

УДК 004.89

Н.М. Кораблев¹, А.А. Фомичев²¹ХНУРЭ, г. Харьков, Украина, korablev@kture.kharkov.ua²ХНУРЭ, г. Харьков, Украина, alexandros.gru@gmail.ru

АВТОМАТИЧЕСКАЯ КЛАССИФИКАЦИЯ ДАННЫХ НА ОСНОВЕ МОДЕЛИ ИСКУССТВЕННОЙ ИММУННОЙ СЕТИ

В статье рассматривается решение задачи автоматической классификации данных на основе модели искусственной иммунной сети. Для повышения скорости иммунного обучения используется пропорциональная мутация на основе аффинности дальнего предка, и конкурентно-целевой отбор клонов на этапе редактирования сети. При выделении кластеров для объектов, которые не могут быть отнесены ни к одному из исходных классов, используются стимулирующие антитела, которые становятся центрами формируемых кластеров.

ИСКУССТВЕННЫЕ ИММУННЫЕ СИСТЕМЫ, КОНКУРЕНТНО-ЦЕЛЕВОЙ ОТБОР КЛОНОВ, ПРОПОРЦИОНАЛЬНАЯ МУТАЦИЯ, СТИМУЛИРУЮЩИЕ АНТИТЕЛА, ЦЕЛЕВЫЕ АНТИГЕНЫ

Введение

Одним из наиболее распространённых способов классификации, использующихся при решении практических задач, является автоматическая классификация, которая сочетает в себе принципы классификации и кластеризации. Автоматическая классификация является разновидностью кластеризации, при которой допускается использование обучающих выборок и формирование дополнительных кластеров для объектов, которые не могут быть отнесены ни к одному из классов, описанных обучающими выборками [1]. При организации самообучения в методах автоматической классификации могут использоваться различные модели организации искусственного интеллекта. Среди таких моделей выделяются модели, функционирующие на основе биологических принципов организации вычислений, в частности, теории искусственных иммунных систем (ИИС). Наиболее распространёнными иммунными моделями, используемыми при классификации, являются [2–5]: 1) модель клонального отбора (алгоритмы ВСА, CLONALG); 2) модель искусственной иммунной сети (алгоритм aiNET); 3) модель ресурсо-ограниченной иммунной системы RLAIIS. При этом алгоритм ВСА характеризуется наилучшей точностью обработки данных при низком быстродействии. Алгоритмы CLONALG и aiNET, независимо от различий в иммунных моделях, на основе которых они функционируют, сопоставимы между собой по точности и быстродействию. При этом данные алгоритмы характеризуются наилучшим быстродействием и сравнительно невысокой точностью при сопоставлении с алгоритмом ВСА. Алгоритм RLAIIS сопоставим по точности с алгоритмом ВСА, и превышает его по быстродействию, однако характеризуется высокой сложностью реализации и модификации.

Модель aiNET имеет ряд преимуществ по сравнению с моделями CLONALG и RLAIIS. Это обуславливается возможностью взаимодействия

антител между собой в процессе иммунного обучения при распознавании антигенов, что предоставляет больше возможностей для редактирования сети антител, не усложняя работу данной модели (как это происходит в RLAIIS).

В данной работе для решения задачи классификации предлагается модифицированная модель aiNETmc, которая использует принцип конкурентного целевого отбора клонов на этапе супрессии сети и принцип стимулирующих антител на этапах иммунного обучения и выделения кластеров. Использование данных подходов в модели aiNET позволяет повысить точность обработки данных на этапе иммунного обучения, и классификации без потери быстродействия, свойственного алгоритму aiNET.

1. Постановка задачи

Дано множество классифицируемых объектов и множество классов $CL\{cl_1; \dots; cl_k\}$, представленных объектами обучающей выборки $SMP\{s_1; \dots; s_m\}$. Для описания классифицируемых объектов и объектов обучающей выборки используется одинаковый формат представления данных. При классификации исходных объектов происходит восстановление объектов обучающей выборки в процессе иммунного обучения вследствие клонирования, мутации и отбора популяций классифицируемых объектов. Применительно к теории ИИС, классифицируемые объекты представляются популяцией антител $AB\{ab_1; \dots; ab_n\}$, а объекты обучающей выборки – популяцией антигенов $AG\{ag_1; \dots; ag_m\}$. В качестве меры близости между антигенами и антителами используется значение аффинности между ними [2, 3]:

$$aff(ab_i, ag_j) = (1 + d(ab_i, ag_j))^{-1}, \quad (1)$$

где $aff(ab_i, ag_j)$ – аффинность между i -м и j -м объектами, а $d(ab_i, ag_j)$ – евклидово или манхэттенское расстояние между ними. Необходимо путем модификации модели aiNET разработать иммунную

модель автоматической классификации, характеризующуюся высокой точностью классификации без снижения быстродействия.

2. Модификация aiNET для автоматической классификации объектов

Основными особенностями модели aiNET являются: 1) популяционная обработка антигенов, 2) использование множества клеток иммунной памяти и 3) использование трехуровневого механизма супрессии для редактирования иммунной сети. Кроме того, в связи с использованием оператора супрессии иммунной сети, в aiNET происходит произвольное формирование дополнительных антител. Поэтому использование модели aiNET для решения задачи автоматической классификации неэффективно, т.к. она по точности группировки объектов значительно (на 25-40 %) уступает модели ВСА. Для повышения точности модели aiNET при автоматической классификации объектов предлагается ее модификация aiNETmc. Основными особенностями предлагаемой модели aiNETmc являются:

- разделение популяции классифицируемых антител на три множества: 1) антитела, которые могут быть классифицированы без проведения иммунного обучения; 2) антитела, которые определяют принадлежность к одному из исходных классов в результате иммунного обучения; 3) антитела, для которых необходимо проведение кластеризации;
- использование принципа конкурентно-целевого отбора клонов;
- использование пропорционального клонирования с определением минимального количества клонов;
- использование обратно-пропорциональной мутации на основе аффинности дальнего предка с увеличением аффинностей при определении коэффициента мутации;
- использование супрессии клонов и клонированных антител с популяцией антигенов;
- использование принципа стимулирующих антител при кластеризации;
- отказ от проведения дополнительного разброса и трехуровневой супрессии.

Работа модели aiNETmc разделяется на три основных этапа: 1) подготовительный этап; 2) классификация объектов при использовании обучающих выборок; 3) выделение кластеров для неклассифицированных объектов.

Для упрощения представления и модификации иммунных моделей и алгоритмов в ИИС используется понятие иммунного оператора [1, 2] как некоторой функции или совокупности одинаковых действий, производимых над популяцией антител или антигенов. Среди наиболее распространенных иммунных операторов выделяются операторы

клонирования антител, клонального отбора, старения, супрессии, мутации, первичного отбора и др. [2, 3]. В соответствии с этим, процесс классификации объектов на основе предлагаемой модели aiNETmc на формальном уровне может быть описан следующим образом:

$$\begin{aligned}
 & aiNETmc(AG, AB, C, \min) = \\
 & \left[\begin{array}{l} \text{Presentation}(AG) \rightarrow \\ \rightarrow NATCalculation(AG) \rightarrow \\ \rightarrow \text{Presentation}(AG, AB) \end{array} \right]^{PREP} \rightarrow \\
 & \rightarrow \left[\begin{array}{l} \text{TrgSelection}(AG, AB') \rightarrow \\ \begin{array}{l} \left. \begin{array}{l} \text{Cloning}(ab'_i, \min) \rightarrow \\ \rightarrow Mutation(CL_i) \rightarrow \\ \rightarrow \text{Presentation}(AG_i^T, CL_i) \rightarrow \\ \rightarrow \text{Supression}(AB'_i, CL_i) \end{array} \right\} \begin{array}{l} \rightarrow \\ \rightarrow \\ \rightarrow \\ \rightarrow \end{array} \end{array} \\ \rightarrow \text{ClassSelection}(AB') \end{array} \right]^{CLASS} \rightarrow (2) \\
 & \rightarrow \left[\begin{array}{l} \text{NetPresentation}(AB'') \rightarrow \\ \rightarrow \text{StimAbSelection}(AB'_s) \rightarrow \\ \begin{array}{l} \left. \begin{array}{l} \text{Cloning}(ab'', \min) \rightarrow \\ \rightarrow Mutation(CL_i) \rightarrow \\ \rightarrow \text{Presentation}(AB'', CL_i) \rightarrow \\ \rightarrow \text{Supression}(AB'', CL_i) \end{array} \right\} \begin{array}{l} \rightarrow \\ \rightarrow \\ \rightarrow \\ \rightarrow \end{array} \end{array} \\ \rightarrow \text{ClusterSelection}(AB'_s) \end{array} \right]^{CLUST} \\
 & \rightarrow \text{NetFormation}(AB', AB''),
 \end{aligned}$$

где *PREP* – подготовительный этап классификации; *CLASS* – этап классификации с контролируемым обучением; *CLUST* – этап кластеризации неклассифицированных антител; *Presentation(AG)* – оператор определения аффинностей внутри популяции антигенов; *NATCalculation(AG)* – оператор определения критериев отбора антител и пороговых значений аффинности *NAT*; *Presentation(AG, AB)* – оператор определения аффинностей между классифицируемыми антителами и обучающими антигенами; *TrgSelection(AG, AB')* – оператор определения целевых антигенов для классифицируемых антител; *Cloning(ab'_i, min)* – оператор клонирования антител; *min* – минимально допустимое количество клонов; *Mutation(CL_i)* – оператор мутации клонов; *Presentation(AG_i^T, CL_i)* – оператор представления целевых антигенов клонам после мутации; *Supression(AB'_i, CL_i)* – оператор межпопуляционной супрессии клонов и клонированных антител; *NetPresentation(AB'')* – оператор сетевого взаимодействия антигенов, для которых проводится кластеризация; *StimAbSelection(AB'')* – оператор определения стимулирующих антител при классификации; *Cloning(ab'', min)* – оператор

клонирования неклассифицированных антител; $Pr esentation(AB'', CL_i)$ – оператор представления иммунной сети клоном; $Supression(AB'', CL_i)$ – оператор внутривыборочной супрессии клонов и клонированных антител; $ClusterSelection(AB''_s)$ – оператор определения кластеров из стимулирующих антител; $NetFormation(AB', AB'')$ – оператор формирования классифицированной иммунной сети; C – критерий завершения иммунного обучения.

Подготовительный этап является первым этапом работы модели aiNETmc, выполняемым при классификации исходных антител. На данном этапе происходит определение критериев отбора антител для разделения исходных объектов на несколько множеств антител, классификация которых производится различными способами. Следует отметить, что в случае отсутствия обучающих выборок, представленных популяциями антигенов, содержащих информацию об исходных классах, этап классификации не выполняется, и все множество исходных объектов подвергается кластеризации. Для разделения исходной популяции объектов на множестве антител используются критерий средней аффинности внутри популяции антигенов (обучающей выборки), а также средние аффинности между антигенами, принадлежащими одному исходному классу.

На этапе классификации с контролируемым обучением происходит определение классов для антител, которые могут быть классифицированы с помощью объектов обучающей выборки, при этом у данных антител средняя аффинность с популяцией антигенов сопоставима со средней аффинностью внутри популяции антигенов.

На этапе кластеризации антител, которые не могут быть классифицированы с помощью объектов обучающей выборки, происходит выделение стимулирующих антител, характеризующихся максимальными аффинностями с множеством кластеризуемых объектов. После этого за счет клонирования и мутации происходит группировка неклассифицированных объектов с выделенными стимулирующими антителами. После проведения кластеризации происходит слияние сети антител и уточнение границ классов и кластеров.

Операторы $Pr esentation(AG)$ и $NATCalculation(AG)$ используются для вычисления средней аффинности NAT_{AG} для популяции антигенов, и аффинностей NAT_{AG_i} для определения средних аффинностей между антигенами, принадлежащими общим исходным классам. Данными величинами являются предельные пороговые аффинности NAT [2, 3], которые определяются в виде:

$$NAT_{AG} = \frac{1}{m} \sum_{i=0}^m aff(ag_i, AG), \quad (3)$$

где NAT_{AG} – предельная аффинность для популяции антигенов AG (обучающей выборки), m – общее количество антигенов, а $aff(ag_i, AG)$ – средняя аффинность между i -м антигеном и остальными антигенами популяции, которая определяется в соответствии с выражением:

$$aff(ag_i, AG) = \frac{1}{m-1} \sum_{j=0, j \neq i}^m aff(ag_i, ag_j). \quad (4)$$

Кроме общего порогового значения NAT_{AG} , определяемого для всей популяции антигенов, для выделения классифицируемых антител определяются значения NAT_{AG_i} для антигенов каждого исходного класса в соответствии с (3).

После определения пороговых аффинностей NAT для популяции антигенов происходит определение аффинностей между антителами и антигенами путем вызова оператора $Pr esentation(AG, AB)$. При этом каждое антитело определяет аффинности со всеми антигенами обучающей выборки. Это происходит для того, чтобы для каждого антитела можно было выделить среднюю аффинность с популяцией антигенов $aff(ab_i, AG)$, а также средние аффинности с множествами антигенов, характеризующих исходные классы $aff(ab_i, AG_j)$. После этого, на основании выделенных аффинностей, происходит разделение исходной популяции антител на два основных множества:

- 1) антитела AB' , которые могут быть классифицированы на основе обучающей выборки;
- 2) антитела AB'' , для которых в результате кластеризации произойдет формирование дополнительных классов (кластеров).

Условием отнесения антитела к множеству объектов, классифицируемых при использовании обучающей выборки антигенов, является условие:

$$\forall ab'_i \in AB' : aff(ab'_i, AG) \cdot \lambda \geq NAT_{AG}, \quad (5)$$

где $aff(ab'_i, AG)$ – средняя аффинность i -го антитела с популяцией обучающих антигенов; λ – коэффициент роста исходных классов; NAT_{AG} – пороговая аффинность для популяции антигенов.

В случае, если условие (5) не выполняется, то антитело не может быть классифицировано с помощью обучающей выборки, и оно относится к множеству антител, для которых необходимо проведение кластеризации.

После выделения множества антител, которые могут быть классифицированы с помощью антигенов, данное множество повторно разделяется. При этом происходит выделение множества антител, которые могут быть классифицированы без обучения, и множества антител, для классификации которых необходимо проведение иммунного обучения. Антитела могут быть классифицированы без проведения дополнительного обучения при выполнении условия:

$$\forall ab_i^* \in AB^* : \exists AG_j, aff(ab_i^*, AG_j) \cdot \lambda \geq NAT_{AG_j}, \quad (6)$$

где $aff(ab_i^*, AG_j)$ – средняя аффинность i -го антигена с множеством антигенов, принадлежащих j -му классу; λ – коэффициент роста исходных классов; NAT_{AG_j} – пороговая аффинность для популяции антигенов, принадлежащих j -му классу.

Таким образом, после проведения подготовительного этапа популяция исходных антител разделяется на три множества:

- 1) антитела AB^* , которые могут быть классифицированы без проведения обучения;
- 2) антитела AB' , для классификации которых необходимо проведение иммунного обучения;
- 3) антитела AB'' , которые не могут быть отнесены ни к одному из исходных классов, представленных антигенами, для которых необходимо проведение кластеризации.

На этапе классификации *CLASS* происходит определение классов для антител из множеств AB^* и AB' . При этом для классификации множества антител AB' используется процесс иммунного обучения. Для ускорения процесса иммунного обучения в модели aiNETmc используется метод конкурентно-целевого отбора антител на этапе супрессии [6–8]. В соответствии с этим, в начале обучения происходит определение целевых антигенов для классифицируемых объектов за счет работы оператора *TrgSelection*(AG, AB'). Использование целевых антигенов повышает скорость иммунного обучения за счет сокращения количества операций определения аффинности между антителами и антигенами. При определении целевых объектов для каждого антитела определяется множество антигенов, которые характеризуются максимальными значениями аффинностей. При этом целевыми антигенами становятся все антигены, для которых аффинность с антителом удовлетворяет выражению:

$$aff(ab_i', ag_j) \geq NAT_{\min}, \quad (7)$$

где $aff(ab_i', ag_j)$ – аффинность между i -м антителом, которое может быть классифицировано с помощью антигенов обучающей выборки и j -м антигеном, NAT_{\min} – минимальная пороговая аффинность для одного из исходных классов. Следует отметить, что на подготовительном этапе классификации при определении пороговой аффинности для популяции антигенов NAT_{AG} , происходит определение пороговых аффинностей NAT_{AG_k} для антигенов каждого исходного k -го исходного класса. В соответствии с этим, NAT_{\min} определяется как минимальное значение NAT_{AG_k} одного из исходных классов.

После определения целевых антигенов для каждого классифицируемого антитела из множества AB' происходит последовательный вызов операторов клонирования, мутации, представления целевых антигенов клоном и супрессии.

Работа оператора клонирования антител *Cloning*(ab_i') заключается в формировании некоторого множества идентичных копий клонируемого объекта. В модели aiNETmc на этапе иммунного обучения используется оператор пропорционального клонирования с минимальным порогом. В соответствии с этим количество клонов, формируемых антителом, прямо пропорционально его средней аффинности с целевыми антигенами, при этом, если полученное количество клонов меньше некоторого минимально допустимого значения, то результирующее количество клонов будет равно данному минимуму. Таким образом, количество клонов для каждого клонируемого антитела определяется в соответствии с выражением:

$$n_i = \begin{cases} aff(ab_i', AG_i^t) \cdot N, & \text{при } aff(ab_i', AG_i^t) \cdot N \geq \min \\ \min, & \text{при } aff(ab_i', AG_i^t) \cdot N < \min, \end{cases} \quad (8)$$

где n_i – количество клонов, формируемых i -м антителом при клонировании; $aff(ab_i', AG_i^t)$ – средняя аффинность между клонируемым антителом ab_i' и множеством целевых антигенов AG_i^t данного антитела; N – общее количество антител в исходной популяции; \min – минимально допустимое количество клонов, формируемых антителом при клонировании.

Для изменения признаков клонов в aiNETmc используется оператор обратно-пропорциональной мутации *Mutation*(CL_i) на основе аффинностей дальнего предка. Определение коэффициента мутации на основе аффинности дальнего предка предложено в [6, 7] и применяется для повышения скорости иммунного обучения. При мутации клонов используется увеличение аффинностей при определении коэффициента мутации [3, 6, 7]. При этом аффинность дальнего предка и аффинность мутирующего клона переопределяются следующим образом:

$$aff(cl_i, AG_i^t) = \begin{cases} 2 \cdot aff(cl_i, AG_i^t), & \text{при } aff(cl_i, AG_i^t) < 0.5 \\ aff(cl_i, AG_i^t), & \text{при } aff(cl_i, AG_i^t) \geq 0.5. \end{cases} \quad (9)$$

Увеличение аффинности производится в связи с тем, что рост функции аффинности на интервале [0.5; 1] неравномерен, поэтому для повышения скорости иммунного обучения необходимо изменить принцип определения коэффициента мутации клонов.

В связи с тем, что в модели aiNETmc используется обратно-пропорциональная мутация клонов, коэффициент мутации определяется в виде:

$$\mu = rand[1 - aff(ab_i^*, AG_i^t); 1 - aff(cl_i, AG_i^t)], \quad (10)$$

где $aff(ab_i^*, AG_i^t)$ – аффинность дальнего предка клонированного i -го антитела к целевым антигенам; $aff(cl_i, AG_i^t)$ – аффинность клона к целевым антигенам.

При этом изменение признаков клона во время мутации происходит следующим образом:

$$cl_i.attr = cl_i.attr \pm \mu \cdot \left(aff(cl_i, AG_i^t)^{-1} - 1 \right), \quad (11)$$

где $cl_i.attr$ – признак i -го клона, который подвергнут действию оператора мутации; μ – коэффициент мутации, определяемый в соответствии с (10). Следует отметить, что знак \pm в выражении (11) определяется случайным образом.

После проведения мутации клонов происходит определение аффинностей между клонами и целевыми антигенами посредством вызова оператора $Presentation(AG_i^t, CL_i)$.

Для редактирования множества клонов и клонированных антител в модели aiNETmc используется оператор супрессии $Supression(AB'_i, CL_i)$ между популяциями клонов, клонированных антител и антигенов. При этом происходит добавление клонированного антитела в популяцию мутированных клонов, после чего производится отбор одного объекта, характеризующегося максимальной аффинностью с одним из целевых антигенов. Полученный в результате супрессии объект заменяет клонированное антитело, которое удаляется вместе с остальными клонами.

Оператор $ClassSelection(AB')$ используется для определения классов для антител из множеств AB^* и AB' по результатам проведенного иммунного обучения. При этом принадлежность к одному из исходных классов для каждого антитела, относящегося к одному из данных множеств, определяется по принадлежности антигена, с которым данное антитело имеет максимальную аффинность. Данный оператор завершает этап классификации с контролируемым обучением.

Первым оператором, используемым на этапе кластеризации антител, принадлежащих множеству AB'' , является оператор представления иммунной сети $NetPresentation(AB'')$. В процессе работы данного оператора происходит определение аффинностей между антителами, которые не могут быть классифицированы с помощью объектов обучающей выборки. Поэтому при классификации данных объектов происходит формирование новых классов (кластеров) без непосредственного использования популяции обучающих антигенов. После определения аффинностей между антителами множества AB'' происходит отбор стимулирующих антител путём вызова оператора $StimAbSelection(AB''_s)$. Данные объекты стимулируют неклассифицированные антитела на клонирование и мутацию для формирования из них новых классов. При этом стимулирующими антителами становятся все антитела из множества AB'' , которые характеризуются максимальными аффинностями с другими неклассифицированными антителами, однако расстояние между каждой

парой стимулирующих антител должно быть больше минимального значения NAT для одного из исходных классов, представленных популяцией антигенов.

Таким образом, полученное множество стимулирующих антител заменяет обучающие антигены для неклассифицированных антител. После этого начинается процесс иммунного обучения для популяции неклассифицированных антител, не ставших стимулирующими антителами. В соответствии с этим для каждого антитела из AB'' , не являющегося стимулирующим, последовательно выполняются клонирование, мутация клонов и супрессия.

Оператор клонирования антител $Cloning(ab'', min)$ из множества AB'' функционирует аналогично тому, как это происходит для антител, которые могут быть классифицированы с помощью объектов обучающей выборки. При представлении стимулирующих антител клоном после мутации в результате вызова оператора $Presentation(AB'', CL_i)$ происходит определение аффинности между ними. После этого, из всего множества клонов в процессе внутрипопуляционной супрессии $Supression(AB'', CL_i)$ определяется клон с максимальной аффинностью к одному из стимулирующих антител.

Кластеризация антител происходит после проведения иммунного обучения в результате работы оператора $ClusterSelection(AB''_s)$, когда для каждого антитела из множества AB'' определяется принадлежность к одному из новых классов по максимальной аффинности с одним из стимулирующих антител.

Завершающим этапом работы aiNETmc является вызов оператора формирования классифицированной сети антител $NetFormation(AB', AB'')$, которая получается путём объединения множеств AB^* , AB' и AB'' , полученных на подготовительном этапе работы модели. Таким образом, в результате работы модели aiNETmc производится не только классификация исходных объектов, но и формирование новых классов для объектов, которые не могут быть отнесены ни к одному из исходных классов.

3. Экспериментальные исследования

Тестирование предлагаемой модели классификации производилось на нескольких наборах исходных объектов. При этом наборы классифицируемых объектов различаются между собой по количеству объектов и по количеству признаков, характеризующих каждый объект. Характеристики наборов классифицируемых объектов представлены в табл. 1.

Каждый набор объектов использовался в качестве контрольной выборки для оценки точности

классификации иммунных алгоритмов. Основные различия между данными наборами состоят в количестве классов, размерности матрицы признаков и количестве объектов в наборе.

Таблица 1

Наборы классифицируемых объектов

Идентификатор	Количество элементов		Размерность матрицы признаков
	объектов	классов	
Тест 1	500	2	3×3
Тест 2	500	5	3×3
Тест 3	5000	20	6×6
Тест 4	20000	50	9×9

Для сравнения предлагаемой модели с другими иммунными моделями, применяемыми для решения задачи классификации наборов исходных объектов, приведенных в табл. 1, использовались модели CLONALG, ВСА и aiNET [2-5].

Таблица 2

Результаты классификации

Алгоритм		Тест 1	Тест 2	Тест 3	Тест 4
CLONALG	В	62.5 %	62.8 %	54.9 %	44.6 %
	К	2	5	21	53
	Т	70.9 %	71.0 %	69.2 %	64.5 %
ВСА	В	100 %	100 %	100 %	100 %
	К	2	5	20	50
	Т	99.8 %	99.7 %	99.1 %	98.7 %
aiNET	В	58.7 %	59.0 %	51.1 %	41.2 %
	К	2	5	24	46
	Т	63.9 %	64.1 %	56.5 %	52.7 %
aiNETmc	В	58.9 %	59.3 %	51.9 %	41.8 %
	К	2	5	20	50
	Т	99.1 %	99.3 %	98.5 %	97.9 %

Результаты классификации наборов данных, приведенных в табл. 1, с помощью перечисленных иммунных моделей, приведены в табл. 2, в которой используются следующие условные обозначения: «В» – характеризует быстродействие иммунного алгоритма, т.е. общее время, затраченное на классификацию исходной популяции антител; «К» – характеризует количество классов, выделенных в результате классификации; «Т» – характеризует специфичность классифицируемых антител по отношению к антигенам и точность их распределения по классам.

По результатам классификации исходных наборов данных, приведенных в табл. 1, алгоритм ВСА характеризуется наилучшим быстродействием при наилучшей точности классификации. Поэтому время, затрачиваемое на классификацию наборов данных алгоритмом ВСА, является абсолютным максимальным значением (100 %). При этом алгоритм ВСА характеризуется максимальным значением точности классификации, т.к. при обработке контрольных наборов объектов, результаты,

полученные при использовании данного алгоритма, практически полностью соответствуют эталонным результатам классификации (отклонение – 0.2 %). Поэтому ВСА является эталонным методом не только по времени работы, но и по точности классификации.

Исходя из результатов классификации наборов данных, приведенных в табл. 2, алгоритм CLONALG по быстродействию превосходит ВСА в среднем на 38-50 %, однако уступает ему по точности классификации на 30 %. Кроме того, использование CLONALG приводит к выделению неверного количества классов при увеличении количества объектов, либо увеличению количества признаков, характеризующих данные объекты. Следует отметить, что, несмотря на сравнительно невысокую точность, алгоритм CLONALG значительно превосходит эталонный алгоритм ВСА по быстродействию, что и обуславливает большую популярность CLONALG при решении различных практических задач.

При рассмотрении характеристик алгоритма aiNET, полученных в результате классификации исходных наборов данных, приведенных в табл. 2, данный алгоритм характеризуется минимальной точностью классификации, но максимальным быстродействием, т.е. для классификации указанных наборов объектов данному алгоритму требуется на 40-60 % меньше времени, чем эталонному алгоритму ВСА. Однако при использовании данного алгоритма значительно снижается точность результатов классификации (на 35 % по сравнению с ВСА). Кроме того, при увеличении количества классифицируемых объектов, либо увеличении количества их признаков, использование алгоритма aiNET приводит к повышению количества ошибок при определении количества классов для объектов, которые не могут быть отнесены ни к одному из исходных классов.

Из результатов классификации наборов объектов, приведенных в табл. 2, следует, что алгоритм aiNETmc по быстродействию сопоставим с алгоритмом aiNET, но значительно превосходит его по точности и сопоставим по данному параметру с алгоритмом ВСА. При этом по быстродействию алгоритм aiNETmc незначительно отличается от алгоритма aiNET (в среднем на 0,8-1 %), независимо от увеличения количества классифицируемых объектов, или увеличения количества признаков, характеризующих данные объекты. По быстродействию aiNETmc превосходит ВСА и CLONALG, но незначительно уступает лишь алгоритму aiNET. Следует отметить, что алгоритм aiNETmc по точности превосходит aiNET и CLONALG в среднем на 33-34 %. Таким образом, алгоритм aiNETmc является наиболее сбалансированным алгоритмом по сравнению с эталонными ВСА, CLONALG и aiNET.

Выводы

В работе рассмотрено решение актуальной задачи автоматической классификации объектов на основе иммунного подхода. Предложенная модель классификации aiNETmc является модификацией модели искусственной иммунной сети aiNET. Для повышения быстродействия модель aiNETmc использует разделение исходного набора классифицируемых объектов на несколько множеств. Это позволяет снизить временные затраты на иммунное обучение для объектов, аффинности которых позволяют их классифицировать при использовании исходного набора классов. Использование принципа конкурентно-целевого отбора на этапе супрессии клонов и клонированных антител приводит к повышению скорости иммунного обучения на этапе классификации при использовании объектов обучающей выборки.

Для повышения скорости выделения кластеров из множества объектов, которые не могут быть отнесены ни к одному из исходных классов, в модели aiNETmc используется принцип стимулирующих антител, которые характеризуются максимальными аффинностями к остальным неклассифицированным антителам. При определении стимулирующих антител используются значения средних пороговых аффинностей между антигенами исходных классов. Благодаря этому все формируемые кластера не сливаются, т.е. не содержат общих объектов между собой и с исходными классами.

Из результатов экспериментов, проведенных при тестировании aiNETmc, следует, что предложенная модель превосходит большинство существующих иммунных моделей классификации по точности и сопоставима по данной характеристике с моделью ВСА. При этом модель aiNETmc по быстродействию и скорости иммунного обучения сопоставима с моделью aiNET, которая характеризуется наилучшим быстродействием. Таким образом, предложенная модель aiNETmc характеризуется хорошей сбалансированностью по основным характеристикам по сравнению с существующими иммунными моделями, используемыми при решении задачи автоматической классификации.

Список литературы: 1. *Gan G.* Data Clustering: Theory, Algorithms, and Applications / G. Gan, C. Ma, J. Wu // ASA-SIAM, Alexandria. – 2007. – 488 p. 2. *Dasgupta D.* Immuno-

logical computation, theory and applications / D. Dasgupta, L.F. Nino – CRC Press, 2009. – 298 p. 3. *Timmis J.* An Overview of Artificial Immune Systems / J. Timmis, T. Knight, L.N. de Castro, E. Hart // Natural Computation, Springer. – 2004. – P. 55-86. 4. *Watkins A.* New Classifier Based on Resource Limited Artificial Immune Systems / A. Watkins, L. Boggess // CEC '02. – 2002. – Vol. 02. – P. 146–151. 5. *Muwang C.* Application of Artificial Immune System Approach in MRI Classification / C. Muwang, C.T. Kuo, G.H. Chang // EURASIP Journal on Advances in Signal Processing. – 2008. – P. 208-212. 6. *Кораблев Н.М.* Классификация объектов на основе искусственных иммунных систем / Н.М. Кораблев, А.А. Фомичев // Системи обробки інформації: зб. наук. праць. – 2010. – Вип. 6 (87). – С. 13-17. 7. *Фомичев А.А.* Оптимизация оператора мутации для иммунных методов группировки данных / А.А. Фомичев, С.А. Стец // Радиоэлектроника и молодёжь в XXI веке. – 2013 – Том 5. – С. 252-253. 8. *Кораблев Н.М.* Исследование иммунных операторов в задаче кластеризации объектов / Н.М. Кораблев, А.А. Фомичев // Бионика интеллекта: науч.-техн. журнал. – 2010. – №1 (72). – С. 70-74.

Поступила в редколлегию 04.07.2014

УДК 004.89

Автоматична класифікація даних на основі моделі штучної імунної мережі / М.М. Кораблєв, О.О. Фомічов // Біоніка інтелекту: наук.-техн. журнал. – 2014. – № 2 (83). – С. 100–106.

У статті досліджується проблема автоматичної класифікації даних на основі моделі штучної імунної мережі. На етапі навчання системи використовується пропорційна мутація з посиленням на основі афінності далекого предка, і конкурентно-цільовий відбір клонів на етапі редагування мережі. При виділенні кластерів для об'єктів, які не можуть бути віднесені до жодного з вихідних класів, використовуються стимулюючі антитіла, які стають центрами сформованих кластерів.

Табл. 2. Бібліогр.: 8 найм.

UDC 004.89

Automatic classification data based on the model of artificial immune network / N.M. Korabljev, A.A. Fomichev // Bionics of Intelligence: Sci. Mag. – 2014. – № 2 (83). – P. 100–106.

This paper examines the problem of automatic classification of data based on the model of artificial immune network. During training system used is proportional to the mutation with increased affinity based on distant ancestor, and competitive selection of target clones on the editing stage network. When you select a cluster of objects that can not be attributed to any of the original class are used stimulating antibodies, which are formed by the centers of the clusters.

Tab. 2. Ref. : 8 items.