

УДК 004.89

Н.М. Кораблев, А.А. Фомичёв

Харьковский национальный университет радиоэлектроники, Харьков

КЛАССИФИКАЦИЯ ОБЪЕКТОВ НА ОСНОВЕ ИСКУССТВЕННЫХ ИММУННЫХ СИСТЕМ

В работе рассматривается алгоритм классификации объектов, работающий по принципам искусственной иммунной системы. Для решения задачи классификации предлагается использование метода приоритетного последовательного клонирования с конкурентно-целевым отбором клонов, общих критериев отбора клеток и метод дополнительного разброса в ограниченных областях при восстановлении классов. В случае неполной классификации объектов производится формирование дополнительных классов.

Ключевые слова: аффинность, критерий отбора, приоритетное последовательное клонирование, конкурентно-целевой отбор, дополнительный разброс в ограниченных областях.

Введение

Классификация представляет собой процесс разбиения объектов на заранее установленные группы (классы) по подобию их признаков [1]. Процесс классификации объектов включает в себя два основных этапа:

- 1) обучение;
- 2) определение принадлежности объектов к классам.

На этапе обучения формируются классы, к которым впоследствии производится определение принадлежности объектов.

Задача классификации может быть решена различными методами, как классическими [1, 2], так и на основе искусственных нейронных сетей, иммунных систем и др. [3 – 7]. Особенностью иммунных методов обработки данных является возможность комбинирования различных подходов и формирование гибридов, сочетающих в себе свойства классических и современных методов. Среди иммунных методов классификации объектов можно выделить два основных подхода:

- 1) классификация на основе клонального отбора;
- 2) классификация на основе теории иммунной сети [3 – 6].

В [3] предложен иммунный метод классификации на основе клонального отбора. На этапе отбора клонов в [3] производится большое количество избыточных вычислений, а дополнительный разброс производится случайным образом, что не обеспечивает быстрого восстановления классов.

Использование иммунного метода «областей распознавания» (ARBs) в [5] обеспечивает хорошую точность классификации, однако не отличается высокой скоростью работы.

В [6] используется модифицированный метод клонального отбора, обеспечивающий автоматиче-

скую классификацию объектов с большой степенью точности. Метод [6] обеспечивает более быструю классификацию, чем [3], однако не устраняет проблему избыточных вычислений, а также повышает сложность реализации.

В предлагаемом методе иммунной классификации объектов на основе принципа клонального отбора используются новые подходы к решению задач клонирования и отбора клонов, организации дополнительного разброса и принципу отнесения объектов к классам, что позволяет ускорить этап обучения и автоматизировать процесс классификации.

Постановка задачи

Дано множество объектов

$$OBJ\{obj_1; \dots; obj_k\}, \quad i = \overline{1, N},$$

каждый из которых описывается матрицей признаков $MO_{n \times m}$, и множество классов $CL\{cl_1; \dots; cl_z\}$, $j = \overline{1, N}$, каждый из которых описывается матрицей признаков $MC_{n \times m}$. Применительно к теории искусственных иммунных систем классы представляются популяцией антиген $AG(ag_1; \dots; ag_z)$, а исходные объекты – популяцией антител $AB(ab_1; \dots; ab_k)$.

В качестве меры близости между объектами m_i используется критерий аффинности [3 – 7]:

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

где Af_{ij} – аффинность между i и j объектами;

d_{ij} – евклидово расстояние между ними.

Необходимо классифицировать исходное множество объектов $OBJ\{obj_1; \dots; obj_k\}$ по классам $CL\{cl_1; \dots; cl_z\}$ на основе иммунного подхода, используя в качестве основной меры близости объектов и классов критерий (1).

Иммунный алгоритм классификации

Работу иммунного алгоритма классификации можно условно разделить на три основных процесса: восстановление исходной популяции антиген, определение принадлежности исходных объектов

$$OBJ\{obj_1; \dots; obj_k\}$$

к классам множества

$$CL\{cl_1; \dots; cl_z\}$$

и, в случае невозможности определения классов для некоторого множества объектов, формирование новых классов.

Формально процесс иммунной классификации объектов может быть описан следующим образом:

$$\begin{aligned} fclass(ag_1; \dots; ag_z) &= result(rconst(ab_1; \dots; ab_k), \\ &class(ab'_1; \dots; ab'_{k'}), update(ag_1; \dots; ag_z)) \\ &\text{при } \exists(ag_1; \dots; ag_z), k = k' \\ fclass(ag_1; \dots; ag_z) &= result(rconst(ab_1; \dots; ab_k), \\ &class(ab'_1; \dots; ab'_{k'}), fclust(ab''_1; \dots; ab''_{k''}), \\ &cupdate(ag'_1; \dots; ag'_z)) \\ &\text{при } \exists(ag_1; \dots; ag_z), k = k' + k'', k' > 0 \\ fclass(ag_1; \dots; ag_z) &= result(fclust(ab_1; \dots; ab_k), \\ &cupdate(ag_1; \dots; ag_z)) \end{aligned} \quad (2)$$

в остальных случаях

где $rconst(ab_1; \dots; ab_n)$ – восстановление антиген (обучение) путем применения иммунных операторов отбора, клонирования, мутации и старения популяции антител,

$class(ab'_1; \dots; ab'_{k'})$ – определение принадлежности объектов к классам,

$update(ag_1; \dots; ag_z)$ – обновление параметров классов,

$fclust(ab''_1; \dots; ab''_{k''})$ – кластеризация объектов, классы которых не были определены,

$cupdate(ag'_1; \dots; ag'_z)$ – добавление новых классов, сформированных после кластеризации объектов [8].

Таким образом, иммунная классификация объектов происходит следующим образом:

Этап 1. Проверка наличия классов. Если классы заданы – переход к этапу 2, иначе – этап 4.

Этап 2. Восстановление антиген (классов).

Этап 3. Определение принадлежности исходных объектов к классам по результатам процедуры восстановления, если принадлежность для всех объектов определена – конец, иначе – переход к 4-му этапу – кластеризации.

Этап 4. Кластеризация неклассифицированных антител.

Этап 5. Обновление классов по результатам кластеризации (добавление новых классов).

Способ представления объектов и классов влияет на организацию вычислений и работу иммунных операторов. Классифицируемые объекты и классы представляются матрицами одинакового размера, где каждый столбец представляет определенную группу параметров.

На рис. 1 представлен общий вид матриц параметров объектов и классов.

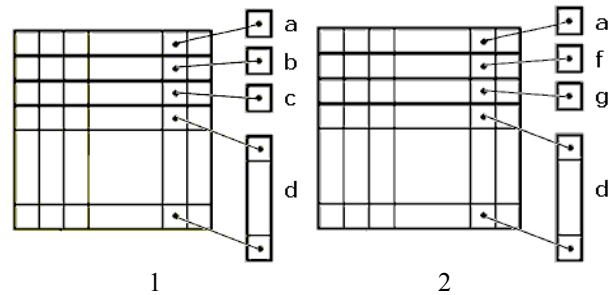


Рис. 1. Представление матриц параметров объектов и классов:
1 – матрица представления класса;
2 – матрица представления объекта

На рис. 1 используются следующие условные обозначения:

- a – размерность вектора параметров,
- b – вес вектора параметра в определении принадлежности объекта,
- c – допустимые отклонения параметров вектора,
- d – признаки объекта/класса (координаты, цвет и др.),
- f – идентификатор принадлежности группы параметров объекта классу,
- g – общий идентификатор класса, которому принадлежит объекта.

Основное различие между представлением классов и объектов заключается в том, что классы помимо параметров своего центра имеют строку весов параметров (b) и строку допустимых отклонений (c)

Процесс восстановления классов

$$reconstr(ab_1; \dots; ab_k)$$

происходит аналогично [7], однако в [7] для представления клеток и кластеров использовались не матрицы, а вектора параметров, поэтому организация вычислений в рассматриваемом методе иммунной классификации объектов отличается от [7].

Для обеспечения отбора антител (первичного и после дополнительного разброса) используется критерий средней аффинности антиген CSel [7]:

$$CSel = \frac{\sum_{i=1}^n AF_{iAG}}{n}, \quad (3)$$

где AF_{iAG} – средняя аффинность антигена ag_i со всеми антигенами выборки $AG(ag_1; \dots; ag_n)$, которая определяется следующим образом:

$$AF_{iAG} = \frac{\sum_{j=1}^n af_{ij}}{n}. \quad (4)$$

Здесь af_{ij} – аффинность антигена ag_i и антигена ag_j . Это упрощает процедуру отбора и повышает вероятность восстановления антиген на последующих этапах иммунного алгоритма.

Функция (3) определения критерия отбора клеток $CSel$ в [7] использовалась для работы с векторами параметров объектов. При работе с матрицами параметров объектов для каждого столбца необходимо определение отдельного критерия отбора. Поэтому при отборе клеток используется вектор критериев отбора

$$\overline{CSel}(csel_1; \dots; csel_n),$$

где n – количество столбцов матриц классов и объектов.

При отборе антител и клонов для каждой клетки в соответствии с (4) формируется вектор средних аффинностей с популяцией антиген

$$\overline{AF_{iAG}}(af_{iAG1}; \dots; af_{iAGn}),$$

который сопоставляется с вектором \overline{CSel} . Клетка проходит отбор в том случае, если его более половины средних значений вектора $\overline{AF_{iAG}}$ удовлетворяет условию (5) с соответствующими значениями вектора критериев отбора \overline{CSel} :

$$AF_{iAG} \geq 98\% \times CSel. \quad (5)$$

Для решения задачи клонирования и отбора используется приоритетное последовательное клонирование с конкурентно целевым отбором клонов [8].

Процедура клонирования и отбора представлена следующим образом (6):

$$\begin{aligned} \text{clon}(ab_1; \dots; ab_n) &= \text{clon}(\text{create}(ab_1; \dots; ab_k), \\ &\text{cmut}(ab'_1; \dots; ab'_k), \text{csel}(ab''_1; \dots; ab''_k)), \\ (ab'_1; \dots; ab'_k) &= \text{create}(ab_1; \dots; ab_k), \\ (ab''_1; \dots; ab''_k) &= \text{mut}(ab'_1; \dots; ab'_k), \end{aligned} \quad (6)$$

где $\text{create}(ab_1; \dots; ab_k)$ – функция создания клонов;

$\text{cmut}(ab'_1; \dots; ab'_k)$ – функция мутации клеток;

$\text{csel}(ab''_1; \dots; ab''_k)$ – функция отбора клонов.

Суть последовательного клонирования [7] состоит в том, что для каждого антитела выделяется все максимально возможное на данном этапе работы

количество клонов. Такая организация повышает вероятность восстановления антиген и ускоряет работу алгоритма.

Конкурентно-целевой отбор [7] используется для уменьшения количества отобранных клонов и заключается в том, что для каждого клонируемого антитела определяется расстояние поиска целевых антиген r :

$$r = k \times \sqrt{\frac{\text{width} \times \text{height}}{n}}, \quad (7)$$

где width – ширина,

height – высота области поиска объектов,

n – количество антиген,

k – коэффициент увеличения области.

Для каждого i -го клона после клонирования и мутации объекта определяется матрица аффинностей с целевыми антигенами $MAF_{i \text{ width} \times \text{height}}$, в которой строки являются векторами аффинностей $\overline{AF_{iAG}}$ с антигенами, расположенными в области поиска r .

После этого среди строк матрицы определяется строка с максимальными аффинностями с учетом весов групп параметров. Антиген, которому соответствует данная строка, является целевым антигеном.

Отбор клонов осуществляется путем конкуренции клеток по целевому признаку [7]. Последовательное клонирование антител может привести к ситуации, когда для клонирования некоторого множества антител клонов может оказаться недостаточно. В этом случае для организации клонирования предлагается ввести приоритеты клонирования. Это выражается в том, что объекты, которым не хватило клонов для клонирования, на следующем шаге будут обладать повышенным приоритетом и клонироваться в первую очередь.

Для организации дополнительного разброса предлагается формирование ограниченных областей разброса [7]. Основной отличительной чертой процесса восстановления антиген при классификации от [7] является повторение мутации восстановившего антитела до тех пор, пока оно не восстановит все параметры антигена полностью, либо до достижения заданного порога повторения цикла восстановления.

После восстановления антиген определяются принадлежности объектов по идентификаторам целевых антиген последней популяции антител. В случае, если после проведения процедуры восстановления классов существуют объекты с неопределенным классом, т.е. объекты, антитела которых не восстановили антигены, начинается формирование новых классов путем кластеризации оставшихся антител [7].

Классы, сформированные после кластеризации неклассифицированных объектов, добавляются во множество исходных классов.

Процесс иммунной классификации объектов может быть представлен следующим образом:

1. Определение вектора критериев отбора антител.
2. Определение областей дополнительного разброса.
3. Отбор антител.
4. В случае удаления всех антител переход к 7.
5. Цикл восстановления антител:
 - 5.1. Отбор антител полученных в результате дополнительного разброса на предыдущей итерации
 - 5.2. Удаление клонированных не восстанавливающих антител.
 - 5.3. Приоритетное последовательное клонирование с конкурентно-целевым отбором клонов.
 - 5.4. Установка приоритетов клонирования для всех не клонированных антител.
 - 5.5. Осуществление дополнительного разброса.
 - 5.6. Проверка критерия останова, переход к 5.1.

6. Определение классов для исходных объектов по результатам восстановления.

6.1. Установка идентификаторов классов объектам.

6.2. Определение неклассифицированных объектов, в случае их существования – переход к 7, иначе – к 9.

7. Кластеризация объектов с неопределенными кластерами [7].

8. Обновление классов по результатам кластеризации.

9. Конец

В результате работы иммунного алгоритма классификации производится определение принадлежности объектов к заданным классам и формирование новых классов в результате кластеризации объектов, не принадлежащих ни одному из исходных классов.

Экспериментальные результаты

Тестирование предлагаемого алгоритма классификации производилось в области размером 300×270 .

В работе использовалось 3 класса и 1000 объектов, сформированных случайным образом в трех областях разброса (рис. 2).

После формирования популяции антител в соответствии с (3) был вычислен вектор критериев отбора клеток \overline{CSel} , и произведено определение областей дополнительного разброса.

После первичного отбора антител путем реализации операторов клонирования, мутации, старения

и дополнительного разброса осуществляется восстановление антиген.

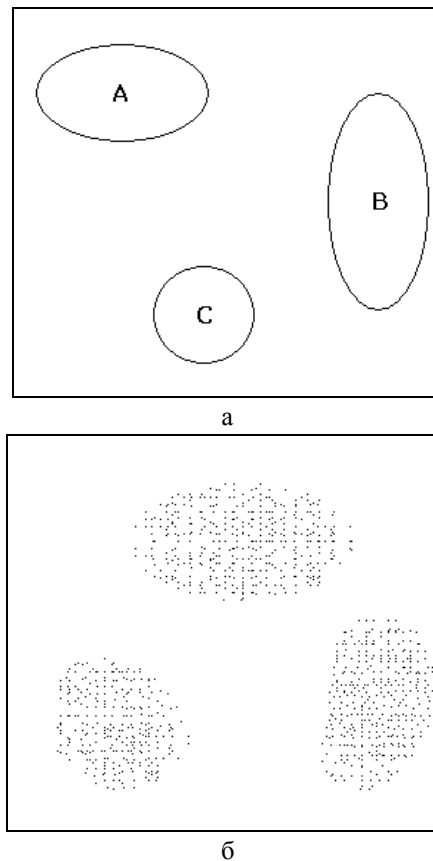


Рис. 2. Входные данные для классификации:
а – исходные классы;
б – исходные объекты

После восстановления исходной популяции антиген начинается определение принадлежности объектов к классам. На восстановление классов потребовалось 6 проходов цикла восстановления, однако 10,2% количества исходных объектов не определили принадлежность к какому-либо из классов, вследствие чего начался этап их кластеризации. По результатам кластеризации был сформирован класс «D», в который вошли неклассифицированные объекты. Результаты работы алгоритма иммунной классификации представлены на рис. 3.

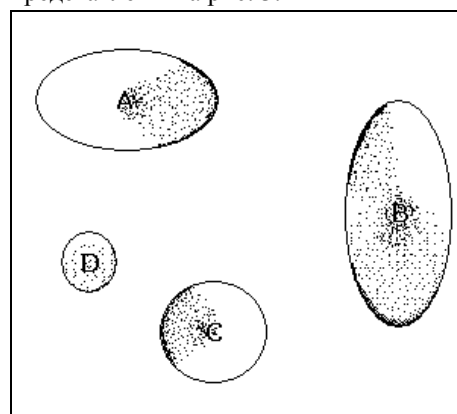


Рис. 3. Результат классификации объектов

Таким образом, для классификации объектов предлагаемому алгоритму потребовалось небольшое количество времени благодаря использованию метода приоритетного последовательного клонирования с конкурентно-целевым отбором клонов и увеличению точности дополнительного разброса объектов.

Выводы

В работе рассмотрено решение актуальной задачи классификации объектов на основе искусственных иммунных систем.

Предложенный метод классификации отличается от существующих решений основных задач на этапе восстановления классов.

Особенностью рассмотренного метода иммунной классификации объектов является также его способность к самообучению путем добавления классов после кластеризации объектов, не принадлежащих ни одному из исходных классов.

Использование новых методов работы иммунных операторов, таких как приоритетное последовательное клонирование и дополнительный разброс в ограниченных областях разброса ускоряет процесс восстановления классов и, следовательно, и сам процесс классификации объектов.

Предложенный метод классификации, несмотря на достаточную сложность реализации, обеспечивает быстрое определение классов для классифицируемых объектов и при незначительной модификации может использоваться для решения различных задач обработки данных (распознавания образов, идентификации 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 // *IAENG International Journal of Applied Mathematics*. – 1998. – 36:2, IJAM. – P. 136-141.
3. 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.
4. Gentile C. *A New Approximate Maximal Margin Classification Algorithm* / C. Gentile // *Journal of Machine Learning Research*. – 2001. – 2. – P. 213-242.
5. Watkin A. *A New Classifier Based on Resource Limited Artificial Immune Systems* / A. Watkins, L. Boggess // *CEC '02*. – 2002. – Vol. 02. – P. 146-151.
6. 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, issue 7. – P. 1665-1686.
7. Кораблёв Н.М. *Исследование иммунных операторов в задаче кластеризации объектов* / Н.М. Кораблёв, А.А. Фомичёв // *Бионика интеллекта*. – 2010. – № 1 (72). – С. 70-74.

Поступила в редакцию 1.09.2010

Рецензент: д-р техн. наук, проф. С.Г. Удовенко, Харьковский национальный университет радиоэлектроники, Харьков.

КЛАСИФІКАЦІЯ ОБ'ЄКТІВ НА ОСНОВІ ШТУЧНИХ ІМУННИХ СИСТЕМ

М.М. Корабльов, О.О. Фомічов

В роботі розглядається алгоритм класифікації об'єктів, що працює за принципами штучної імунної системи. Для вирішення задачі класифікації пропонується використання методу пріоритетного послідовного клонування із конкурентно-цільовим відбором клонів, загальних критеріїв відбору клітин та метод додаткового розкиду у визначених областях при відновлюванні класів. У випадку неповної класифікації об'єктів формуються додаткові класи.

Ключові слова: афінність, критерій відбору, пріоритетне послідовне клонування, конкурентно-цільовий відбір, додатковий розкид у визначених областях.

OBJECTS CLASSIFICATION USING ARTIFICIAL IMMUNE SYSTEMS

N.M. Korablev, A.A. Fomichov

The given work is devoted to the designing of the objects classification algorithm based on artificial immune system approach. To solve the problem of classification it proposed to use priority sequential cloning with competitive targeting cloning selection and some selection and the method of additional scatter in the limited areas in the recovery of the classes. In the case of partial classification of objects is the formation of additional classes.

Keywords: affinity, selection criterion, priority sequential cloning, competitive targeting cloning selection, additional dispersion in the limited areas.