

# ПОСТРОЕНИЕ ОПТИМАЛЬНОЙ ИММУНОСЕТЕВОЙ МОДЕЛИ ДЛЯ ПРОГНОЗИРОВАНИЯ СВОЙСТВ НЕИЗВЕСТНЫХ ЛЕКАРСТВЕННЫХ СОЕДИНЕНИЙ НА ОСНОВЕ МУЛЬТИАЛГОРИТМИЧЕСКОГО ПОДХОДА

Г. А. Самигулина, З. И. Самигулина\*

Институт проблем информатики и управления Министерства образования и науки Республики Казахстан, 050010, Алма-Ата, Казахстан

\* Казахский национальный технический университет им. К. И. Сатпаева  
050013, Алма-Ата, Казахстан

---

УДК 004.89

Разрабатываются методы построения оптимальной иммуносетевой модели для прогноза свойств неизвестных химических соединений на основе искусственных иммунных систем и мультиалгоритмического подхода с целью выявления перспективных лекарственных веществ. Выделение информативных дескрипторов и построение оптимальной иммуносетевой модели осуществляются на основе методов факторного анализа и искусственных нейронных сетей.

**Ключевые слова:** интеллектуальная технология, искусственные иммунные системы, лекарственные препараты, прогнозирование фармакологической активности, ошибки обобщения, факторный анализ, нейросетевой подход.

Article is devoted to the development of the method of the construction an optimal immune network model for the prediction the properties of the unknown chemical compounds on the based on artificial immune systems (AIS) and multi algorithmic approach for the purpose of discovery perspective medicinal substances. Separation of the informative descriptors and construction of the optimal immune network model realized on the based on methods of the factor analyzes and artificial neural networks.

**Key words:** intellectual technology, artificial immune systems, medical products, forecasting of pharmacological activity, a generalisation error, the factorial analysis, the neural network approach.

**Введение.** Компьютерный молекулярный дизайн новых лекарственных препаратов является важнейшей составляющей процесса исследования биологической активности химических соединений и представляет собой совокупность подходов, методов и программ, которые применяются для моделирования [1].

Используется концепция взаимосвязи молекулярной структуры и биологической активности химических соединений. Существующие методы, такие как QSAR (quantitative structure activity relationships), QSPR (quantitative structure property relationships), искусственные нейронные сети и др., разработаны для ускорения и оптимизации поиска новых биологически активных веществ. Актуальным является создание новых нетрадиционных подходов, предназначенных для решения этой сложной проблемы.

При изучении связи между структурой и активностью соединений необходимо обрабатывать огромные массивы данных с использованием последних достижений в области интеллектуальных систем и суперкомпьютеров, что осложняется междисциплинарным характером исследований. При разработке интеллектуальных систем прогнозирования свойств

лекарственных препаратов используются новейшие результаты в области биохимии, молекулярной биологии, компьютерного моделирования, хемометрики, фармакологии, прикладной математики и искусственного интеллекта.

При нетрадиционном подходе для обработки огромных массивов данных используются принципы имитационного моделирования естественных систем. В последнее время широкое распространение получили нейронные системы и эволюционные алгоритмы.

**1. Подход искусственных иммунных систем.** Особое внимание уделяется искусственным иммунным системам (artificial immune system) [2], основанным на обработке информации молекулами белков и иммунологической реакции организма при вторжении чужеродных антигенов. В настоящее время экспериментально подтверждено существование иммунных сетей в организме человека. За счет взаимодействия белков иммунной системы между собой создаются иммунные сети. Преимуществом данных систем являются распределенность, самообучаемость, отсутствие централизованного контроля, самоорганизация и эволюция, малые вычислительные ресурсы, возможность параллельной обработки информации.

В подходе искусственных иммунных систем (ИИС) существует несколько направлений: на основе теории отрицательного отбора, теории клональной селекции, теории иммунной сети. Перспективным считается подход, в котором рассматриваются механизмы молекулярного узнавания [3]. В этом случае базовым элементом является формальный пептид. Под формальным пептидом понимается математическая абстракция свободной энергии белковой молекулы от ее пространственной формы. Это относительно новое направление в искусственном интеллекте, с использованием которого разработан ряд приложений.

При иммуносетевом моделировании возникают следующие проблемы:

- выбор структуры иммунной сети;
- уменьшение времени, затрачиваемого на обучение иммунной сети;
- выбор оптимальной структуры иммунной сети;
- решение задачи выделения информативных признаков;
- уменьшение ошибки обобщения иммунной сети;
- повышение достоверности прогноза;
- работа в реальном масштабе времени;
- распараллеливание вычислительных алгоритмов;
- разработка автоматизированной системы построения технологических цепочек для различных критериев качества на основе формального языка интеллектуальной технологии обработки многомерных данных ИИС;
- разработка операционных маршрутов пользователя и создание библиотек для иммуносетевого моделирования;
- визуализация данных.

В подходе ИИС под сетью связывания понимается любая последовательность связываний формальных пептидов. Для построения иммуносетевой модели необходимо создать формальный пептид, который будет служить эталоном лекарственного вещества с нужными свойствами. Экспертами выбирается известное вещество, которое описывается с помощью дескрипторов.

Особенностью данного подхода ИИС является возможность использования различных дескрипторов (рис. 1). Например, химическое вещество можно описать дескрипторами элементарного уровня (ДЭУ), дескрипторами структурной формулы (ДСФ), дескрипторами

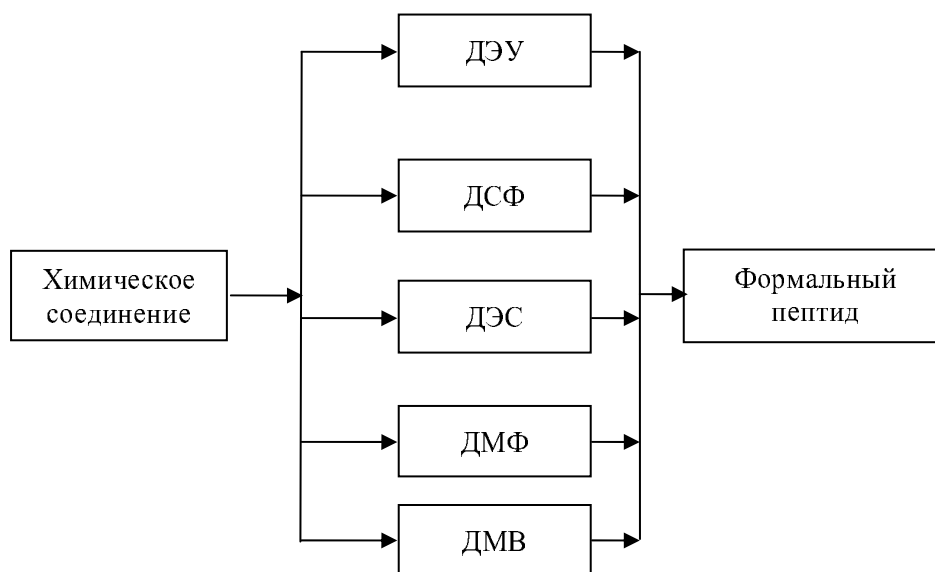


Рис. 1. Описание химического вещества с помощью различных дескрипторов

электронной структуры (ДЭС), дескрипторами молекулярной формы (ДМФ), дескрипторами межмолекулярных взаимодействий (ДМВ).

Приведем пример использования дескрипторов структурной формулы для построения формального пептида. Рассмотрим класс седативных лекарственных препаратов. Барбитураты (barbiturates) — это группа лекарственных средств, производных барбитуровой кислоты ( $\text{CONHCOCH}_2\text{CONH}$ ), оказывающих угнетающее влияние на центральную нервную систему [4]. Данные препараты обладают снотворным, противосудорожным и наркотическим действием.

Разные барбитураты имеют различную продолжительность действия на организм. Существуют препараты с длительной (барбитал, фенобарбитал, барбитал-натрий), средней (циклобарбитал, барбамил, этаминал-натрий) и короткой (гексобарбитал) продолжительностью действия. Более подробные данные о структурах химических соединений барбитуратов и их влиянии на организм человека приведены в работе [5]. С помощью этих данных можно построить формальные пептиды — эталоны для трех классов: сильнодействующие, слабые и умереннодействующие снотворные. Указанная классификация необходима для решения задачи распознавания образов на основе искусственных иммунных систем.

**2. Постановка задачи.** Необходимо разработать процедуру построения оптимальной иммуносетевой модели для прогнозирования свойств и активности неизвестных лекарственных соединений на основе дескрипторов (соответствующих структуре исследуемого химического соединения) и мультиалгоритмического подхода.

**Определение.** В разрабатываемой интеллектуальной технологии иммуносетевого моделирования лекарств под оптимальной структурой иммунной сети понимается сеть, построенная на основе весовых коэффициентов выделенных информативных дескрипторов, которые наиболее полно описывают рассматриваемое химическое соединение. Критерием является максимальное сохранение информации при использовании минимального количества дескрипторов.

При построении оптимальной иммуносетевой модели для выделения информативных дескрипторов одновременно используется несколько подходов и алгоритмов. По результатам проведенных вычислений выбирается наилучший алгоритм, т. е. набор информативных

дескрипторов, который содержит наименьшее количество различного рода ошибок и более пригоден для дальнейшего прогнозирования.

Предпочтителен тот алгоритм, при использовании которого ошибка обобщения минимальная. Под ошибкой обобщения  $E$  понимается ошибка, допускаемая моделью, основанной на данных, не использованных при обучении иммунной сети. При построении модели основной целью является уменьшение ошибки обобщения, так как малость ошибки обучения не гарантирует малость ошибки обобщения.

Можно рассчитать относительную ошибку обобщения  $E^*$ , которая показывает, во сколько раз величина ошибки обучения меньше величины ошибки обобщения.

Для выделения информативных дескрипторов предлагается использовать алгоритмы факторного анализа и нейросетевого подхода.

При построении оптимальной структуры иммунной сети на основе факторного анализа (ФА) используется метод главных компонент [6]. Данные вычисления осуществляются с помощью программы SPSS [7].

### **3. Выделение информативных дескрипторов на основе нейросетевого подхода.**

В случае использования нейронных сетей (НС) для выбора оптимальной сети используется программа NeuroShell, разработанная компанией AI Trilogy (Ward System Group). С помощью этой программы, которая является универсальным пакетом, можно решать широкий спектр сложных задач. Применение данного пакета программ целесообразно при проведении предварительной обработки данных и выделении информативных дескрипторов.

Программный продукт NeuroShell имеет следующие преимущества:

- простота интерфейса и легкость использования;
- наличие русифицированной версии;
- работа с большими массивами данных (внутренний формат поддерживает 65 535 рядов и 32 767 столбцов);
- возможность реализации достаточно большого числа архитектур нейронных сетей;
- работа в режиме реального времени;
- возможность использования для различных приложений.

Приведем алгоритм выделения информативных признаков на основе программы NeuroShell.

*Шаг 1.* Необходимо открыть рабочее окно программы NeuroShell и выбрать раздел “Нейронные сети для профессионала”. Данный раздел программы позволяет выбрать архитектуру нейронной сети различной сложности для достижения контроля за процессом обработки данных. До начала работы проект сохраняется в рабочей директории.

*Шаг 2.* Осуществляется импорт данных. Возможна загрузка через стандартные форматы (текстовые файлы, двоичные файлы) и специальные форматы (файлы электронных таблиц и т. д.). Программа позволяет загрузить базу данных, например с расширением .xls. После импорта данные заносятся в электронную таблицу.

*Шаг 3.* В разделе “Предобработка” доступна функция “Преобразование символов” для перекодирования текста. Следующий раздел “Правила” позволяет создавать систему правил для предобработки данных, например устанавливать пороговое значение. На основе полученных данных возможно построение графиков (по всем примерам, внутри примера, корреляционной точечной диаграммы и т. д.).

*Шаг 4.* На следующем этапе осуществляется построение сети. Сначала выделяются входы и выходы, затем выбираются максимальное и минимальное значения данных, а также рассчитываются среднее значение и стандартное отклонение. Далее выделяется тестовый

набор данных. В разделе “Проектирование” необходимо выбрать архитектуру и параметры сети. После этого настраиваются основные параметры обучения и критерии останова. Различными цветами выделены сети, рекомендуемые для предсказания (желтый цвет), и сети, рекомендуемые для классификации (зеленый цвет). После выбора архитектуры сети необходимо настроить тип выбора примеров, вариант подстройки весов, а также возможность заполнения пропущенных данных.

*Шаг 5.* Далее осуществляется обучение сети до момента достижения соответствия критерию останова, выполняется расчет текущей средней ошибки и минимальной средней ошибки. Пользователь может посмотреть структуру нейросети. Далее можно открыть окно просмотра весов связи нейросети. Для выявления закономерностей в весах связи можно построить цветную диаграмму Хинтона.

*Шаг 6.* На основе анализа обученной нейронной сети рассчитывается относительный показатель важности для каждой входной переменной сети, являющийся грубой мерой важности по сравнению с другими переменными сети. Наибольший показатель соответствует наилучшему вкладу в работу сети.

Таким образом, программа рассчитывает относительные показатели важности, т. е. весовые коэффициенты дескрипторов. При проведении дальнейших расчетов малоинформативные дескрипторы удаляются.

**4. Интеллектуальная технология иммуносетевого моделирования и прогнозирования свойств новых лекарственных соединений.** Приведем алгоритм иммуносетевого моделирования лекарств на основе мультиалгоритмического подхода (рис. 2).

*Шаг 1.* Описание химических соединений дескрипторами, выбор вида дескрипторов.

*Шаг 2.* Построение с помощью экспертов ФП-эталонов для различных классов.

*Шаг 3.* Построение оптимальной структуры иммунной сети на основе методов факторного анализа и нейросетевого подхода.

*Шаг 4.* Обучение иммунной сети.

*Шаг 5.* Формирование ФП-образов.

*Шаг 6.* Решение задачи распознавания образов на основе сингулярного разложения матриц (singular value decomposition (SVD)).

*Шаг 7.* Оценка энергетических погрешностей [8, 9] и выбор (по минимуму ошибки обобщения) наилучшего алгоритма для прогнозирования.

*Шаг 8.* Прогноз свойств и активностей неизвестных химических соединений.

*Шаг 9.* Классификация новых химических соединений по их свойствам и отбор соединений-кандидатов в лекарства.

*Шаг 10.* Проведение дальнейших исследований.

В связи с повышением требований к качеству фармакологической продукции общепризнанным перспективным направлением является разработка методов оценки безопасности и эффективности лекарственных препаратов на основе технологий биоинформатики и компьютерного конструирования лекарств. Особая роль в этих исследованиях уделяется разработке методов определения безопасности лекарственных препаратов и их влияния на организм человека.

При выделении информативных дескрипторов в процессе построения оптимальной структуры иммунной сети на основе различных алгоритмов (факторного анализа и нейронных сетей) возникает проблема распараллеливания вычислений для уменьшения времени, затрачиваемого на реализацию предлагаемой технологии на основе подхода искусственных иммунных систем.



Рис. 2. Схема иммуносетевого моделирования лекарственных препаратов

**5. Параллельные вычисления при иммуносетевом моделировании биологической активности органических соединений.** Преимуществом предлагаемой интеллектуальной технологии является возможность использования параллельных вычислений и суперкомпьютеров, так как обработка огромных массивов данных требует применения последних научных достижений в этой области. Использование параллельных вычислительных алгоритмов позволяет существенно сократить временные и финансовые затраты при компьютерном моделировании лекарственных препаратов.

При решении задачи распознавания образов предлагается использовать следующий алгоритм распараллеливания:

- после предварительной обработки с помощью экспертов формируются пептиды-эталонны для всех рассматриваемых классов;
- каждый эталон записывается на отдельный процессор;
- из базы данных формируются пептиды-“образы”;
- каждый образ сравнивается с эталонами на всех процессорах параллельно, т. е. сингулярное разложение матриц и нахождение энергии связи осуществляются одновременно для всех эталонов. Эти вычисления проводятся параллельно для определения минимальной энергии связи и решения задачи распознавания образов, что позволяет существенно уменьшить время, необходимое для обработки информации и прогнозирования свойств химических соединений.

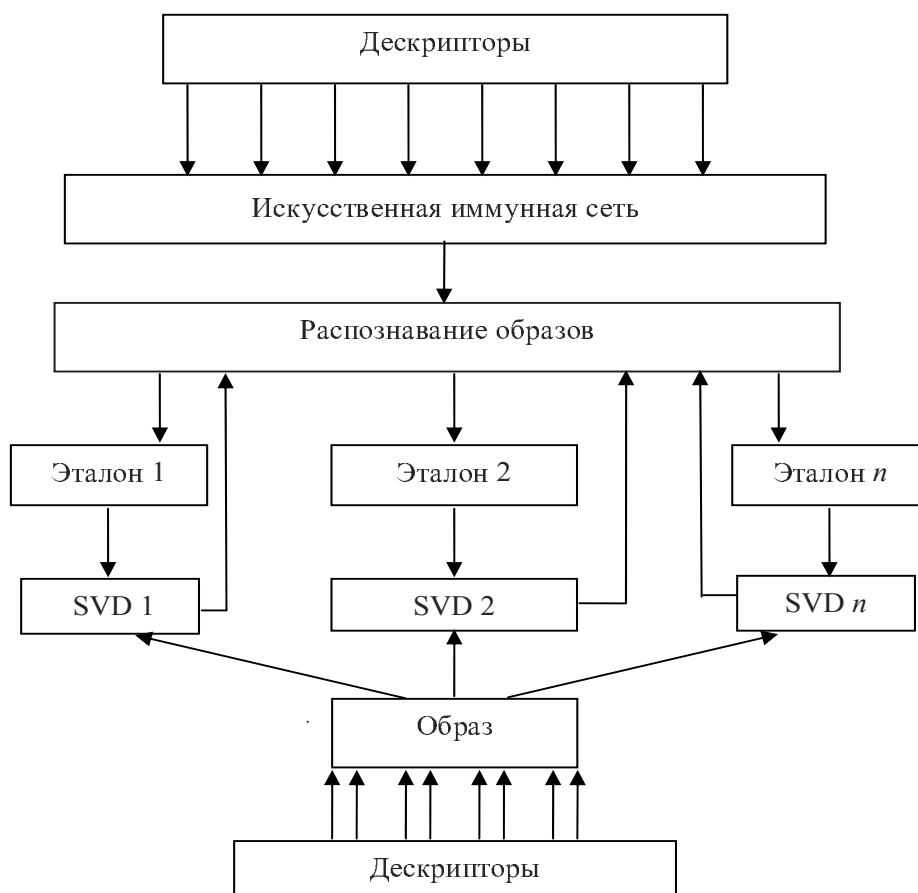


Рис. 3. Распараллеливание вычислений искусственной иммунной системы

На рис. 3 представлена схема распараллеливания вычислений при иммуносетевом моделировании лекарственных препаратов ( $n$  — количество эталонов химических соединений с известными свойствами (количество классов); SVD — сингулярное разложение матриц). По окончании решения задачи распознавания образов для всех эталонов одновременно выполняется оценка энергетических погрешностей.

**Заключение.** Областью применения новой интеллектуальной технологии является компьютерное прогнозирование фармакологической активности органических соединений с целью выявления перспективных лекарственных веществ. Разработанные алгоритмы позволяют уменьшить временные и финансовые затраты на проведение исследований, предшествующих клиническим испытаниям. Разработанная интеллектуальная технология может применяться при поиске химических соединений в базах данных и планировании синтеза новых химических соединений с высокой фармакологической активностью.

При решении задач компьютерного молекулярного дизайна лекарственных препаратов на основе иммуносетевого моделирования используются мультиалгоритмический подход и многопроцессорная вычислительная система с параллельной реализацией этих алгоритмов.

Преимуществом подхода искусственной иммунной системы является возможность (при определенных условиях, определяемых экспертами) использования в формальном пептиде одновременно несколько уровней дескрипторов. Кроме того, можно включить параметры математической модели, учитывающей влияние данного препарата на организм человека.

Оценки энергетических ошибок на основе гомологов [8] позволяют повысить достоверность прогноза искусственной иммунной системы, так как предоставляют возможность распознавать пептиды с почти одинаковыми параметрами на границах нелинейно разделенных классов.

На программное обеспечение по разработке технологии иммуносетевого моделирования для компьютерного молекулярного дизайна лекарственных препаратов получено авторское свидетельство о государственной регистрации объекта интеллектуальной собственности [11].

## Список литературы

1. КУВИНЫ Г. В поисках новых соединений-лидеров для создания лекарств // Рос. хим. журн. 2006. № 2. С. 5.
2. ИСКУССТВЕННЫЕ иммунные системы и их применение / Под ред. Д. Дасгупты. М.: Физматлит, 2006.
3. ТАРАКАНОВ А. О. Formal peptide as a basic of agent of immune networks: from natural prototype to mathematical theory and applications // Proc. of the 1st Intern. workshop of Central and Eastern Europe on multi-agent systems. St. Petersburg, 1999. P. 37.
4. БЕЛИКОВ В. Г. Фармацевтическая химия. Астана: Комитет по правам интелект. собственности Мин-ва юстиции Республики Казахстан, 2007.
5. СТЬЮПЕР Э. Машинный анализ связи химической структуры и биологической активности / Э. Стьюпер, У. Брюггер, П. Джурс. М.: Мир, 1982.
6. ИБЕРЛА К. Факторный анализ. М.: Статистика, 1980.
7. САМИГУЛИНА Г. А., ЧЕБЕЙКО С. В. Разработка технологий иммуносетевого моделирования для компьютерного молекулярного дизайна лекарственных препаратов // Вестн. Харьк. ун-та. Темат. вып. Информатика и моделирование. 2011. № 17. С. 142–148.
8. SAMIGULINA G. A. Development of the decision support systems on the basis of the intellectual technology of the artificial immune systems // Automat. Remold Control. 2012. V. 74, N 2. P. 397–403.
9. САМИГУЛИНА Г. А., ЧЕБЕЙКО С. В. Интеллектуальная технология прогнозирования фармакологической активности органических соединений на основе искусственных иммунных систем // Тр. Междунар. 3D Интернет-конф. “Математическое и компьютерное моделирование в биологии и химии. Перспективы развития”, Казань, 28–30 мая 2012 г. Казань: ВМиИТ, 2012. С. 152–155.
10. САМИГУЛИНА Г. А., САМИГУЛИНА З. И. Разработка методов построения оптимальной иммуносетевой модели для прогноза свойств неизвестных лекарственных соединений // Тр. Междунар. науч.-практ. конф. “Современные научные достижения — 2013”, Прага (Чехия), 27 янв. — 5 февр. 2013 г. Прага: Наука и образование, 2013. С. 46.
11. А. с. Разработка технологии иммуносетевого моделирования для компьютерного молекулярного дизайна лекарственных препаратов (программа для ЭВМ) / Г. А. Самигулина, З. И. Самигулина. Опубл. 28.03.11. № 473.

*Самигулина Галина Ахметовна — д-р техн. наук, доц., зав. лабораторией  
Института проблем информатики и управления Министерства образования  
и науки Республики Казахстан; e-mail: galinasamigulina@mail.ru;  
Самигулина Зарина Ильдусовна — PhD-докторант Казахского национального  
технического университета и.м. К. И. Сатпаева; e-mail: zarinasamigulina@mail.ru*

Дата поступления — 15.03.13