

СПЕЦИАЛИЗИРАН НАУЧЕН СЪВЕТ  
ПО ИНФОРМАТИКА И ПРИЛОЖНА МАТЕМАТИКА  
ПРИ ВАК

Анатоли Маринов Начев

АДАПТИВНИ ART2 НЕВРОННИ МРЕЖИ

АВТОРЕФЕРАТ  
на  
ДИСЕРТАЦИЯ

за получаване на образователната и научна степен

доктор

научен ръководител:

ст. н. с. II ст. д-р. Александър Геров

рецензенти:

1. проф. д-р. Петър Бърнев
2. ст. н. с. II ст. д-р. Александър Григоров

София, 1999

Дисертационният труд е обсъден и допуснат до защита на .....г.  
на разширено заседание на секция "Изкуствен интелект" към Институт  
по Математика и Информатика, БАН.

Докторантът работи Шуменския Университет "Еп. Константин Прес-  
лавски", Факултет по Математика и Информатика, катедра Информа-  
тика.

Защитата на дисертационният труд ще се състои  
на ..... от ..... часа  
в зала ..... на ..... при .....  
на разширено заседание на СНС по информатика и приложна матема-  
тика при ВАК.

Автор: Анятоли Маринов Начев

Заглавие: Адаптивни ART2 невронни мрежи

## Мотиви

Мотивите за написването на този труд са свързани с работата на автора върху проблемите на машинното обучение, както и интереса към изследване и моделиране на някои психо-физиологични процеси на централната нервна система.

Изследванията в областта на машинното обучение са свързани най-вече с проблема как една система може да придобие знания, които тя не притежава. Не е изненадващо, че относително малко внимание е отделено на обратния проблем: как една система може да се освободи от знания, които вече притежава. Среждайки се при хората, този феномен се нарича забравяне и се разглежда като нежелано пропадане на паметта. Негативният образ на забравянето не е основателен, защото това е полезен процес, улесняващ възприемането на информацията. Според Пол Скот [55] обучението е процес, в който се конструира организирано представяне на опита, а забравянето е процес, в който част от това организирано представяне се реорганизира или разрушава. По такъв начин двата процеса взаимно се допълват, а окончателно представеният опит е резултат от съвместното им действие.

Механизмите на забравянето заслужават изучаване не по-малко от тези на запознаването с информация, тъй като те заедно установяват резултата от обучението. В своята същност забравянето е свързано с промяна на структурата на знанията, което частично или напълно ограничава достъпа до някои елементи в паметта. Това включва и процеси, които отслабват паметта или изолират фрагменти от съхранената информация. Тези промени могат да бъдат наблюдавани. Маркович и Скот [47] описват обстоятелствата, при които е по-добре да се премахне един елемент в паметта, отколкото да се припомни. Според тях проблемът най-добре се разбира, когато обучението се разгледа от икономическа перспектива. Би могло да се приеме, че всяка форма на обучение е всъщност форма на инвестиция. Една система по време на обучение инвестира известно количество ресурси в запознаването и поддържането на някои знания. В полза на системата е единствено обучение, при което възвръщаемостта от тази инвестиция е положителна. Така цената на запознаването и поддържането на знания не бива да надхвърля цената на ресурса, отделен

вграждащи в себе си механизмите на забравянето. Така се избягва ползването на допълнителни надзираващи подсистеми и нови характеристики на изкуствените неврони, като в същото време се постига простота и биологична правдоподобност.

Предлаганото от автора адаптивно обучение допринася за решаването на един проблем в *ART* моделите, а именно блокирането на системата при изразходване на ресурсите си. Практическата приложимост се вижда най-вече при решаване на някои класове задачи, свързани с работа в силно зашумена среда или променливо се входно пространство.

Този труд е оформен в четири основни части.

Глава първа прави кратък обзор на основните понятия, термини и архитектури на невронни мрежи, които се използват в следващите глави. В началото са разглеждани фундаменталните понятия и термини на невронните мрежи, както и сравнение между изчислителните възможности на невронните мрежи и конвенционалните компютри с Фон Нойманова архитектура. Следва кратко въведение в адаптивната резонансна теория и произтичащите от нея модели на невронни мрежи. Специално внимание се отделя на *ART1* [12] и *ART2* [13] архитектурите, тъй като те въвеждат основните принципи, върху които се изграждат последващите *ART* архитектури. Прави се кратък обзор на един механизъм за адаптивност към променливо се входно пространство, който се прилага към модифицираните от Фрицке самоорганизиращи се карти на Кохонен [23], [24], както и една модификация на *ART1* с променлива сила на връзките между невроните [4]. В Глава 3 тези модели се съпоставят с предлагания от автора механизъм на адаптивно обучение.

Глава 2 прави кратък обзор на онези елементи, понятия и теории в неврологията и когнитивната психология, които са мотивирали автора да създаде биологично правдоподобни обучаващи правила на невронни мрежи от класа *ART2*. Описаните понятия и теории се разискват и в Глава 3 в контекста на невронните мрежи за да покажат биологичната правдоподобност на предложения адаптивен *ART2* модел. В главата се очертават се трите основни класически теории на забравянето - теорията за спонтанния разпад на следите, интерференсната теория и теорията за репресията, както и една от по-новите теории в когнитивната психология - мултикомпонентната теория на следите в паметта. Разглеждат се някои въпроси, свързани със структурата и метаболизма на биологичните синапси, имащи пряко отношение към предлагания от автора адаптивен *ART2* модел.

Глава 3 въвежда нов модел на *ART2* невронна мрежа, наречен адаптивен, който внася допълнителни характеристики в обучението на класическите *ART2* невронни мрежи. Във връзка с коментираните в Глава 2 особености на обучаващия процес се предлагат нови диференциални

уравнения, описващи обучаващите правила на *ART2* мрежа, а коректността в работа на системата и поведенческите характеристики на модела се доказват чрез две лема и една теорема. В главата се описва влиянието на адаптивните обучаващи правила върху динамиката на невронната мрежа и произтичащите от това особености. Въведното адаптивно обучение се коментира в контекста на машинното самообучение. Представя се невронен симулатор, базиран на адаптивния *ART2* модел. Разглежда се биологичната правдоподобност на модела в контекста на понятията и теориите, описани в Глава 2.

Глава четвърта прави експериментална оценка на адаптивния *ART2* модел. Прави се анализ на мрежовите параметри и значението им в поведението на мрежата, както същевременно се разглежда начина за тяхното настройване. Описани са множество експериментални резултати в тази връзка. Специално внимание е отделено на новия мрежов параметър  $\lambda$ . Експериментално се демонстрира, че класическият *ART2* модел се явява частен случай на въведения адаптивен модел. Демонстрирани са и по-добрите качества на адаптивния модел при решаване на задачи, свързани с работа в силно зашумена среда или менищо с входно пространство. В края на главата се демонстрира практическата приложимост на адаптивния модел при решаване на задача за класификация на субекти от фондовия пазар с цел автоматично филтриране на информация и подпомагане при вземане на решения.

Заключението на дисертацията обобщава накратко получените резултати и публикациите на автора във връзка с темата на дисертацията.

## Глава 1

### Невронни мрежи и адаптивна резонансна теория

В тази глава се прави кратък обзор на основните понятия, термини и архитектури на невронни мрежи, което е необходимо за изложението на следващите глави.

#### 1.1 Въведение в теорията на невронните мрежи

В началото са разгледани фундаменталните понятия и термини на невронно-мрежовото пресмятане, както и сравнение между изчислителните възможности на невронните мрежи и конвенционалните компютри с Фой Нойманова архитектура.

#### 1.2 Адаптивна резонансна теория

Адаптивната резонансна теория (*adaptive resonance theory, ART*) [29], [14], [12] е разработена от Стивън Гросберг в Бостънския Университет като теория за обучението на мозъка. Неговите изследвания върху експериментални данни от психологически тестове формират идеята, че индивидуалното поведение е в съответствие с въздействието на околната среда [28]. Гросберг разглежда теорията си от гледна точка на мрежа. Той приема, че поведенческите елементи трябва да се представят като възли, а моделът трябва да е способен постепенно да присъединява тези възли в обучението при кодиране на придобития опит. По-точно *ART* казва, че нашите научени очаквания фокусират вниманието върху сензорната информация. Начинът, по който те са съхранени в дълговременната памет (*long term memory, LTM*) позволява новото придобито знание, временно съхранено в кратковременната памет (*short term memory, STM*) да бъде автоматично включено в *LTM*. *ART* дава решение на проблема за предпазване на съществуващите знания при обучението на нови

(стабилност), като същевременно се запазва способността за обучение на нова информация (пластичност). Този проблем, известен още като "*stability – plasticity dilemma*" се решава от малко модели на изкуствени невронни мрежи.

Друга отличителна черта на теорията е, че се прави разлика между сензорната активност като входен източник на информация и *STM* активност като реакция на получената от входа информация. Мотивите за такова представяне идват от проведените психологически експерименти, които показват, че *STM* следите в паметта се разпадат по-бавно от въвеждането на нови входни сензорни сигнали. В тази теория *LTM* следите се разглеждат като колекция от променящи се величини, които на основата на научения опит могат да различат кои *STM* сигнали трябва да достигнат техните възли.

Разпознаването на образец в една *ART* система може да се опише като частен случай на по-общ когнитивен процес на подбор на хипотези, тестване, търсене, класификация и обучение.

Принципите, произтичащи от адаптивната резонансна теория са използвани в голям брой различни модели на невронни мрежи и техни модификации. Основните типове сред тях са *ART1* и *ART2*, които предлагат паралелна архитектура за стабилно самоорганизиращи се разпознаватели на двоични и аналогови входни образци. Моделът *ART1* се използва в случая на двоични входни образци, а по-развитата му версия – *ART2* се прилага както за двоични, така и за аналогови входни образци. Тези два модела развиват основните принципи, върху които се градят и други типове невронни мрежи, базирани на адаптивната резонансна теория.

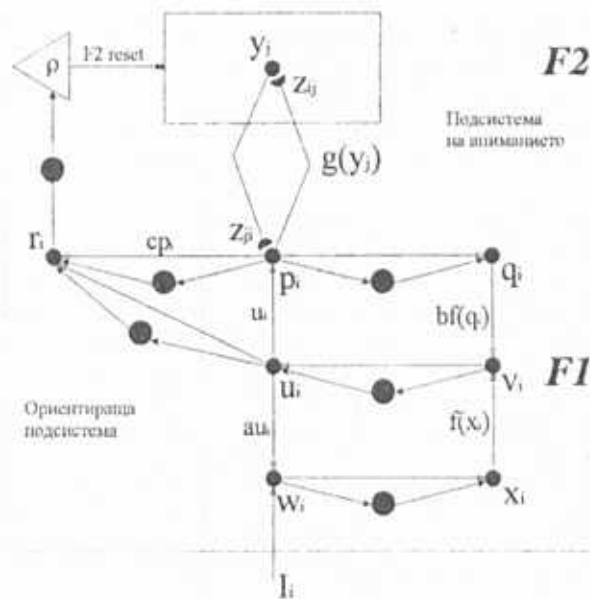
### 1.3 Моделът *ART1*

В главата се дискутира накратко основният модел *ART1* [12], [3], върху чиито принципи лежат много от *ART* моделите.

Разгледани са и някои модификации на автора върху *ART1* невронни мрежи, ползващи следи с променлива сила [4], [52].

### 1.4 Моделът *ART2*

В модела *ART2* са заложили идеи, които се явяват продължение на тези в *ART1*. Една от съществените разлики между двата е, че *ART2* работи както с двоични, така и с аналогови входни образци.



Фигура 1.1: Класическа ART2 архитектура. Белите стрелки представят тилковите (възбуждащи) сигнали към съответните слоеве, а черните - метничките (подтискащи) сигнали, реализиращи нормирането на сигналния вектор с евклидова норма.

### 1.4.1 Структура

Мрежата се състои от *подсистема на вниманието* и *ориентираща подсистема* (фигура 1.1) [13]. Подсистемата на вниманието е изградена от две полета **F1** и **F2** (фигура 1.1), напълно свързани едно с друго в права и обратна посока чрез теглови връзки, обособяващи двата адаптивни филтъра: *прав* (*bottom – up*, *BU*) и *обратен* (*top – down*, *TD*). На фигурата те са означени с черни полуокръжности. Всеки от филтрите кодира разпознаващите категории чрез промяна на теглата (назречени още *следци*) на дълговременната памет (*LTM*). **F1** се нарича още *поле за представяне на характеристики*, а **F2** – *поле за представяне на категориите*.

В модела ART2 полето за представяне на характеристиките **F1** включва няколко обработващи нива. За разлика от модела ART1, постъпващите входни сигнали и тези, получени от полето **F2** се получават на различни места в **F1**. Причината за това е необходимостта от допълнително



обработване на сигналите с цел усилване на характерните им особености и подтискане на шума. Въпреки, че при *ART2* полето *F1* е по-сложно отколкото в *ART1*, обучаващите правила на *ART2* са по-прости.

Полето *F1* се състои от три подполета, наречени *горно* (най-близкото до *F2*), *средно* и *долно* (фиг.1.1). Буквите *p*, *q*, *u*, *v*, *w* и *x* представят векторите на активност, получени в съответното подполе. Например значението  $p_i$  служи за показване на активността на *i*-тия възел на съответния слой.

Във фигура 1.1 малките черни кръгове показват наличие на възбуждаща *STM* активност на това място. Невронните слоеве, пораждащи тази активност не са означени, но те неявно присъстват на съответните места.

Големите черни кръгове представят активностите, които предават подтискащи сигнали успоредно на възбуждащите. Предназначението им е да нормират с евклидова норма възбуждащия вектор чрез подтискащи сигнали, които са право пропорционални на дължината му.

Полето *F2* се състои от един невронен слой, чиито възли са подложени на взаимни състезателни въздействия по схемата *on-center*, *off-surround*. Всеки възел е свързан със себе си чрез възбуждащи рекурентни връзки (*on-center*) и е свързан с всички други възли в полето чрез подтискащи връзки (*off-surround*). Въпреки, че *ART2* и по-специално *F2* най-често е описвана в алгоритмична форма, важно е да се осъзнае, че много аспекти от нейното поведение са моделирани върху особеностите на състезателно-кооперативната природа на това поле. То е конфигурирано така, че да увеличава контраста на постъпващия образец от *F1*. Увеличаването на контраста води до реализирането на избор чрез възела с максимална активност, докато всички останали възли в полето имат активност нула.

Друга особеност на полето *F2* е механизма, чрез който възлите могат да бъдат въздействани от траен *reset*, т.е. да придобият трайна неактивност и излизане от състезанието до момента, в който системата изпадне в състояние на резонанс.

Ориентиращата подсистема (фиг. 1.1) е предназначена да получава и сравнява два образца, носещи информация съответно за постъпващия в мрежата вход, и наученото очакване, кодирано в *LTM*. Мярката за това съответствие се определя от дължината на вектора *r*, която пък се сравнява с параметъра  $\rho$ , наречен *параметър на бдителност*. Резултатът от работата на подсистемата контролира изпращането на *reset* сигнал към *F2*.

## 1.4.2 STM уравнения на F1

Потенциалът на STM активността  $V_i$  на  $i$ -тия възел в който и да е F1 обработващ етап се подчинява на уравнението за потенциала на клетъчната мембрана [13], [36]:

$$\delta \frac{d}{dt} V_i = -AV_i + (1 - BV_i)J_i^+ - (C + DV_i)J_i^-$$

за ( $i = 1 \dots M$ ).

Означението  $J_i^+$  бележи сумарния възбуждащ вход към  $i$ -тия възел, а  $J_i^-$  е сумарния подтискащ вход. Параметъра  $\delta$  представя отношението между времето за STM релаксация и времето за LTM релаксация. Счита се, че

$$0 < \delta \ll 1.$$

В модела ART2 се приема, че са в сила ограниченията [13]:

$$B \equiv 0, C \equiv 0, \delta \rightarrow 0.$$

Следователно функцията на активност за кой да е възел в полето F1 има най-общо вида

$$V_i = \frac{J_i^+}{A + DJ_i^-}. \quad (1.1)$$

Като частен случай на (1.1) при ограниченията [13]:

$$A \equiv \varepsilon, D \equiv 1, 0 < \varepsilon \ll 1$$

и във връзка с фигура 1.1 могат да се дефинират уравненията (1.2) – (1.7), които характеризират STM активностите  $p_i$ ,  $q_i$ ,  $u_i$ ,  $v_i$ ,  $w_i$  и  $x_i$ , получени в полето F1.

$$p_i = u_i + \sum_j g(y_j)z_{ji} \quad (1.2)$$

$$q_i = \frac{p_i}{\varepsilon + \|p\|} \quad (1.3)$$

$$u_i = \frac{v_i}{\varepsilon + \|v\|} \quad (1.4)$$

$$v_i = f(x_i) + bf(q_i) \quad (1.5)$$

$$w_i = I_i + au_i \quad (1.6)$$

$$x_i = \frac{w_i}{\epsilon + \|w\|} \quad (1.7)$$

В тези уравнения  $\| \cdot \|$  означава евклидова норма, а  $y_j$  е *STM* активността на  $j$ -тия *F2* възел.

Сигналната функция  $f$  в уравнение (1.5) може да бъде определена като [13]:

$$f(x) = \begin{cases} \frac{2\theta x^2}{x^2 + \theta^2} & , \text{ ако } 0 \leq x \leq \theta \\ x & , \text{ ако } x > \theta \end{cases} \quad (1.8)$$

или

$$f(x) = \begin{cases} 0 & , \text{ ако } 0 \leq x \leq \theta \\ x & , \text{ ако } x > \theta \end{cases} \quad (1.9)$$

Графиката на функцията  $f(x)$  от (1.8) може да бъде изместена надясно, правейки  $f(x) = 0$  за малки стойности на  $x$ , както е в (1.9).

### 1.4.3 *STM* уравнения на *F2*

Полето за представяне на категории *F2* е организирано също така, както и в модела *ART1* [12]. Ключова особеност на полето *F2* е увеличението на контраста на постъпващата *STM* активност. Другото важно качество на възлите в това поле е способността им да се подтискат трайно винаги когато ориентиращата подсистема изпрати *reset* сигнал към тях.

Увеличението на контраста се реализира чрез състезание в *F2*. Победителят от него се определя от екстремния случай на увеличаване на контраста. *F2* прави избор, когато съответния възел е получил най-голям общ вход, с което той погасява активността на всички останали възли. Нека към  $j$ -тия *F2* възел постъпва филтрираната през правия адаптивен филтър *F1* *STM* активност. Ако с  $T_j$  се бележи сумарния вход, то

$$T_j = \sum_i p_i z_{ij}, \quad (1.10)$$

където ( $j = M + 1, \dots, N$ ). Тогава се казва, че *F2* прави избор, ако  $J$ -тия *F2* възел стане максимално активен, докато всички останали възли са подтиснати

$$T_J = \max\{T_j : j = M + 1, \dots, N\} \quad (1.11)$$

Реализирането на *reset* в *F2* може да стане по няколко начина, един от които е използване на шлюзова двуполюсна мрежа в *F2* [29]. Когато неособен възбуждащ вход се появи в шлюзовото двуполюсно поле

на  $F2$ , възлите се подтискат пропорционално на тяхното  $STM$  ниво на активност, което всъщност реализира *reset*. Нещо повече, това подтискане трае през цялото време на подаване на  $STM$  образец на  $F1$ . Такава подтискаща вълна достига до  $F2$  чрез ориентиращата подсистема при наличие на достатъчно голямо несъответствие в  $F1$ .

Когато  $F2$  прави избор, динамиката на полето може да се опише по следния начин:

$$g(y_j) = \begin{cases} d, & \text{ако } j = J, T_j = \max\{T_j : \text{неподтисканати от reset}\} \\ 0, & \text{в останалите случаи} \end{cases} \quad (1.12)$$

От (1.12) следва, че (1.2) има вида

$$p_i = \begin{cases} u_i, & \text{ако } F2 \text{ е неактивен} \\ u_i + z_j d, & \text{ако } F2 \text{ е активен} \end{cases} \quad (1.13)$$

#### 1.4.4 $LTM$ уравнения

Обучението в състояние на резонанс променя създадения преди това код базирайки се на онази нова информация, която входния образец съдържа, или пък инициализира код, който служи за обучение в първоначално необвързаното множество от възли. Ако е намерено адекватно съответствие, което е възможно още в първия цикъл, представянето на категориите би могло да се донастрои, включвайки новия образец. Ако не е намерено съответствие и пълният капацитет за кодиране не е изчерпан, ще бъде създадена нова категория с помощта на предварително необвързани  $LTM$  следи, които кодират създадения от входните сигнали  $STM$  образец.

Обучаващите правила на  $ART2$  модела се използват за промяна на следите както в правия, така и в обратния адаптивен филтър. Те се активират тогава, когато системата изпадне в състояние на резонанс за да предизвикат приближаване на образаца, кодиран в  $LTM$  следите на активния  $F2$  възел, към  $STM$  образаца на горното ниво в  $F1$ .

Обучаващите правила се представят чрез следните диференциални уравнения [13]:

$$F2 \rightarrow F1: \quad \frac{d}{dt} z_{ji} = [p_i - z_{ji}]g(y_j), \quad (1.14)$$

$$F1 \rightarrow F2: \quad \frac{d}{dt} z_{ij} = [p_i - z_{ij}]g(y_j). \quad (1.15)$$

Ако  $F2$  е направил избор чрез един активен възел  $J$ , от равенства (1.12), (1.13), (1.14) и (1.15) следва

$$\frac{d}{dt}z_{Ji} = [p_i - z_{Ji}]d = \left[\frac{u_i}{1-d} - z_{Ji}\right]d(1-d), \quad (1.16)$$

$$\frac{d}{dt}z_{iJ} = [p_i - z_{Ji}]d = \left[\frac{u_i}{1-d} - z_{iJ}\right]d(1-d), \quad (1.17)$$

където  $0 < d < 1$ .

Освен това за всяко  $j \neq J$

$$\frac{d}{dt}z_{ij} = 0, \quad (1.18)$$

$$\frac{d}{dt}z_{ji} = 0. \quad (1.19)$$

По-късно ще бъдат разгледани началните стойности на  $LTM$  следите.

#### 1.4.5 Уравнения на ориентиращата подсистема

При съпоставянето на двоични образци сходството между тях може да бъде пресметнато чрез отчитане на съвпадащите битове. Поради тази причина  $ART1$  архитектурите използват  $STM$  образаца на полето  $F1$  и входния образец за да намерят брой на единиците им без да е необходимо присъствието на самите образци в ориентиращата подсистема. За разлика от това обаче, пресмятането на сходството при аналогови образци изисква информация за самите образци. Степента на сходство за  $ART2$  архитектурата от фигура 1.1 между един  $STM$  образец на  $F1$  и активен  $LTM$  образец се определя от вектора  $r = (r_1, \dots, r_M)$ , където

$$r_i = \frac{u_i + cp_i}{\varepsilon + \|u\| + \|cp\|}. \quad (1.20)$$

Ориентиращата подсистема предизвиква *reset* на  $F2$  винаги, когато един входен образец е активен и

$$\frac{\rho}{\varepsilon + \|r\|} > 1, \quad (1.21)$$

където параметърът на бдителност  $\rho$  е ограничен от неравенството

$$0 < \rho < 1.$$

За простота в разглежданията се приема че  $\varepsilon = 0$ . По такъв начин от уравнения (1.7), (1.4) (1.3) следва, че

$$\|x\| = \|u\| = \|q\| = 1.$$

Векторът  $r$  е способен да се справи с всички особености и изисквания за удовлетворяване на съоставянето в ориентиращата подсистема. Тъй като векторът  $p$  се явява сума на  $u$  и  $z_j$ , съответствието между  $p$  и  $u$  може да се изрази чрез  $u$  и  $z_j$ . Мярката за съответствие между тях би могла да се изрази чрез  $\cos(u, z_j)$ . Поради това, че  $J$ -тия  $F2$  възел остава активен само ако  $\rho \leq \|r\|$  и тъй като  $\rho < 1$ , това означава, че активността може да се поддържа или ако  $\cos(u, z_j) = 1$ , или ако  $\|z_j\| = 0$ . Не би могло да има *reset*, ако  $STM$  вектора  $u$  е достатъчно успореден на  $LTM$  вектора  $z_j$  или ако  $LTM$  следите на обратния адаптивен филтър са малки.

## 1.5 Предимства на *ART* моделите

Най-важните предимства на *ART* пред останалите модели на невронни мрежи са, че могат да работят добре в автономна среда, да се обучават бързо и същевременно стабилно в постоянно променяща се обкръжаваща среда докато се изразходи целия капацитет на мрежата. Ето и по-важните характеристики на *ART* моделите според Карпънтър: [14]

- Обучение в реално време
- Самоорганизиране и самообучение
- Работа в изменчива обкръжаваща среда
- Самостабилизиране на паметта при произволен брой входове
- Ефективно използване на целия капацитет на паметта
- Поддръжка на пластичност в несочакван свят
- Активно, фокусиращо регулиране на обучението
- Бавно или бързо обучение
- Обучение във фаза на приблизително съвпадение
- Саморегулиращо се тестване на хипотези
- Бързо адаптивно търсене на най-добро сравнение
- Бърз директен достъп до кодовете на познати събития
- Променлив критерий за грешка (параметър на бдителност)

## 1.6 Самоорганизиращи се карти на Кохонен с променлива клетъчна структура

В главата се прави кратък обзор на един механизъм за адаптивност към променящо се входно пространство, който се прилага към модифицираните от Фридзе самоорганизиращи се карти на Кохонен. Подходът на тази разработка е залегнал и в реализацията на проекта *SPIN* (*Spatial Preception to Identification with Neural networks*) [39] за динамична класификация на 3D сканирани изображения. Специално внимание е отделено на забравянето като средство за постигане на адаптивност, а използваният подход ще бъде сравнен с резултатите от този труд.

## Глава 2

### Обучението и забравянето в когнитивната психология

Тази глава е посветена на онези елементи, понятия и теории в неврологията и когнитивната психология, които са мотивирали автора да създаде биологично правдоподобни обучаващи правила на невронни мрежи от класа ART2. Описаните тук понятия и теории ще бъдат разисквани и в следващата глава в контекста на невронните мрежи за да покажат биологичната правдоподобност на предложения адаптивен ART2 модел.

#### 2.1 Въведение

Загубата на възможност за възстановяване и използване на първоначално кодираната информация е онова, което обикновено се нарича *забравяне* [41]. Алън Бадли [6] очертава три основни класически теории на забравянето - теорията за *спонтанния риад на следите*, *интерференсната теория* и теорията за *репресията*, чиито основни идеи се разглеждат накратко в главата.

#### 2.2 Биологичните синапси

Както синаптичните връзки играят ключова роля във функционирането на биологичните нервни системи, така и организацията на връзките между изкуствените неврони е определяща за поведението на невронните мрежи. Разглеждат се някои важни факти за структурата и функционирането на биологичния модел – синаптичната връзка.



## 2.3 Класически теории на забравянето

### 2.3.1 Кривата на Ебингхаус

Несъмнено едно от най-важните открития в историята на когнитивната психология 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 [60]. Анализиранието 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 Кивц [40] кривата на Ебингхаус може да бъде добре апроксимирана чрез функцията:

$$Y(t) = ab^{-t}, \quad (2.1)$$

където  $Y(t)$  е мярката на съхранената информация в момента  $t$  след обучението, зависещо от параметрите  $a$  и  $b$ . Лесно се вижда, че от (2.1) следва

$$\ln Y(t) = \ln a - (\ln b)t = a_0 - a_1 t, \quad (2.2)$$

където  $a_0 = \ln a$  и  $a_1 = \ln b$ . Следователно

$$e^{\ln Y(t)} = e^{a_0 - a_1 t},$$

$$Y(t) = A_0 e^{-a_1 t}. \quad (2.3)$$

Уравнение (2.3) представя съхранението на запомнената информация като експоненциално намаляваща функция на времето.

### 2.3.2 Теория на спонтанния разпад

Теорията на спонтанния разпад е въведена формално от Гесталт [60] като обобщение на онези представи за когнитивните процеси, които са възприети в психологията на дадения етап. Поради невъзможността да бъдат доказани експериментално, изводите в тази теория са приемани за коректни по-скоро интуитивно.

Според теорията на спонтанния разпад всяко обучение се кодира чрез формиране на относително непостоянни следи в паметта. Колкото по-силно е първоначалното обучение, толкова "по-дълбока" ще бъде тази следа и толкова по-дълго ще остане в паметта. От друга страна периодичното припомняне или използване на информацията, кодирана в следите ще ги

преактивира и засили [64]. Ако с течение на времето следите останат неизползвани, те стават неясни и представеният в тях спомен започва да се изгубва. Някои автори дори отиват по-далеч и разглеждат паметта като вид орган на тялото. Точно както някои наши мускули и органи могат да отслабват функциите си поради недостатъчна активност, така и паметта би могла да отслабва поради тази причина. От тук следва, че за да бъде една информация запомнена, тя трябва да бъде срещана отново, практикувана или припомняна [22].

### 2.3.3 Интерференсна теория

Тази теория се явява алтернативно схващане на теорията за спонтанния разпад. Основното в нея е схващането, че забравянето се причинява от активното въздействие на постъпващата информация върху паметта. С други думи колкото повече време минава след възприемането и запомнянето на едно знание, толкова повече се увеличава възможността да се случи въздействие върху паметта.

Независимо от това, че интерференсната теория се възприема като единна теория, съществуват два обособени класа въздействия и съответно подтеории.

Първата и същевременно по-забележителна форма на въздействие върху паметта се изразява в подтискащия ефект на новото обучение върху припомнянето на стари спомени. Тази форма на "обратно" въздействие върху заучената информация носи името *ретроактивна интерференция*. Както в ежедневието, така и в лабораторни условия се наблюдава друг тип въздействие - подтискане на запаметяването на новата информация от заучената стара. Това въздействие е известно под името *проактивна интерференция*. Независимо от това, че са фактически различни, ретроактивната и проактивната интерференция са двете страни на един и същ процес. Дори в лабораторни условия при строг контрол на обучението е трудно да се изолират проявите на двата типа въздействие и така да се получат независими резултати [64].

Важна част от интерференсната теория е обяснението на механизмите, водещи до загуба на информация. Изказаните в тази връзка различни обяснения и гледни точки попадат в четири основни групи.

1. **Състезателност на спомените.** Според тази гледна точка механизмът на забравяне се свързва с подтискане и блокиране на старата информация, което е резултат на състезание между спомените. Ефектът от това е, че по-силните спомени изпреварват търсените, независимо че могат да са нежелани.

2. **Отучване.** Така наречената *теория на отучването* разглежда възможното първоначално объркване при възстановяването на спомени като част от общия процес на забравяне. Разграничаването на старата и нова информация се предшества от отслабване на по-ранните следи, което може да доведе до окончателното забравяне на изпреварените в състезанието следи. С други думи обучението с нова информация предизвиква "отучване" на старата [50].
3. **Достъпност на спомените.** Ефектът от новото обучение се възприема като претрупване във вече и без това пренаселената памет. Ако е налице голямо пренаселване с множество подобни един на друг спомени, тогава процесът на търсене в паметта е по-продължителен и някои от нейните елементи стават по-малко достъпни.
4. **Качествени промени.** Тази теория за въздействие върху старата информация се основава на взаимодействиящата същност на паметта. Новите спомени припокриват части от старите, давайки една смесена картина, в която има елементи както от новите, така и от старите.

Наличието на четири различни модела поставя въпроса кой от тях е верен. Уингфилд [64] смята, че забравянето е достатъчно сложен процес и всяка от гледните точки, давайки частичното му обяснение, има значителна заслуга за изграждане на общата картина.

### 2.3.4 Теория за репресията (мотивираното забравяне)

Теорията на забравянето в следствие на репресия е развита от Фройд [25]. Тя разглежда един от известните източници на забравяне - мотивираното забравяне, но нейното приложение е ограничено само върху загубата на определен тип информация - тази с отрицателно емоционално съдържание.

## 2.4 Мултикомпонентна теория на следите

През последните години придобиват популярност и някои по-нови теории в когнитивната психология, изградени като всеобхватни модели на човешкото обучение и памет. Такава е *мултикомпонентната теория на следите в паметта*, създадена от Гордън Бауър [11]. В дисертацията са разгледани само онези части от теорията, които разглеждат механизмите на забравянето.

Предполага се, че възприемането на дадена информация се съпровожда със съхраняването и в следи, но не под формата на буквални входни

стимули, а на никакви техни кодирани представяния. Счита се, че съхранената информация е или под формата на *първичен код*, чрез който възприеманото събитие е разпознато, или като *вторичен код*, съдържащ асоциирана с първичния код информация. Приема се, че даден елемент на паметта е вектор с  $N$  компоненти. Ако се приеме, че едно обучение е съхранено като вектор от информационни компоненти, тогава забравянето на никаква информация (поради каквато и да е причина) би могло да се разгледа като изтриване или смяна на стойността на някоя от компонентите в оригиналния вектор. Докато трае процесът на забравяне следите продължават да се разрушават, а векторът носи все по-малко информация от първоначалното обучение.

След редица изводи Бауър получава функцията на припомняне

$$C(t) = [r(t) + (1 - r(t))g]^N.$$

Изразът  $r + (1 - r)g$  представя вероятността за коректно припомняне на коя да е компонента в следата. Тогава вероятността за коректно припомняне на цялата следа се свежда до припомнянето на всичките  $N$  компонента, т.е.  $r + (1 - r)g$  се повдига на  $N$ -та степен. Функцията на запамятаване  $r(t)$  се изразява от уравнението

$$r(t) = J + (1 - J)a^t,$$

където  $J$  и  $a$  са положителни константи. Отново след преобразувания (логаритмуване и антилогаритмуване на второто събираемо) може да се покаже, че запамятаването е експоненциално намаляваща функция на времето:

$$r(t) = J + C_1 e^{-C_2 t}. \quad (2.4)$$

Тук  $J$ ,  $C_1$  и  $C_2$  са положителни константи.

Бауър смята, че тестовите, проведени в Станфордския университет са дали добро приближение към теоретичните предположения на горните уравнения [11].

Независимо от многообразието на теориите, всички те приемат факта, че задържането на запомнената информация е експоненциално намаляваща функция на времето. Тази особеност, установена за пръв път от Ебингхаус, присъства математически формулирана в студиите на Хинц [40], флукуационната теория на стимулите на Естес [21] и мултикомпонентната теория на следите в паметта на Бауър [11].

Този факт ще има съществено значение в Глава 3 при моделиране на забравянето и вграждането му като част от поведението на невронните мрежи.

## Глава 3

### Адаптивен *ART2* модел

#### 3.1 Въведение

В тази глава авторът въвежда нов модел изкуствени невронни мрежи, базирани на адаптивната резонансна теория (adaptive resonance theory, *ART*) и по-специално на един клас от тях - *ART2*. Стремежът е да се реализира биологично правдоподобно моделиране на онези когнитивни процеси, които са свързани с частично отслабване на паметта, както и с възможността системата да се освободи сама от информация, която може да се счита за излишна и несъществена. Предлагането на нови обучаващи правила, наречени адаптивни, е мотивирано от разбиранията и представите за тези процеси в биологичните нервни системи, описани накратко в Глава 2. Моделът е наречен *адаптивен ART2 модел*.

Качествата, които отличават адаптивното обучение допринасят за преодоляване блокирането на системата в класическия *ART2* модел при изчерпване на мрежовите ресурси за представяне на категории. Важността на този проблем се проявява особено при решаване на някои класови задачи, свързани с работа в променящо се входно пространство, често появяващи се шумове или при значително многообразие на обкръжаващата среда.

В главата се описва влиянието на адаптивните обучаващи правила върху динамиката на *ART2* невронни мрежи и произтичащите от това особености на модела. Въведното адаптивно обучение се коментира в контекста на машинното самообучение.

Представя се невронен симулатор, базиран на адаптивния *ART2* модел, както и някои негови особености.

## 3.2 Обучение и адаптивност

Една важна отличителна черта от дейността на биологичните нервни системи е наличието на процес, който предизвиква отслабване на синаптичните връзки при определени условия. Описаните в Глава 2 представи за физиологичните процеси в синаптичната връзка показват, че намаляването на активността на клетката за сравнително продължителен интервал от време намалява активността на метаболитните промени в нея. Наблюдаваните явления като отслабване, а понякога и изчезване на синапси от самосебе си се обясняват именно чрез намаляване на общата активност на клетката [49]. Разглежда се правилото на Хеб за обучението и в контекста на тези разсъждения.

## 3.3 Адаптивни обучаващи правила на *ART2*

Постигането на адаптивност към обкръжаващата среда в поведението на една невронна мрежа е свързано с вграждането на допълнителен механизъм, който предизвиква постепенно освобождаване на несъществената информация. Неточността и субективността в преценката за това доколко едно знание е несъществено предполага наличието на критерий за оценка. Когнитивната психология приема, че най-важния признак за това е степента на употреба на знанието при генерирането на реакция на нервната система. Прилагането на този критерий към изкуствените невронни мрежи ще доведе до това, че ако една мрежа бъде поставена в различна обкръжаваща среда, тя би имала различни критерии за полезност на знанията. Следователно дадено знание в една ситуация би било полезно, а в друга - не. Използваната информация в даден период на мрежова активност може да стане неизползваема на един по-късен етап. Системата би трябвало да се освободи от нея, давайки възможност на нова информация да изразходва освободените от това ресурси.

Идеята за освобождаване на информация по критерия "неизползвана дълго време" се използва при реализирането на системата *SPIN*, тъй като мотивите за това произтичат от схващанията в когнитивната психология, разгледани в предишната глава. Този модел обаче е свързан с добавянето на сложни характеристики на възлите и надзираващи подсистеми [23], [24], [39], [53], които освен че са биологично неправдоподобни, са твърде сложни за хардуерна реализация.

Класическият *ART2* модел притежава до известна степен способността да се адаптира към плавно променяща се среда. Това се проявява само при условие, че обучението е бавно, комбинирано с не голям параметър на бдителност. Така категориите консервативно и постепенно следват

несъществено променящото се входно пространство, преобучавайки съществуващите категории. Проблемите обаче се проявяват, когато някое от горните условия не е изпълнено. Ако типът на решаваната задача илисква ползването на голям брой тесни категории, тогава параметъра за бдителност е висок и всяко подаване на образец с малко отклонение от съществуващите категории ще предизвика създаване на нова категория. Причината за появата на такива образци би могла да е значителна промяна на входното пространство, както и наличието на значителни шумове във входните образци. По време на работата си системата създава категории, които представят рядко срещани прояви на обръжаващата среда, или пък нейни изчезнали характеристики в резултат от промяната и. Тези категории увеличават вероятността от изчерпване на ресурсите и блокиране на системата, като се има предвид, че класическият *ART2* модел не е в състояние сам да се освободи тях. Начинът по който това може да се направи е да се вгради допълнителен механизъм, който да реализира "забравяне" на тези ненужни категории.

Анализирайки обучаващите правила на класическия *ART2* модел се вижда, че състоянието на резонанс предизвиква промяна на следите само на онзи възел, чиито сигнали са предизвикали резонанса. Информацията, съхранена в останалите следите на двата адаптивни филтъра остава непроменена. За да може системата да се освободи от най-неизползваните си знания е необходимо да се направи така, че всички категории да попаднат в конкурентна среда, аналогично на схващанията за оцеляване на информацията в биологичните нервни системи. Това означава, че всички следите без изключение трябва да отслабват, докато тече обучението на онази категория, която е предизвикала резонанса. Трябва да се има предвид, че повторното използване на отслабените следите трябва да опресни силата им. По такъв начин често използваните категории оцеляват, а неизползваните непрекъснато губят силата си, докато се премине определена граница, след което следите могат да бъдат преобучени на нова категория.

Основната идея при този модел е схващането, че забравянето е самоиндуциращо се обучение с обратен знак. Вграждането на този механизъм в обучаващите правила би освободил модела от допълнителни подсистеми или свойства на възлите. Необходимо е при конструирането на новите обучаващи правила да се удовлетворят следните изисквания:

- Да се съхранява класическото *ART2* обучение на победилния възел. Това гарантира съхранението на произтичащите от класическия *ART2* модел функционални особености.

- Да се отслабва силата на следите във всички непобедили възли. Ако съхраняването на информацията се окаже експоненциално намаляваща функция, този модел би изглеждал биологично правдоподобен във връзка с разглежданията в Глава 2.
- Всяко отслабване на следите да съхранява кодираната информацията. В противен случай този процес би довел до изкривяване или разрушаване на придобитите знания, а оттам – нестабилност на системата. Съхраняване на кодираната информация означава, че  $n$ -мерните вектори на адаптивните филтри могат да се скъсват, но трябва да запазват посоката си.
- Всеки ресурс на *LTM* да се освобождава след достатъчно дълго неизползване. Това означава, че за всяка категория има предел на неупотреба, след който съответният възел се счита за свободен.
- Да се въведе нов мрежов параметър, с чиято помощ се регулира степента на отслабване на следите.

Изброевите изисквания гарантират стабилната работа на системата, като същевременно внасят допълнителна адаптивност на модела.

Авторът предлага нови обучаващи правила, наричани за краткост *адаптивни*, които се представят от следните диференциалните уравнения:

$$\frac{dz_{ji}}{dt} = g(y_j)[p_i - z_{ji}] - \lambda[d - g(y_j)]z_{ji}, \quad (3.1)$$

$$\frac{dz_{ij}}{dt} = g(y_j)[p_i - z_{ij}] - \lambda[d - g(y_j)]z_{ij}, \quad (3.2)$$

където  $0 \leq \lambda \ll 1$  е нов мрежов параметър на *запаметяване*, определящ способността на системата да съхранява информация продължително време.

От (1.2) и (1.12) следва, че

$$p_i = \begin{cases} u_i & , \text{ ако } F2 \text{ е неактивен} \\ u_i + dz_{ji} & , \text{ ако } J^{-\text{та}} F2 \text{ възел е активен} \end{cases}$$

и следователно

$$\frac{dz_{ji}}{dt} = \begin{cases} g(y_j)[p_i - z_{ji}] & , \text{ ако } g(y_j) = d \\ -\lambda z_{ji} d & , \text{ ако } g(y_j) = 0 \end{cases}, \quad (3.3)$$

$$\frac{dz_{ij}}{dt} = \begin{cases} g(y_j)[p_i - z_{ij}] & , \text{ ако } g(y_j) = d \\ -\lambda z_{ij} d & , \text{ ако } g(y_j) = 0 \end{cases}. \quad (3.4)$$



Сравнението на уравнения (1.14) и (1.14) с уравнения (3.3) и (3.4) показва, че адаптивните правила обучават следите на победилia  $F2$  възел точно както класическите, но освен това съществува обучение и на всички останали следи чрез диференциалните уравнения

$$\frac{dz_{ij}}{dt} = -\lambda z_{ij} d, \quad (3.5)$$

$$\frac{dz_{ji}}{dt} = -\lambda z_{ji} d. \quad (3.6)$$

От (3.3) и (3.4) следва, че при  $\lambda = 0$  адаптивните правила стават еквивалентни на класическите, т.е. адаптивният модел се явява обобщение на класическия.

За целите на по-нататъшните разглеждания ще бъдат дефинирани следните понятия:

- Един  $F2$  възел се нарича *активен*, когато неговите сигнали предизвикват състояние на резонанс на ART2 мрежа в отговор на произволен входен образец.
- Един  $F2$  възел се нарича *неактивен*, когато той не е активен възел в отговор на произволен входен образец.
- Един  $F2$  възел се нарича *активен в интервал от време*  $[t_0, t_k]$ , когато той е активен за поне един входен образец, подаден по време на този интервал.
- Един  $F2$  възел се нарича *неактивен в интервал от време*  $[t_0, t_k]$ , когато той не е бил неактивен през целия интервал.

Във връзка с тези дефиниции може да се говори и за интервал на неактивност на даден  $F2$  възел.

### 3.4 Отслабване на LTM следите

Двойственият характер на обучението, реализирано в правила (3.1) и (3.2) се определя както от промяната на следите в активния  $F2$  възел, така и от намаляването на теглата на всички неактивни  $F2$  възли.

Имайки предвид, че моделът ART2 активира обучението само в състояние на резонанс, обучаващите правила (3.1) и (3.2) променят LTM следите съобразно (3.5) и (3.6) [54].

**Лема 1** *Нека една ART2 невронна мрежа притежава обучаващи правила (3.1) и (3.2) при  $0 < \lambda < 1$ . Тогава ненулевите LTM следи на всички неактивни в кратък интервал възли експоненциално намаляват стойността си.*

### Доказателство:

Нека разгледаме обучение в отговор на произволен входен образец, протичащо в интервала от време  $[t_0, t_1]$ . Нека означенията

$$z(t_0) = z^0,$$

$$\frac{dz}{dt} = -\lambda z d$$

описват обучението на коя да е ненулева *LTM* следа на неактивен *F2* възел чрез обучаващи правила (3.1) и (3.2). Тогава решението на горното диференциално уравнение се представя във вида:

$$z(t) = z^0 e^{(-\lambda d)(t-t_0)} \quad t \in [t_0, t_1]. \quad (3.7)$$

Нека сега разгледаме обучението в отговор на последователност от  $k$  входни образца, което продължава в интервала  $[t_0, t_k]$ , състоящ се от  $k$  подинтервала  $[t_{i-1}, t_i]$ ,  $i = 1 \dots k$ . По време на всеки такъв подинтервал напълно приключва обучението на един възел – активния *F2* възел на съответния входен образец.

Игнорирайки времето между обученията приемаме, че интервала  $[t_0, t_k]$  е непрекъснат. От (3.7) следва, че крайната стойност на една неактивна *LTM* следа в даден подинтервал е начална стойност за следващия подинтервал. Тогава  $z(t)$  е непрекъснатата функция на времето в период на неактивност, т.е. уравнение (3.7) е валидно за непрекъснатия интервал  $[t_0, t_k]$ . От (3.7) и неравенства  $0 < \lambda < 1$  и  $0 < d < 1$  следва, че

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

■

Експоненциалният характер на отслабване на *LTM* следите може да се счита биологично правдоподобен във връзка с представите за тези процеси, описани в Глава 2.

## 3.5 Съхраняване на кодираната информация

Отслабването на следите, принадлежащи на неактивните *F2* възли трябва да съхрани кодираната в тях информация, тъй като противното води до разрушаване на придобитите знания, постоянна нестабилност на мрежата и нежелани резултати по време на работа.

За да бъде съхранена информацията в  $LTM$  векторите на неактивните възли, те не бива да променят посоката си. Това означава, че отслабването трябва само да ги скъсява и по този начин да мащабира кодиращия образ, като същевременно запазва пропорциите му.

Подобна трансформация е реализирана в слоевете  $F0$  и  $F1$  на класическия  $ART2$  модел, когато  $n$ -мерният вектор на един входен образец циркулира, претърпявайки нормализация. Тази промяна в силата на сигналите не води до промяна на кодираната в тях информация. Обучаващите правила (3.1) и (3.2) притежават способността да съхраняват информацията в  $LTM$  векторите на неактивните възли, запазвайки по този начин стабилността на системата. Това се доказва от следващата лема [54].

**Лема 2** *Нека  $ART2$  невронна мрежа притежава обучаващи правила (3.1) и (3.2) при  $0 < \lambda < 1$ . Тогава всеки неактивен  $F2$  възел съхранява кодираната в следите му информация.*

**Доказателство:**

Нека  $\tilde{z}$  и  $\bar{z}$  са две произволно избрани следи на един неактивен  $F2$  възел. Нека  $[t_0, t_1]$  е интервала, в който протича обучението след подаване на произволен входен образец. От (3.7) следва, че за кой да е момент  $t \in [t_0, t_1]$

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

Следователно отслабването на теглата не влияе върху отношението между стойностите на две произволно избрани  $LTM$  следи. От тук следва, че  $LTM$  векторът, представящ категорията на съответния възел се скъсява, но запазва посоката си. Следователно и кодираната в него информация се съхранява.

■

Може да се направи извода, че

**Следствие 1** *Нека  $ART2$  невронна мрежа притежава обучаващи правила (3.1) и (3.2). Тогава всеки неактивен в произволен краен интервал  $F2$  възел съхранява кодираната в следите му информация.*

Доказателството е очевидно.

■

### 3.6 Теорема за освобождаване на ресурса

Когато отслабването на  $LTM$  следите на един възел достигне критична стойност, адаптивният  $ART2$  модел счита този възел за потенциално

*свободен*. Смисълът на това понятие е, че възела поддържа все още принадлежащата му категория, но е готов да я замени с друга. Когато той спечели състезанието в  $F2$ , постъпващите сигнали в ориентиращата подсистема преодоляват прага на бдителност и мрежата изпада в състояние на резонанс с последващо обучение. По такъв начин ориентиращата подсистема не прави разлика между потенциално свободен възел и първоначално необвързан възел. Така възлите могат да бъдат обучавани многократно на съществено различни една от друга категории, а невронната мрежа може да се адаптира към променяща се обкръжаваща среда при непроменящ обема си ресурс. Способността на адаптивните обучаващи правила да освобождават неизползваните дълго време ресурси се доказва от следващата теорема [54].

**Теорема 1 (за освобождаване на ресурса)** *Нека ART2 невронна мрежа притежава обучаващи правила (3.1), (3.2) и параметри  $0 < \rho < 1$ ,  $0 < \lambda < 1$ . Тогава за всеки обвързан  $F2$  възел съществува краен интервал на неактивност, след който възелът се счита за потенциално свободен.*

**Доказателство:**

Ориентиращата подсистема предизвиква състояние на резонанс винаги, когато съществува активен входен образец и е удовлетворено неравенството

$$\frac{\rho}{\|r\|} \leq 1, \quad (3.8)$$

където  $r \equiv (r_1 \dots r_M)$ .

Имайки предвид, че  $\|x\| = \|u\| = \|q\| = 1$ , от (1.20) следва

$$\|r\| = \frac{(1 + 2\|cp\| \cos(u, p) + \|cp\|^2)^{\frac{1}{2}}}{1 + \|cp\|}. \quad (3.9)$$

При активност на  $J$ -ти  $F2$  възел в отговор на подаден входен образец, от уравнение (1.2) следва, че  $p = u + z_J d$ . Почленното скалярно умножение с  $u$ , както и повдигането му на квадрат води до равенства

$$\|p\| \cos(u, p) = 1 + \|z_J d\| \cos(u, z_J), \quad (3.10)$$

$$\|p\| = (1 + 2\|z_J d\| \cos(u, z_J) + \|z_J d\|^2)^{\frac{1}{2}}. \quad (3.11)$$

Обучаващите правила (3.1) и (3.2) предизвикват отслабване на  $LTM$  теглата на всеки възел в неговия интервал на неактивност. Разглеждаме  $z_J$  като векторна функция на времето. От (3.9), (3.10) и (3.11) следва, че

$$\|r(t)\| = \frac{((1+c)^2 + 2(1+c)\|z_J(t)cd\|\cos(u, z_J(t)) + \|z_J(t)cd\|^2)^{\frac{1}{2}}}{1 + (c^2 + 2c\|z_J(t)cd\|\cos(u, z_J(t)) + \|z_J(t)cd\|^2)^{\frac{1}{2}}}. \quad (3.12)$$

Нека един произволно избран *LTM* образец е неактивен в интервала  $[t_0, t_k]$  и е кодиран чрез *BU LTM* вектора  $z^J \equiv (z_{1J} \dots z_{MJ})$  и *TD LTM* вектора  $z_J \equiv (z_{JM+1} \dots z_{JN})$ . Разглеждаме динамиката на  $z_J$  при удължаване на неактивния интервал за  $t \in [t_0, t_k]$ . От лема 1 следва, че

$$\lim_{t, t_k \rightarrow \infty} \|z_J(t)\| = 0.$$

От лема 2 следва, че

$$\lim_{t, t_k \rightarrow \infty} \cos(u, z_J(t)) = D,$$

където  $D$  е константа и  $-1 \leq D \leq 1$ .

След граничен преход в (3.12) получаваме

$$\lim_{t, t_k \rightarrow \infty} \|r(t)\| = \frac{((1+c)^2)^{\frac{1}{2}}}{1 + (c^2)^{\frac{1}{2}}} = 1. \quad (3.13)$$

От тук следва, че за произволна фиксирана стойност на  $\rho$ ,  $0 < \rho < 1$  съществува момент  $\bar{t}$ ,  $t_0 < \bar{t}$  такъв, че за всяко  $t'$ :  $\bar{t} < t'$

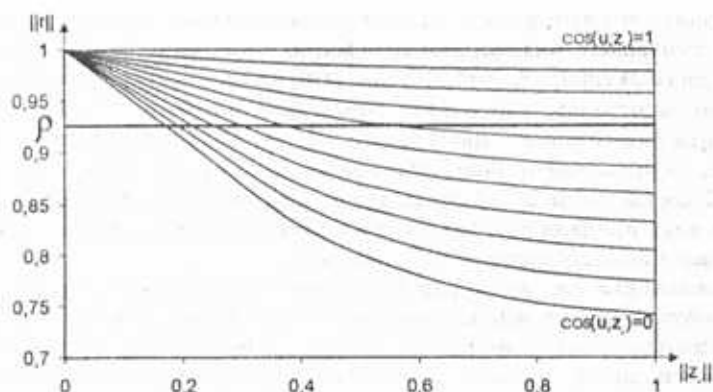
$$\|r(t')\| > \rho.$$

Следователно използването на неактивен в интервала  $[t_0, \bar{t}]$  *F2* възел след момента  $\bar{t}$  предизвиква резонанс и подмяната на кодираната в следите на възела информация с тази на входния образец.

■

### 3.7 Компромисът обучение – забравяне

Отслабването на следите на неактивните възли в съответствие с обучаващи правила (3.1) и (3.2) оказва пряко влияние върху работата на ориентиращата подсистема и оттам върху поведението на мрежата като цяло. Това влияние се предизвиква от промяната на дължината на вектора  $r$  след отслабването на следите на неактивните възли. Обучаващите правила предизвикват съксяване на векторите, представлящи неактивните категории както в правия, така и в обратния адаптивен филтър. Това от своя страна намалява влиянието на заученото от мрежата очакване и



Фигура 3.1: Графика на  $\|r\|$  като функция на  $\|z_j\|$  при  $c = 0.1$  и  $d = 0.9$ .

увеличава влиянието на входния образец. В резултат на това постъпващите в ориентиращата подсистема вектори, стават нарастващо успоредни, което води до увеличаване на дължината на  $r$ . Графично този процес е показан на Фигура 3.1. Изобразените криви представят дължината на вектора  $r$  като функция на дължината на  $z_j$  (умножена с константите  $c$  и  $d$ ). Кривите са получени от уравнение 3.12 при различни стойности на  $\cos(u, z_j(t))$ , т.е. степен на сходство между постъпващия входен образец и наученото очакване, представено от вектора на съответната категория. На фигурата е изобразен прага  $\rho$ , над който системата изпада в резонанс. Отслабването на следите след неактивност съответства на движението по някоя от изобразените криви в посока отлясно надясно. Вижда се, че отслабването на следите неизбежно води до превръщането на възела в потенциално свободен, в което се заключава смисъла на доказаната теорема.

Отслабването на следите в неактивните възли на правилен адаптивен филтър води до промяна на поведението на мрежата. Съксяването на вектора, представящ неактивна категория намалява скаларното произведение на този вектор с  $r$ . Тъй като силата на подаваните към  $F2$  сигнали играе решаваща роля за определянето на неврона-победител, отслабването на следите би могло да промени реда на избор на побеждаващите възли в  $F1 - F2$  цикъла.

Нека разгледаме ситуация, при която един входен образец се причислява към дадена категория при първия опит, активирайки съответния  $F2$

възел. Да приемем, че по-късно следите на този възел отслабват поради неизползване. В случай, че същият входен образец се подаде към мрежата, постъпващите сигнали към този възел могат да се окажат недостатъчни за да го активират. Тогава търсенето на съответствие минава през неуспешно съпоставяне с други категории докато изборът отново се спре на дадената категория и съпоставянето успее. По такъв начин припомнянето на заучена информация изисква един по-продължителен търсец процес. Това би могло да се счита за биологично правдоподобно, особено ако се имат предвид разгледаните в Глава 2 теории за състезателност на спомените и за достъпност на спомените.

Разсъждавайки изобщо върху биологичното правдоподобие на адаптивните обучаващи правила, може да се отбележи, че постигането на експоненциално отслабване на следите не е единственият признак за такова правдоподобие. Начинът, по който се реализира отслабването на възлите би могъл да се съпостави с интерференсната теория и по-точно с ретроактивното подтискане на придобити знания от новопостъпващата информация. По същия начин адаптивните обучаващи правила отслабват следите в адаптивните филтри само по време на обучение, т.е. на възприемане на новата информация. Разбира се постигането на биологично правдоподобие не е самоцел при разработването на адаптивния *ART2* модел, защото липсата на единна психологична теория, обясняваща функционирането на биологичните нервни системи превръща такъв стремеж в рисковано занимание. Всъщност акцентът при разработката на този модел пада най-вече върху постигането на мрежова пластичност при решаване на някои класове задачи.

Наличието на паралелизъм в отслабването на неактивните възли в полето *F2* позволява моделирането на тези процеси чрез обобщени мержи на Петри [30], [31], [32].

## 3.8 Организация и структура на адаптивния *ART2* симулатор

### 3.8.1 Адаптивен *ART2* симулатор

За целите на изследванията в този труд е създаден симулатор на невронна мрежа, построена върху *ART2* модел с предпроцесорен слой *F0*. В симулатора е вградена динамика, определена от уравнения (1.2) – (1.7), които характеризират *STM* активностите  $p_i$ ,  $q_i$ ,  $u_i$ ,  $v_i$ ,  $w_i$  и  $x_i$ , пресмятани в полетата *F0* и *F1*, както и уравнение (1.9) за функцията  $f(x)$ . Уравнения (1.10) – (1.13) описват динамиката полето *F2*, а (1.20) и (1.21) реализират особеностите на ориентиращата подсистема.

Обучаващите правила се представят чрез диференциалните уравнения (1.16) – (1.19), които се решават с числен метод. Използван е метода на Рунге-Кута от четвърти ред, където задачата на Коши за диференциално уравнение от първи ред е

$$y' = f(x, y),$$
$$y(x_0) = y_0.$$

Грешката при пресмятане е от порядък  $O(h^5)$  при големината на стъпката  $h$ .

Обучението се извършва итеративно, като след всяка стъпка се изчислява сумарната разлика на *STM* активностите на *F2* и *F1* между две последователни стъпки. Счита се, че обучението е завършило, ако едновременно са изпълнени двете условия:

- сумарната разлика на *STM* активностите във *F1* между две последователни стъпки е под прага *ac\_dif*
- сумарната разлика на *STM* активностите във *F2* между две последователни стъпки е под прага *z\_dif*

Стойностите на *ac\_dif* и *z\_dif* играят важна роля в поведението на *ART2*, поради което се включват като параметри на мрежата. Счита се, че *F1* и *F2* са стабилизирани, когато разликата на *STM* активностите на две последователни итерации падне под стойността на даден праг  $\mu$ . За малки стойности на  $\mu$  равновесието може да изисква много итерации. Така стойността на  $\mu$  определя обучението като *бързо* или *бавно*.

Бързото обучение може да се дефинира като режим, при който се постига строго равновесие ( $\mu$  е заложено много малко) при подаване на входен образец.

За разлика от него бавното обучение се дефинира като режим, при който се използва малък брой итерации. Това предполага, че мрежата като цяло ще се обучава бавно на новите постъпващи образци, запазвайки до известна степен старите си знания. Този консерватизъм я предпазва от бързо разрушаване на научените знания, "доверявайки" се на новите образци. Подобно поведение е реализирано и в модификации на *ART1* модела чрез промяна на обучаващите правила [4].

Тъй като бавното обучение има значително по-голяма биологична правдоподобност и приложимост, то е избрано за обучаващ модел на описания симулатор. Стойностите на параметрите и цялостната настройка на модела за конкретно приложение са разгледани в следващата глава.

Системата е програмирана на езика *C++*, а кодът е показан в Приложение А на дисертацията.



### 3.8.2 Функциониране

Преди подаването на някакъв вход към *ART2* мрежа се приема, че всички клетки имат активност нула, теглата на правия адаптивен филтър са инициализирани с еднакви малки стойности, а теглата на обратния адаптивен филтър са нули. Подаването на входният образец активира вътрешните цикли в полето *F0*, които приключват след достигане на заложен чрез параметър критерий за стабилност. Това означава, че при циркулирането един *STM* образец се променя незначително или не се променя. Стабилизирането на това поле активира изхода едновременно към полето *F1* и ориентиращата подсистема чрез *STM* образец  $q$ , който представя входния образец чрез нормиран вектор, претърпял допълнителна обработка за подтискане на шума и увеличение на контраста.

Тъй като в този момент всички възли на *F2* имат активност нула, полето *F1* получава вход само откъм *F0*. Сигналите циркулират в трите подполета до установяване на стабилност. Ако параметрите на *F0* и *F1* съвпадат, стабилността се постига веднага след първия цикъл. Тогава *STM* образца  $p$  подава сигнали към полето *F2* през правия адаптивен филтър.

Тъй като *F1* и *F2* са напълно свързани чрез теглови връзки (и двете посоки), правата теглова матрица филтрира сигнала на вектора  $p$ , подавайки резултатен вектор към *F2*. Този вектор би могъл да се изрази и като произведение на тегловата матрицата и вектора  $p$ . Началните стойности на правата теглова матрица биха могли да са както случайно избрани малки стойности, така и еднакви малки стойности.

Достигайки до полето *F2*, сигналите активират неговите възли, влизайки във вътрешното циклиране на връзките *on-center*, *off-surround*. Стабилизирането на полето предполага, че възелът с максимален сумарен входен сигнал е увеличил активността си, подтискайки останалите. Съществено в случая е да се отбележи, че само един от възлите е активен. Алгоритмичната реализация на този модел предполага, че при наличието на няколко еднакви максимални активности се избира една от тях съобразно някаква стратегия (в случая е възприета "първия максимален"). Така изходът на *F2* е *STM* активност, която се представя чрез вектор с една компонента със стойност  $d$ , а всички останали са нули.

Тъй като връзките между *F2* и *F1* чрез обратния адаптивен филтър са инициализирани с нулеви тегла, първоначално липсва сигнал от *F2* към *F1*. Това предполага, че двата *STM* образца, постъпващи в ориентиращата подсистема са почти или изцяло сходни. Поради това, че мярката за сходство между тях се изразява чрез дължината на вектора  $r$ , тя надхвърля параметъра на бдителност  $\rho$ . В тази ситуация ориентиращата подсистема не генерира *reset*-сигнал. Сходството на постъпващите в нея сигнали предизвикват резонанс, което активира обучението в *LTM*.

Промени в адаптивните филтри стават само когато търсенето завърши с успешно съпоставяне, водещо системата до състояние на резонанс. Обучението, което по същество е самообучение, се изразява в промяна на теглата на правил и обратен адаптивен филтър. По време на обучението тегловите вектори към и от активния  $F2$  възел стават "нарастващо успоредни" към текущия вектор  $p$ . Всички останали вектори, които не са свързани с възела-победител не се обучават.

От гледна точка на прави адаптивен филтър това обучение увеличава шансовете на този  $F2$  възел да се активира при подаване на същия или подобен входен образец. От друга страна обаче, след обучението обратния адаптивен филтър представя прототипа на категорията, или звученото очакване.

Да предположим, че активността на мрежата е установена на нула и се подава втори входен образец, подобен (но не идентичен) на първия. Още веднъж входния образец се установява в  $F1$ , подавайки сигнал към  $F2$ . Присмаме, че избора в  $F2$  е паднал върху същия възел, който се е активирал от първия образец. След обучението, за разлика от първия случай, вектора  $p$  става амалгама от постъпващите сигнали  $i$  и  $v$  в права посока и сигналите от обратния адаптивен филтър, представящи заучения прототипен образец. Полученият по такъв начин вектор не е съвсем еднакъв с входния образец поради съществуващата разлика между първото и второ представяне.

Ако в ориентиращата подсистема степента на сходство е достатъчно голяма за да надхвърли параметъра на бдителност, тогава се предизвиква резонанс и обучение на теглата както преди. Ако обаче степента на сходство е под  $\rho$ , тогава се изпраща *reset* сигнал към всички възли в  $F2$  с цел установяване активността им на нула. Тъй като един единствен възел е активен в този момент, само той ще бъде ефективно и трайно подтиснат. По такъв начин възелът се изключва от бъдещи състезания до момента, в който мрежата изпадне в състояние на резонанс. Този *reset* сигнал автоматично подтиква подсистемата на вниманието да продължи паралелното търсене на нови категории. След това състезанието се възобновява, но вече без участието на подтиснатите по описания начин възли. Като резултат от това друг  $F2$  възел се активира. Ако той е необвързан, т.е. никога не е участвал в обучение, тогава следва резонанс и обучение. В случай, че е обвързан, резонанс и обучение ще има само, ако тестът за бдителност премине успешно.

За всеки следващ входен образец се прилага същата процедура. Тестването продължава, докато се намери адекватно съпоставяне или се създаде нова категория. Извлеченото научно очакване и ориентиращата подсистема са единствените неща, чрез които *ART* мрежата поддържа активна регулация на обучаващия процес.

## Глава 4

### Експериментална оценка на адаптивния *ART2* модел

#### 4.1 Въведение

Тази глава е посветена на някои детайли, свързани със софтуерната реализация на адаптивна *ART2* невронна мрежа, както и на резултатите от нейната работа.

В началото се разглеждат въпроси, свързани с настройката на създадения невронен симулатор, тъй като тази дейност е решаваща за правилното функциониране на системата.

В главата се описват експерименталните резултати, показващи подобрите качества на адаптивното обучение в сравнение с класическото *ART2* обучение при решаване на някои класове задачи.

Разглежда се и едно приложение на адаптивния *ART2* модел за класификация на субекти от фондовия пазар.

#### 4.2 Невронен симулатор

За нуждите на експериментите е използван симулатор на невронна мрежа, реализиращ адаптивния *ART2* модел. Системата е програмирана с отворен и опростен интерфейс, както и с възможности за вграждане в други програмни системи като тяхно ядро. Кодът на програмата е написан на езика *C++* и е представен в Приложение А. За установяване на работоспособността на симулатора е използвано оригиналното тестващо средство, предложено от *Carpenter* и *Grossberg* в [13] и представено в Приложение Б. Разработен е втори симулатор на невронна мрежа, базиран на класическата *ART2* архитектура с предпроцесорен слой. Предназначението му е да се използва за оценка и сравняване на адаптивния и класическия *ART2* модел.

### 4.2.1 Основни параметри и настройка

Една невронна мрежа, базирана на *ART2* архитектура демонстрира голяма чувствителност към промяна на параметрите си. Ефективността и зависи до голяма степен от подходящата настройка на мрежовите параметри [5], чрез които може да се влияе върху поведението на мрежата и възможността ѝ да удовлетворява изискванията на поставената задача. В главата е дискутирана настройката на всеки един от мрежовите параметри, като се дава препоръчителен алгоритъм за начина, по който да се прави това. Постигането на максимална чувствителност на мрежата при  $\rho = 0.95$  за споменатото тестващо множество се получава при стойности на параметрите  $a = b = 10$ ,  $c = 0.1$ ,  $d = 0.9$ ,  $\theta = 0.2$ ,  $z_{ij0} = 0.5$ ,  $ac\_dif = 0.001$ ,  $z\_dif = 0.01$ ,  $h = 0.1$ ,  $nsteps = 100$ .

### 4.2.2 Сравнение на класическия и адаптивния *ART2* модел

Освен описаните по-горе параметри, адаптивният *ART2* модел разполага с допълнителен мрежов параметър -  $\lambda$ , определящ степента с която отслабват следите на неактивните възли. Намирането на подходящи стойности на  $\lambda$  е особено важно за стабилността на системата, защото твърде голямото отслабване на следите в адаптивните филтри освобождава възлите дори на категории, които са били използвани често. Ако стойностите на  $\lambda$  са твърде малки, тогава ефектът от отслабването на следите няма да е забележим и мрежата ще има поведението на класическия *ART2* модел. Подходящите стойности за  $\lambda$  се определят от изискването за брой категории, които ще бъдат формирани, честота на използване на всяка категория и изискването за "разумност" в забравянето, съобразено с конкретната задача.

Евентуалното загубване на една категория в резултат от забравянето не е фатално за системата дори когато тя има нужда от нея, защото би могла да се обучи на тази категория отново, стига съответните образци да се появят на входа на мрежата. Такава ситуация не изглежда страшна, защото се наблюдава често при биологичните нервни системи.

Експериментите, които целят сравняване на класическия и адаптивния модел могат да се разделят на три основни групи [54]:

- Установяване на идентичност в поведението на класическия и адаптивния модел при  $\lambda = 0$  или за твърде малки стойности на параметъра. Това торетично произтича от заключението, че класическите обучаващи правила се явяват частен случай на адаптивните при  $\lambda = 0$ .
- Тестване способността на адаптивния модел да освобождава ресурсите от категории, които дълго време не са били използвани.

- Способност на адаптивния *ART2* модел да работи в променящо се входно пространство без излишен разход на ресурси. Отпадането на съществуващи категории от входното пространство поради промяна на средата в която системата работи води до излишно съхраняване на вече непотребна информация. Класическият *ART2* модел не може да се освободи от нея, а при продължителна работа се увеличава и вероятността за блокиране на системата поради изчерпване на ресурсите.

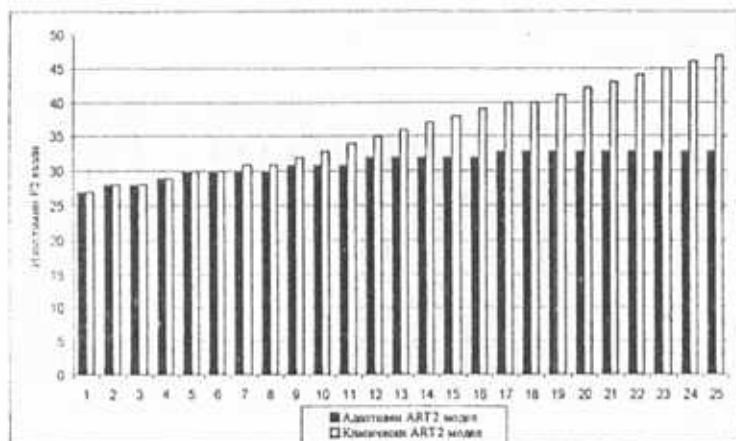
**Първата група експерименти** показва, че за  $0 \leq \lambda \leq 0.0015$  няма разлика между резултатите от класификацията на класическия и адаптивния модел.

**Втората група експерименти** използва същото тестващо множество, но към него са добавени 2 типа допълнителни входни образци: образци, за които е доказано че създават нови категории; силно зашумени образци от тестващото множество, получени както с внасяне на излишък от сигнали, така и отнемане на сигнали. Създадените от допълнителните образци категории могат да се считат за непотребни и биха могли да се породят от погрешно зададени или неспецифични за решаваната задача образци, силни шумове при подаване на образците или пропадане на част от сигналите. Това, което обединява всички тези случаи е обстоятелството, че породените от тях категории много рядко се използват или изобщо не се използват повторно. По тази причина всеки допълнителен образец се среща след 500 основни образца. За целта са създадени 25 входни файла от 501 образца, съдържащи 500 случайно наредени повтарящи се основни образци, а накрая е добавен един от допълнителните. Така образуваната серия от 12525 входни образци е подадена и на двата симулатора, с параметри  $\rho = 0.975$ ,  $a = b = 10$ ,  $c = 0.1$ ,  $d = 0.9$ ,  $\theta = 0.2$ ,  $z_{tj0} = 0.5$ ,  $ac\_dif = 0.001$ ,  $z\_dif = 0.01$ ,  $h = 0.01$ ,  $nsteps = 100$  и  $\lambda = 0.002$  за адаптивния модел.

Резултатите показват, че след цялата серия класическия и адаптивния модел класифицират по един и същ начин основното входно множество в 26 категории. Използването на  $F^2$  възли по време на серията за всеки от моделите е показано на Фигура 4.1.

Вижда се, че след шестия входен файл адаптивният модел започва да използва потенциално свободни възли за новото обучение, докато класическият модел продължава да изразходва необвързани възли, запазвайки непотребните категории.

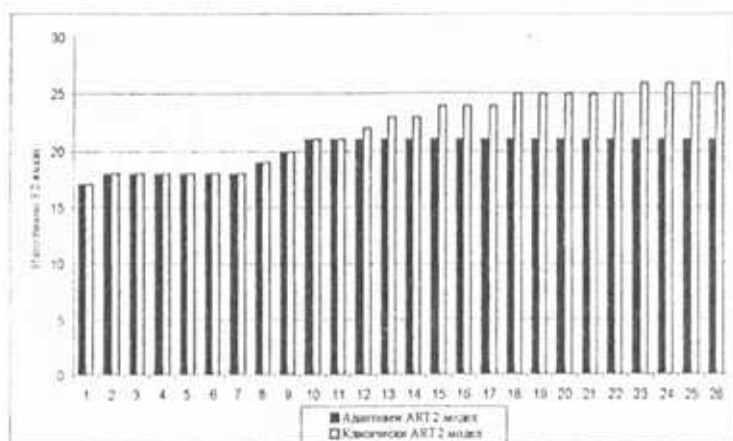
**Третата група експерименти** отново използва тестващото множество от Приложение Б, но този път 50-те входни образца са разделени по



Фигура 4.1: Използвани F2 възли по време на класификация на 12525 образци, разделени в 25 групи по 501 образци. Всяка група съдържа по един силно зашумен (нетипичен) образец.

случаи начин на две групи по 25, наречени  $SP1$  и  $SP2$ . Първоначално на мрежата се подава  $SP1$  и тя се самообучава на него. В последствие  $SP1$  започва постепенно да се променя в посока към  $SP2$ , а в края на тази промяна  $SP1$  се превръща в  $SP2$ . Промяната се реализира на 26 стъпки, като на всяка стъпка (без първата) един от образците на  $SP1$  отстъпва мястото си на образец от  $SP2$ . Трябва да се отбележи, че случайният характер на формиране на  $SP1$  и  $SP2$  би могъл да раздели образците на една и съща категория в  $SP1$  и  $SP2$ . Следователно една стъпка в промяната на  $SP1$  към  $SP2$  не води непременно до внасянето на качествено нови образци.

За целите на експериментите са създадени 26 файла по описания по-горе начин и са подавани последователно на двата симулатора при същите параметри. Както класическият, така и адаптивният модел класифицират цялата серия от образци в 13 категории, но адаптивният модел ползва значително по-малко ресурси за това. Както се вижда на Фигура 4.2, след стъпка 11 адаптивният модел изразходва ресурсите по-пестеливо, ползвайки потенциално свободните възли. Това разкрива неспособността на класическият ART2 модел да се освобождава от вече ненужните му категории и решението, което предлага адаптивният модел.



Фигура 4.2: Използване  $F2$  взли по време на класификация на променящо се изходно пространство. Промяната на  $SP1$  към  $SP2$  е на 26 стъпки, като всяка стъпка съответства на 500 изходни образеца.

### 4.3 Класификация на субекти от фондовия пазар

Адаптивният  $ART2$  модел е използван при решаване на задача за класификация на субекти от фондовия пазар с цел отчитане на техните характеристики в краткосрочен и дългосрочен план, както и подпомагане при вземане на решение за инвестиране в акции. Филтрирането на големия поток информация, комбинирано с присъщото на невронните мрежи свойство да установяват сходство между входните си образци са качествата, които правят системата привлекателна за решаването на тази задача.

Оценката на субектите във фондовия пазар се прави чрез анализ на данните, получени от официалния бюлетин на Българската Фондова Бурса (БФБ). Използвани са данните на 60 последователни бюлетина на БФБ за 280 икономически субекта, чиито акции се търгуват в борсата. За целите на задачата се въвеждат допълнителни показатели, които разглеждат неяви отношения в тези данни. Показателите могат да бъдат причислени към две групи. Първата включва онези от тях, които отразяват данните само от един (текущия) бюлетин и поради това са наречени *текущи показатели*. Втората група е от *дългосрочни показатели*, които описват отношения и проявяващи се тенденции през сравнително

по-дълъг период от време. За формирането им се налага използване на информация от всички бюлетени на БФБ в разглеждания период.

Текущите показатели включват относителен дивидент (ОД), рисков фактор при инвестиция (РФИ), относителна приходност (ОП), ценова устойчивост (ЦУ), абсолютен процент на промяна (АПП) и относителна разлика в оферти (ОРО). Дългосрочните показатели включват абсолютна средна промяна (АСП), усреднена цена (УЦ) и относителна амплитуда (ОА). Всеки входен образец за невронната мрежа е 9-компонентен вектор, съдържащ показателите на един икономически субект.

Системата позволява групирането на сходни субекти, като дава възможност на потребителя да зададе желана тежест на всеки от използваните показатели и с това да формира групи от субекти, чиито характеристики носят конкретен икономически смисъл. Това се постига чрез задаване на нови девет *мащабиращи параметъра* на мрежата – реални числа, които съответстват на деветте показателя. Входните образци претърпяват предварителна обработка, като всяка компонента на входния образец се умножава със съответния мащабиращ параметър. Така във входния образец изпъква една част от данните, а друга се подтиска. Подходящият избор на мащабиращи параметри е онова, което дава икономически смисъл на критерия за сходство между елементите, както и възможност за филтриране на потока от информация.

Експериментите са проведени с двата симулатора при мрежови параметри  $\rho = 0.995$ ,  $a = b = 10$ ,  $c = 0.1$ ,  $d = 0.9$ ,  $\theta = 0.1$ ,  $z_{ij0} = 0.5$ ,  $ac\_dif = 0.001$ ,  $z\_dif = 0.01$ ,  $h = 0.1$ ,  $nsteps = 100$  и  $\lambda = 0.003$  за адаптивния модел.

В първата група от експерименти е постигната подходяща класификация при доминиране както на всеки показател поотделно, така и на комбинация от показатели. По такъв начин едновременното усилване на параметрите РФИ, ЦУ и АСП (ОД = 1, РФИ = 5, ОП = 1, ЦУ = 5, АПП = 0.01, ОРО = 0.3, АСП = 0.1, УЦ = 0.00001, ОА = 0.3) позволява класификация на икономическите субекти по признака "стабилност" на инвестицията.

Втората група експерименти използва данните на първите 30 от общо 60 последователни бюлетина на БФБ за начално обучение на мрежата при параметри  $\rho = 0.995$ ,  $a = b = 10$ ,  $c = 0.1$ ,  $d = 0.9$ ,  $\theta = 0.1$ ,  $z_{ij0} = 0.5$ ,  $ac\_dif = 0.001$ ,  $z\_dif = 0.01$ ,  $h = 0.1$ ,  $nsteps = 100$  и  $\lambda = 0.003$  за адаптивния модел. Мащабиращите параметри имат стойности ОД=1, РФИ=1, ОП=1, ЦУ=1, АПП=0.01, ОРО=0.3, АСП=0.02, УЦ=0.00001, ОА=0.3. Към всеки от двата симулатора последователно се подават данните от останалите 30 бюлетина в режим на обучение. Отчита се също и принадлежността на образците към категориите и изразходваните  $F^2$  възли.



Резултатите показват идентична класификация при класическия и адаптивния модел, но адаптивният модел демонстрира икономичност в използването на ресурсите.

## Глава 5

### Заклучение

Тук ще бъдат обобщени накратко резултатите от този труд.

В дисертацията авторът въвежда нов модел невронни мрежи, базиран на адаптивната резонансна теория (adaptive resonance theory, *ART*) и по-точно модела *ART2* [13]. Отслабването на синаптичните връзки при продължителна неактивност и произтичащите от това процеси на постепенно забравяне са онези характеристики на обучението, които са намерили отражение в новите обучаващи правила, наречени *адаптивни*. Те дават името на предложениия от автора модел, наречен *адаптивен ART2 модел*.

Характеристиките на адаптивните *ART2* невронни мрежи допринасят и за преодоляване на един от недостатъците на класическия *ART2* модел, а именно блокирането му при изчерпване на мрежовите ресурси. Това прави адаптивните *ART2* невронни мрежи по-ефективни при решаването на някои класове задачи.

В дисертацията са постигнати следните научни и научно-приложни приноси:

1. Направен е анализ на съществуващите техники за моделиране на когнитивните процеси, свързани с отслабване и освобождаване на придобити знания в самообучаващи се невронни мрежи [23], [24], [Nac96], [Nac97], [Nac97b] и са показани техните предимства и недостатъци.
2. Предлага се нетрадиционен подход за моделиране на забравянето без добавяне на нови характеристики на възлите и надзираващи подсистеми, с което се постига простота и ефективност на модела.
3. Конструирани са нови диференциални уравнения, представлящи обучаващите правила [Nac98a], като с помощта на две лема и една теорема са доказани някои техни важни характеристики [Nac98b]. Доказва се теоретично, че класическият *ART2* модел се явява частен случай на адаптивния при стойност на мрежовия параметър  $\lambda = 0$  [Nac98a].

4. Въведният адаптивен *ART2* модел демонстрира биологично правдоподобие не само по отношение на експоненциалното отслабване на следите, но и по отношение на ретроактивното подтискане в интерференсната теория, теорията за състезателност на спомените и теорията за достъпност на спомените [Nac98b].
5. Програмно е реализиран симулатор на невронна мрежа, базиран на адаптивния *ART2* модел и разширена архитектура с предпроцесорен слой [Nac98b].
6. Експериментално е анализиран адаптивния *ART2* модел и са показани неговите предимства пред класическия при решаване на някои класове задачи. Представено е и едно приложение на адаптивния *ART2* модел за филтриране на информация от фондовия пазар.

Резултатите от дисертационния труд са публикувани в шест статии, представени в приложения списък.

- [16] Carpenter, G., S. Grossberg, D. Rosen. Fuzzy ART: Fast Stable Learning and Categorization of Analog Patterns by an Adaptive Resonance System. *Neural Networks*, 4, 759-771, 1991
- [17] Ceraso, J. The Interference theory of Forgetting. *Scientific American*, 217, 117-124, 1967
- [18] Churchland, P. S., T. J. Sejnowski. *The Computational Brain*. MIT Press, Cambridge, MA, 1993
- [19] Corsini, R. *Encyclopedia of Psychology*, Wiley-Interscience, New York, vol. 2, p.383, 1984
- [20] Deutsch, A. The Psychological Basis of Memory. *Academic Press*, New York, p.59-73, 1983
- [21] Estes, W. Statistical Theory of Spontaneous Recovery and Regression. *Psychol. Rev.*, vol.62, 207-223, 1955
- [22] Eysenck, H. *Encyclopedia of Psychology*, vol 1, Search Press, London, 1985
- [23] Fritzke, B. Unsupervised Clustering with Growing Cell Structures, *Proc. of the IJCNN'91 Seattle (IEEE)*, 1991.
- [24] Fritzke, B. Let It Grow - Self-Organizing Feature Map with Problem Dependent Cell Structure, *Proc. of the ICANN'91, Helsinki*, 1991.
- [25] Freud, S. *On Aphasia, a Critical Study*, International University Press, New York, 1941
- [26] Grossberg, S. Adaptive Pattern Classification and Universal Recoding, I: Parallel Development and Coding of Neural Feature Detectors. *Biological Cybernetics*, 23, 121-134, 1976
- [27] Grossberg, S., A Theory of Human Memory: Self Organization and Performance of Sensory-Motor Codes, Maps, and Plans. In R. Rosen and F. Snell (Eds.) *Progress in Theoretical Biology*, Vol. 5, New York, Academic Press, 1978
- [28] Grossberg, S. Associative and Competitive Principles of Learning and Development - The Temporal Unfolding and Stability of STM and LTM Patterns. in *Competition and Cooperation in Neural Nets*, Springer Verlag, New York, 1982
- [29] Grossberg, S. How Does a Brain Build a Cognitive Code? *Neurocomputing. Foundations of research*, Ed., MIT Press, Massachusetts, p. 347-399, 1988
- [30] Hadjyisky, L., K. Atanassov. Theorem for Representation of the Neural Networks by Generalized Nets. *AMSE Review*, Vol.12, No 3, 47-54, 1990
- [31] Hadjyisky, L., K. Atanassov. A Generalized Net, Representing the Elements of Neural Network Set. *AMSE Review*, Vol. 14, No 4, 55-59, 1990
- [32] Hadjyisky, L. K. Atanassov. Generalized Nets Representing the elements of Neural Networks in Applications of Generalized Nets. (K. Atanassov, Ed.), *World Scientific*, Singapore, New Jersey, London, 49-67, 1993.
- [33] Hamnerstrom, D., W. Henry, M. Kuhn. *Neurokomputer systems for neural network applications*. Prentice-Hall, Englewood Cliffs, NJ, 1992
- [34] Hartley, K., Seeing the Need for ART. *Sci. News* 132, 14, 1987
- [35] Hebb D. *The Organization and Behaviour*, New York : Witey, p.50, 1949.

- [36] Hodgkin, A., A. Huxley, A Quantitative Description of Membrane Current and Its Applications to Conductions and Excitation in Nerve. *J. Psychol. London*, 117, 500, 1952
- [37] Houston, J. *Fundamentals of Learning and Memory*, Academic Press, New York, p.306, 1981
- [38] Jenkins, J., K., Dallenbach Oblivescens During Sleep and Weaking. *American Journal of Psychology*, 35:605-612
- [39] Keuchel, H.,E. Putcamer, U. Zimmer. Learning and Forgetting Surface Classification with Dynamic Neural Networks, *Proc. of the ICANN'93*, Amsterdam, IX, 1993.
- [40] Kintsch, W. *Learning, Memory, and Conceptual Processes*. John Wiley & Sons, New York, p.139:143, 1970
- [41] Klatsky, R. *Human Memory: Structures and Processes*, W.H. Freeman and Company, New York, p.125, 1992
- [42] Kohonen, T. *Self-Organization and Associative Memory*, Springer, Berlin, 1984.
- [43] Kohonen, T. *Statistical Pattern Recognition Revisited*, *Advanced Network Computers*, R. Eckmiller (ed.), 1990.
- [44] Kolodzy, P. *Multidimensional Machine Vision Using Neural Networks*, *Proceedings of First International Conference on Neural Networks*, San Diego, IEEE, New York, 1987
- [45] Krose, B., P. Smagt. *An Introduction to Neural Networks*. Fourth edition, p.54, Amsterdam, 1991
- [46] Lajtha, A., J. Toth. *Instability of Cerebral Proteins*. *Biochemical and Biophysical Research Communication*, vol.23, 294-298, 1976.
- [47] Marcovich S., P. Scott. *The Role of Forgetting in Learning*. *Proc. of the 5-th International Conference on Machine Learning*. Ann Arbor, MI:Morgan Kaufman, 1988
- [48] McGeoch, J., A. Irion *The Psychology of Human Learning*, 2nd ed. Mc Kay, New York, 1952
- [49] MacGregor, R. J. *Neural and Brain Modeling*. San Diego: Academic Press, 1987
- [50] Melton, A., J. Irwin. *The Influence of Degree of Interpolated Learning on Retroactive Inhibition and the Overt Transfer of Specific Responses*. *American Journal of Psychology*, 53:173-203, 1940
- [51] Murre, J. M., *Neurosimulators*. In M.A. Arbib, ed., *Handbook of Brain Research and Neural Networks*. MIT Press, Cambridge, MA, 1995.
- [52] Nachev A. *Plasticity of ART Neural Networks in Case of Considerably Change of the Input Space*, *Proc. in Automatics and Informatics'97*, vol.6, Sofia, 1997
- [53] Nachev A., N. Griffith, A. Gerov. *Dynamic Learning in ART2 Neural Networks*. *Proc. of XXVII Spring conference of UBM, Mathematics and Education in Mathematics*, 195-201, Sofia, 1998
- [54] Nachev A., N. Griffith, A. Gerov. *Dynamic Learning - An Approach to Forgetting in ART2 Neural Networks*, *Lecture Notes in Artificial Intelligence 1480*, AIMS'98, Springer Verlag, Heidelberg, 1998

- [55] Scott, P. Learning: The Construction of a Posterior Knowledge Structures. Proc. of 3-th National Conference on AI. Washington, DC: Morgan Kaufman, 1983
- [56] Schalkoff, R., Artificial Neural Networks. McGraw-Hill International Editions, New York, 1997
- [57] Shepherd, G. M. Neurobiology, New York: Oxford University Press, 1988
- [58] Simon, H. Why Should Machines Learn? In: Michalski, R., J. Carbonell and T. Mitchell (eds), Machine Learning: Artificial Intelligence Approach, vol. I, Tigoa, 1983
- [59] Sjöstrand, J. Rapid Axoplasmic Transport of Labelled Proteins in the Vagus and Hypoglossal Nerves of the Rabbit. Experimental Brain Research, Vol. 8, 105-112, 1969
- [60] Solso, R. L. Theories in Cognitive Psychology: The Loyola Symposium. John Wiley & Sons, New York, p.3, 1984
- [61] Talland, A. Disorders of Memory and Learning, Penguin Books, London, p.55, 1982
- [62] Underwood, B. Interference and Forgetting, Psychological Review, 64, 49-60. Fig. 3, p.53, 1957
- [63] Widrow, B. R. G. Winter. Neural Nets for Adaptive Filtering and Adaptive Pattern Recognition. IEEE Computer, 21:25-39, March 1988
- [64] Wingfield, A. Human learning and Memory: An Introduction, Harper & Row Publishers Inc., p.302, New York, 1979
- [65] Woodworth, R.S., H. Schlosberg Experimental Psychology. Rev. ed. New York: Holt, Rinehart and Winston, 1954