

УДК 004.032.26

DOI: 10.26102/2310-6018/2019.26.3.029

И.Л. Каширина, К.А. Федутинов  
**ПОСТРОЕНИЕ РЕШАЮЩИХ ПРАВИЛ С ПОМОЩЬЮ  
НЕЙРОННОЙ СЕТИ ARTMAP**

*ФГБОУ ВПО «Воронежский государственный университет»,  
Воронеж, Россия*

*В статье рассматривается нейросетевая архитектура ARTMAP, совместимая с символическим представлением, основанным на IF- THEN правилах. В частности, знания, полученные в процессе обучения сети ARTMAP, могут быть трансформированы в компактный набор решающих правил для классификации исходных данных, которые могут быть проанализированы экспертами предметной области, по аналогии с интерпретируемыми методами машинного обучения, такими, как деревья решений или линейная регрессия. Аналогичным образом, знания в априорной области, представленные в форме правил IF- THEN могут быть преобразованы в нейросетевую архитектуру ARTMAP. Наличие предварительного набора правил, используемых при инициализации сети, повышает точность классификации и эффективность обучения. Исходный набор правил может быть дополнен с помощью алгоритма обучения ARTMAP. Каждое правило, сформированное в процессе обучения сети имеет коэффициент достоверности, который можно интерпретировать как его важность или полезность. Описание архитектуры, алгоритмов обучения и функционирования сети ARTMAP для извлечения правил представлено в терминах предложенной авторами ранее обобщенной модели сетей семейства ART.*

**Ключевые слова:** нейронная сеть, машинное обучение, теория адаптивного резонанса, ARTMAP, извлечение правил.

## ВВЕДЕНИЕ

Машинное обучение имеет большой потенциал для решения широкого круга задач в самых разных областях. Но нейронные сети обычно работают в режиме “черного ящика”, то есть не объясняют свои предсказания, что зачастую является препятствием для использования этих методов [1]. В некоторых случаях исследователи ограничены только простыми интерпретируемыми моделями. Такая ситуация возникает, например, в медицинских приложениях, когда речь идет о прогнозировании диагноза пациента, и цена ошибки может быть высокой. В таких случаях нельзя полагаться на одного аналитика с его неинтерпретируемым алгоритмом машинного обучения, и ценность этого алгоритма начинает определяться не только точностью прогноза, но и способностью понимать и объяснять логику его работы, чтобы, в случае необходимости, внести коррективы со стороны экспертов в данной области. В таких случаях используются простые, интерпретируемые модели, такие как деревья принятия решений, правила принятия решений и линейная регрессия, точность которых, как правило, недостаточно высока.

В данной статье обсуждается нейросетевая архитектура, относящаяся

к классу сетей семейства ART (адаптивной резонансной теории). Сеть ARTMAP, являющаяся одним из представителей данного семейства и в общем случае предназначенная для решения задачи классификации, позволяет получать явное представление каскадов знаний, основанных на правилах [2]. Во время обучения сети ARTMAP новые категории распознавания (правила) создаются динамически, дополняя существующий набор правил. Благодаря присущим сетям семейства ART свойству стабильности-пластичности [3], обучение в ARTMAP не стирает существующие знания. С использованием обобщенной процедуры извлечения правил ARTMAP, конечные состояния системы могут быть преобразованы в компактный набор решающих правил. Это позволяет проводить прямое сравнение исходных экспертных знаний с уточненными правилами. Кроме того, каждое выделенное правило может быть связано с фактором достоверности, который указывает на его важность или полезность. Это позволяет ранжировать и оценивать полученные правила. В целом, процедуры уточнения и оценки правил ARTMAP образуют нейросетевую парадигму для формирования решающих правил (Рисунок 1).

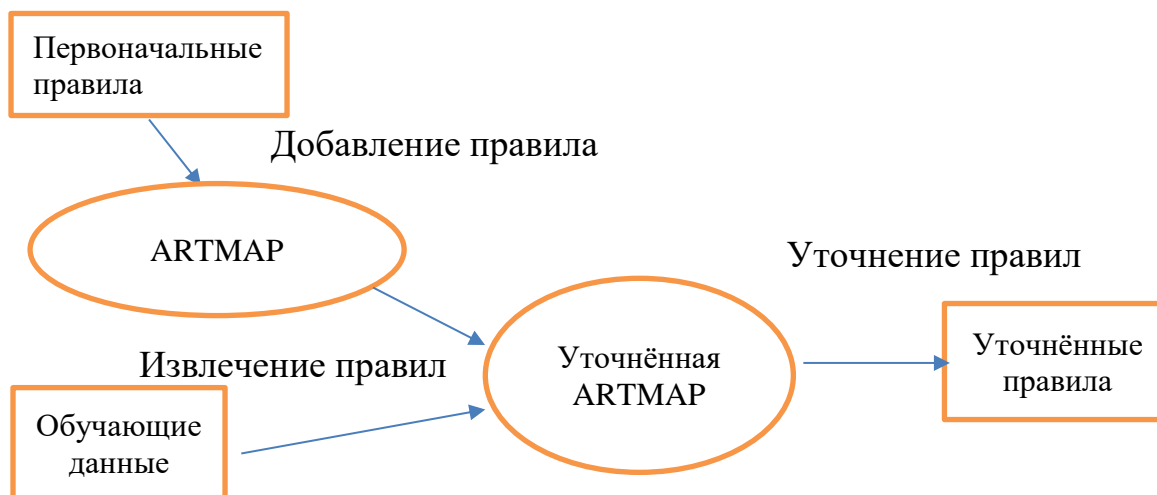


Рисунок 1 – Cascade ARTMAP для построения решающих правил.

Одной из проблем, связанных с практическим использованием сети ARTMAP для построения решающих правил, является то, что в литературных источниках уделяется большое внимание концепции ее функционирования, но не дается четкого представления об архитектуре модели и не приводятся структурированных алгоритмов ее обучения [2,3]. В данной статье сеть ARTMAP, извлекающая решающие правила, описана в терминах обобщенной модели функционирования сетей семейства ART, предложенной авторами в [4].

## 1. МАТЕРИАЛЫ И МЕТОДЫ

### 1.1. Архитектура сети ARTMAP для извлечения правил

Как и другие сети семейства ART, сеть ARTMAP представляет собой онлайн систему реального времени, не нуждающуюся в отдельном этапе предварительного обучения, т.е. система функционирует в зависимости от текущей входной среды, одновременно обучаясь. Сеть ARTMAP состоит из двух модулей Fuzzy ART: ART<sub>a</sub> и ART<sub>b</sub>, объединенных с использованием слоя ассоциативной памяти  $F^{ab}$ , как показано на Рисунке 2 и основанных на принципах нечеткой логики.

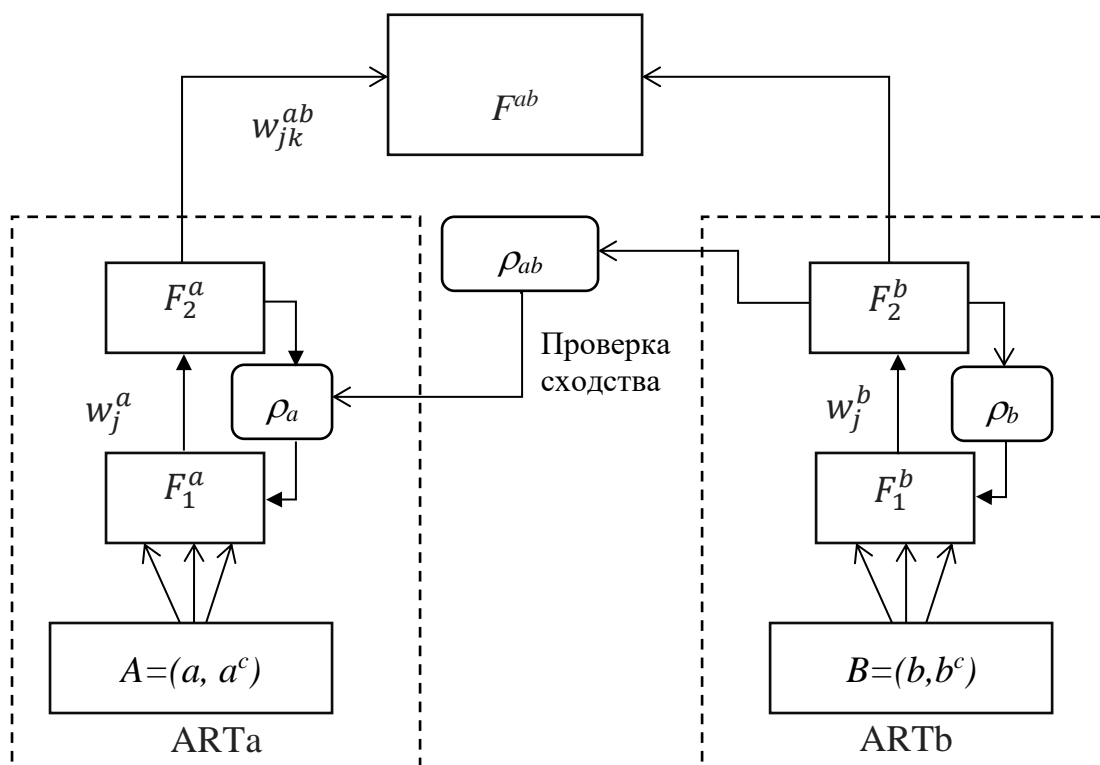


Рисунок 2 – Архитектура сети ARTMAP

### 1.2. Каскадное представление правил

Считается, что набор правил образует каскад, когда одно из следствий правила также является посылкой другого правила. Атрибуты, которые имеют две роли, называются промежуточными атрибутами, в отличие от входных атрибутов, которые используются только в качестве посылки и выходных атрибутов, которые используются только как следствие. Через промежуточные атрибуты использование одного правила может привести к запуску другого правила на более позднем этапе процесса логического вывода. Промежуточные атрибуты и каскады правил полезны для разложения задач таким образом, чтобы только на каждом уровне каскада использовался только небольшой набор простых правил.

Сеть ARTMAP, нацеленная на извлечение правил, объединяет слой  $F_1^a$  входных атрибутов и слой  $F_1^b$  выходных атрибутов в том смысле, что и  $F_1^a$

и  $F_1^b$  будут обрабатывать входные, выходные и промежуточные атрибуты. Рассмотрим два правила, которые образуют простой двухуровневый каскад правил:

Правило 1: IF A and B THEN C,

Правило 2: IF C and D THEN E,

где A, B и D - входные атрибуты; C - промежуточные атрибуты; и E - выходные атрибуты. Все атрибуты (A, B, C, D и E) представлены как в  $F_1^a$ , так и в  $F_1^b$  (Рисунок 3). Для правила 1 категория слоя  $F_2^a$  используется для запоминания A и B, при этом она связана с категорией слоя  $F_2^b$ , предсказывающей C. Аналогично для правила 2, категория слоя  $F_2^a$  используется для запоминания C и D, и при этом она связана с с категорией слоя  $F_2^b$ , предсказывающей E.

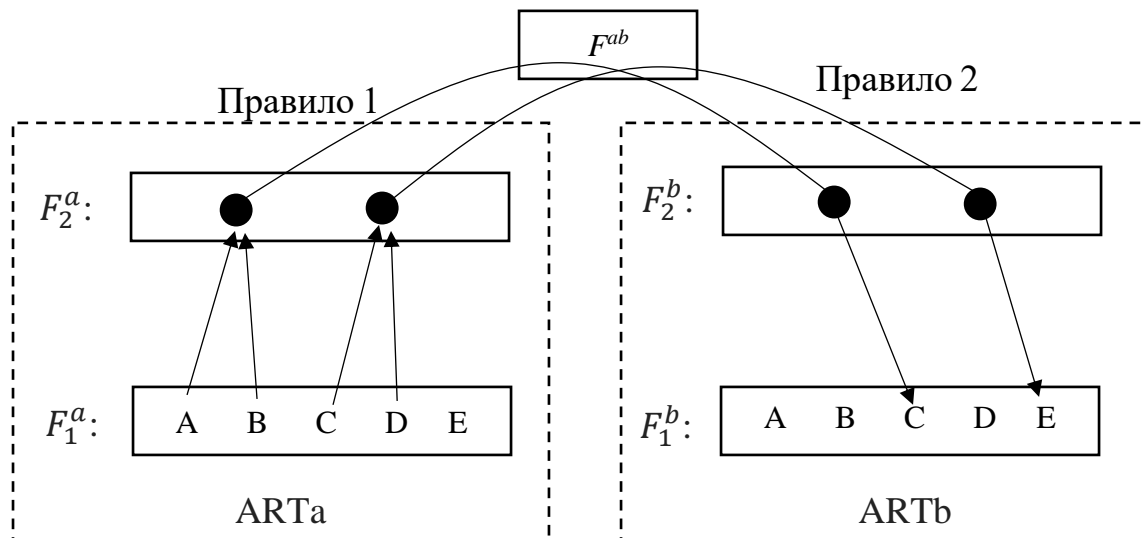


Рисунок 3 – Представление каскада простых правил в сети ARTMAP.

### 1.3. Обобщенная модель[4]

Кратко общие этапы работы сетей семейства ART имеют следующий вид.

Этап 1. Инициализация сети и ее параметров.

Этап 2. Предварительная обработка (препроцессинг) входных векторов.

Этап 3. Выбор категории (на основе функции выбора) - первоначальная оценка входного вектора с использованием некоторой функции сходства и активация наиболее похожего нейрона из слоя прототипов.

Этап 4. Проверка сходства (на основе функции соответствия) более детальный анализ сходства с возможным возвратом к этапу 3 и рассмотрением оставшихся прототипов.

Этап 5. Обучение (с использованием функции обучения) - точная

настройка весов прототипа кластера, вызвавшего реакцию функции выбора. Если ни один из существующих прототипов не прошёл проверку функцией соответствия, активируется новый нейрон и создается новая категория на основе нераспознанного вектора.

Этап 6. Постобработка (постпроцессинг) нейронного слоя - например, объединение или отключение некоторых категорий в соответствии с необходимыми условиями.

## 2. РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

### 2.1. Обучение сети

Рассмотрим реализацию этапов 2-5 данной модели для сети ARTMAP. При этом, так как сеть ARTMAP состоит из двух модулей архитектуры ART (ARTa и ARTb), некоторые этапы алгоритма для этих модулей рассмотрим отдельно (под буквами а и b соответственно)

Этап 2. Рассмотрим правила инициализации всех основных элементов сети.

а) Обозначим через  $A$  и  $B$  входные векторы слоев  $F_1^a$  и  $F_1^b$  соответственно. Как и в сети Fuzzy ART, в сети ARTMAP к входным векторам применяется комплементарное кодирование:  $M$ -мерный вектор превращается в  $2M$ -мерный путём дополнения его  $M$  компонентами, что сохраняет его амплитудную информацию. Таким образом, входные образы сети ARTMAP имеют вид:  $A=(a, a^c)$ , где  $a_i^c = 1 - a_i$ ;  $B=(b, b^c)$ , где  $b_i^c = 1 - b_i$ .

б) Обозначим через  $x^a = (x_i^a, x_h^a, x_o^a, x_i^{ac}, x_h^{ac}, x_o^{ac})$  и  $x^b = (x_i^b, x_h^b, x_o^b, x_i^{bc}, x_h^{bc}, x_o^{bc})$  –  $2M$ -мерные векторы активности слоев  $F_1^a$  и  $F_1^b$  соответственно, где  $x_i^a$  и  $x_i^b$  –  $M_i$ -мерные входные векторы этих слоев;  $x_h^a$  и  $x_h^b$  –  $M_h$ -мерные векторы промежуточных атрибутов;  $x_o^a$  и  $x_o^b$  –  $M_o$ -мерные выходные векторы; и  $x_i^{ac}$ ,  $x_i^{bc}$ ,  $x_h^{ac}$ :  $x_h^{bc}$ ,  $x_o^{ac}$  и  $x_o^{bc}$  – векторы дополнительных атрибутов, соответствующих комплементарному кодированию. Обозначим через  $y^a$  и  $y^b$  аналогичные векторы активности слоев  $F_2^a$  и  $F_2^b$  и через  $x^{ab}$  вектор активности слоя  $F^{ab}$ .

с) Обозначим через  $w_j^a$  и  $w_j^b$   $2M$ -мерные векторы весовых коэффициентов  $j$ -й категории в слоях  $F_2^a$  и  $F_2^b$ , через  $w_j^{ab}$  весовой вектор от  $j$ -й категории слоя  $F_2^a$  к нейронам слоя  $F^{ab}$  (рис.1). Первоначально все весовые коэффициенты инициализируются единицами. Это означает, что категории пока не зафиксированы, и все узлы  $F_2^a$  не связаны с какими-либо классами. Вектор активности слоя  $F_1^a$  содержит входной вектор  $A$ :  $x^a=A$ .

д) Обозначим через  $S_j$   $2M$ -мерный вектор области видимости  $j$ -й категории в слое  $F_2^a$ . Вектор области видимости определяет входные атрибуты, относящиеся к категории  $F_2^a$ , и позволяет более точно рассчитать функцию соответствия этой категории. На этапе инициализации для  $j$ -й категории слоя  $F_2^a$  компоненты вектора области видимости  $S_j$  задаются следующим образом:

$$s_{ji} = \begin{cases} 1, & \text{если } j - \text{я категория содержит } i - \text{й входной атрибут} \\ 0, & \text{иначе} \end{cases}$$

е) Сеть ARTMAP имеет следующий набор гиперпараметров: параметры функции выбора  $\alpha_a > 0$  и  $\alpha_b > 0$ ; параметры скорости обучения  $\beta_a \in [0, 1]$  и  $\beta_b \in [0, 1]$ ; параметры сходства  $\rho_a \in [0, 1]$  и  $\rho_b \in [0, 1]$ . Во время инициализации сети ARTMAP, предназначенной для извлечения правил, полагается  $\rho_b = 1$  для точной классификации выходов и  $\rho_a < 1$  для обобщения входных данных.

Этап 3а. Для каждого  $j$ -го нейрона слоя  $F_2^a$  функция выбора  $T_j^a$  определяется следующим образом

$$T_j^a = \frac{|x^a \wedge w_j^a|}{\alpha_a + |w_j^a|}, \quad (1)$$

где нечеткий оператор AND ( $\wedge$ ) определяется следующим образом  $(p \wedge q)_j = \min(p_j, q_j)$ , а норма  $|p| = \sum_i p_i$ . Считается, что система

осуществила выбор категории, если определена категория  $J$ , такая, что

$$T_J^a = \max_j T_j^a. \quad (2)$$

Если максимум достигается более, чем на одном значении, выбирается категория с наименьшим индексом. При выборе  $j$ -й категории полагается  $y_j^a = 1$ , и  $y_j^a = 0$  для  $j \neq J$ .

Этап 4а. Резонанс возникает, если функция соответствия  $m_j^a$  выбранного нейрона  $J$  отвечает критерию соответствия:

$$m_j^a = \frac{|x^a \wedge w_j^a \wedge S_j|}{|x^a \wedge S_j|} \geq \rho_a, \quad (3)$$

где обобщенная нечеткая операция  $\wedge$  определяется следующим образом:

$$(p_1 \wedge p_2 \wedge \dots \wedge p_N)_j = \min(p_{1j}, p_{2j}, \dots, p_{Nj}) \quad (4)$$

В противном случае происходит сброс соответствия, при котором значение функции выбора  $T_j^a$  устанавливается равным 0 в течение всего времени обработки данного входного сигнала  $A$ , чтобы предотвратить повторный выбор той же категории. Этап 3 повторяется, выбирается еще один новый индекс  $J$ .

Этап 5а. Если текущий входной вектор  $A$  не похож ни на одну из существующих категорий, т.е.  $w_{JK}^{ab} = 1$  для всех нейронов слоя  $F^{ab}$ , для него создается новая категория. При этом веса нейрона  $J$  слоя  $F_2^a$  модифицируются следующим образом:

$$w_j^a := \beta_a (x^a \wedge w_j^a) + (1 - \beta_a) w_j^a \quad (5)$$

Этап 5б. Для входного вектора  $B$  аналогичным образом выбирается категория в слое  $F_2^b$ . Выбранный узел  $K$  слоя  $F_2^b$  запоминает шаблон слоя  $F_1^b$  в соответствии с формулой:

$$w_K^b := \beta_b (x^b \wedge w_K^b) + (1 - \beta_b) w_K^b \quad (6)$$

Категория  $J$  слоя  $F_2^a$  затем связывается с категорией  $K$  слоя  $F_2^b$  через промежуточное поле карты ART:

$$w_{jK}^{ab} = 1, w_{jk}^{ab} = 0 \text{ для } k \neq K. \quad (7)$$

Если выбранный нейрон слоя  $F_2^a$  соответствует уже существующей категории  $J$ , то вектором активности слоя  $F^{ab}$  становятся его весовые коэффициенты. Таким образом выход  $x^{ab}$  определяется следующим образом:

$$x_j^{ab} = w_j^{ab} \quad (8)$$

Этап 3б. Для каждого нейрона  $k$  слоя  $F_2^b$  функция выбора  $T_k^b$  определяется следующим образом:

$$T_k^b = x_k^{ab}$$

Система снова делает выбор некоторой категории с номером  $K$ , где

$$T_K^b = \max_k T_k^b.$$

Когда выбрана  $K$ -я категория, полагается  $y_K^b = 1$  и  $y_k^b = 0$  для  $k \neq K$ . Выходы активированного нейрона слоя  $F_1^b$  полагаются равными весовым коэффициентам слоя  $F_2^b$ :

$$x^b = w_K^b \quad (9)$$

Если нейрон слоя  $F_1^b$  активирован выбором категории в  $F_2^b$ , условие завершения вывода проверяется вычислением выходного сигнала  $g$ :

$$g = \sum_{i=1}^{M_0} (x_{oi}^b + x_{oi}^{bc}) \quad (10)$$

Вывод делается всякий раз, когда для данного входа определен какой-либо выходной атрибут, в этом случае  $g > 0$ . Если вывод не сделан, т.е.  $g = 0$ , вектор активности  $x^a$  обновляется с учетом вектора  $x^b$  по формуле:

$$x^a := x^a \vee x^b, \quad (11)$$

где нечеткая операция OR ( $\vee$ ) определяется следующим образом:

$$(p \vee q)_j = \max(p_j, q_j). \quad (12)$$

Этап 4б. Если вывод сделан, то есть  $g > 0$ , функция соответствия  $m_K^b$  прогноза  $x^b$  и входного вектора  $B$  слоя  $F_1^b$  вычисляется следующим образом:

$$m_K^b = \frac{|x^b \wedge B|}{|B|} \quad (13)$$

Если полученное значение  $m_K^b$  удовлетворяет критерию соответствия модуля ART<sub>b</sub> ( $m_K^b \geq \rho_b$ ), возникает резонанс. Активированные нейроны слоев  $F_2^a$  и  $F_2^b$  обучаются по формулам, аналогичным формулам (5) и (6) соответственно. Несовпадение прогнозов активирует процесс проверки сходства (см. Рисунок 1). С помощью процедуры поиска минимальных совпадений определяется нейрон  $j$ , который имеет минимальное значение функции соответствия. Функция выбора  $T_j^a$  нейрона  $j$  устанавливается

равной нулю. Величина параметра  $\rho_a$  в модуле  $ART_a$  повышается до чуть более высокого значения, чем соответствие, достигнутое нейроном  $j$ :

$$\rho_a := \max\{\rho_a, \min_j m_j^a + \epsilon\}. \quad (14)$$

Идеальное соответствие возникает, когда система пытается увеличить  $\rho_a$  до значения, равного единице. Идеальное соответствие в  $ART_a$  в сочетании с несоответствием в  $ART_b$  указывает на существование противоречивых знаний, где идентичные входные наборы связаны с разными результатами классификации.

После проверки сходства новый цикл работы сети повторяется с третьего этапа.

### 3.2 Извлечение правил

Из сети ARTMAP правила могут быть получены проще, чем из других сетей, в которой роли скрытых слоев явно не выражены. В сети ARTMAP каждый нейрон в слое  $F_2^a$  представляет собой категорию распознавания входных шаблонов  $ART_a$ . Через промежуточный слой  $F^{ab}$  каждый такой нейрон связан с категорией  $ART_b$  в слое  $F_2^b$ , которая, в свою очередь, кодирует результат классификации. Обученные весовые векторы веса слоя  $F_2^a$ , запоминают набор правил, связывающих входные данные с результатами (Рисунок 3). Количество правил равно количеству нейронов слоя  $F_2^a$ , которые сформировались во время обучения.

Как правило, большие наборы данных заставляют ARTMAP генерировать слишком много правил для их практического применения. Задача извлечения правил состоит в том, чтобы выбрать небольшой набор правил для хорошо прогнозируемых категорий и описать их в понятной форме. Для оценки каждой категории рассчитывается коэффициент достоверности, который определяет точность правила, соответствующего этой категории. Удаление категорий распознавания с низкой степенью достоверности, созданных на нетипичных примерах, приводит к созданию сетей меньшего размера. Удаление избыточных весов в весовом векторе категории уменьшает количество посылок в соответствующем правиле.

Коэффициент достоверности рассчитывается для каждой категории слоя  $F_2^a$  с точки зрения частоты встречаемости этой категории в обучающей выборке и точности классификации объектов, относящихся к данной категории. Для оценки коэффициента достоверности, для каждой  $j$ -й категории слоя  $F_2^a$  вычисляются два показателя: доля  $c_j$  соотношенных категории  $j$  обучающих векторов и доля  $s_j$  объектов, правильно классифицированных категорией  $j$  (относительно количества всех объектов, относящихся к данной категории).

Коэффициент достоверности узла  $j$  вычисляется следующим образом

$$k_j = \gamma c_j + (1 - \gamma) s_j, \quad (15)$$

где  $\gamma \in [0, 1]$  — это весовой коэффициент.



## ЗАКЛЮЧЕНИЕ

Присущие сетям ART характеристики, в частности, возможности быстрого и инкрементного обучения, в котором входные данные непрерывно используются для расширения знаний в существующей модели и совместимость со знаниями на основе правил, представляют собой вычислительную парадигму, принципиально отличающуюся от других систем машинного обучения. Такие алгоритмы могут найти применение для обработки потоковых данных или больших данных, позволяя “на лету”, без предварительной процедуры обучения, извлекать из этих данных знания и формулировать эти знания на понятном языке правил. При этом по мере обработки новых объемов данных правила будут корректироваться и уточняться, при этом с каждым правилом будет связан коэффициент достоверности, показывающий актуальность этого правила в данный момент времени.

## ЛИТЕРАТУРА

1. Molnar C. Interpretable Machine Learning [Электронный ресурс]: Perfect-bound Paperback. – 2019. – 318 P. – Режим доступа: <https://christophm.github.io/interpretable-ml-book/> (дата обращения: 08.09.2019)
2. Tan A.-H. Cascade ARTMAP: Integrating Neural Computation and Symbolic Knowledge Processing. // IEEE Trans.on Neural Networks, 1997, vol. 8, n.2.
3. Carpenter, G.A., Grossberg, S. Adaptive Resonance Theory// The Handbook of Brain Theory and Neural Networks. Cambridge, MA: MIT Press. - 2003, pp. 87-90.
4. Каширина И.Л., Федутинов К.А. Кластеризация непрерывного потока данных на основе обобщенной модели нейронной сети семейства ART//Системы управления и информационные технологии. - 2018. Т. 71. № 1. С. 33-39.
5. Carpenter G.A., Grossberg S., Reynolds J.H. ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network // Neural Networks. – 1991. – № 4. – P. 565-588.
6. Каширина И.Л., Львович Я.Е., Сорокин С.О. Нейросетевое моделирование формирования кластерной структуры на основе сетей ART // Информационные технологии. 2017. Т. 23. № 3. С. 228-232.
7. Каширина И.Л., Федутинов К.А. Применение сети Fuzzy ARTMAP в интеллектуальных системах обнаружения вторжений// Моделирование, оптимизация и информационные технологии. 2018. Т. 6. № 3 (22). С. 243-257.

I.L. Kashirina, K.A. Fedutinov  
**CONSTRUCTION OF DECISION RULES USING THE ARTMAP  
NEURAL NETWORK**

*Voronezh State University, Voronezh, Russia*

*This article discusses the ARTMAP neural network architecture, compatible with a symbolic representation based on IF-THEN rules. In particular, the knowledge gained during the training of the ARTMAP network can be transformed into a compact set of decision rules for classifying the source data, which can be analyzed by domain experts, by analogy with interpreted machine learning methods, such as decision trees or linear regression. Similarly, knowledge in the a priori area presented in the form of IF-THEN rules can be transformed into the ARTMAP neural network architecture. The presence of a preliminary set of rules used in the initialization of the network increases the accuracy of classification and the effectiveness of training. The original set of rules can be supplemented using the learning algorithm ARTMAP. Each rule formed in the process of learning a network has a confidence factor that can be interpreted as its importance or usefulness. The architecture, training algorithms and functioning of the ARTMAP network for the extraction of rules are described in terms of the previously proposed generalized model of networks of the ART family proposed by the authors.*

**Keywords:** neural network, machine learning, adaptive resonance theory, ARTMAP, rule extraction.

#### REFERENCES

1. Molnar C. Interpretable Machine Learning [Elektronnyy resurs]: Perfect-bound Paperback. – 2019. – 318 R. – Rezhim dostupa: <https://christophm.github.io/interpretable-ml-book/> (data obrashcheniya: 08.09.2019)
2. Tan A.-H. Cascade ARTMAP: Integrating Neural Computation and Symbolic Knowledge Processing. // IEEE Trans.on Neural Networks, 1997, Vol. 8, n.2.
3. Carpenter, G.A., Grossberg, S. Adaptive Resonance Theory// The Handbook of Brain Theory and Neural Networks. Cambridge, MA: MIT Press. - 2003, pp. 87-90.
4. Kashirina I.L., Fedutinov K.A. Klasterizatsiya nepreryvnogo potoka dannykh na osnove obobshchennoy modeli neyronnoy seti semeystva ART//Sistemy upravleniya i informatsionnye tekhnologii. - 2018. Vol. 71. No. 1. pp. 33-39.
5. Carpenter G.A., Grossberg S., Reynolds J.H. ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network // Neural Networks. – 1991. – Vol. 4. – pp. 565-588.
6. Kashirina I.L., L'vovich Ya.E., Sorokin S.O. Neyrosetevoe modelirovanie formirovaniya klasternoy struktury na osnove setey ART // Informatsionnye tekhnologii. 2017. Vol. 23. No. 3. pp 228-232.
7. Kashirina I.L., Fedutinov K.A. Primenenie seti Fuzzy ARTMAP v intellektual'nykh sistemakh obnaruzheniya vtorzheniy// Modelirovanie, optimizatsiya i informatsionnye tekhnologii. 2018. Vol. 6. No. 3 (22). pp. 243-257.