Construction of Optimal Immune Network Model Based on Swarm Intelligence Algorithms for Computer-aided Design of New Drugs

Galina A. Samigulina and Zhazira A. Massimkanova

Institute of Information and Computational Technologies, Almaty, Kazakhstan
galinasamigulina@mail.ru
masimkanovazh@gmail.com

Abstract. Nowadays the development of information technologies based on bioinspired intellectual approaches for the computer design of new drugs and forecasting of their properties is an urgent task. The research is devoted to the development of an intellectual information system for conducting scientific researches and forecasting the structure-property/activity relationship of new drugs based on artificial immune systems approach. In accordance with the concept of multi-algorithmic approach, the construction of an optimal immune network model and the allocation of informative descripts are carried out using swarm intelligence algorithms: modified algorithms of ant colonies and particle swarms. The developed information system allows selecting the best algorithm for preliminary data processing, in which after immune network modeling, the value of generalization error will be minimal. The use of multi-algorithm approach at immune network modeling of drugs requires the systematization of used algorithms and the creation of an integrated ontological model, which allows structuring the input and outputting data. There is presented an example of the database of sulfanilamides with different pharmacological activity, also modeling results and comparative analysis of the use of various algorithms of swarm intelligence.

Keywords: Swarm intelligence · Drug design · Optimal immune network model.

1 Introduction

Nowadays modern methods of artificial intelligence are widely used in pharmacology for computer modeling of new drug compounds with pre-defined properties. The creation and investigation of chemical compounds is associated with the processing of multidimensional data. The development of information technologies based on intellectual approaches for processing and analysis a large data sets and solution of forecasting is an actual problem. Neural networks [1], genetic algorithms [2], artificial immune systems (AIS) [3, 4], swarm intelligence algorithms [5] and others are widely used in medicine. The article [6] presents the joint use of modified AIS and a partial

In: S. Belim et al. (eds.): OPTA-SCL 2018, Omsk, Russia, published at http://ceur-ws.org

least square regression method for breast cancer diagnosis. The proposed approach has a significant effect on the classification accuracy for clinical diagnosis and can be used to solve detection problems. The research [7] describes artificial immune recognition algorithm, which shows the highest classification accuracy for tuberculosis diagnosis. The method can be applied for any diagnostics, the classification accuracy will be high, especially for large data sets.

Forecasting quantitative structure-activity relationship (QSAR) of drugs and identifying the links between compounds structure and their activity is an actual problem in pharmacology. One of the important steps in the process of forecasting structure-property/activity relationship is the selection of informative descriptors for reducing the size of a descriptor space.

Nowadays bioinspired intellectual approaches to solve optimization problems are actively developed. Swarm intelligence algorithms, such as ant colony optimization (ACO) and particle swarm optimization (PSO), which based on animal and insect life behavior to find the shortest path between food source and their nests, are applied in QSAR modeling [8]. The article [9] describes optimization algorithms (evolutionary algorithms, particle swarm algorithms) to determine the best position of a ligand in protein-ligand docking. The research is executed using program AutoDock. The results show the effectiveness of particle swarm algorithm, especially for highly flexible ligands. The work [10] presents a new approach, which is called two-step swarm intelligence. The main idea of the algorithm is to divide the heuristic search into two stages. At the first stage, agents create solutions that are used as the initial states in the second stage. This algorithm is used in joint with ACO and PSO algorithms. The obtained experimental results demonstrate that the two-step swarm intelligence improves the characteristics of used algorithms. In research [11], there is proposed PSO algorithm to search an optimal number of features and reduce the dimensionality of spectral image data. The study [12] deals with particle swarm and firefly algorithm for diagnosing a tumor in images of magnetic resonance and computer tomography. In article [13], there is studied a joint use of chaotic optimization algorithm and PSO algorithm to improve the classification accuracy, which are used in the selection of data sets with certain pharmacodynamic properties of drug. The experimental results present that the proposed method has good learning performance, strong generalization ability and classification accuracy.

At construction the optimal immune network model for forecasting QSAR of drugs, there is relevant to use an ontological approach, which allows structuring the input and output data, take into account the features of functioning and interconnection, and save the time and computing resources while developing the information system. In paper [14], an improved PSO algorithm based personalized ontology model is described. The model creates personalized user profiles and finds information about users from local repositories. The experimental results show that the proposed PSO algorithm based personalized ontology model is effective in comparison with other models. In article [15], there is introduced PSO algorithm based on semantic relations and tested on the engineering applications. The experimental results demonstrate that for small data sets the optimization ability of PSO algorithm based on the semantic relations is better than classical algorithms. The algorithm allows finding the

optimal value for short period. The work [16] provides PSO algorithm in ontology repository for semantic web service selection.

The following structure of the article is proposed: Section 2 describes problem statement of the research and the immune network technology for forecasting of QSAR of chemical compounds. Section 3 presents the creation of ontological models of swarm intelligence algorithms. Section 4 is devoted to the development an information system of forecasting for conducting scientific researches "SIIM" (Swarm intelligence for immune network modeling) and the description of the database of sulfonamides with different duration of action. Section 5 presents the modeling results using the chemical compounds of sulfanilamide group as an example and comparative analysis of used algorithms. At the end of the article, conclusion and references are presented.

2 Problem Formulation and Solution Methods

The problem statement is formulated as follows: it is necessary to develop an information system of forecasting for conducting scientific researches "SIIM" (Swarm intelligence for immune network modeling) for creation an optimal immune network model of drug compounds of sulfanilamide group based on swarm intelligence algorithms: modified algorithms of ant colonies and particle swarms.

Definition: optimal immune network is a network constructed based on the weight coefficients of the selected informative descriptors and most fully characterizing the considered chemical compound. The criterion of optimization is the storage of maximum information at a minimum number of descriptors [3, 17].

The intellectual technology for forecasting the properties of new chemical compounds consists of the stages of preliminary data processing, immune network training, image recognition, energy error estimation and selection of candidates of drug compounds [18]. Preliminary data processing includes the description of chemical compounds in the form of descriptors, normalization, verification of the completeness and reliability of descripts, also the reduction of low informative descriptors. The selection of informative descriptors is performed using swarm intelligence algorithms based on multi-algorithm approach [19], which allows to use several algorithms.

The development of integrated ontological model (OM) allows to study of subject domain of AIS in detail and to analyze of swarm intelligence algorithms deeply. The use of modern ontological editors and the creation of OM facilitate the solution of problem of selection informative descriptors and the construction of an optimal immune network model. As a tool for developing OM, there is chosen the ontology editor Protégé [20].

3 The Creation of Ontological Models

The integrated OM of immune network technology, which consists of OM of preliminary data processing, OM of image recognition and OM of energy error estimation of AIS has been proposed. The integrated OM is presented in the form of a tuple of sets:

$$OM_{INT} = <OM_{PR}, OM_{IR}, OM_{EEE}>,$$

where OM_{PR} – OM of preliminary data processing;
OM_{IR} – OM of image recognition based on AIS;
OM_{EEE} – OM of energy error estimation of AIS.

The ontological model of preliminary data processing consists of ontological models of algorithms of ant colony and particle swarms:

$$OM_{PR} = <OM_{ACO}, OM_{PSO}>,$$

where OM_{ACO} – OM of algorithms of ant colony;
OM_{PSO} – OM of algorithms of particle swarms.

There are many modifications of classical swarm intelligence algorithms for preliminary data processing. Ant colony algorithm has several modifications such as AntSrank, Max-min ant system, Elitist ant system, etc. In addition, particle swarm algorithm has following modifications: CoPSO, Fully informed PSO, Inertia Weighted PSO, etc. Table 1 shows OM of swarm intelligence algorithms, OM of image recognition, OM of energy error estimation.

Table 1. Content of ontological models.

Ontological model	Content
Ontological model of algorithms of ant colony	Algorithms of ant colony: – AntSrank algorithm. – Max-min ant system algorithm. – Elitist ant system algorithm. – Ant-Q algorithm. – Classical BasicACO algorithm: Generation of population size of agents. Random permutation of agents. Initialize the amount of pheromone. Calculation of fitness function. Determination the amount of pheromone. Permutation of agents. Delay the amount of pheromone. Update local and global amount of pheromone. Check stop condition. Save global best position (gbest).

Table 1 (continued)

Ontological model of algorithms of particle swarms	Algorithms of particle swarms: – CoPSO algorithm. – Fully informed PSO algorithm. – Inertia Weighted PSO algorithm. – Time-Varying Inertia Weighted PSO algorithm. – Classical BasicPSO algorithm: Generation of population size of agents. Generation of random position and random velocity of agents. Calculation of fitness function. Determination the best position of agents. Migration of agents. Update position and velocity of agents. Update the best position of agents. Check stop condition. Save global best position (gbest) of agents.
Ontological model of image recognition	Implementation of AIS technology and the creation of an optimal immune network model [21, 22]: - creation of matrixes of standards and matrixes of images formed from time series (descriptors). - training of AIS with the teacher. - singular value decomposition (SVD). - determination of binding energies between formal peptides. - solution of the problem of image recognition based on the determination of the minimum value of the binding energy and forecasting.
Ontological model of energy error estimation of AIS	Energy error estimation of AIS [21, 23]: - Averaging of the potentials by the homologies. - Calculation of the average amount of standard deviations between the native structure energy and the energy of the randomly chosen stacking of the chain. - Determination of prediction risk factors.

Swarm intelligence algorithms at selection of informative descriptors show different results depend on the size and quality of data, the availability of independent parameters and the optimality criteria. There are no universal algorithms for preliminary data processing. The advantage of using a multi-algorithmic approach is the possibility of choosing swarm intelligence algorithm, which allows to create an immune network model with the best prognostic properties and shows the minimum value of generalization error of AIS. Image recognition and energy error the estimation of AIS are described in work [21].

4 Information System of Forecasting for Conducting Scientific Researches «SIIM»

Information system of forecasting for conducting scientific researches «SIIM» [24] is used for selection informative descriptors at preliminary data processing. The information system is developed in programming language Python 3.6. At the first step of the information system, there is connected a database of descriptors of chemical compounds. The database is displayed on the left screen of the interface. At next step swarm intelligence algorithm (ant colony optimization or particle swarm optimization) is chosen. The fields for input coefficients is displayed. A coefficients are introduced depending on selected algorithm. For example, for particle swarm optimization algorithm the coefficients are as followings: population size, iteration numbers, weight and velocity. After introducing all coefficients it is need to click "Run" button for calculation. The processing of multidimensional data is performed, the allocation of informative descriptors and the construction of an optimal immune network model are implemented. Modeling results are displayed on the right screen of the interface. By comparing the results of AIS prediction, there is chosen swarm intelligence algorithm with the minimum value of generalization error.

As the database has been used a database of descriptors of sulfanilamide group with pre-defined pharmacological properties on the basis of the resource Mol-instincs and PubChem [25]. The database consists more than 1500 descriptors. Table 2 shows a fragment of the sulfanilamide database, which are classified into short acting, medium acting and long acting sulfanilamides.

Table 2. A fragment of the sulfanilamide database.

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 aromatic bonds
short_acting	27,00	0,37	0,37	0,07	0,15	…	0,43
medium_acting	24,00	0,33	0,42	0,13	0,08	…	0,13
short_acting	31,00	0,35	0,42	0,10	0,10	…	0,19
long_acting	35,00	0,34	0,40	0,11	0,11	…	0,31
long_acting	36,00	0,42	0,39	0,06	0,11	…	0,41
short_acting	33,00	0,36	0,42	0,06	0,12	…	0,29
medium_acting	28,00	0,36	0,39	0,11	0,11	…	0,32
short_acting	46,00	0,33	0,54	0,04	0,07	…	0,17
…	…	...	…	…	…	…	…
long_acting	30,00	0,37	0,40	0,07	0,13	…	0,29

5 Modeling Results and Comparative Analysis

At modeling sulfanilamides based on PSO algorithm the population size is 100, iteration number is 50, c1 (weight) = 1, c2 (velocity) = 2, report frequency equal to 50. As a result, there were selected 49 informative descriptors from 1500 ones. Figure 1 shows the graph of selected informative descriptors based on PSO algorithm.

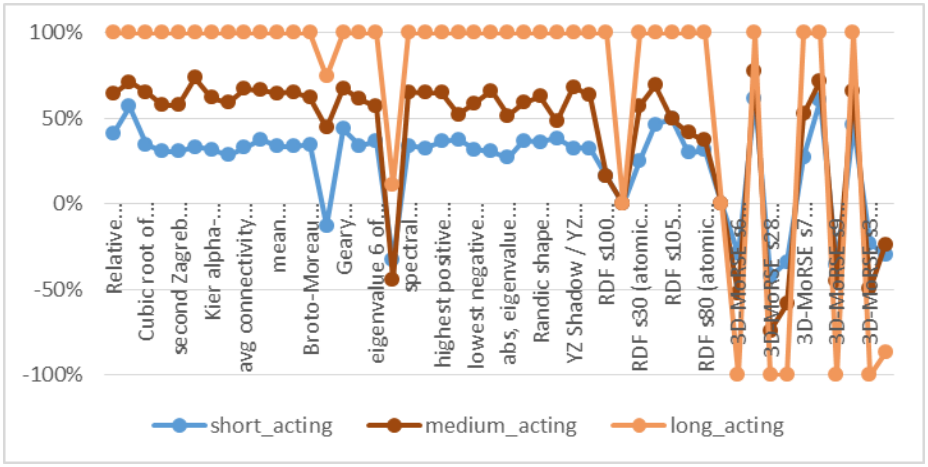

Fig. 1. The visualization of the selected informative descriptors based on PSO algorithm with a population size of 100

If at modeling the population size is 200, then, as a result, there are selected 11 informative descriptors from 1500 ones (Fig. 2).

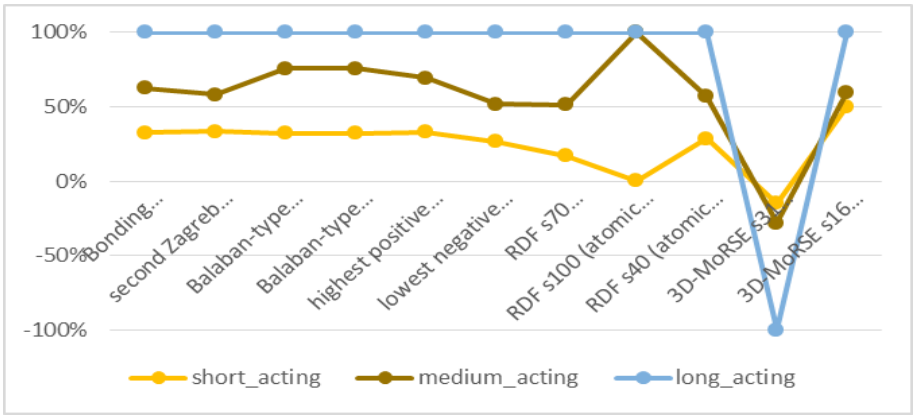

Fig. 2. The visualization of the selected informative descriptors based on PSO algorithm with a population size of 200

From comparison purposes, both algorithms have the same iteration numbers and population sizes. The comparative analysis allows to define optimization algorithm with the best performance and low execution time. Table 3 shows a comparison of modeling results based on algorithms of ant colonies and particle swarms.

Table 3. Comparative analysis of the sulfonamides modeling results

Coeffecients	Ant colony algorithm		Coeffecients	Particles swarm algorithm	
Population size	100	200	Population size	100	200
Iteration numbers	50		Iteration numbers	50	
Amount of pheromone	1		Weight	1	
Evaporation of pheromone	2		The velocity of the particle	2	
Report frequency	50		Report frequency	50	
	Sulfonamides modeling results				
General amount of descriptors	1500	1500	General amount of descriptors	1500	1500
The amount of informative descriptors	25	17	The amount of informative descriptors	49	11

The amount of population size and iterations affect the effectiveness of swarm intelligence algorithms. A large number of populations and iterations allows agents to explore descriptor space more detailed and reduce the number of selected informative descriptors. The running time of PSO algorithm is 8 seconds when the population size is equal to 100 particles. The running time of ACO algorithm is 20 seconds with population size of 100. The modeling results show, PSO algorithm is considered as best optimization method with minimum execution time.

6 Conclusion

Therefore, the study of structure-property/activity relationship of drug compounds, the development of new non-traditional intellectual approaches of QSAR and the computer molecular design of drugs with pre-defined properties are one of the most actual and main tasks of modern pharmacology aimed at reduction the time and cost of creating new drugs. The developed information system for conducting scientific researches “SIIM” with the use of immune network technology for forecasting structure-property/activity relationship of chemical compounds and multi-algorithmic approach, which allows integrating different methods of artificial intelligence to solve the problem of computer molecular design of new drugs with pre-defined properties. The application of multi-algorithmic approach with the use of modified algorithms of ant colonies and particle swarms allows performing preliminary data processing efficiently, allocating the informative set of descriptors and creating an optimal immune

network model. At developing an information system, the use of ontological models allows structuring data and analyzing the hidden interactions between descriptors. The advantage of immune network modeling technology is the energy error estimation by homologues [21], which allows to separate chemical compounds with almost identical structure, but belonging to different classes of pharmacological activity.

The work was carried out according to the grant of the CS MES RK on the theme: "Development and analysis of databases for the information system for prediction the "structure-property" dependence of drug compounds based on artificial intelligence algorithms" (2018-2020).

References

1. Zamani, M., Kremer, S.C.: Neural networks in bioinformatics. In: Handbook on Neural Information Processing. vol. 49, pp. 505-525, Springer-Verlag, Berlin Heidelberg (2013)
   Wong, K.: Evolutionary Algorithms: Concepts, Designs, and Applications in Bioinformatics. Nature-Inspired Computing: Concepts, Methodologies, Tools, and Applications, https://www.igi-global.com/chapter/evolutionary-algorithms/161025, [accessed 23 March 2018]
2. Tarakanov, A., Nicosia, G.: Foundations of immunocomputing. In: Proceedings of the 1-st IEEE Symposium on Foundations of Computational Intelligence (FOCI'07). pp. 503–508. IEEE, Honolulu (2007)
3. Dasgupta, D.: Recent advances in artificial immune systems: models and applications. Applied Soft Computing Journal 11, 1574–1587 (2011)
4. Das, S., Abraham, A., Konar, A.: Swarm intelligence algorithms in bionformatics. Studies in Computational Intelligence (SCI) 94, 113-147 (2008)
5. Zhao, W., Davis C.E.: A modified artificial immune system based pattern recognition approach an application to clinic diagnostics. Artificial Intelligence in Medicine 52(1), 1–9 (2011)
6. Shamshirband, S., Hessam, S., Javidnia, H., Amiribesheli, M., Vahdat, S., Petković, D., Gani, A., Kiah, M.L.: Tuberculosis disease diagnosis using artificial immune recognition system. International Journal Medical Science 11(5), 508–514 (2014)
7. Goodarzi, M., Dejaegher, B.: Feature selection methods in QSAR studies. Journal AOAC Int. 95(3), 636–651 (2012)
8. Guo, L., Yan, Z., Zheng, X., Hu, L., Yang, Y., Wang, J.: A comparison of various optimization algorithms of protein-ligand docking programs by fitness accuracy. Journal Molecular modeling 20(7). http://doi.org/10.1007/s00894-014-2251-3, [accessed 20-January-2018]
9. Gómez, Y., Bello, R., Puris, A., García, M.M.: Two step swarm intelligence to solve the feature selection problem. Journal of Universal Computer Science 14, 2582–2596 (2008)
10. Monteiro, S.T., Kosugi, Y.: Applying particle swarm intelligence for feature selection of spectral imagery. http://citeseerx.ist.psu.edu/doi=10.1.1.726.8470, [accessed 20 November 2017]
11. Thamaraichelvi, B., Yamuna, G.: Hybrid firefly swarm intelligence based feature selection for medical data classification and segmentation in SVD - NSCT domain. International Journal of Advanced Research. http://dx.doi.org/10.21474/IJAR01/1544, [accessed 10 January 2018]

12. Liu, F., Zhou, Z.: A new data classification method based on chaotic particle swarm optimization and least square-support vector machine. Chemometrics and intelligent laboratory systems, 147–156 (2015)
13. Reha, S., Geetharani, M.: Swarm intelligence based fuzzy with personalized ontology model for web information gathering. International Journal of Advances in Computer Science and Technology 3(1). http://warse.org/pdfs/2014/ijacst07312014.pdf, [accessed 12 December 2017]
14. Shao, L., Bai, Y., Qiu, Y., Du, Zh.: Particle swarm optimization algorithm based on semantic relations and its engineering applications. Systems Engineering Procedia 5, 222–227 (2012)
15. Yaamini, S.K., Suganya, I.: Semantic Web service selection using particle swarm optimization. International Journal of Engineering Trends and Technology (IJETT) 4(603) (2013)
16. Samigulina, G.A.: Immune Network Modeling Technology for Complex Objects Intellectual Control and Forecasting System: Monograph. Science Book Publishing House, USA (2015)
17. Samigulina, G.A., Samigulina, Z.I.: Immune network technology on the basis of random forest algorithm for computer-aided drug design. Bionformatics and biomedical engineering, 50-61, Spain (2017)
18. Samigulina, G.A., Samigulina, Z.I.: The construction of an optimal immune network model for forecasting the properties of unknown drug compounds on the basis of multialgorithmic approach. Problems of informatics 2, 21-29 (2013)
19. Samigulina, G.A., Massimkanova, Z.A.: Ontological models of swarm intelligence algorithms for immune network modeling of drugs. Bulletin of Al-Farabi KazNU 1(93), 92-104 (2017)
20. Samigulina, G.A.: Development of the decision support systems on the basis of the intellectual technology of the artificial immune systems. Automatic and remote control. 74(2), 397-403, Springer (2012)
21. Samigulina, G.A., Samigulina, Z.I.: Development of the technology of immune network modeling for computer-aided molecular design of drugs (computer program). Certificate of state registration of the object of intellectual property No. 473, Astana (2011)
22. Samigulina, G.A., Samigulina, Z.I.: Energy errors estimation of artificial immune system on homologues (computer program). Certificate of state registration of the object of intellectual property No. 396, Astana (2008)
23. Samigulina, G.A., Massimkanova, Zh.A.: Swarm intelligence for immune network modeling “SIIM” (computer program). Certificate of state registration of the object of intellectual property No. 319, Astana (2017)
24. https://www.molinstincts.com, [accessed 21 April 2017]