МАТЕМАТИКА И МАТЕМАТИЧЕСКО ОБРАЗОВАНИЕ, 1999
MATHEMATICS AND EDUCATION IN MATHEMATICS, 1999
Proceedings of Twenty Eighth Spring Conference of
the Union of Bulgarian Mathematicians
Montana, April 5–8, 1999

АДАПТИВНИ НЕВРОННИ МРЕЖИ, БАЗИРАНИ НА МОДЕЛА ART2

Анатоли Начев, Александър Геров,
Даниела Челебиева

Тази статия предлага един подход за моделиране на когнитивния процес забравяне в ART2 невронни мрежи. Показана е биологичната правдоподобност на модела чрез съпоставянето му с основните теории в когнитивната психология и математическите изводи в тях. Ефективността на модела е доказана с приложените леми и теорема. Представено е приложение, решаващо задача за филтриране информация от официалния фондов пазар.

1. Въведение. Внасянето на допълнителни характеристики в обучението на един клас изкуствени невронни мрежи – ART2 има за цел пълно или частично решаване на един проблем, свързан с функционирането на класическия ART2 модел. Става дума за блокирането на системата при изчерпване на ресурсите за представяне на категории. При някои класове задачи, свързани с работа в променящо се входно пространство или често появяващи се шумове, създаващи фалшиви категории, се наблюдава значително количество безполезно изразходвани ресурси. Може да се формулира подобен проблем при работа на мрежата в обкръжаваща среда, която притежава голямо, даже потенциално безкрайно многообразие в проявите си. Поради ограничения обем на ресурсите, мрежата не е в състояние да отрази изцяло входното пространство и трябва да работи с някакво негово подмножество. Такъв е случая с човешкия мозък, а решението на проблема при него се нарича забравяне. Моделирането на този процес в ART2 архитектурата цели както постигане на пластичност към обкръжаващата среда, така и биологическо правдоподобие. Очакването е да се реализира механизъм, с чиято помощ системата сама да се освободи от онази акумулирана информация, която може да се счита за излишна или несъществена.

2. Забравянето и когнитивната психология. Когнитивната психология разглежда процеса на забравяне като неотменна част от дейността на мозъка, тясно свързан и допълващ процеса на обучение. Несъмнено едно от най-важните открития в тази област е установяването на количествените характеристики на обучението, запаметяването и степента на забравяне. Получената от Ебингхаус крива на запаметяването отчита броя на запомнените елементи след обучение като процент от целия брой на заучените елементи. Анализирането на кривата разкрива отрицателното ускорение на забравянето. По-голямата част от изгубването на информация става през първите няколко часа след обучението, като след това степента на забравяне намалява значително.

Според Кинщ [7] кривата на Ебингхаус може да бъде добре апроксимирана чрез функцията $Y(t) = ab^{-t}$, където $Y(t)$ е мярката на съхранената информация в момента t след обучението, зависещо от параметрите a и b. Функцията може да се преобразува във вида

$$Y(t) = A_0 e^{-a_1 t} \tag{1}$$

където $a_0 = \ln a$ и $a_1 = \ln b$. Това уравнение представя съхранението на запомнената информация като експоненциално намаляваща функция на времето.

В когнитивната психология кривата на забравянето се свързва със *закона на Джост*, според който „*...ако две асоциации са с еднаква сила, но на различна възраст, по-старата изгубва силата си по-бавно във времето*".

Една от първите и популярни теории, отнасяща се за тези процеси е теорията за *спонтанния разпад на следите*, приемаща времето като единствен фактор за забравяне на информацията. Опитите на Дженкинс и Даленбах, както и на други психолози след тях стимулират създаването на друга теория – *интерференсната теория*. Според нея най-важната причина за загубата на информация е активното въздействие върху паметта. То би могло да вземе формата на влияние от старо обучение върху възприемането на ново (проактивно подтискане), или влияние на ново обучение върху припомнянето на старо (ретроактивно подтискане). Въпреки аргументите „за" и „против", двете теории по-скоро взаимно се допълват, отколкото отричат при обясняването на фактите, свързани с процесите на обучението и забравянето.

Една от известните съвременни теории в когнитивната психология е *мултикомпонентната теория на следите в паметта* на Боуър [6]. Тя предлага математически обоснован модел на забравянето и теоретични заключения, които са експериментално потвърдени. След поредица от изводи и преобразувания Боуър предлага за правдоподобна следната функция на запаметяване [6]:

$$r(t) = \frac{c}{c+f} + \left(1 - \frac{c}{c+f}\right)(1-f-c)^t$$

Константата f бележи вероятността за промяна на структурите в паметта, водеща от начално до нулево състояние на коя да е компонента, а c е вероятността за промяна, водеща от нулево до първоначално състояние на коя да е компонента. За константите е в сила ограничението $c + f < 1$. След полагане $J = \dfrac{c}{c+f}$ и $a = (1 - f - c)$ и преобразувания се получава:

$$r(t) = J + C_1 e^{-C_2 t} \tag{2}$$

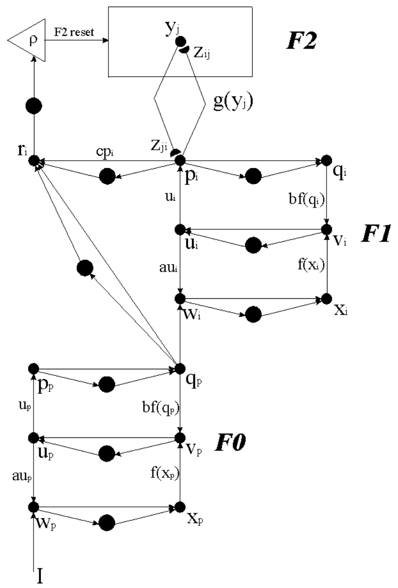
където J, C_1 и C_2 са положителни константи.

Трябва да се отбележи, че това уравнение съвпада с уравнението на забравянето във *флуктуационната теория на стимулите* на Естес.

Независимо от многообразието на теориите, всички те приемат факта, че запаметяването е експоненциално намаляваща функция на времето. Тази особеност, установена за пръв път от Ебингхаус присъства математически формулирана в студиите на Кинщ, флуктуационната теория стимулите на Естес и мултикомпонентната теория на следите в паметта на Боуър.

3. Адаптивни ART2 обучаващи правила. Адаптивните резонансни архитектури са невронни мрежи, които в реално време самоорганизират стабилни кодове за разпознаване на образци. Еволюирането на модела ART1 води до формирането на по-сложната версия ART2 , която както и ART1 функционира като самоорганизиращ се категоризатор.

При разработването на ART2 симулатор е използвана разширена архитектура, показана на фигура 1. В сравнение с класическия ART2 модел [4], тя притежава подобри характеристики при класификация. Допълнителното предпроцесорното поле $F0$ има структура, която наподобява тази на полето $F1$. Липсата на обратни сигнали позволява на $F0$ да реализира онези функции на $F1$, които са свързани с обработката на входния образец, изключвайки влиянието на заученото очакване. Тази предварителна обработка стабилизира полето $F1$ при минимален брой вътрешни цикли.



Фигура 1. ART2 архитектура с предпроцесорно поле $F0$.

Динамиката на системата е описана от уравненията на STM активностите в полетата $F0$, $F1$ и $F2$ в [1]. Активностите на r_i и p_i се представят от

(3) $$r_i = \frac{u_i + cp_i}{\|r\| + \|cp\|}$$

(4) $$p_i = \begin{cases} u_i & \textit{ако } F2 \textit{ е неактивен} \\ u_i + z_{Ji}d & \textit{ако } J\text{-}^{mu}F2 \textit{ възел е активен} \end{cases}$$

Диференциалното уравнение

(5) $$\frac{dz}{dt} = g(y_j)[p_i - z] - \lambda[d - g(y_j)]z$$

представя обобщено предлаганите нови обучаващи правила на системата. То се решава числено по метода на Рунге-Кута от 4$^{\text{ти}}$ ред [3]. Преобразуването му във вида

$$\frac{dz_{ji}}{dt} = \begin{cases} g(y_j)[p_i - z_{ji}] & \text{ако } g(y_j) = d \\ -\lambda z_{ji} d & \text{ако } g(y_j) = 0 \end{cases} \tag{6}$$

показва, че механизма на забравяне е вграден в самото обучение. Уравнението в случая $g(y_j) = 0$ се представя обобщено от

$$\frac{dz}{dt} = -\lambda z d\,, \qquad z(t_0) = z^0 \tag{7}$$

което описва забравянето като обучение с „обратен знак“, разпростиращо се върху всички LTM следи с изключение на активните в текущото реалното обучение.

Този подход се отличава с простотата на реализиране и липсата на необходимост от сложни подсистеми, осигуряващи преструктурирането на възлите и теглата. Подобни недостатъци притежават например модификациите на Фрицке върху мрежите на Кохонен [5].

Съществен аргумент в полза на предложения механизъм е неговата биологична правдоподобност. Решението на (7) се дава от уравнение (8). Забелязва се съответствието на уравнение (8) с (1) и (2). По такъв начин моделирането на забравянето като експоненциално намаляваща функция на времето съответства на представите в психологията. Начинът по който е реализиран механизма носи част от идеите на теорията за спонтанния разпад, както и на ретроактивното подтискане в интерференсната теория.

Важно е да се отбележи, че проявите на когнитивните процеси в човешкия мозък са достатъчно многообразни и комплицирани за да няма все още единна и завършена теория, която ги обяснява. От друга страна опростеното моделиране на мозъчните функции чрез парадигмите на невронните мрежи предполагат и липсата на абсолютно биологично правдоподобие на предложения модел. Разбира се остава открит въпроса необходимо ли е изобщо да се цели такова правдоподобие. Във всички случаи обаче, простотата в реализацията и постигнатата ефективност говорят в полза на предложените идеи.

Коректното функциониране на модела ART2 при ползването на новите обучаващи правила, както и ефективността на правилата при реализиране на забравянето се гарантира от динамиката на класическия ART2 модел, допълнен от две леми и една теорема, представени в Приложение А.

4. Експерименти при клъстеризация на входното пространство. Естественото отпадане на онази част от придобитите знания, които могат да се считат за несъществени води до промяна в поведението на мрежата. Тя повишава пластичността си към обкръжаващата среда, демонстрирайки адекватност и ресурсна обезпеченост при решаване на класове задачи, свързани със зашумяване или непрекъсната промяна на входното пространство [2].

Симулатора е приложен към задача за автоматична клъстеризация на субектите във фондовия пазар чрез анализ на данните, получени от официалния бюлетин на Българската Фондова Борса. Целта е филтриране на информацията за конкретни параметри на инвеститорски интерес. Въпреки, че входната информация носи сама по себе си данни за икономическите субекти, целите на задачата изискват включването на допълнителни показатели, представящи неявни отношения между тях. Най-общо показателите за оценка могат да бъдат причислени към две групи.

Първата група се състои от *текущи показатели*, включващи относителна цена на инвестиция, относителен дивидент, рисков фактор при инвестиция, относителна приходност, ценова устойчивост, абсолютен процент на промяна и относителна разлика в оферти.

Втората група се състои от *дългосрочни показатели*, описващи тенденции през сравнително по-дълъг период от време. Такива са показателите абсолютна средна промяна, усреднена цена и относителна амплитуда.

Приложена е предварителна обработка на данните за усилване на сигнала от някои показатели и подтискането на други в зависимост от техния икономически смисъл. Информацията се подава на симулатора под формата на 250 10-мерни вектора за всеки бюлетин. Постигната е настройка на мрежовите параметри, при които субектите се групират в категории, които обобщават конкретния инвеститорски интерес. Така например, при доминиране на показателя относителна цена на инвестиция входното пространство се разделя в четири категории при мрежови параметри $\rho = 0.98$, $a = b = 10$, $c = 0.1$, $d = 0.9$, $\theta = 0.2$, $z_{ij}^0 = 0.5$. Постигната е класификация в три категории при комбинации по признак рискови вложения.

Резултатите от симулацията показват отпадане на информацията за субекти, неучастващи в бюлетините продължително време. Степента на това отпадане зависи пряко от заложените стойности на параметъра λ.

Приложение А.

Лема 1. *Нека ART2 невронна мрежа притежава обучаващо правило (5). Тогава всяка ненулева и неактивна в краен интервал LTM следа експоненциално намалява стойността си.*

Доказателство: Нека разгледаме обучение в отговор на произволен входен образец, протичащо в интервала $[t_0, t_1]$. Уравнение (7) описва обучението на коя да е неактивна LTM следа, а решението е

$$z(t) = z^0 \, e^{(-\lambda d)(t-t_0)} \qquad t \in [t_0, t_1] \tag{8}$$

Нека обучението в отговор на последователност от k входни образеца продължава в интервала $[t_0, t_k]$, състоящ се от k подинтервала $[t_{i-1}, t_i]$, $i = 1 \dots k$. Интервала $[t_0, t_k]$ може да се приеме като непрекъснат, игнорирайки времето между обученията, тъй като тогава няма промяна на адаптивните филтри. От (8) следва, че крайната стойност на една неактивна LTM следа в даден подинтервал е начална стойност за следващия. Следователно (8) е валидно за непрекъснатия интервал $[t_0, t_k]$. От (8) и неравенства $0 < \lambda < 1$ и $0 < d < 1$ следва

$$\lim_{t_k, t \to \infty} z(t) = 0$$

Лема 2. *Нека ART2 невронна мрежа притежава обобщено обучаващо правило (5). Тогава всеки неактивен по време на обучение LTM възел съхранява кодираната в него информация.*

Доказателство: Нека $\tilde{z}$ и $\bar{z}$ са две произволно избрани следи на неактивен по време на обучение LTM възел и $[t_0, t_1]$ е интервал, в който протича обучението. От (8) следва, че $t \in [t_0, t_1]$

$$\frac{\tilde{z}(t)}{\bar{z}(t)} = \frac{\tilde{z}^0 e^{(-\lambda d)(t-t_0)}}{\bar{z}^0 e^{(-\lambda d)(t-t_0)}} = \frac{\tilde{z}^0}{\bar{z}^0} = const$$

Следователно отслабването на теглата не влияе върху отношението между стойностите на две произволно избрани LTM следи, а LTM вектора, представящ категорията се скъсява, но запазва направлението си като съхранява кодираната в него информация.

Следствие 1. *Нека ART2 невронна мрежа притежава обобщено обучаващо правило (5). Тогава всеки неактивен в произволен краен интервал LTM възел съхранява кодираната в него информация.*

Теорема 1 за освобождаване на ресурса. *Нека ART2 невронна мрежа притежава обобщено обучаващо правило (5), и параметри $0 < \rho < 1$, $0 < \lambda < 1$. Тогава за всеки обвързан LTM възел съществува краен интервал на неактивност, след който възелът се счита за свободен.*

Доказателство: Ориентиращата подсистема предизвиква състояние на резонанс винаги, когато съществува активен входен образец и е удовлетворено неравенството $\frac{\rho}{\|r\|} \leq 1$, където $r \equiv (r_1 \dots r_M)$. Имайки предвид, че $\|x\| = \|u\| = \|q\| = 1$, от (3) следва

$$\|r\| = \frac{\sqrt{1 + 2\,\|cp\| \cos(u,p) + \|cp\|^2}}{1 + \|cp\|} \tag{9}$$

При активност на J-ти $F2$ възел, от уравнение (4) следва, че $p = u + z_J d$. При скаларното му умножение с u и повдигането му на квадрат следва, че $\|p\| \cos(u,p) = 1 + \|z_J d\| \cos(u, z_J)$ и $\|p\| = \sqrt{1 + 2\,\|z_J d\| \cos(u, z_J) + \|z_J d\|^2}$

Разглеждаме z_J като векторна функция на времето. От (9) следва

$$\|r(t)\| = \sqrt{\frac{(1+c)^2 + 2(1+c)\,\|z_J(t)cd\| \cos(u, z_J(t)) + \|z_J(t)cd\|^2}{1 + (c^2 + 2c\,\|z_J(t)cd\| \cos(u, z_J(t)) + \|z_J(t)cd\|^2)}} \tag{10}$$

Нека произволен LTM образец е неактивен в интервала $[t_0, t_k]$, кодиран чрез $BU\ LTM$ вектор $z^J \equiv (z_{1J} \dots z_{MJ})$ и $TD\ LTM$ вектор $z_J \equiv (z_{JM+1} \dots z_{JN})$. За $t \in [t_0, t_k]$ от лема 1 следва, че $\lim_{t,t_k \to \infty} \|z_J(t)\| = 0$, а от лема 2 следва, че $\lim_{t,t_k \to \infty} \cos(u, z_J(t)) = D$, където D е константа и $-1 \leq D \leq 1$. След граничен преход в (10) получаваме

$$\lim_{t,t_k \to \infty} \|r(t)\| = \frac{\sqrt{(1+c)^2}}{1 + \sqrt{c^2}} = 1 \tag{11}$$

Следователно за произволна фиксирана стойност на ρ , $0 < \rho < 1$ съществува момент $\tilde{t}$, $t_0 \prec \tilde{t}$ такъв, че за $\forall\, t' :\ \tilde{t} \prec t'$, $\|r(t')\| > \rho$. Ако след момента $\tilde{t}$ произволно избран входен образец използва неактивен в $[t_0, \tilde{t}\,]$ LTM възел той предизвиква състояние на резонанс, което води до освобождаване на възела от кодираната в него информация, с което теоремата е доказана.

ЛИТЕРАТУРА

[1] А. Начев, Д. Димитрова. Параметри на ART2 Невронни Мрежи. *Автоматика и Информатика'98*, **2**, (1998), 67-71.

[2] A. Nachev, N. Griffith, A. Gerov. Dynamic Learning – An Approach to Forgetting in ART2 Neural Networks. Lecture Notes in Artificial Intelligence, vol. **1480**, Proc. AIMSA'98, Heidelberg, Springer-Verlag-Germany, 1998.

[3] A. Nachev, N. Griffith, A. Gerov. Dynamic Learning in ART2 Networks. *Mathematics and Education oin Math.*, **27** (1998), 195-201.
[4] G. Carpenter,S. Grossberg. ART2: Self-Organization of Stable Category Recognition Codes for Analog Input Patterns. *Applied Optics*, **26** (1987), 4916-4930.
[5] B. Fritzke. Unsupervised Clustering with Growing Cell Structures. Proc. of the IJCNN'91 Seattle (IEEE), 1991.
[6] G. Bower. Human memory: Basic Processes. New York, Academic Press, 1977.
[7] W. Kintsch. Learning, Memory, and Conceptual Processes. New York, John Wiley & Sons, 1970.

А. Начев
кат. Информатика, ФМИИ
Шуменски Университет
9700 Шумен
e-mail nachev@main.infotel.bg,
nachev@fmi.uni-shoumen.acad.bg

Александър Геров
ИМИ, ул. Акад. Г. Бончев 8
1113 София
e-mail gerov@math.bas.bg

Д. Челебиева
кат. Информатика, ФМИИ
Шуменски Университет
9700 Шумен

ADAPTIVE NEURAL NETWORK BASED ON ART2 ARCHITECTURE

Anatoly Nachev, Alexander Gerov,
Daniela Chelebieva

This paper consider an approach to madel forgetting in ART2 neural networks as an important cognitive process. Main theories of cognitive psychology and some mathematical conclusions imply biological plaussibility of the method. Reliability of model dynamics is proved by lemmas and a theorem. The paper describes simulation results from an application concerned with stock exchange information filtering.