

ПОВЫШЕНИЕ ЭФФЕКТИВНОСТИ ИММУННЫХ МЕТОДОВ КЛАССИФИКАЦИИ ОБЪЕКТОВ С ПОМОЩЬЮ ЦЕЛЕВОГО КЛОНАЛЬНОГО ОТБОРА

Рассматривается иммунный алгоритм классификации объектов, работающий на основе модели отбора клонов при использовании последовательного клонирования с конкурентно-целевым отбором. Основное внимание уделяется операторам мутации и первичного отбора, принципам отбора целевых антигенов. Для классификации антител, не прошедших первичный отбор, предлагается использование антител-делегатов с известной принадлежностью к классам.

1. Введение

В настоящее время в сфере моделирования искусственного интеллекта большое внимание уделяется системам обработки информации, функционирующим на основе биологических принципов организации вычислений. В связи с этим большим исследовательским интересом пользуются модели искусственных иммунных систем (ИИС) и их практические реализации. Различные иммунные модели и методы успешно применяются для решения задач классификации и кластеризации данных [1, 2]. Процесс классификации объектов на основе ИИС разделяется на два этапа: на первом этапе происходит восстановление признаков классов исходными объектами с помощью иммунных операторов клонирования и отбора клонов, на втором этапе производится определение классов для объектов, не отнесенных ни к одному классу.

Наиболее часто используемыми иммунными методами классификации являются методы, функционирующие на основе модели отбора клонов, и методы на основе модели иммунной сети [1, 2]. При применении метода CLONALG [1, 2] для классификации объектов обеспечивается высокая скорость работы при низкой точности классификации. При использовании метода ВСА [1] для классификации обеспечивается высокая точность при очень больших затратах времени. Применение метода aiNet [1, 2] характеризуется большей точностью классификации, в отличие от CLONALG, но меньшей, чем точность ВСА. Кроме того, метод aiNet является более медленным по сравнению с методом CLONALG.

В [3, 4] предложен метод классификации, основным отличием которого является приоритетное последовательное клонирование с конкурентно-целевым отбором клонов, что приводит к сокращению избыточных вычислений, присущих перечисленным иммунным методам. Это приближает его по точности к методу ВСА, а по времени, затрачиваемого на классификацию – к методу aiNet. Основными недостатками метода, предложенного в [3, 4], является высокая сложность реализации, не оптимальная мутация, которая приводит к увеличению времени классификации. Предлагаемый в работе метод иммунной классификации объектов является модификацией метода, предложенного в [3, 4]. Изменениям подверглись оператор мутации и способ определения целевых антигенов, что приводит к повышению точности и сокращению времени классификации.

2. Постановка задачи

Дано множество классифицируемых объектов, представленных популяцией антител $AB\{ab_1; \dots; ab_m\}$, и множество классов, представленных популяцией антигенов $AG\{ag_1; \dots; ag_n\}$. Форматом представления антител и антигенов могут быть матрицы признаков, либо множество векторов признаков переменной длины, при этом каждый столбец матрицы (каждый вектор признаков) представляет собой отдельную группу признаков антигена или антитела. Каждая группа признаков характеризуется длиной, т.е. количеством признаков в группе, и весом, т.е. степенью важности данной группы признаков при определении принадлежности к какому-либо классу. При этом веса и длины соответствующих групп признаков одинаковы как для антител, так и для антигенов.

В качестве меры близости между антигенами и антителами используется аффинность [1, 2]:

$$\text{aff}(\text{ag}_i; \text{ab}_j) = (1 + d(\text{ag}_i; \text{ab}_j))^{-1}, \quad (1)$$

где $\text{aff}(\text{ag}_i; \text{ab}_j)$ – аффинность между i -м антигеном и j -м антителом; $d(\text{ag}_i; \text{ab}_j)$ – евклидово расстояние между ними.

Следует отметить, что определение аффинности с помощью (1) применяется только для метрических признаков. Кроме того, выражение (1) используется в простейшем случае, когда классы и объекты характеризуются только одной группой признаков – одним вектором признаков. В случае, когда применяются матрицы признаков, либо совокупность векторов переменной длины, аффинность между антигенами и антителами определяется следующим образом:

$$\text{aff}(\text{ag}_i; \text{ab}_j) = \sum_{c=1}^k w_c (1 + d(\text{ag}_i; \text{ab}_j)_c)^{-1}, \quad (2)$$

здесь k – общее количество групп признаков, т.е. количество столбцов в матрице или векторов переменной длины; w – вес группы признаков; $d(\text{ag}_i; \text{ab}_j)_c$ – евклидово расстояние между признаками данной группы для i -го антигена и j -го антитела.

Цель исследования. Необходимо разработать иммунный метод классификации объектов, устраняющий проблему избыточности вычислений, характерной для методов ВСА [1] и aiNET [1, 2], и проблему больших временных затрат на классификацию, присущую методу ВСА, без потери точности определения классов.

3. Метод иммунной классификации

Предлагаемый метод классификации объектов является модификацией метода, предложенного в [3, 4]. Изменениям подверглась последовательность вызовов иммунных операторов и принципы организации их работы. Предлагаемый метод назван алгоритмом целевого клонального отбора (targeting clonal selection algorithm – TACSA). Формально на уровне иммунных операторов метод TACSA описывается следующим образом:

$$\begin{aligned} \text{TACSA}(\text{AG}, \text{AB}, p) = & \text{AgPresentation}(\text{AG}, \text{AB}) \rightarrow \text{FirstSelection}(\text{AB}) \rightarrow \text{TrgDetection}(\text{AB}) \rightarrow \\ & \rightarrow \text{AB}[\text{Cloning}(\text{ab}, \text{CL}) \rightarrow \text{Mutation}(\text{CL}) \rightarrow \text{TrgPresentation}(\text{ab}_{\text{TRG}}, \text{CL}) \rightarrow \\ & \rightarrow \text{ClonSelection}(\text{CL}) \rightarrow \text{Apoptosis}(\text{ab}, \text{cl})] \text{TermTest}(p), \end{aligned} \quad (3)$$

где $\text{AgPresentation}(\text{AG}, \text{AB})$ – оператор представления антигенов антителам для определения аффинностей; $\text{FirstSelection}(\text{AB})$ – оператор первичного отбора антител; $\text{TrgDetection}(\text{AB})$ – оператор определения целевых антигенов и антител-делегатов; $\text{Cloning}(\text{ab}, \text{CL})$ – клонирование антител, прошедших первичный отбор; $\text{Mutation}(\text{CL})$ – оператор мутации клонов; $\text{TrgPresentation}(\text{ab}_{\text{TRG}}, \text{CL})$ – оператор представления целевых антигенов клонам; $\text{ClonSelection}(\text{CL})$ – конкурентный целевой отбор клонов; $\text{Apoptosis}(\text{ab}, \text{cl})$ – оператор старения антител; $\text{TermTest}(p)$ – процедура проверки критерия останова.

Завершение работы метода TACSA может происходить при выполнении одного из следующих условий: 1) достижение полной специфичности – состояния, при котором все антитела стали специфичными какому-либо антигену в ходе клонирования и мутации; 2) достижение предельного количества популяций антител. Специфичным антигену является антитело, которое в ходе мутации полностью приняло значения признаков антигена. В методе TACSA для повышения скорости обработки популяции антител используются антитела-делегаты. Под антителом-делегатом понимается антитело с хорошей аффинностью к антигенам, которое представляет группу ближайших к нему антител с худшей аффинностью к антигенам. Следует отметить, что операторы представления антигенов, первичного отбора и определения целевых антигенов и антител-делегатов вызываются только один раз для исходной популяции антител и не используются в дальнейшем.

Оператор представления антигенов $\text{AgPresentation}(\text{AG}, \text{AB})$ реализует несколько функций: определение критерия первичного отбора антител и определение аффинностей между

антигенами и антителами (2). Критерием отбора антител является значение средней аффинности между антигенами $CSel$ [3, 4]:

$$CSel = \frac{1}{n} \sum_{i=1}^n aff(AG, ag_i), \quad (4)$$

где $aff(AG, ag_i)$ – аффинность i -го антигена со всеми остальными антигенами, определяемая в соответствии с (2).

Оператор первичного отбора $FirstSelection(AB)$ вызывается для выделения антител исходной популяции, у которых аффинность с популяцией антигенов удовлетворяет следующему условию [3, 4]:

$$aff(AG, ab_i) \geq f \cdot CSel, \quad (5)$$

здесь $aff(AG, ab_i)$ – аффинность i -го антитела с популяцией антигенов AG , f – коэффициент первичного отбора антител, значения которого предлагается выбирать в диапазоне $[0,85; 0,95]$.

Оператор определения целевых антигенов и антител-делегатов $TrgDetection(AB)$ используется для выделения целевых антигенов для антител, прошедших первичный отбор на основании аффинностей, определённых на этапе представления антигенов антителам. Антиген становится целевым для антитела и его клонов в случае выполнения следующего условия:

$$aff(ag_i, ab_j) \geq r \cdot CSel, \quad (6)$$

где $aff(ag_i, ab_j)$ – аффинность i -го антигена с j -м антителом; r – коэффициент сокращения области поиска целевых антигенов, значения которого предлагается выбирать в диапазоне $[1,5; 1,9]$. Если антитело не имеет ни одного целевого антигена – оно не проходит первичный отбор. Для всех антител, не прошедших первичный отбор, выбирается антитело-делегат из числа антител, прошедших первичный отбор. Таким образом, в результате первичного отбора и определения целевых антигенов исходная популяция антител разделяется на две части: множество антител-делегатов, имеющих хотя бы один целевой антиген, над которыми будут производиться дальнейшие действия (клонирование, мутация, старение), и множество антител, не прошедших первичный отбор либо не имеющих ни одного целевого антигена и определивших антитело-делегат по наилучшей аффинности. Антитела, не прошедшие этап первичного отбора, не подвергаются клонированию, мутации и действиям других иммунных операторов. Их принадлежность к какому-либо классу определяется по результатам классификации их антител-делегатов, т.е. антитело, не прошедшее этап первичного отбора, принадлежит тому же классу, что и его антитело-делегат. Антитело-делегаты определяются таким же образом, как и целевые антигены. В случае, если антитело не первичного отбора не может определить антитело-делегат в соответствии с (6), выбор антитело-делегата осуществляется по наилучшей аффинности с каким-либо антителом, прошедшим первичный отбор и имеющим целевые антигены. Использование антител-делегатов сокращает объем избыточных вычислительных операций, возникающих на этапе клонирования и мутации в других иммунных методах.

Оператор клонирования $Cloning(ab, CL)$ используется для клонирования антител, прошедших первичный отбор, при этом применяется пропорциональное клонирование [1], при котором количество клонов, создаваемых антителом, прямо пропорционально его средней аффинности с целевыми антигенами. Следует отметить, что антитела, ставшие специфичными какому-либо антигену, не подвергаются действию операторов клонирования и старения. В других иммунных методах [1, 2] такие антитела называют клетками памяти.

Оператор мутации $Mutation(CL)$ используется для внесения изменений в признаки клонов антител для достижения состояния специфичности клонов какому-либо целевому антигену. При этом важную роль играет коэффициент мутации, который определяет характер изменений признаков клонов. В существующих иммунных методах [1, 2] наилучший

эффект даёт обратно пропорциональная мутация, при которой коэффициент мутации определяется следующим образом:

$$\mu = \text{rand}(0; 1 - \text{aff}(\text{AG}; \text{ab}_i)), \quad (7)$$

где $\text{aff}(\text{AG}; \text{ab}_i)$ – аффинность i -го клонированного антитела с популяцией антигенов.

Данный подход не оптимален, поскольку даже при использовании направленной мутации [1] достижение состояния специфичности клонами во время мутации маловероятно, что обуславливается нулевой нижней границей определения коэффициента мутации. Суть данной проблемы заключается в том, что в случае, если значение коэффициента мутации будет близко к нижней границе – изменения в клонах при мутации будут очень незначительными, в результате чего для достижения состояния специфичности потребуется большое количество популяций антител, что в свою очередь приведёт к увеличению вычислений.

В предлагаемом методе TACSA нижняя граница при определении коэффициента мутации является значением аффинности антитела-предка из предшествующего поколения антител. Определим такую нижнюю границу коэффициента мутации как аффинность дальнего предка. При этом коэффициент мутации предлагается определять в соответствии с выражением:

$$\mu = \text{rand}(1 - \text{aff}(\text{ab}_{\text{TRG}}; \text{ab}'_i); 1 - \text{aff}(\text{ab}_{\text{TRG}}; \text{ab}_i)), \quad (8)$$

где $\text{aff}(\text{ab}_{\text{TRG}}; \text{ab}'_i)$ – аффинность i -го антитела, породившего клонированное антитела текущей популяции с целевыми антигенами, а $\text{aff}(\text{ab}_{\text{TRG}}; \text{ab}_i)$ – аффинность i -го антитела, для клонов которого определяется коэффициент мутации. Следует отметить, что поскольку для клонов, сформированных исходной популяцией антител, дальних предков не существует, при определении коэффициента мутации используется нулевая нижняя граница. Применение аффинностей дальних предков при определении коэффициента мутации начинается со второй популяции антител до завершения работы TACSA. Таким образом, решается проблема предельно малых значений коэффициента мутации, характерная для других иммунных методов [1, 4].

Оператор представления целевых антигенов $\text{TrgPresentation}(\text{ab}_{\text{TRG}}, \text{CL})$ используется для определения аффинностей между мутированными клонами и целевыми антигенами. Таким образом, вместо всей популяции антигенов во взаимодействие с клонами вступают только небольшая группа целевых антигенов, что приводит к снижению количества вычислительных операций.

Оператор $\text{ClonSelection}(\text{CL})$ используется для проведения отбора клонов и определения одного клона с наилучшей аффинностью с одним из своих целевых антигенов. Механизм конкурентно-целевого отбора описан в [3, 4] и не подвергается модификации в TACSA.

Оператор старения $\text{Apoptosis}(\text{ab}, \text{cl})$ используется для замены клонированного антитела клоном, оставшимся в результате конкурентно-целевого отбора. Клонированное антитело заменяется своим клоном в том случае, если его аффинность хуже аффинности клона. В противном случае клон удаляется, и антитело остается для клонирования на следующей популяции.

Процедура проверки критерия останова $\text{TermTest}(p)$ применяется для завершения работы метода, которое может произойти в случае достижения предельного количества популяций антител, либо достижения состояния полной специфичности антител.

Следует отметить, что в TACSA как и в BSA каждое антитело обрабатывается последовательно, т.е. для каждого антитела, прошедшего процедуру первичного отбора, вызывается последовательность операторов клонирования, мутации, представления целевых антигенов клонам, отбора клонов и старения. В методах CLONALG [1] и aiNet [1] происходит обработка всей популяции антител, что повышает сложность их реализации, увеличивает количество вычислительных операций и увеличивает время, затрачиваемое на классификацию.

Процесс классификации с помощью метода TACSA представляется следующим образом:

1. Определение критерия отбора антител (4).
2. Первичный отбор антител (5).
3. Определение целевых антигенов и антител-делегатов (6).
4. Для каждого антитела, прошедшего первичный отбор, выполняется:
 - 4.1. Пропорциональное клонирование.
 - 4.2. Мутация клонов на основе аффинности предыдущих поколений (8).
 - 4.3. Представление целевых антигенов клонам.
 - 4.4. Отбор клонов.
 - 4.5. Старение клонированного антитела.
5. Проверка критерия останова, в случае его достижения – переход к 6, иначе переход к 4.
6. Конец.

В результате работы метода TACSA для множества исходных классифицируемых объектов принадлежности к классам определяются по одному из перечисленных ниже способов:

- специфичные антигенам антитела принадлежат тому же классу, что и антигены;
- не специфичные антигенам антитела принадлежат тому классу, к которому относится большинство его целевых антигенов;
- антитела, не прошедшие процедуру первичного отбора, относятся к тем же классам, что и их антитела-делегаты.

Таким образом, всё множество исходных объектов будет классифицировано независимо от прохождения процедуры первичного отбора.

Экспериментальные результаты

Экспериментальные исследования производилась для методов ВСА [1], aiNET [1, 2], CLONALG [1, 2], метода, предложенного в [3, 4], предлагаемого метода TACSA, а также метода kNN [5] с количеством ближайших соседей, равным 3. При исследовании использовались три набора входных данных: первый набор состоял из 100 объектов и 5 классов, в котором каждый объект представлялся вектором из 5 признаков; второй набор – из 300 объектов и 10 классов, в котором объекты представлялись матрицей 3×5 признаков; третий набор содержал 500 объектов, представленных матрицами 5×5 признаков и 15 классами.

В качестве критерия останова использовалось предельно допустимое количество популяций антител (100). Результаты классификации приведены в процентном отношении к ее эталонным результатам. Это значит, что для каждого набора классифицируемых объектов существует заранее известный результат классификации, который является эталонным по точности (100%) для исследуемых методов. При оценке быстродействия среди исследуемых методов был выбран самый медленный, который и является эталонным (100%) по отношению к остальным (таким является метод ВСА). После этого для остальных методов классификации проводилось сопоставление их времени работы с временем работы эталонного метода, на основе которого формируется конечное значение. Результаты классификации по точности и временным затратам приведены в таблице.

Таблица

Метод	I набор данных		II набор данных		III набор данных	
	Точность, %	Время, %	Точность, %	Время, %	Точность, %	Время, %
kNN(3)	93	12,78	91	12,52	88	12,39
CLONALG	62	41,24	61	42,01	52	41,67
BCA	95	100,00	94	100,00	91	100,00
aiNET	78	76,39	79	75,98	76	76,03
метод [4,5]	92	65,17	91	64,81	89	65,29
TACSA	96	56,11	96	55,83	95	56,42

По результатам классификации различными методами видно, что иммунные методы классификации требуют для работы больше времени, чем метод kNN, однако некоторые из них превосходят его по точности даже на небольших наборах данных. Среди иммунных

методов наибольшим быстродействием выделяется CLONALG, однако он отличается также и наименьшей точностью при определении классов. ВСА является наиболее медленным методом классификации, однако он даёт очень высокую точность определения принадлежности объектов к классам. Метод, предложенный в [3, 4], по точности классификации сопоставим с методом ВСА, но превосходит его по быстродействию на 35-40 %. Предлагаемый метод ТАССА превосходит по быстродействию и по точности большинство иммунных методов, использующихся для классификации. Кроме того, по точности работы метод ТАССА превосходит метод kNN в среднем на 4-8%, особенно при работе с большими наборами объектов, и сопоставим только с ВСА.

Выводы

Научная новизна. В предложенном методе классификации объектов ТАССА на основе искусственных иммунных систем используются новые подходы к мутации и классификации объектов, наиболее удалённых от классов. Благодаря применению аффинностей дальних предков при определении коэффициента мутации повышается быстродействие метода за счёт сокращения количества популяций антител и объёмов вычислений при определении принадлежности объектов классам. Для объектов, располагающихся на большом расстоянии от классов, используются антитела-делегаты, которые выбираются из множества объектов, прошедших процедуру первичного отбора и имеющих целевые антигены. Принадлежность объектов к классам, не прошедших первичный отбор, определяется принадлежностью их антител-делегатов, что ведет к сокращению объёмов вычислений и ускорению классификации объектов. Кроме того, предложенный метод ТАССА более прост в реализации и модификации для решения различных практических задач, что устраняет недостатки метода, изложенного в [3, 4].

Предложенный метод классификации по точности и скорости работы превосходит большинство иммунных методов, являющихся наиболее часто используемыми при решении различных практических задач, в том числе задач классификации и кластеризации данных.

Список литературы: 1. *Dasgupta D.*, Immunological computation, theory and applications / D. Dasgupta, L. F. Nino – CRC Press, 2009. 298 p. 2. *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. 3. *Кораблёв Н.М.* Классификация объектов на основе искусственных иммунных систем / Н.М. Кораблёв, А.А. Фомичёв // Системы обработки информации. 2010. № 6 (87). С. 13-17. 4. *Кораблёв Н.М.* Исследование иммунных операторов в задаче кластеризации объектов / Н.М. Кораблёв, А.А. Фомичёв // Бионика интеллекта. 2010. № 1 (72). С. 70-74. 5. *Fayyad U.M.* Advances in knowledge discovery and data mining, / U.M. Fayyad, G.P. Shapiro / AAAI Press, 1996. 356 p.

Поступила в редколлегию 16.12.2012

Кораблёв Николай Михайлович, д-р техн. наук, профессор кафедры ЭВМ ХНУРЭ. Научные интересы: интеллектуальная обработка информации на основе искусственных иммунных систем. Увлечения и хобби: футбол, плавание, танцы. Адрес: Украина. 61000, Харьков. пр Ленина, 14, тел. (050)-530-85-72, email: korablev@kture.kharkov.ua

Фомичёв Александр Александрович, ассист. кафедры ЭВМ ХНУРЭ. Научные интересы: моделирование искусственных иммунных систем. Увлечения: музыка, танцы. Адрес: Украина. 61000, Харьков. пр Ленина, 14, тел. (093)-332-11-73, email: alexandros_1985@mail.ru