Design Efficient Technologies for Context Image Analysis in Dialog HCI Using Self-Configuring Novelty Search Genetic Algorithm

Evgeny Sopov and Ilia Ivanov
Institute of Computer Science and Telecommunications, Siberian State Aerospace University, Krasnoyarsk, Russia

Keywords: Image Analysis, Classification, Multiobjective Genetic Algorithm, Novelty Search.

Abstract: The efficiency of HCI systems can be sufficiently improved by the analysis of additional contextual information about the human user and interaction process. The processing of visual context provides HCI with such information as user identification, age, gender, emotion recognition and others. In this work, an approach to adaptive model building for image classification is presented. The novelty search based upon the multi-objective genetic algorithm is used to stochastically design a variety of independent technologies, which provide different image analysis strategies. Finally, the ensemble based decision is built adaptively for the given image analysis problem.

1 INTRODUCTION

Modern information and communication technologies are highly interactive. There is a problem of designing effective HMIs (Human machine interface, HMI), which include a contextual analysis of the user's activities, dialog speech interfaces, etc. Today, virtually all modern communication and computer technologies have the capability of capturing the user's image (using the camera on computers and smartphones, ATMs, vending machines services, vehicles and many others). The user image analysis can provide HMI with additional information to improve the efficiency of human-computer interaction. In particular, image analysis is able to obtain such information as user anthropological data (gender, age, ethnicity, etc.). This work is focused on the particular analysis problem of estimating a person's age based on a photo.

In general, existing image analysis technologies are realized using two image-processing stages: the pre-processing stage to highlight different image patterns and the recognition stage to solve the given analysis problem. Both stages become a great challenge, as the user image in real applications is not usually well suited to automated computer processing. Moreover, there is no a priori effective technique for highlighting image patterns. Finally, the practical usage of HMI leads to various image analysis problems, which can dynamically change over time. The design of intelligent adaptive methods for generating and tuning effective image analysis technologies for each particular application seems to be a good idea.

The genetic algorithms (GA) are a well-known and highly effective approach to solving many search and adaptation problems (Goldberg, 1989), and are also widely used to configure intelligent information technologies (Shabalov *et al.*, 2011). In this work multi-objective GA is used to design simultaneously pre-processing and recognition technologies. The self-configuring version of GA is used to eliminate the need for choosing the correct GA structure and tuning of parameters.

In the field of statistics and machine learning, ensemble methods are used to improve decision making. The collective solution of multiple learning algorithms provides better predictive performance than could be obtained from any of the constituent learning algorithms (Opitz and Maclin, 1999). It is clear that similar algorithms will lead to identical failures, so we need to support a diversity of technologies of an ensemble. In this work the novelty search approach is used to generate various image analysis technologies, which represent various analysis strategies. The advantage of the novelty search is that the obtained solutions may be new, previously unknown and even non-obvious (Lehman and Stanley, 2008).

Sopov E. and Ivanov I..
Design Efficient Technologies for Context Image Analysis in Dialog HCI Using Self-Configuring Novelty Search Genetic Algorithm.
DOI: 10.5220/0005147108320839
In *Proceedings of the 11th International Conference on Informatics in Control, Automation and Robotics* (ASAAHMI-2014), pages 832-839
ISBN: 978-989-758-040-6

2 RELATED WORK

The idea to obtain additional information from a user image to adapt the dialog interaction in HMI is not new, but has still not been sufficiently investigated. The main reason is the problem of discovering a generalization for the methods applied to recognize human features. In particular, there are the following problems in the field of estimating a person's age (Geng *et al.*, 2007; Sithu *et al.*, 2013):

- Different people grow old in different ways, thus there is no universal methodology of age recognition;
- Image pre-processing, feature extraction and principal component analysis involve operations like human features detection (eyes, nose, mouth, eyebrows, etc.), which are complex problems themselves;
- There is no universal technique for highlighting image patterns for age recognition. There exist many local approaches based on using different filters for image processing;
- The image of a person is obtained each time under non-standard, non-ideal conditions (lighting, shooting angle, excess objects and artefacts and so on).

The way to overcome these problems is to design technology that efficiently combines the feature extraction from the image and the recognition system.

The list of image pre-processing filters is finite and known. The problem is to select the proper ones and to construct an efficient combination of filters.

The given extracted features are used in the recognition system. There are many approaches to complete learning and classification: statistical and mathematical methods, machine learning, heuristics and others. Many researchers in the field of image analysis consider the Convolutional Neural Networks (CNNs) as the one of the most powerful approaches to image recognition problem solving (Bishop, 1995; Simard *et al.*, 2003).The CNN is based on the idea of Deep Learning that uses many levels of information representation for complex relationships in data modelling. This leads to the following advantages:

- Local perception. Only a part of the given image is processed by each neuron.
- Weights selection. A small number of weights is used to model a large number of links in data.
- Generalization. CNN ignores the image noise and finds image invariants.

As is known, the structure and parameter adjustment of a neural network is a complex problem, which, nevertheless, can be solved using an adaptive stochastic procedure like a genetic algorithm. This work is focused on a special type of search algorithm called a novelty search algorithm.

The novelty search approach instead of searching for a final objective, rewards instances with functionality significantly different from what has been discovered before. Thus, instead of a traditional objective function, evolution employs a novelty metric. The novelty metric measures how different an individual is from other individuals, so creating a constant pressure to do something new. The novelty search was successfully applied to many search problems in various domains (Lehman and Stanley, 2011).

Applying the idea of a novelty search to the problem of CNN design, it is possible to obtain many different recognition technologies, even those that are not a priori obvious for an expert in a field of CNN.

3 INTELLEGENT TECHNOLOGY OF IMAGE ANALYSIS

The process of image analysis for each person starts with some standard pre-processing steps like face detection, canvas rotation/orientation, image levels correction, transformation to grayscale (if necessary), etc. We will assume that these operations are well studied, universal and not in the scope of this work.

The image after the pre-processing steps are applied is represented in some resolution M by M pixels (also can be viewed as a square matrix of size M). It is obvious that a higher resolution leads to possibly better feature extraction, but a more complex recognition process. A set of initial images forms a training set for CNN. Unfortunately the initial image contains no human features in an explicit form, thus we need to apply some operations to extract features to be appended to the training set.

Usually special image filters are used to highlight different image patterns. Most of them are based on different methods to find the edges of objects using contrast, light gradient, blurring, etc. After a filter is applied, an additional processed image of M by M resolution is obtained, which is appended to the training set.

As we use the GA to design the image analysis technology, we will discuss the GA solution representation. The binary GA is used. The chromosome of an individual consists of two parts: filter combination and CNN structure representations.

It is assumed that the set of filters consists of K items. We use the first K position of chromosome to encode filter combination. The i^{th} position is equal to one if the i^{th} filter is applied, and zero otherwise.

The number of maps and the size of kernels define the CNN structure. Usually two types of CNN are used: 3 layer (the sequence of Convolutional (C-layer), Sub-sampling (S-layer) and Fully-connected (F-layer)) and 5 layer (C-, S-, C-, S- and F-layer). As the full-connected layer is defined by the preceding convolution and sub-sampling, it is necessary to define only the parameters of the convolutional and the sub-sampling layer. Thus, the second part of the chromosome should encode the following set of integer parameters: (1) in a case of 3 layer CNN, (2) in a case of 5 layer CNN.

$$\langle CL, c_1, c_2, \ldots c_{CL}, SL, s_1, s_2, \ldots s_{SL}\rangle \tag{1}$$

$$\left\langle CL_1, c_1^1, \ldots c_{CL_1}^1, SL_1, s_1^1, \ldots s_{SL_1}^1, \; CL_2, c_1^2, \ldots c_{CL_2}^2, SL_2, s_1^2, \ldots s_{SL_2}^2 \right\rangle \tag{2}$$

where CL is the number of maps of C-layer, SL is the number of maps of S-layer, c_i and s_i is the size of kernel of i^{th} map.

The quality measure for GA individuals (or fitness function) is directly based on a standard metric of accuracy (the percentage of correct predicted samples). In real-world dialog HMI problems, it is not necessary to estimate the exact age, but to classify some age interval. The exact age estimation is a great challenge even for a human and we usually divide people into groups like child, teenager, adult under 30 y.o., 30-45 y.o., etc. Thus, the used age intervals reduce the number of classes in the recognition problem, give more reliable results of prediction and make the decision-making process closer to the human one.

The fitness evaluation implies two steps. First, the binary solution is decoded to obtain the structures of combinations of image filters and CNN, i.e. an image analysis technology is designed. Second, the learning process is performed with the training set and the accuracy of the age estimation is evaluated.

The weight adjustment algorithm for CNN is not a focus of this work; any efficient technique can be applied (e.g. genetic algorithms, evolutionary strategies, particle swarm optimization, gravitational search algorithm, etc.).

4 NOVELTY SEARCH GENETIC ALGORITHM

The proposed approach suggests designing a variety of image analysis technologies with different properties to be applied in an ensemble. As previously mentioned, one of the ways to improve a collective decision in an ensemble is to include technologies as diverse as possible.

The GA search based upon only the recognition accuracy criteria may lead to very similar or even the same optimal solutions. Thus to support the solution diversity additional search criteria should be used. In this work, we apply the idea of the novelty search.

To implement the search for a novelty, the search algorithm should focus on the functionality diversity of solutions (in terms of the given problem) instead of the objectives as in the standard approach.

The functionality of the image analysis technologies designed in this work is based on two things:

- The combinations of image filters applied produce different feature sets extracted from the original image of the person. As a result, the same CNN is trained with different training data.
- Different structures of CNN implement dissimilar convolutions and sub-samplings, so CNNs process the same image in many ways using various links in data and unique generalization.

Using the solution representation proposed in section 3, we can control the functionality diversity directly by preserving population diversity. Thus, we can formalize the following additional search criteria for GA (3):

$$\rho(X) = \frac{1}{T}\sum_{j=1}^{T} dist_H(X, X_j) \to max \tag{3}$$

where X_j is the j^{th}-nearest neighbor of X with respect to the distance metric $dist_H$, $dist_H$ is the Hamming distance for binary representation.

The number of T nearest neighbour is chosen experimentally and its value is much less than the binary solution dimension (chromosome length) because of the neighbourhood structure in the binary space.

The novelty criteria $\rho(X)$ can be viewed as the sparseness at point X. It is clear that if the average distance to the nearest neighbours of the given point is large then point X is in a sparse area and it is a novel solution. Otherwise, if the average distance is small, the point is in a dense region.

There exist some techniques for preserving population diversity (e.g. crowding, sharing, niching, etc.). They are widely used for finding multiply optima (Mahfoud, 1995), but seem to be unsuitable for a novelty search as their fitness function is based on objectives.

We state the problem of the design of image analysis technologies as an optimization problem with two criteria: recognition accuracy and novelty search. Thus, we have to use a multi-objective technique instead of the standard GA. In this work, we use an efficient approach called the self-configuring coevolutionary multi-objective genetic algorithm (SelfCOMOGA). The SelfCOMOGA was introduced in (Ivanov and Sopov, 2013a) and its effectiveness was investigated in (Ivanov and Sopov, 2013b).

One of the ways to develop self-configuring search algorithms is the co-evolutionary approach (Ficici, 2004).

Most of the works related to the co-evolutionary multi-objective GA use the cooperative co-evolution scheme, which implies the problem decomposition of variables, objectives, population functionality, etc. The fitness evaluation is based on the fitness values of many individuals. Such a scheme shows good performance, but requires the fine-tuning of algorithm parameters, decomposition and cooperation.

In the case of competitive co-evolution, algorithms (populations) evolve independently over some time. Then their performance is estimated. Computational resources are redistributed to the most effective algorithms. In addition, random migrations of the best solutions are presented. The competitive co-evolution scheme eliminates the necessity to define an appropriate algorithm for the problem as the choice of the best algorithm is performed automatically during the run (Goh, 2007).

It is clear that applying the co-evolution scheme provides the self-configuration for the given multi-objective problem. Using different multi-objective search approaches with fundamentally different properties, the co-evolution algorithm is able to improve the use of their individual advantages and minimize the effects of the disadvantages.

The general SelfCOMOGA scheme is as follows:

Step 1. To define a set of multi-objective algorithms included in the co-evolution.

Step 2. To perform algorithms run over some time (called the adaptation period).

Step 3. To estimate the performance for each algorithm over the adaptation period.

Step 4. To redistribute the computational resources (population) and perform a new adaptation period (go to the step 2).

We will discuss the SelfCOMOGA steps in detail.

The first step defines the search strategies in the form of the algorithms included. There are at least three ways to form the algorithm set:

- A pre-defined set of algorithms with special performance features, which can be applied to multi-objective problem solving:
- All appropriate multi-objective algorithms;
- A random selection of algorithms from the set of all possible algorithms combinations.

The first way requires the involvement of a priori knowledge of the problem and appropriate algorithms. So it is not applicable to complex real-world problems and contradicts the idea of the self-configuring GA. The second approach is very expensive because of the number of all algorithm combinations (tens and hundreds). As shown in (Ivanov and Sopov, 2013b), the third strategy provides acceptable efficiency on average, and it is not necessary to involve additional information about the algorithms and their properties. This option provides the best concept of the self-configuring GA.

The adaptation period is a parameter of the co-evolution algorithm. Numerical experiments show that the parameter value is individual for each problem and depends on the performance criteria of the algorithm. Moreover, the value depends on the limitation of the computational resource (total number of fitness evaluations). It should be noted that the algorithm is not sensitive to the parameter on average with sufficient computational resources.

The key point of any co-evolutionary scheme is the performance evaluation of the individual algorithm. Since the performance is estimated using the Pareto concept, a direct comparison of algorithms is not possible, so well-known approaches cannot be used. In various studies the following criteria are proposed: the distance (closeness) of the obtained solutions set to the true Pareto set and the uniformity of the set of obtained solutions. It is obvious that in real-world problems the first criterion is not applicable because the true Pareto set is a priori unknown. In this work we use the following criteria combined into two groups.

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. The pool of all algorithm solutions is created and non-dominated sorting is performed. Finally, the number of non-dominated solutions corresponding to each algorithm is counted.
- Criterion 2 is the uniformity (dispersion) of non-dominated solutions. The average variance of distances between individuals is computed using the coordinates (for the Pareto set) or criteria (for Pareto front).

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. The improvement is completed if the current solutions dominate the previous ones, even its number has decreased.

The last stage is the redistribution of resources. New population sizes are defined for each algorithm. All algorithms give to the "winner" algorithm a certain percentage of their population size. Each algorithm has a minimum guaranteed resource that is not distributed.

In many co-evolutionary schemes each new adaptation period begins with the same starting points ("the best displaces the worst"). This approach is not suitable for multi-objective problems due to the requirement of the diversity of Pareto solutions. In this work, we apply the probability sampling scheme "substitution by rank selection". Solutions with a high rank after non-dominated sorting have a greater probability value to be transferred to the best algorithms. With decreasing rank, the probability of selection decreases linearly, but the chance to be selected remains for all solutions.

The co-evolutionary algorithm stop criteria are similar to the standard GA: fitness evaluation restriction, a number of generations with no improvement (stagnation), etc.

5 EXPERIMENTAL RESULTS

The proposed approach was investigated with a widely used dataset of human facial images labelled with their true age. The dataset is collected by researchers of the University of Michigan and provided by the University of Texas in Dallas (the full data set can be found in the Internet in http://agingmind.utdallas.edu/facedb). The goal is to design an image analysis technology for automatic age recognition. Either this problem may be viewed as a classification problem (with the finite number of age groups) or as a regression problem (the age of a person is a real number).

The data contains more than 1000 different facial images of humans of different genders and ethnic groups, and the people also have different emotions.

The original resolution was 640 by 180 pixels. A face detector is applied to every photo to select the informative region of the image. Colour images are converted to grayscale. After the feature extraction filters are applied, the resolutions of all images are reduced to 100 by 100 pixels for the CNN training.

We use the following encoding of image analysis technologies designed using the Self-configuring Novelty Search Genetic Algorithm (4):

$$X = (x_1, \dots, x_l, x_{l+1}, \dots, x_{l+m}) \quad (4)$$

where X is a GA binary solution, $x_i \in \{0,1\}$. l is the number of bits encoding the filters applied, m is the number of bits encoding the CNN structure. The total length of the chromosome is $n=l+m$.

We start with the list of filters which are used to highlight image patterns. The list, based on studies in (Zhou and Chellappa, 2004), contains the following: Logarithmic, Canny, Zerocross, Prewitt, Sobel, Roberts and LBP.

Thus the value of l is equal to 7 (total number of all possible filters combinations is $2^7 = 128$). The i^{th} position is equal to one, if the i^{th} filter is applied, and zero otherwise. The solution with all l positions equal to zero corresponds to the case that no filters are applied; the training set contains the initial image only.

The example of an arbitrary initial image taken from the data set is presented in Figure 1. The examples of the same image after face detection and filter are applied are in Figure 2 and Figure 3.

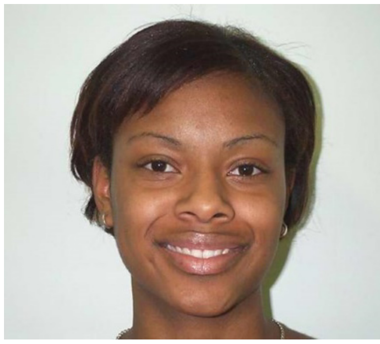

Figure 1: The example of an initial image.

Figure 2: The example of canny filter applied.

Figure 3: The example of zerocross filter applied.

A new image, obtained after a filter is applied forms new data in the training set. There the smallest size of the training set is in the case of no filter being used, the largest one is in the case of all filters being applied.

The next step is that we encode the structure of the CNN that are trained with the given extracted features. We use the CNN with 5 layer of the following order: C-, S-, C-, S- and F-layer.

Starting with the *l+1* position of the solution chromosome, we use the following encoding:

- 5 bits to encode the number of maps of the first convolutional layer. The max number is $2^5 = 32$.
- 6 bits 32 times (summary $6 \cdot 32 = 192$ bits) to encode the kernel size for each map. If the number of maps after decoding is less than 32, the corresponding parts of size encoding are not used in the CNN building.
- 5 bits 32 times (160 bits) to encode the subsampling kernels for each convolutional layer map.
- 4 bits to encode the number of maps of the second convolutional layer (maximum 16 maps for each map after subsampling).
- 2 bits 16 times (32 bits) – convolutional kernels sizes.
- 2 bits 16 times (32 bits) – subsampling kernels sizes.

As result, the total length of the CNN encoding part is 425 bits. The total chromosome length is 432 bits.

The initial parameters for the SelfCOMOGA are:

- The list of multi-objective search algorithms contains Vector Evaluated Genetic Algorithm (VEGA), Fonseca and Fleming's Multi-objective Genetic Algorithm (FFGA), Niched Pareto Genetic Algorithm (NPGA), Non-dominated Sorting Genetic Algorithm (NSGAII) and Strength Pareto Evolutionary Algorithm (SPEA) (Tan et al., 2001; Zhoua et al., 2011). All algorithms use the self-configured parameters as performed in (Semenkin and Semenkina, 2012).
- Total population size is 400 individuals.
- The initialization is random.
- The crossover operators: one-point, two-point and uniform.
- The probability of mutation operator: low, mean and high.
- The maximum size of the Pareto set (for SPEA): 50 individuals.
- The niching parameter (niche radius) (for FFGA and NPGA): 1, 3 and 5.
- The size of a comparison set (for NPGA): 3, 5 and 7.

The different image analysis technologies discovered by GA are applied in the form of ensemble with averaging of individual predictions.

To estimate the designed CNN performance we have divided the given age range into the following 7 intervals:

- 18-23 years old,
- 23-27 years old,
- 25-36 years old,
- 32-49 years old,
- 46-60 years old,
- 58-70 years old,
- >70 years old.

The training set is randomly divided into two samples (80% and 20%) to perform the training and test. The error is a standard accuracy.

The result of the SelfCOMOGA run is a set of designed technologies, which are effective with respect to two criteria: accuracy and novelty (diversity). The result Pareto front set is shown in Figure 4.

As we can see, the SelfCOMOGA provides a rather uniform distribution over the Pareto set, so we obtain a good variance of image analysis technologies. The Pareto set contains the age estimation technologies that both have good accuracy, but their structures are similar and there is poor accuracy with a unique structure. As we previously assumed, the novel solutions provide a completely different analysis strategy and may improve total efficiency when applied in an ensemble with more accurate solutions.

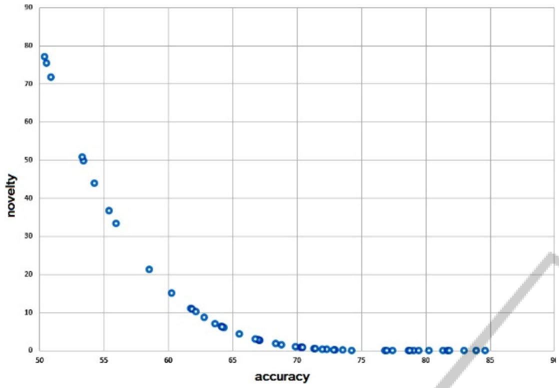

Figure 4: Pareto front set distribution.

Also we have compared the proposed approach with the following algorithms:

- 7 algorithms are the combination of a filter and manually (experimentally) tuned CNN.
- The averaging ensemble mentioned above 7 algorithms.
- 7 algorithms + ensemble, where the CNN is substituted with Support Vector Machines (SVM). SVMs are widely and successfully used for many complex classification and recognition problems in machine learning (Byun and Lee, 2002).

The results of numerical experiments are shown in Table 1. The algorithms are grouped: self-configuring, CNN-based and SVM-based.

Table 1: Age estimation accuracy.

Algorithm	Training sample	Test sample
Self-configuring technology design	85.37	81.03
Logarithmic+CNN	72.45	68.26
Canny+CNN	70.32	66.88
Zerocross+CNN	72.11	65.63
Prewitt+CNN	60.87	56.17
Sobel+CNN	58.90	54.46
Roberts+CNN	58.42	55.04
LBP+CNN	59.93	58.13
7 CNN ensemble	78.87	71.23
Logarithmic+SVM	64.15	63.74
Canny+ SVM	71.97	64.13
Zerocross+ SVM	70.00	66.21
Prewitt+ SVM	56.37	54.02
Sobel+ SVM	56.57	52.11
Roberts+ SVM	52.39	51.34
LBP+ SVM	53.54	50.87
7 SVM ensemble	70.55	68. 43

As we see, the ensemble approach for both CNN and SVM shows slightly better performance than the best algorithm in a group (CNN or SVM). The self-configuring approach shows significantly better results. We have applied the ANOVA (Wilcoxon-Mann-Whitney) tests to confirm the statistical significance of the difference in results.

6 CONCLUSIONS

The self-configuring novelty search genetic algorithm for the design of image analysis technologies is proposed in this work. We have demonstrated the efficiency of the approach with respect to the particular problem of age recognition. The approach can be implemented for any other image analysis problem raised in the field of dialog HCI as we do not apply specific information about the problem.

The computational costs of the approach are higher than of the standard algorithms, but the results are performed automatically in an adaptive way for a certain problem.

So we can recommend the proposed approach for the design of image analysis when the problem is poorly-studied, no a priori information on the problem has been presented or there are no appropriate experts in the field of computer image analysis.

In further investigations, we intend to examine the approach with other particular problems and to integrate the approach with dialog HCI systems to provide the design of technologies for certain HCI requirements.

ACKNOWLEDGEMENTS

Research is performed with the financial support of the Ministry of Education and Science of the Russian Federation within the federal R&D programme (project RFMEFI57414X0037).

The authors express their gratitude to Mr. Ashley Whitfield for his efforts to improve the text of this article.

REFERENCES

Bishop, C.M. (1995). *Neural Networks for Pattern Recognition*. Oxford University Press.

Byun, H., Lee, S.W. (2002). Applications of Support Vector Machines for Pattern Recognition: A Survey. In: *Proceedings of the First International Workshop on Pattern Recognition with Support Vector Machines.* pp. 213-236.

Ficici, S.G. (2004). *Solution Concepts in Coevolutionary Algorithms.* A Doctor of Philosophy Dissertation / S.G. Ficici. - Brandeis University.

Geng, X., Zhou, Z.-H., Smith-Miles, K. (2007). Automatic age estimation based on facial aging patterns. In: *IEEE Trans. On PAMI* 29(12), pp. 2234-2240.

Goh, C.K. (2007). A competitive-cooperation coevolutionary paradigm for multi-objective optimization. In: *22nd IEEE International Symposium on Intelligent Control, Singapore*. pp. 255-260.

Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley Publishing Co., Inc., Reading, Mass.

Ivanov, I., Sopov, E. (2013a). Self-configured genetic algorithm for multi-objective decision making support. *Vestnik SibSAU (scientific journal of Siberian State Aerospace University)* 1(47). pp. 30-35. - In Russian, abstract in English.

Ivanov, I., Sopov, E. (2013b). Investigation of the self-configured coevolutionary algorithm for complex multi-objective optimization problem solving. *Control systems and information technologies,* 1.1 (51). pp. 141-145. - In Russian, abstract in English.

Lehman, J., Stanley, K.O. (2008). Improving Evolvability through Novelty Search and Self-Adaptation. In: *Proceedings of the 2011 IEEE Congress on Evolutionary Computation (CEC 2011).* IEEE, pp. 2693-2700.

Lehman, J., Stanley, K.O. (2011). Abandoning objectives: Evolution through the search for novelty alone. *Evol. Comp*. 19(2), pp. 189-223.

Mahfoud, S.W. (1995). A Comparison of Parallel and Sequential Niching Methods. In: *Proceedings of the Sixth International Conference on Genetic Algorithms.* pp. 136-143.

Opitz, D., Maclin, R. (1999). Popular ensemble methods: An empirical study, *Journal of Artificial Intelligence Research* 11, pp. 169–198.

Semenkin, E.S., Semenkina, M.E. (2012). Self-configuring Genetic Algorithm with Modified Uniform Crossover Operator. Advances in Swarm Intelligence. *Lecture Notes in Computer Science 7331. – Springer-Verlag, Berlin Heidelberg*. pp. 414-421.

Shabalov, A., Semenkin, E., Galushin, P. (2011). Automatized Design Application of Intelligent Information Technologies for Data Mining Problems. In: *Joint IEEE Conference "The 7th International Conference on Natural Computation & The 8th International Conference on Fuzzy Systems and Knowledge Discovery", FSKD'2011,* pp. 2596-2599.

Simard, P.Y., Steinkraus, D., Platt, J.C. (2003). Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis. In: *International Conference on Document Analysis and Recognition (ICDAR), IEEE Computer Society*.

Sithu, U., Shyama, D., Imthiyas, M.P. (2013). Human age prediction and classification using facial image. *International Journal on Computer Science and Engineering*, 5(05).

Tan, K.C., Lee, T.H., Khor, E.F. (2001). Evolutionary Algorithms for Multi-Objective Optimization: Performance Assessments and Comparisons. In: *Proceedings of the 2001 IEEE Congress on Evolutionary Computation Seoul, Korea*. pp. 979-986.

Zhou, S.K., Chellappa, R. (2004). Illuminating Light Field: Image-Based Face Recognition Across Illuminations and Poses. In: *Proc. Sixth IEEE Int'l Conf. Automatic Face and Gesture Recognition*, pp. 229-234.

Zhoua, A., Qub B.Y., Li, H., et al. (2011). Multiobjective evolutionary algorithms: A survey of the state of the art. Swarm and Evolutionary Computation, 1 (2011). pp. 32-49.