

МЕТОД ВЫБОРА НАПРАВЛЕНИЯ ДЛЯ УСКОРЕНИЯ КЛАССИФИКАЦИИ С ПОМОЩЬЮ ВЕРОЯТНОСТНОЙ НЕЙРОННОЙ СЕТИ АДАПТИВНОГО РЕЗОНАНСА

Д.В. Постарнак

аспирант, e-mail: dmi1852@yandex.ru

Тюменский государственный университет

Аннотация. В статье представлен новый метод, обеспечивающий ускорение классификации посредством класса нейронных сетей, базирующегося на теории адаптивного резонанса. Реализация метода возможна в рамках новой модели адаптивного резонанса, имеющей название «Вероятностная модель АРТ». Метод полезен в задачах, требующих максимального быстрогодействия в поддержке принятия решений, в задачах распознавания химических соединений движущихся жидкостей, например.

Ключевые слова: нейрон, теория адаптивного резонанса, слой распознавания, синаптический вес.

1. Введение

Существуют задачи распознавания образов, которые требуют максимального быстрогодействия в поддержке принятия решений, например, задачи распознавания химических соединений движущихся жидкостей. Статья содержит описание нового метода, обеспечивающего ускорение распознавания посредством класса нейронных сетей, базирующегося на теории адаптивного резонанса. Реализация метода возможна в рамках новой модели адаптивного резонанса, имеющей название «Вероятностная модель АРТ».

2. Описание вероятностной модели АРТ

В ходе исследования различных видов нейронных сетей адаптивного резонанса для реализации метода выбора направления разработана вероятностная модель АРТ [1-5] (АРТ-модель). Название класса нейронных сетей на основе теории адаптивного резонанса АРТ происходит от английской аббревиатуры АРТ — adaptive resonance theory. Вероятностная АРТ-модель разработана автором на основе модели АРТ-1, подробно описанной в работе [6]. АРТ-1 имеет два слоя: слой сравнения и слой распознавания (рис. 1).

Остальные блоки модели АРТ-1 отражают технические аспекты её реализации, поэтому в данной работе они не рассматриваются. Слой сравнения состоит

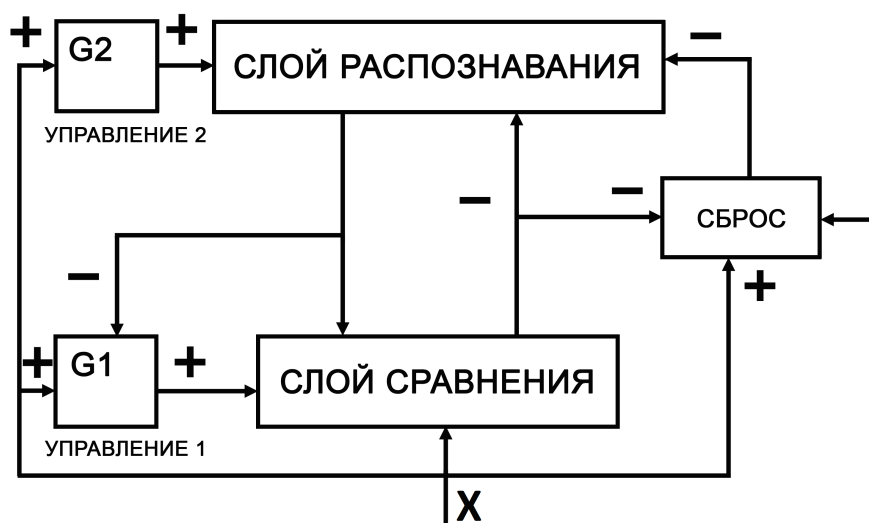


Рис. 1. Схема модели ART-1

из входных нейронов (в дальнейшем — входы). Его функция состоит в том, чтобы принимать поступающий в модель образ и сравнивать его с имеющимися образами в текущей конфигурации модели. Входной образ является бинарным: сигнал каждого входа — 1 или 0.

Каждый нейрон слоя распознавания связан со всеми входами. Активация нейрона слоя распознавания осуществляется в том случае, когда входы в общей совокупности несут образ, наиболее соответствующий образу, закреплённому за этим нейроном. Активированный нейрон называется *нейроном-победителем*. При этом выходной сигнал нейрона слоя распознавания должен соответствовать уровню, зависящему от соотношения образа активированного нейрона и текущего образа. Это одна из функций слоя распознавания. Если это условие не выполняется, процесс определения активированного нейрона повторяется, но уже без участия предыдущего нейрона. О такой ситуации говорят как об отсутствии адаптивного резонанса. Если ни один из нейронов слоя распознавания не выдал необходимого уровня сигнала, тогда слой дополняется новым нейроном, за которым будет зафиксирован поступивший образ. Если же активированный нейрон выдал достаточный уровень сигнала, тогда происходит адаптивный резонанс, то есть синаптические веса активированного нейрона модифицируются с целью обобщения предыдущего образа, закреплённого за данным нейроном, с текущим образом. То есть нейрон объединяет в себе возможности для распознавания нескольких максимально схожих образов. Такой нейрон в дальнейшем именуется как *сработавший нейрон*.

ART-1 была выбрана для исследования в результате критического анализа, описанного в статье [7]. Критерием выбора стало наличие возможности самообучения и принятия решения при поступлении на вход незнакомого образа.

В вероятностной ART-модели слой распознавания представлен в виде ориентированного графа без циклов (рис. 2).

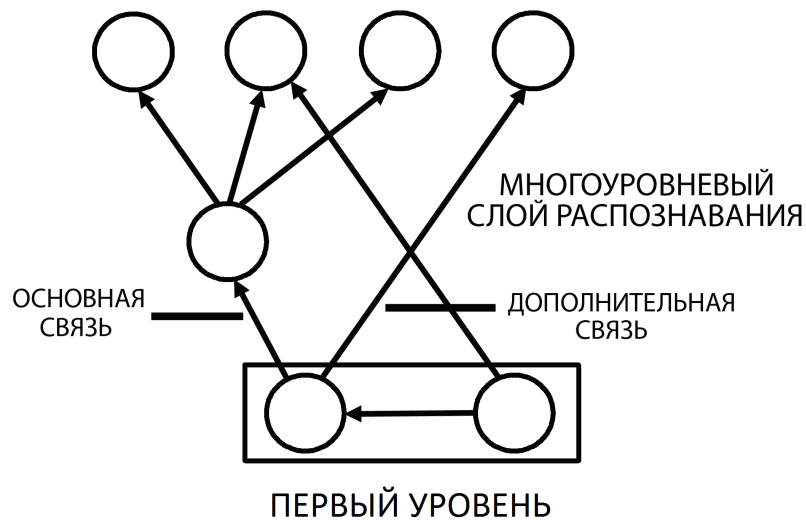


Рис. 2. Многоуровневый слой распознавания

За каждым нейроном слоя распознавания в этой модели закреплён не весь образ, а его конкретный участок (*локальное рецептивное поле*). Этот граф также содержит набор вершин, составляющих так называемый *первый уровень* (исток графа). С рассмотрения нейронов первого уровня всегда начинается классификация.

Преимущество вероятностной АРТ-модели перед АРТ-1 – в возможности использовать данные о вероятностях появления определённых фрагментов в соответствующих местах *глобального рецептивного поля* (области фигурирования образа).

3. Алгоритм метода выбора направления

Под направлением понимается порядок рассмотрения нейронов слоя распознавания. Алгоритм представлен следующим образом:

1. Вычисляются выходные сигналы нейронов первого уровня. Конкурировать по уровню сигнала могут только нейроны с общим *локальным рецептивным полем*:

$$F = \sum_l b_l s_l, s \in \{1, 0\}, \quad (1)$$

где F – выходной сигнал нейрона; l – порядковый номер входа, связанного с нейроном слоя распознавания; s – сигнал входа; b – вес связи между входом и нейроном слоя распознавания.

2. Для *нейрона-победителя* необходимо проверить неравенство, определяю-

шее степень соответствия образа, закреплённого за нейроном, входному образу:

$$\sum_l t_l / \sum_l s_l > \rho, \quad \rho < 1, \tag{2}$$

где l — номер входа, связанного с текущим нейроном слоя распознавания; s — сигнал входа ($s \in \{0, 1\}$); t — элемент образа, закреплённого за нейроном слоя распознавания (отражение того, что в идеальном случае должно быть на месте соответствующего входа s). Величина ρ является переменным параметром нейронной сети, $\rho \in (-\infty, 1)$. Высокий параметр ρ приведёт к повышению числа нейронов слоя распознавания, а низкий параметр ρ приведёт к высокой степени обобщения образов в нейронах.

3. Если степень соответствия образа нейрона входному образу удовлетворительна (неравенство (2) выполняется), это означает, что нейрон сработал, то есть стал частью ответного результата нейронной сети. При этом необходимо уточнить для *сработавшего нейрона* вектор весов связей по формуле, аналогичной той, которая применяется для модели АРТ-1 [6]:

$$b < \frac{W}{W - 1 + \sum_l s_l}, \quad W > 1, \tag{3}$$

где $W \in (1, \infty)$ — параметр, определяющий диапазон между максимальным и минимальным весом (имеет смысл при технической реализации); l — номер входа, связанного с нейроном слоя распознавания; s — сигнал входа. Следует отметить, что образ, закреплённый за *сработавшим нейроном* (элементы t на рис. 3), модифицируется путём присвоения элементам t значений соответствующих элементов s при условии $t \neq 0$ (принцип логического «и»).

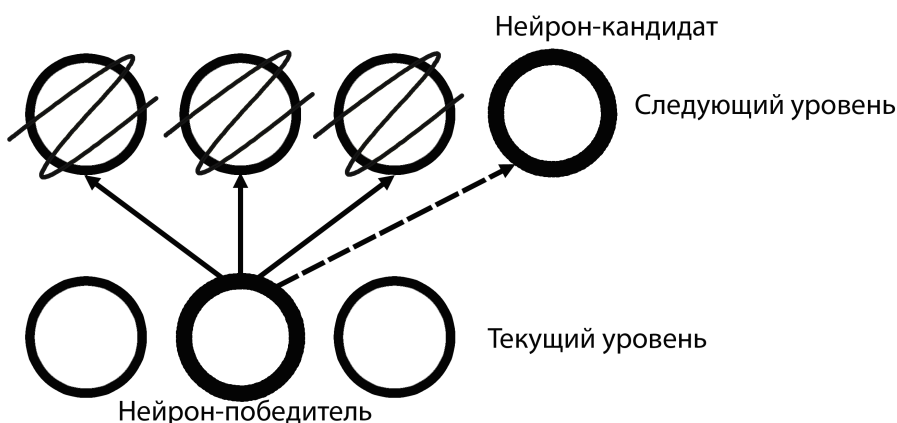


Рис. 3. Элементы t

4. Если степень соответствия образа нейрона входному образу неудовлетворительна (неравенство (2) не выполняется), тогда повторно вычисляются сигналы нейронов слоя распознавания, но без учёта предыдущего нейрона (возврат к шагу 1).

5. Если степень соответствия образа нейрона входному образу неудовлетворительна ни для одного нейрона, тогда в текущем уровне необходимо выделить *нейрон-кандидат* (нейрон, претендующий на то, чтобы быть добавленным в слой распознавания) (рис. 4).

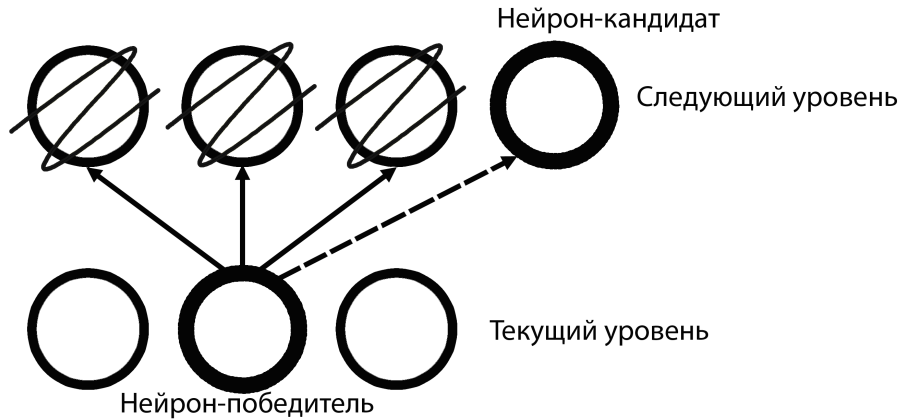


Рис. 4. Выделение нейрона-кандидата

Точный образ *нейрона-кандидата* (элементы t) равен фрагменту входного образа для текущего нейрона. Нейрон-кандидат должен быть связан с предыдущим сработавшим нейроном. Шаги 6 и 7 не являются обязательными.

6. После того как нейрон-кандидат сформирован, необходимо проверить его на возможность замены другим, уже имеющимся, нейроном. Для этого необходимо вычислить характеристики подобия образов K между имеющимся нейроном (любым) и нейроном-кандидатом:

$$K = \sum_l t_l \& s_l, \quad (4)$$

где t — элемент образа текущего сравниваемого нейрона; s — элемент образа нейрона-кандидата, $\&$ — логическое «и».

7. Если все характеристики $K < \varphi$, тогда *нейрон-кандидат* может быть добавлен в систему (рис. 5).

В противном случае вместо *нейрона-кандидата* будет использоваться текущий сравниваемый нейрон (рис. 6).

Величина $\varphi > 0$ является переменным параметром нейронной сети и зависит от желаемой степени сходства образов. При добавлении *нейрона-кандидата* в состав нейронов слоя распознавания классификация не останавливается, поскольку результат может быть завершён с помощью словарного уточнения образа без учёта нового нейрона (завершение классификации по неполному набору сработавших нейронов, с уточнением оставшихся нейронов по словарю).

8. Для *сработавшего нейрона* (или для нового нейрона) вычисляется величина E^* , которая является *скалярным видом образа*. *Скалярный вид образа*

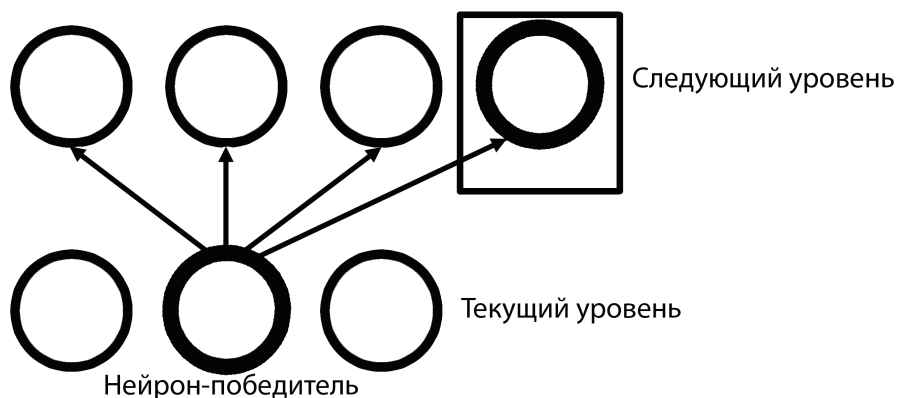


Рис. 5. Добавление нейрона-кандидата в систему

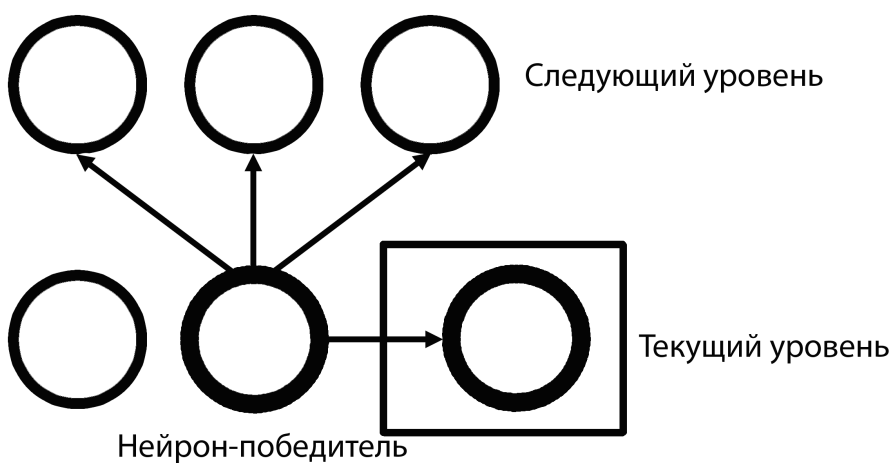


Рис. 6. Отмена нейрона-кандидата

— это целое число, однозначно определяющее весь образ:

$$E^* = \sum_l s_l \times 2^{l-1}, \tag{5}$$

где E^* — скалярный вид образа;

l — порядковый номер входа, связанного с нейроном слоя распознавания;

s — сигнал входа.

9. Осуществляется построение *определяющих векторов* H для каждого нейрона *следующего уровня* (т.е. для нейрона, связанного с последним *сработавшим нейроном*). Каждый определяющий вектор должен соответствовать равенству:

$$E^* = E + \sum_j h_j, h_j \in \{2^u; 0\}, j \in [1 : V], u \in [0 : \infty), \tag{6}$$

где E^* — скалярный вид образа, вычисленный на предыдущем шаге;
 E — скалярный вид образа, вычисленный для *сработавшего нейрона* на предыдущем шаге, но на более ранних этапах работы нейронной сети и закреплённый (сохранённый) за связанным нейроном, который на ранних этапах был следующим *сработавшим*. При первом использовании алгоритма данный шаг пропускается, поскольку все E равны нулю;
 h — элемент *определяющего вектора*;
 j — порядковый номер элемента *определяющего вектора*;
 V — число входов *локального рецептивного поля сработавшего нейрона*.

Необходимо отметить, что при добавлении *нейрона-кандидата* в состав слоя распознавания, величина E этого нейрона должна быть равна *скалярному виду образа* предыдущего *сработавшего нейрона*.

10. Сигналы нейронов *следующего уровня* (нейроны, связанные с последним *сработавшим* связью от него) в случае, когда для этих нейронов формула (6) даёт **определяющий вектор** с минимумом элементов больше нуля вычисляются в первую очередь. Это осуществляется при возврате к шагу 1. Можно ориентироваться не на минимум, а на диапазон, охватывающий близкие к минимуму варианты значений. При срабатывании очередного нейрона его хранимая величина E получает значение *скалярного вида образа* предыдущего *сработавшего нейрона*. Необходимо отметить, что использование именно *скалярного вида образа* вместо вектора обусловлено более высокой потенциальной производительностью данного алгоритма, так как скалярные величины при компьютерной реализации обрабатываются быстрее, чем векторы.

Алгоритм завершается при получении набора *сработавших нейронов*, достаточного для классификации или для её завершения с помощью словарного уточнения.

4. Структура слоя распознавания

Каждый нейрон слоя распознавания вероятностной АРТ-модели представлен совокупностью компонентов Q :

$$Q = \langle G, I, J, Z \rangle. \quad (7)$$

Здесь G — вектор, содержащий набор порядковых номеров нейронов слоя распознавания, связанных с текущим нейроном (связь направлена от текущего нейрона):

$$G = (x_1, x_2, \dots, x_n), \quad E > 0, \quad (8)$$

где x_i ($i = 1, 2, \dots, n$) — порядковый номер нейрона слоя распознавания, связанного с текущим.

Вектор I хранит информацию о том, от каких нейронов существует связь к текущему (в отличие от компонентов вектора G здесь направление связей к текущему нейрону, а не от него), причём каждый порядковый номер связанного нейрона сопровождается скалярным видом образа E_i ($i = 1, 2, \dots, n$), сохранённым на предыдущих этапах работы нейронной сети. Скалярный вид

образа, как было обозначено выше, определяет скалярную величину, которая однозначно выражает образ, закреплённый за нейроном слоя распознавания.

$$I = (E_1, y_1, E_2, y_2, \dots, E_n, y_n), \quad E \geq 0, \quad (9)$$

где y_i ($i = 1, 2, \dots, n$) — порядковый номер нейрона слоя распознавания, связанного с текущим нейроном; E_i ($i = 1, 2, \dots, n$) — хранимый скалярный вид образа.

Вектор J в формуле (7) определяет набор входов, весовые коэффициенты связи нейрона слоя распознавания с входом и элементы образа, на который настроен нейрон (идеального образа) по отношению к соответствующим входам:

$$J = (l_1 b_1 t_1, l_2 b_2 t_2, \dots, l_n b_n t_n), \quad n = V, \quad (10)$$

где l_i ($i = 1, 2, \dots, n$) — порядковый номер входа; b_i ($i = 1, 2, \dots, n$) — весовой коэффициент связи нейрона слоя распознавания с входом; t_i ($i = 1, 2, \dots, n$) — элемент образа, на который настроен нейрон; V — число входов локального рецептивного поля (эти поля могут отличаться по размеру при условии, что они не перекрываются, и полностью охватываются соответствующими нейронами, конкурирующими между собой).

Компонент Z — это скаляр, который определяет принадлежность нейрона к первому уровню (если равен 1):

$$Z \in \{1, 0\}. \quad (11)$$

Если модель не содержит ни одного нейрона в слое распознавания, то необходимо для каждого целого образа определить набор нейронов слоя распознавания, *локальные рецептивные* поля которых не будут пересекаться. Связь между нейронами должна полностью охватывать распознаваемый образ. При этом всё пространство входных нейронов должно быть использовано. В качестве нейронов первого уровня необходимо использовать те, которые отражают отличия образов. Остальные характеристики должны быть следующие:

- вектор I нулевой;
- элементы b вектора J задаются по формуле, аналогичной той, которая применяется для присвоения начальных значений нейронам в модели АРТ-1 [6]:

$$b < \frac{W}{W - 1 + V}, \quad W > 1, \quad (12)$$

где $W \in (1, \infty)$ — параметр, определяющий диапазон между максимальным и минимальным весом (имеет смысл при технической реализации); V — число входов *локального рецептивного поля*, закреплённых за текущим нейроном слоя распознавания;

- все элементы t вектора J равны 1.

5. Обоснование метода выбора направления

Для лучшего восприятия метода рассмотрим задачу распознавания рукописных римских цифр «I», «V» и «X». Для каждого распознаваемого символа

сформирована совокупность из 81 входного нейрона (9x9). За каждой группой конкурирующих нейронов слоя распознавания (с общим *локальным рецептивным полем*) закреплено поле размером 3x3. На рис. 7 эти поля разделены жирной линией.

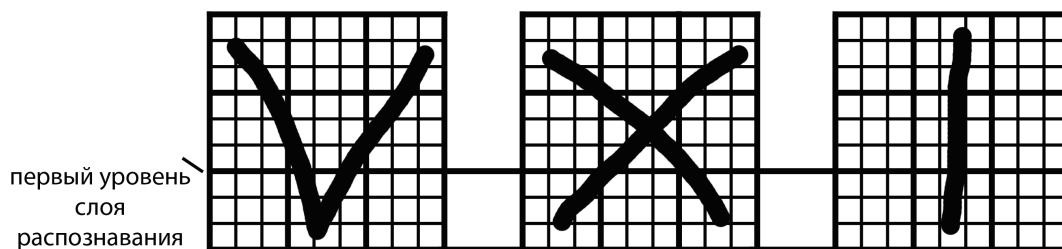


Рис. 7. Примеры локальных рецептивных полей

Процесс распознавания начинается с *первого уровня*. Очевидно, что, если активизируется только крайний правый нейрон среди нейронов *первого уровня*, как показано на рис. 8, то вероятнее всего, что распознаваемый символ является символом «X», поскольку символы «V» и «I» в среднестатистическом случае будут активировать средний нейрон *первого уровня*.

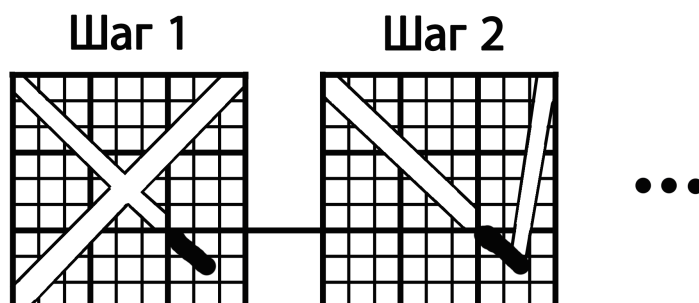


Рис. 8. Вероятность появления образа

Следовательно, дальнейшее рассмотрение нейронов для символа «X» следует производить в тех областях глобального рецептивного поля, в которых этот символ чаще всего представлен. Если вероятность не оправдывается, тогда переходим к следующему шагу и будем рассматривать области, которые соответствуют другому образу, имеющему менее распространённый вариант отображения, например символ «V», поскольку очевидно, что для текущего результата распознавания в правом нейроне первого уровня символ «I» является наименее вероятным.

Таким образом, применение метода выбора направления целесообразно, а конкретная степень увеличения скорости распознавания (по сравнению с моде-

лью ART-1) зависит от распознаваемых образов, то есть от предметной области, в которой данная нейронная сеть может функционировать. Метод, в целом, основан на анализе распределения вероятностей.

Заключение

В работе представлен метод выбора направления, обеспечивающий ускорение классификации. Разработан алгоритм, реализующий этот метод. В результате найден вариант ускорения классификации бинарных образов в нейронных сетях класса ART.

ЛИТЕРАТУРА

1. ART 2: Self-organization of stable category recognition codes for analog input patterns / Gail A. Carpenter and Stephen Grossberg // *Applied Optics*. 1987. Vol. 26, № 23.
2. ART 3: Hierarchical Search Using Chemical Transmitters in Self-Organizing Pattern Recognition Architectures / Gail A. Carpenter and Stephen Grossberg // *Neural Networks*. 1990. Vol 3. P. 129–152.
3. Fuzzy ARTMAP: A Neural Network Architecture for Incremental Supervised Learning of Analog Multidimensional Maps / Gail A. Carpenter, Stephen Grossberg, Natalya Markuzon, John H. Reynolds, David B. Rosen // *Transactions on neural networks*. 1992. Vol 3, № 5.
4. ARTMAP: Supervised Real-Time Learning and Classification of Nonstationary Data by a Self-Organizing Neural Network / Gail A. Carpenter, Stephen Grossberg, John H. Reynolds // *Neural Networks*. 1991. Vol 4. P. 565–588.
5. Fuzzy ART: Fast Stable Learning and Categorization of Analog Patterns by an Adaptive Resonance System / Gail A. Carpenter, Stephen Grossberg, David B. Rosen // *Neural Networks*. 1991. Vol. 4. P. 759–771.
6. Терехов С.А. Лекции по теории и приложениям искусственных нейронных сетей / Лаборатория Искусственных Нейронных Сетей НТО-2. Снежинск : ВНИИТФ, 1998. 66 с.
7. Постарнак Д.В. Критический анализ моделей нейронных сетей // *Вестник ТюмГУ*. 2012. № 4. С. 162–167.

**DIRECTION CHOICE METHOD FOR INCREASING SPEED
OF CLASSIFICATION USING PROBABILISTIC ART MODEL****D.V. Postarnak**

Postgraduate Student, e-mail: dmi1852@yandex.ru

Tumen State University

Abstract. This article presents a new method that increases recognition speed using a class of adaptive resonance theory neural networks. The method implementation is possible within a new adaptive resonance model named as "Probabilistic ART model". The method is useful in the tasks demanding the fastest decision-making, e.g., in recognition of chemical compounds in the moving liquids.

Keywords: neuron, recognition layer, synaptic weight, adaptive resonance theory.