

ПОДХОДЫ К РЕШЕНИЮ ЗАДАЧИ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ НА ОСНОВЕ ИСКУССТВЕННЫХ ИММУННЫХ СИСТЕМ

М. Х. Дадаханов

*Наманганский государственный университет,
г. Наманган, Узбекистан*

Рассматриваются подходы к решению задачи классификации на основе искусственных иммунных систем. 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 актуальными задачами. В работе рассмотрены обобщенные условия сходимости иммунных алгоритмов.

1. Введение

Биологическая иммунная система (ИС) представляет собой сложную распределенную адаптивную систему интеллектуальной обработки информации. ИС защищает организм от инородных вирусов и инфекций. Иммунная система способна к обучению, обладает памятью, решает задачи поиска и классификации. ИС обрабатывает огромные объемы информации, поэтому алгоритмы, созданные природой, оказались эффективными и в математических задачах поиска и распознавания.

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, то есть в возможности произвольного связывания (так называемого молекулярного узнавания) посредством определения минимальной энергии связи между формальными пептидами.

Новая вычислительная процедура, называемая искусственные иммунные системы, основанная на принципах иммунной системы, обладает способностью обучаться новой информации, запоминать ранее полученную информацию и осуществлять распознавание образов и анализ данных на основе принципов биомолекулярного узнавания в высоко распределенной манере. Эти системы предлагают мощные и робастные возможности обработки больших массивов информации для решения сложных задач. Строгий математический базис ИИС основан на биологическом прототипе иммунной сети и понятиях формального протеина и формальной иммунной сети (ФИС).

Эти математические модели были названы формальной иммунной системой или иммунокомпьютерингом на основе свойств сингулярного разложения произвольных матриц.

Отдельные статьи по ИИС начали появляться в 1980-х годах, однако в отдельном направлении ИИС выделились только к середине 90-х годов с появлением работ Форреста, Дасгупты, Ханта и Кука. Первая книга про искусственные иммунные системы вышла в 1998 году под редакцией Дипанкара Дасгупты [3].

Чужеродные агенты, находясь в организме, производят молекулы, называемые антигенами. Большая часть антигеном может быть распознана специальными клетками – В-лимфоцитами, которые циркулируют в кровеносной и лимфатической системах в ожидании столкновения с антигенами. После того как антиген взаимодействует с

антителами В-лимфоцита, стимулируется процесс клонирования лимфоцита. Этот процесс называется клональным отбором. Процесс клонирования В-лимфоцитов в результате взаимодействия с антигенами называется иммунным ответом [4].

Предварительные операции включают в себя создание баз данных на основе статистических временных рядов, характеризующих рассматриваемую систему; создание баз знаний на основе мнений экспертов; нормирование входных параметров; удаление пробелов в таблицах данных; операции по считыванию информации из баз данных, баз знаний и т.д.

Основные операции заключаются в создании иммунной сети, обучении иммунной сети с учителем и т. д.

Вспомогательные операции состоят из таких как: выделение информационно - ценных признаков и снижение размерности анализируемого пространства признаков; создание оптимальной структуры иммунной сети; тестирование; оценка энергетических погрешностей ИИС при распознавании образов и т. д.

База знаний содержит сведения, которые отражают закономерности, существующие в рассматриваемой предметной области, позволяют выводить новые знания и прогнозировать потенциально возможные состояния исследуемой области; сведения о структуре и содержании базы данных; сведения по языку общения; метазнания, определяющие способы представления и переработки знаний. В базу знаний помещаются как общедоступные данные, так и знания эксперта в данной предметной области, вычислительные алгоритмы реализации процедур иммунокомпьютинга, результаты группировки и автоматической классификации, а также интерпретация результатов вычислений.

2. Иммунный алгоритм решения задач классификации

Нечеткие логические уравнения задачи классификации (типа если <вход>, то <выход>) вместе с функциями принадлежности нечетких термов позволяют принимать решение с использованием следующего алгоритма:

1. Фиксируются значения параметров состояния объекта:

$$X = [x_1, \dots, x_n].$$

2. Определяются значения функций принадлежности $\mu^j(x_i^*)$, при фиксированных значениях параметров x_i^* , $i = \overline{1, n}$.

3. Используя логические уравнения, вычисляются значения функций принадлежности $\mu^{r_j}(x_1^*, x_2^*, \dots, x_n^*)$ при векторе состояния $X^* = (x_1^*, x_2^*, \dots, x_n^*)$.

4. Определяется решение r_j^* , для которого:

$$\mu^{r_j^*}(x_1^*, x_2^*, \dots, x_n^*) = \max_{j=1, n} [\mu^{r_j}(x_1^*, x_2^*, \dots, x_n^*)].$$

Суть обучения состоит в подборе таких параметров функций принадлежности, которые минимизируют различие между результатами нечеткой аппроксимации и реальным поведением объекта.

Определим задачу классификации как отображение $f(X) \rightarrow \{1, \dots, c\}$ любого образа X в одно из целых чисел $1, \dots, c$, которые представляют классы.

Задача классификации может быть сформулирована следующим способом.

Дано:

- число классов c ;

- набор из m обучающих образов: X_1, X_2, \dots, X_m ;

- класс любого обучающего образа:

$$f(X_1) = c_1, \dots, f(X_m) = c_m;$$

- произвольный n -мерный вектор Z .

Найти:

Класс вектора Z : $f(Z) = ?$

Процесс обучения состоит из следующих этапов:

1. Сформировать обучающую матрицу $A = [X_1, \dots, X_m]^T$ размерности $m \times n$.

2. Вычислить максимальное сингулярное число s , а также левый и правый сингулярные векторы L и R обучающей матрицы по следующей итеративной (эволюционной) схеме:

$$L_{(0)} = [1, \dots, 1]^T,$$

$$R^T = L_{(k-1)}^T A, \quad R_{(k)} = R / |R|, \quad \text{где } |R| = \sqrt{r_1^2 + \dots + r_n^2},$$

$$L = AR_{(k)}, \quad L_{(k)} = L / |L|, \quad \text{где } |L| = \sqrt{l_1^2 + \dots + l_n^2},$$

$$s_{(k)} = L_{(k)}^T AR_{(k)}, \quad k = 1, 2, \dots,$$

до выполнения условия $|s_{(k)} - s_{(k-1)}| < \varepsilon$,

$$s = s_{(k)}, \quad L = L_{(k)}, \quad R = R_{(k)}.$$

3. Хранить сингулярное число s .

4. Хранить правый сингулярный вектор R (как «антитело-пробу»).

5. Для всякого $i=1, \dots, m$ хранить компоненту l_i левого сингулярного вектора L и класс c_i соответствующий обучающему образу X_i .

6. Для всякого n -мерного образа Z вычислить его энергию связи с R :

$$w(z) = Z^T R / s.$$

7. Выбрать l_i , которая имеет минимальное расстояние d с w :

$$d = \min_i |w - l_i|, \quad i=1, \dots, m,$$

и считать класс c_i искомым классом образа Z .

В общем виде один шаг работы иммунного алгоритма можно представить следующим образом [2]:

$$\begin{aligned} \forall i \in \{1, \dots, m\} : x_i' &= \text{mut}(\text{clon}(x_1, \dots, x_n)) \\ (x_1'', \dots, x_k'') &= \text{aging}(x_1, \dots, x_n, x_1', \dots, x_m') \quad (1) \\ (y_1, \dots, y_n) &= \text{sel}(x_1'', \dots, x_k'') \end{aligned}$$

где $(x_1, \dots, x_n) \in X^n$ – текущая популяция антител; (x_1', \dots, x_m') – популяция антител, возникающая в результате клонирования и мутации; (x_1'', \dots, x_k'') – антитела, которые удаляются из популяции; (y_1, \dots, y_n) – антитела, добавляемые в текущую популяцию.

Как следует из (1), ИА присущи следующие операторы: клонирование, мутация, старение и селекция. Рассмотрим подробнее эти операторы.

Оператор клонирования генерирует новое поколение копий антител в будущей популяции. Известны следующие основные операторы клонирования [3]: а) статический оператор клонирования, который просто копирует каждую В-клетку, производя переходную популяцию; б) пропорциональный оператор клонирования, который клонирует В-клетки

пропорционально их антигенной схожести; в) оператор вероятностного клонирования, в соответствии с которым В-клетки выбираются из текущего поколения в зависимости от вероятности клональной селекции.

Оператор мутации действует в зависимости от имеющейся популяции клонов, применяя к каждому антителу определенное количество одиночных мутаций, осуществляемых случайным образом. Можно выделить следующие способы мутации [4]:

1. Статическая мутация. Количество мутаций зависит от минимизируемой функции f , поэтому антитело в каждый момент времени будет подвергаться определенному числу мутаций.

2. Пропорциональная мутация. Количество мутаций антитела пропорционально соответствующему значению.

3. Обратно пропорциональная мутация. Количество мутаций антитела обратно пропорционально соответствующему значению.

4. Круглая мутация. Каждое антитело подвергается мутации круглого сочетания.

Оператор старения устраняет старые особи. Статический оператор старения использует возрастной параметр для максимального количества поколений антител, которым разрешено оставаться в популяции. Когда антитело старше, оно удаляется из системы, даже если оно может оказаться вполне пригодным на последующих итерациях.

При клональной экспансии клонированное антитело наследует возраст его родителя. После этапа мутации только те антитела, которые получили высшее значение аффинности, получают возраст, равный 0. Элитный вариант этого оператора получается путем взятия наилучших антител популяции в поколение с возрастом, равным 0.

Элитный вариант этого оператора получается взятием лучшего антитела из популяции в поколение.

Оператор селекции заменяет наихудшие антитела в популяции новыми случайными антителами.

3. Заключение

В работе рассмотрены обобщенные условия сходимости ИА в зависимости от используемых иммунных операторов. Анализ сходимости ИА основан на выполнении двух условий:

1) в результате мутации можно достичь оптимальное состояние из неоптимального за один шаг;

2) как только оптимальное состояние будет найдено, оно сохранится в популяции и не будет утеряно. Показано, что только операторы мутации и селекции, которые могут вносить изменения в антитела, способствует поиску оптимума.

Литература

1. Finkelstein A. V., Gutin A.M., Badretdinov A. Y // Proteins. - 1995. - V.23. - P.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.P.186-188.
3. Dasgupta D., Artificial Immune Systems and Their Applications, Springer-Verlag, 1998.
4. Gaber J., Bakhouya M., An Immune Inspired-based Optimization Algorithm: Application to the Traveling Salesman Problem, Université de Technologie de Belfort-Montbéliard, 2007.