

УДК 004.89



ОБРАБОТКА РЕЗУЛЬТАТОВ ТЕСТИРОВАНИЯ ОБУЧАЮЩИХСЯ С ИСПОЛЬЗОВАНИЕМ ИСКУССТВЕННЫХ ИММУННЫХ СИСТЕМ

Н.М. Кораблев¹, А.А. Фомичев²

¹ХНУРЭ, г. Харьков, Украина, korablev@kture.kharkov.ua

²ХНУРЭ, г. Харьков, Украина, alexandros_1985@mail.ru

В работе рассматривается решение задачи обработки результатов тестирования обучающихся на основе искусственных иммунных систем как задача классификации объектов. Спецификой работы используемого метода классификации обучающихся является использование аффинности как единого критерия определения близости объектов и классов.

ИСКУССТВЕННЫЕ ИММУННЫЕ СИСТЕМЫ, ПРИОРИТЕТНОЕ ПОСЛЕДОВАТЕЛЬНОЕ КЛОНИРОВАНИЕ, КОНКУРЕНТНО-ЦЕЛЕВОЙ ОТБОР КЛОНОВ, ОБЛАСТЬ ПОИСКА ЦЕЛЕВЫХ АНТИГЕН

Введение

В настоящее время в условиях информатизации общества большое значение приобретает компьютеризация сферы образования. В учебных заведениях широко распространяется компьютерное тестирование учащихся, исследуются возможности улучшения его качества. Задача обработки ответов, полученных в результате тестирования, может рассматриваться как задача классификации объектов. При тестировании под объектами классификации подразумеваются обучающиеся, которые характеризуются набором ответов на контрольные вопросы тестов. Процесс классификации разделяется на два этапа: 1) обучение, и 2) определение принадлежности к классам [1, 2].

Задача классификации объектов может быть решена различными методами: классическими [1-4, 6, 8, 9], на основе искусственных иммунных систем [5, 7, 10], на основе других биологических принципов, использующихся в задачах обработки данных. Среди иммунных методов, использующихся при классификации, выделяются два основных подхода: 1) классификация на основе отбора клонов [5, 10]; 2) классификация на основе теории иммунной сети [7].

Основными недостатками существующих иммунных методов классификации являются сложность их реализации и большое количество избыточных вычислений, появляющихся на различных этапах работы иммунных операторов.

Классификация объектов, предложенная в [10], осуществляется на основе отбора клонов, где предложены новые подходы к решению задач клонирования, отбора клонов и определения принадлежности объектов к классам, что позволяет ускорить этапы обучения и классификации, а также и повысить их точность.

Постановка задачи

Пусть обучающиеся представляются множеством объектов $OBJ\{obj_1; \dots; obj_n\}$, $n = \overline{1, N}$, каждый из которых характеризуется своим идентификатором

Id_i и набором признаков $P\{p_1; \dots; p_m\}$, $m = \overline{1, N}$. Результаты тестирования представляются множеством классов $CL\{cl_1; \dots; cl_k\}$, $k = \overline{1, N}$, каждый из которых характеризуется идентификатором Id_j , признаками, характеризующими центр класса $C\{c_1; \dots; c_m\}$ и допустимыми отклонениями от центра $\Delta\{\delta_1; \dots; \delta_m\}$. Признаками в данном случае являются ответы обучающихся, полученные при тестировании.

Рассматриваемый метод классификации обучающихся используется при обработке результатов тестирования владельцев оружия в компании «СТРАЖ», г. Харьков. Вопросы в тестах разбиты на четыре основных темы (по 20-25 вопросов в каждой), которым присвоены различные цены ошибок (от 0,5 до 1,5 баллов). Для классификации результатов тестирования предусмотрено три класса – «сдал», «необходимо собеседование» и «не сдал». Исходные условия для проведения обработки результатов тестирования приведены в табл. 1 и 2.

Таблица 1

Цены ошибок по темам тестирования

Тема	Цена ошибки
тема 1	1.5
тема 2	1
тема 3	0.5
тема 4	0.5

Таблица 2

Характеристика классов

Классы	Ошибки по темам				Всего баллов
	T1	T2	T3	T4	
«Сдал»	2	3	4	4	10
«Собеседование»	4	5	6	6	17
«Не сдал»	> 4	> 5	> 6	> 6	> 17

При использовании иммунного подхода для классификации объекты представляются популяцией антител $AB(ab_1; \dots; ab_n)$, а классы – популяцией антигенов $AG(ag_1; ag_2; ag_3)$. В качестве меры близости между ними используется критерий аффинности Aff_{ij} [8, 10]:

$$Aff_{ij} = (1 + d_{ij})^{-1}, \quad (1)$$

где Aff_{ij} – аффинность между i -м объектом и j -м классом; d_{ij} – евклидово расстояние между ними.

По результатам тестирования необходимо классифицировать обучающихся в соответствии с критерием (1) и с исходными установками на основе использования ИИС-алгоритма [10].

Алгоритм классификации для обработки результатов тестирования

Для классификации обучающихся на основе ИИС-алгоритма зададим форматы представления объектов и классов, которые оказывают влияние на работу иммунных операторов и объемы вычислений (рис. 1).

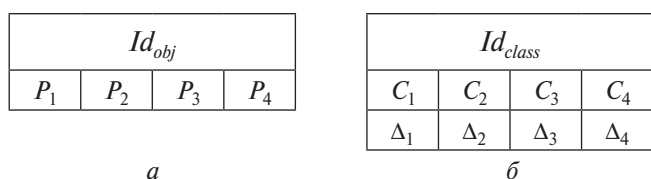


Рис. 1. Представление объекта (а) и класса (б) для классификации результатов тестирования

Здесь Id_{obj} – идентификатор объекта, Id_{class} – идентификатор класса, P_i – признак объекта, характеризующий количество допущенных ошибок по вопросам i -й темы, C_i – признак центра класса, характеризующий допустимое количество ошибок по вопросам i -й темы, а Δ_i – отклонение от центра C_i , допустимое для вхождения объекта в Id_{class} .

При использовании ИИС-алгоритма для обработки результатов тестирования его работу можно условно разделить на два основных процесса:

- 1) восстановление исходной популяции антигенов;
- 2) определение принадлежности исходных объектов $OBJ\{obj_1; \dots; obj_n\}$ к классам множества $CL\{cl_1; \dots; cl_k\}$.

Формально процесс классификации результатов тестирования может быть описан следующим образом:

$$fclass(ab_1; \dots; ab_n) = result(rconst(ab_1; \dots; ab_n), class(ab'_1; \dots; ab'_n)), \quad (2)$$

где $rconst(ab_1; \dots; ab_n)$ – восстановление антигенов (обучение) путем применения иммунных операторов отбора, клонирования, мутации и старения популяции антител, $class(ab'_1; \dots; ab'_n)$ – определение принадлежности объектов к классам.

Изначально для данной задачи все объекты относятся к классу «не сдал», т.е. данный класс является универсумом по отношению к двум другим исходным классам. Поэтому при определении критериев отбора антител параметры класса «не сдал» не учитываются, и все антитела, не прохо-

дящие первичный отбор, будут отнесены к классу «не сдал». Процесс восстановления классов $rconst(ab_1; \dots; ab_n)$ формально описывается следующим образом:

$$\begin{aligned} reconstr(ag_1; \dots; ag_n) &= reconstr(sel(ab_1; \dots; ab_n), \\ &clon(ab'_1; \dots; ab'_n), ageing(ab''_1; \dots; ab''_n)) \\ (ab'_1; \dots; ab'_n) &= sel(ab_1; \dots; ab_n) \\ (ab''_1; \dots; ab''_n) &= clon(ab'_1; \dots; ab'_n) \end{aligned} \quad (3)$$

где $sel(ab_1; \dots; ab_n)$ – функция отбора антител, $clon(ab'_1; \dots; ab'_n)$ – функция клонирования, мутации и отбора клонов для популяции антител, а $ageing(ab''_1; \dots; ab''_n)$ – функция старения антител.

Для обеспечения отбора популяции антител используется критерий средней аффинности между антигенами $CSel$ [10]:

$$CSel = \frac{\sum_{i=1}^n AF_{iAG}}{n}, \quad (4)$$

где AF_{iAG} – средняя аффинность антигена ag_i со всеми остальными антигенами.

Антитела проходят этап отбора $sel(ab_1; \dots; ab_n)$ аналогично отбору, предложенному в [10]. Для решения задачи клонирования и отбора клонов $clon(ab_1; \dots; ab_n)$ используется приоритетное последовательное клонирование с конкурентно-целевым отбором клонов. Его суть состоит в том, что для каждого антитела для клонирования выделяется всё максимально возможное на данном этапе работы количество клонов. Антитела клонируются последовательно. При последовательной организации клонирования на определенном этапе работы может возникнуть ситуация, при которой количества клонов для клонирования некоторого множества антител недостаточно. Это обуславливает необходимость использования приоритетов при клонировании клеток. Повышенный приоритет клонирования присваивается антителам, для клонирования которых количества клонов оказалось недостаточно. Поэтому такие антитела при клонировании следующей популяции будут клонироваться в первую очередь. Данный подход может быть представлен следующей последовательностью шагов:

Шаг 1. Определение количества клонов клонирования антител.

Шаг 2. Поиск антитела для клонирования в соответствии с их приоритетами.

Шаг 3. Формирование области поиска целевых антигенов.

Шаг 4. Клонирование отобранного антитела.

Шаг 5. Мутация полученных клонов.

Шаг 6. Определение целевых антигенов для полученных мутированных клонов.

Шаг 7. Конкуренция между клонами за общие целевые антигены.

Шаг 8. Добавление оставшихся клонов к поколению антител.

Для сокращения количества целевых антигенов определяются специальные области поиска целей r . Для этого используются значения аффинностей между объектами и классами:

$$r = \frac{aff_{iAG} \max + aff_{iAG} \min}{k}, \quad (5)$$

где $aff_{iAG} \max$ – максимальная аффинность i -го антитела с популяцией антиген; $aff_{iAG} \min$ – минимальная аффинность i -го антитела с антигенами; k – нормирующий коэффициент.

Отбор клонов производится путем конкуренции клеток по степени близости (аффинности) к общему целевому антигену. Таким образом, в результате конкурентного отбора из множества клонов антитела отбирается объект, имеющий наибольшую аффинность с целевым антигеном.

После восстановления антигенов антителами с помощью оператора $rconst(ab_1; \dots; ab_n)$ определяются принадлежности объектов к классам по идентификаторам целевых антигенов последней популяции антител, не восстановивших в процессе клонирования и мутации антигены. Поскольку к классу «не сдал» относятся все объекты, не вошедшие в другие классы, ситуации существования объектов с неопределенным классом не возникает. Поэтому после клонирования и мутации популяции антител нет необходимости в использовании каких-либо дополнительных процедур для формирования классов для неклассифицированных объектов. Таким образом, процесс обработки результатов тестирования может быть представлен следующим образом:

1. Определение критерия отбора антител (объектов).
2. Первичный отбор антител.
3. Цикл восстановления антиген:
 - 3.1. Удаление клонированных антител, которые не восстановили антигены.
 - 3.2. Последовательное приоритетное клонирование с конкурентно-целевым отбором клонов.
 - 3.3. Установка приоритетов клонирования для всех не клонированных антител.
 - 3.4. Проверка критерия останова, переход к шагу 3.1.
4. Определение классов для исходных объектов по результатам восстановления.
5. Конец.

Таким образом, в результате работы иммунного алгоритма производится классификация обучающихся по результатам тестирования. При этом процесс классификации упрощается, поскольку нет необходимости в формировании новых классов для неклассифицированных антител.

Экспериментальные результаты

Тестирование предлагаемого метода классификации производилось на выборке из 20 объектов (обучающихся) при исходных установках классификации (табл. 1 и 2). Результаты тестирования обучающихся представлены в табл. 3.

Таблица 3

Исходные данные для классификации (результаты тестирования обучающихся)

Объекты	Количество ошибок по темам			
	Тема 1	Тема 2	Тема 3	Тема 4
объект 1	4	2	3	5
объект 2	2	2	4	2
объект 3	3	4	8	6
объект 4	1	0	2	3
объект 5	6	4	7	8
объект 6	2	2	4	5
объект 7	4	1	1	5
объект 8	3	2	2	4
объект 9	4	2	2	6
объект 10	3	1	2	4
объект 11	3	7	4	9
объект 12	2	1	2	3
объект 13	0	0	3	4
объект 14	0	2	4	5
объект 15	3	2	3	4
объект 16	2	0	2	5
объект 17	4	1	2	2
объект 18	3	3	5	0
объект 19	2	6	4	2
объект 20	7	3	2	6

В соответствии с исходными установками (таблицы 1 и 2) был определен критерий отбора антител C_{Sel} по (4). После проведения первичного отбора антител было проведено восстановление антигенов (классов). Большое значение в работе алгоритма приобретает порог восстановления клеток, после которого начинается определение принадлежностей к классам. Это связано с большим количеством вычислений, которые производятся на различных этапах работы цикла восстановления, что может существенно увеличить время выполнения всего алгоритма. Поэтому в данном примере порог восстановления был установлен равным 80%. Для классификации объектов по данному методу потребовалось 4 прохода цикла восстановления $rconst(ab_1; \dots; ab_n)$. Результаты обработки тестов представлены в табл. 4.

В результате классификации ответов обучающихся из группы в 20 человек к классу «сдал» было отнесено 50% обучающихся (10 человек), к классу «необходимо собеседование» отнесено 30% обучающихся (6 человек) и 20% обучающихся (4 человека) отнесено к классу «не сдал». При использовании предложенного ИИС-алгоритма была произведена классификация обучающихся по результатам тестирования. Для обработки ре-

зультатов тестирования было затрачено небольшое количество времени.

Таблица 4

Результаты классификации обучающихся

Объекты	Результат кластеризации
объект 1	«Собеседование»
объект 2	«Сдал»
объект 3	«Собеседование»
объект 4	«Сдал»
объект 5	«Не сдал»
объект 6	«Сдал»
объект 7	«Собеседование»
объект 8	«Сдал»
объект 9	«Собеседование»
объект 10	«Сдал»
объект 11	«Не сдал»
объект 12	«Сдал»
объект 13	«Сдал»
объект 14	«Сдал»
объект 15	«Сдал»
объект 16	«Сдал»
объект 17	«Собеседование»
объект 18	«Собеседование»
объект 19	«Не сдал»
объект 20	«Не сдал»

Выводы

В работе рассмотрено решение актуальной задачи классификации обучающихся путем обработки результатов тестирования на основе использования искусственных иммунных систем. Классификация обучающихся выполнена путем восстановления антигенов (классов) с последующим определением принадлежности к ним антител (объектов).

Предложенный иммунный метод классификации отличается от существующих аналогов практически по всем решениям основных задач на этапе восстановления антигенов. Для повышения эффективности конкурентно-целевого отбора клонов предложен новый подход к определению области поиска целей.

Проведены экспериментальные исследования на примере классификации обучающихся владельцев оружия по результатам их тестирования в компании «СТРАЖ» (г. Харьков), которые указали на высокую эффективность предложенного метода классификации объектов на основе искусственных иммунных систем.

При незначительной модификации данный подход может использоваться для решения различных задач обработки данных, таких как идентификация, распознавание образов и др.

Список литературы: 1. *Fayyad U.M.* Advances in knowledge discovery and data mining / U.M. Fayyad, G.P. Shapiro // AAAI Press. – 1996. – 356 p. 2. *Tang Z.* A New Class Based Associative Classification Algorithm / Z. Tang, Q. Liao // International Journal of Applied Mathematics. – 1998. – P. 136-141. 3. *Matthew N.* Comparative Analysis of Serial Decision Tree Classification Algorithms / N. Matthew, G. Shiva // International Journal of Computer Science and Security. – Vol. 3, 2009. – 230 p. 4. *Afsa F.A.* A pruned fuzzy k-nearest neighbor classifier with application to electrocardiogram based cardiac arrhythmia recognition / F.A. Afsar, M.U. Akram, M. Arif // Multitopic Conference: proceedings. – 2008. – P. 143-148. 5. *Muwang C.* Application of Artificial Immune System Approach in MRI Classification / C. Muwang, C.T. Kuo, C.Y. Lin, G.H. Chang // EURASIP Journal on Advances in Signal Processing. – 2008. – P. 208-212. 6. *Gentile C.* A New Approximate Maximal Margin Classification Algorithm / Gentile C. // Journal of Machine Learning Research. – 2001. – Vol.2. – P. 213-242. 7. *Watkins A.* A New Classifier Based on Resource Limited Artificial Immune Systems / A. Watkins, L. Boggess // CEC '02. – 2002. – Vol. 02. – P. 146-151. 8. *Zhang L.* Applications of artificial immune systems in remote sensing image classification / L. Zhang, Y. Zhong, P. Li // International Journal of Remote Sensing. – 2007. – Vol. 28, Is. 7, – P. 1665-1686. 9. *Duda R.O.* Pattern classification / R.O. Duda, P.E. Hart, D.G. Stork // Wiley & Sons. – 2001. – 738 p. 10. *Кораблёв, Н.М.* Классификация объектов основе искусственных иммунных систем [Текст] / Н.М. Кораблёв, А.А. Фомичёв // Системы обработки інформації. – 2010. – Вип. 6 (87). – С. 13-17.

Поступила в редколлегию 4.03.2011.

УДК 004.89

Обробка результатів тестування тих, що навчаються, з використанням штучних імунних систем / М.М. Кораблёв, О.О. Фомічов // Біоніка інтелекту: наук.-техн. журнал. – 2011. – № 1 (75). – С. 84–87.

У даній роботі розглядається рішення задачі обробки результатів тестування учнів на основі штучних імунних систем як задачі класифікації об'єктів. Специфікою роботи використовуваного методу класифікації тих, що навчаються, є використання афінності як єдиного критерію визначення близькості об'єктів і класів.

Табл. 4. Іл. 1. Бібліогр.: 10 найм.

UDK 004.89

Processing test results studying using artificial immune systems / N.M. Korablev, A.A. Fomichov // Bionics of Intelligence: Sci. Mag. – 2011. – № 1 (75). – P. 84–87.

In this paper we consider the solution to the problem of processing the results of testing students based on artificial immune systems as the problem of classifying objects. Specificity of the method used, the classification of students is the use of affinity as the single criterion for determining the proximity of objects and classes.

Tab. 4. Fig. 1. Ref.: 10 items.