

Г.А. Самигулина<sup>1</sup>, Ж.А. Масимканова<sup>2</sup><sup>1</sup> *Институт информационных и вычислительных технологий КН МОН РК,  
Алматы, Казахстан*<sup>2</sup> *Казахский национальный университет им. аль-Фараби, Алматы, Казахстан*

## РАЗРАБОТКА МОДИФИЦИРОВАННОГО АЛГОРИТМА НА ОСНОВЕ КООПЕРАТИВНОГО АЛГОРИТМА РОЯ ЧАСТИЦ, ИСКУССТВЕННЫХ ИММУННЫХ СИСТЕМ И ОНТОЛОГИЧЕСКОГО ПОДХОДА

**Аннотация.** Статья посвящена разработке модифицированного алгоритма на основе кооперативного алгоритма роя частиц и искусственных иммунных систем для прогнозирования зависимости «структура-свойство/активность» (QSAR) лекарственных соединений. Построение оптимального набора дескрипторов выполняется на основе кооперативного алгоритма роя частиц. Распознавание образов и прогноз молекул-кандидатов в лекарственные препараты осуществляется с использованием искусственных иммунных систем (ИИС). Разработанный модифицированный алгоритм позволяет сократить время обучения ИИС и повысить точность прогнозирования за счет построенного оптимального набора данных. Для систематизации данных и облегчения создания программного обеспечения построена OWL модель кооперативного алгоритма роя частиц в редакторе онтологий Protégé. Приведены результаты моделирования с использованием базы данных сульфаниламидов.

**Ключевые слова:** Smart-технология прогнозирования, зависимость «структура-свойство/активность» (QSAR), кооперативный алгоритм роя частиц, искусственные иммунные системы, OWL модель, лекарственные препараты

### Введение

В настоящее время актуальна разработка и внедрение эффективных Smart-технологий на основе интеллектуальных методов для обработки и анализа химических данных и прогнозирования зависимости «структура-свойство/активность» (QSAR) новых лекарственных соединений. Разработка нового лекарственного препарата занимает порядка 10 лет и стоит приблизительно 2 558 млрд. долларов США [1]. Несмотря на высокий уровень инвестиций, только 13% отбирается в кандидаты лекарственных препаратов. Широкое применение получили Smart-технологии для прогнозирования биологической активности, токсичности, физико-химических свойств, побочных эффектов и др. Активно ведутся исследования [2] с использованием нейронных сетей, алгоритмов роевого интеллекта, искусственных иммунных систем, генетических алгоритмов и др. Широко используются алгоритмы роевого интеллекта для предварительной обработки химических данных, выделения информативных дескрипторов

и построения оптимальной иммуносетевой модели [3]. Применение этих алгоритмов позволяет облегчить процесс прогнозирования и отбора новых химических соединений с желаемыми свойствами в потенциальные кандидаты лекарственных препаратов.

Существует ряд работ по данной теме. Статья [4] посвящена сравнению метода роя частиц (PSO, particleswarmoptimization) и генетического алгоритма для создания QSAR модели циклооксигеназы (COX, cyclooxygenase). В результате сравнения PSO алгоритм выбран для отбора дескрипторов, которые показывают активности селективность химического соединения к COX. В работе [5] предлагаются эволюционные алгоритмы PSO алгоритм для прогнозирования свойств химических соединений. Исследования выполнены с использованием программного продукта AutoDock. Сравнение результатов моделирования показывает эффективность применения PSO алгоритма. Исследования [6] посвящены совместному использованию хаотического алго-

ритма оптимизации (COA, Chaoticoptimizationalgorithm) и PSO алгоритма для улучшения точности прогноза при выборе наборов данных с определенными фармакологическими свойствами. Результаты эксперимента показывают, что предложенная модель имеет хорошую обучаемость и повышает точность прогнозирования. В статье [7] сравниваются интеллектуальные методы (k-ближайших соседей, логической регрессии, метод опорных векторов и искусственных иммунных систем) для определения индицирующего потенциала фосфолипида. Наилучшие прогнозы получены с помощью искусственных иммунных систем и метода опорных векторов.

Проведенный анализ современных публикаций по данной тематике подтверждает актуальность применения модифицированных алгоритмов роя частиц и искусственных иммунных систем для прогнозирования зависимости «структура – свойство/активность» химических соединений.

**Постановка задачи** формулируется следующим образом: необходимо разработать модифицированный алгоритм на основе кооперативного алгоритма роя частиц, искусственных иммунных систем и онтологического подхода для прогнозирования зависимости «структура-свойство/активность» новых лекарственных препаратов сульфаниламидной группы.

Для более эффективного и детального исследования многомерного пространства, построения оптимальной иммунносетевой модели, уменьшения времени обучения ИИС, повышения точности прогнозирования разработан модифицированный алгоритм (CPSO-AIS) на основе кооперативного алгоритма роя частиц (CPSO, Cooperativeparticleswarmoptimization) [8] и подхода искусственных иммунных систем (AIS, Artificialimmunesystem).

*Алгоритм CPSO-AIS:*

**Шаг 1.** Решается задача построения информативного набора дескрипторов на основе кооперативного алгоритма роя частиц.

**Шаг 2.** Выполняется инициализация пространства поиска. Создаются популяции агентов основного роя и подчинённых роев. Производится случайная генерация начальных положений и скоростей всех агентов.

**Шаг 3.** Выполняется параллельное вычисление фитнес-функций каждого подчинённого роя. Определяются лучшие положения агентов подчинённых роев ( $p_g^S$ ). Сравниваются лучшие положения агентов подчинённых роев.

**Шаг 4.** Передаются лучшие положения в основной рой. Обновляются положения ( $x_i^M$ ) и скорости ( $v_i^M$ ) агентов основного роя по формулам [9]:

$$v_i^M(t+1) = wv_i^M(t) + r_1c_1 + \Phi r_2c_2(p_g^M - x_i^M(t)) + (1 - \Phi)r_3c_3(p_g^S - x_i^M(t)),$$

$$x_i^M(t+1) = x_i^M(t) + v_i^M(t), \quad i = 1, 2, \dots, n,$$

где  $t$  – итерация,  $w$  – вес инерции,  $M$  – мастер-рой,  $S$  – подчинённые рои,  $r_1, r_2, r_3$  – равномерно распределенные случайные числа в интервале  $[0, 1]$ ,  $c_1, c_2, c_3$  – коэффициенты ускорения,  $p_g^M$  – лучшее положение частицы в мастер-рое,  $p_g^S$  – лучшее положение частицы в подчинённых роях.

**Шаг 5.** Осуществляется миграция всех агентов по формуле:

$$\phi = \begin{cases} 0 & Gbe t^S < Gbe t^M, \\ 0.5 & Gbe t^S = Gbe t^M, \\ 1 & Gbe t^S > Gbe t^M. \end{cases}$$

где  $\phi$  – фактор миграции,  $Gbe t^S$  – значение фитнес-функции, определенное для  $p_g^S$ ,  $Gbe t^M$  – значение фитнес-функции, определенное для  $p_g^M$ .

**Шаг 6.** Обновляются положения и скорости агентов подчинённых роев. Проверяются условия завершения. Сохраняется лучшее положение агента основного роя ( $p_g^M$ ).

**Шаг 7.** Осуществляется редукция малоинформативных дескрипторов. Формируется БД выделенного информативного набора дескрипторов.

**Шаг 8.** Решается задача распознавания образов на основе иммунносетевых

алгоритма.

*Шаг 9.* Построение оптимальной иммуносетевой модели на основе выделенного набора дескрипторов. Под оптимальной структурой иммунной сети понимается сеть, построенная на основе весовых коэффициентов выделенных информативных дескрипторов, наиболее полно характеризующих рассматриваемое химическое соединение. Критерием является максимальное сохранение информации при минимальном количестве дескрипторов.

*Шаг 10.* Создание матриц эталонов и матриц образов, сформированных из временных рядов дескрипторов.

*Шаг 11.* Обучение ИИС с учителем.

*Шаг 12.* Сингулярное разложение матриц (SVD). Определение энергий связи. Оценка энергетических погрешностей по гомологичным белкам [10].

*Шаг 13.* Распознавание образов на основе определения минимума значения энергии связи.

*Шаг 14.* Прогноз свойств и отбор кандидатов новых химических соединений.

Для анализа применяемых алгоритмов, структуризации данных и создания программного обеспечения актуальна разработка онтологических (OWL, WebOntologyLanguage) моделей. На рисунке 1 представлена OWL модель CPSO алгоритма в редакторе онтологий Protégé. В алгоритме создается один основной рой (Generation\_MaterSwarm) и несколько подчинённых роев (Generation\_SlaveSwarms). Начальные положения (Random\_Positions) и скорости (Random\_Velocities) агентов задаются случайно. Далее параллельно вычисляются фитнес-функции каждого подчинённого роя и сравниваются их лучшие значения (Comparison\_Gbests). Затем обновляются скорости и положения агентов основного роя. Далее агенты перемещаются, обновляют положения (Update\_SlaveSwarms\_Positions) и скорости (Update\_SlaveSwarms\_Velocities) подчинённого роя. Проверяются условия завершения и после каждой итерации сохраняется лучшее значение агента

основного роя (MasterSwarm\_Gbest).

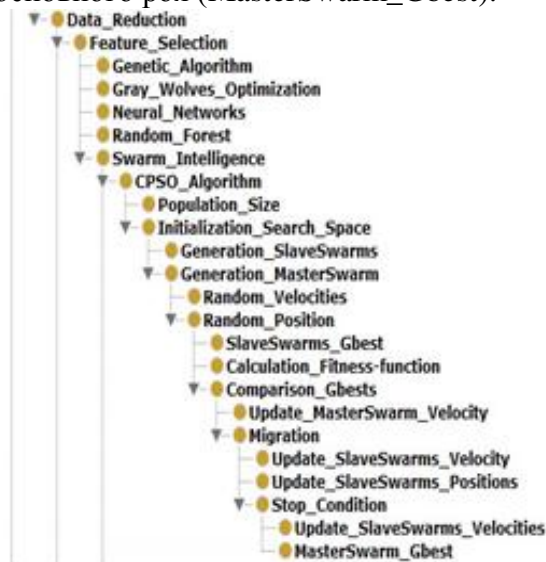


Рисунок 1 – OWL модель CPSO алгоритма

Предложенный CPSO алгоритм отличается возможностью детального исследования пространства поиска. В алгоритме у каждого подчинённого роя существует отдельная память, что позволяет каждому рою действовать самостоятельно. Агенты не стремятся к локальному лучшему значению, что решает проблему преждевременной сходимости. Достоинством CPSO алгоритма является параллельное действие агентов подчинённого роя, которое позволяет увеличить скорость вычисления алгоритма. Предложенный алгоритм отличается простотой программной реализации.

Ниже приведены результаты моделирования. В качестве базы данных для исследования эффективности предложенного метода используются дескрипторы сульфаниламидов на базе ресурсов Molinstincs и PubChem, которая состоит из более 1500 дескрипторов разных уровней. Рассмотренные химические соединения классифицированы на сульфаниламиды короткого действия, средней длительности действия и длительного действия. Используются такие дескрипторы (Табл. 1, приложение 1), как: количество атомов, количество атомов карбона, относительное количество атомов водорода, молекулярный

вес, количество связей, гравитационный индекс и др.

Алгоритм реализован на языке программирования Python. На рисунке 2 (приложение 2) представлен график выбранных информативных дескрипторов с использованием CPSO алгоритма на примере лекарственных соединений сульфаниламидной группы. В результате из 1500 дескрипторов были выбраны 60 информативных дескрипторов.

После отбора информативных дескрипторов согласно алгоритму CPSO-AIS осуществляется прогноз зависимости «структура-свойство/активность» новых лекарственных препаратов.

Работа выполнена по гранту КН МОН РК на тему: "Разработка и анализ баз данных для информационной системы прогнозирования зависимости "структура-свойство" лекарственных соединений на основе алгоритмов искусственного интеллекта" (2018-2020 гг.)

### **Список литературы**

- 1 Zhong F., Xing J., Li X., etc. Artificial intelligence in drug design // *Science China Life Sciences.* – 2018. – Vol. 61, Issue 10. – P. 1191-1204.
- 2 Dobchev A., Dimitar G., Pillai G., etc. In Silico Machine Learning Methods in Drug Development // *Current Topics in Medicinal Chemistry.* – 2014. – Vol. 14, No 16. – P. 1913-1922.
- 3 Samigulina G.A., Massimkanova Zh.A. Construction of optimal immune network model based on swarm intelligence algorithms for computer design of new drugs // *Proceedings of VII International Conference «Optimization problems and their applications, OPTA-2018».* – Omsk, 2018. – P.349-358.
- 4 Lü J., Shen Q., Jiang J., Shen G., Yu R. QSAR analysis of cyclooxygenase inhibitor using particle swarm optimization and multiple linear regression // *Journal of Pharmaceutical and Biomedical Analysis.* – 2004. – Vol. 35, Issue 4. – P. 679-687.
- 5 Jia Y., Chen W., Zhang J., Li J. Generating Software Test Data by Particle Swarm Optimization // *Materials of Asia-Pacific Conference on Simulated Evolution and Learning.* - 2014. - P. 37-47.
- 6 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.* – 2015. – P.147-156.
- 7 Ivanciuc O. Weka machine learning for predicting the phospholipidosis inducing potential // *Curr. Top. Med. Chem.* – 2008. – Vol. 8(18). P. 1691-1709.
- 8 Hayashida T., Nishizaki I., Sekizaki S., Koto S. Cooperative Particle Swarm Optimization in Distance-Based Clustered Groups // *Journal of Software Engineering and Applications.* – 2017. – Vol. 10. – P. 143-158.
- 9 Niu B., Zhu Y., He X. A Multi-population Cooperative Particle Swarm Optimizer for Neural Network Training // *Advances in Neural Networks. Lecture Notes in Computer Science.* – 2006. – Vol. 3971. – P. 570-576.
- 10 Samigulina G.A., Samigulina Z.I. Modified immune network algorithm based on the Random Forest approach for the complex objects control // *Artificial Intelligence Review.* – 2018. – P. 1-17.

*Принято в печать 16.03.2019*

**Приложение 1**

Таблица 1 – Фрагмент базы данныхс ульфаниламидами.

Class	Number of atoms	Relative number of C atoms	Relative number of O atoms	Relative number of aromatic bonds	Molecular weight	...	Gravitationindex (allbonds)
short_acting	27,00	0,37	0,07	0,43	250,28	...	1980,80
short_acting	33,00	0,36	0,06	0,29	278,33	...	1916,90
short_acting	31,00	0,35	0,10	0,19	267,30	...	1874,90
short_acting	27,00	0,33	0,07	0,21	270,32	...	1856,90
short_acting	46,00	0,33	0,04	0,17	311,44	...	2335,90
short_acting	33,00	0,36	0,06	0,29	278,33	...	1916,90
short_acting	46,00	0,33	0,04	0,17	311,44	...	2335,90
medium_acting	24,00	0,33	0,13	0,13	214,24	...	1408,30
medium_acting	35,00	0,34	0,11	0,31	310,33	...	2401,50
medium_acting	28,00	0,36	0,11	0,32	253,28	...	1949,40
medium_acting	31,00	0,35	0,10	0,31	267,30	...	2111,70
Long_acting	35,00	0,34	0,11	0,31	310,33	...	2401,50
Long_acting	36,00	0,42	0,06	0,41	314,36	...	2118,10
Long_acting	31,00	0,35	0,10	0,19	280,30	...	1960,30
Long_acting	30,00	0,37	0,07	0,29	264,30	...	1843,00

**Приложение 2**

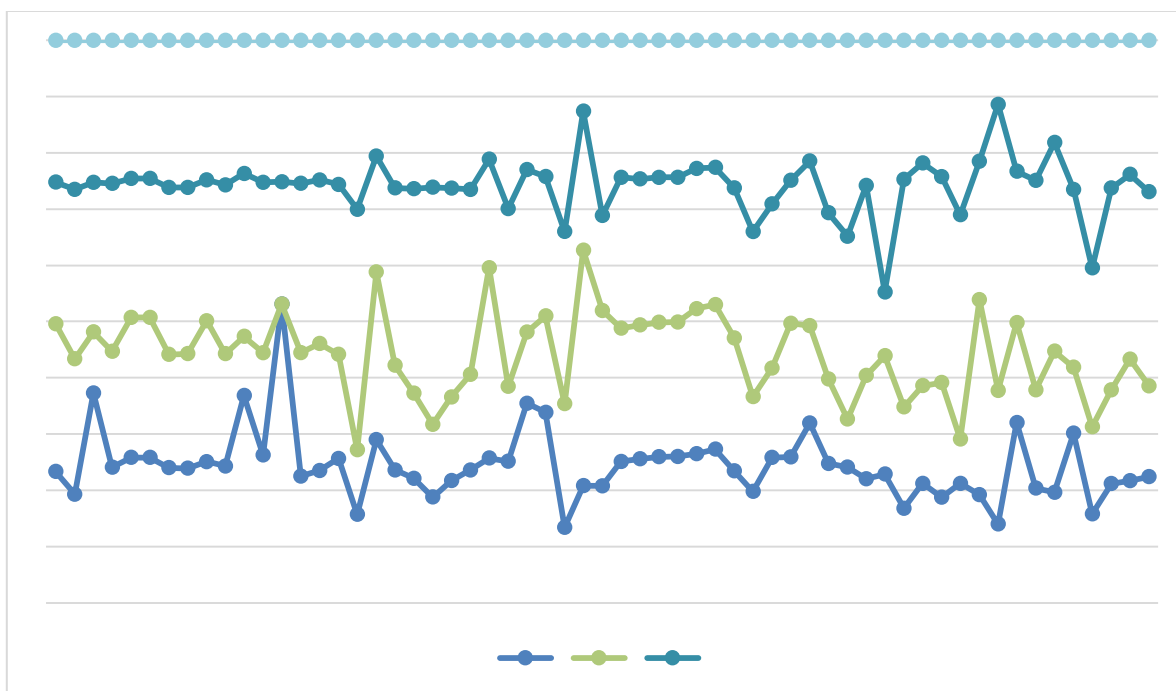


Рисунок 2 – Визуализация выбранных информативных дескрипторов на основе CPSO алгоритма

Г.А. Самигулина<sup>1</sup>, Ж.А. Масимканова<sup>2</sup>

<sup>1</sup> *Институт информационных и вычислительных технологий КН МОН РК,  
Алматы, Казахстан*

<sup>2</sup> *Казахский национальный университет им. аль-Фараби, Алматы, Казахстан*

## РАЗРАБОТКА МОДИФИЦИРОВАННОГО АЛГОРИТМА НА ОСНОВЕ КООПЕРАТИВНОГО АЛГОРИТМА РОЯ ЧАСТИЦ, ИСКУССТВЕННЫХ ИММУННЫХ СИСТЕМ И ОНТОЛОГИЧЕСКОГО ПОДХОДА

**Аннотация.** Статья посвящена разработке модифицированного алгоритма на основе кооперативного алгоритма роя частиц и искусственных иммунных систем для прогнозирования зависимости «структура-свойство/активность» (QSAR) лекарственных соединений. Построение оптимального набора дескрипторов выполняется на основе кооперативного алгоритма роя частиц. Распознавание образов и прогноз молекул-кандидатов в лекарственные препараты осуществляется с использованием искусственных иммунных систем (ИИС). Разработанный модифицированный алгоритм позволяет сократить время обучения ИИС и повысить точность прогнозирования за счет построенного оптимального набора данных. Для систематизации данных и облегчения создания программного обеспечения построена OWL модель кооперативного алгоритма роя частиц в редакторе онтологий Protégé. Приведены результаты моделирования с использованием базы данных сульфаниламидов.

**Ключевые слова:** Smart-технология прогнозирования, зависимость «структура-свойство/активность» (QSAR), кооперативный алгоритм роя частиц, искусственные иммунные системы, OWL модель, лекарственные препараты

Г.А. Самигулина<sup>1</sup>, Ж.А. Масимканова<sup>2</sup>

<sup>1</sup> *ҚР БҒМ ҒК Ақпараттық және есептеуіш технологиялар институты,  
Алматы, Қазақстан*

<sup>2</sup> *әл-Фараби атындағы Қазақ Ұлттық Университеті, Алматы, Қазақстан*

## ОНТОЛОГИЯЛЫҚ ӘДІС, ЖАСАНДЫ ИММУНДЫҚ ЖҮЙЕЛЕР ЖӘНЕ КООПЕРАТИВТІК ҮЙІР АЛГОРИТМІНІҢ НЕГІЗІНДЕ ЖАҢАРТЫЛҒАН АЛГОРИТМ ҚҰРУ

**Аннотация.** Мақала дәрілік қосылыстардың «құрылым-қасиет/белсенділік» тәуелділігін (QSAR) болжау үшін кооперативтік үйір алгоритмі мен жасанды иммундық жүйелер негізінде жаңартылған алгоритм құруға арналған. Дескрипторлардың оңтайлы жиынтығын құру кооперативтік үйір алгоритмінің негізінде жүзеге асырылады. Дәрі-дәрмектің кандидат-молекулаларын тану және болжау жасанды иммундық жүйелер (ЖИЖ) арқылы орындалады. Құрастырылған жаңартылған алгоритм оңтайлы деректер жиынтығының арқасында оқу уақытын қысқартуға және болжамдаудың дәлдігін арттыруға мүмкіндік береді. Деректерді жүйелеу және бағдарламалық жасақтаманы құруды жеңілдету үшін Protégé онтологиялық редакторында кооперативтік үйір алгоритмінің OWL моделі құрылды. Сульфаниламид деректер базасын пайдалана отырып моделдеу нәтижелері келтірілді.

**Түйін сөздер:** Болжамдаудың Smart-технологиясы, «құрылым-қасиет/белсенділік» тәуелділігі (QSAR), кооперативтік үйір алгоритмі, жасанды иммундық жүйелер, OWL модель, дәрі-дәрмектер

**G.A.Samigulina<sup>1</sup>, Zh.A. Massimkanova<sup>2</sup>**

<sup>1</sup>*Institute of Information and Computational Technologies CS MES RK,  
Almaty, Kazakhstan*

<sup>2</sup>*Al-Farabi Kazakh National University, Almaty, Kazakhstan*

**THE DEVELOPMENT OF MODIFIED ALGORITHM BASED ON  
COOPERATIVE PARTICLE SWARM OPTIMIZATION, ARTIFICIAL IMMUNE  
SYSTEMS AND ONTOLOGICAL APPROACH**

**Abstract.** The article is devoted to the development of modified algorithm based on cooperative particle swarm optimization and artificial immune systems for forecasting quantitative «structure-property/activity» relationship (QSAR) of medicinal compounds. The construction of an optimal set of descriptors has been performed on the basis of a cooperative particle swarm optimization algorithm. Pattern recognition and prognosis of candidate molecules of drugs has been carried out using artificial immune systems (AIS). The developed modified algorithm allows to reduce learning time of AIS and improve forecasting accuracy due to the constructed optimal data set. For systematization data and simplification of creating software the OWL model of cooperative particle swarm optimization has been built in Protégé ontology editor. The simulation results using the sulfonamide database are given.

**Keywords:** Smart-technology for forecasting, «structure-property/activity» relationship (QSAR), cooperative particle swarm optimization, artificial immune systems, OWL model, drugs