Feature detection and 3D patch extraction are applied to video streams and the dimensionality of video features is reduced by applying PCA. From audio streams prosodic and mel-frequency cepstrum coefficients (MFCC) are extracted. After feature extraction, the authors construct separate codebooks for audio and video modalities by applying the K-means algorithm in Euclidean space. Finally, multiclass support vector machine (SVM) classifiers are applied to audio and video data, and decision-level data fusion is performed by applying Bayes sum rule. By building the classifier on audio features the authors received an average accuracy of 67.39%, using video features gave an accuracy of 74.15%, while combining audio and visual features on the decision level improved the accuracy to 80.27%.

Kahou et al. [12] described the approach they used for submission to the 2013 Emotion Recognition in the Wild Challenge. The approach combined multiple deep neural networks including deep convolutional neural networks (CNNs) for analyzing facial expressions in video frames, deep belief net (DBN) to capture audio information, deep autoencoder to model the spatio-temporal information produced by the human actions, and shallow network architecture focused on the extracted features of the mouth of the primary human subject in the scene. The authors used the Toronto Face Dataset, containing 4,178 images labelled with basic emotions and with only fully frontal facing poses, and a dataset harvested from Google image search which consisted of 35,887 images with seven expression classes. All images were turned to grayscale of size 48x48. Several decision-level data integration techniques were used: averaged predictions, SVM and multi-layer perceptron (MLP) aggregation techniques, and random search for weighting models. The best accuracy they achieved on the competition testing set was 41.03%.

In the work by Cruz et al. [13] the concept of modelling the change in features is used, rather than their simple combination. First, the faces are extracted from the original images, and Local Phase Quantization (LPQ) histograms are extracted in each n by n local region. The histograms are concatenated to form a feature vector. The derivative of features is computed by two methods: convolution with the difference of Gaussians (DoG) filter and the difference of feature histograms. A linear SVM is trained to output posterior probabilities. and the changes are modelled with a hidden Markov model. The proposed method was tested on the Audio/Visual Emotion Challenge 2011 dataset, which consists of 63 videos of 13 different individuals, where frontal face videos are taken during an interview where the subject is engaged in a conversation. The authors claim that they increased the classification rate on the data by 13%.

In [14] the authors exploit the idea of using electroencephalogram, pupillary response and gaze distance to classify the arousal of a subject as either calm, medium aroused, or activated and valence as either unpleasant, neutral, or pleasant. The data consists of 20 video clips with emotional content from movies. The valence classification accuracy achieved is 68.5 %, and the arousal classification accuracy is 76.4 %.

Busso et al. [15] researched the idea of acoustic and facial expression information fusion. They used a database recorded from an actress reading 258 sentences expressing emotions. Separate classifiers based on acoustic data and facial expressions were built, with classification accuracies of 70.9% and 85% respectively. Facial expression features include 5 areas: forehead, eyebrow, low eye, right and left cheeks. The authors covered two data fusion approaches: decision level and feature level integration. On the feature level, audio and facial expression features were combined to build one classifier, giving 90% accuracy. On the decision level, several criteria were used to combine posterior probabilities of the unimodal systems: maximum – the emotion with the greatest posterior probability in both modalities is selected; average –the posterior probability of each modality is equally weighted and the maximum is selected; product – posterior probabilities are multiplied and

the maximum is selected; weight– different weights are applied to the different unimodal systems. The accuracies of decision-level integration bimodal classifiers range from 84% to 89%, product combining being the most efficient.

III. METHODOLOGY

In this work we propose the application of multi-objective approach to classification algorithms parameter optimization and to feature selection.

The proposed multi-objective optimization approach to feature selection belongs to the class wrapper methods. It was compared to principal components analysis method and single-objective optimization approach to feature selection. We designed the optimization based feature selection methods in the following way. The input variables were constructed as binary vectors of length m, where m is the initial number of dataset features. Each bit of such a binary vector takes value of either 1 or 0, where 1 means that the corresponding feature is selected for further use, and 0 means that it is not used, respectively. In single-objective problem statement, classification rate serves as a maximization criterion. Classification rate is defined as follows:

$$R = (N_c/N) \times 100\ \% \quad (1)$$

where N_c is the number of correctly classified instances, N is the total number of dataset instances, R is the classification rate.

In multi-objective problem statement, we add the second criterion - the number of selected features. This criterion is to be minimized. The idea behind this is that the models based on smaller amount of features are usually simpler, thus generally more preferable. Support vector machine algorithm was chosen as a classification algorithm.

We selected the class of evolutionary algorithms to solve optimization tasks in our feature selection problem statement. This choice was made because evolutionary algorithms proved to be good at locating the global optimum. They are particularly helpful when there is no information about the surface of the optimized function. In the case of feature selection problem we do not have any prior information regarding the dependency between the effectiveness and the subset of selected features. That is why we consider evolutionary algorithms a good choice.

We used co-evolutionary genetic algorithm (GA) in single-objective optimization approach for feature selection. This algorithm combines several standard GAs with different parameter values. These standard GAs work in parallel on the same optimization problem. After every fixed number of iterations GAs change individual solutions among each other, and save the best solutions. This workflow helps to find global optimum without having to configure GAs parameters explicitly. GAs effectiveness highly depends on choosing the optimal parameter values for each distinctive problem. We have used the co-evolution scheme to perform self-tuning of the GA parameters.

In multi-objective optimization approach for feature selection SPEA algorithm was used. Optimization algorithms parameter values are presented in Table 1. We indicate crossover and mutation probability as low. The quantitative values of these probabilities are calculated by the following formula:

$$p = 1/(k \times |P|) \quad (2)$$

where $k=3$ for our experiments, but may take on another integer value, $|P|$ is the population size, p is the probability value.

TABLE I. GENETIC ALGORITHMS PARAMETER VALUES

Genetic algorithm parameter	Value
Population size	50
Number of iterations	50
Crossover probability	Low
Crossover type	Uniform
Mutation probability	Low
Maximum size of external set (SPEA)	50
Adaptation interval	5
Penalty (% of population size)	10%
Minimal guaranteed population size (% of initial population size)	10%

The multi-objective optimization approach was also applied to neural networks parameter optimization. We compared it to the single-objective optimization approach.

We chose feed-forward neural network as a classification algorithm for several reasons. First, neural networks proved to be successfully applied to real-world image analysis problems. Second, neural networks are sensitive to parameter configuration, so this is a good chance to test the effectiveness of the proposed approaches. We used a single-layer neural network with variable number of neurons. The activation function type was sigmoid.

In single-objective optimization approach the input variables include the overall number of network neurons and the number of iterations for network training. The input variables vary in the following borders: number of network neurons $N_n = [2, 50]$, number of network training iterations $N_t = [2{,}200]$. The classification rate obtained by the corresponding network serves as a maximization criterion.

The multi-objective approach is quite similar in formulation, the difference is that the second optimization criterion is added - the number of network neurons. This second criterion is equal to the first input variable, and it is to be minimized, because neural networks with less neurons are simpler, thus more preferable.

The class of evolutionary algorithms was selected for solving the optimization problems formulated above. For single-objective optimization we used co-evolutionary GA. For multi-objective optimization four different algorithms

were used: SPEA, NSGA-II, VEGA, and Self-configuring Multi-objective Genetic algorithm (SelfCOMOGA) [8].

The SelfCOMOGA combines the advantages of SPEA, NSGA-II and VEGA optimizers. These optimizers work in parallel and share resources after every fixed number of iterations. Also non-dominated sorting procedure is performed in order to locate optimal solutions. The SelfCOMOGA is a hybrid of the island model in GA, competitive and cooperative coevolution schemes. It's main conception is as follows.

The total population is divided into disjoint subpopulations of equal size. The portion of the population is called the computational resource. Each subpopulation corresponds to certain multi-objective GA and evolves independently (corresponds to the island model). After a certain period, which is called the adaptation period, the performance of individual algorithms is estimated and the computational resource is redistributed (corresponds to the competitive coevolution). Finally, random migrations of the best solutions are presented to equate start positions of GAs for the run with the next period (corresponds to the cooperative coevolution).

The following criteria are used for estimating performance of a single algorithm in the SelfCOMOGA. The first group includes the static criteria (the performance is measured over the current adaptation period). Criterion 1 is the percentage of non-dominated solutions. Criterion 2 is the uniformity (dispersion) of non-dominated solutions. The second group contains the dynamic criteria (the performance is measured in a comparison with previous adaptation periods). Criterion 3 is the improvement of non-dominated solutions. The solutions of the previous and current adaptation period are compared. More detailed information on the SelfCOMOGA can be found in [8].

In the single-objective optimization approach as a result we get the single neural network with optimal parameters. Whereas in the multi-objective approach we get the Pareto set of neural networks with optimal parameters. In order to provide a single output, and make possible the comparison of single and multi-objective approaches, the Pareto set neural networks obtained as a result of multi-objective optimization were combined into an ensemble. Three ensemble classifiers output fusion schemes were applied to compare their effectiveness:

- Voting;
- Average class probabilities - posterior class probabilities for each class are averaged over all ensemble classifiers;
- SVM meta-classification – training dataset is divided into two parts, the first part is used to train the ensemble classifiers. The output posterior class probabilities of all ensemble classifiers are treated as input variables, and the second part of the training dataset is used to train an auxiliary SVM meta-classifier, which outputs the resulting class label.

We implemented all optimization algorithms mentioned in this section in C# programming language. Also we used an R implementation of neural network, support vector machine and principal component analysis algorithms.

IV. Emotion Recognition Benchmark Problem

Emotion recognition problem was used as a benchmark classification problem for comparing the approaches to dimensionality reduction and learning algorithms parameter optimization described above.

SAVEE emotion database was used as a source of raw data used for solving the problem. The database includes 480 video recordings of 4 male speakers reading a predefined set of phrases imitating 7 basic emotions: anger, disgust, fear, happiness, neutral, sadness, surprise. Database emotion classes distribution is given in Fig. 1.

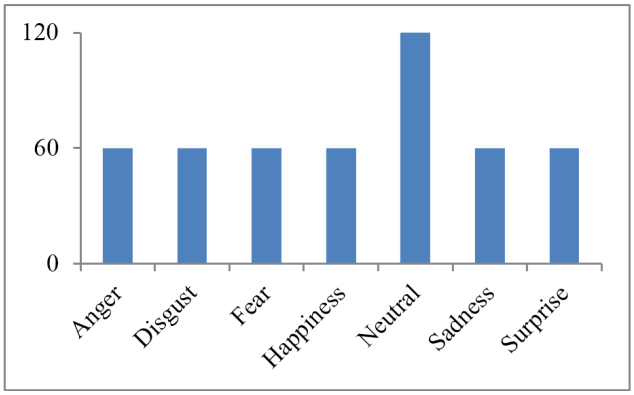

Fig. 1. Database emotion classes distribution

In order to build quantitative models, audio and video features have been extracted from raw audio-video recordings. Audio features were extracted with openSMILE - open source software for audio and visual features extraction [9]. Video features were extracted using 3 algorithms:

- Quantized Local Zernike Moments (QLZM) [10];
- Local Binary Patterns (LBP) [18];
- Local Binary Patterns on Three Orthogonal Planes (LBP-TOP) [19].

QLZM and LBP-TOP algorithms for video features extraction were used because they are state-of-the-art and were successfully applied to image analysis problems by other researchers. LBP algorithm was selected because it is a classic algorithm of image processing, and serves as a good lower bound effectiveness estimate.

QLZM and LBP algorithms extract features from every single video frame in a video sequence, whereas LBP-TOP deals with spatio-temporal space which includes several consecutive frames. Feature vectors obtained with QLZM and LBP were averaged over the whole video sequence. Extracted audio and video features were combined into single overall dataset in order to check if combining audio and visual information helps to improve the classification rate. The number of extracted audio and video features can be found in Table 2. Audio features extracted by openSMILE software are labeled as "audio" in the table.

The combined audio-visual dataset containing features extracted by openSMILE, QLZM, LBP and LBP-TOP is labeled "audio+video". As can be seen, the number of features in the constructed datasets is quite high, which makes the dimensionality reduction procedure reasonable.

V. Experimental Results

The approaches to feature selection and learning algorithms parameter optimization described in section 3, were applied at solving the emotion recognition problem described in section 4.

Results of dimensionality reduction research can be found in Table 2. Three approaches were compared - PCA, feature selection by means of single-objective optimization and multi-objective optimization, as well as using the unreduced initial number of features. As can be observed, the multi-objective optimization approach to feature selection provided better classification rate on QLZM, LBP-TOP, audio, and audio-visual dataset (4 out of 5 datasets), while losing 0.3% to single-objective optimization approach on LBP dataset. These experimental results prove the effectiveness of the proposed multi-objective approach.

TABLE II. EMOTION CLASSIFICATION RATE (%) FOR VARIOUS DATASETS AND DIMENSIONALITY REDUCTION APPROACHES USING SUPPORT VECTOR MACHINE

Dataset	Number of features	Classification rate / reduced number of features			
		All features	Principal components analysis	Feature selection	
				Single-objective optimization	Multi-objective optimization
QLZM	656	10.506	21.458 / 36	20.208 / 301	**24.911** / 319
LBP-TOP	177	22.847	32.017 / 10	40.278 / 77	**45.694** / 90
LBP	59	20.486	23.75 / 4	**25.972** / 33	25.694 / 31
Audio	991	28.542	35.923 / 131	38.095 / 476	**39.702** / 484
Audio + video	1883	19.732	31.718 / 180	33.661 / 902	**35.893** / 885

The experiments on neural network parameter optimization were conducted for every available dataset, various optimization algorithms, and various classifiers output fusion schemes. The example of the obtained Pareto optimal set of neural networks is shown in Table 3. Table 4 presents the summarized results on neural networks parameter optimization. As can be seen, combining the obtained Pareto optimal networks from Table 3 to an ensemble resulted in the increase of effectiveness up to 39.76%.

The results suggest that multi-objective optimization approach to neural network parameter optimization applied to emotion recognition problem is more effective than single-objective optimization approach, because it outperforms it on all five datasets. We cannot give certain advice regarding which multi-objective optimization algorithm to use, because all of them provided the best classification rates on different datasets. SVM meta-classifier output fusion scheme seems to be the most effective technique of aggregating ensemble classifiers outputs, because it provided the best results on four out of five datasets. The fact that the most effective optimization algorithms differ for different datasets proves once more that SVM-meta classification fusion scheme is invariant to optimization algorithm used, thus being the robust approach.

TABLE III. EXAMPLE OF PARETO OPTIMAL SET OF NEURAL NETWORKS IN MULTI-OBJECTIVE PROBLEM STATEMENT

No.	Number of neurons	Number of training iterations	Classification rate
1	10	119	10.88
2	12	20	29.49
3	13	113	30.34
4	24	150	33.38
5	14	73	33.69
6	11	119	15.88
7	29	100	35.38
8	39	144	32.02
9	15	51	15.44
10	23	74	27.89

We summarize the results in Table 5. It contains information about the best achieved emotion classification rates as well as the data and methodologies that were used to obtain them. Performing feature selection by multi-objective optimization approach on LBP-TOP dataset provided the best classification rate of 45.7 %. The baseline model that predicts the most frequent class in the dataset for all instances provides the rate of 25 %. Taking into account the complexity of the emotion recognition problem, we consider that the results are successful.

VI. Conclusions

The problems of dimensionality reduction, feature selection, and learning algorithms parameter configuration stay one of the most important issues in machine learning problems and applications.

In this work we proposed the multi-objective optimization approach to feature selection, and to neural networks parameter optimization. The proposed approach was tested on emotion recognition problem.

According to the obtained results, multi-objective optimization approach applied to feature selection provided higher classification rate than single-objective optimization approach by 2.8% on average over various datasets. We also found out that multi-objective approach is 5.4% more effective than PCA algorithm, and 13.9% more effective than not doing dimensionality reduction at all. According to the obtained results the multi-objective optimization approach to feature selection is the most effective approach for the emotion recognition problem. Our advice is to use it in further works on emotion recognition. The proposed approach may also be helpful in other associated machine learning tasks.

We also applied multi-objective optimization approach to neural networks parameter optimization. The obtained results prove that the ensembles of neural networks with Pareto optimal parameters provide better classification rate than the single optimal neural network found by single-objective optimization approach. The average difference in effectiveness is 7.1% in favor of multi-objective approach. Our advice is to use SVM meta-classification fusion scheme because it provided the best results on 4 datasets out of 5. But further research needs to be done in order to prove this fusion scheme effectiveness. Also, we plan to try other ensemble output fusion schemes in order to compare them to the existing ones and draw conclusions.

ACKNOWLEDGMENT

The research was supported by President of the Russian Federation grant (MK-3285.2015.9).

TABLE IV. CLASSIFICATION RATE (%) FOR VARIOUS EMOTION RECOGNITION PROBLEM STATEMENTS

Optimization Algorithm (number of objectives)	Ensemble Classifiers Output Fusion Scheme	Data				
		Audio	QLZM	LBP	LBP-TOP	Audio + video
Co-evolutionary GA (1)	-	35.923	21.458	23.75	32.917	31.718
SPEA (2)	Voting	31.012	16.319	16.667	34.167	27.292
	Average class probabilities	16.994	10.903	16.458	**39.583**	14.256
	SVM meta-classifier	28.631	16.042	18.264	34.583	25.06
NSGA-2 (2)	Voting	29.226	21.181	19.236	33.403	24.554
	Average class probabilities	29.435	14.722	16.667	17.639	23.571
	SVM meta-classifier	**39.762**	11.528	17.5	38.125	34.94
VEGA (2)	Voting	33.839	17.5	24.514	32.639	22.5
	Average class probabilities	27.262	24.306	20.069	21.042	15.119
	SVM meta-classifier	38.899	13.958	29.167	36.736	**37.292**
SelfCOMOGA (2)	Voting	26.577	20.347	33.125	36.25	19.94
	Average class probabilities	23.244	15.935	25.417	22.708	17.768
	SVM meta-classifier	36.518	**26.756**	**38.333**	36.319	29.405

TABLE V. THE SUMMARY RANKING OF METHODOLOGIES AND DATA THAT ACHIEVE THE BEST EMOTION CLASSIFICATION RATE

Rank	Methodology	Data	Classification rate, %
1	Feature selection, multi-objective optimization	LBP-TOP	45.694
2	Neural network optimization, NSGA-2, SVM meta-classifier fusion scheme	Audio	39.762
3	Feature selection, multi-objective optimization	Audio	39.702
4	Neural network optimization, SPEA, average class probabilities fusion scheme	LBP-TOP	39.583
5	Neural network optimization, SelfCOMOGA, SVM meta-classifier fusion scheme	LBP	38.333
6	Neural network optimization, VEGA, SVM meta-classifier fusion scheme	Audio + video	37.292
7	Feature selection, multi-objective optimization	Audio + video	35.893

REFERENCES

[1] E. Zitzler and L. Thiele, "An evolutionary algorithm for multiobjective optimization: the strength Pareto approach," Swiss Federal Institute of Technology, Zurich, Switzerland, TIK-Report No. 43, May 1998.

[2] K. Deb, A. Pratap, S. Agarwal and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," IEEE Trans. on Evolutionary Computation, Vol. 6, No. 2, April 2002.

[3] J. D. Schaffer, "Multiple objective optimization with vector evaluated genetic algorithms," Proc. of the 1st International Conference on Genetic Algorithms, 1985, pp. 93-100.

[4] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," Journal of Machine Learning Research 13, 2012, pp. 281-305.

[5] H. Larochelle, Y. Bengio, J. Louradour and P. Lamblin, "Exploring strategies for training deep neural networks," Journal of Machine Learning Research 1, 2009, pp. 1-40.

[6] D. W. Opitz and J. W. Shavlik, "Generating accurate and diverse members of a neural-network ensemble," 1996.

[7] C. Smith and Y. Jin, "Evolutionary multi-objective generation of recurrent neural network ensembles for time series prediction," Neurocomputing, Vol. 143, 2, November 2014, pp. 302-311.

[8] I. Ivanov and E. Sopov, "Design Efficient Technologies for Context Image Analysis in Dialog HCI Using Self-Configuring Novelty Search Genetic Algorithm," the 11th International Conference on Informatics in Control, Automation and Robotics, ICINCO 2014. Vienna, Austria, 2014, pp. 832-839.

[9] F. Eyben, M. Wullmer and B. Schuller, "OpenSMILE - the Munich versatile and fast open-source audio feature extractor," In Proceedings ACM Multimedia (MM), ACM, Florence, Italy, ISBN 978-1-60558-933-6, pp. 1459-1462, 25.-29.10.2010.

[10] E. Sariyanidi, H. Gunes, M. Gokmen and A. Cavallaro, "Local zernike moment representation for facial affect recognition," BMVC'13.

[11] M. Rashid, S. A. R. Abu-Bakar and M. Mokji, "Human emotion recognition from videos using spatio-temporal and audio features," Vis Comput (2013),29: 1269-1275.

[12] S. E. Kahou, C. Pal, X. Bouthillier, P. Froumenty, C. Gulcehre, R. Memisevic, P. Vincent, A. Courville and Y. Bengio, "Combining modality specific deep neural networks for emotion recognition in video," ICMI'13, December 9-13, 2013, Sydney, Australia.

[13] A. Cruz, B. Bhanu and N. Thakoor, "Facial emotion recognition in continuous video," In Proceedings of the 21st International Conference on Pattern Recognition (ICPR 2012), November 11-15, 2012, Tsukuba, Japan.

[14] M. Soleymani, M. Pantic and T. Pun, "Multimodal emotion recognition in response to videos," IEEE Transactions on affective computing, vol. 3, no. 2, April-June, 2012.

[15] C. Busso, Z. Deng, S. Yildirim, M. Bulut, C. M. Lee, A. Kazemzadeh, S. Lee, U. Neumann and S. Narayanan, "Analysis of emotion recognition using facial expressions," Speech and Multimodal Information, University of Southern California, Los Angeles.

[16] T. M. Phuong, Z. Lin and R. B. Altman, "Choosing SNPs using feature selection," Proceedings / IEEE Computational Systems Bioinformatics Conference, CSB. IEEE Computational Systems Bioinformatics Conference, pp. 301-309, 2005.

[17] B. Duval, J.-K. Hao and J. C. Hernandez Hernandez, "A memetic algorithm for gene selection and molecular classification of an cancer," In Proceedings of the 11th Annual conference on Genetic and evolutionary computation, GECCO '09, pp. 201-208, New York, NY, USA, 2009.

[18] T. Ojala, M. Pietikäinen and D. Harwood, "A comparative study of texture measures with classification based on feature distributions," Pattern Recognition 19(3): 51-59.

[19] G. Zhao, and M. Pietikäinen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," IEEE Trans. Pattern Analysis and Machine Intelligence 29(6): 915-928.