

Г. А. САМИГУЛИНА, З. И. САМИГУЛИНА

**РАЗРАБОТКА ПАКЕТА ПРИКЛАДНЫХ ПРОГРАММ
ОБРАБОТКИ, АНАЛИЗА И ПРОГНОЗА ДАННЫХ
ИСКУССТВЕННЫМИ ИММУННЫМИ СЕТЯМИ**

В настоящее время проблема моделирования биологических принципов обработки информации является одной из основных задач искусственного интеллекта. Разработка нетрадиционных биологических подходов при решении широкого круга проблем, начиная от защиты информации и заканчивая проблемами прогнозирования динамических свойств нелинейных объектов управления в реальном масштабе времени, являются весьма актуальными задачами во всем мире. Последние научные достижения в биофизике, молекулярной биологии и биохимии показали, что процессы, происходящие при обработке информации естественными системами, и принципы их функционирования имеют единую информационную основу – генетический код и универсальный алфавит, состоящий из молекул белков. Белки играют огромную роль в природе. Необходим тщательный анализ и осмысление процессов обработки информации белками в естественных системах, что позволит глубже понять механизмы функционирования информационных систем и покажет дальнейшие пути развития данной области науки.

Белки – это биополимеры сложного строения [1], макромолекулы которых представляют собой остатки аминокислот или полипептидных цепей, соединенных между собой пептидной связью. Именно пространственная структура белка определяет химические, биологические и функциональные свойства белка. Свойства белка могут сильно изменяться при замене хотя бы одной аминокислоты. Это связано с тем, что изменение конфигурации пептидных цепей ведет к другим условиям

образования пространственной структуры белка, которая определяет все его функции в организме. Решающая роль в белках принадлежит не отдельным аминокислотным остаткам, а их сочетаниям. В процессе эволюции сформировались механизмы отбора белковых структур, которые способны сворачиваться в определенные трехмерные нативные структуры, чья энергия существенно ниже, чем энергия альтернативных структур. Основной чертой белковых последовательностей, определяющей их физические свойства, является повышенная стабильность нативной (функциональной) структуры и существование большой щели между энергией нативной структуры и минимальной энергией неверно свернутых структур.

Приведенные свойства естественных белковых структур служат основой биологического подхода искусственных иммунных систем (ИИС) [2].

Основная идея заключается во взаимодействии между белками иммунной системы человека и чужеродными антигенами, т. е. в возможности произвольного связывания (так называемого молекулярного узнавания) посредством определения минимальной энергии связи между формальными пептидами, т.е. определения нативной структуры.

Постановка задачи: необходимо разработать пакет прикладных программ (ППП) обработки, анализа и прогноза многомерных данных в реальном масштабе времени на основе интеллектуальной технологии ИИС.

Структура интеллектуальной технологии обработки информации на основе иммунных сетей основана на модульном подходе. Каждый из этапов данной технологии представляет собой отдельный модуль.

Разработанное программное обеспечение PERTID предназначено для реализации поставленной задачи с помощью искусственных иммунных сетей. Такой пакет программ реализован на языке программирования DELPHI 7.0.

В состав предложенного программного продукта входит девять взаимосвязанных подпрограмм, выполняющих определенные задачи.

Входными характеристиками ИИС являются временные ряды, составленные из различных параметров, описывающих рассматриваемую систему. Вся необходимая информация заносится в базы данных (БД) и базы знаний (БЗ). Для построения иммунной сети в качестве признаков берутся поля таблиц БД.

Структурная схема пакета прикладных программ PEPTID обработки информации искусственной иммунной системой приведена на рис. 1.

Рис. 1. Структурная схема пакета прикладных программ PEPTID

Рис. 2. Пример нормировки входных признаков

Предложенный пакет программ состоит из следующих подпрограмм: NORM, INF_PRIZ, OPT_NET, CLAS_EXP, LEARN, MIN_ENERG, ENERG_ERRORS, K_PRIDIC, GRAFIC.

После считывания исходной информации из баз данных и баз знаний в подпрограмме NORM осуществляется нормировка входных параметров.

Предварительная обработка данных является важным этапом любого анализа данных. Наиболее распространено нормирование или масштабирование данных. Основной целью нормирования является приведение всех данных к сопоставимому виду. Так как обычно параметры, характеризующие состояние рассматриваемой системы, измеряются в разных единицах, то результат существенно зависит от выбора масштабов измерения.

Подпрограмма преобразует элементы каждого вектора таким образом, чтобы математическое ожидание было равно нулю, а дисперсия – единице.

Разработанный удобный интерфейс (рис. 2) позволяет графически наглядно отобразить результаты нормирования. Из примера видно, что вся совокупность данных после нормирования сместилась относительно начала координат.

Подпрограмма INF_PRIZ предназначена для выделения информативных признаков на основе факторного анализа и метода главных компонент.

Суть факторного анализа заключается в том, чтобы на основе исследования корреляционных взаимосвязей признаков находить причины, которые определяют эти взаимосвязи. Метод главных компонент применяется в целях извлечения наибольшей информации из исходных данных путем устранения избыточности.

Подпрограмма решает задачу снижения размерности анализируемого пространства признаков и отбор наиболее информативных признаков с помощью метода главных компонент на основе вращения собственного вектора.

Подпрограмма OPT_NET осуществляет построение оптимальной структуры иммунной сети по весовым коэффициентам информативных признаков и редукцию малоинформативных признаков.

Способность ИИС обрабатывать большой объем информации неизбежно приводит к увеличению времени на обучение иммунной сети. В результате возможна такая ситуация, когда время, необходимое на обучение иммунной сети, превосходит все разумные для поставленной задачи временные ресурсы.

Включение в модель признаков, мало влияющих на выходной параметр, затрудняет ее использование, ведет к избыточности информации и возрастанию вычислительных ресурсов. Все это приводит к ситуации, когда прогнозирование по избыточной модели менее эффективно, чем по модели из оптимального числа признаков, обладающих максимальной

информативностью. Подпрограмма CLAS_EXP предназначена для реализации алгоритмов классификации решений, предложенных экспертами. Подпрограмма LEARN осуществляет обучение иммунной сети с учителем по выбранным сформированным экспертом эталонам для соответствующих классов. Подпрограмма MIN_ENERG предназначена для решения задачи распознавания образов на основе определения минимальной энергии связи между формальными пептидами. Подпрограмма ENERG_ERRORS служит для оценки энергетических ошибок при решении задачи распознавания образов и обработки многомерной информации на основе анализа свойств гомологичных пептидов.

Определение нативной укладки белковой цепи, соответствующей минимальному значению энергии связи между пептидами, затруднено из-за различных погрешностей (неполнота данных, ошибки измерения, корреляция данных).

В подпрограмме K_PREDIC осуществляется расчет коэффициентов риска прогнозирования на основе Z-факторов.

Заключительная подпрограмма GRAFIC позволяет сгруппировать данные для визуального представления и наглядно выделить зоны неоднозначного решения для проведения дополнительных процедур.

На основе описанного пакета прикладных программ созданы следующие приложения:

интеллектуальная экспертная система анализа и прогнозирования асимптотической устойчивости в среднем квадратическом стохастической системы управления [3];

интеллектуальная экспертная система анализа и прогнозирования динамических свойств интервально-заданной системы управления [4];

интеллектуальная экспертная система дистанционного обучения в среде Internet на основе ИИС [5];

биометрическая система контроля доступа на основе аутентификации рукописной подписи [6];

ИИС анализа состояния чумной триады по характеристикам штаммов чумного микроба и прогнозирование начала эпизоотического процесса [7].

Результаты проведенных экспериментов показывают, что предложенная интеллектуальная технология и разработанное программное обеспечение в целом обладают высокой эффективностью и точностью распознавания образов с использованием искусственных иммунных сетей.

ЛИТЕРАТУРА

1. *Finkelstein A.V., Gutin A.M., Badretdinov A.Y.* // Proteins. 1995. V. 23. P. 151-162.
2. *Tarakanov A.O.* Formal peptide as a basic of agent of immune networks: from natural prototype to mathematical theory and applications // Proceeding of the I Int. workshop of central and Eastern Europe on Multi - Agent Systems, 1999. P. 186-188.
3. *Самигулина Г.А.* Интеллектуальная система прогнозирования асимптотической устойчивости в среднем квадратическом стохастической системы управления // Труды IX международной конференции «Интеллектуальные системы и компьютерные науки». М.: МГУ, 2006. Т. II. С. 265-269, <http://intsys.msu.ru/science/conference/>.
4. *Самигулина Г.А.* Исследование интервальных систем управления с применением подхода квазирасщепления // Математический журнал. Алматы, 2004. Т. 4, № 1(11). С. 155-160.
5. *Самигулина Г.А., Чебейко С.В.* Интеллектуальная информационная технология виртуального обучения на основе искусственных иммунных систем // Вестник КазНТУ. Алматы, 2005. № 1(45). С. 13-18.

6. *Самигулина Г.А.* Биометрическая система контроля доступа по рукописной подписи на основе искусственных иммунных систем // Вестник КазНТУ. Алматы, 2005. № 5. С. 142-149.
7. *Samigulina G.A., Chebeiko S.V.* Technology of elimination errors the energy estimations of Artificial Immune Systems of the forecasting plague // Proceeding on the sixth international conference on Computational Intelligence and Natural Computation (CINC 2003), Cary, North Carolina, USA. 2003. P. 1693-1696.