

УДК 004.93'1

DOI 10.18413/2411-3808-2018-45-4-709-717

СЕТЬ АДАПТИВНО-РЕЗОНАНСНОЙ ТЕОРИИ С МНОГОУРОВНЕВОЙ ПАМЯТЬЮ

NETWORK OF ADAPTIVE-RESONANCE THEORY WITH MULTILEVEL MEMORY

Д.Г. Буханов, В.М. Поляков
D.G. Bukhanov, V.M. Polyakov

Белгородский государственный технологический университет им. В.Г. Шухова,
Россия, 308012, Белгород, ул. Костюкова, 46

Belgorod State Technological University named after V.G. Shukhov,
46 Kostyukova St., Belgorod, 308012, Russia

E-mail: db.old.stray@gmail.com

Аннотация

В работе проведен анализ искусственной нейронной сети на основе адаптивно-резонансной теории с непрерывными входными сигналами. Выявлены недостатки сети и проблемы при ее использовании. Основной проблемой, которая решается в работе, является низкая скорость поиска активного резонирующего нейрона памяти в поле F2, что приводит к менее точному распознаванию объектов систем реального времени. Для решения проблемы предложена новая модель поля распознавания F2 ART-2 сети, которая представляет собой древовидную структуру с рекуррентно изменяющимся параметром сходства для каждого последующего уровня. На каждом уровне происходит увеличение параметра сходства, что приводит к последовательному поиску активного резонирующего нейрона. Проведен сравнительный анализ временных характеристик предложенной сети и классической реализации. Экспериментально доказано преимущество предложенной модели сети на основе адаптивно-резонансной теории.

Abstract

The analysis of the artificial neural network based on the adaptive-resonance theory with continuous input signals is described in the paper. Reviewed general shortcomings of the network and the problems with its using. The main problem is low search speed of the active resonating memory neuron in the field F2, which leads to less accurate recognition of objects in real-time systems. To solve the problem, a new model of the recognition field F2 of the ART-2 network is proposed, which is a tree structure, with a recurrently changing similarity parameter for each subsequent level. At each level, the similarity parameter increases, that leads to a sequential search for an active resonating neuron. A comparative analysis of the time characteristics of the proposed network and the classical implementation is carried out. The advantages of the proposed network model based on the adaptive resonance theory are experimentally proved. Network with the proposed modification shows better speed than the classic ART-2 without loss of accuracy.

Ключевые слова: искусственные нейронные сети, адаптивно-резонансная теория, кластеризация, интеллектуальный анализ данных.

Keywords: artificial neural networks, adaptive resonance theory, clustering, data mining.

Введение

В настоящее время все больше задач автоматизированного и автоматического распознавания технических состояний объекта решаются с использованием интеллектуальных методов анализа [Григорьева и др., 2017; Хорсова 2016]. Одними из таких методов являются методы с применением искусственных нейронных сетей (ИНС) [Некрасова,

2015]. Развитие применения ИНС происходит по нескольким направлениям [Мельникова, Михайличенко, 2017]: сфера техники и телекоммуникаций [Saeks, 1997], информационные технологии при работе с текстами, методы распознавания текстов [Wang et al, 2012], также получили широкое распространение использование нейронных сетей для определения тональности текста [Будыльский, Подвесовский, 2015; Kim, 2014]. Следующим направлением является применение ИНС в области экономики и финансов [Heaton et al, 2016], рекламы и маркетинга [Wang et al, 2016].

ИНС могут использоваться для распознавания образов, представленных как графическими изображениями [Григорьева и др., 2017; Корсунов и др., 2015], так и числовыми параметрами [Вовченко и др., 2016]. Для распознавания образов, представленных числовыми параметрами, хорошо себя зарекомендовала сеть, описанная Гроссбергом и Карпентером [Carpenter, Grossberg, 1987; Carpenter et al, 1991]. Предложенная ими сеть адаптивной резонансной теории (АРТ) с дискретными входами получила свое распространение при диагностировании цифровых устройств [Дмитриенко и др., 2004]. Из-за гибкости структуры сети существуют работы, где ее используют для управления автоматическими системами [Grosspietsch, Silayeva, 2012]. Существуют модификации АРТ-2, которые могут работать с нечеткими параметрами входов [Majeed et al, 2018]. Архитектура сети, которая работает с непрерывными значениями входных изображений, получила краткое название АРТ-2. Структуры таких сетей состоят из трех типов полей: входного поля сравнения, выходного поля распознавания и решающего поля (модуля сброса), который формирует управляющее воздействие на поле распознавания. Основным принцип работы этих сетей предполагает нахождение соответствия восходящего сенсорного сигнала ожидаемому нисходящему сигналу. Одно из возможных применений такой сети – распознавание текущего состояния компьютерной сети для диагностирования возможных отказов [Буханов и др., 2017].

1. Классическое представление сети АРТ-2. На рис. 1 изображена структура ИНС АРТ-2. Она состоит из трех основных полей $F1$, $F2$ и G [Дмитриенко и др., 2004]. Поле $F1$ можно представить кортежем $F1 = \{S, W, X, V, U, Q, P, WN, VN, PN\}$, где первые семь элементов представляют собой слои нейронов обработки АРТ-2, а последние три – нормализующие элементы. Поле $F2$ состоит из распознающих нейронов Y_j ($j = 1..m$, где m – количество различных классов образов), $F2 = \{Y\}$. Поле G представляет собой кортеж, состоящий из управляющих нейронов R_i (где $i = 1..n$, где n – количество входных параметров), $G = \{R, RN\}$, где RN нормализующий элемент.

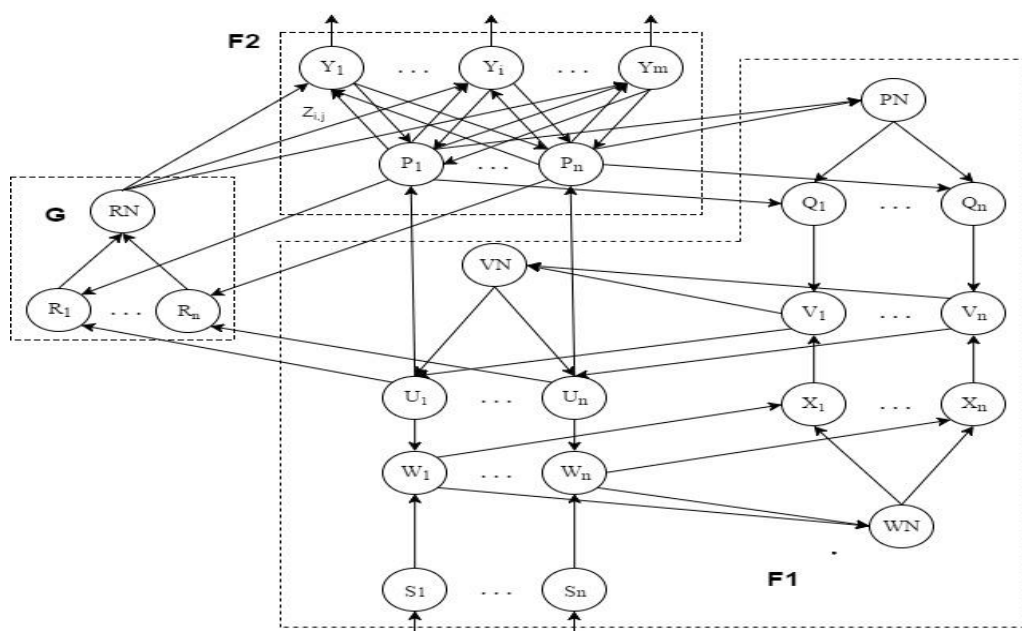


Рис. 1. Структура АРТ-2

Fig. 1. Structure of the ART-2 network

В работах [Carpenter, Grossberg, 1987; Carpenter et al, 1991] доказаны теоремы, которые описывают принципы работы сетей АРТ. Основными следствиями теорем являются:

- процесс поиска ранее обученного образа является устойчивым, т. е. после определения выигравшего нейрона в сети не будет возбуждений других нейронов, только сигнал сброса может дать возможность им активироваться;
- процесс обучения устойчив, обучение весов нейрона-победителя не приведет в дальнейшем к переключению на другой нейрон.

Процессы обучения и распознавания при такой структуре сети не имеют принципиальных отличий [Grossberg, 2013.]. Единственным отличием является то, что при обучении происходит изменение весов связей z_{ij} , организованными распознающими нейронами. А при распознавании изменение весов не происходит.

Алгоритм работы сети состоит из следующих шагов.

Шаг 1. На вход слоя S подаются вектор сигналов, состоящий из n элементов. Нейроны слоя W воспринимают сигналы слоя S и суммируют их с выходными сигналами слоя U .

$$w_i = s_i + au_i;$$

$$i = \overline{1..n}.$$

Шаг 2. Выходные сигналы w_i нейронов слоя W поступают на входы элементов слоя X и в нормализующий элемент WN , который вычисляет норму:

$$\|w\| = \sqrt{\sum_{i=1}^n w_i^2};$$

$$x_i = \frac{w_i}{e + \|w\|};$$

$$i = \overline{1..n};$$

где e – небольшая положительная константа, предотвращающая деление на ноль.

Шаг 3. Выходные сигналы слоя V определяются выражением:

$$v_i = f(x_i) + bf(q_i);$$

$$i = \overline{1..n};$$

где $f()$ – пороговая функция, для подавления шумовых сигналов:

$$f(x) = \begin{cases} x, & \text{если } x > \theta; \\ 0, & \text{иначе;} \end{cases}$$

где θ – порог шума.

Шаг 4. Выходные сигналы v_i нейронов слоя V поступают на входы элементов слоя U и в нормализующий элемент VN , который вычисляет норму:

$$\|v\| = \sqrt{\sum_{i=1}^n v_i^2};$$

$$u_i = \frac{v_i}{e + \|v\|};$$

$$i = \overline{1..n}.$$

Шаг 5. Сигналы p_i однозначно определяются u_i -ми. Выходные сигналы нейронов P -слоя подаются на нормализующий элемент PN и группу нейронов Q .

$$\|p\| = \sqrt{\sum_{i=1}^n p_i^2};$$

$$q_i = \frac{p_i}{e + \|p\|};$$

$$i = \overline{1..n}.$$

Затем выполняются шаги 3–5. Устойчивое состояние в поле F1 устанавливается после выполнения двух итераций.

Шаг 6. Выходные сигналы из слоя P поступают на распознающий слой Y, и вычисляется максимальный элемент y_i .

$$y_j = \sum_{i=1}^n z_{ij} p_i; \quad (1)$$

$$j = \overline{1..m};$$

$$i_{\max} = \max(Y);$$

затем находится выходной сигнал из слоя Y:

$$p_i = u_i + z_{i_{\max} i} d;$$

$$i = \overline{1..n};$$

где z – веса связей от нейронов слоя P к нейронам слоя Y.

Шаг 7. Рассчитываются сигналы элементов управляющего слоя.

$$r_i = \frac{u_i + cp_i}{e + \|u\| + c\|p\|}. \quad (2)$$

Нормализующий элемент RN рассчитывает следующий выходной сигнал.

$$\|r\| = \sqrt{\sum_{i=1}^n r_i^2}. \quad (3)$$

Шаг 8. Если $\|r\| \geq \rho$, (где ρ – параметр соответствия ожидаемого результата текущему, который изменяется в интервале $[0;1]$), то считается, что текущий активный нейрон слоя Y с индексом i_{\max} поля F2 является победителем. Затем происходит дообучение его весов.

$$z_{i_{\max} i} += d(1-d) \left(\frac{u_i}{1-d} - z_{i_{\max} i} \right);$$

$$i = \overline{1..n}.$$

Если $\|r\| < \rho$, то текущий активный нейрон слоя Y замораживается и далее в соревновании не участвует.

Шаг 9. Если не все нейроны слоя Y заморожены – переход к шагу 6, иначе создается новый нейрон в слое Y, m увеличивается на 1. Считается, что данный нейрон будет резонировать с новым образом, поэтому веса вычисляются следующим способом:

$$z_{mi} = d(1-d) \frac{u_i}{1-d}.$$

На рис. 2 показан график изменения значения веса z_1 резонирующего нейрона слоя Y при фиксированном приближенном выходном сигнале поля F1 ($U_1 \approx 0,62$).

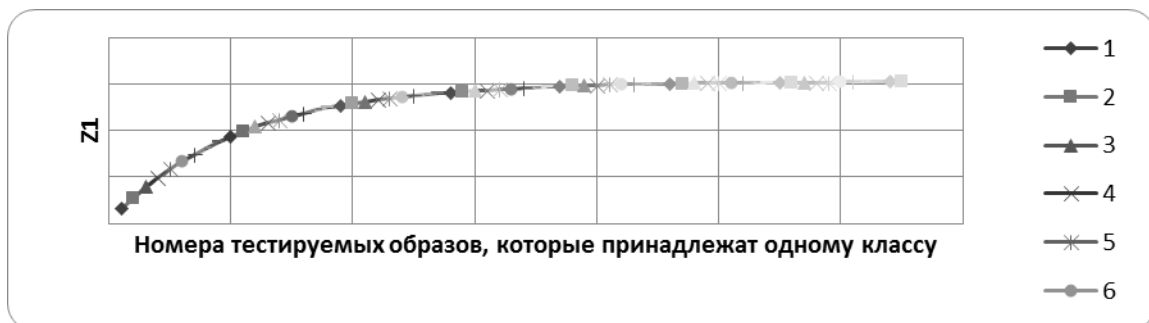


Рис. 2. График изменения весов z_1
Fig. 2. Changes of z_1 weights

Данный график показывает, что изменения весов памяти поля F2 стремится к некоторому пределу, это подтверждает теорему, сформулированную в работе [Carpenter, Grossberg, 1987], о конечности и устойчивости процесса обучения.

При использовании классической реализации сети были выявлены следующие проблемы:

1. Значительное увеличение времени распознавания при обучении сети большим количеством разных векторов данных с большим параметром сходства, так как под каждый класс образов создается отдельный нейрон в слое Y.

2. Невозможность распараллеливания процесса распознавания в виду того, что память, которая представлена матрицей весовых коэффициентов z, чувствительна к последовательности подачи распознаваемых образов.

3. Долгий поиск активного нейрона Y-слоя, так как поиск осуществляется только на основе максимального u_i (1), который характеризует частоту подачи образов. Чем чаще образ из класса изображений был подан в сеть, тем больше является u_j .

2. Структура АРТ-2 с самоорганизующейся памятью. Рассмотрим предложенную структуру ИНС АРТ-2 с модифицированной памятью (АРТ-2m). Она состоит из трех основных полей F1, F2, G. Поле F1 состоит из тех же элементов, что и в классической реализации АРТ-2, описанной выше в разделе 1.

Исходя из теории адаптивного резонанса, после проверки всех нейронов Y-слоя происходит добавление нового элемента u_m . Проверка соответствия восходящего сигнала к активному нисходящему сигналу из слоя Y происходит следующим образом:

$$\frac{\sqrt{\sum_{i=1}^n (u_i + cp_i)^2}}{\sqrt{(u_i)^2 + c\sqrt{(p_i)^2}}} > \rho;$$

где $p_i = u_i + z_{i \max} d; z_{i \max} = d(1 - d) \left(\frac{u_i}{1 - d} - z_{i \max} \right); i = \overline{1..n}$.

Отсюда вытекает утверждение, что резонанс между памятью и восходящим сигналом будет тогда и только тогда, когда u_i будет пропорционален z_i :

$$u_i \equiv z_i; i = \overline{1..n}$$

Это утверждение и теорема, экспериментально подтвержденная выше, позволяют организовать многоуровневую память с различными параметрами соответствия на каждом из уровней. Таким образом, поле F2 можно представить в виде слоев распознающих нейронов Y с разным уровнем детализации образов и M-слоев – семантических нейронов, связывающих Y различных уровней, $F2 = \{Y, M\}$. Тогда поле G представляет собой кортеж, состоящий из управляющих нейронов R_i (где $i = \overline{1..n}$, n – количество входных параметров), RN представляет собой нормализующий элемент, $Riter_p$ – пороговый элемент ($p = \overline{1..kc}$, kc – количество уровней памяти), $G = \{R, Riter, RN\}$. На рис. 3 представлена структура полей F2 и G с модифицированной памятью.

Для уменьшения времени поиска активного нейрона u_i структуру поля F2 можно представить в виде дерева памяти с конкретизацией меры сходства на каждом из уровней дерева Riter, причем обучение памяти, находящейся на одном уровне, не затрагивает веса лежащих выше элементов.

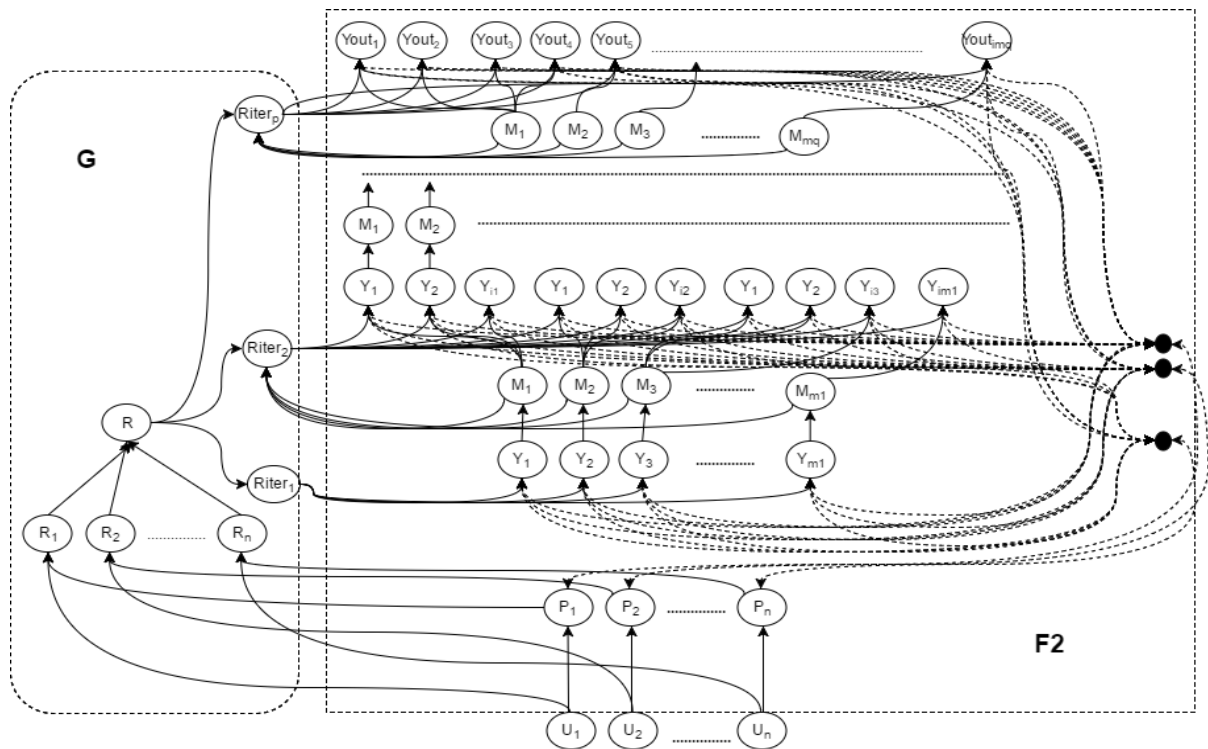


Рис. 3. Структура полей F2 и G с самоорганизующейся памятью
 Fig. 3. F2 and G field structure with self-organized memory

Алгоритм работы сети состоит из следующих шагов.

Шаги 1–5. При работе в поле F1 не имеют отличий от классической реализации, описанной выше.

Шаг 6. Требуется рассчитать $Riter$ для каждого уровня. Чем меньше разница между параметром сходства и $\|r\|$, тем больше количество различных нейронов Y -слоя будет создано, соответственно, требуется построить рекуррентную зависимость нахождения параметра сходства для каждого уровня дерева памяти. В работе была использована следующая рекуррентная зависимость с начальной мерой сходства 0.5:

$$Riter_1 = 0.5;$$

$$Riter_{k+1} = Riter_k + 0.75(1 - Riter_k);$$

$$i = \overline{1..kc}.$$

Шаг 7. Выходные сигналы из слоя P поступают на распознающий слой Y^k и вычисляется максимальный элемент y_j^k .

$$y_j^k = \sum_{i=1}^n z_{ij}^k p_i;$$

$$i \max^k = \max(Y^k);$$

$$p_i = u_i + z_{i \max^k}^k d;$$

$$i = \overline{1..n};$$

где z^k – веса связей от нейронов слоя P к нейронам слоя Y^k .

Шаг 8. Затем рассчитываются сигналы элементов управляющего слоя по формулам (2,3).

Шаг 9. Если $\|r\| \geq Riter_k$, то считается, что текущий активный нейрон слоя Y^k поля F2 является победителем. Затем происходит дообучение его весов и переход к шагу 10.

$$z_{i \max^k}^k += d(1 - d) \left(\frac{u_i}{1 - d} - z_{i \max^k}^k \right).$$

Если $\|r\| < Riter_k$, то текущий активный нейрон слоя Y^k замораживается и далее в соревновании не участвует. Если все нейроны текущего слоя Y^k заморожены или полностью отсутствуют, то выполняется переход к шагу 11.

Шаг 10. Переход на следующий уровень памяти, связанный с текущим активным нейроном слоя Y^k связями M^k .

Слой M^k хранит семантические связи нейронов слоя Y^k с Y^{k+1} . Если $k \leq k_c$, то выполняется шаг 7, иначе считается, что активный нейрон Y^{k+1} -слоя является тем, в весах которого и хранится конечный результат распознавания.

Шаг 11. Далее происходит создание нового нейрона слоя Y^k , увеличение m и обучение его весов по следующему правилу:

$$z_{mi}^k = d(1-d) \frac{u_i}{1-d}.$$

Предложенная структура сети АРТ-2 обладает следующими особенностями.

1. Поиск активного нейрона, в весах которого хранится образ, происходит теперь не только на основе частоты поданных изображений соответствующих классов, но и на основе семантических связей, формирующихся в момент обучения сети. Это значительно позволяет снизить количество проверок, так как вся память теперь представлена в виде древовидной структуры.

2. Возможность распараллеливания вычислений в поле F2, которая обусловлена невозможностью возбуждения нейрона Y^k -слоя, который семантически не связан с текущим.

3. Экспериментальные исследования временных характеристик работы АРТ-2m. Эксперимент проводился при 10 входах. В качестве данных для анализа были взяты случайные числа, принадлежащие интервалу (0;1000]. Оценивалась скорость обучения.

На рис. 4 представлены графики зависимостей скорости обучения от количества поданных разных образов для классической реализации нейронной сети АРТ-2 и предложенной АРТ-2m, с многоуровневой памятью.

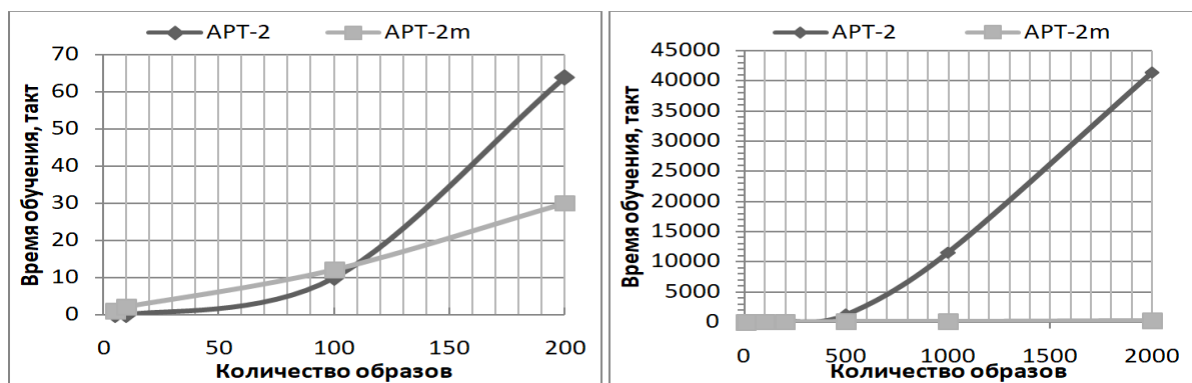


Рис. 4. Сравнительный анализ времени обучения АРТ-2 и АРТ-2m
 Fig. 4. Comparative analysis of the ART-2 and ART-2m network training time

Из результатов проведенных экспериментов видно, что при использовании предложенной структуры памяти время работы сети на выбранных интервалах изменяется линейно. Это обусловлено наличием семантических связей между разными уровнями Y -слоев при поиске активного нейрона. Время обучения на малой выборке для сетей АРТ-2m происходит медленней, так как требуется дополнительное время для построения дерева. Но как только количество различных поданных образов превышает произведение высоты дерева на среднее количество элементов в слое, АРТ-2m показывает лучшее время.

На рис. 5 изображена зависимость времени обучения АРТ-2m от количества входных образов.

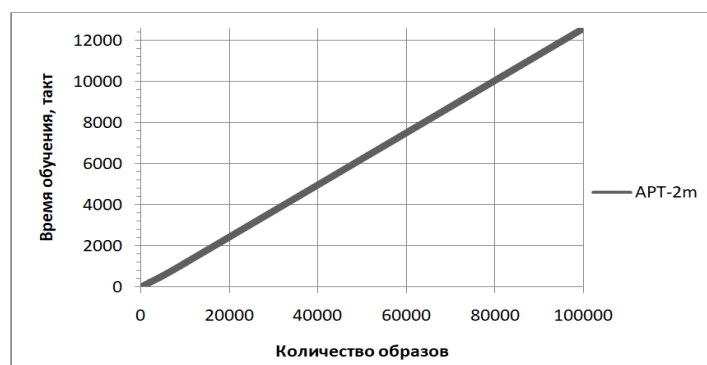


Рис. 5. Зависимость времени обучения АРТ-2m от количества входных образов

Fig. 5. Dependence of ART-2m training time on the number of input images

Из графика видно, что с ростом числа входных образов время обучения также растет линейно. Это позволяет использовать предложенную модификацию сети АРТ-2m для распознавания большего количества различных векторов данных, чем при использовании АРТ-2.

Заключение

Для устранения недостатков классической реализации АРТ-2 была предложена модифицированная модель сети адаптивно-резонансной теории с многоуровневой памятью АРТ-2m. Благодаря своей древовидной структуре памяти она позволяет уменьшить время поиска ранее запомненного образа. Таким образом, можно повысить точность распознавания путем добавления новых уровней памяти с более жестким параметром сходства. Архитектура сети АРТ-2m позволяет выполнять классификацию образов значительно быстрее, что делает возможным применять ее в системах реального времени.

В дальнейшем планируется разработка алгоритма параллельной работы сети АРТ-2m и проведение экспериментов при параллельной работе сети. Также планируется применить ее при диагностировании состояния компьютерной сети.

Работа выполнена при финансовой поддержке РФФИ грант № 16-07-00487 и при поддержке Фонда содействия развитию малых форм предприятий в научно-технической сфере (по программе У.М.Н.И.К.)

Список литературы

References

1. Будыльский Д.В., Подвесовский А.Г. 2015. Исследование применимости моделей глубокого обучения для задачи аспектного анализа тональности текстовых сообщений. Вестник Брянского государственного технического университета, 3: 117.

Budytsky D.V., Podvesovsky A. 2015. About research of applicability of models of deep training for a task of the aspect analysis of tonality of text messages. The Bulletin of the Bryansk state technical university, 3: 117. (in Russian)

2. Буханов Д.Г., Поляков В.М., Смакаев А.В. 2017. Определение состояния компьютерной сети на основе использования нейронных сетей АРТ. Вестник БГТУ им. В.Г. Шухова, 7: 157–162.

Bukhanov D.G., Poljakov V.M., Smakaev A.V. 2017. Opredelenie sostojanija komp'juternoj seti na osnove ispol'zovanija nejronnyh setej ART. Vestnik BGTU im. V.G. Shuhova, 7: 157–162. (in Russian)

3. Вовченко Р.А., Гайворонский Г.В., Беляев Н.Д. 2016. Перцептрон как основа современных инс. Новая наука: Теоретический и практический взгляд, 2: 129–131.

Vovchenko R.A., Gajvoronskij G.V., Beljaev N.D. 2016. Perceptron kak osnova soveremnyh ins. Novaja nauka: Teoreticheskij i prakticheskij vzgljad, 2: 129–131. (in Russian)

4. Григорьева Е.Ю., Мисюра М.А., Палиенко А.Н. 2017. Современные глубокие сверточные нейронные сети. Молодой ученый, 4: 153–156.

Е.Ю. Grigor'eva, М.А. Misjura, А.Н. Palienko. 2017. Sovremennye glubokie svetochnye nejronnye seti. Molodoj uchenyj, 4: 153–156. (in Russian)

5. Дмитриенко В.Д., Терёхина В.М., Заковоротный А.Ю. 2004. Вычислительное устройство для распознавания режимов функционирования динамических объектов. Вестник Национального технического университета Харьковский политехнический институт. Серия: Информатика и моделирование, 34: 70–81.

Dmitrienko V.D., Terjohina V.M., Zakovorotnyj A.Ju. 2004. Vychislitel'noe ustrojstvo dlja raspoznavanija rezhimov funkcionirovanija dinamičeskikh objektov. Vestnik Nacional'nogo tehničeskogo universiteta Har'kovskij politehničeskij institut. Serija: Informatika i modelirovanie, 34: 70–81. (in Russian)

6. Корсунов Н.И., Лысых К.В., Торопчин Д.А. 2015. Метод обучения перцептрона распознаванию текстовых символов при зашумлениях. Научные ведомости БелГУ, 13(210): 99–103.

Korsunov N.I., Lysyh K.V., Toropchin D.A. 2015. Metod obučeniya perseptrona raspoznavaniyu tekstovyh simvolov pri zashumlenijah. Nauchnye vedomosti BelGU, 13(210): 99–103. (in Russian)

7. Мельникова А.А., Михайличенко Ж.В. 2017. Развитие искусственных нейронных сетей. 27: 128.

Melnikova A.A., Mykhaylychenko Zh.V. 2017. Development of artificial neural networks. 27: 128. (in Russian)

8. Некрасова Е.И. 2015. Искусственные нейронные сети. Наука настоящего и будущего, 12: 36.

Nekrasova E.I. 2015 Artificial neural networks. Science of the present and future, 12: 36. (in Russian)

9. Платонов В.В., Семенов П.О. 2016. Обнаружение сетевых атак в компьютерных сетях с помощью методов интеллектуального анализа данных. Интеллектуальные технологии на транспорте, 4(8): 16–21.

V.V. Platonov, Semenov P.O. 2016. Obnaruzhenie setevyh atak v komp'juternyh setjah s pomoshh'ju metodov intellektual'nogo analiza dannyh. Intellektual'nye tehnologii na transporte, 4(8): 16–21. (in Russian)

10. Хорсова А.В. 2016. Интеллектуальный анализ данных. Теория и практика современной науки, 7: 356–359.

Horsova A.V. 2016. Data mining. Theory and practice of modern science, 7: 356–359. (in Russian)

11. Carpenter G.A., Grossberg S. 1987. ART 2: Self-organization of stable category recognition codes for analog input patterns. Applied optics, 26(23): 4919–4930.

12. Carpenter G.A., Grossberg S., Rosen D.B. 1991. ART 2: An adaptive resonance algorithm for rapid category learning and recognition. Neural networks, 4(4): 493–504.

13. Grossberg S. 2013. Adaptive Resonance Theory: How a brain learns to consciously attend, learn, and recognize a changing world. Neural Networks. 37: 1–47.

14. Grosspietsch K.E., Silayeva T.A. 2012. Modified ART network architectures for the control of autonomous systems. International Conference on Product Focused Software Process Improvement. Springer, Berlin, Heidelberg, 1: 300–319.

15. Heaton J.B., Polson N.G., Witte J.H. 2016. Deep learning in finance. arXiv preprint arXiv:1602.06561.

16. Kim Y. 2014. Convolutional neural networks for sentence classification. arXiv preprint arXiv:1408.5882.

17. Majeed S., Gupta A., Raj D., Rhee F.C.H. 2018. Uncertain fuzzy self-organization based clustering: interval type-2 fuzzy approach to adaptive resonance theory. Information Sciences, 424: 69–90.

18. Rowley H.A., Baluja S., Kanade T. 1998. Neural network-based face detection. IEEE Transactions on pattern analysis and machine intelligence, 20(1): 23–38.

19. Saeks R. 1997. LoFLYTE-A neurocontrols testbed. 35th Aerospace Sciences Meeting and Exhibit, 1: 85.

20. Wang T., Wu D., Coates A. 2012. End-to-end text recognition with convolutional neural networks. Pattern Recognition (ICPR), 21st International Conference on IEEE, 1: 3304–3308.

21. Wang W.C., Silva M.M.S., Moutinho L. 2016. Modelling Consumer Responses to Advertising Slogans through Artificial Neural Networks. International Journal of Business and Economics, 15(2): 89.