

Г.А. Самигулина¹ , З.И. Самигулина² ¹Институт информационных и вычислительных технологий,
г. Алматы, Казахстан²Казахстанско-Британский Технический Университет,
г. Алматы, Казахстан

РАЗРАБОТКА ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ ДЛЯ РЕАЛИЗАЦИИ ИНТЕЛЛЕКТУАЛЬНОЙ ТЕХНОЛОГИИ ПРОГНОЗИРОВАНИЯ ЗАВИСИМОСТИ «СТРУКТУРА-СВОЙСТВО» ЛЕКАРСТВЕННЫХ СОЕДИНЕНИЙ НА ОСНОВЕ МОДИФИЦИРОВАННОГО АЛГОРИТМА ИСКУССТВЕННЫХ ИМ- МУННЫХ СИСТЕМ

Аннотация. В настоящее время фармакологическая отрасль активно развивается за счёт внедрения последних достижений в области искусственного интеллекта (ИИ). Ведущие мировые производители лекарственных препаратов сотрудничают с известными фирмами, занимающимися разработкой новых инновационных алгоритмов искусственного интеллекта с целью снижения себестоимости процесса производства лекарств и сокращения временных ресурсов. В связи с чем разработка современной интеллектуальной технологии, позволяющей обрабатывать большой массив химической информации и прогнозировать фармакологические свойства лекарственных соединений является актуальной задачей.

Прогнозирование зависимости «структура-свойство» (Quantitative Structure-Activity Relationship, QSAR) лекарственных соединений осуществляется с помощью модифицированного алгоритма искусственных иммунных систем (Artificial Immune Systems, AIS) и алгоритма опыления цветов (Flower Pollination Algorithm, FPA). Алгоритм опыления цветов FPA применяется для решения задачи выделения информативных дескрипторов, описывающих структуру химических соединений. Представлена архитектура программного обеспечения для реализации предложенной интеллектуальной технологии с помощью агентно-ориентированного и онтологического подходов для мультиагентной платформы JADE. Построена онтологическая модель модифицированного алгоритма FPA-AIS в редакторе онтологий Protégé с учётом механизма оценки эффективности модифицированного алгоритма FPA-AIS на основе характеристик: ошибка классификации, точность, полнота и f-мера.

Ключевые слова: интеллектуальная технология, зависимость «структура-свойство» QSAR лекарственных соединений, искусственные иммунные системы, алгоритм опыления цветов, агентно-ориентированный подход, онтологическая модель.

Введение

Последние тенденции в области развития фармакологической отрасли привели к значительному росту сотрудничества известных производителей лекарств и организаций, занимающихся разработкой и внедрением ИИ. Например, швейцарская фармакологическая компания Roche заключила соглашение с фирмой Exscientia, специализирующейся в области ИИ для разработки доклинических препаратов. Также крупнейшая компания IBM успешно предоставляет суперкомпьютер Watson фармакологической компании Pfizer для поиска противораковых препаратов. Вне-

дрение методов ИИ нацелено на снижение временных и финансовых затрат при отборе кандидатов в лекарства, поскольку характерной особенностью исследований в данной области является большое количество химической структурной информации, эффективный анализ которой и последующая редукция данных позволяет достичь более высокой точности прогнозирования фармакологических свойств лекарственных соединений.

Хорошо зарекомендовали себя методы искусственного интеллекта для решения задачи выделения информативных признаков, такие как: алгоритм оптимиза-

ции серых волков (Grey wolf optimization, GWO) [1], генетический алгоритм (Genetic Algorithm, GA) [2], алгоритм роя частиц (Particle Swarm Optimization, PSO) [3], алгоритм опыления цветов (Flower Pollination Algorithm, FPA) [4] и т.д.

Особый интерес вызывает алгоритм опыления цветов, впервые предложенный учёным YangXin-She в 2012 году [5]. FPA является метаэвристическим алгоритмом, имитирующим процесс опыления цветов в природе. В настоящее время существует много исследований, посвящённых применению FPA в области биоинформатики и медицины. Например, в работе [6] представлен улучшенный алгоритм опыления цветов для идентификации белков. Работа [7] посвящена исследованиям сосудов сетчатки на основе алгоритма FPA с поиском по эталону. В исследованиях [8] алгоритм опыления цветов используется для идентификации личности по энцефалограмме. Работа [9] посвящена применению алгоритма FPA для настройки нейронной сети при классификации болезней сердца. В статье [10] представлен бинарный алгоритм опыления цветов и его приложения для решения задачи выделения информативных признаков. Моделирование осуществлялось на основе эталонных данных в области биоинформатики, таких как Tox-171, AR10P и т.д.

Таким образом, FPA является перспективным современным метаэвристическим алгоритмом, который в настоящее время успешно применяется для решения задачи выделения информативных признаков и может использоваться при разработке модифицированных алгоритмов прогнозирования на основе AIS.

Постановка задачи исследований

Постановка задачи исследований формулируется следующим образом: необходимо разработать программное обеспечение на мультиагентной платформе JADE для реализации интеллектуальной технологии прогнозирования зависимости «структура-свойство» лекарственных со-

единений на основе модифицированного алгоритма искусственных иммунных систем FPA-AIS, агентно-ориентированного и отнологического подходов.

Модифицированный алгоритм FPA-AIS

Так как в настоящее время не существует универсальных алгоритмов для решения задачи прогнозирования актуальна разработка модифицированных алгоритмов искусственного интеллекта, которые можно применять к базам данных дескрипторов, рассматриваемых лекарственных соединений различных по размерности и характеру данных.

Рассмотрим процесс функционирования алгоритма опыления цветов для решения задачи выделения информативных дескрипторов, описывающих структуру химических соединений. В основе алгоритма FPA лежит природный процесс опыления цветов для дальнейшего воспроизводства вида. Опыление может быть осуществлено переносом пыльцы за счёт опылителей (пчёл, насекомых, птиц и т.д.), а также с помощью самоопыления. Данные формы опыления называются биотическая и абиотическая. На рисунке 1 представлен процесс опыления цветов в природе.

При разработке алгоритма опыления цветов выделены следующие положения [5, 11]:

- процесс глобального опыления осуществляется на основе биотической формы, когда пыльца от одного цветка переносится на рыльце пестика другого цветка за счёт опылителя, совершающего полёт Леви [15]; локальное опыление рассматривается как форма самоопыления; постоянство цветка – это процесс, при котором отдельные опылители посещают определённые виды цветов; локальное и глобальное опыление происходит с вероятностью $p \in [0,1]$.

Псевдокод алгоритма опыления цветов представлен следующим образом [4,5]:

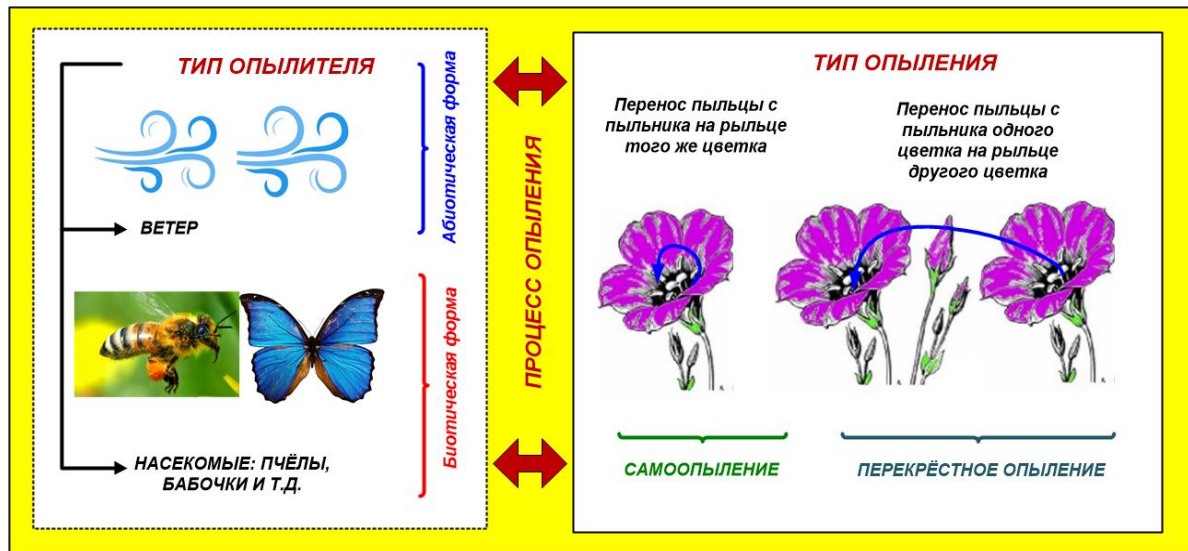


Рисунок 1 – Процесс опыления цветов

Инициализация параметров с вероятностью $p \in [0,1]$;

Случайная генерация начальной популяции цветов;

Оценка начальной популяции и поиск текущего лучшего решения $gbest$;

While (пока не будет достигнут критерий останова) do

For each flower

If $\text{rand}() < p$

Глобальноеопыление:

$$(x_i^{t+1} = x_i^t + L(x_i^t - gbest));$$

// на основе шагов Леви

else

Выбор двух случайных событий x_j^t и x_k^t ;

Локальное опыление:

$$(x_i^{t+1} = x_i^t + \epsilon (x_j^t - x_k^t));$$

endif

обновление новых позиций;

обновление решений с учётом лучшего решения;

endifor

Сохранить текущее лучшее решение;

endwhile

Рассмотрим модифицированный алгоритм *FPO – AIS* на основе алгоритма опыления цветов и искусственных иммунных систем.

Алгоритм: *FPO – AIS*.

Шаг 1. Формирование базы данных дескрипторов лекарственных соединений.

Шаг 2. Классификация БД дескрипторов по фармакологическим свойствам. Создание базы знаний на основе заключений экспертов онтологической модели.

Шаг 3. Построение оптимального набора дескрипторов на основе FPA.

Шаг 4. Решение задачи прогнозирования QSARна основе алгоритма искусственных иммунных систем [12].

*Шаг 5.*Отбор кандидатов химических соединений с заданными фармакологическими свойствами.

*Шаг 6.*Оценка эффективности алгоритма FPA-AIS.

Достоинствами предложенного алгоритма являются: выделение информативных признаков, повышение точности моделей прогнозирования зависимости «структура-свойство» лекарственных соединений, сокращение временных и финансовых ресурсов при создании новых лекарственных препаратов.

Разработка программного обеспечения на мультиагентной платформе JADE для интеллектуальной технологии прогнозирования QSARна базаагентно-ориентированного и онтологического подходов

Интеллектуальная технология прогнозирования зависимости «структура-свойство» лекарственных соединений состоит из различных этапов, реализация которых удобна на основе агентно-

ориентированного подхода [13]. Современные мультиагентные системы состоят из совокупности автономных агентов, которые функционируют в программной среде, могут взаимодействовать друг с другом и осуществлять параллельную обработку данных. Так как анализ баз данных дескрипторов химической информа-

ции требует значительных временных и вычислительных ресурсов актуально использование агентной платформы JADE с открытым исходным кодом. Общая архитектура программного обеспечения, реализующего интеллектуальную технологию на основе платформы JADE представлена на Рисунке 2.

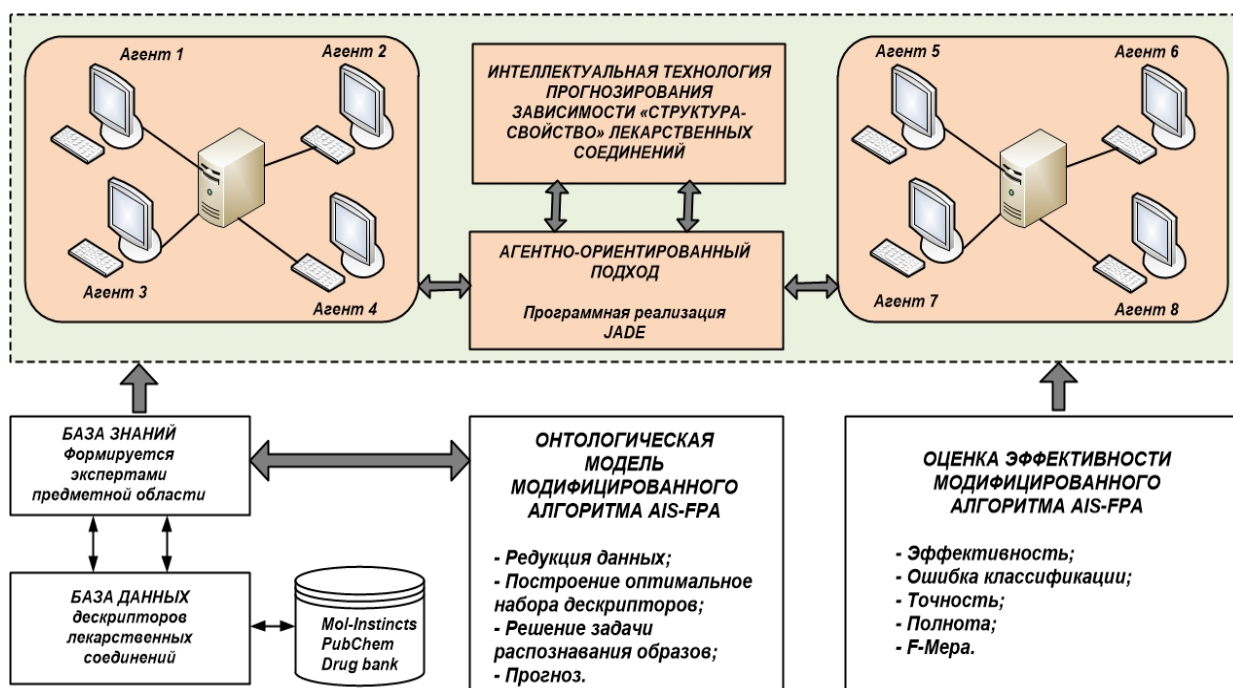


Рисунок 2 – Архитектура программного обеспечения для прогнозирования зависимости «структура-свойство» лекарственных соединений на основе агентно-ориентированного и онтологического подходов

Програмная среда состоит из взаимодействующих агентов, которые реализуют поставленные цели, способны принимать решения и связываться с другими агентами.

Система подключается к базе знаний и к онтологической модели, содержащей сведения экспертов и онтологической модели системы, а также базе данных дескрипторов рассматриваемого химического соединения. База данных дескрипторов формируется на основе мировых банков данных химической

информации, таких как Mol-Instincts, PubChem, Drugbank и т.д.

Разработанная модель прогнозирования на основе модифицированного алгоритма FPA-AIS описывается с помощью онтологической модели, реализованной в редакторе онтологий Protégé. На Рисунках 3 и 4 представлены токсонимия классов модифицированного алгоритма FPA-AIS и визуализация структурной схемы онтологической модели.



Рисунок 3–Токсономия классов онтологической модели модифицированного алгоритма FPA-AIS

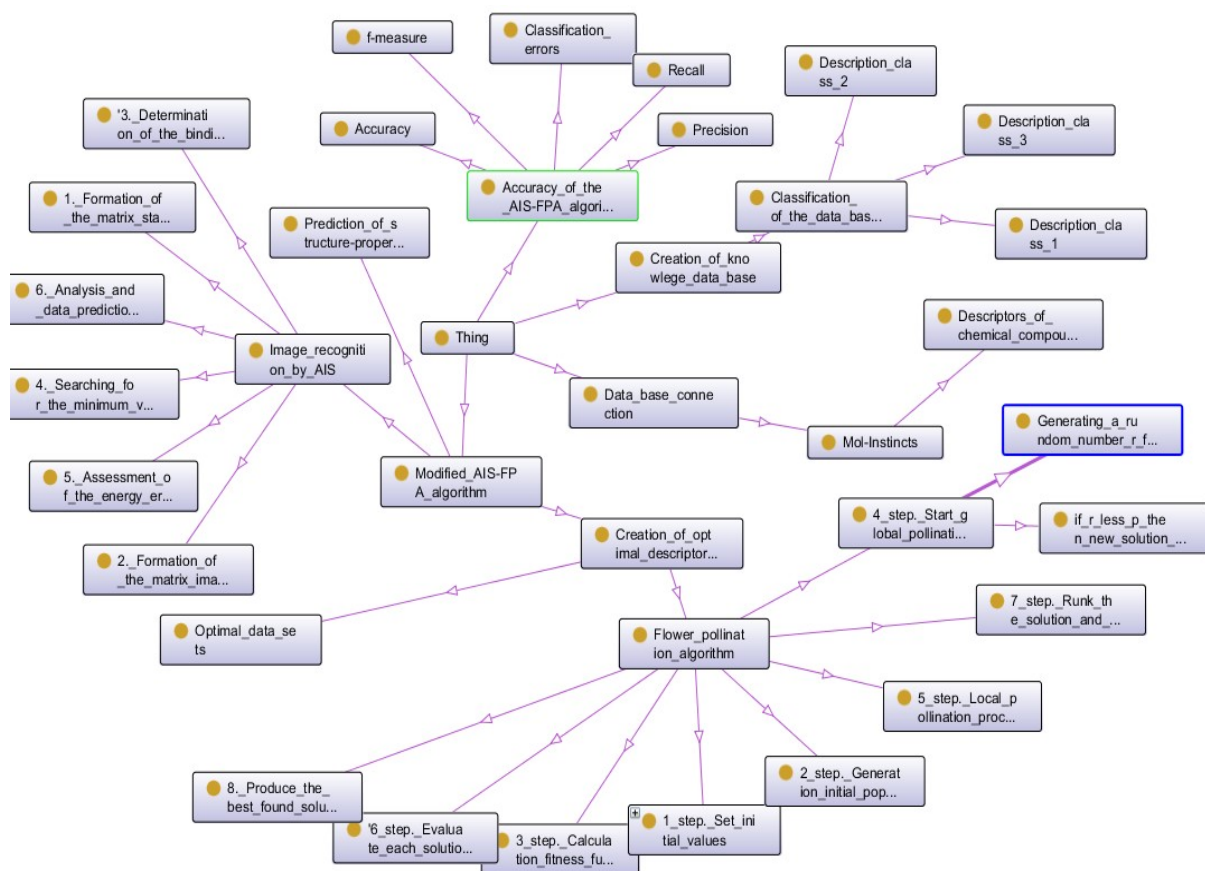


Рисунок 4 – Структурная схема онтологической модели модифицированного алгоритма FPA-AIS

Эффективность разработанной модели прогнозирования оценивается с помощью характеристик: accuracy, classification error, precision, recall, f-measure и др.

Таким образом, агентно-ориентированный и онтологический подходы являются удобными современными инструментами для реализации разработанной интеллектуальной технологии прогнозирования зависимости «структура-свойство» лекарственных соединений на основе модифицированных алгоритмов искусственного интеллекта. Данный подход позволяет обрабатывать большие массивы химической информации, формировать оптимальные наборы дескрипторов, повышать качество моделей прогнозирования и сокращать временные ресурсы при создании кандидатов в лекарственные соединения.

Работа выполнена при финансовой поддержке КН МОН РК в рамках научного проекта №AP05130019 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 Mirjalili S., Mirjalili S.M., Lewis A. Grey wolf optimizer // *Advances in Engineering software*. – 2014. – Vol. 69. – P. 46-61.
- 2 Saidi R., Neir W., Essoussi N. Feature Selection Using Genetic Algorithm for Big Data // *AMLT A*. – Springer, 2018. – P. 352-361.
- 3 Qiu C.A novel multi-swarm particle swarm optimization for feature selection // *Genetic Programming and Evolvable*. – Springer, 2019. – P. 1-27.
- 4 Abdel-Basset M., Shawky L.A. Flower pollination algorithm: a comprehensive review // *Artificial Intelligence Review*. – Springer, 2019. – Vol. 52., №4. – P. 1-25.
- 5 Yang X.-S. Flower Pollination Algorithms // *Nature-Inspired Optimization Algorithms*. – 2014. – P. 155-173.
- 6 Lei X., Fang M., Chen L. Improved flower pollination algorithm for identifying essential proteins // *Biology, Medicine, Computer Science. System Biology*. – 2018. – Vol. 12. - №4. – P. 130-166.

- 7 Emary E., Zawbaa H.M., Hassaien A.E., Parv B. Multi-objective retinal vessel localization using flower pollination search algorithm with pattern search // *Advanced Data Anal Classification*. – 2017. – Vol. 11. - №3. – P. 611-627.

- 8 Rodrigues D., Silva G.F., Papa J.P., Marana A.M., Yang X.S. EEG – based person identification through binary flower pollination algorithm // *Expert system application*. – 2016. – Vol. 62. – P. 81-90.

- 9 Yazid M.H., Talib M.S., Satria M.H. Flower pollination neural network for heart Disease classification // *Materials Science and Engineering*. – 2019. – Vol. 551. – P. 1-5.

- 10 Rodrigues D., Yang X.S., De Souza A.N., Papa J.P. Binary flower pollination algorithm and its application to feature selection // *Recent advances in swarm intelligence and evolutionary computation*. – Springer, 2015. – P. 85-100.

- 11 Буркеев Е.М. Прогнозирование потребления нефти с использованием нейронной сети. Магистерская диссертация. - Санкт-петербургский государственный университет. – Санкт-Петербург, 2017. – 50 с.

- 12 Samigulina G.A., Samigulina Z.I. Modified immune network algorithm based on the Random Forest approach for the complex objects control // *Artificial intelligence review*. – Springer, 2019. – Vol. 52. - №4. – P. 2457-2473.

- 13 Samigulina G.A., Samigulina Z.I. Development of multi-agent technology for prediction of the «structure-property» dependence of drugs on the basis of modified algorithms of artificial immune systems // *Proceedings of International Work Conference on Bioinformatics and Biomedical Engineering, IWBBIO 2018, April 25-27*. – Spain, Granada. – 2018. – P. 1-2.

- 14 Samigulina G.A., Samigulina Z.I. Immune network technology on the basis of Random Forest algorithm for computer aided drug design // *Lecture Notes in Computer Science. Proceedings of the Conference on Bioinformatics and Biomedical Engineering (IWBBIO 2017)*. – Granada (Spain): Springer, 26-28 April 2017. Part 1. – P. 50-61.

- 15 Pavlyukevich I. Levy flights, non-local search and simulated annealing // *Journal Computational Physics*. – 2007, №226. – P. 1830-1844.

Принято в печать 21.12.2019

Г.А. Самигулина¹, З.И. Самигулина²

¹Институт информационных и вычислительных технологий,
г. Алматы, Казахстан

²Казахстанско-Британский Технический Университет,
г. Алматы, Казахстан

РАЗРАБОТКА ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ ДЛЯ РЕАЛИЗАЦИИ ИНТЕЛЛЕКТУАЛЬНОЙ ТЕХНОЛОГИИ ПРОГНОЗИРОВАНИЯ ЗАВИСИМОСТИ «СТРУКТУРА-СВОЙСТВО» ЛЕКАРСТВЕННЫХ СОЕДИНЕНИЙ НА ОСНОВЕ МОДИФИЦИРОВАННОГО АЛГОРИТМА ИСКУССТВЕННЫХ ИМ- МУННЫХ СИСТЕМ

Аннотация. В настоящее время фармакологическая отрасль активно развивается за счёт внедрения последних достижений в области искусственного интеллекта (ИИ). Ведущие мировые производители лекарственных препаратов сотрудничают с известными фирмами, занимающимися разработкой новых инновационных алгоритмов искусственного интеллекта с целью снижения себестоимости процесса производства лекарств и сокращения временных ресурсов. В связи с чем разработка современной интеллектуальной технологии, позволяющей обрабатывать большой массив химической информации и прогнозировать фармакологические свойства лекарственных соединений является актуальной задачей.

Прогнозирование зависимости «структура-свойство» (Quantitative Structure-Activity Relationship, QSAR) лекарственных соединений осуществляется с помощью модифицированного алгоритма искусственных иммунных систем (Artificial Immune Systems, AIS) и алгоритма опыления цветов (Flower Pollination Algorithm, FPA). Алгоритм опыления цветов FPA применяется для решения задачи выделения информативных дескрипторов, описывающих структуру химических соединений. Представлена архитектура программного обеспечения для реализации предложенной интеллектуальной технологии с помощью агентно-ориентированного и онтологического подходов для мультиагентной платформы JADE. Построена онтологическая модель модифицированного алгоритма FPA-AIS в редакторе онтологий Protégé с учётом механизма оценки эффективности модифицированного алгоритма FPA-AIS на основе характеристик: ошибка классификации, точность, полнота и f-мера.

Ключевые слова: интеллектуальная технология, зависимость «структура-свойство» QSAR лекарственных соединений, искусственные иммунные системы, алгоритм опыления цветов, агентно-ориентированный подход, онтологическая модель.

Г.А. Самигулина¹, З.И. Самигулина²

¹Ақпараттық және есептеу іштехнологиялар институты
Алматы қ., Қазақстан

²Қазақстан-Британ техникалық университеті
Алматы қ., Қазақстан

ЖАСАНДЫ ИММУНДЫҚ ЖҮЙЕЛЕРДІҢ МОДИФИКАЦИЯЛАНҒАН АЛГОРИТМІНІҢ НЕГІЗІНДЕГІ ДӘРІЛІК ҚОСЫЛЫСТАРДЫҢ «ҚҰРЫЛЫМ-ҚАСИЕТ» ТӘУЕЛДІЛІГІН БОЛЖАУДЫҢ ИНТЕЛЛЕКТУАЛДЫ ТЕХНОЛОГИЯСЫН ЖҮЗЕГЕ АСЫРУҒА АРНАЛҒАН БАҒДАРЛАМАЛЫҚ ЖАБДЫҒЫН ЖАСАУ

Аннотация. Қазіргі таңда фармакология саласы жасанды интеллекттің (ЖИ) соңғы жетістіктерін енгізе отырып белсенді даму үстінде. Дәрі өндіру үдерісінің құндылығын төмендету мен уақыт қорын қысқарту мақсатында дәрілік препараттарды өндіруші алдыңғы қатарлы әлемдік өндірушілер жасанды интеллекттің жаңа инновацияланған алгоритмдерін жасайтын танымал фирмалармен бірге қызмет атқарып келеді. Осыған байланысты ауқымды

химиялық ақпараттарды өңдеу және дәрілік қосылыстардың фармакологиялық қасиеттерін болжауға мүмкіндік беретін замануи жасанды интеллектуалды технологияларын жасау өзекті мәселе болып табылады.

Дәрілік қосылыстардың «құрылым-қасиет» (Quantitative Structure-Activity Relationship, QSAR) және тәуелділігін болжау гүлдердің тозаңдануы алгоритмінің (Flower Pollination Algorithm, FPA) жасанды иммунды жүйелердің модификацияланған алгоритмінің көмегімен жүзеге асырылады. Гүлдердің тозаңдануы алгоритмі FPA химиялық қосылыстардың құрылысын сипаттайтын ақпаратты дескрипторлардың мәселерін шешуде қолданылады. JADE мультиагентті платформасына арналған онтологиялық және агентті-бағдарланған тәсілдерінің көмегімен ұсынылған интеллектуалды технологияларды жүзеге асыру үшін бағдарламалық жабдықтаудың құрылысы ұсынылды. Толықтық, f-өлшем, дәлдік және классификацияның қатесі сияқты сипаттамалардың негізінде FPA-AIS модификацияланған алгоритмінің бағалау механизмінің тиімділігін ескере отырып Protégé онтологиялық редакторінде FPA-AIS модификацияланған алгоритмінің онтологиялық моделі тұрғызылды.

Түйін сөздер: интеллектуалды технология, QSAR дәрілік қосылыстардың «құрылым-қасиет» тәуелділігі, жасанды иммунды жүйелер, гүлдердің тозаңдану алгоритмі, агентті-бағдарланған тәсіл, онтологиялық модель.

G.A. Samigulina¹, Z.I. Samigulina²

*¹Institute of information and computing technologies,
c. Almaty, Kazakhstan*

*²Kazakh-British Technical University,
c. Almaty, Kazakhstan*

DEVELOPMENT OF THE SOFTWARE FOR IMPLEMENTATION OF INTELLECTUAL TECHNOLOGY FOR FORECASTING DEPENDENCE «STRUCTURE-PROPERTIES» OF MEDICINAL COMPOUNDS BASED ON MODIFIED ARTIFICIAL IMMUNE SYSTEMS ALGORITHM

Abstract. Nowadays the pharmacological industry is actively developing due to the introduction of the latest achievements in the field of artificial intelligence (AI). The world's leading drug manufacturers are collaborating with well-known companies involved in the development of new innovative artificial intelligence algorithms in order to reduce the cost of the drug production process and time resources. In this connection, the development of modern intellectual technology that allows processing a large array of chemical information and predicting the pharmacological properties of medicinal compounds is an actual task. The prediction of the Quantitative Structure-Activity Relationship (QSAR) of drug compounds is carried out using a modified Artificial Immune Systems (AIS) and the Flower Pollination Algorithm (FPA). The FPA algorithm is used to solve the problem of extracting informative descriptors describing the structure of chemical compounds. The software architecture for the implementation of the proposed intellectual technology using agent-based and ontological approaches for the multi-agent platform JADE is presented. The ontological model of the modified FPA-AIS algorithm is constructed in the Protégé ontology editor, taking into account the mechanism for evaluating the effectiveness of the modified FPA-AIS algorithm based on the characteristics: classification error, accuracy, precision, recall and f-measure.

Keywords: intellectual technology, QSAR structure-property relationship of drug compounds, artificial immune systems, flower pollination algorithm, agent-based approach, ontological model.