

УДК 004.89:004.4

Г.А. Самигулина

## МОДЕЛИРОВАНИЕ ТЕХНОЛОГИИ ДИСТАНЦИОННОГО ОБУЧЕНИЯ НА ОСНОВЕ ИСКУССТВЕННЫХ ИММУННЫХ СЕТЕЙ

*Разрабатывается интеллектуальная образовательная технология на основе иммуносетевого моделирования, которая позволяет анализировать значительные массивы данных в режиме реального времени и адаптировать учебный процесс к конкретному пользователю. Достоинством предложенной интеллектуальной технологии является возможность оперативного управления процессом и качеством обучения, индивидуальный подход к каждому обучающемуся, повышение достоверности прогнозирования результатов обучения путем уменьшения ошибок обобщения искусственной иммунной сети.*

### Введение

Одним из важнейших факторов развития информационно-образовательного пространства является использование сети Интернет для широкого доступа различных групп населения к образовательным ресурсам, поэтому проблема развития электронного (дистанционного) образования весьма актуальна не только для Казахстана, но и для других стран.

Министерство образования и науки Республики Казахстан в соответствии с Государственной программой развития образования на 2005–2010 гг. проводит целенаправленную работу по созданию информационно-инновационной образовательной системы, программы развития и поддержки электронного обучения. Широкое внедрение интернет-технологий в Республике Казахстан приводит к необходимости интеграции информационно-образовательных ресурсов и услуг, развитию нетрадиционных инновационных подходов и образовательных технологий, отвечающих последним требованиям времени. Данные образовательные технологии должны быть направлены на повышение качества получаемых образовательных услуг, развитие индивидуального подхода к каждому обучаемому и возможности оперативного управления процессом электронного обучения. Разработка образовательных порталов, построенных на этих принципах, является первостепенной задачей для Республики Казахстан.

Более чем в 100 странах мира широко применяется виртуальная образовательная среда MOODLE (Modular Object-Oriented Dynamic Learning Environment – модульная объектно-ориентированная динамическая учебная среда). Особенностью и преимуществом этой среды по сравнению с другими системами дистанционного обучения является то, что она распространяется в открытом коде. Это дает возможность встраивать в нее новые компоненты. Национальным центром информатизации Казахстана в рамках проекта ЮНЕСКО «Дистанционное образование с применением информационно-коммуникационных технологий (ИКТ) для средних школ отдаленных и уязвимых регионов» разработан портал MOODLE НЦИ (URL: <http://moodle.nci.kz>).

Данная статья посвящена исследованиям, направленным на разработку интеллектуальной системы управления дистанционным обучением (ИСУДО), которая может быть применена в MOODLE.

В настоящее время весьма актуальным является поиск новых эффективных нетрадиционных методов обучения, а также инновационных подходов к осуществлению контроля качества усвоения. Одно из главных требований к данным системам – способность анализировать огромные массивы данных в режиме реального масштаба времени для эффективного управления процессом обучения.

Известно, что эффективность процесса получения знаний на основе систем дистанционного обучения зависит от многих факторов, таких, как возраст обучающегося, память, наличие личной заинтересованности, базовое образование, психологический портрет, способность к восприятию информации тем или иным путем и т. д. В связи с этим огромное значение имеет возможность оперативного динамического изменения учебной программы, методики и способов подачи учебного материала в зависимости от степени усвоения пройденного. Необходимы механизмы

быстрого анализа текущей ситуации в процессе получения знаний каждым обучающимся. Контроль и оперативное управление процессом получения знаний в среде Интернет требуют анализа большого числа индивидуальных признаков, характеризующих каждую личность. Это необходимо для учета особенностей обучающихся и оперативной корректировки предлагаемого учебного материала в зависимости от темпов обучения. Достаточно большой набор информативных признаков позволяет более точно отразить интеллектуальный потенциал каждого обучающегося. С учетом этого потенциала строится программа обучения и т. д.

Разработанная ИСУДО направлена на интенсификацию процесса, индивидуальный подход, повышение качества обучения, развитие логического мышления и навыков творческой работы (посредством выполнения тестовых заданий), позволяет оперативно корректировать индивидуальную программу обучения на основе комплексной оценки знаний для различных классов, выделенных экспертами.

В связи с этим актуально применение биологического подхода искусственных иммунных систем (ИИС), построенных на принципах обработки информации молекулами белков.

### 1. Принципы построения ИИС

Под ИИС понимаются информационные методологии, использующие понятия теоретической иммунологии для решения различных прикладных задач.

Процессы, происходящие при обработке информации естественными системами, и принципы их функционирования поражают своей эффективностью, экономичностью и быстродействием [1]. Прежде всего вызывает интерес способность данных систем решать многомерные задачи огромной вычислительной сложности в реальном времени. При обработке информации молекулами белков большое значение имеют пространственная конфигурация белков, принципы самосборки белков, их комплексов и сетей. Принципы действия механизмов репарации, т. е. исправления ошибок в процессе функционирования иммунной системы, позволяют строить совершенно новые алгоритмы распознавания образов на основе искусственных иммунных сетей.

Разработка высокоэффективных ИИС находится на начальном этапе, поэтому они недостаточно распространены, но имеют огромный потенциал развития. Прежде всего это связано с тем, что до конца неизвестны все механизмы функционирования естественных иммунных систем. В микробиологии, биохимии, биофизике в последнее время получены новые весомые результаты по данным вопросам, тем не менее многое еще не изучено. Еще одной из трудностей в создании ИИС является междисциплинарный характер исследований, когда решение проблемы находится на стыке очень многих наук, что, в свою очередь, сильно тормозит развитие этой области.

В ИИС используется идея взаимодействия между белками иммунной системы человека и чужеродными антигенами, т. е. возможность молекулярного узнавания посредством определения минимальной энергии связи между формальными пептидами (белками) и решения задачи распознавания образов. ИИС могут обучаться новой информации, обладают памятью и способностью к распознаванию образов. Иммунные сети строятся на основе полученной информации в виде временных рядов.

Математическая основа подхода ИИС заключается во введении понятия формального пептида (ФП) как математической абстракции свободной энергии белковой молекулы от ее пространственной формы, описанной в алгебре кватернионов. А.О. Тараканов предложил математическую модель формального пептида (простейшего белка, который содержит до 50 аминокислот) [2].

Формальным пептидом называют упорядоченную пятерку

$$P = \langle n, U, Q, V, v \rangle,$$

которая включает следующие компоненты:

- количество звеньев  $n > 0$ ;
- множество торсионных углов  $U = \{\varphi_k, \psi_k\}, k = 1, \dots, n$ , где  $-\pi \leq \varphi_k \leq \pi, -\pi \leq \psi_k \leq \pi$ ;
- множество единичных кватернионов  $Q = \{Q_0, Q_k\}$ , где кватернионы  $Q_k = Q_k(\varphi_k, \psi_k)$  и результирующий кватернион ФП  $Q_0$  определяются как их произведение:  $Q_0 = Q_1 Q_2 \dots Q_n$ ;

– множество коэффициентов  $V = \{v_{ij}\}, i = 1, 2, 3, 4, j \geq i$  ;  
– функцию  $v$  (без индекса), определенную на элементах результирующего кватерниона  $Q_0$  квадратичной формой  $v = -\sum_{j \geq i} v_{ij} q_i q_j$  .

На сегодняшний день получено достаточно результатов по применению ИИС [3–5], однако не до конца исследованы вопросы выбора структуры иммунной сети. Включение в модель признаков, мало влияющих на выходной параметр, затрудняет ее использование, ведет к избыточности информации и возрастанию вычислительной сложности решаемых задач. Все это обуславливает возникновение ситуации, когда прогнозирование по избыточной модели менее эффективно, чем по модели из оптимального числа признаков, обладающих максимальной информативностью. Поэтому при разработке интеллектуальной информационной технологии по анализу и прогнозированию многомерных данных на основе ИИС в реальном масштабе времени особенно актуальной является задача оптимизации иммунной сети. Успешное развитие ИИС напрямую зависит от создания алгоритмов, позволяющих обеспечить наиболее эффективное обучение иммунной сети за минимально короткое время. Применяются методы предварительной обработки данных и факторный анализ (метод главных компонент) [6] с целью извлечения наибольшей информации из исходных данных путем устранения избыточности.

Еще одна серьезная проблема заключается в оценке энергетических погрешностей, которые неизбежно возникают при обработке многомерной совокупности данных иммунными сетями и решении задачи распознавания образов на основе ИИС. Погрешности энергетических оценок возникают за счет неполноты данных, их корреляции и т. д. и существенно влияют на достоверность прогноза. Особенно это актуально для пептидов, имеющих приблизительно одинаковые параметры и находящихся на границах классов. Из-за энергетических ошибок они могут быть отнесены не в свой класс. Проблема сильно усложняется по мере увеличения размерности системы [7].

Число входных и выходных элементов ИИС определяется условиями рассматриваемой задачи, зависит от количества входных признаков и классов.

## 2. Постановка задачи исследования и алгоритм ее решения

Задача исследования заключается в разработке ИСУДО и осуществлении комплексной оценки знаний студентов на основе биологического подхода ИИС с целью оперативного управления процессом обучения в реальном масштабе времени и получения качественного индивидуального образования в среде Интернет.

Входными характеристиками ИИС являются временные ряды, составленные из индивидуальных признаков каждого обучающегося, которые определяются тестированием по изучаемой дисциплине, психологами и т. д. Вся необходимая информация содержится в базах данных (БД) и базах знаний (БЗ). Для построения иммунной сети в качестве признаков берутся поля таблиц БД.

На выходе ИИС имеем комплексную оценку знаний, группировку обучающихся по классам и прогноз качества полученного образования. Классы определяются экспертами и соответствуют определенным знаниям, практическим навыкам, творческим способностям, логическому мышлению и т. д. В зависимости от принадлежности обучающегося к определенному классу можно оценить его интеллектуальный потенциал и в соответствии с ним оперативно предоставить индивидуальную программу обучения.

Несомненным достоинством подхода ИИС к решению задачи оценки знаний является возможность комплексного решения данной проблемы [8]. Количество индивидуальных признаков, характеризующих каждого обучающегося, может быть достаточно большим (100–200). Это позволяет осуществить индивидуальный подход к каждому конкретному обучающему с целью лучшего усвоения пройденного материала, развития творческих способностей, присущих именно данному психологическому типу личности.

По индивидуальным признакам, характеризующим обучающихся, для определенного класса экспертами формируются матрицы эталонов:

$$A_1, A_2, A_3, \dots, n,$$

где  $n$  – количество классов.

После сингулярного разложения данных матриц получаем правые и левые сингулярные векторы  $\{x_1, y_1\}$ ,  $\{x_2, y_2\}$ , ... эталонных матриц. Затем формируется множество матриц, рассматриваемых как образы:

$$B_1, B_2, B_3, B_4, \dots, m,$$

где  $m$  – количество обучающихся.

Согласно подходу ИИС энергия связи между формальными пептидами может быть представлена в виде

$$W_1 = -x_1^T B y_1, W_2 = -x_2^T B y_2, W_3 = -x_3^T B y_3, W_4 = -x_4^T B y_4, \dots,$$

где  $T$  – символ транспонирования.

С помощью сингулярных разложений исходных матриц определяются энергии связи. Известно, что нативная (функциональная) укладка белковой цепи соответствует минимуму энергии связи, поэтому минимальное значение энергии связи определяет класс, к которому принадлежит данный образ:

$$k : W_k = \min \{W_1, W_2, W_3, W_4, \dots\}.$$

Структура ИСУДО с использованием иммуносетевого моделирования основана на модульном подходе. Каждый из этапов данной технологии представляет собой отдельный модуль.

Разработка ИСУДО на основе ИИС подразумевает реализацию следующих подсистем [8, 9]:

– *информационной*, осуществляющей разработку методов и средств хранения информации, разработку баз данных и знаний. Включает поисковые системы, электронные учебники по различным дисциплинам, электронные библиотеки, справочки, каталоги и т. д.;

– *интеллектуальной*, осуществляющей обучение иммунной сети, обработку многомерных данных в режиме реального времени на основе ИИС. Применение алгоритма оценок энергий связи на основе свойств гомологичных пептидов позволяет уменьшить ошибку обобщения при функционировании интеллектуальной системы и производить обучение студентов в соответствии с их индивидуальными особенностями;

– *обучающей*, реализующей методы, средства и формы подачи обучающей информации, адаптированной к конкретному пользователю с учетом его индивидуальных характеристик. Составляется график выполнения объема требуемых работ и сроки его реализации;

– *контролирующей*, предназначенной для комплексной оценки знаний обучающегося с целью оперативной корректировки программы и процесса обучения.

Ниже приведен алгоритм реализации ИСУДО в среде Интернет на основе ИИС [7]:

*Шаг 1.* Тестирование обучающихся по изучаемым дисциплинам, по различным методикам у психологов, специалистов с целью получения индивидуальных признаков.

*Шаг 2.* Создание баз данных на основе индивидуальных признаков обучающихся.

*Шаг 3.* Создание с помощью экспертов баз знаний.

*Шаг 4.* Нормирование и центрирование индивидуальных признаков обучающихся.

*Шаг 5.* Отбор информативных индивидуальных признаков обучающихся с помощью факторного анализа на основе вращения собственного вектора (метода главных компонент). Используется стандартный пакет прикладных программ SPSS 16.0.

*Шаг 6.* Построение оптимальной структуры иммунной сети по весовым коэффициентам информативных индивидуальных признаков. Редукция малоинформативных признаков.

*Шаг 7.* Классификация обучающихся по уровням знаний, практическим навыкам, творческим способностям, логическому мышлению на основе мнений экспертов.

*Шаг 8.* Формирование матриц эталонов для каждого класса по информативным индивидуальным признакам обучающихся, сворачивание их в квадратные матрицы для улучшения специфичности узнавания, сингулярное разложение данных матриц (используется MATLAB) и

определение правых и левых сингулярных векторов. Матрицы эталонов рассматриваются как антигены.

*Шаг 9.* Обучение ИИС по эталонам с учителем.

*Шаг 10.* Формирование матриц образов по информативным индивидуальным признакам обучающихся, сворачивание их в квадратные матрицы для улучшения специфичности узнавания. Матрицы образов рассматриваются как антитела.

*Шаг 11.* Определение минимальной энергии связи (энергии нативной структуры) между формальными пептидами – антигенами и антителами и решение задачи распознавания образов.

*Шаг 12.* Оценка энергетической погрешности ИИС на основе свойств гомологичных белков и расчет Z-факторов [10].

*Шаг 13.* Расчет коэффициентов риска прогнозирования.

*Шаг 14.* Комплексная оценка знаний, группировка обучающихся по классам, прогноз качества получаемого образования каждым обучающимся и оперативное управление процессом дистанционного обучения в среде Интернет.

Достоинством данного алгоритма является оценка энергетических погрешностей на основе свойств гомологичных белков при решении задачи распознавания образов, позволяющая исключить ошибочное попадание пептидов не в свой класс в случае, если они имеют схожие характеристики. Нативная структура белковой цепи, соответствующая минимуму энергии связи, является для каждого класса определенной и позволяет оценить принадлежность гомологов к какому-либо классу решений. Особенно это свойство ценно для образов, которые находятся на границах классов. Например, обучающимся, которые имеют схожие индивидуальные признаки, но относятся к разным классам, могут быть предложены ошибочные траектории обучения. Предложенная технология существенно уменьшает погрешности энергетических оценок, повышает достоверность прогноза и качество обучения.

### **3. Описание технологии обработки информации иммунными сетями в виде формального представления нетерминального алфавита**

Обработка многомерных данных ИИС возможна путем реализации различных технологических цепочек [7]. Условия функционирования ИСУДО на основе ИИС существенно влияют на состав и последовательность технологических операций, которые необходимы для успешной работы.

Технологические операции при обработке многомерной информации ИСУДО можно разделить на предварительные, основные, вспомогательные.

*Предварительные операции* включают в себя создание БД на основе индивидуальных признаков обучающихся, нормирование входных параметров, удаление пробелов, операции по считыванию информации из БД и т. д.

*Основные операции* заключаются в создании иммунной сети и ее обучении.

*Вспомогательные операции* состоят из выделения информативных индивидуальных признаков обучающихся и снижения размерности анализируемого пространства признаков, создания оптимальной структуры иммунной сети, тестирования, оценки энергетических погрешностей и т. д.

Ниже предлагается описание технологии обработки информации иммунной системой в виде формального представления нетерминального алфавита или словаря. Конкретные операции описываются с помощью определенных букв.

Нетерминальный алфавит (словарь) состоит из следующих подгрупп:

<P> – <Предварительные операции> – <Операции сбора данных, создание баз данных, считывания, преобработки данных>

<O> – <Основные операции> – <Создание, чтение иммунной сети, обучение сети>

<B> – <Вспомогательные операции> – <Тестирование, упрощение и создание оптимальной структуры иммунной сети, определение весовых коэффициентов входных параметров, оценка энергетических погрешностей, сохранение сети >

<R> – <Преобработка> – <Изменение структуры базы данных, преобразование нечисловых полей, поиск и заполнение пробелов, нормирование>

<L> – <Обучение иммунной сети> – <Обучение сети>

<Y> – <Упрощение структуры иммунной сети> – <Упрощение структуры сети, сокращение малоинформативных признаков, создание оптимальной структуры сети>

<N> – <Нормирование> – <Преобразование данных для приведения значений к определенному диапазону>

<T> – <Тестирование> – <Тестирование по примеру и по выборке>

После описания подгрупп рассматриваются конкретные технологические операции, которые входят в состав подгрупп для того, чтобы с наименьшими затратами ресурсов выбирать определенную технологическую последовательность в зависимости от требований, необходимых для эффективного решения поставленной задачи.

Технологические операции представляются буквами терминального алфавита (словаря), например:

< $r_d$ > – <считывание баз данных> – <открытие файла данных>

< $r_p$ > – <открытие сети> – <чтение сохраненной сети>

< $m$ > – <создание сети> – <выбор структуры сети, выбор входных параметров (индивидуальных признаков обучающихся)>

< $i$ > – <определение значимости входных параметров> – <расчет весовых коэффициентов информативных индивидуальных признаков обучающихся, показывающих их значимость при определении выходного параметра>

< $y_i$ > – <удаление входных параметров> – <редукция малоинформативных индивидуальных признаков обучающихся>

< $l_y$ > – <обучение иммунной сети> – <обучение сети с учителем>

< $l_v$ > – <самообучение сети> – <обучение сети без учителя>

< $f$ > – <преобразование полей базы данных> – <преобразование нечисловых полей базы данных>

< $n_d$ > – <нормирование данных> – <преобразование данных таким образом, чтобы математическое ожидание было равно нулю, а дисперсия – единице>

Приведем пример слова, принадлежащего данному языку:

< $r_d f n_d r_p i r_d l_y y_i i$ > – <Открыли базу данных, выполнили преобразование полей и нормирование, открыли сеть, определили значимость параметров, открыли базу данных, обучили сеть с учителем, провели редукцию входных параметров, определили значимость параметров>.

С помощью экспертов и разработанного формального языка описания технологии обработки данных иммунными сетями формируются базы знаний, которые позволяют систематизировать процесс с единых теоретических позиций и используются при выборе оптимальной технологической цепочки и автоматизированном написании программ.

## Заключение

Особенностями предложенной интеллектуальной технологии на основе иммуносетевого моделирования и ИСУДО являются:

– способность системы глубоко анализировать скрытые (латентные) взаимодействия между индивидуальными признаками и основополагающие факторы, влияющие на них;

– сокращение времени на обучение иммунной сети за счет построения оптимальной структуры и редукции малоинформативных индивидуальных признаков, несущих существенные погрешности;

– уменьшение погрешностей энергетических оценок (ошибок обобщения), повышение достоверности прогноза, оперативная корректировка процесса получения знаний и обеспечение качественного индивидуального дистанционного обучения в среде Интернет.

На разработанное программное обеспечение получены два авторских свидетельства о государственной регистрации объекта интеллектуальной собственности, зарегистрированные в Комитете по правам интеллектуальной собственности Министерства юстиции Республики Казахстан.

Результаты исследования и разработок, включая авторские инструментальные средства, используются на лекционных, практических и лабораторных занятиях по дисциплинам «Искусственные иммунные системы», «Новые информационные технологии», «Экспертные и интеллектуальные системы» в учебном процессе Центрально-Азиатского университета на кафедре «Вычислительная техника и информационные технологии».

### Список литературы

1. Молекулярная биология клетки. В 3 т. / Б. Альбертс [и др.]. – М. : Мир, 1986.
2. Tarakanov, A.O. Formal peptide as a basic of agent of immune networks: from natural prototype to mathematical theory and applications / A.O. Tarakanov // Proc. of the I Int. workshop of central and Eastern Europe on Multi-Agent Systems (CEEMAS'99). – St. Petersburg, 1999. – P. 281–292.
3. Искусственные иммунные системы и их применение : сборник статей / под ред. В. Дасгупты ; пер. с англ. А.А. Романюхи, С.Г. Руднева под ред. А.А. Романюхи. – М. : Физматлит, 2006. – 344 с.
4. De Castro, L.M. Artificial immune systems: A new computational intelligence approach / L.M. De Castro, J.I. Timmis. – London : Springer-Verlag, 2003. – 357 p.
5. An Overview of Artificial Immune Systems / J.I. Timmis [et al.] // Computation in Cells and Tissues: Perspectives and Tools for Thought, Natural Computation Series. – Springer, 2004. – P. 51–86.
6. Иберла, К. Факторный анализ / К. Иберла. – М. : Статистика, 1980. – 398 с.
7. Самигулина, Г.А. Разработка интеллектуальных экспертных систем управления на основе искусственных иммунных систем / Г.А. Самигулина. – Алматы : ИПИУ МОН РК, 2010. – 252 с.
8. Самигулина, Г.А. Интеллектуальная экспертная система дистанционного обучения на основе искусственных иммунных систем / Г.А. Самигулина // Информационные технологии моделирования и управления. – 2007. – № 9 (43). – С. 1019–1024.
9. Самигулина, Г.А. Разработка дистанционной образовательной технологии на основе искусственной иммунной системы / Г.А. Самигулина // Открытое образование. – 2008. – № 6 (71). – С. 52–58.
10. Самигулина, Г.А. Разработка интеллектуальных экспертных систем прогнозирования и управления на основе искусственных иммунных систем / Г.А. Самигулина // Проблемы информатики. – 2010. – № 1. – С. 15–22.

Поступила 14.04.10

*Институт проблем информатики и управления  
Министерства образования и науки Республики Казахстан,  
Алматы, ул. Пушкина, 125  
e-mail: galinasamigulina@mail.ru*

**G.A. Samigulina**

### MODELING THE REMOTE EDUCATION TECHNOLOGY USING ARTIFICIAL IMMUNE NETWORKS

An intelligent educational technology based on immune network modeling which allows to analyze huge data files in real time and to adapt an educational process to a specific user is developed. The advantages of the suggested intelligent technology include the possibility of operative control of the process and the quality of training, an individual approach to each trainee, improved reliability of forecasting the training results by reducing the generalization errors of the artificial immune network.