

МОДИФИКАЦИЯ ИММУННОГО МЕТОДА RLAI5 ДЛЯ АВТОМАТИЧЕСКОЙ КЛАССИФИКАЦИИ ОБЪЕКТОВ

УДК 004.89

КОРАБЛЕВ Николай Михайлович

д.т.н, профессор кафедры электронных вычислительных машин Харьковского национального университета радиоэлектроники.

Научные интересы: интеллектуальная обработка информации с использованием искусственных иммунных систем.

e-mail: korablev@kture.kharkov.ua

ФОМИЧЕВ Александр Александрович

ассистент кафедры электронных вычислительных машин Харьковского национального университета радиоэлектроники.

Научные интересы: классификация и кластеризация данных с использованием искусственных иммунных систем.

e-mail: alexandros_1985@mail.ru

КУШНАРЕВ Максим Владимирович

ассистент кафедры электронных вычислительных машин Харьковского национального университета радиоэлектроники.

Научные интересы: распознавание образов, детектирование и анализ вредоносных программ с использованием искусственных иммунных систем.

e-mail: mauxion@gmail.com

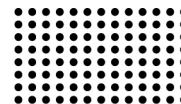
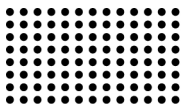
ВВЕДЕНИЕ

Классификация данных является процессом группировки разрозненных объектов в однородные группы (классы) по однородности их признаков [1]. Под однородностью признаков понимают наличие у нескольких объектов одного класса общих свойств, определяющих их сходство. Это сходство является основанием для отнесения таких объектов к одному классу при использовании определенного критерия, применяемого для измерения степени схожести между ними [1, 2].

Существует два основных вида классификации [1]: 1) классификация с контролируемым обучением (обучением с учителем); 2) классификация с неконтролируемым обучением (классификация без учителя или кластеризация). В случае классификации с контролируемым обучением предполагается использование набора известных классов, заданных в явном виде, либо использование обучающих или проверяющих выборок. При кластеризации задача

группировки объектов сводится к выделению кластеров, информация о которых изначально неизвестна. При этом обучающие выборки не используются. В некоторых случаях в качестве исходной информации, кроме набора классифицируемых объектов, используется количество формируемых классов [1-3].

При решении практических задач наиболее часто используется автоматическая классификация объектов [1]. При таком способе классификации информация о классах может быть априорно известной, либо определяться путём кластеризации. Кроме того, при автоматической классификации обычно используется механизм самообучения, который заключается в переопределении параметров классов и их количества по результатам обработки нескольких различных наборов объектов. При организации самообучения в алгоритмах автоматической классификации объектов могут использоваться различные модели и методы, в том числе, основанные на биологических принципах организации вычислений. Среди таких моделей



выделяются модели, функционирующие на основе теории искусственных иммунных систем (ИИС). В настоящее время существует множество иммунных моделей, используемых при классификации, из которых наиболее распространёнными являются [2-4]: 1) модель иммунной сети (метод aiNet); 2) модель клонального отбора (методы CLONALG, VCA); 3) модель ресурсно-ограниченной иммунной системы (метод RLAIIS), разработанной для решения задач классификации и распознавания образов. В процессе самообучения группируемые антитела в ходе клонирования и мутации воспроизводят признаки антигенов (классов). Антитело становится специфичным антигену в том случае, если значения их признаков совпадают. Идеальным условием завершения классификации является состояние полной специфичности антител антигенам, при котором признаки каждого антитела в иммунного обучения соответствуют признакам какого-либо антигена.

Существующие иммунные методы классификации, разделяются на две группы: 1) методы, в которых классификация происходит после достижения состояния специфичности популяций антител и антигенов; 2) методы, в которых классификация происходит в процессе достижения состояния специфичности. Иммунные методы aiNet, CLONALG и VCA относятся к первой группе. Основным преимуществом методов aiNet и CLONALG по сравнению с VCA является высокое быстродействие. Однако они уступают методу VCA в точности. В методе RLAIIS формируемые классы представляются множеством распознающих областей ARB (Artificial Recognition Ball), которым принадлежат антитела или антигены. Несмотря на высокую сложность реализации, по сравнению с другими иммунными методами, он характеризуется более высокой точностью классификации. Основным недостатком метода RLAIIS является низкое быстродействие по сравнению с методами aiNET или CLONALG.

В работе предлагается модифицированный метод RLAIISm, который отличается от оригинального RLAIIS способом организации сетевого взаимодействия антигенов, способом организации конкуренции ARB-классов и принципом сжатия сети. Изменение оператора мутации и принципа сжатия популяции антител предназна-

чены для повышения быстродействия метода RLAIIS без потери точности классификации объектов.

ПОСТАНОВКА ЗАДАЧИ

Пусть исходные данные для классификации представляются множеством m объектов $OBJ\{obj_1; \dots; obj_m\}$, которые в зависимости от постановки задачи классификации могут быть представлены наборами антигенов или антител. В случае классификации с контролируемым обучением исходные классы представляются популяцией n антигенов $AG\{ag_1; \dots; ag_n\}$, а классифицируемые объекты – популяцией m антител $AB\{ab_1; \dots; ab_m\}$. В результате классификации антиген передает принадлежность к классу (ARB), специфичному антителу. В случае автоматической классификации антигенами является множество исходных объектов, а антителами – набор динамически формируемых и изменяемых объектов, используемых для определения центров кластеров (ARB-классов).

Форматом представления классифицируемых объектов являются матрицы признаков, либо множество векторов признаков переменной длины. При этом каждый столбец матрицы является группой признаков объекта, которая характеризуется количеством атрибутов и весом, используемым при классификации.

Для определения специфичности антигенов и антител используется аффинность [2-4]:

$$\text{aff}(ag_i; ab_j) = (1 + d(ag_i; ab_j))^{-1}, \quad (1)$$

где $\text{aff}(ag_i; ab_j)$ – аффинность между i -м антигеном и j -м антителом, $d(ag_i; ab_j)$ – евклидово расстояние между ними [3] (в случае метрических признаков), либо хеммингово расстояние (в случае неметрических признаков) [3]. Следует отметить, что аффинность между иммунными объектами изменяется в диапазоне (0; 1]. При этом специфичность между антителом и антигеном достигается в случае максимального значения аффинности между объектами.

В случае, когда используются матрицы признаков, либо совокупность векторов переменной длины, аффинность между антигенами и антителами определяется следующим образом [2-4]:

$$\text{aff}(ag_i; ab_j) = \sum_{c=1}^k w_c (1 + d(ag_i; ab_j)_c)^{-1}, \quad (2)$$

где k – общее количество групп признаков, т.е. количество столбцов в матрице или векторов переменной длины, w_c – вес группы признаков, $d(ag_i; ab_j)_c$ – евклидово либо хеммингово расстояние между признаками данной группы для i -го антигена и j -го антигена.

Необходимо разработать иммунный метод классификации данных на основе метода RLAIS, устраняющий проблему избыточности вычислений, характерную для BSA и RLAIS, без потери быстродействия, характерного для aiNET и CLONALG.

МОДИФИКАЦИЯ МЕТОДА RLAIS

Предлагаемый метод классификации функционирует на основе метода RLAIS [3, 4], однако в его работу вносится ряд модификаций: использование стимулирующих антител для повышения скорости иммунного обучения и выделения центров классов; применение мутации на основе аффинности дальнего предка, в соответствии с [5]; а также использование конкурентно-целевого отбора клонов, в соответствии с [2]. Следует отметить, что метод RLAIS сочетает особенности как модели клонального отбора, так и модели искусственной иммунной сети, характеризующейся взаимодействием между антителами. В соответствии с этим модель иммунной сети позволяет использовать механизмы взаимодействия между антителами для стимуляции клонирования, мутации или сжатия сети. Таким образом, из популяции антител можно выделить антитела, характеризующиеся высокими аффинностями к антигенам, стимулирующие клонирование остальной популяции антител. Такие антитела являются стимулирующими по отношению к популяции исходных антител. Стимулирующие антитела в результате взаимодействия с остальными антителами формируют области стимуляции – множества антител, характеризующиеся высокими аффинностями к стимулирующим антителам.

Использование стимулирующих антител и их областей стимуляции упрощает процесс выделения ARB-классов и выбора их центров. Кроме того, стимулирующие антитела используются для

повышения скорости иммунного обучения. Следует отметить, что в большинстве иммунных методов клонированию подвергается вся популяция антител, независимо от их аффинностей к антигенам (BSA, RLAIS), либо некоторая их часть, характеризующаяся высокими аффинностями к антигенам (CLONALG, aiNET). При этом антитела, не подлежащие клонированию, не обрабатываются и удаляются из популяции в результате старения или супрессии сети. Использование стимулирующих антител позволяет решить эти проблемы.

После определения областей стимуляции для каждого стимулирующего антитела формируется множество антител, подвергающихся действию его стимуляции. При этом клонирование производится только для множества стимулирующих антител. В результате отбора клонов для каждого стимулирующего антитела определяется один клон, характеризующийся максимальной аффинностью к антигенам, который может заменить стимулирующее антитело на этапе старения популяции. Одной из особенностей использования стимулирующих антител является обработка клонов на этапе клонального отбора. В результате такого клонального отбора количество выбранных клонов не меньше количества антител, находящихся в области стимуляции. Это производится для упрощения процесса старения, в ходе которого антитела из областей стимуляции могут быть заменены наилучшими клонами стимулирующего антитела.

Функционирование модифицированного метода RLAISm разделяется на ряд основных этапов:

- 1) подготовительный этап;
- 2) этап иммунного обучения и выделения начальных ARB-классов;
- 3) этап конкуренции между сформированными ARB-классами и их слияние;
- 4) классификация антигенов, для которых отсутствуют специфичные антитела.

В ходе выполнения подготовительного этапа происходит определение пороговой аффинности между объектами NAT (Natural Affinity Threshold) в соответствии с [2, 3], а также вычисление аффинностей между антителами и антигенами. На этапе иммунного обучения и выделения ARB-классов происходит

выделение стимулирующих антител, являющихся начальными центрами ARB, и их областей стимуляции, а также их клонирование, мутация, отбор клонов и старение. На этапе конкуренции между ARB происходит определение областей конкуренции за антитела между ARB-классами, уточнение их границ и слияние классов, имеющих минимальное количество антител с ближайшими крупными классами. Определение классов для антигенов, не имеющих специфичных антител, является завершающим этапом классификации.

Для упрощения разработки и модификации иммунных методов обработки данных их условно разбивают на совокупность иммунных операторов, каждый из которых выполняет определённую функцию по отношению к иммунным объектам [2], например,

оператор клонирования применяется для клонирования антител, оператор мутации – для мутации клонов и т.д. Иммунные операторы разделяются на две группы: 1) универсальные операторы – операторы, использование которых не зависит от выбора иммунной модели (операторы клонирования, мутации и др.); 2) операторы, зависящие от выбора модели (операторы клонального отбора, супрессии и др.). Поскольку метод RLAIsm использует принципы моделей клонального отбора и иммунной сети, в нём могут присутствовать операторы клонального отбора, супрессии и старения. В соответствии с этим, модифицированный метод RLAIsm на уровне иммунных операторов может быть представлен следующим образом:

$$\begin{aligned}
 \text{RLAIsm}(AB, AG) &= \left[\begin{array}{l} \text{NatCalculation}(AG) \rightarrow \\ \text{Ag Presentation}(AB, AG) \end{array} \right] \Rightarrow \left[\begin{array}{l} \text{StAbSelection}(AB, AG, NAT) \rightarrow \\ \left. \begin{array}{l} \text{StAbCloning}(AB, CL) \rightarrow \\ \text{Mutation}(CL) \rightarrow \\ \text{ClSelection}(CL) \rightarrow \\ \text{Apoptosys}(AB, CL) \end{array} \right\}^T \\ \text{ARBSelection}(AB, ARB) \end{array} \right] \Rightarrow \\
 &\Rightarrow \left[\begin{array}{l} \text{StimCalculation}(ARB, AG) \rightarrow \\ \text{SpacesSelection}(ARB, AB) \rightarrow \\ \text{ARBComparison}(ARB, AB) \rightarrow \\ \text{ARBMerging}(ARB, AB) \end{array} \right] \Rightarrow \text{ARB Presentation}(AG, AB, ARB),
 \end{aligned} \tag{3}$$

где $\text{NatCalculation}(AG)$ – оператор определения порогового значения NAT, используемого при определении стимулирующих антител; $\text{Ag Presentation}(AB, AG)$ – оператор представления антигенов антителам и выделения целевых антигенов для каждого антитела [2]; $\text{StAbSelection}(AB, NAT)$ – оператор выделения стимулирующих антител и их областей стимуляции; $\text{StAbCloning}(AB, CL)$ – оператор клонирования стимулирующих антител; $\text{Mutation}(CL)$ – оператор мутации клонов; $\text{ClSelection}(CL)$ – оператор отбора клонов; $\text{Apoptosys}(AB, CL)$ – оператор старения антител; $\text{ARBSelection}(AB, ARB)$ – оператор выделения ARB-классов; $\text{StimCalculation}(ARB, AG)$ – оператор

определения уровней стимуляции для ARB-классов; $\text{SpacesSelection}(ARB, AB)$ – оператор определения областей конкуренции между классами; $\text{ARBComparison}(ARB, AB)$ – оператор сопоставления уровней стимуляции между ARB-классами и конкуренции за антитела в выделенных областях; $\text{ARBMerging}(ARB, AB)$ – оператор слияния классов; $\text{ARB Presentation}(ARB, AG)$ – оператор представления ARB-классов антигенам, не имеющим специфичных антител; T – критерий остановки иммунного обучения.

Оператор определения критерия отбора антител $\text{NatCalculation}(AG)$ используется в исходном методе RLAIsm [3, 4]. При этом значение NAT является средней аффинностью внутри популяции антигенов. В

дальнейшем данный параметр используется при выделении стимулирующих антител и определении их областей стимуляции. В соответствии с [3, 4] значение NAT определяется следующим образом:

$$NAT = \frac{1}{n-1} \left(\sum_{i=1}^n \sum_{j=1}^n \text{aff}(ag_i, ag_j) - n \right), \quad (4)$$

где n – общее количество антигенов.

Работа оператора представления антигенов антителам $AgPresentation(AB, AG)$ заключается в определении аффинности между антителами и антигенами в соответствии с (2). Поскольку в предлагаемом модифицированном методе RLAIsm используется конкурентно-целевой отбор клонов, на этапе представления антигенов антителам в соответствии с [2] для каждого антитела определяются целевые антигены. Это делается для сокращения количества вычислительных операций при последующем отборе клонов и старении.

Оператор $StAbSelecion(AB, NAT)$ используется для выделения стимулирующих антител и определения их областей стимуляции. Стимулирующими являются антитела с максимальными аффинностями к антигенам, аффинности между которыми меньше порогового значения NAT. После выделения стимулирующих анти-

тел происходит определение аффинностей стимулирующих антител к остальным антителам. Антитела входят в область стимуляции в том случае, если их аффинность со стимулирующими антителами превышает пороговое значение NAT. После определения областей стимуляции для каждого целевого антитела происходит определение целевых антигенов. При этом множество целевых антигенов стимулирующего антитела формируется путём включения множеств целевых антигенов всех антител из его области стимуляции.

Действию оператора клонирования $StAbCloning(AB, CL)$ подвергаются только стимулирующие антитела. При этом в соответствии с [2] происходит пропорциональное клонирование. Количество формируемых клонов прямо пропорционально максимальной аффинности стимулирующего антитела к одному из целевых антигенов. Следует отметить, что при определении количества клонов минимальным значением является количество антител из области стимуляции. Таким образом, количество клонов, формируемое стимулирующим антителом, определяется следующим образом:

$$N_{clones} = \begin{cases} m \cdot \text{aff}_{\max}(ab_i^s, AG^t), & \text{если } m \cdot \text{aff}_{\max}(ab_i^s, AG^t) > sn_i \\ sn_i, & \text{если } m \cdot \text{aff}_{\max}(ab_i^s, AG^t) \leq sn_i \end{cases} \quad (5)$$

где m – общее количество антител, $\text{aff}_{\max}(ab_i^s, AG^t)$ – максимальная аффинность i -го стимулирующего антитела ко множеству его целевых антигенов, sn_i – количество антител в области стимуляции.

Для изменения признаков клонов используется обратно-пропорциональная мутация на основе аффинности дальнего предка, предложенного в [5]. Основное отличие данного подхода от существующих методов мутации заключается в способе определения коэффициента мутации. В существующих иммунных методах определение коэффициента при обратно-пропорциональной мутации зависит только от аффинности клонированного антитела, и определяется в соответствии с выражением ниже:

$$\mu = \text{rand}[0; 1 - \text{aff}(ab_i, AG)], \quad (6)$$

где $\text{aff}(ab_i, AG)$ – аффинность клонированного i -го антитела к популяции антигенов.

Следует отметить, что минимальное значение коэффициента мутации μ равно нулю, вследствие чего он может быть определён как предельно малое значение, что негативно скажется на мутации признаков клонов и приведёт к увеличению времени иммунного обучения. Для устранения данной проблемы при определении μ нижняя граница в выражении (6) является аффинностью дальнего предка. Дальним предком для клона является антитело, сформировавшее антителородитель в ходе клонирования предыдущей популяции антител. Таким образом, определение коэффициента мутации происходит следующим образом:

$$\mu = \text{rand}[1 - \text{aff}(ab_i^*, AG^t); 1 - \text{aff}(ab_i, AG^t)], \quad (7)$$

где $\text{aff}(ab_i^*, AG^t)$ – аффинность дальнего предка клонированного i -го антитела к целевым антигенам; $\text{aff}(ab_i, AG^t)$ – аффинность клонированного антитела к целевым антигенам.

При использовании оператора случайной мутации изменение признака клона производится следующим образом:

$$cl_i.attr = cl_i.attr \pm \mu \cdot (\text{aff}(ab_i, AG)^{-1} - 1), \quad (8)$$

где $cl_i.attr$ – признак i -го клона, который подвергнут действию оператора мутации; μ – коэффициент мутации, определяемый в соответствии с (7). Следует отметить, что при случайной мутации знак \pm в выражении (8) определяется случайным образом.

В предлагаемом методе классификации используется оператор конкурентно-целевого отбора клонов $ClSelection(CL)$, предложенный в [2].

Оператору старения $Apoptosis(AB, CL)$ подвергаются все антитела популяции, независимо от значений их аффинности к антигенам. При этом клоны, выделенные в результате клонального отбора, заменяют антитела, находящиеся в областях стимуляции в случае, если их аффинности к целевым антигенам превосходят аффинности антител, находящихся в областях стимуляции.

Оператор формирования классов из выделенных стимулирующих антител $ARBSelection(AB, ARB)$ используется для определения границ формируемых классов. В результате работы данного оператора, границы классов ARB расширяются за счёт антител, у которых аффинность с центром данного класса превышает значение NAT . Следует отметить, что на данном этапе работы предлагаемого метода антитела могут принадлежать сразу нескольким классам.

Определение уровня стимуляции, в ходе работы оператора $StimCalculation(ARB, AG)$, осуществляется следующим образом:

$$sl_j = \frac{1}{z} \cdot \sum_{i=1}^z \text{aff}(ab_i, AG), \quad (9)$$

где sl_j – уровень стимуляции j -го ARB -класса; z – количество антител, принадлежащих j -му ARB -классу; $\text{aff}(ab_i, AG)$ – аффинность i -го антитела, принадлежащего данному ARB -классу, с популяцией целевых антигенов.

Оператор определения областей конкуренции между ARB -классами $SpacesSelection(ARB, AB)$ используется для определения антител, находящихся в нескольких областях стимуляции различных стимулирующих антител, т.е. принадлежащих нескольким классам.

Работа оператора конкуренции ARB -классов за антитела $ARBComparison(ARB, AB)$ заключается в сопоставлении уровней стимуляции между ARB -классами, имеющими общие области конкуренции, а также сопоставлении аффинностей антител, находящихся в данных областях, к антителам, принадлежащим конкурирующим классам. Антитело, находящееся в области конкуренции нескольких классов, определяет принадлежность к одному из них по максимальной стимуляции ARB -класса и средней аффинности к принадлежащим ему антителам. В результате работы данного оператора каждое антитело принадлежит только одному классу из множества сформированных классов.

Оператор слияния классов $ARBmerging(ARB, AB)$ используется для сокращения количества ARB -классов, имеющих минимальное количество антител. При этом данные классы поглощаются ближайшими по аффинности классами, имеющими большее количество антител и больший уровень стимуляции.

Оператор представления классов антигенам $ARBRepresentation(ARB, AG)$ используется для классификации антигенов, не имеющих специфичного антитела, т.е. антигенов, признаки которых не были восстановлены антителами в процессе иммунного обучения. Определение класса для таких антигенов происходит аналогично тому, как это делается для антител на этапе конкуренции ARB -классов.

Предлагаемый метод классификации $RLAISm$ состоит из следующих основных шагов:

1. Инициализация и нормализация данных.
 - 1.1. Формирование исходных антител.
 - 1.2. Определение значения пороговой аффинности NAT .
 - 1.3. Представление антител популяции антигенов.
2. Выделение ARB .
 - 2.1. Определение стимулирующих антител.
 - 2.2. Иммунное обучение (выполняется для каждого

стимулирующего антитела).

2.2.1. Пропорциональное клонирование.

2.2.2. Обратное-пропорциональная мутация клонов.

2.2.3. Представление антигенов клонам и конкурентно-целевой отбор клонов.

2.2.4. Старение стимулирующего антитела и антител, находящихся в его области стимуляции.

2.2.5. Проверка критерия останова.

2.3. Определение границ ARB-классов.

3. Конкуренция ARB-классов за антитела.

3.1. Определение уровней стимуляции.

3.2. Определение областей конкуренции классов.

3.3. Сравнение уровней стимуляции конкурирующих классов.

3.4. Слияние классов с минимальным количеством антител.

4. Классификация антигенов, не имеющих специфичных антител.

5. Конец.

В качестве критерия останова предлагаемого метода используется условная специфичность, определяемая максимальным количеством популяций антител, используемых для восстановления исходных антигенов, аналогично тому, как это происходит в методах [2-4].

ЭКСПЕРИМЕНТАЛЬНЫЕ ИССЛЕДОВАНИЯ

Тестирование предлагаемого метода автоматической классификации производилось на нескольких наборах исходных объектов. В первом случае, для наглядности результатов классификации, в качестве исходных данных использовалось изображение клеток крови (рис 1 а). При этом объекты представляются двумя группами признаков: координатами (x, y) и цветовыми компонентами (r, g, b).

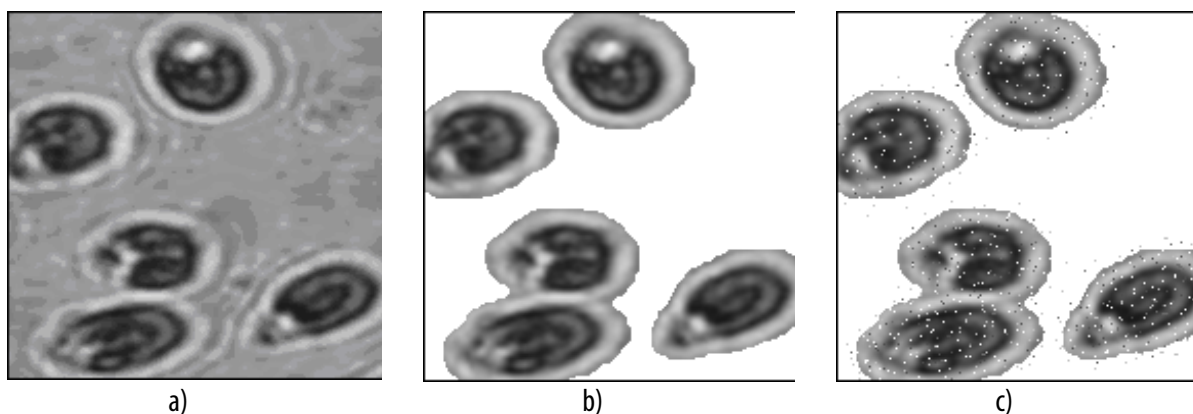


Рисунок 1 – Классификация изображения клеток крови: а) входные данные; б) результаты удаления шумов; в) результаты восстановления антигенов при классификации

Следует отметить, что перед обработкой изображения и выделения классифицируемых объектов была произведена очистка изображения от посторонних объектов и шумов, характеризующихся заданными цветовыми параметрами. Результаты удаления шумов из классифицируемого изображения приведены на рисунке 1 б. По результатам предварительной обработки исходного изображения было выделено 15 566 объектов. При классификации данные объекты формируют популяцию антигенов. Исходные антитела формируются случайным образом, при этом их количество равно количеству антигенов. Среди исходных антител определяются стимулирующие

антитела, которые подвергаются клонированию. В ходе клонирования и мутации стимулирующие антитела повышают аффинности к антигенам и становятся специфичными антителами иммунной памяти в том случае, если их признаки в ходе мутации восстановили признаки какого-либо антигена. При классификации было выделено пять классов и восстановлено 89.43% исходных антигенов. Результаты обработки исходных антигенов представлены на рисунке 1 с.

Для сравнения предлагаемого метода с другими иммунными методами, используемыми для классификации, данная задача решалась на основе методов CLOANG, BCA, aiNET и RLAIIS [3, 4]. Результаты

классификации исходного изображения перечисленными иммунными методами представлены в табл.1. При этом, при определении точности группировки для каждого метода производится сопоставление результатов классификации с эталонной классификацией группируемых объектов. В случае полного совпадения результатов данный метод получает наивысшую оценку точности –100% от эталонного разбиения.

Таблица 1 –

Результаты классификации изображения

Характеристики	CLONALG	BCA	aiNET	RLAIS	RLAISm
Время обучения	43 %	100 %	39 %	51 %	45 %
Количество классов	5	5	5	5	5
Точность	62.72 %	92.31 %	54.39 %	83.94 %	89.43 %

По результатам классификации изображения клеток крови (рис 1) для анализа производительности исследуемых методов был выделен метод, которому потребовалось больше всего времени для обработки исходных объектов. Таким методом оказался метод BCA. После этого, относительно результатов BCA, были определены временные затраты на классификацию остальными перечисленными методами. Следует отметить, что метод BCA характеризуется также и наибольшей точностью классификации. Методы CLONALG и aiNET уступают методу BCA по точности на 30-40 %, но превосходят его по быстродействию на 60-65 %. Метод RLAIIS уступает методу BCA по точности на 10 %, но превосходит по быстродействию на 50 % и является наиболее сбалансированным методом по сравнению с aiNET и CLONALG. Предлагаемый модифицированный метод RLAIISm превосходит метод RLAIIS в среднем на 6 % по каждой характеристике (быстродействие и точность), которые используются для оценки иммунных методов классификации.

Для оценки характеристик метода RLAIISm кроме изображения клеток крови, приведенного на рис. 1, использовались также и другие наборы данных, представленные в табл.2. При этом каждый набор классифицируемых данных характеризуется количеством объектов и размерностью матриц признаков. В соответствии с этим в случае, если

размерность матрицы признаков характеризуется значением 3×3 , то каждый классифицируемый объект и каждый объект обучающей выборки описывается девятью признаками.

Таблица 2 –

Характеристики наборов данных

Характеристики	Набор 1	Набор 2	Набор 3
Количество объектов (размер популяции)	500	5000	20000
Размерность матрицы признаков	3×3	6×6	9×9

Результаты классификации наборов данных, представленных в табл. 2, с помощью иммунных методов, перечисленных выше, представлены в таблице 3, в которой используются несколько условных обозначений: «Время» – характеризует быстродействие иммунного метода, т.е общее время, затраченное на классификацию исходной популяции антител; «Классы» – характеризует количество классов, выделенных в результате классификации; «Точность» – характеризует специфичность классифицируемых антител по отношению к антигенам и точность их распределения по классам. По результатам классификации различных наборов данных (табл. 2) метод BCA характеризуется наихудшим быстродействием и наилучшей точностью классификации. Использование метода aiNet характеризуется наилучшим быстродействием при невысокой точности формирования классов, при этом точность aiNET снижается при увеличении количества классифицируемых объектов, либо увеличении их размерности. Метод CLONALG сопоставим с методом aiNET по быстродействию и точности классификации также, как это имело место при классификации изображения клеток крови (табл.1). Метод RLAIIS характеризуется наилучшей сбалансированностью по основным характеристикам также, как это следует из табл. 1. По точности классификации данный метод превосходит методы aiNET и CLONALG в среднем на 15-25 % и уступает методу BCA на 3-4 %. По количеству выделенных классов результаты классификации методом RLAIIS совпадают с результатами классификации методом BCA. При этом метод RLAIIS по быстродействию уступает методам aiNET и CLONALG на

8-15 % и превосходит метод ВСА на 30-40 %. Таким образом, использование метода RLAIIS для классификации или кластеризации является более целесообразным, по отношению к другим существующим иммунным методам.

Таблица 3 –

Результаты классификации наборов данных

Метод		Набор 1	Набор 2	Набор 3
CLONALG	Время	62.8 %	54.9 %	44.6 %
	Классы	3	6	12
	Точность	71.3 %	71.0 %	71.2 %
BCA	Время	100 %	100 %	100 %
	Классы	3	5	11
	Точность	100 %	100 %	100 %
aiNET	Время	59.0 %	51.1 %	41.2 %
	Классы	3	7	16
	Точность	64.5 %	64.1 %	64.3 %
RLAIS	Время	70.4 %	67.2 %	59.5 %
	Классы	3	5	11
	Точность	95.5 %	95.7 %	95.6 %
RLAISm	Время	64.1 %	56.2 %	45.9 %
	Классы	3	5	11
	Точность	97.2 %	97.1 %	97.3 %

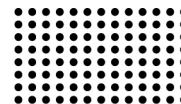
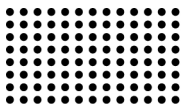
Из результатов классификации, приведенных в табл. 3 следует, что метод RLAIISm превосходит метод RLAIIS по точности и быстродействию, и совпадает с ним по количеству сформированных кластеров. При этом по точности модифицированный метод RLAIISm превосходит исходный RLAIIS в среднем на 2 % независимо от количества классифицируемых объектов и их размерности. По быстродействию метод RLAIISm превосходит метод RLAIIS на 6-12 %, при этом следует отметить, что увеличение быстродействия RLAIISm по отношению к RLAIIS прямо пропорционально увеличению количества классифицируемых объектов и увеличению количества их признаков. В сравнении с методами CLONALG и aiNET, метод RLAIISm уступает им по быстродействию в среднем на 2-4 %, но превосходит их по точности классификации на 20-30 %, независимо от количества классифицируемых объектов. Относительно метода ВСА модифицированный метод RLAIISm уступает ему по точности на 2-3%, но

значительно превосходит его по быстродействию – на 35-45 %, при этом данное соотношение наблюдается при увеличении количества классифицируемых объектов либо увеличении размерности.

Таким образом, предлагаемый метод RLAIISm сопоставим с методом ВСА по точности классификации, а по быстродействию – с методами CLONALG и aiNET. Кроме того, предложенный метод классификации превосходит исходный метод RLAIIS по быстродействию и точности в среднем на 3-4 %.

Выводы. В работе рассмотрено решение актуальной задачи классификации объектов на основе ИИС. В предложенном методе классификации RLAIISm для повышения скорости формирования классов используются концепция стимулирующих антител, что позволяет сократить количество вычислительных операций, выполняемых при клонировании, мутации и сжатии сети антител, характерных для существующих иммунных методов интеллектуальной обработки данных. Кроме того, для ускорения иммунного обучения в предложенном методе при изменении признаков клонов применяется мутация на основе аффинности дальнего предка в сочетании с конкурентно-целевым отбором клонов. Использование аффинности дальнего предка в качестве нижнего порога при определении коэффициента мутации повышает вероятность формирования клонов, специфичных целевым антигенам в ходе мутации, что приводит к сокращению количества популяций антител, необходимого для проведения иммунного обучения. Использование конкурентно-целевого отбора при клональном отборе сокращает количество вычислительных операций, характерное для других операторов редактирования популяции антител. Организация старения антител в областях стимуляции за счёт клонов стимулирующих антител повышает скорость иммунного обучения благодаря сокращению количества операций определения аффинности к антигенам на этапах мутации и клонального отбора.

Формирование результирующих ARB-классов в рассмотренном методе производится за счёт стимулирующих антител и их областей стимуляции, что не противоречит основным принципам функционирования метода RLAIIS. Таким образом использование стимулирующих антител производится



не только для сокращения вычислительных операций и повышения быстродействия метода, но и для упрощения процесса выделения начальных центров ARB-классов или кластеров. В соответствии с основными принципами работы метода RLAIIS а также других методов, функционирующих на основе модели иммунной сети, стимулирующие антитела взаимодействуют не только с антителами из областей стимуляции, но и между собой. Взаимодействие стимулирующих антител производится для организации конкуренции за общие антитела, находящиеся на пересечении их областей стимуляции. Таким образом, благодаря конкуренции между стимулирующими антителами, в результате классификации каждый иммунный объект принадлежит только одному ARB-классу или кластеру.

Из результатов проведённых исследований и

экспериментов следует, что предложенный метод RLAIISm превосходит большинство иммунных методов классификации по точности классификации и сопоставим с методом ВСА. При этом метод RLAIISm по быстродействию и скорости иммунного обучения незначительно уступает методам aiNET и CLONALG. Таким образом, предложенный метод RLAIISm характеризуется хорошей сбалансированностью по основным характеристикам по сравнению с существующими иммунными методами, реализованными на основе модели клонального отбора и иммунной сети. Следует отметить, что метод RLAIISm при незначительной модификации может использоваться не только при классификации и кластеризации данных, но и при решении задач распознавания, прогнозирования и др.

ЛИТЕРАТУРА:

1. G. Gan, C. Ma, J. Wu – «Data Clustering: Theory, Algorithms, and Applications», ASA-SIAM, Alexandria, 2007, 488 p.
2. Korablyov N.M. Klassificacia ob'ectov na osnove iscusstvennix immunnix sistem /Korablyov N.M., Fomichev A.A. //Sistemi obrobki informacij. – 2010 – Vip.6 (87). – S.13-17.
3. Dasgupta D. Immunological computation, Theory and applications /D. Dasgupta, L.F. Nino – Taylor & Francis Group, 2009. – 278 p.
4. Watkins A.A. New Classifier Based on Resource Limited Artificial Immune Systems /A. Watkins, L. Boggess //CEC '02. –2002. – Vol. 02. – P.146-151.
5. Fomichev A.A. Optimizacia operatora mutacii dl'a immunnih metodov gruppirovki danih /Fomichev A.A., Stez S.A. //Radioelektronika i molodej v XXI veke – 2013 – Tom 5. – S. 252-253.

Рецензент: д.т.н., проф. Кривуля Г.Ф.,
Харьковский национальный университет радиоэлектроники.