

G.A. SAMIGULINA, Doctor of Technical Sciences, Head of the lab. "Intelligent Systems of Control and Forecasting", Institute of Information and Computational Technologies, Almaty, Kazakhstan,
Zh.A. MASSIMKANOVA, Master, Junior researcher of the lab "Intelligent Systems of Control and Forecasting", Institute of Information and Computational Technologies, Almaty, Kazakhstan

COMPUTER MODELING OF NEW DRUGS BASED ON THE METHODS OF SWARM INTELLIGENCE AND IMMUNE NETWORK MODELING

The article is devoted to the development of an intellectual information system for conducting scientific research and for predicting the dependence "structure-property/activity" of new drugs based on algorithms of swarm intelligence and immune-network modeling. Selection of informative descriptors is carried out on the basis of a multi-algorithmic approach, which allows selecting the algorithm in which the generalization error will be minimal after immune network modeling. Algorithms are distinguished by the presence of customizable parameters that affect the productivity of an intelligent information system. Figs.: 2. Refs.: 11 titles.

Keywords: medications, swarm intelligence, information system, immunosetting modeling, information descriptors.

The statement of the problem. Designing of new medicinal compounds is a complex process consisting of several stages, which takes a lot of time and requires significant labor and financial costs. The study of chemical compounds involves the processing of multidimensional data sets. The development of computer technology made it possible to apply the methods and algorithms of bioinformatics for the processing and analysis of accumulated chemical data. The application of these methods is relevant for the construction of databases of medicinal compounds, for predicting the physical, chemical and biological properties of substances.

Nowadays the development of modern non-traditional information technologies for the processing of large amounts of data is urgently needed in order to improve the efficiency of the process of drugs creation. Analysis of the structure of compounds with given properties allows to predict the structure of a new compound. Prediction of the structure-property/activity relationship of drug substances and the identification of relationships between the structure of compounds and their activity [1] is an actual

problem in pharmacology.

In prediction tasks the description of the system's behavior based on the optimal set of descriptors has a great importance. Unnecessary and redundant data reduces system productivity. One of the main stages of the process of information extracting from a large volume of accumulated data is the stage of preliminary processing of the initial data, which includes the allocation of informative features. Creation of an adequate model based on the optimal set of descriptors increases the probability of making the right decisions.

The statement of the problem is formulated as follows: it is necessary to solve the problem of informative descriptors selection on the basis of algorithms of the swarm intelligence (algorithm of the ant colony and the particle swarm method) for further immune network modeling of the sulfanilamide group medicines with the given properties. The purpose of an optimal set of descriptors selection is to reduce the dimension of the descriptor space. Actual is the use of modern QSAR (Quantitative structure-activity relationship) methods of drug design based on intellectual approaches that can significantly simplify the process of selecting new chemical compounds with prescribed properties for drug candidates.

Literature review. Today artificial intelligence methods are successfully used to detect new patterns in chemical data. These include neural networks, evolutionary algorithms, artificial immune systems, algorithms of swarm intelligence, etc. At present methods of swarm intelligence are promising directions and are widely used in many fields of medicine and pharmacology. The methods of swarm intelligence, based on the behavior modeling of living beings, include: the methods of ants and bee colony, the algorithm of a particle swarm, cuckoo, gray wolves, bats, gravitational search, etc. The most common of them are the ant colony algorithm and the particle swarm method. The proposed algorithms more effectively explore the search space, spending less time than the methods of full search.

In article [2] there is considered a model based on the ant algorithm for predicting the tertiary structure of a protein. The algorithm is investigated and is focused on proving its convergence to a global solution. This allows for further research to construct procedures for adaptive selection of parameters at each iteration of the algorithm, depending on the nature of its convergence. This method makes it possible to reduce the number of parameters arriving at the input of the algorithm. In work [3] there is proposed a new computer molecular design for the development of solvents based on the algorithm of efficient ant colony optimization (EACO). The article [4] is devoted to ACO-MLR (Ant Colony Optimization-Multiple linear regression) for the development of QSPR (Quantitative structure-property relationship)

algorithm, in which the main problem is the multidimensional descriptor space.

In work [5] there are considered algorithms of swarm intelligence for solving certain problems of bioinformatics, such as data clustering, prediction of the protein structure and molecular docking. In article [6] there are studied a chaotic optimization algorithm and a particle swarm optimization algorithm to improve the classification accuracy, which are used in the selection of data sets with certain pharmacodynamic properties of the drug. The experimental results show that the proposed method has good learning ability, the ability to generalize and the accuracy of classification.

In work [7] there are considered and discussed particle swarm algorithms in problems of secondary prediction of structure, gene association into clusters, energy minimization and protein modeling. The article [8] presents a comparative analysis of the particle swarm method and the bee algorithm as the solution of the protein structure prediction problem. A new method of PSOVina (Particle Swarm Optimization Vina), which combines PSO with the effective local search method of Broyden-Fletcher-Goldfarb-Shannon (BFGS) is proposed in the article [9].

A new algorithm is considered in work [10], which defines associative rules for classification based on swarm intelligence. The main idea is to develop an algorithm that can demonstrate good productivity at the selection of data sets with uneven distribution of classes. The results show that the proposed method works best for weakly and strongly unstructured data sets.

Immune network technology of predicting QSAR of chemical compounds. In work [11] there was developed an immune-network technology for predicting the structure-property/activity dependency of chemical compounds, which consists of preliminary data processing stage, the immune network training stage, image recognition stage, the stage of energy errors evaluation and the prediction of the pharmacological properties of chemical compounds. Intellectual technology based on immune network modeling allows analyzing hidden interactions between descriptors. At the stage of preliminary data processing there is carried out the normalization, completeness check and reliability of the descriptors. The selection of informative descriptors is performed on the basis of algorithms of swarm intelligence in accordance with the concept of multi-algorithm approach, at which several algorithms are used. After the immune network modeling based on the comparison of the prediction results, an algorithm with the least generalization error is selected. The application of algorithms of swarm intelligence allows to reduce the time for immune network training by creating an optimal immune network model and to exclude low-information descriptors.

There was developed an integrated structural scheme of an intelligent information system (Fig. 1), which consists of modules of preliminary processing, image recognition and energy error evaluation. The preprocessing module includes the selection of informative descriptors based on the algorithm of the ant colony and the method of the particle swarm.

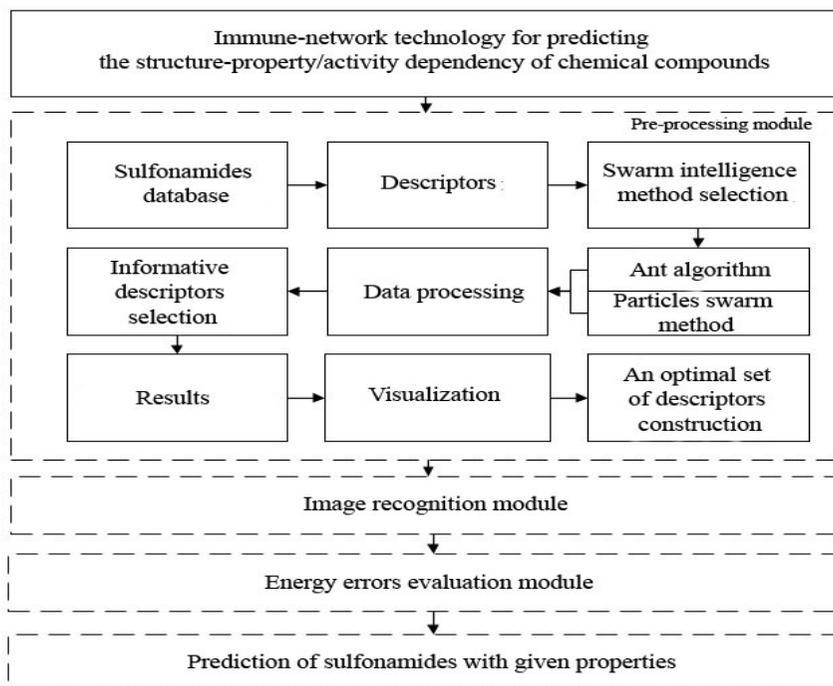


Fig. 1. Structural diagram of intellectual immune network technology

Processing module for chemical compounds "SIIM" (Swarm Intelligence for Immune network Modeling). Developing the primary data processing module "SIIM" there was used the programming language Python 3.6 using PyQt. The following algorithm for the functioning of the "SIIM" system has been developed.

Algorithm:

Step 1. Connecting the database.

Step 2. Choosing a method of swarm intelligence (an ant algorithm or a particle swarm method).

Step 3. Entering the coefficients (number of populations, number of iterations, weight, speed).

Step 4. Data processing and selection of informative descriptors.

Step 5. Construction of an optimal set of descriptors.

Step 6. Output of the result of the algorithm.

Description of the sulfonamides database. Intellectual technology of immune network modeling includes work with chemical compounds descriptors databases. As an example there were used 15 chemical compounds of sulfonamides and 1500 descriptors of different levels. Chemical compounds were classified into short-acting, medium-acting and long-acting sulfanilamides. There were considered such descriptors as: number of atoms, relative number of carbon atoms, relative number of hydrogen atoms, molecular weight, number of single bonds, gravitation index and others.

Experimental results. In the developed intellectual information system "SIIM" (Fig. 2) with the help of "Data connection" button, the database is selected and is displayed in the first window. Then, there is carried out an algorithm selection (an algorithm of an ant colony or a particle swarm method) and coefficients input. In the algorithm of the ant colony the following coefficients are used: the number of population, the number of iterations, the amount of pheromone, the pheromone vaporization, report frequency. After starting, modeling results are displayed in the second window. The system has the ability to visualize in the form of a graph.

The screenshot shows the SIIM software interface with two main optimization panels: Ant Colony Optimization and Particle Swarm Optimization. The Ant Colony Optimization panel has parameters: Population size (100), Number of generations (50), c1 (1), and c2 (2). The Particle Swarm Optimization panel has Report frequency (50). Below these is a 'Database connection' table with 15 rows of chemical classes and their descriptors. To the right is a 'Run' table with 15 rows of optimization results.

Database connection						Run																														
Class	Number of atoms	relative number of C atoms	relative number of H atoms	relative number of O atoms	relative number of N atoms	relative number of S atoms	relative number of Cl atoms	relative number of F atoms	relative number of Br atoms	relative number of I atoms	relative number of P atoms	relative number of B atoms	relative number of Si atoms	relative number of Ti atoms	relative number of Zr atoms	relative number of Hf atoms	relative number of Ta atoms	relative number of Nb atoms	relative number of Mo atoms	relative number of Cr atoms	relative number of Mn atoms	relative number of Fe atoms	relative number of Co atoms	relative number of Ni atoms	relative number of Cu atoms	relative number of Zn atoms	relative number of Ga atoms	relative number of Ge atoms	relative number of As atoms	relative number of Se atoms	relative number of Br atoms	relative number of Kr atoms	relative number of Xe atoms	relative number of Rn atoms		
1 short_acting	27.00	0.37	0.37	0.07	0.15	0.04	1	1.24	-0.38	-0.19	522.00	0.00	0.35	-0.07																						
2 short_acting	33.00	0.36	0.42	0.06	0.12	0.03	2	0.22	-0.06	-0.23	839.00	-0.14	0.77	-0.02																						
3 short_acting	31.00	0.35	0.42	0.10	0.10	0.03	3	1.51	0.37	-0.06	720.00	-0.25	-0.20	-0.03																						
4 short_acting	27.00	0.33	0.37	0.07	0.15	0.07	4	2.59	-0.36	0.03	502.00	-0.29	-0.29	0.15																						
5 short_acting	46.00	0.33	0.54	0.04	0.07	0.02	5	0.09	-0.30	0.08	1040.00	0.08	0.57	-0.07																						
6 short_acting	33.00	0.36	0.42	0.06	0.12	0.03	6	0.22	-0.06	-0.23	839.00	-0.14	0.77	-0.02																						
7 short_acting	46.00	0.33	0.54	0.04	0.07	0.02	7	0.09	-0.30	0.08	1040.00	0.08	0.57	-0.07																						
8 medium_acting	24.00	0.33	0.42	0.13	0.08	0.04	8	2.47	0.48	0.06	321.00	-0.16	-0.53	-0.14																						
9 medium_acting	35.00	0.34	0.40	0.11	0.11	0.03	9	0.45	-0.23	0.02	1050.00	-0.03	0.54	0.00																						
10 medium_acting	28.00	0.36	0.39	0.11	0.11	0.04	10	0.90	-0.57	-0.05	584.00	0.05	0.52	-0.17																						
11 medium_acting	31.00	0.35	0.42	0.10	0.10	0.03	11	0.05	0.00	0.14	625.00	-0.02	0.39	-0.14																						
12 long_acting	35.00	0.34	0.40	0.11	0.11	0.03	12	0.45	-0.23	0.02	1050.00	-0.03	0.54	0.00																						
13 long_acting	36.00	0.42	0.39	0.06	0.11	0.03	13	1.11	0.28	-0.25	1289.00	-0.04	-0.04	0.09																						
14 long_acting	31.00	0.35	0.39	0.10	0.13	0.03	14	0.69	0.21	-0.11	817.00	-0.11	0.11	-0.05																						
15 long_acting	30.00	0.37	0.40	0.07	0.13	0.03	15	1.47	0.04	-0.27	666.00	-0.23	-0.21	0.05																						

Fig. 2. Sulfanilamides modeling results based on the ant algorithm

Fig. 2 shows the results of sulfanilamides modeling based on the ant algorithm. During the modeling the size of the population is 100, the number of iterations is 50, c_1 (the amount of pheromone) = 1, c_2 (pheromone evaporation) = 2, report frequency = 50. As a result, from 1500 descriptors there were selected 25 informative ones.

Similarly to this the intelligent system works with the particle swarm method [12]. There is selected that algorithm which after the immune network modeling shows the best prognostic results.

Conclusion. In this article there was solved the task of informative sulfanilamides descriptors selection for constructing an optimal immune network model and further immune network modeling of new drugs in accordance with the concept of multi-algorithm approach. During the developing of intelligent system for predicting the pharmacological activity of drug compounds based on immune network modeling, the main task is to reduce the generalization error and to improve the quality of the prognosis. Nowadays, there are no universal methods for constructing an optimal set of descriptors.

The advantage of multi-algorithm approach usage is the possibility of choosing an algorithm with the best prognostic properties after immune network modeling. The results of modeling differ from each other depending on the size and type of data, the availability of independent descriptors, the optimality criteria and stopped criteria. The presented intellectual immune network technology of scientific research allows to analyze hidden interactions between descriptors, to reduce financial and computing resources.

The research is conducted on the grant of the Science Committee of the Ministry of Education and Science of the Republic of Kazakhstan (No. GR 0115PK00549) on the theme: Computer molecular design of medicines based on immune network modeling (2015-2017).

References:

1. Golla, S., Neely, B., Whitebay, E., Madihally, S., Robinson, R., and Gasem, K. (2002). "Virtual design of chemical penetration enhancers for transdermal drug delivery", *Chem. Biol. Drug Design*, pp. 478-487.
2. Niu, D., Wang, Y., and Wu, D.D. (2010). "Power load forecasting using support vector machine and ant colony optimization", *Expert Systems with applications*, Vol. 37, pp. 2531-2539.
3. Liu, Y., Wang, G., Chen, H., Zhao, Z., Zhu, X., and Liu, Z. (2011). "An adaptive fuzzy ant colony optimization for feature selection", *Journal of Computational Information Systems*. Vol. 7, pp. 1206-1213.
4. Morteza, A., Kobra, Z., and Azam, B. (2013). "Ant colony optimization as a descriptor selection in QSPR modeling: Estimation of the k-max of anthraquinones-based dyes". *Journal of Saudi Chemical Society*, Vol. 20, Supplement 1, pp. 2-5.
5. Agrawal, S., and Silakari, S. (2015). "A review on application of Particle Swarm Optimization in Bioinformatics", *Current bioinformatics*, Vol. 10, pp. 401-413.

6. Liu, F., and Zhou, Z. (2015). "A new data classification method based on chaotic particle swarm optimization and least square-support vector machine", *Chemometrics and intelligent laboratory systems*, Vol. 147, pp. 147-156.
7. Zakaria, N., and Abdullah, R.A. (2013). "Comparison between Artificial Bee Colony and Particle Swarm Optimization Algorithms for Protein Structure Prediction Problem", *Neural Information Processing*, Vol. 8227, pp. 331-340.
8. Marcus, C., Fong, S., and Siu, W.I. (2015). "PSOVina: The hybrid particle swarm optimization algorithm for protein-ligand docking", *Journal of bioinformatics and computational biology*, Vol. 13, Issue 3, pp. 154-157.
9. Allias, N., Mohamed-Noor, M., Ismail, N., and de Silva, K. (2013). "A Hybrid Gini PSO-SVM Feature Selection: An Empirical Study of Population Sizes on Different Classifier", *Proceedings of First International Conference on Artificial Intelligence, Modelling & Simulation*, IEEE, pp. 94-97.
10. Das, S., Abraham, A., and Konar, A. (2015). "Swarm Intelligence Algorithms in Bioinformatics". www.softcomputing.net/cib-web.pdf.
11. Samigulina, G.A., and Samigulina, Z.I. (2017). "Immune Network Technology on the basis of random forest algorithm for computer-aided drug design". *Bionformatics and biomedical engineering*, pp. 50-61.

Статтю представив д-р техн. наук., проф. НТУ "ХПІ" Леонов С.Ю.

Поступила (received) 21.11.2017

Samigulina Galina, Dr. Sci. Tech,
Institute of Information and Computational Technologies,
Str. Pushkin, 125, Almaty, Kazakhstan, 050010,
Tel:+7(777)244-43-67, e-mail: galinasamigulina@mail.ru

Massimkanova Zhazira, master,
Institute of Information and Computational Technologies
Str. Pushkin, 125, Almaty, Kazakhstan, 050010,
Tel:+7(707) 834-61-77, e-mail: masimkanovazh@gmail.com

УДК 004.89:004.4

Комп'ютерне моделювання нових лікарських препаратів на основі алгоритмів ройового інтелекту і імунносетевого моделювання / Самігуліна Г.А., Масімканова Ж.А. // Вісник НТУ "ХПІ". Серія: Інформатика та моделювання. – Харків: НТУ "ХПІ". – 2017. – № 50 (1271). – С. 88 – 95.

Стаття присвячена розробці інтелектуальної інформаційної системи ведення наукових досліджень і прогнозуванню залежності структура-властивість / активність нових лікарських препаратів на основі алгоритмів ройового інтелекту і імунносетевого моделювання. Відбір інформативних дескрипторів здійснюється на основі мультіалгоритмічного підходу, який дозволяє вибрати той алгоритм, при якому після імунносетевого моделювання помилка узагальнення буде мінімальною. Алгоритми відрізняються наявністю настроюються параметри, які впливають на продуктивність інтелектуальної інформаційної системи. Ил.: 2. Библиорг.: 11 назв.

Ключові слова: лікарські препарати; ройовий інтелект; інформаційна система; імунносетевое моделювання; інформаційні дескриптори.

УДК 004.89:004.4

Компьютерное моделирование новых лекарственных препаратов на основе алгоритмов роевого интеллекта и иммунносетевого моделирования / Самигулина Г.А., Масимканова Ж.А. // Вестник НТУ "ХПИ". Серія: Інформатика и моделирование. – Харьков: НТУ "ХПИ". – 2017. – № 50 (1271). – С. 88 – 95.

Статья посвящена разработке интеллектуальной информационной системы ведения научных исследований и прогнозирования зависимости структура-свойство / активность новых лекарственных препаратов на основе алгоритмов роевого интеллекта и иммунносетевого моделирования. Отбор информативных дескрипторов осуществляется на основе мультиалгоритмического подхода, который позволяет выбрать тот алгоритм, при котором после иммунносетевого моделирования ошибка обобщения будет минимальной. Алгоритмы отличаются наличием настраиваемых параметры, которые влияют на производительность интеллектуальной информационной системы. Ил.: 2. Библиогр.: 11 назв.

Ключевые слова: лекарственные препараты; роевой интеллект; информационная система; иммунносетевое моделирование; информационные дескрипторы.

УДК 004.89:004.4

Computer modeling of new drugs based on the methods of swarm intelligence and immune network modeling / Samigulina G.A., Massimkanova Zh.A. // Herald of the National Technical University "KhPI". Subject issue: Information Science and Modelling. – Kharkov: NTU "KhPI". – 2017. – №. 50 (1271). – P. 88 – 95.

The article is devoted to the development of an intellectual information system for conducting scientific research and for predicting the dependence "structure-property/activity" of new drugs based on algorithms of swarm intelligence and immune-network modeling. Selection of informative descriptors is carried out on the basis of a multi-algorithmic approach, which allows selecting the algorithm in which the generalization error will be minimal after immune network modeling. Algorithms are distinguished by the presence of customizable parameters that affect the productivity of an intelligent information system. Figs.: 2. Refs.: 11 titles.

Keywords: medications, swarm intelligence; information system; immunosetting modeling; information descriptors.