



ПРОГРАММНО-ТЕХНИЧЕСКИЕ КОМПЛЕКСЫ

Ан.А. ОЛЕЙНИК, Ал.А. ОЛЕЙНИК, С.А. СУББОТИН

УДК 004.93

АГЕНТНЫЕ ТЕХНОЛОГИИ ДЛЯ ОТБОРА ИНФОРМАТИВНЫХ ПРИЗНАКОВ

Ключевые слова: агент, коллективный интеллект, мультиагентный метод, оптимизация, отбор признаков.

ВВЕДЕНИЕ

При решении задач распознавания, прогнозирования, классификации, диагностики и управления возникает необходимость построения математических моделей по данным обучающей выборки [1]. Включение в синтезируемую модель малоинформативных (с низкими оценками значимости) или избыточных (несущих избыточную информацию) признаков приводит к повышению ее сложности и снижению адекватности, что, в свою очередь, увеличивает время вычисления выходного параметра построенной модели, понижает вероятность принятия правильных решений с ее помощью, а также снижает ее интерпретируемость [1, 2]. Поэтому до момента построения моделей, описывающих исследуемые зависимости, на практике целесообразно понижать размерность признакового пространства, что позволит сократить время построения модели, повысить ее адекватность и восприятие пользователем, а также исключить необходимость измерения величин, не влияющих на выходной параметр.

Цель настоящей работы — разработка метода отбора информативных признаков, основанного на мультиагентном подходе и позволяющего выполнять сокращение размерности пространства признаков для построения моделей исследуемых объектов, процессов или систем.

ПОСТАНОВКА ЗАДАЧИ

Пусть задана обучающая выборка исходных данных в виде $\langle X = \{X_1, X_2, \dots, X_L\} = \{X_i\}, Y = \{y_1, y_2, \dots, y_m\} = \{y_p\} \rangle$, где X — исходный набор значений признаков, характеризующих рассматриваемый объект или процесс; Y — массив значений выходного параметра в заданной выборке; $X_i = \{x_{ip}\}$ — i -й признак в исходной выборке, $i = 1, 2, \dots, L$, x_{ip} — значение i -го признака для p -го экземпляра (наблюдения, образа) выборки, $p = 1, 2, \dots, m$; y_p — значения прогнозируемого параметра для p -го экземпляра; L — общее количество признаков в исходном наборе; m — количество экземпляров выборки.

Тогда постановка задачи отбора информативных признаков [1, 3] может быть представлена следующим образом (идеализированная постановка задачи отбора признаков): выделить комбинацию признаков X^* из исходного массива данных $\langle X, Y \rangle$, при которой достигается минимум заданного критерия оценивания J набора признаков:

$$J(X^*) = \min_{X \in XS} J(Xe),$$

© Ан.А. Олейник, Ал.А. Олейник, С.А. Субботин, 2012

где Xe — элемент множества XS ; $J(Xe)$ — критерий оценивания значимости набора признаков элементов Xe ; XS — множество всех возможных комбинаций признаков, полученное из исходного набора признаков X .

Как правило, используется идеализированная постановка задачи отбора информативных признаков. Тем не менее существует также классическая постановка, при которой необходимо отобрать из множества исходных L признаков комбинацию, состоящую не более чем из L_0 признаков ($L_0 < L$) и обеспечивающую достижение оптимума заданного критерия:

$$J(X^*) = \min_{Xe \in XS, |Xe| \leq L_0} J(Xe),$$

где $|Xe|$ — количество элементов во множестве Xe .

Кроме того, задача отбора информативных признаков может заключаться в поиске набора признаков минимального размера, обеспечивающего достижение заданного значения критерия оценивания значимости набора признаков:

$$|X^*| = \min_{Xe \in XS, J(Xe) < \varepsilon} |Xe|,$$

где ε — заданное значение критерия оценивания набора признаков J .

Таким образом, полученный в результате отбора признаков оптимальный набор X^* , обладая достаточной информативностью, наиболее полно отражает исследуемый объект или процесс. При этом из исходного набора X исключаются [2, 4]:

- незначимые признаки, не влияющие на выходной параметр;
- избыточные признаки, значения которых зависят от других признаков.

Такие признаки не приводят к улучшению качества прогнозирования по синтезируемой модели.

АНАЛИЗ МЕТОДОВ ОТБОРА ИНФОРМАТИВНЫХ ПРИЗНАКОВ

В настоящее время для отбора информативных признаков используются различные методы: полного перебора, эвристические, отбора признаков по максимуму значения критерия индивидуальной значимости и др.

Однако применение методов полного и сокращенного перебора [3] требует оценки значительного количества комбинаций признаков, составленных из исходной совокупности, что делает невозможным использование такого подхода при большом количестве признаков в исходном наборе ввиду огромных вычислительных затрат.

Методы эвристического поиска [4] недостаточно эффективны из-за неоптимальности жадной стратегии поиска, последовательно добавляющей или удаляющей по одному признаку, в результате чего получаемый набор признаков содержит избыточные признаки, коррелирующие с другими признаками в наборе. Кроме того, при отборе признаков из массива данных большой размерности эвристический поиск также требует значительных затрат на оценку наборов признаков.

Поскольку при решении задач оценивания и классификации, как правило, приходится оперировать с системой статистически зависимых признаков, информативность набора которых не выражается через информативность отдельных признаков [1, 3], зачастую нецелесообразно использовать ранжирование признаков по вычисленной индивидуальной оценке.

Для выделения комбинаций информативных признаков целесообразно использовать стохастические методы, которые в отличие от приведенных выше детерминированных более эффективно исследуют пространство поиска, затрачивая при этом существенно меньше времени, чем методы полного и сокращенного перебора в случаях большого количества признаков в обучающей выборке. Такие методы не склонны к закликиванию в локальных оптимумах и не выдвигают дополнительных требований к виду целевой функции.

К методам стохастического поиска относятся мультиагентные методы интеллектуальной оптимизации, которые имеют бионическую природу, т.е. моде-

лируют поведение живых существ, характеризующихся коллективным интеллектом [2, 5, 6]. К мультиагентным методам интеллектуальной оптимизации относятся метод муравьиных колоний (Ant Colony Optimization, ACO) [7], метод пчелиной колонии (Bee Colony Optimization, BCO) [8], метод оптимизации с группировкой частиц (Particle Swarm Optimization, PSO) [5], метод оптимизации на основе моделирования перемещения бактерий (Bacteria Foraging Optimization, BFO) [9]. Данные методы уже широко применяются при решении различных задач оптимизации [5–11].

Однако известные мультиагентные методы разрабатывались, как правило, для решения оптимизационных задач определенного типа, например метод муравьиных колоний создавался для решения задачи коммивояжера. Кроме того, модели поведения живых существ, которые лежат в основе данных методов, — неполные и не вполне отвечают своим естественным аналогам. В частности, недостаточно эффективен механизм воспроизводства (создания нового множества решений), что приводит к продолжительным расчетам при выполнении оптимизационного процесса.

Поэтому актуальной является разработка метода отбора информативных признаков, основанного на мультиагентном подходе и не имеющего недостатков существующих методов.

МУЛЬТИАГЕНТНЫЙ МЕТОД С КОСВЕННОЙ СВЯЗЬЮ МЕЖДУ АГЕНТАМИ

Как отмечалось выше, одним из мультиагентных методов является метод оптимизации на основе моделирования перемещения бактерий BFO, который применялся для решения различных задач оптимизации: поиска оптимального потока энергии, анализа независимых компонентов, настройки пропорционально-интегрально-дифференциального (ПИД) контроллера, задачи динамического распределения ресурсов и др. [1, 5, 9–11]. Результаты проведенных экспериментов по решению тестовых задач [1, 9–11] показали, что применение метода BFO эффективно и перспективно, поскольку полученные с его помощью результаты характеризуются более высокой или достаточной точностью по сравнению с традиционными методами.

Метод BFO основан на моделировании поведения бактерии *E. Coli* [9]. Данная бактерия может перемещаться прямолинейно, кувыряться, скользить, участвовать в воспроизводстве [9]. Эти и некоторые другие аспекты существования бактерий *E. Coli* легли в основу метода BFO. С помощью этого метода также моделируется процесс хемотаксиса, представляющий собой двигательную реакцию бактерии в ответ на появление в среде аттрактанта (вещество, привлекающее бактерии) или репеллента (вещество, отпугивающее бактерий) [9, 10]. В естественных условиях аттрактантами являются полезные для бактерий вещества, а репеллентами — вредные.

Метод BFO предназначен для поиска минимума функции $J(X)$, $X \in R^P$, при неизвестном градиенте $\nabla J(X)$, где X — позиция агента (бактерии) в пространстве поиска R^P , а с помощью $J(\theta)$ моделируются полезные и вредные свойства среды, т.е. $J(X)$ характеризует местонахождение аттрактантов и репеллентов. Таким образом, $J < 0$, $J = 0$, $J > 0$ означает, что агент находится в полезной, нейтральной или вредной среде соответственно.

Пусть $P(j, k, l) = \{X^i(j, k, l), i = 1, 2, \dots, S\}$ описывает позицию каждого члена популяции S агентов на j -м хемотаксическом шаге, k -м шаге воспроизводства и на l -м событии исключения-рассеивания; d_{attract} — глубина аттрактанта (полезных веществ) клетки и w_{attract} — мера ширины аттрактанта. Клетки взаимно исключаются посредством локального потребления и за счет того, что не являются пищей одна для другой; $h_{\text{repellent}} = d_{\text{attract}}$ — концентрация вредного вещества (репеллента) и $w_{\text{repellent}}$ — мера ширины репеллента. Таким образом, можно ис-

пользовать функции $J_{cc}^i(X)$, $i=1,2,\dots,S$, для моделирования сигналов между клетками посредством выделения агентами аттрактанта и репеллента [9, 10]:

$$J_{cc}(X) = \sum_{i=1}^S J_{cc}^i = \sum_{i=1}^S \left[-d_{\text{attract}} \exp \left(-w_{\text{attract}} \sum_{j=1}^p (X_j - X_j^i)^2 \right) \right] + \sum_{i=1}^S \left[h_{\text{repellent}} \exp \left(-w_{\text{repellent}} \sum_{j=1}^p (X_j - X_j^i)^2 \right) \right],$$

где $X = [x_1, \dots, x_p]^T$ — точка в пространстве оптимизации.

Особенностями такого подхода к расчету $J_{cc}(X)$ являются:

- значение $J_{cc}(X)$ не зависит от значения целевой функции в точке θ ;
- за счет варьирования коэффициентов w_{attract} и $w_{\text{repellent}}$ можно влиять на поведение популяции агентов в процессе поиска: при малых значениях w_{attract} и больших значениях $w_{\text{repellent}}$ популяция склонна к большому рассеиванию в пространстве поиска, в противном случае, т.е. при больших значениях w_{attract} и малых значениях $w_{\text{repellent}}$, популяция склонна к группировке в небольших областях поиска.

Очевидно, что сила выделения химических веществ бактериями зависит от окружающей среды, т.е. бактерия, находящаяся в среде с высокой концентрацией полезных веществ, будет выделять более сильный аттрактант, чем та же бактерия в среде с низкой концентрацией полезных веществ. Поэтому в данном методе используется функция $J_{ar}(\theta)$ для моделирования взаимодействия между ячейками с учетом особенностей окружающей среды: $J_{ar}(X) = \exp(M - J(X))J_{cc}(X)$, где M — настраиваемый параметр. Следовательно, для поиска оптимума необходимо минимизировать выражение $(J(i, j, k, l) + J_{ar}(X^i(j, k, l)))$, что позволит агентам (бактериям) находить полезные вещества, избегая областей с вредными веществами и в то же время незначительно приближаясь к другим ячейкам.

Однако в методе BFO недостаточно эффективен механизм воспроизводства, который предназначен для образования нового множества решений, что приводит к продолжительным расчетам при выполнении оптимизационного процесса.

Для устранения указанного недостатка в разработанном мультиагентном методе с косвенной связью между агентами предлагается расширить модель метода BFO путем внедрения генетических операторов отбора, скрещивания и мутации. Для этого предлагается использовать такие генетические операторы:

- для отбора — пропорциональный отбор;
- для скрещивания — случайный выбор родительской пары с арифметическим скрещиванием;
- для мутации — простая мутация.

В рассмотренном методе взаимосвязь между агентами происходит за счет моделирования косвенной связи между ними, поскольку агенты моделируют выделение вещества в зависимости от качества среды, после чего следующие агенты при своем перемещении учитывают свойства среды, которые, в свою очередь, зависят как от объективных факторов (значение целевой функции), так и от выделенных другими агентами веществ. Таким образом, связь между агентами происходит не за счет непосредственного обмена информацией между ними, а с помощью информации, получаемой ими через окружающую среду.

С учетом аналогии с естественными процессами, а также предложенных модификаций разработанный мультиагентный метод с косвенной связью между агентами предусматривает выполнение таких этапов:

- инициализация;
- моделирование хемотаксиса;
- воспроизводство;
- исключение и рассеивание;

- проверка критериев останова;
- перезапуск агентов.

Таким образом, предложенный метод содержит этапы, которые свойственны известным методам оптимизации (инициализация и проверка критериев останова), мультиагентным методам (воспроизводство и перезапуск агентов) и естественным процессам перемещения бактерий (моделирование хемотаксиса, исключение и рассеивание).

При инициализации в разработанном методе задаются параметры, влияющие на его работу: S — количество агентов, N_{re} — количество шагов воспроизводства, N_s — количество шагов-повторений на одном хемотаксическом шаге, N_c — количество хемотаксических шагов, N_{ed} — количество событий исключения-рассеивания; P_{ed} — вероятность рассеивания. Затем случайным образом распределяются начальные значения X^i , $i = 1, 2, \dots, S$, в пространстве поиска и рассчитываются начальные значения целевой функции для каждого агента J^i . Затем для каждого из них выполняется пересчет: $J^i = J^i + J_{ar}(X^i)$.

На этапе моделирования хемотаксиса для каждого агента выполняются операторы кувыркивания, перемещения и скольжения. Данный этап может быть выполнен по следующему алгоритму.

Шаг 1. Установить: $i = i + 1$.

Шаг 2. Кувыркивание. Моделирование кувыркивания достигается за счет генерации вектора случайных чисел $\varphi(j) \in R^P$:

$$\varphi = \frac{\Delta}{\sqrt{\Delta^T \Delta}},$$

где Δ — вектор случайных чисел в интервале $[-1; 1]$. Вектор φ представляет собой множество длин для соответствующих измерений.

Шаг 3. Перемещение. Рассчитывается новое положение i -го агента по формуле

$$X^i(j+1, k, l) = X^i(j, k, l) + C(i)\varphi(j),$$

где $C(i) > 0$ — размер шага в определенном направлении, позволяющий моделировать процесс кувыркивания.

Для новой позиции $X^i(j+1, k, l)$ рассчитывается соответствующее значение целевой функции $J(i, j+1, k, l)$. Установить, что $J(i, j+1, k, l) = J(i, j+1, k, l) + J_{ar}(X^i(j+1, k, l))$. Если $J(i, j+1, k, l) < J_{best}$, то установить, что $J_{best} = J(i, j+1, k, l)$, $X_{best} = X^i(j+1, k, l)$.

Шаг 4. Скольжение. Если в позиции $X^i(j+1, k, l)$ значение $J(i, j+1, k, l)$ лучше, чем в позиции $X^i(j, k, l)$, т.е. выполняется условие $J(i, j+1, k, l) < J(i, j, k, l)$, то реализуется следующий хемотаксический шаг с тем же вектором φ и в том же направлении (переход к шагу 3), такое повторение может повторяться N_s раз. Если условие не выполняется, то переход к шагу 5.

Шаг 5. Если $i < S$, то выполняется переход к шагу 1, в противном случае — переход к следующему этапу метода.

Моделирование процесса хемотаксиса выполняется N_c раз.

Этап воспроизводства предназначен для создания нового множества агентов из существующего набора. На данном этапе менее приспособленные агенты уничтожаются, а более приспособленные получают возможность размножаться. Для этого предлагается использовать эволюционные операторы отбора, скрещивания и мутации.

Для отбора агентов с целью генерации новых агентов предлагается использовать пропорциональный отбор, алгоритм выполнения которого представлен ниже.

Шаг 1. Найти среднее значение целевой функции $J_{\text{ср}}$ популяции как среднее арифметическое значений целевых функций всех особей:

$$J_{\text{ср}} = \frac{1}{S} \sum_{i=1}^S J^i.$$

Шаг 2. Для каждого агента вычислить отношение, определяющее вероятность быть отобранным к скрещиванию:

$$P_s(i) = \frac{J^i}{J_{\text{ср}}}.$$

Шаг 3. В зависимости от величины $P_s(i)$ сформировать массив агентов, допущенных к скрещиванию, путем стохастического остаточного отбора: если $P_s(i) < 1$, то агент считается хорошо приспособленным и допускается к скрещиванию.

Для генерации новых агентов на основе агентов, отобранных с помощью оператора отбора, выполняется случайный выбор родительских пар по следующему алгоритму.

Шаг 1. Для выбора пары родителей задать вероятность скрещивания P_c . Величина P_c зависит от решаемой задачи, и в общем случае такую вероятность целесообразно выбирать [2] из интервала $[0,6; 0,99]$.

Шаг 2. Пронумеровать произвольно всех агентов, отобранных для скрещивания.

Шаг 3. Выбрать первого родителя. Для этого, начиная с первого, перебирать всех агентов популяции до тех пор, пока случайно выбираемое число из интервала $[0; 1]$ не будет меньше, чем P_c . Когда это условие выполнится для одного из агентов, этот элемент станет первым родителем.

Шаг 4. Продолжить просмотр популяции, начиная со следующего после первого родителя решения, и продолжать до тех пор, пока снова случайно выбранное число не будет меньшим P_c . Агент, для которого выполнится такое условие, будет вторым родителем.

Описанным образом составляются пары до тех пор, пока не будет выбрано нужное количество пар родителей.

Далее выполняется арифметическое скрещивание: на основе двух родителей X^1 и X^2 создаются два потомка: $X^{1п}$ и $X^{2п}$, значение i -х координат $x_i^{1п}$ и $x_i^{2п}$ которых рассчитываются по формулам

$$x_i^{1п} = k \cdot x_i^1 + (1-k)x_i^2, \quad x_i^{2п} = k \cdot x_i^2 + (1-k)x_i^1,$$

где $k \in [0; 1]$ — некоторый действительный коэффициент, который выбирается случайным образом на каждой итерации.

Для организации поиска в разных областях поискового пространства и выхода из локальных оптимумов на этапе воспроизводства агентов выполняется оператор простой мутации.

Шаг 1. Выбрать случайным образом координату, которая будет меняться.

Шаг 2. В заданном интервале допустимых значений координаты выбрать новое значение, которое отличается от текущего.

Целью этапа исключения и рассеивания является уменьшение вероятности заикливания в локальных оптимумах. В естественных природных условиях жизнь популяции бактерий (используемых в качестве агентов в предложенном методе) в окружающей среде может изменяться либо постепенно (например, путем потребления полезных веществ), либо внезапно под воздействием других факторов. Может произойти так, что все бактерии в области погибнут или колония бактерий будет рассеяна в другую часть окружающей среды. Данный эффект может помешать возможному хемотаксическому прогрессу, но в то же время и

помочь, поскольку в случае рассеивания агенты разместятся около хороших источников с полезными веществами. Исключение и рассеивание способствуют понижению вероятности стагнации, т.е. заикливания в локальном оптимуме, что часто наблюдается в известных градиентных методах оптимизации.

В соответствии с данным подходом каждый агент с вероятностью P_{ed} размещается в случайно выбранной точке пространства поиска. Таким образом, проверяется условие $U^i < P_{ed}$, где U^i — случайное число в интервале $[0; 1]$ для i -го агента. Если данное условие выполняется, то агент помещается в позицию $X^i(j, k, l)$, полученную случайным образом.

После выполнения N_{ed} раз этапов моделирования хемотаксиса, воспроизводства, исключения и рассеивания происходит сохранение лучшего решения J_{best} и соответствующей позиции X_{best} , в которой достигается лучшее решение J_{best} .

Поиск оптимума прекращается, если достигнуты критерии останова метода, в частности, если найдено приемлемое решение или в случае, если были выполнены все циклы для всех агентов.

Если критерии останова не достигнуты, то выполняется этап перезапуска агентов: выбираются новые случайные позиции для каждого агента X^i , $i = 1, 2, \dots, S$, рассчитываются соответствующие значения целевой функции J^i , $i = 1, 2, \dots, S$, после чего выполняется этап моделирования хемотаксиса.

В отличие от традиционных методов многомерной оптимизации предложенный мультиагентный метод с косвенной связью между агентами имеет следующие преимущества:

- не предъявляет каких-либо требований к целевой функции, т.е. функция может быть мультимодальной, негладкой и иметь разрывы;
- для поиска оптимума целевой функции не требуется рассчитывать градиенты целевой функции, т.е. она может быть недифференцируемой;
- является методом глобального поиска, в то время как большинство классических методов оптимизации представляют собой методы локального поиска;
- имеет достаточное количество параметров, которые настраиваются, что способствует его лучшей адаптации для решения той или иной практической задачи оптимизации;
- является мультиагентным, за счет чего достигается сокращение времени выполнения поиска, поскольку последний выполняется некоторым количеством агентов параллельно;
- хорошо расширяется, т.е. в его работу можно легко внедрять новые операторы и тем самым создавать гибридные системы.

ОТБОР ИНФОРМАТИВНЫХ ПРИЗНАКОВ НА ОСНОВЕ МУЛЬТИАГЕНТНОГО ПОДХОДА

Для решения задачи отбора информативных признаков на основе предложенного мультиагентного метода с косвенной связью между агентами последние должны осуществлять оптимизацию в пространстве поиска $X^F = (x_1, x_2, \dots, x_L)$, $x_i = \{0, 1\}$, где L — общее количество признаков, характеризующих выборку. Таким образом, каждая точка пространства поиска представляет собой битовую строку, координаты которой позволяют определить, считается ли признак информативным. В случае информативности признака соответствующая координата равна единице, в противном случае признак считается неинформативным.

В качестве целевой функции $J(X)$ предлагается использовать ошибку модели, синтезированной на основе оцениваемой комбинации признаков. Модели, используемые для оценивания наборов признаков, могут быть регрессионными, полиномиальными, нейросетевыми, нечеткими, нейро-нечеткими и др. [1, 2].

При выполнении перемещения в таком пространстве поиска необходимо учитывать его особенности, т.е. полученная новая позиция не должна выходить за пределы пространства поиска. Поэтому новое положение i -го агента в разраба-

тываемом мультиагентном методе отбора признаков предлагается рассчитывать по следующей формуле:

$$X^i(j+1, k, l) = \begin{cases} 1, & \text{если } X^i(j, k, l) + C(i) \cdot \varphi(j) \geq 1; \\ \text{round}(X^i(j, k, l) + C(i) \cdot \varphi(j)), & \\ & \text{если } 0 < X^i(j, k, l) + C(i) \cdot \varphi(j) < 1; \\ 0, & \text{если } X^i(j, k, l) + C(i) \cdot \varphi(j) \leq 0. \end{cases} \quad (1)$$

Здесь $X^i(j, k, l)$ — позиция i -го агента на j -м хемотаксическом шаге, k -м шаге воспроизводства и на l -м событии исключения-рассеивания; φ — множество длин для соответствующих измерений; $C(i) > 0$ — размер шага в определенном направлении, позволяющий моделировать процесс кувыркивания.

Тогда величины, определяющие процесс движения в пространстве поиска, должны принимать следующие значения:

- переменные, определяющие размер шага в определенном направлении для i -го агента: $C(i) \in (0; 1)$;
- вектор случайных чисел на j -м шаге хемотаксиса: $\varphi(j) = \{-1; 0; +1\}$.

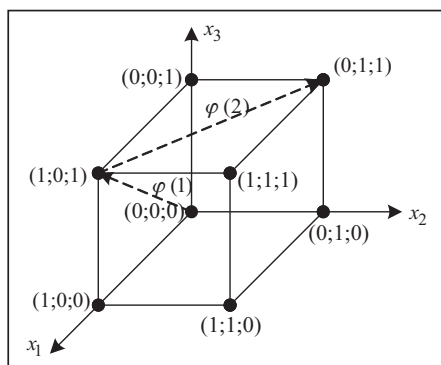


Рис. 1. Пример перемещения в пространстве поиска из трех признаков

На рис. 1 приведен пример перемещения агентов в пространстве поиска из трех признаков, если агент в начале поиска находился в точке $(0; 0; 0)$, вектор размера шага для данного агента $C = (1; 1; 1)$, вектор случайных чисел на первом шаге хемотаксиса $\varphi(1) = (1; 0; 1)$, а на втором шаге $\varphi(2) = (0; 1; 1) - (1; 0; 1) = (-1; 1; 0)$.

Тогда мультиагентный метод с косвенной связью между агентами для отбора информативных признаков можно представить в виде следующей последовательности шагов.

Шаг 1. Инициализация. Задать параметры метода: $S, N_{re}, N_s, N_c, N_{ed}, P_{ed}$.

Случайным образом распределить начальные значения $X^i, i = 1, 2, \dots, S$, в пространстве поиска. Рассчитать начальные значения целевой функции для каждого агента J^i и выполнить для них пересчет: $J^i = J^i + J_{ar}(X^i)$. Установить счетчики циклов в 0: $l = 0, k = 0, j = 0$, где l ($l = 1, 2, \dots, N_{ed}$) — счетчик событий исключения-рассеивания; k ($k = 1, 2, \dots, N_{re}$) — счетчик шагов воспроизводства; j ($j = 1, 2, \dots, N_c$) — счетчик хемотаксических шагов.

Шаг 2. Увеличить значение счетчика событий исключения-рассеивания: $l = l + 1$.

Шаг 3. Увеличить значение счетчика шагов воспроизводства: $k = k + 1$.

Шаг 4. Увеличить значение счетчика хемотаксических шагов: $j = j + 1$.

Шаг 5. Моделирование хемотаксиса.

Шаг 5.1. Установить, что $i = i + 1$, где i — счетчик агентов в популяции $i = 1, 2, \dots, S$.

Шаг 5.2. Кувыркивание и перемещение. Создается вектор $C(i) = \text{rand}(0; 1)$, где $\text{rand}(0; 1)$ — случайное действительное число в диапазоне $[0; 1]$. Генерация вектора случайных чисел $\varphi(j) = \text{rand_int}(-1; +1)$, где $\text{rand_int}(-1; +1)$ — случайное целое число в диапазоне $[-1; +1]$. Рассчитывается новое положение i -го агента $X^i(j+1, k, l)$ по формуле (1), учитывающей особенности задачи отбора признаков и обеспечивающей постоянное нахождение агентов в необходимом пространстве поиска.

Рассчитывается значение целевой функции $J(i, j+1, k, l)$. Устанавливается, что $J(i, j+1, k, l) = J(i, j+1, k, l) + J_{ar}(X^i(j+1, k, l))$. Если $J(i, j+1, k, l) < J_{best}$, то установить $J_{best} = J(i, j+1, k, l), X_{best} = X^i(j+1, k, l)$.

Шаг 5.3. Скольжение. Пока выполняется условие $J(i, j+1, k, l) < J(i, j, k, l)$, повторять шаг 5.2. Такое повторение может выполняться N_s раз. Если условие не выполняется, то перейти к шагу 5.4.

Шаг 5.4. Если $i < S$, то выполнить переход к шагу 5.1, в противном случае — к шагу 6.

Шаг 6. Если $j < N_c$, то перейти к шагу 4, в противном случае — к шагу 7.

Шаг 7. Воспроизводство. Используя такой же подход, как и в базовом мультиагентном методе с косвенной связью между агентами, выполнить формирование нового множества агентов.

Шаг 8. Если $k < N_{re}$, то перейти к шагу 3, в противном случае — к шагу 9.

Шаг 9. Исключение и рассеивание. Проверить условие $U^i < P_{ed}$. Если оно выполняется, то агент помещается в позицию $X^i(j, k, l)$, полученную случайным образом.

Шаг 10. Если $l < N_{ed}$, то перейти к шагу 2, в противном случае — к шагу 11.

Шаг 11. Выбрать и сохранить лучшее решение J_{best} и соответствующую позицию X_{best} .

Шаг 12. Выполнить проверку условий окончания поиска. Если найдено приемлемое решение или выполнены все циклы для всех агентов, то перейти к шагу 14, в противном случае произвести перезапуск агентов (шаг 13).

Шаг 13. Перезапуск агентов: выбираются новые случайные позиции для каждого агента X^i , $i = 1, 2, \dots, S$, рассчитываются соответствующие значения целевой функции J^i , $i = 1, 2, \dots, S$. Счетчики циклов обнуляются: $j = 0$, $k = 0$, $l = 0$.

Шаг 14. Останов.

Разработанный мультиагентный метод с косвенной связью между агентами для отбора информативных признаков позволяет в условиях взаимной зависимости признаков выполнять выделение наиболее значимой комбинации признаков и, в отличие от известных методов отбора признаков, не склонен к закликиванию в локальных оптимумах, не использует жадную стратегию поиска и не выдвигает дополнительных требований к виду целевой функции.

ЭКСПЕРИМЕНТЫ И РЕЗУЛЬТАТЫ

Предложенный мультиагентный метод отбора признаков был программно реализован на языке пакета Matlab. Для проверки эффективности применения этого метода и разработанного программного обеспечения решалась задача отбора информативных признаков для синтеза моделей в задачах диагностики хронического бронхита [2].

Для построения распознающих моделей использовалась экспериментально полученная выборка данных исследований функционально-морфологических свойств лейкоцитов и тромбоцитов, сосудистого фактора, системы комплемента и циркулирующих иммунных комплексов (ЦИК) периферической крови 205 пациентов.

Анализировалось состояние 28 диагностических признаков: x_1 — пол, x_2 — возраст, x_3 — содержание лейкоцитов, в том числе: x_4 — палочкоядерных, x_5 — сегментоядерных, x_6 — эозинофилов, x_7 — моноцитов, x_8 — лимфоцитов, x_9 — фагоцитарное число нейтрофилов, x_{10} — фагоцитарное число моноцитов, x_{11} — фагоцитарный индекс нейтрофилов, x_{12} — фагоцитарный индекс моноцитов, x_{13} — индекс завершенности фагоцитоза нейтрофилов, x_{14} — индекс завершенности фагоцитоза моноцитов, x_{15} — НСТ-тест, отражающий способность нейтрофилов генерировать активные формы кислорода; x_{16} — процент НСТ положительных нейтрофилов, x_{17} — НСТ-стимулированный, показывающий резервные возможности нейтрофилов генерировать активные формы кислорода в условиях их дополнительного стимулирования, x_{18} — процент НСТ-положительных нейтрофилов, реагирующих на стимулирование; x_{19} — количество

тромбоцитов, x_{20} — количество зрелых тромбоцитов, x_{21} — количество активных тромбоцитов, x_{22} — уровень комплемент (СН₅₀), x_{23} — уровень ЦИК, осаждаемых 3%-м полиэтиленгликолем (ПЭГ), x_{24} — уровень ЦИК, осаждаемых 4%-м ПЭГ, x_{25} — размер ЦИК, x_{26} — активность миелопероксидазы нейтрофилов, x_{27} — активность катионных белков нейтрофилов, x_{28} — уровень сосудистого фактора Виллебранда.

Каждому экземпляру выборки было сопоставлено значение целевого параметра — диагноза соответствующего пациента (1 — болен, 0 — не болен).

Выделение комбинации признаков выполнялось с использованием следующих известных методов: эвристического поиска, анализа главных компонент, группового учета аргументов, канонической модели генетического поиска, мультиагентных методов отбора информативных признаков с прямой и непрямой связью между агентами, а также с помощью разработанного метода отбора информативных признаков с косвенной связью между агентами. Приведем значения параметров работы мультиагентного метода с косвенной связью между агентами: $S = 20$; $N_{re} = 4$; $N_s = 4$; $N_c = 20$; $N_{ed} = 2$; $P_{ed} = 0,25$; $d_{attract} = 0,1$; $w_{attract} = 0,2$; $h_{repellant} = d_{attract}$; $w_{repellant} = 10$. Поскольку оптимизационный процесс большинства из перечисленных методов носит вероятностный характер, поиск оптимальных решений при проведении экспериментов выполнялся сто раз, после чего на основе полученных результатов рассчитывались средние значения исследуемых параметров. Работа методов оценивалась по следующим критериям:

- количество обращений к целевой функции (ЦФ) N_{fit} , необходимое для достижения результата с требуемой точностью, определяет вычислительную сложность и эффективность оптимизационного процесса: чем меньше обращений к ЦФ для нахождения оптимального решения, тем эффективнее оптимизационный процесс;
- ошибка работы метода, определяемая на основе известного оптимального или допустимого значения исследуемого параметра J и рассчитываемая по формуле

$$\varepsilon = |J^{**} - J^*|,$$

где J^{**} — найденное рассматриваемым методом значение исследуемого параметра J ; J^* — известное оптимальное или допустимое значение исследуемого параметра;

- время работы оптимизационного процесса T , необходимое для достижения оптимального решения, является традиционным критерием оценивания работы методов оптимизации;
- количество информативных признаков k ;
- математического ожидания M найденного оптимума при проведении ста запусков работы метода;
- дисперсия D найденного оптимального решения при ста запусков поиска.

В качестве целевой функции $J(X)$ использовалась вероятность принятия ошибочных решений по двухслойной нейронной сети прямого распространения, синтезированной на основе признаков оцениваемой точки пространства поиска и содержащей пять нейронов на первом слое и один нейрон на втором. Все нейроны имели логсигмоидную функцию активации, а в качестве дискриминантных функций использовались взвешенные суммы.

В табл. 1 представлены значения характеристик предложенных мультиагентных и известных методов отбора информативных признаков при диагностике хронического бронхита. Как видим, комбинации признаков, полученные с помощью традиционных методов отбора признаков, характеризуются худшими значениями критерия оценивания информативности по сравнению с комбинациями, найденными путем применения предложенного мультиагентного метода с косвенной связью между агентами. Это обусловлено неоптимальностью традиционных поисковых процедур, при которых большая часть пространства поиска

остаётся неисследованной. Кроме того, модели, построенные на основе признаков, отобранных с использованием предложенного мультиагентного метода, характеризуются меньшей ошибкой, что свидетельствует о его эффективности при применении на практике и может объясняться лучшим изучением пространства поиска, обусловленным особенностями предложенного метода.

Таблица 1

Методы отбора признаков	Значения характеристик методов отбора признаков					
	ε	N_{fit}	T	k	M	D
Метод анализа главных компонент	0,0398	–	126	8	–	–
Метод группового учета аргументов	0,0301	9862	4721	9	–	–
Канонический метод эволюционного поиска	0,0167	10000	4898	9	0,0298	0,00001
Метод поочередного добавления и удаления признаков	0,0402	2984	2361	8	–	–
Мультиагентный метод с непрямой связью между агентами	0,0128	4968	2782	8	0,01288	0,00012
Мультиагентный метод с прямой связью между агентами	0,0132	4984	2823	8	0,01325	0,00007
Мультиагентный метод с косвенной связью между агентами	0,0124	5242	3129	8	0,01242	0,00004

На рис. 2 приведен график зависимости достигнутого значения критерия оценивания информативности набора признаков от номера итерации для различных методов отбора признаков.

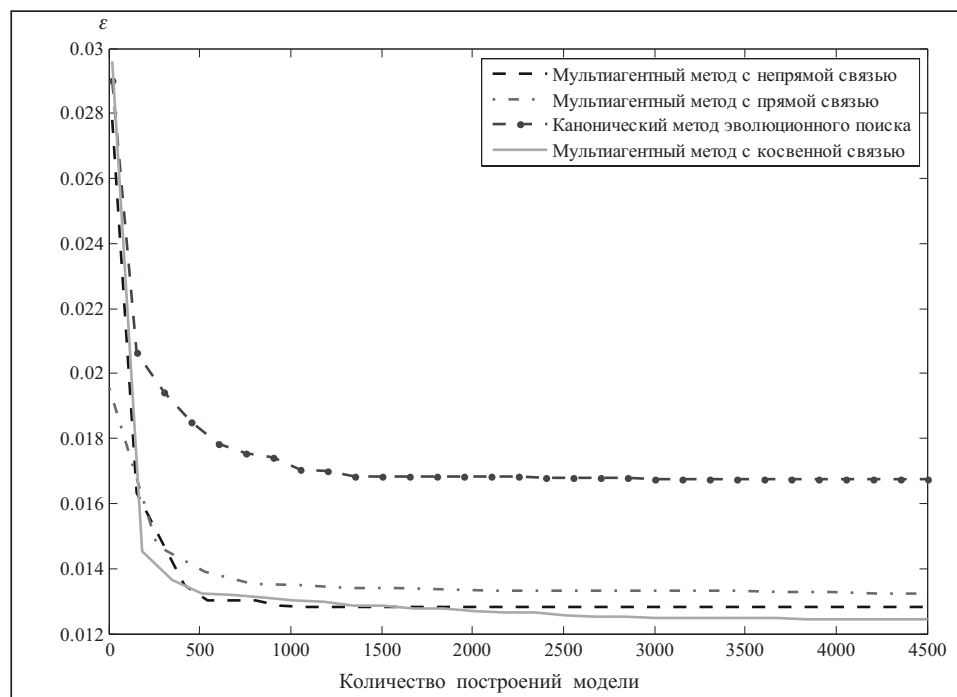


Рис. 2. График зависимости достигнутого значения ЦФ ε от номера итерации для различных методов отбора признаков

Как видно из рис. 2, худшие результаты показал канонический метод эволюционного поиска. Мультиагентные методы с прямой, непрямой и косвенной связью между агентами получили сходные результаты. Однако мультиагентный метод с непрямой связью быстрее находит решение, чем мультиагентные методы с косвенной и прямой связью между агентами, в то время как мультиагентный метод с косвенной связью между агентами обеспечивает итоговое решение большей точности. Такие тенденции могут объясняться тем, что мультиагентный метод с непрямой связью ориентирован на решение задач дискретной оптимизации, к которым относится задача отбора информативных признаков, в то время как мультиагентный метод с косвенной связью между агентами хоть и работает медленнее, но исследует пространства поиска более детально.

Значения весовых коэффициентов нейронной сети, полученной на основании признаков путем применения мультиагентного метода с косвенной связью между агентами, представлены в табл. 2.

Таблица 2

Номер слоя	Номер нейрона в слое	Значение весовых коэффициентов нейросети на входах нейрона с номерами								
		0	1	2	3	4	5	6	7	8
1	1	12,383	-29,38	11,18	13,79	10,35	-7,91	-7,115	29,81	-13,765
	2	-29,77	5,531	-39,11	21,89	21,79	-1,473	-19,89	17,57	39,053
	3	18,821	-5,155	26,798	-20,58	4,1709	24,738	-3,715	7,9964	-49,53
	4	57,701	112,73	-497,4	-174,5	241,16	785,24	-241,8	-422,2	-508,85
	5	16,454	-100,7	-13,24	35,465	43,37	93,237	-17,06	104,71	48,537
2	1	-1677	-1328	433,3	1768	-85,8	1238,7	—		

Таким образом, применение предложенного мультиагентного метода отбора признаков позволяет получать модели с большей точностью и с меньшим количеством признаков по сравнению с известными методами отбора признаков.

ЗАКЛЮЧЕНИЕ

Решена задача отбора информативных признаков на основе мультиагентной оптимизации.

Научная новизна работы:

- предложен мультиагентный метод с косвенной связью между агентами для оптимизации многомерных нелинейных целевых функций, в котором на этапе воспроизводства агентов используются эволюционные операторы отбора, скрещивания и мутации, что позволяет более эффективно исследовать пространство поиска по сравнению с известными мультиагентными методами и уменьшить время, необходимое на поиск. Разработанный метод не выдвигает дополнительных требований к виду целевой функции, не требует вычисления производных целевой функции и является методом глобального поиска;

- разработан метод отбора информативных признаков на основе мультиагентного подхода с косвенной связью между агентами, который позволяет в условиях взаимной зависимости признаков выполнять выделение наиболее значимой комбинации признаков и, в отличие от известных методов отбора, не склонен к закликиванию в локальных оптимумах, не использует жадную стратегию поиска и не выдвигает дополнительных требований к виду целевой функции.

Практическая ценность результатов работы:

- создано программное обеспечение, реализующее предложенный мультиагентный метод с косвенной связью между агентами;

- решена задача отбора информативных признаков для диагностики хронического бронхита.

Работа выполнена в рамках госбюджетной НИР «Информационные технологии автоматизации распознавания образов и принятия решений для диагностики в условиях неопределенности на основе гибридных нечеткологических, нейросетевых и мультиагентных методов вычислительного интеллекта» кафедры программных средств Запорожского национального технического университета (номер государственной регистрации 0109U007673).

СПИСОК ЛИТЕРАТУРЫ

1. Прогрессивные технологии моделирования, оптимизации и интеллектуальной автоматизации этапов жизненного цикла авиадвигателей / А.В. Богуслаев, Ал.А. Олейник, Ан.А. Олейник и др.; под ред. Д.В. Павленко, С.А. Субботина. — Запорожье: ОАО «Мотор Сич», 2009. — 468 с.
2. Субботін С.О., Олійник А.О., Олійник О.О. Ітеративні, еволюційні та мультиагентні методи синтезу нечіткологічних і нейромережних моделей / Під ред. С.О. Субботина. — Запоріжжя: ЗНТУ, 2009. — 375 с.
3. Dash M., Liu H. Feature selection for classification // *Intelligent Data Analysis*. — 1997. — N 1. — P. 131–156.
4. Guyon I., Elisseeff A. An introduction to variable and feature selection // *J. of Machine Learning Research*. — 2003. — N 3. — P. 1157–1182.
5. Beni G., Wang J. Swarm intelligence // *Annual Meeting of the Robotics Society: Proceedings of Seventh International Conference*. — Tokyo: RSJ Press, 1989. — P. 425–428.
6. Bonabeau E., Dorigo M., Theraulaz G. *Swarm intelligence: from natural to artificial systems*. — New York: Oxford University Press, 1999. — 320 p.
7. Dorigo M., Maniezzo V., Colorni A. The Ant System: Optimization by a colony of cooperating agents // *IEEE Transactions on Systems, Man, and Cybernetics*. — 1996. — Part B, N 26 (1). — P. 29–41.
8. Sumpter D.J., Broomhead D.S. Formalising the link between worker and society in honey bee colonies // *Lecture Notes In Computer Science: Proceedings of the First International Workshop on Multi-Agent Systems and Agent-Based Simulation*. — MABS'98, LNAI, 1998. — P. 95–110.
9. Passino K.M. Biomimicry of bacterial foraging for distributed optimization and control // *IEEE Control System Magazine*, 2002. — N 3 (22). — P. 52–67.
10. Analysis and design of intelligent systems using soft computing techniques / Ed. P. Melin, O.R. Castillo, E.G. Ramirez, J. Kacprzyk. — Heidelberg: Springer, 2007. — 855 p.
11. Kim D.H., Cho C.H. Bacterial foraging based neural network fuzzy learning // *Proceedings of the 2nd Indian International Conf. on Artificial Intelligence (IICAI-2005)*. — Pune: IICAI, 2005. — P. 2030–2036.

Поступила 07.09.2010