Genetic Algorithms in the Domain of Personalized Nutrition

Petri Heinonen[1*], Esko K. Juuso[2]

[1]Nutri-Flow Oy, Villiperäntie 5, FI-90410 Oulu, Finland,; *petri.heinonen@nutri-flow.fi

[2]Control Engineering, Environmental and Chemical Engineering,
Faculty of Technology, P.O. BOX 4300, FI-90014 University of Oulu, Finland

SNE 32(1), 2022, 47-54, DOI: 10.11128/sne.32.tn.10597
Received: 2020-11-10 (Selected EUROSIM 2019 Postconference Publication); Revised: 2022-02-15; Accepted: 2022-02-25
SNE - Simulation Notes Europe, ARGESIM Publisher Vienna, ISSN Print 2305-9974, Online 2306-0271, www.sne-journal.org

Abstract. Lifestyle related public health problems are common around the world. Personal nutrient guidance is a tool for promoting healthier lifestyles. Most of the applications available on the market are based on energy only, and a reliable individual assessment and guidance is given by licensed nutritionists. Nutri-Flow has a novel approach into personalized nutrition guidance with Fuzzy Expert System (FES) enhanced with Genetic Algorithms (GA) optimization. While FES assesses the foods and beverages added into a search space, GA is used to find the level of intake for them. The optimization problem is to minimize the distance to ideal nutrient intake levels, and to keep the level of change in a feasible level and take into account other nutrition variables. In this study, the suitability of GA was assessed. Also, the performance the GA was evaluated and evolved. The objective function is presented, and the overall results were evaluated numerically if the system was feasible in the domain of nutrition. The nutritional aspect is not in the scope of this study.

Introduction

According to FinHealth 2017 study [1], obesity is still one of the key public health problems among all age groups in Finland. It is also stated in the study that healthy diet is one solution to prevent many key public health problems. It was found that there is a need to promote healthy lifestyle.

There are available national nutrition and food recommendations. Food pyramids and plate models are given as a general guidance towards a balanced diet. General guidelines and recommended intake levels for each nutrient are introduced nationally, too [2]. In this study, all nutritional values are achieved from Finnish nutrition and food recommendations, which are based on Nordic recommendations [2]. To find the nutrient composition in a diet, there is a need also for a food composition database. National food composition databases are used for this purpose, such as Fineli [3] in Finland, Livsmedeldatabasen [4] in Sweden, USDA [5] in the USA, etc. These databases are collections of foods, beverages, and recipes with their averaged nutrient composition.

A lot of data is available for healthy eating habits, therefore, experts are needed to put all together for assessment and guidance for individuals. In the Internet, there are available thousands services and applications to monitor and balancing the diet. Those are too simplified approaches, since they take into account only the energy in a very complex problem domain. If only the energy is taken into account, the rest of the variables, protective nutrient levels, will deteriorate and might lead even to malnutrition. Typical diets are shown in Figure 1:

- **Diet 1**: The sufficient intake of protective nutrients is accompanied by an excess intake of energy;
- **Diet 2**: The insufficient intake of protective nutrients is combined with the excess intake of energy;
- **Diet 3**: Insufficient intake of protective nutrients with sufficient intake of energy;
- **Diet 4**: Nutrient dense diet with ideal proportion of intake of protective nutrients and energy. [6]

The eating habits have widely moved to excess energy (Diet 2) and the energy-only approach has neglected the importance of the protective nutrients (Diet 3). Diet 4 is needed to balance these problems.

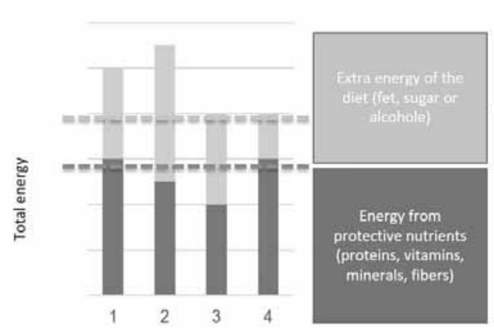

Figure 1: Differences between typical diets [6].

For individual nutritional guidance, there are four main steps to take into account when developing an algorithm for the problem domain. First, there should be a food record or a meal diary which is used to assess nutrient intake levels. This step includes imprecision and uncertainty due to human error while filling in a record with kitchen units. Second step is to assess a personal nutrition recommendation based on the current national recommendation. Third step is to assess the needed level of change to balance the diet. Fourth step is to generate the guidance as foods and beverages with their portion sizes.

In the previous research [7, 8], a Fuzzy Expert System was developed to handle the imprecision and uncertainty present in the system. It was also discussed in the previous publications that an optimization algorithm is needed to assess the portion sizes. GA was selected initially for testing. GA was selected due to the complexity of the problem. GA is widely used with complex real-world problems [9, 10, 11].

This research combines computational intelligence (Section 1) for personalized nutrition guidance with focus on genetic algorithms (Section 2). The operation is demonstrated in a test case (Section 3) and the results are presented in Section 4. Conclusions and future research are discussed in Section 5.

1 Background

1.1 Data Acquisition

Nutrition experts have generated test cases with typical personal data and a meal diary example for the study. In this study, a test case called Sum Mikko, is used. The test case is 41 years old, thus national recommendations for males of ages between 31 and 60 are used to calculate the personal recommendation. The personal data are acquired from Nutri-Flow directly. The meal record of one day represents one week average including seven meals with 16 different foodstuffs. Nutrient intake levels are calculated based on the food record and Fineli Food Composition Database [3]. No personal data from Nutri-Flow database is used in this study.

1.2 Fuzzy Expert System

Fuzzy Expert System (FES) has the key role to handle the imprecision and assesses the personalized dietary guidance. FES is adapted to personal recommendation, by applying the personal nutrition recommendation values. Nutrient intake levels are converted into the fuzzy domain with membership functions. Three linguistic fuzzy variables are used to define the intake levels: too little, ideal, too much. Personal nutrient recommendation values are used to tune the membership functions. Most of the 30 nutrients to take into account have three recommendation values: Lower Intake Level (LI), Recommended In-take Level (RI), and Upper Intake Level (UI). The fuzzy variables are defined with values A,B and C (Figure 2) are tuned with LI, RI, and UI, respectively.

The linguistic variables are applied to form the knowledge base for the fuzzy inference machine. Knowledge of nutrition experts was acquired and coded into the knowledge base as rules, for example

IF vitamin C is too little AND fiber is too little
THEN fruits and berries group IS add

The rules are mapping the guidance directly into foodstuff level from the nutrient level. The inference is Mamdani type, therefore, the output is also fuzzy with similar sets of linguistic fuzzy variables. The used variables are “reduce”, “no action” and “add”. The desired output for each food is a crisp number representing the direction and importance of recommended change.

1.3 Optimization problem

The output of FES is a list of foods and beverages with the defuzzified value expressing the direction and importance of change. The objective is to find the best way to balance a diet. To achieve this objective, portion sizes of the foods and beverages for the guidance should be assessed. As discussed in [7], the problem domain is complex. There are 30 nutrient variables and energy to be taken in account. Also, foodstuffs and recipes present a complex composition of sources of each nutrient.

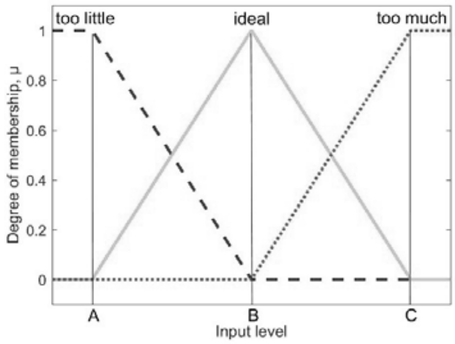

Figure 2: Guidance mapping represents recommendations as fuzzy numbers.

A diet can be altered in various ways to reach similar composition in the nutrient domain. Therefore, there can be solutions which have the same fitness value or the fitness values are very close to each other. The system must be feasible in the field of the nutrition domain, which brings more constrains to be taken into account. E.g. the difference between recommended diet and initial diet cannot be too drastic. The diet should be altered with smaller steps if it was originally far away from the recommendations. Also, other nutrition related variables should be taken into account. The main objective of the optimization algorithm is to minimize distance from ideal nutritional composition, keep the recommended step in desired magnitude, and take into account the other tuning variables.

As discussed in [7], traditional optimization methods are not suitable due to the nature of the problem. The initial approach was to study if GA is a suitable solution.

2 Genetic Algorithms

Genetic Algorithms belong to the group of Evolutionary Algorithms. The roots and ideology of GA are based on Darwin's theory of natural selection. The theory was introduced in 1975, and the terminology was closely adapted from natural genetics. [12]. A single solution is called a chromosome, which has locus bind variables, genes. When the theory was published, a binary approach for coding the results were applied. It was discussed that binary coding is not always the best way with real world problems where the high precision makes the chromosomes long and the algorithm gets inefficient [13].

A population is a set of chromosomes, which is set in a competitive environment to find the global optimum. Each chromosome is evaluated through an objective function. The calculated fitness value is used to rank the solutions. The population is evolved by applying GA operators in every iteration round. The optimization starts with generating the population, where the selection and crossover operators are applied to combine genetic material for the new population. With a mutation operator, random variations are introduced into the population. This helps to prevent stopping on local minimas. Crossover and mutation operations might lead to the loss of the best solutions, the elitism operator can be used to copy the best solutions into the new population directly [12, 13].

There has been a lot of development with the GA since the theory was published. In the literature, there is a large number of problem specific approaches for the operators [14, 15]. The variables for the operators have been traditionally found by trial and error. There are also studies which provide tools for finding the best configuration [15].

2.1 Configuration

The configuration of GA has a strong effect on convergence and on finding a global optimum. Convergence time and feasible solutions are key factors when applying GA in an online service.

Population Population is a set of possible solutions. The size of the population is an important variable. Too small a population might lead to insufficient divergence between the solutions and the optimum is not reached. Too large a population might lead to slow convergence, since the evolving needs more iterations and the objective function must evaluate more solutions. It is discussed in [16] that, the longer the chromosome is, the larger the number of individuals in population is needed.

The initial population is usually generated randomly within the given constraints. There are also statistical methods available. If the population is generated randomly, it is recommended to run the GA with different initial populations, since the initial guess might not always lead to the global optimum [17].

The size of the population can be static or vary between the iterations [18].

Coding The results can be coded into chromosomes in binary or real-valued domain. The size of search space and accuracy level of the results should be used when selecting the coding. Real-value coding is used widely in real world optimization when the size of search space is big and higher accuracy is required on the results. With real-value coding there is no need for result mapping which reduces need for computational resources. However, it has been discussed that real-value coding has problems to yield good results always.

Crossover Crossover is the core operator to evolve population towards better solutions. The good genetic material is distributed between generations. The good genetic material from the population is found with a selection method. The selection is done usually by the roulette wheel method or by the tournament selection method.

The tournament selection method selects randomly chromosomes from the population and the fitness values are compared. The better chromosome is selected into a mating pool for the crossover method. The tournament selection method has a configuration parameter k, which defines the number of selected chromosomes from the population. Typical value for k is two.

The crossover operator is applied to mating population to form offspring. Crossover probability, P_c, is a design parameter for the crossover operator. The parameter is used to determine if current mating population chromosomes are combined with the crossover operator or directly to the offspring population.

The selection of the crossover operator depends on the coding and the problem. For real-value coded chromosomes, non-uniform and uniform crossover operators are applied. The operation of non-uniform crossover operators depends on the age of the population, and the uniform crossover operators operate in the same way in every generation. Arithmetic crossover operator combines two parent chromosomes to two offspring chromosomes as shown in

$$y_i^1 = \alpha_i x_i^1 + (1-\alpha_i^1) x_i^2, \tag{1}$$
$$y_i^2 = \alpha_i x_i^2 + (1-\alpha_i^1) x_i^1, \tag{2}$$

where α_i are uniform random numbers. In the non-uniform crossover, the parameter α_i can vary between the iterations, and in uniform crossover the value is constant. There are several studies presenting different approaches and new development on crossover operators.

Mutation Optimization can stop at local minimums if the diversity of the population is low. With random variation in the population, new solutions are found and the diversity will grow. Mutation operation is applied to prevent from stopping at local minimums. With correct design parameters, the mutation operator can be used efficiently. Mutation probability P_m controls how strong effect the mutation brings into the population. A too low value does have only a very little or no effect and too big value could lead to the loss of good genetic material and slowing down the convergence rate. Mutation operator with real-coded chromosomes can be uniform or non-uniform.

Elitism GA operators alter the genetic data of a population towards to better solutions. The best solution is possible to be lost during the iterations. The elitism operator is used to prevent this to happen. The elitism saves the one or several best chromosomes and transfers them directly into new population. If the population size is static, usually the best chromosome replaces the worst chromosome in the new population.

3 Genetic Algorithms in Personalized Nutrition Guidance

The main objective is to find a feasible solution to balance a diet. Initial values for the test case are acquired from Nutri-Flow software. The test environment is developed in Matlab and Genetic Algorithms solver is applied.

3.1 Configuration

Configuration used in [7] is applied in this study, too. Configuration parameters are presented in Table 1. Real-value coding is selected due to search space. Intake levels vary between the foodstuffs in a great level; e.g. cinnamon one teaspoon vs. 1000 g water.

Parameter	Value
Population size	100
P_c	0.8
k	2
P_m	0.01
Elite individuals	5% of the population
Maximum iterations	500

Table 1: GA configuration parameters.

Name	Lower limit	Upper limit
Bread, graham	0	0
Potato, peeled, cooked	**180**	**540**
Minced meat, beef 17% brown sauce, no fat	0	0
Salad, buffet, no dressing	**50**	**150**
Rye bread	0	0
Macaroni casserole, beef-pork, milk 1.5% fat	0	0
Bread, wheat	0	0
Coffee drink, brewed	600	770
Margarine 40%, industrial average	0	0
Cheese, hard cheese, fat 24-27%	**0**	**52**
Margarine 60%	0	0
Fat-free milk, vitamin-D 1 μg	**0**	**510**
Water	**170**	**3000**
Ketchup	0	0
Cider, sweet, 4,7 vol% alcohol	**0**	**1000**
Cookie, oatmeal, industrial	0	0
Bell pepper	**0**	**200**
Orange, peeled	**0**	**800**
Tangerine, peeled	**0**	**450**
Kiwi fruit, peeled	**0**	**300**
Apple, peeled	**0**	**800**
Wok vegetables	**0**	**500**

Table 2: Search space in the test case.

3.2 Initial state

Sum Mikko has a meal diary for one day filled in NutriFlow database. The state of the diet can be presented as membership degrees for each 30 nutrients. The search space has 22 foods, and nine of them should be added and three reduced according to FES output.

3.3 Search space

The search space is generated following output of FES and meal diary of test case Sum Mikko. The search space is presented in Table 2. Foodstuffs to add or reduce are shown in bold. The direction of change is inherited from FES output. The constrains for the intake levels are defined for foods to add with original input M_i, and upper intake recommendation M_u as $[M_i, M_u]$, and foods to reduce with zero and M_i as $[0, M_i]$. Foods with no action needed, constrains are exactly at M_i, $[M_i, M_i]$.

3.4 Objective function

The objective is to balance a diet. In this study, the fitness is evaluated applying membership grades. The distance to ideal state is minimized when membership grades for too little μ_l and too much μ_l are minimized. Calculation of fitness value, F_μ, for nutrients is presented in

$$F_\mu = \sum_1^n (b_{l,n}\mu_{l,n} + b_{u,n}\mu_{u,n}) \tag{3}$$

where n is nutrient Id, $b_{l,n}$ is a weight factor for too low intake level for nutrient n, $\mu_{u,n}$ is a membership grade for too little for nutrient n, $b_{u,n}$ is a weight factor for too much intake level for nutrient n, and $\mu_{u,n}$ is a membership grade for too little for nutrient n.

According to experts in nutrition, too drastic changes are difficult to follow. The balancing should be done in smaller steps. Minimizing fitness value F_d for food intakes take into account the step size for the guidance. Calculation of F_d is presented in

$$F_d = \sum_1^n |d_m|, \tag{4}$$

where m is Id for a foodstuff and d_m is distance from the initial diet. Other nutrition related factors, such as the amount of vegetables, or even the carbon footprint if available, should be taken in account when assessing guidance. F_o is reserved for other variables to be calculated in the fitness value as presented in

$$F_o = \sum_{1}^{n} |k_p|, \tag{5}$$

The objective function minimizes all the fitness values F_μ, F_d, and F_o as presented in

$$MIN(aF_\mu + bF_d + bF_o), \tag{6}$$

where a, b and c are tuning factors for each component. Tuning factors are applied to enhance the importance of the fitness values. The main factor is naturally the nutrient state, and the two other fitness values are used to guide the optimization towards to desired and feasible result.

3.5 Matlab model

This study applies Matlab Optimization Toolbox and Genetic Algorithms solver for testing the system. All the data used in the optimization are first acquired from Nutri-Flow software into Matlab workspace. The results are saved numerically and graphically for further evaluation. Custom objective function is applied according to equations presented in Section 3.4.

4 Results

Convergence of calculations and effects on foodstuff level are used for assessing the operation.

4.1 GA performance

Convergence was recorded on all the test rounds. However, the best result was not reached on all rounds within 500 iteration rounds. Most of the test rounds were stopped by Matlab algorithm after no improvement during last 50 iterations. The stopping point was recorded between 100 and 200 iterations except two tests where the maximum number of iterations was the stopping criteria.

The population size was kept at 100 which was sufficient for the current test case which had 22 genes in each chromosome. It is recognized that the chromosome size varies between the test cases and between the evaluation periods. The population size needs further testing with a larger test case set.

The convergence speed is the fastest at the beginning of the iterations and drops fast as shown in Figure 3.

GA design variables have an effect on the convergence speed and finding the global optimum. For further testing, different values for design variables should be tested.

Id	Initial	GA	Difference
1	0.36	0.53	0.17
2	0.50	0.52	0.02
3	0.35	0.47	0.12
4	0.97	0.61	-0.36
5	0.32	0.39	0.08
6	N/A	N/A	N/A
7	0.00	0.83	0.83
8	0.00	0.69	0.69
9	0.00	0.78	0.78
10	0.92	0.98	0.05
11	0.93	0.84	-0.09
12	0.77	0.87	0.10
13	1.00	0.98	-0.02
14	0.31	0.63	0.32
15	0.68	1.00	0.32
16	1.00	0.94	-0.06
17	0.93	0.93	0.00
18	1.00	0.99	-0.01
19	1.00	1.00	0.00
20	0.64	0.89	0.24
21	N/A	N/A	N/A
22	0.57	0.93	0.37
23	0.81	0.75	-0.06
24	0.54	0.68	0.14
25	0.86	0.92	0.07
26	0.72	0.80	0.08
27	0.44	0.58	0.14
28	1.00	1.00	0.00
29	0.73	0.90	0.16
30	0.99	0.99	0.00

Table 3: Membership grade for ideal input for initial and GA recommendation with difference.

4.2 GA output

The overall assessment is done by evaluating the output of GA on the nutrient level and on the foodstuff level. Only a numerical evaluation is carried in this study. The nutrient level analysis is done by comparing the membership grades for each nutrient. Table 3 presents the results of one GA optimization run. The status of most nutrients has improved, but one value has dropped significantly and the status of five nutrients has

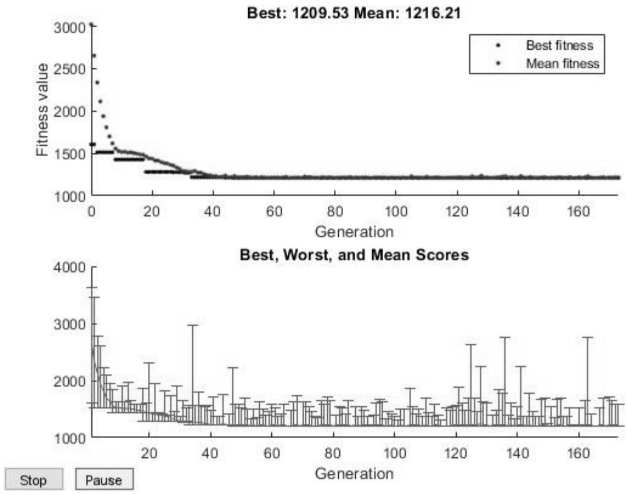

Figure 3: GA optimization with test case Sum Mikko.

moved slightly away from the original ideal status. Two values, added sugar and salt intake are not taken in account in this study directly.

Default feasible range for intake levels for nutrients is set at $[0.5, 1]$ for ideal level. Conclusion of nutrient level evaluation is that the result is at the feasible level. Actual guidance is given as foods and beverages as presented in Table 4. All the recommended values are in the feasible level and the direction of the change is correct. Assessment of level of change should be done as portions. A weekly food plan should be able to generate from the recommendation.

5 Conclusions and Discussion

The test was carried out for one test case. The results are promising, and the nutritional status was improved with the GA optimization. It is important to keep the nutrient intake levels in the feasible range. More studies are needed with a larger test set.

References

[1] Koponen P, Borodulin K, Lundqvist A, Sääksjärvi K, Koskinen S, editors. *Health, functional capacity and welfare in Finland – FinHealth 2017 study, Report 4/2018 (abstract in English)*. 8th ed. Helsinki: THL; 2018. 247 p. http://urn.fi/URN:ISBN:978-952-343-105-8.

[2] Nordic Council of Ministers. *Nordic Nutrition Recommendations 2012*. 5th ed. Copenhagen: Nordisk Ministerråd; 2014. 627 p. http://dx.doi.org/10.6027/Nord2014-002

[3] National Institute for Health and Welfare. *Nutrition Unit: Fineli. Finnish food composition data-base. Release 19*. Helsinki: THL; 2018. www.fineli.fi

[4] Livsmedelsverket. *The National Food Agency food database. Version 2017-12-15*. http://www.slv.se/SokNaringsinnehall/

[5] US Department of Agriculture, Agricultural Research Service, Nutrient Data Laboratory. *USDA National Nutrient Database for Standard Reference. Legacy. Version Current*. USDA; April 2018. https://www.ars.usda.gov/nea/bhnrc/ndl

[6] Nutrition experts team. *Nutrient Density*. Oulu: Nutri-flow; 2018.

[7] Heinonen P, Mannelin M, Iskala H, Sorsa A, Juuso E. Development of a Fuzzy Expert System for a Nutri-tional Guidance Application. In Carvalho P, Dubois D, Kaymak K, Sousa JCM, editors. *Proceedings of 2009 IFSA World Congress / 2009 EUSFLAT Conference*; 2009; Lisboa, Portugal. 1685-1690. ISBN: 978-989-95079-6-8.

[8] Heinonen P, Juuso EK. Development of a Genetic Algorithms Optimization Algorithm for a Nutritional Guidance Application. In Juuso E, Dahlquist E, Leiviskä K, editors. *The 9th EUROSIM Congress on Modelling and Simulation, EUROSIM 2016, The 57th SIMS Conference on Simulation and Modelling SIMS 2016, number 142*; 2018; Oulu, Finland. Linköping: Linköping University Electronic Press. 755–761. doi: 10.3384/ecp1714255.

Name	Recommendation	Level of change
Bread, graham	100	0
Potato, peeled, cooked	**287**	**107**
Minced meat, beef 17% brown sauce, no fat	180	0
Salad, buffet, no dressing	**76**	**26**
Rye bread	60	0
Macaroni casserole, beef-pork, milk 1,5% fat	350	0
Bread, wheat	60	0
Coffee drink, brewed	770	0
Margarine 40%, industrial average	62	0
Cheese, hard cheese, fat 24-27%	**6**	**-46**
Margarine 60%	12	0
Fat-free milk, vitamin-D 1 μg	**203**	**307**
Water	**718**	**548**
Ketchup	30	0
Cider, sweet, 4,7 vol% alcohol	**114**	**-886**
Cookie, oatmeal, industrial	25	0
Bell pepper	**41**	**41**
Orange, peeled	**151**	**151**
Tangerine, peeled	**207**	**207**
Kiwi fruit, peeled	**87**	**87**
Apple, peeled	**269**	**269**
Wok vegetables	**32**	**32**

Table 4: Personalized nutritional guidance in foodstuff level.

[9] Al-Obaidi MA, Li J-P, Kara-Zaïtri C, Mujtaba IM. Optimisation of reverse osmosis based wastewater treatment system for the removal of chlorophenol using genetic algorithms. *Chemical Engineering Journal*. 2017; 316: 91-100. doi: 10.1016/j.cej.2016.12.096.

[10] Abdelaziz M. Distribution network reconfiguration using an geletic algorithm with varying population size. *Electric Power Systems Research*. 2017; 142: 9-11. doi: 10.1016/j.epsr.2016.08.026.

[11] Francescomarino CD, Dumas M, Federici M, Ghidini C, Maggi FM, Rizzi W, Simonetto L. Genetic algorithms for hyperparameter optimization in predictive business process monitoring. *Information Systems*. 2018; 74: 67-83. doi: 10.1016/j.is.2018.01.003.

[12] Holland JH. *Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence*. Ann Arbor: University of Michigan Press; 1975.

[13] Goldberg DE. *Genetic Algorithms in Search, Optimization and Machine Learning*. Boston: Addison-Wesley; 1989.

[14] Singh RK, Panchal VK, Singh BK. A review on Genetic Algorithm and Its Applications. *2018 Second International Conference on Green Computing and Internet of Things (ICGCIoT)*; 2018. 376-380. doi: 10.1109/ICGCIoT.2018.8753030.

[15] Yuan B, Gallagher M. A hybrid approach to parameter tuning in genetic algorithms. *2005 IEEE Congress on Evolutionary Computation*; 2005. 1096-1103. doi: 10.1109/CEC.2005.1554813.

[16] Goldberg DE. Sizing Populations for Serial and Parallel Genetic Algorithms. *3rd International Conference on Genetic Algorithms*; 1989. 70-79.

[17] Diaz-Gomez P, Hougen DF. Initial Population for Genetic Algorithms: A Metric Approach. *2007 International Conference on Genetic and Evolutionary Methods, GEM 2007*; 2007. 8 p.

[18] Davis L. *Handbook of Genetic Algorithms*. New York: Van Nostrand Reinhold; 1991. 385 p.