

УДК 681.5

Г.А.САМИГУЛИНА

## ТЕХНОЛОГИЯ ПОСТРОЕНИЯ ОПТИМАЛЬНОЙ СТРУКТУРЫ ИСКУССТВЕННОЙ ИММУННОЙ СЕТИ

Обработка многомерных данных на основе искусственных иммунных систем ведет к ряду серьезных проблем влияющих на достоверность прогноза. Работа посвящена построению оптимальной структуры искусственной иммунной сети с целью уменьшения времени на ее обучение.

Стремительное развитие современных информационных технологий в области Интеллектуальных Систем (ИС) нового поколения способствует разработке новых нетрадиционных биологических подходов при обработке многомерных данных. Особенный интерес во всем мире уделяется Искусственным Иммунным Системам (ИИС).

Математические основы данного подхода представлены в [1]. Базовым элементом является пептид (белок). Биологическим прототипом является иммунная система и обработка информации молекулами белков.

Белки это биополимеры сложного строения [2], 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. Иммунная система человека в настоящее время изучена значительно лучше, чем мозг и обладает всеми свойствами искусственного интеллекта - памятью, способностью к обучению и принятию решений в незнакомой ситуации.

При разработке интеллектуальной информационной технологии по анализу и прогнозированию данных на основе ИИС в реальном масштабе времени особенно актуально решение задачи оптимизации иммунной сети.

Способность ИИС обрабатывать большой объем информации неизбежно приводит к увеличению времени на обучение иммунной сети. В результате возможна такая ситуация, что време-

мя необходимое на обучение иммунной сети превосходит все разумные для поставленной задачи временные ресурсы.

Таким образом, успешное развитие ИИС на прямую зависит от создания алгоритмов, которые бы позволили наиболее эффективное обучение иммунной сети за минимально короткое время.

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.

Постановка задачи формулируется следующим образом: необходимо разработать технологию построения оптимальной структуры искусственной иммунной сети по весовым коэффициентам выделенных информативных признаков с целью эффективного обучения сети за минимальное время.

В связи с данной задачей используются различные методы предварительной обработки исходных данных и факторного анализа для выделения основополагающих факторов, уменьшения анализируемого пространства признаков и снижения времени на обучение. Разработан алгоритм выделения оптимальных, т.е. наиболее информационно - ценных подмножеств информативных признаков из общего множества признаков и редукции размерности анализа данных. Факторный анализ [3] применяется с целью извлечения наибольшей информации из исходных данных путем устранения избыточности. Это эффективный многомерный метод, позволяющий представить исходную матрицу данных через линейные ком-

бинации ортогональных независимых векторов, когда полученное ортогональное пространство меньшей размерности несет полную информацию об исходной матрице данных. Входными характеристиками ИИС являются временные ряды, составленные из признаков описывающих рассматриваемую систему. Вся необходимая информация содержится в базах данных (БД). Для построения иммунной сети в качестве признаков берутся поля таблиц БД. Введем некоторые определения, необходимые для дальнейшего изложения.

**Определение 1.** Под Иммунной Сетью (ИС) в подходе ИИС будем понимать временные ряды информативных признаков, характеризующие рассматриваемую сложную систему и выполняющие определенные функции.

**Определение 2.** Под оптимальной структурой иммунной сети понимается сеть, построенная на основе весовых коэффициентов выделенных информативных признаков, наиболее полно характеризующих состояние рассматриваемой системы в зависимости от основополагающих факторов, влияющих на функционирование этой системы.

Обобщенный алгоритм построения оптимальной структуры иммунной сети состоит из восьми шагов.

Алгоритм:

Шаг 1. Создание баз данных

Шаг 2. Нормирование входных данных.

Шаг 3. Выделение информативных признаков с помощью метода главных компонент.

Шаг 4. Выделение и анализ основополагающих факторов на основе развития варимаксного подхода.

Шаг 5. Редукция малоинформационных признаков.

Шаг 6. Визуальное представление данных.

Шаг 7. Ранжирование информативных признаков в зависимости от весовых коэффициентов.

Шаг 8. Построение оптимальной структуры иммунной сети для дальнейшего решения задачи распознавания образов на основе ИИС.

Так как признаки, характеризующие состояние системы измеряются в разных единицах, то результат может существенно зависеть от выбора масштаба измерения. Поэтому необходима

предварительная обработка данных, т.е. переход к безмерным величинам с помощью нормирования [4].

Методы предварительной обработки многомерных данных предназначены для перевода данных в наиболее информативное для исследуемой задачи представление.

Пусть исходная совокупность данных записана в виде матрицы  $A = (a_{ij})$ ,  $i, j = 1, \dots, n$  размерности ( $m \times n$ ). Для этого преобразуем элементы каждого вектора таким образом, чтобы математическое ожидание было равно нулю, а дисперсия единице. Матрица  $A$  подвергается нормированию.

Новая матрица стандартизованных переменных  $X$  записывается из элементов:

$$x_{ij} = \frac{x_{ij} - m_j}{s_j}, \quad (1)$$

где  $m_j$  - среднее значение исходных элементов  $j$ -го вектора,  $s_j$  - стандартное отклонение исходных элементов  $j$ -го вектора, которое вычисляется по

$$\text{формуле: } s_j = \left( \frac{1}{N-1} \sum_{i=1}^n (x_{ij} - m_j)^2 \right)^{\frac{1}{2}}, \quad (2)$$

где  $n$  - общее число измерений.

Применяя формулы (1), (2), вычислим матрицу стандартизованных переменных  $X$ .

На рис.1. приведен пример нормирования входных данных, из которого видно, что вся совокупность данных после преобразования сместилась относительно начала координат.

Программное обеспечение написано на языке программирования DELPHI 7.

Задача снижения размерности анализируемого признакового пространства и отбор наиболее информативных признаков решена с помощью метода главных компонент на основе вращения собственного вектора. Данный метод [3] заключается в том, чтобы повернуть все оси  $n$ -мерной совокупности данных таким образом, чтобы первая ось соответствовала направлению наибольшей дисперсии в данных, а каждая последующая представляла максимум остаточной дисперсии. Таким образом, находятся ортогональные оси, которые представляют направления максимума дисперсии, при этом сохраняется исходная структура данных.

Определим базисное пространство  $R$ . Совокупность осей координат можно повернуть на угол  $\theta$ , изменяя ориентацию совокупности точек

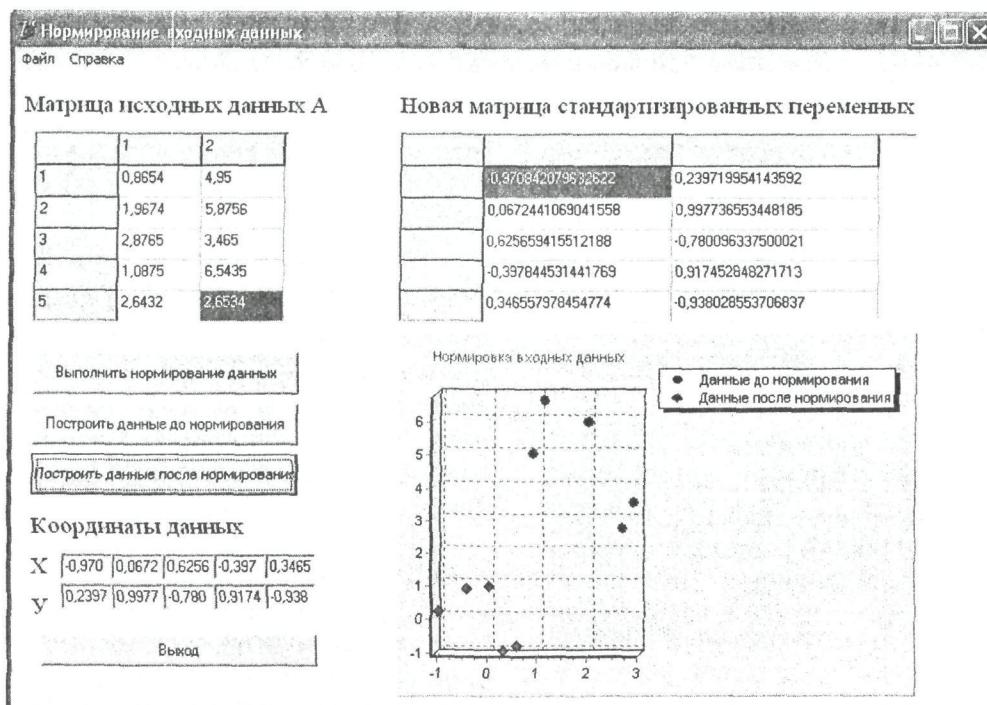


Рис.1. Пример нормирования входных данных

относительно осей.

Для того, чтобы повернуть против часовой стрелки оси координат на угол  $\theta$  необходимо умножить матрицу преобразования  $R^T$  на исходные данные  $A$ . Тогда координаты новой матрицы  $B$  в матричной форме будут записаны в виде:

$$B = R^T A \quad (3)$$

Матрица преобразования  $R^T$  в двумерном пространстве имеет вид:

$$R^T = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix}. \quad (4)$$

Для того, чтобы найти ортогональные оси, представляющие направления максимума дисперсии необходимо рассчитать корреляционную матрицу данных  $C$ , которая после поворотного преобразования осей становится диагональной:

$$C = \frac{1}{N-1} (X^T X), \quad (5)$$

где  $N$  - число столбцов в  $X$ .

Пусть  $Y = B^T$ ,  $X = A^T$ , тогда получим:

$$Y = XR, \quad Y^T = R^T X^T. \quad (6)$$

Необходимо найти матрицу преобразования  $R^T$  такую, чтобы, применив ее к матрице  $X$ ,

получить новую систему координат  $Y$ , которая удовлетворяет выражению:

$$Y^T Y = R^T X^T X R = R^T C R = \Lambda, \text{ где } \Lambda \text{ - диагональная матрица.}$$

Необходимо чтобы выполнялось условие:

$$CR = \lambda R, \quad (7)$$

Тогда получим:

$$(C - \lambda I)R = 0, \quad (8)$$

где  $\lambda$  - скалярные, диагональные элементы в  $\Lambda$ .

Задача будет иметь решение при

$$|C - \lambda I| = 0. \quad (9)$$

После нахождения решения для  $\lambda$  найдем матрицу преобразования  $R$ . Затем мы можем повернуть исходные оси таким образом, чтобы дисперсии ассоциировались с новыми осями, вычислить координаты данных в новой системе координат и начертить данные.

Чтобы лучше понять механизмы функционирования сложной системы, необходим тщательный анализ собственных векторов и их собственных значений (нагрузок), которые показывают относительный вклад признаков в каждом собственном векторе. Фактически они показывают, сколько процентов информации содержит тот или иной признак в соответствующем собственном

векторе. Информация содержится в одном или более собственных векторах, которые имеют наибольшую дисперсию и поэтому погрешности, содержащиеся в собственных векторах с наименьшими собственными числами, могут быть отброшены.

После анализа признаков необходимо произвести редукцию тех признаков, которые лежат ближе к началу координат и являются наименее информативными.

Предложенные процедуры использовались при создании следующих приложений на основе ИИС:

- интеллектуальной экспертной системе анализа и прогнозирования асимптотической устойчивости в среднем квадратическом стохастической системы управления [5];

- интеллектуальной экспертной системе анализа и прогнозирования динамических свойств интервально-заданной системы управления [6];

- интеллектуальной экспертной системе дистанционного обучения в среде Internet на основе ИИС;

- биометрической системе контроля доступа на основе аутентификации рукописной подписи;

- ИИС анализа состояния чумной триады по характеристикам штаммов чумного микробы и прогнозирования начала эпизоотического процесса [7].

#### ЛИТЕРАТУРА

1. Сб.-Tarakanov A.O. Formal peptide as a basic of agent of immune networks: from natural prototype to mathematical theory and applications. Proc. of the 1<sup>st</sup>. Int. Workshop of Central and Eastern Europe on Multi-Agent Systems, 1999, P.46-51.

2. Сб.-Finkelstein A.V., Gutin A.M., Badretdinov A.Y.// Proteins. 1995. V.23. P.151-162.

3. Кн.-К. Иберла. Факторный анализ. М.: Статистика, 1980, 380с.

4. Сб.-Самигулина Г.А., Чебейко С.В. Выделение информативных признаков штаммов чумного микробы для прогнозирования эпизоотий чумы. //Проблемы нелинейного анализа в инженерных системах. Казань, КГТУ, 2004, вып.1(20), том 10, с. 83-95.

5. Сб.-Самигулина Г.А. Интеллектуальная система прогнозирования асимптотической устойчивости в среднем квадратическом стохастической системы управления. //Труды IX Международной конференции Интеллектуальные системы и компьютерные науки, Москва, МГУ им.Ломоносова, 2006, том II, <http://intsys.msu.ru/science/conference/>.

6. Сб.-Самигулина Г.А. Исследование интервальных систем управления с применением подхода квазиразщепления. //Математический журнал. Алматы: Институт математики. 2004. Т.4, N1(11). с.155-160.

7. Сб.-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.p.1693-1696.

#### Резюме

Жасанды иммундық жүйелердің негізіндегі көпөлшемді мәліметтерді өндеу, болжамның дұрыстығына өсер ететін бірқатар маңызды мәселелерге өкеледі. Бұл жұмыс оқыту мерзімін азайту мақсатымен жасанды иммундық жүйенің онтайлы құрылымын құруға арналған.

#### Summary

The processing of the multidimensional data on the basis of the artificial immune systems result to the number of the serious problems which influence on the veracity of the forecast. The article is dedicated to construction of optimal structure of the artificial immune network with the purpose of reduction the time on its training.

*Институт проблем информатики  
и управления МОН РК, г.Алматы*

*Поступила 6.12.07*