

MULTI-AGENT SCIENTIFIC RESEARCH SYSTEM FOR PREDICTING DEPENDENCE „STRUCTURE–PROPERTIES“ OF DRUG COMPOUNDS BASED ON MODIFIED ALGORITHMS OF ARTIFICIAL IMMUNE SYSTEMS

G. A. Samigulina, Z. I. Samigulina*

Institute of Information and Computational Technologies,
050010, Almaty, Kazakhstan

*Kazakh-British Technical University,
050010, Almaty, Kazakhstan

The article deals with the issues of creating multi-agent Smart-system for conducting scientific research for computer molecular design of new drugs with specified properties and prediction of the „structure-property“ relationship (QSAR) based on modified algorithms of artificial immune systems and other bio-inspected approaches of artificial intelligence. The main advantages and disadvantages of using various intelligent algorithms when building a Smart — system are given.

One of the problems with computer-aided molecular design of drugs is the „paradox of similarity“, when the compounds differ structurally quite insignificantly, but have completely different properties. For example, an optically active substance and its mirror isomer can vary significantly in biological activity. Therefore, it is particularly relevant to develop new non-traditional approaches of artificial intelligence and algorithms that provide the ability to recognize chemical compounds with almost the same structure, but with completely different properties.

The structure of a multi-agent Smart-system for conducting scientific research has been developed based on modified algorithms of artificial immune systems and the functioning of agents has been described.

The work was carried out according to the grant of the Committee Science of the Ministry of Education and Science of the Republic of Kazakhstan (2018–2020), on the topic „Development and analysis of databases for the information system of predicting dependence „structure–property“ of drug compounds based on artificial intelligence algorithms“.

Key words: multiagent Smart-system, drugs, molecular design, prediction of the „structure–property“ relationship (QSAR), artificial immune systems, modified artificial intelligence algorithms.

References

1. ROY K., KAR S., DAS R.N. Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment. 2015. NY: Academic Press, [Electron. Res.]: <https://www.elsevier.com/books/understanding-the-basics-of-qsar-for-applications-in-pharmaceutical-sciences-and-risk-assessment/roy/978-0-12-801505-6> (access date 06.02.2019).
2. ROY K. Advances in QSAR Modeling. Applications in Pharmaceutical, Chemical, Food, Agricultural and Environmental Sciences. 2017. Springer, [Electron. Res.]: <http://www.springer.com/in/book/9783319568492> (access date 06.02.2019).
3. GAL'BERSHTAM N. M., BASKIN I. I., PALYUTIN V. A., ZEMFIROV N. S. Nejronny'e seti kak metod poiska zavisimostej struktura–svojstvo organicheskikh soedinenij // Uspekhi khimii. 2003. N 72 (7). S. 706–727.

4. IVANCIUC O. Structure-activity relationships in aquatic toxicology with artificial immune systems: Mechanism of toxic action classification of polar and nonpolar narcotic pollutants with CLONALG (clonal selection algorithm) // *J. Molecular Design*. 2007. N 6. P. 106–114.
5. CHUNLIN LIANG, LINGXI PENG. An Automated Diagnosis System of Liver Disease using Artificial Immune and Genetic Algorithms // *Journal of Medical Systems*, 2013. DOI:10.1007/s10916-013-9932-9.
6. ÜMIT YILMAZ, HELMI MD RAIS, SAID JADID ABDULKADIR. Hybrid Swarm Intelligence Algorithms with Ensemble Machine Learning for Medical Diagnosis // *Proceedings of 4th International Conference on Computer and Information Sciences (ICCOINS)*, 2018.
7. ZHU H., WU J., GU J. Studies on Immune Clonal Selection Algorithm and Application of Bioinformatics // *International Journal of Intelligent Engineering and Systems*. 2015. V. 8. N 1. P. 10–16.
8. SOSNIN S. B., RADCHENKO E. V., Palyulin V. A., ZEFIROV N. S. Obobshchennyj fragmentnyj podkhod v issledovaniyakh QSAR/QSPR // *Doklady Akademii nauk. Khimiya*. 2015, T. 463. N 3. S. 297–300.
9. GHASEMI, PÉREZ-SÁNCHEZ; MEHRI, PÉREZ-GARRIDO. Neural network and deep-learning algorithms used in QSAR studies: merits and drawbacks // *Drug Discovery Today*. 2018. N 23 (10). P. 1784–1790. DOI:10.1016/j.drudis.2018.06.016. PMID29936244.
10. YAVUZ B. Z., SERTKAYA C., YURTAY N. Prediction of secondary structures of hemoglobin using clonal selection algorithm // *Proceedings of the 7 th Int. Work. Comput. Sci. Eng.* 2017. P. 1387–1391.
11. SAMIGULINA G. A., SAMIGULINA Z. I. Immune network technology on the basis of Random Forest algorithm for computer aided drug design // *J. 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.
12. LUCIJA BREZOČNIK, IZTOK FISTER JR., VILI PODGORELEC. Swarm Intelligence Algorithms for Feature Selection: A Review // *J. Applied Sciences*. 2018. N 8 (9). P. 1521. [Electron. Res.]: <https://doi.org/10.3390/app8091521>.
13. RAMÍREZ M. R., RAMÍREZ MORENO H. B., ROJAS E. M., HURTADO C., VÁZQUEZ NÚÑEZ S. O. Multi-Agent System Model for Diagnosis of Personality Types // *Proceedings of the 12th International Conference Agents and Multi-agent Systems: Technologies and Applications. Smart Innovation, Systems and Technologies book series (SIST)*. Springer, 2018. Vol. 96. P. 209–214. DOI: 10.1007/978-3-319-92031-3_19.
14. IVANOVIC M., NINKOVIC S. Personalized HealthCare and Agent Technologies // *Proceedings of the 11th International Conference Agents and Multi-agent Systems: Technologies and Applications. Smart Innovation, Systems and Technologies book series (SIST)*. Springer: KES-AMSTA, 2017. V. 74. P. 3–11.
15. GERMAN S., SHIN S., TSOURDOS A. // *Proceeding of the IEEE 24th Mediterranean Conference on Control and Automation (MED)*. 2016. P. 1020–1025.
16. SAMIGULINA G. A., MASSIMKANOVA ZH. A. Multiagent system of recognize on the basis of modified algorithms of swarm intelligence and immune network modeling // *Proceedings of the 12th International Conference Agents and Multi-agent Systems: Technologies and Applications (AMSTA-18)*. Australia: Springer, 20–22 June 2018. P. 199–208.
17. 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.
18. 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 (IWBBIO2018). Spain, Granada, 25–27 April 2018. P. 1–2.

19. Samigulina G. A., Samigulina Z. I. Razrabotka tekhnologii immunosetevogo modelirovaniya dlya komp'yuternogo molekulyarnogo dizajna lekarstvenny'kh preparatov (programma dlya E'VM). Svidetel'stvo o gosudarstvennoj registraczii prav na ob'ekt avtorskogo prava v Komitete po pravam intellektual'noj sobstvennosti MYu RK. Astana, 28 marta 2011. N 473. 23 s.

МУЛЬТИАГЕНТНАЯ СИСТЕМА ВЕДЕНИЯ НАУЧНЫХ ИССЛЕДОВАНИЙ ДЛЯ ПРОГНОЗИРОВАНИЯ ЗАВИСИМОСТИ „СТРУКТУРА–СВОЙСТВО“ ЛЕКАРСТВЕННЫХ СОЕДИНЕНИЙ НА ОСНОВЕ МОДИФИЦИРОВАННЫХ АЛГОРИТМОВ ИСКУССТВЕННЫХ ИММУННЫХ СИСТЕМ

Г. А. Самигулина, З. И. Самигулина*

Институт информационных и вычислительных технологий,
050010, г. Алма-Ата, Республика Казахстан,

*Казахстанско-Британский Технический Университет,
050010, г. Алма-Ата, Республика Казахстан

УДК 004.89

DOI: 10.24411/2073-0667-2019-00010

В статье рассматриваются вопросы создания мультиагентной Smart-системы ведения научных исследований для компьютерного молекулярного дизайна новых лекарственных препаратов с заданными свойствами и прогнозирования зависимости „структура–свойство/активность“ (QSAR, Quantitative Structure-Activity Relationship) на основе модифицированных алгоритмов искусственных иммунных систем и других биоинспирированных подходов искусственного интеллекта. Приведены основные достоинства и недостатки применения различных интеллектуальных алгоритмов при построении Smart–системы. Разработана структура мультиагентной Smart-системы ведения научных исследований и описано функционирование агентов.

Ключевые слова: мультиагентная Smart-система, лекарственные препараты, молекулярный дизайн, прогнозирование зависимости „структура–свойство“ (QSAR), искусственные иммунные системы, модифицированные алгоритмы искусственного интеллекта.

Введение. Компьютерный молекулярный дизайн новых лекарственных препаратов с заданными свойствами является одной из основных и актуальных задач современной фармакологии [1]. Стремительный рост в последнее время неконтролируемого приема лекарств, в частности антибиотиков, приводит к резистентности (невосприимчивости) организма человека к данным химическим соединениям, и вследствие этого многие лекарственные препараты вызывают побочные отрицательные эффекты, такие как аллергические реакции, снижение эффективности лечения, несовместимость с другими лекарствами и т. д. Возникает острая необходимость в создании новых высокоэффективных лекарственных соединений различного назначения. Однако традиционная разработка лекарств — достаточно сложная и дорогостоящая процедура, которая занимает очень длительный период времени, включая клинические испытания и т. д. Поэтому разработка инновационных технологий прогнозирования зависимости „структура–свойство/активность“ (QSAR,

Работа выполнена по гранту Комитета Науки Министерства Образования и Науки Республики Казахстан АР05130019(2018–2020 гг.) по теме: „Разработка и анализ баз данных для информационной системы прогнозирования зависимости „структура–свойство“ лекарственных соединений на основе алгоритмов искусственного интеллекта“.

Quantitative Structure-Activity Relationship) лекарственных соединений с использованием последних достижений искусственного интеллекта, применения различных модифицированных биосперированных алгоритмов и современных компьютерных технологий является актуальной и перспективной задачей, нацеленной на сокращение финансовых затрат и времени при процедуре отбора кандидатов новых химических соединений с заданными фармакологическими свойствами для дальнейших исследований [2]. Создание инновационной информационной системы ведения научных исследований для молекулярного дизайна лекарственных препаратов актуально для развития фармакологической отрасли и синтеза новых дешевых лекарственных препаратов отечественного производства.

Применением современных подходов искусственного интеллекта для прогнозирования QSAR на основе искусственных нейронных сетей активно занимаются уже более двадцати лет. Например, работа [3] посвящена нейронным сетям как методу поиска зависимостей структура–свойство органических соединений. Также особый интерес представляют искусственные иммунные системы для моделирования QSAR [4].

В настоящее время активно разрабатываются информационные системы для обработки многомерных данных, прогнозирования и диагностики с использованием модифицированных алгоритмов искусственных иммунных систем (AIS, artificial immune systems) и других интеллектуальных алгоритмов. Например, в исследованиях [5] рассматриваются вопросы разработки системы диагностики заболеваний печени на основе комбинации двух методов: подхода искусственного иммунитета и генетического алгоритма. Архитектура информационной системы основана на искусственной иммунной системе. В процедуре обучения системы применяется генетический алгоритм для вмешательства в эволюцию популяции антител. Используется два набора эталонных данных из известного репозитория машинного обучения UCI (UCI Machine Learning Repository). Результаты моделирования показывают, что эта система может быть полезным инструментом для автоматической диагностики заболеваний печени. В работе [6] рассматривается гибридный роевой алгоритм для медицинской диагностики. Основной проблемой в данной области является минимизация ошибки при постановке диагноза, так как существующие методы не всегда показывают хороший результат на всех наборах данных болезней. Поэтому предложена гибридная динамическая система на основе алгоритма колонии муравьев с тремя уровнями обновления, с вейвлет-преобразованием и методом опорных векторов. Предложенный метод оценен с использованием пяти базовых наборов медицинских данных о различных заболеваниях из хранилища UCI и показал достаточно хороший результат. В статье [7] предлагается новый алгоритм искусственных иммунных систем на основе квантового клонального отбора (QCSA, Quantum Clonal Selection Algorithm) для прогнозирования структуры белка. Исследования [8] посвящены дизайну и поиску новых перспективных структур потенциальных лекарственных веществ на основе развития обобщенного фрагментного подхода к анализу количественной связи структуры со свойствами и биоактивностью органических соединений, а также построению предсказательных классификационных моделей QSAR с использованием методов опорных векторов и случайного леса. Полученные модели для прогнозирования мутагенности и температуры вспышки органических соединений применяются при решении задач молекулярного дизайна соединений с заданными свойствами.

В работе [9] рассматриваются достоинства и недостатки применения различных алгоритмов нейронных сетей (NN, neural networks) с использованием инновационных методов глубокого обучения (DL, deep-learning) для исследований QSAR при открытии новых лекарств. Исследования [10] посвящены прогнозированию вторичной структуры белка с

помощью совместного применения алгоритма искусственных иммунных систем на основе клонального отбора и многослойного персептрона. На первом этапе данные обрабатываются с помощью алгоритма клональной селекции (CSA). На втором этапе осуществляется классификация с использованием многослойного персептрона, который является одним из методов глубокого обучения. Моделирование показало, что подготовка данных с помощью CSA до классификации существенно улучшает результаты прогнозирования. В статье [11] предлагается иммунносетевая технология на основе модифицированного алгоритма случайного леса для компьютерного дизайна лекарственных соединений.

Одной из основных проблем при анализе многомерных данных и прогнозировании является выделение информативных признаков (дескрипторов). В настоящее время стремительный рост информации [12], совместное использование и обмен информацией ставят исследователей перед сложной задачей анализа данных и извлечения соответствующей нужной информации из данных. Алгоритмы роевого интеллекта (SI, Swarm Intelligence) набирают популярность при решении задач оптимизации и успешно используются для выделения информативного набора признаков в различных приложениях. Исследования показали, что 60,53 % таких задач решаются в области биомедицины, 28,95 % в области компьютерной инженерии и только 10,53 % в других приложениях.

Активно развиваются в последнее время мультиагентные технологии при реализации интеллектуальных алгоритмов и построении информационных систем различного назначения. Мультиагентные системы (MAS, multiagent systems) обладают рядом достоинств и являются перспективным направлением в медицине и фармакологии. Исследования [13] посвящены построению модели мультиагентной системы для диагностики типов личности. Используется технология интеллектуальных агентов, основанная на анализе базы знаний. В статье [14] рассматриваются вопросы развития персонализированной медицины с использованием современных мультиагентных информационных технологий для пожилых людей. Применение широкого спектра новых технологий, включая носимые 3D-датчики и интеллектуальные среды, дает огромные возможности выявления особенностей и состояний пациентов для своевременной корректировки здоровья пожилых людей. Анализируются основные проблемы при разработке таких систем на основе агентов и способы их решения. В работе [15] разрабатывается система децентрализованного мультиагентного управления на основе подхода искусственных иммунной систем.

Проведенный анализ литературы показал, что применение различных модифицированных алгоритмов искусственных иммунных систем и других современных подходов искусственного интеллекта, а также мультиагентных технологий является актуальным и перспективным направлением при обработке многомерных данных, выделении информативных дескрипторов, построении оптимального набора данных для решения задач прогнозирования и диагностики.

1. Постановка задачи формулируется следующим образом: необходимо разработать информационную Smart-систему ведения научных исследований для прогнозирования зависимости „структура–свойство“ (QSAR) лекарственных соединений на основе модифицированных алгоритмов искусственных иммунных систем и мультиагентного подхода на примере молекулярного дизайна новых лекарственных препаратов сульфаниламидной группы.

В статье под модифицированными алгоритмами понимаются алгоритмы, в которых предварительная обработка данных и редукция малоинформативных дескрипторов осуществляется различными оптимизационными алгоритмами ИИ, а для прогнозирования



Рис. 1. Технология прогнозирования QSAR на основе подходов AIS

QSAR используется подход искусственных иммунных систем. Разработанная мультиагентная технология прогнозирования зависимости „структура–свойство/активность“ лекарственных препаратов позволяет выбирать тот алгоритм предварительной обработки данных (алгоритм случайного леса, генетический алгоритм, алгоритм оптимизации серых волков и т. д.), который после решения задачи распознавания образов на базе различных подходов AIS (иммунносетевого моделирования, клональной селекции и негативного отбора) и оценки ошибки обобщения будет давать наилучший прогностический результат (рис.1).

— **Мультиагентная Smart-система ведения научных исследований для прогнозирования QSAR новых лекарственных препаратов.** Создание модифицированных алгоритмов искусственного интеллекта для обработки структурной химической информации и формирования баз данных лекарственных соединений с заданными свойствами на основе оптимального набора дескрипторов является перспективным и важным направлением в фармакологии.

Так как не существует универсальных алгоритмов прогнозирования и наблюдается рост огромного количества новых подходов и модифицированных алгоритмов искусственного интеллекта, то актуальна разработка определенного инструментария научных исследований для анализа и сравнения различных алгоритмов при решении задач QSAR. Разработка удобной и эффективной специализированной мультиагентной Smart-системы прогнозирования QSAR является достаточно сложной задачей, и ниже приведены необходимые требования, которые должны учитываться при обработке многомерной структурной химической информации:

- Возможность обработки больших объемов данных.
- Удобный и понятный для пользователя интерфейс.
- Модульная структура и способность к расширению системы.
- Достаточно высокая скорость обработки информации за счет применения параллельных технологий вычисления.
- Возможность подключения к современным пакетам прикладных программ и библиотекам по обработке и визуализации больших данных.
- Применение облачных технологий.
- Простота работы с системой без длительного обучения.

Ниже приведен алгоритм прогнозирования QSAR:

Шаг 1. Подключение базы данных дескрипторов структурной химической информации, характеризующих рассматриваемые лекарственные соединения. В качестве примера рассматриваются сульфаниламиды с разной продолжительностью фармакологической активности (короткого, среднего и длительного действия).

Шаг 2. Выбор метода искусственного интеллекта: метода главных компонент (PCA, principal component analysis), алгоритма роя частиц (PSO, particle swarm optimization), алгоритма муравьиной колонии (ACO, ant colony optimization), алгоритма пчелиной колонии (ABC, artificial beecolony optimization), алгоритма серых волков (GWO, grey wolf optimizer) или алгоритма случайного леса (RF, random forest) для предварительной обработки и редукции малоинформативных дескрипторов, а также построения оптимального набора данных.

Шаг 3. Построение оптимальной иммунной модели на основе выделенного набора дескрипторов. Избыточность и малоинформативность дескрипторов снижает качество прогноза, поэтому при отборе новых химических соединений в кандидаты лекарственных препаратов важно учитывать информативность дескриптора.

Шаг 4. Выбор алгоритма искусственных иммунных систем: на основе клонального отбора, негативной селекции или иммунносетевого алгоритма для решения задачи распознавания образов прогнозирования.

Шаг 5. Обучение искусственной иммунной системы с „учителем“ по эталонам, составленным из дескрипторов лекарственных соединений с точно известными свойствами. Эталоны составляются экспертами.

Шаг 7. Распознавание образов на основе выбранного алгоритма искусственной иммунной системы.

Шаг 8. Прогнозирование зависимости „структура–свойство“ химических соединений.

Шаг 9. Оценка прогнозирования ИИС.

Шаг 10. Вывод результата. Отбор кандидатов новых химических соединений с заданными свойствами для дальнейших исследований.

При разработке мультиагентной Smart-системы ведения научных исследований созданы следующие агенты (табл. 1): агент базы данных (АБД), онтологический агент (ОА), агент факторного анализа (АФА), агент роя частиц (АРЧ), агент муравьиных колоний (АМК), агент пчелиной колонии (АПК), агент серых волков (АСВ), агент случайного леса (АСЛ), агент искусственных иммунных систем на основе клонального отбора (АКС), агент иммунносетевого моделирования (АИМ), агент искусственных иммунных систем на основе негативной селекции (АНС), агент оценки прогнозирования (АОП), менеджер агент (МА), агент помощник (АП). Модульная структура мультиагентной Smart-системы поз-

Таблица 1

Описание агентов

Название агента	Описание функции агента
Агент базы данных (АБД)	<ul style="list-style-type: none"> — Разработка баз данных дескрипторов различных уровней лекарственных соединений с применением мировых репозиториев химической информации. — Расчет необходимых дескрипторов с применением современного программного обеспечения (ADAPT, CODESSA, PaDEL-Descriptor и т. д.). — Создание базы данных (БД) дескрипторов сульфаниламидов. — Работа с БД.
Онтологический агент (ОА)	<ul style="list-style-type: none"> — Построение онтологических моделей. — Структурирование входных и выходных данных на основе онтологических OWL (Web Ontology Language) моделей.
Агент факторного анализа (АФА)	<ul style="list-style-type: none"> — Построение информативного набора дескрипторов на основе метода главных компонент. — Факторный анализ.
Агент роя частиц (АРЧ)	<ul style="list-style-type: none"> — Реализация алгоритма роя частиц для построения оптимального набора дескрипторов [16].
Агент муравьиной колонии (АМК)	<ul style="list-style-type: none"> — Выделение информативных дескрипторов с помощью алгоритма муравьиной колонии.
Агент пчелиной колонии (АПК)	<ul style="list-style-type: none"> — Создание оптимального набора данных на основе алгоритма пчелиной колонии.
Агент серых волков (АСВ)	<ul style="list-style-type: none"> — Применение алгоритма серых волков для построения оптимального набора дескрипторов.
Агент случайного леса (АСЛ)	<ul style="list-style-type: none"> — Реализация алгоритма случайного леса для выделения информативных дескрипторов [17].
Агент искусственных иммунных систем на основе клонального отбора (АКС)	<ul style="list-style-type: none"> — Решение задачи распознавания образов с использованием алгоритма AIS клонального отбора.
Агент иммунносетевого моделирования (АИМ)	<ul style="list-style-type: none"> — Реализация иммунносетевого алгоритма для решения задачи распознавания образов.
Агент искусственных иммунных систем на основе негативной селекции (АНС)	<ul style="list-style-type: none"> — Применение алгоритма AIS на основе негативной селекции для решения задачи прогнозирования.
Агент оценки прогнозирования (АОП)	<ul style="list-style-type: none"> — Оценка прогнозирования. — Выбор алгоритмов с наилучшими прогностическими результатами.
Менеджер агент (МА)	<ul style="list-style-type: none"> — Осуществление взаимосвязи между агентами. — Организация передачи информации. — Координация работы агентов.
Агент помощник (АП)	<ul style="list-style-type: none"> — Помощь при выборе алгоритма. — Формирование подсказок при вводе параметров алгоритма. — Поддержка при функционировании в программной среде.

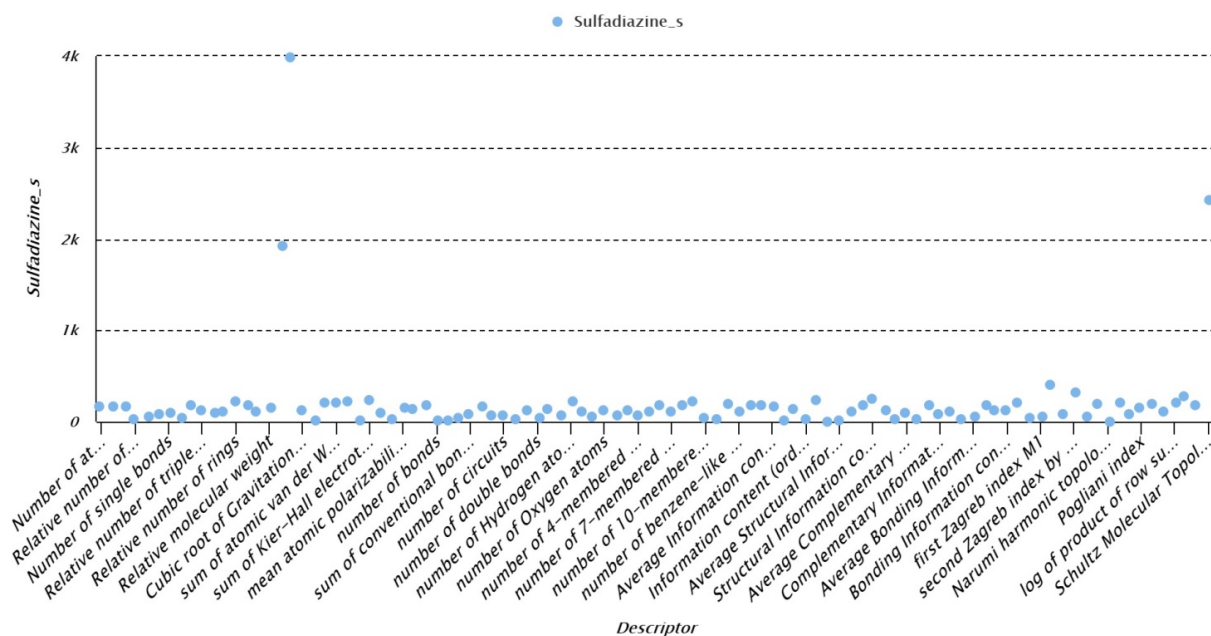


Рис. 2. Фрагмент БД дескрипторов сульфаниламидов

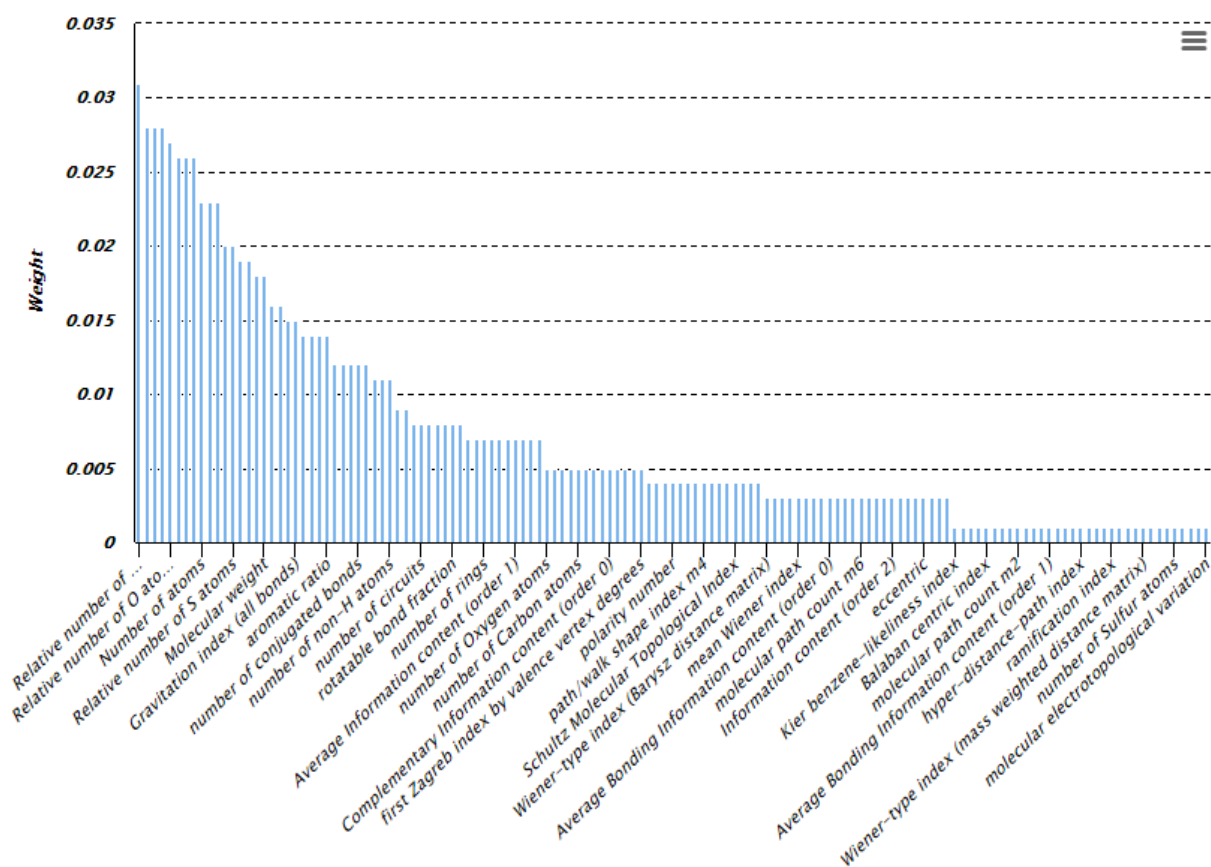


Рис. 3. Результаты предварительной обработки БД сульфаниламидов на основе алгоритма Random Forest

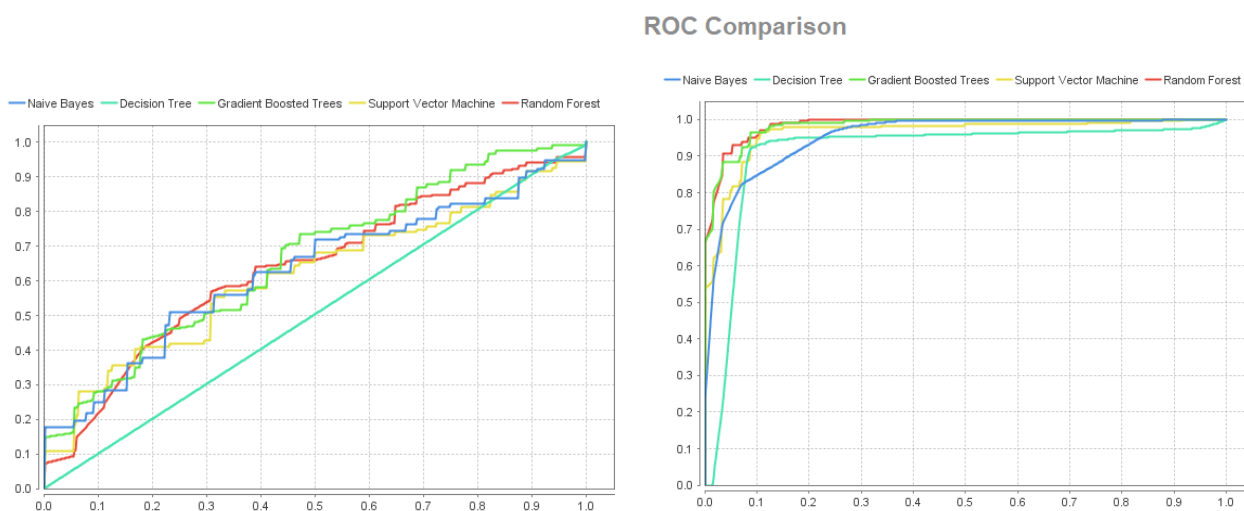


Рис. 4. Сравнительный анализ эффективности прогнозирования QSAR до предварительной обработки данных и после на основе алгоритма Random Forest

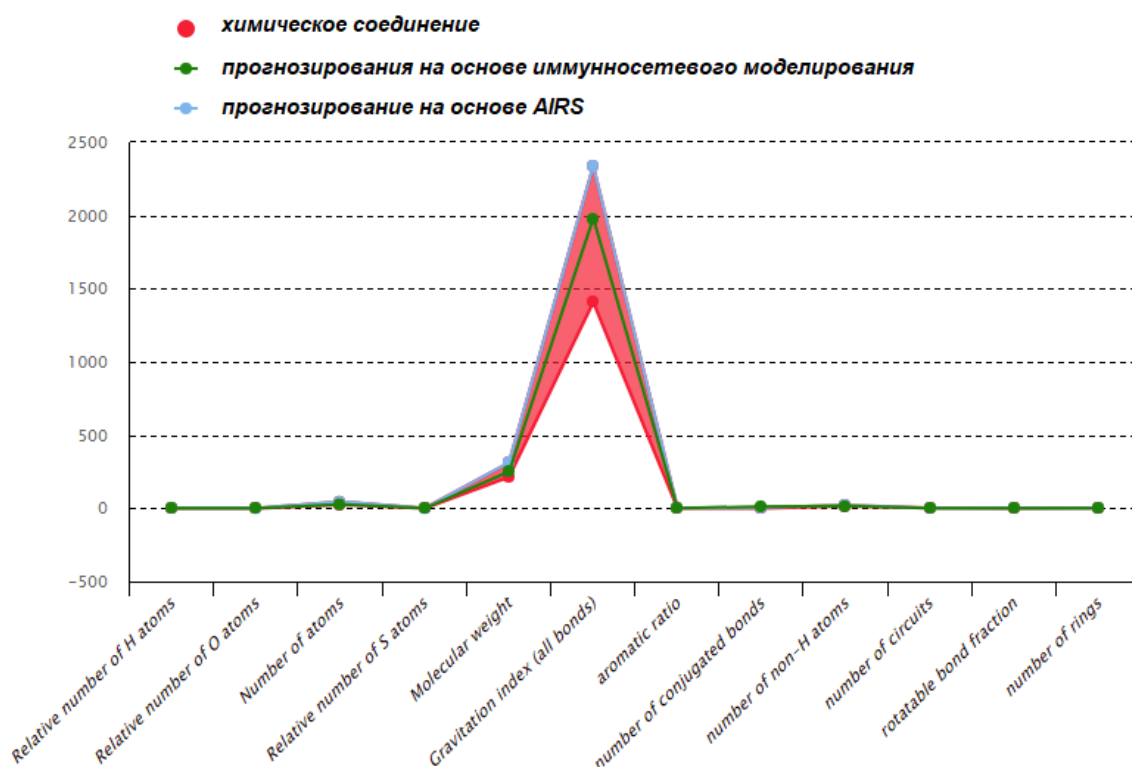


Рис. 5. Прогнозирование QSAR на основе алгоритмов искусственных иммунных систем

воляет при необходимости расширять систему новыми алгоритмами и дополнительными агентами.

Одна из проблем при компьютерном молекулярном дизайне лекарственных препаратов заключается в „парадоксе похожести“, когда соединения различаются структурно совсем незначительно, но имеют совершенно разные свойства. Например, оптически актив-

Таблица 2

Результаты моделирования на основе алгоритма Random Forest

№	Дескриптор	Значение параметра weight
1	Relative number of single bonds	0,031
2	Relative number of H atoms	0,028
3	Relative number of C atoms	0,028
4	Relative number of N atoms	0,027
...
1500	Molecular electrotopological variation	0,001

ное вещество и его зеркальный изомер могут значительно различаться по биологическим свойствам. Поэтому особенно актуальна разработка новых нетрадиционных подходов искусственного интеллекта и алгоритмов, обеспечивающих возможность распознавания химических соединений с почти одинаковой структурой, но имеющих совершенно разные свойства.

— **Результаты моделирования.** Рассмотрим моделирование QSAR на основе базы данных (БД) дескрипторов сульфаниламидов. Сульфаниламиды — это противомикробные препараты с различной продолжительностью действия. База данных составлена с помощью крупнейшего мирового хранилища химической информации Mol-Instincts, в котором на 2019 г. содержится 2,85 млн. химических соединений. Структура химических соединений сульфаниламидов описывается различными дескрипторами. В качестве примера на рис. 2. представлена визуализация фрагмента БД дескрипторов сульфадиазина.

Каждое соединение характеризуется 1500 дескрипторами, в совокупности БД дескрипторов сульфаниламидов состоит из 22500 экземпляров данных. Точность моделей прогнозирования во многом зависит от качества исходных данных. Для построения оптимального набора дескрипторов применяется процедура предварительной обработки данных. В качестве примера рассмотрим эффективность применения алгоритма Random Forest для выделения информативных дескрипторов. Рассчитывается значение параметра weight для каждого дескриптора, чем больше данное значение, тем информативнее дескриптор (табл. 2). Данные с наименьшим показателем подлежат редукции [17].

После редукции малоинформативных дескрипторов получена БД размерностью $R=15 \times 200$, 3000 экземпляров данных. Рассмотрим эффективность применения процедуры предварительной обработки данных на основе алгоритма RF на примере следующих прогностических моделей (рис. 4): Наивный Байесовский алгоритм (Naive Bayes), деревья решений (Decision Tree), случайный лес, деревья решений с градиентным бустингом (Gradient Boosted Trees) и метод опорных векторов (Support Vector Machine).

На рис. 4 показан сравнительный анализ кривых ошибок моделей прогнозирования до предварительной обработки данных на основе RF и после с помощью ROC-анализа (Receiver Operating Characteristic, ROC). Чем ближе кривая к левому верхнему углу, тем лучше модель. Таким образом, по результатам анализа видно, что после процедуры редукции неинформативных дескрипторов качество моделей прогнозирования значительно улучшилось.

Далее на рис. 5 рассмотрено прогнозирование оптимальной базы данных дескрипторов сульфаниламидов на основе иммунносетевого моделирования и алгоритма распознавания искусственной иммунной системой (Artificial Immune Recognition System).

Сравнительный анализ полученных результатов показывает, что после предварительной обработки данных с помощью алгоритма Random Forest наиболее лучший прогностический результат показывает модель на основе иммунносетевого моделирования.

Заключение. Таким образом, разработка мультиагентной Smart-системы [18] ведения научных исследований способствует созданию теоретических основ принципиально новой технологии построения баз данных химических соединений на основе модифицированных алгоритмов AIS [19] для отбора кандидатов в лекарственные препараты с заданными свойствами и внедрения их в реальное фармакологическое производство для дальнейших исследований.

Список литературы

1. ROY K., KAR S., DAS R.N. Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment. 2015. NY: Academic Press, [Electron. Res.]: <https://www.elsevier.com/books/understanding-the-basics-of-qsar-for-applications-in-pharmaceutical-sciences-and-risk-assessment/roy/978-0-12-801505-6> (дата доступа 06.02.2019).
2. ROY K. Advances in QSAR Modeling. Applications in Pharmaceutical, Chemical, Food, Agricultural and Environmental Sciences. 2017. Springer, [Electron. Res.]: <http://www.springer.com/in/book/9783319568492> (дата доступа 06.02.2019).
3. ГАЛЬБЕРШТАМ Н. М., БАСКИН И. И., ПАЛЮТИН В. А., ЗЕМФИРОВ Н. С. Нейронные сети как метод поиска зависимостей структура–свойство органических соединений // Успехи химии. 2003. № 72 (7). С. 706–727.
4. IVANCIUC O. Structure-activity relationships in aquatic toxicology with artificial immune systems: Mechanism of toxic action classification of polar and nonpolar narcotic pollutants with CLONALG (clonal selection algorithm) // J. Molecular Design. 2007. N 6. P. 106–114.
5. CHUNLIN LIANG, LINGXI PENG. An Automated Diagnosis System of Liver Disease using Artificial Immune and Genetic Algorithms // Journal of Medical Systems, 2013. DOI:10.1007/s10916-013-9932-9.
6. ÜMIT YILMAZ, HELMI MD RAIS, SAID JADID ABDULKADIR. Hybrid Swarm Intelligence Algorithms with Ensemble Machine Learning for Medical Diagnosis // Proceedings of 4th International Conference on Computer and Information Sciences (ICCOINS), 2018.
7. ZHU H., WU J., GU J. Studies on Immune Clonal Selection Algorithm and Application of Bioinformatics // International Journal of Intelligent Engineering and Systems. 2015. V. 8. N 1. P. 10–16.
8. СОСНИН С. Б., РАДЧЕНКО Е. В., ПАЛЮЛИН В. А., ЗЕФИРОВ Н. С. Обобщенный фрагментный подход в исследованиях QSAR/QSPR // Доклады Академии наук. Химия. 2015, Т. 463. № 3. С. 297–300.
9. GHASEMI, PEREZ-SANCHEZ; MEHRI, PEREZ-GARRIDO. Neural network and deep-learning algorithms used in QSAR studies: merits and drawbacks // Drug Discovery Today. 2018. N 23 (10). P. 1784–1790. DOI:10.1016/j.drudis.2018.06.016. PMID29936244.
10. YAVUZ B. Z., SERTKAYA C., YURTAY N. Prediction of secondary structures of hemoglobin using clonal selection algorithm // Proceedings of the 7 th Int. Work. Comput. Sci. Eng. 2017. P. 1387–1391.
11. SAMIGULINA G. A., SAMIGULINA Z. I. Immune network technology on the basis of Random Forest algorithm for computer aided drug design // J. 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.

12. LUCIJA BREZOČNIK, IZTOK FISTER JR., VILI PODGORELEC. Swarm Intelligence Algorithms for Feature Selection: A Review // J. Applied Sciences. 2018. N 8 (9). P. 1521. [Electron. Res.]: <https://doi.org/10.3390/app8091521>.

13. RAMÍREZ M. R., RAMÍREZ MORENO H. B., ROJAS E. M., HURTADO C., VAZQUEZ NÚÑEZ S. O. Multi-Agent System Model for Diagnosis of Personality Types // Proceedings of the 12th International Conference Agents and Multi-agent Systems: Technologies and Applications. Smart Innovation, Systems and Technologies book series (SIST). Springer, 2018. Vol. 96. P. 209–214. DOI: 10.1007/978-3-319-92031-3_19.

14. IVANOVIĆ M., NINKOVIĆ S. Personalized HealthCare and Agent Technologies // Proceedings of the 11th International Conference Agents and Multi-agent Systems: Technologies and Applications. Smart Innovation, Systems and Technologies book series (SIST). Springer: KES-AMSTA, 2017. V. 74. P. 3–11.

15. GERMAN S., SHIN S., TSOURDOS A. // Proceeding of the IEEE 24th Mediterranean Conference on Control and Automation (MED). 2016. P. 1020–1025.

16. SAMIGULINA G. A., MASSIMKANOVA ZH. A. Multiagent system of recognize on the basis of modified algorithms of swarm intelligence and immune network modeling // Proceedings of the 12th International Conference Agents and Multi-agent Systems: Technologies and Applications (AMSTA-18). Australia: Springer, 20–22 June 2018. P. 199–208.

17. 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.

18. 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 (IWBBIO2018). Spain, Granada, 25–27 April 2018. P. 1–2.

19. САМИГУЛИНА Г. А., САМИГУЛИНА З. И. Разработка технологии иммуносетевого моделирования для компьютерного молекулярного дизайна лекарственных препаратов (программа для ЭВМ). Свидетельство о государственной регистрации прав на объект авторского права в Комитете по правам интеллектуальной собственности МЮ РК. Астана, 28 марта 2011. № 473. 23 с.



Самигулина Галина Ахметовна — доктор технических наук, Академик МАИН, зав. лаб. „Интеллектуальные системы управления и прогнозирования“ Института информационных и вычислительных технологий, Казахстан

Самигулиной Г. А. разработана иммуносетевая технология построения интеллектуальных систем прогнозирования и управления сложными объектами в условиях неопределенности параметров. Сфера деятельности: интеллектуальные системы промышленной автоматизации, Smart-системы дистанционного обучения для людей с ограниченными возможностями зрения; молекулярный дизайн лекарственных препаратов с заданными свойствами; прогнозирование рис-

ков сложных инвестиционных проектов и др. Опубликовано более 250 научных работ в ближнем и дальнем зарубежье (более 15 публикаций в базах данных ThomsonReuters и Scopus), 6 монографий; 2 учебных пособия, получено 10 авторских свидетельств. В 2012 году Комитетом науки присуждена стипендия „Ученым, внесшим выдающийся вклад в развитие науки и техники“. В 2017 году награждена нагрудным знаком „За заслуги в развитии науки Республики Казахстан“.

Samigulina Galina Ahmetovna — Doctor of Technical Sciences, Academician IAIN, head of the lab. „Intellectual control systems and forecasting“ of the Institute of Information and Computational Technologies, Kazakhstan

An immuno-network technology for building intelligent prediction and control systems for

complex objects under conditions of parameter uncertainty has been developed by Samigulina G. A. Field of activity: intellectual systems of industrial automation, Smart-distance learning systems for people with visual disabilities; molecular design of drugs with desired properties; forecasting the risks of complex investment projects, etc. There are more than 250 scientific papers published in the near and far abroad (more than 15 publications in the Thomson Reuters and Scopus databases), 6 monographs; 2 textbooks, received 10 copyright certificates. The Committee of Science awarded a scholarship to „Scientists who have made outstanding contributions to the development of science and technology“ in 2012 year. She was awarded the badge „For merits in the development of science of the Republic of Kazakhstan“ in 2017 year.



Самигулина Зарина Ильдусовна — ассоц. профессор Казахстанско-Британского технического университета; e-mail: zarinasamigulina@mail.ru

Самигулиной З. И. разработана адаптивная система управления на основе эталонной модели траекторией движения космического аппарата. Создана интеллектуальная технология сбора и обработки данных с объектов промышленной автоматизации на основе микропроцессорной техники фирмы SchneiderElectricиSiemens. Пред-

ложен компонентно-ориентированный подход для автоматизированной системы ведения научных исследований при создании новых лекарственных препаратов.

Опубликовано более 120 научных работ ближнем и дальнем зарубежье (из них 7 статей в базе данных Scopus, 4 статьи, входящие в базу данных ThompsonReuters), 2 монографии; 7 методических указаний для выполнения лабораторных работ, получено 8 авторских свидетельств на программы ЭВМ, 5 актов внедрений.

Samigulina Zarina Ildusovna — Doctor of PhD, associate professor of Information Technologies Faculty, Kazakh-British Technical University.

Samigulina Z. I. developed an adaptive control system based on the standard model of the spacecrafttrajectory of movement. Was created an intelligent technology for data collection and processingfrom industrial automation objects based on microprocessor technology from Schneider Electricand Siemens companies. Was proposed a component-oriented approach for an automated systemfor research conducting at the development of new drugs.

Were published more than 120 scientific papers in the near and far abroad (including 7 articles in the Scopus database, 4 articles included in the Thompson Reuters database), 2 monographs; 7 methodical instructions for the laboratory work, obtained 8 copyright certificates for computer programs, 5 acts of implementation.

Дата поступления — 15.02.2019