

УДК 681.3

UDC 681.3

**СИНТЕЗ ПАРАМЕТРОВ НЕЙРОННОЙ СЕТИ ХОПФИЛДА ДЛЯ РЕШЕНИЯ ЗАДАЧИ РАНЖИРОВАНИЯ В ИНФОРМАЦИОННО-ПОИСКОВЫХ СИСТЕМАХ****SYNTHESIS OF PARAMETERS OF NEURAL NETWORK OF HOPFIELD FOR THE DECISION OF THE PROBLEM OF RANGING IN INFORMATION RETRIEVAL SYSTEMS**

Занин Дмитрий Евгеньевич, аспирант  
*Кубанский государственный технологический университет, Краснодар, Россия*

Zanin Dmitry Evgenievich, postgraduate student  
*Kuban State Technological University, Krasnodar, Russia*

Задача индексирования больших массивов информации предъявляет высокие требования по производительности к ИПС. Таким образом применение нейронной сети позволяет распараллелить задачу, то есть позволяют выполнять ресурсоемкие операции точного ранжирования для документов, имеющих шанс оказаться на достаточно высоких местах в выдаче по результатам более грубой оценки их релевантности.

The problem of indexing of the array of the information demands much on productivity of the information retrieval system. Thus application of a neural network allows to parallelize a problem, that is allow to carry out resource-intensive operations of exact ranging for the documents having chance to appear on enough high places in delivery by results of more rough estimate of their relevance.

Ключевые слова: НЕЙРОННАЯ СЕТЬ ХОПФИЛДА, РАНЖИРОВАНИЕ, ПАРАМЕТРЫ НЕЙРОННОЙ СЕТИ, ЭНЕРГИЯ СЕТИ

Keywords: NEURAL NETWORK OF HOPFIELD, RANGING, PARAMETERS OF NEURAL NETWORK, ENERGY OF NETWORK.

Без эффективного ранжирования результаты поиска в информационно-поисковой системе (ИПС) теряют смысл, так как могут включать в себя ссылки на десятки и сотни тысяч документов. В таких условиях ранжирование носит характер всеобщего проблемного императива, при этом основными проблемами развития научных основ архитектурных принципов ИПС являются недостаточная теоретическая проработка применяемых на практике подходов, сравнительно невысокий уровень использования развитых математических механизмов, отставание теоретических разработок от быстро меняющихся поисковых потребностей пользователей компьютерных сетей.

Необходимость обеспечить высокую производительность системы при большом объеме проиндексированной информации и векторном критерии значимости часто приводит к невозможности применить в ИПС алгоритмы, хорошо зарекомендовавшие себя в экспериментальных исследованиях. В нашем случае высокие требования по

производительности к ИПС обеспечиваются нейросетевым вычислительным базисом, позволяющим распараллеливать задачу, то есть позволяющим выполнять ресурсоемкие операции точного ранжирования для документов, имеющих шанс оказаться на достаточно высоких местах в выдаче по результатам более грубой оценки их релевантности.

Общая последовательность предлагаемого метода ранжирования перед выдачей результата поиска пользователю представлена четырьмя этапами.

1. Динамически определяется множество  $V$  смысловых соответствий (критериев значимости) между информационным запросом и полученным документом (результатом поиска), а также обозначается исходное - подлежащее ранжированию, множество  $D$  найденных документов.

2. Строится таблица соответствия множества  $D$  найденных документов множеству критериев значимости. Таблица определяет исходные данные для решения задачи ранжирования в рамках комбинаторной задачи о назначениях.

3. Формируется нейросетевая модель решения задачи ранжирования на основе динамической нейронной сети Хопфилда с БФВ.

4. Инициализируя нейронную сеть случайными входными векторами получаем искомую последовательность индексов документов в соответствии с заданным множеством критериев релевантности в ИПС.

Рассмотрим нейросетевую интерпретацию задачи ранжирования по множеству критериев как задачи о назначениях, при условии сведения задачи о назначениях к стандартной форме (число групп критериев равно числу ранжируемых документов).

Определим архитектуру нейронной сети, решающую задачу:

$$F(X) = \sum_{j=1}^M \sum_{i=1}^N r_{ji} x_{ji} \rightarrow \max, \quad (1)$$

$$\text{при ограничениях : } \begin{cases} \sum_{j=1}^M x_{ji} \leq 1, i = \overline{1, N}, \\ \sum_{i=1}^N x_{ji} \leq 1, j = \overline{1, M}, \\ x_{ji} \in \{0,1\}, j = \overline{1, M}, i = \overline{1, N} \end{cases} . \quad (2)$$

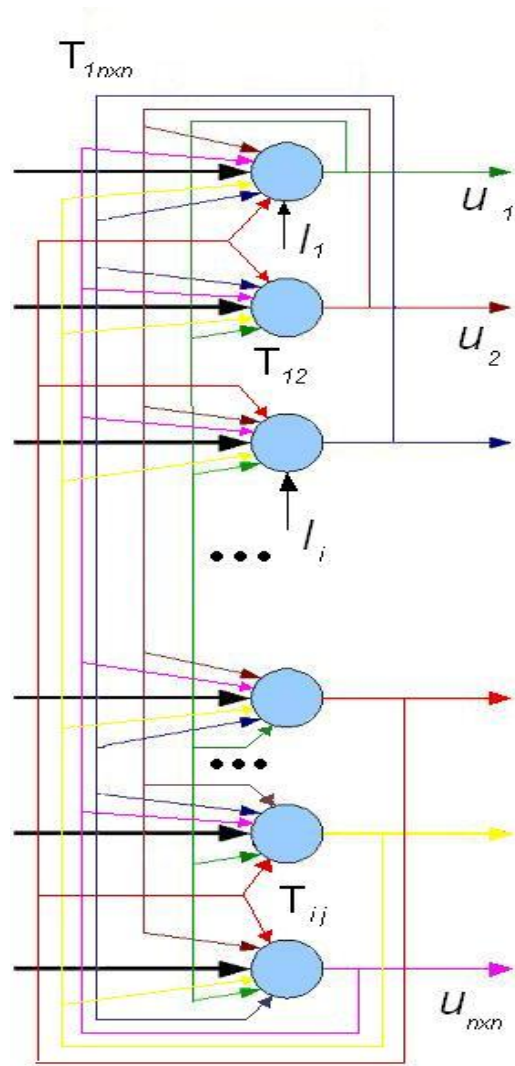
Введем в рассмотрение сеть бинарных нейронов, представляющую собой матрицу размерностью  $n \times n$ , где  $n = N = M$  – число документов или групп критериев.

За основу модели ранжирования может быть взята нейронная сеть (рисунок 1), содержащая обратные связи, по которым переданное возбуждение возвращается к нейронам, и они повторно выполняют свои функции [1-4].

В динамических нейронных сетях неустойчивость проявляется в блуждающей смене состояний нейронов, не приводящей к возникновению стационарных состояний. В общем случае ответ на вопрос об устойчивости динамики произвольной системы с обратными связями крайне сложен и до настоящего времени является открытым [5-7].

Пусть используемая нейронная сеть Хопфилда имеет следующие характеристики (рисунок 1):

1. Один слой элементов (входные элементы, представляющие входной образец, не учитываются).
2. Каждый элемент связывается со всеми другими элементами, но элемент не связывается с самим собой.
3. За один шаг обновляется только один элемент.
4. Элементы обновляются в случайном порядке, но в среднем каждый элемент должен обновляться в одной и той же мере (частоте).
5. Вывод элемента ограничен значениями 0 или 1, т.е функция выхода – бинарная [1].



**Рисунок 1 - Нейронная сеть Хопфилда с БФВ  $u=f(u,I,T)$ , где  $u$  – выходы нейрона,  $I$  – величины смещений,  $T$  – коэффициенты синаптических связей.**

Сеть Хопфилда является рекуррентной в том смысле, что для каждого входного образца выход сети повторно используется в качестве ввода до тех пор, пока не будет достигнуто устойчивое состояние.

Соответствующим образом организованная (запрограммированная) нейросеть после "запуска" меняет свое состояние, постепенно переходя в установившийся режим.

Удобно считать, что сеть Хопфилда не имеет входных элементов, так как входной вектор просто определяет начальные значения активности элементов. Элементы обновляются тогда, когда все элементы передадут свои значения активности по имеющимся взвешенным связям, после чего вычисляется сумма произведений (т.е. берется скалярное произведение).

Значение активности элемента получается на основе использования некоторого правила активизации.

Каждой целочисленной переменной  $x_{ij}$  поставим в соответствие выходной сигнал  $ij$ -го нейрона  $u_{ij}$ , стоящего в  $i$ -й строке и  $j$ -м столбце матрицы сети.

$$(x_{ij} = 1) \Leftrightarrow (u_{ij} = 1), \forall i, j \in \overline{1, n}. \quad (3)$$

На рисунке 2 схематично представлена матрица сети в состоянии покоя, где в виде заштрихованных квадратов изображены нейроны с единичными выходными сигналами.

Совокупность возбужденных нейронов интерпретируется как план назначений.

В соответствии с (3), интерпретируем ограничения (2) и целевую функцию (1), в результате получаем (4)-(7):

$$\sum_{j=1}^n u_{ji} = 1, \forall i \in \overline{1, n}, \quad (4)$$



**Рисунок 2 - План назначений - матрица нейронной сети Хопфилда в состоянии покоя**

$$\sum_{i=1}^n u_{ji} = 1, \forall j \in \overline{1, n}, \tag{5}$$

$$\sum_{j=1}^n \sum_{i=1}^n u_{ji} = n, \tag{6}$$

$$\Phi(u) = \sum_{j=1}^n \sum_{i=1}^n u_{ji} r_{ji} \rightarrow \max, \tag{7}$$

где  $u_{ji}$  – значения выхода нейронной сети Хопфилда (рисунок 1);

$r_{ji}$  – значения матрицы производительности элементы которой  $r_{ji}$  представляют собой релевантность документа с номером  $j$  относительно критерия (группы критериев) с номером  $i$ .

Сконструируем энергетическую функцию  $E^0(u)$ , минимизация которой обеспечивает выполнение ограничений (4)-(6) и решение задачи (7). Построим ее в виде

$$E^0(u) = E_{\Psi}^0(u) + E_{\Phi}^0(u), \tag{8}$$

где последнее слагаемое обеспечивает оптимизацию функции стоимости и с точностью до константы  $F > 0$  однозначно определяется следующим образом [7]:

$$E_{\Phi}^0(u) = -F \sum_{j=1}^n \sum_{i=1}^n u_{ji} r_{ji}, \quad (9)$$

а первое слагаемое обеспечивает выполнение ограничений и может быть построено несколькими способами. Согласно первому из них данный компонент конструируемой энергетической функции имеет вид

$$E_{\Psi}^0(u) = \frac{A}{2} \sum_{j=1}^n \sum_{i=1}^n \sum_{n \neq i} u_{ji} u_{jn} + \frac{B}{2} \sum_{i=1}^n \sum_{j=1}^n \sum_{m \neq j} u_{ji} u_{mi} + \frac{C}{2} \left( \sum_{j=1}^n \sum_{i=1}^n u_{ji} - n \right)^2, \quad (10)$$

где  $A$ ,  $B$  и  $C$  - положительные константы. Первое слагаемое принимает минимальное и равное нулю значение лишь в том случае, если каждая строка матрицы  $\{u_{ij}\}$  содержит не более одной единицы, второе слагаемое принимает минимальное нулевое значение, если каждый столбец данной матрицы содержит не более одной единицы, наконец, третье слагаемое принимает минимальное нулевое значение, если во всей матрице  $\{u_{ij}\}$  содержится ровно  $n$  единиц.

Построенная функция  $E_{\Psi}^0(u)$  достигает своего минимума во всех состояниях, удовлетворяющих совокупности ограничений (4)-(6) и представляющих собой план назначений.

Согласно второму способу построения данного компонента конструируемой энергетической функции будем иметь

$$E_{\Psi}^0(u) = \frac{A}{2} \sum_{j=1}^n \left( \sum_{i=1}^n u_{ji} - 1 \right)^2 + \frac{B}{2} \sum_{i=1}^n \left( \sum_{j=1}^n u_{ji} - 1 \right)^2, \quad (11)$$

где первое слагаемое принимает минимальное нулевое значение только в том случае если в любой строке матрицы  $\{u_{ij}\}$  будет ровно один возбужденный нейрон, а второе - если в любом столбце этой матрицы будет ровно один возбужденный нейрон.

В целом данная функция принимает минимальное нулевое значение только на состояниях, удовлетворяющих ограничениям (4)-(6) и представляющих собой планы назначений.

Суммируя функцию (9) с функцией (10) или (11), сконструируем энергетическую функцию в завершённом виде

$$E^0(u) = \frac{A}{2} \sum_{j=1}^n \sum_{i=1}^n \sum_{n \neq i} u_{ji} u_{jn} + \frac{B}{2} \sum_{i=1}^n \sum_{j=1}^n \sum_{m \neq j} u_{ji} u_{mi} + \frac{C}{2} \left( \sum_{j=1}^n \sum_{i=1}^n u_{ji} - n \right)^2 - F \sum_{j=1}^n \sum_{i=1}^n u_{ji} r_{ji}, \quad (12)$$

или

$$E^0(u) = \frac{A}{2} \sum_{j=1}^n \left( \sum_{i=1}^n u_{ji} - 1 \right)^2 + \frac{B}{2} \sum_{i=1}^n \left( \sum_{j=1}^n u_{ji} - 1 \right)^2 - F \sum_{j=1}^n \sum_{i=1}^n u_{ji} r_{ji} \quad (13)$$

Определим параметры сети, сопоставив одну из полученных функций с энергетической функцией, записанной в общем виде

$$E(u, T, I) = -\frac{1}{2} \sum_{j=1}^n \sum_{i=1}^n \sum_{m=1}^n \sum_{n=1}^n T_{jimm} u_{ji} u_{mn} + \sum_{j=1}^n \sum_{i=1}^n u_{ji} I_{ji}, \quad (14)$$

где  $T_{jimm}$  - коэффициент связи между входом  $ij$ -го нейрона и выходом  $mn$ -го;

$I_{ji}$  - смещение  $ij$ -го нейрона.

В данном выражении для энергетической функции сети умышленно опущен временной параметр в связи с тем, что при определении синапсов и внешних смещений он не играет какой-либо существенной роли как для сетей с дискретным временем, так и для сетей с непрерывным временем. Более того, данным выражением мы будем пользоваться при определении параметров синтезируемых сетей, как с дискретными, так и с непрерывными состояниями. Основанием для этого служит тот факт, что энергетические функции сетей с дискретными и с непрерывными состояниями отличаются только наличием у последних интегрального слагаемого, которое ни от значений синапсов, ни от внешних смещений в явном виде не зависит.



Для того, что бы определить параметры сети в соответствии с построенной энергетической функцией (12), приведем выражение для этой функции к виду (14):

$$E^0(u) = \frac{A}{2} \sum_{j=1}^n \sum_{i=1}^n \sum_{n \neq i} u_{ji} u_{jn} + \frac{B}{2} \sum_{i=1}^n \sum_{j=1}^n \sum_{m \neq j} u_{ji} u_{mi} + \frac{C}{2} \sum_{j=1}^n \sum_{i=1}^n \sum_{m=1}^n \sum_{n=1}^n u_{ji} u_{mn} - Cn \sum_{j=1}^n \sum_{i=1}^n u_{ji} - F \sum_{j=1}^n \sum_{i=1}^n u_{ji} r_{ji} + \frac{C}{2} n^2 \quad (15)$$

и приравняем коэффициенты при линейных и квадратичных членах последнего выражения и энергии (14). Последнее слагаемое из рассмотрения можно исключить, так как оно не зависит от состояния сети.

Сопоставление линейных членов позволит определить значения внешних смещений, а сопоставление квадратичных членов позволит определить синаптические связи между нейронами.

Анализ первого слагаемого сконструированной энергетической функции свидетельствует о том, что любой нейрон сети должен иметь синаптические связи с коэффициентом  $-A$  со всеми нейронами одноименной с ним строки (условие  $m = i$ ) кроме самого рассматриваемого нейрона (условие  $n \neq j$ ).

Второе слагаемое диктует наличие связей с коэффициентом  $-B$  между нейронами одноименного столбца (условие  $n = j$ ) кроме собственной обратной связи (условие  $m \neq i$ ).

Третье слагаемое свидетельствует о том, что все нейроны сети связаны друг с другом синапсами с коэффициентами  $-C$ . Воспользовавшись символом Кронекера  $d_{ji}$ , запишем результирующее выражение для синаптических связей сети в виде

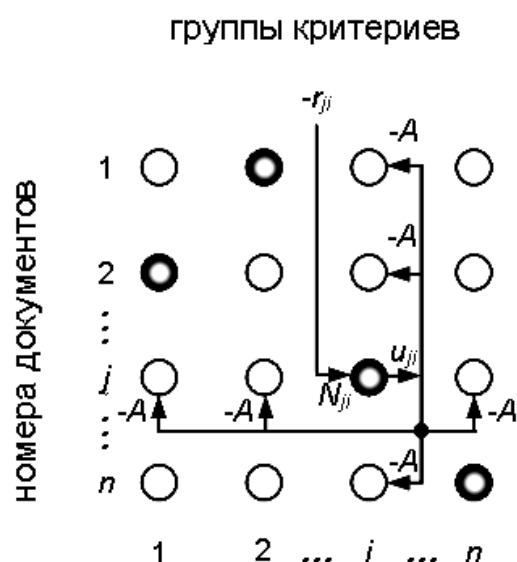
$$T_{jimm} = -A d_{jm} (1 - d_{in}) - B d_{in} (1 - d_{jm}) - C = -A d_{jm} - B d_{in} + (A + B) d_{jm} d_{in} - C, \quad i, j, m, n \in \overline{1, n}. \quad (16)$$

Анализ четвертого и пятого слагаемых сконструированной энергетической функции свидетельствует о том, что на все нейроны сети необходимо подавать внешние смещения в виде

$$I_{ji} = -Cn - F c_{ji}, \quad i, j \in \overline{1, n}. \quad (17)$$

Как правило, в практических задачах принимают  $F = 1$  и  $A = B$ , тогда все ненулевые связи имеют одинаковый вес, равный  $-A$ . Кроме того, анализируя выражения (16) и (17), можно заметить, что наличие глобальных связей с коэффициентом  $-C$  каждого нейрона с каждым в конечном состоянии сети, соответствующем некоторому плану назначений, обеспечивает подачу на любой нейрон со стороны всех других суммарного сигнала, равного  $-Cn$ , который компенсируется постоянным смещением  $-Cn$ . Следовательно, для упрощения структуры синапсов сети глобальными связями с весом  $-C$  и частью смещения  $-Cn$  в первом приближении можно пренебречь. В этом случае упрощенную структуру сети для синтеза оптимального плана оценивания документов путем решения задачи о назначениях можно представить в виде, изображенном на рисунке 3.

Искомая модель нейронной сети с БФВ нейронов содержит матрицу из  $n \times n$  нейронов, на каждый из которых подается внешнее смещение, равное соответствующей производительности  $I_{ji} = -r_{ji}$ , а выходной сигнал любого нейрона  $u_{ji}$ , с коэффициентом  $-A$  подается на входы всех нейронов одноименных с ним строки и столбца.



**Рисунок 3 - Динамическая нейронная сеть релаксирующая к своему энергетическому минимуму, интерпретированного в качестве максимального суммарного значения релевантности в комбинаторной группе критериев**

Для того чтобы не загромождать рисунок, на нем изображены связи и смещение только одного нейрона  $N_{ji}$ . Пример плана назначений представлен совокупностью возбужденных (выделенных жирным) нейронов.

Другой вариант параметров сети для оптимального плана оценивания документов можно получить, используя сконструированную энергетическую функцию в виде (13).

Аналогично проведенной выше процедуре приведем данное выражение к виду (14):

$$\begin{aligned}
 E^0(u) = & \frac{A}{2} \sum_{j=1}^n \sum_{i=1}^n \sum_{n=1}^n u_{ji} u_{jn} + \frac{B}{2} \sum_{i=1}^n \sum_{j=1}^n \sum_{m=1}^n u_{ji} u_{mi} - A \sum_{j=1}^n \sum_{i=1}^n u_{ji} \\
 & - B \sum_{j=1}^n \sum_{i=1}^n u_{ji} - F \sum_{j=1}^n \sum_{i=1}^n u_{ji} r_{ji} + \frac{n}{2}(A + B),
 \end{aligned}
 \tag{18}$$

и сопоставим коэффициенты при линейных и квадратичных членах, отбросив последнее слагаемое. В результате получим:

$$\begin{aligned} T_{jmn} &= -A d_{jm} - d_{in}, \\ I_{ji} &= -(A + B) - F r_{ji}, \quad i, j, m, n \in \overline{1, n}. \end{aligned} \quad (19)$$

Кроме рассмотренного, можно использовать различные комбинации функций (10) и (11) для конструирования энергетической функции  $E^o(u)$ . В результате будем получать различные варианты параметров нейросети.

При построении нейроподобной сети с непрерывными состояниями необходимо обеспечить условия нахождения точек покоя в углах  $n$ -мерного куба ее пространства состояний.

Другим способом обеспечения строгой бинарности выходных сигналов нейронов в устойчивых состояниях является добавление к конструируемой энергетической функции дополнительного слагаемого, достигающего минимального значения на состояниях сети, в которых выходные сигналы нейронов принимают значения 0 или 1.

Примером такой функции может служить следующая [6-8]:

$$E_b^0(u) = \frac{G}{2} \sum_{j=1}^n \sum_{i=1}^n u_{ji} (1 - u_{ji}) = \frac{G}{2} \sum_{j=1}^n \sum_{i=1}^n u_{ji} - \frac{G}{2} \sum_{j=1}^n \sum_{i=1}^n u_{ji}^2, \quad (20)$$

где  $G > 0$  – константа. Добавив данное слагаемое к ранее построенной энергетической функции, например в виде (12), получим

$$\begin{aligned} E^0(u) &= \frac{A}{2} \sum_{j=1}^n \sum_{i=1}^n \sum_{n \neq i} u_{ji} u_{jn} + \frac{B}{2} \sum_{i=1}^n \sum_{j=1}^n \sum_{m \neq i} u_{ji} u_{mi} + \frac{C}{2} \sum_{j=1}^n \sum_{i=1}^n \sum_{m=1}^n \sum_{n=1}^n u_{ji} u_{mn} \\ &- \frac{G}{2} \sum_{j=1}^n \sum_{i=1}^n u_{ji}^2 - Cn \sum_{j=1}^n \sum_{i=1}^n u_{ji} - F \sum_{j=1}^n \sum_{i=1}^n u_{ji} r_{ji} + \frac{C}{2} n_i^2, \end{aligned} \quad (21)$$

откуда определим следующие параметры нейросети:

$$\begin{aligned} T_{jmn} &= -A d_{jm} (1 - d_{in}) - B d_{in} (1 - d_{jm}) - C + G d_{jm} d_{in} = \\ &- A d_{jm} - B d_{in} + (A + B + G) d_{jm} d_{in} - C, \\ I_{ji} &= \frac{G}{2} - Cn - F r_{ji}, \quad i, j, m, n \in \overline{1, n}. \end{aligned} \quad (22)$$

Объединив выражения (20) и (14), можно сконструировать энергетическую функцию в виде:

$$E^0(u) = \frac{A}{2} \sum_{j=1}^n \sum_{i=1}^n \sum_{n=1}^n u_{ji} u_{jn} + \frac{B}{2} \sum_{i=1}^n \sum_{j=1}^n \sum_{m=1}^n u_{ji} u_{mi} - \frac{G}{2} \sum_{j=1}^n \sum_{i=1}^n u_{ji}^2 + \frac{G}{2} \sum_{j=1}^n \sum_{i=1}^n u_{ji} -$$

$$A \sum_{j=1}^n \sum_{i=1}^n u_{ji} - B \sum_{j=1}^n \sum_{i=1}^n u_{ji} - F \sum_{j=1}^n \sum_{i=1}^n u_{ji} r_{ji} + \frac{n}{2}(A + B), \quad (23)$$

откуда искомые параметры сети определяются следующим образом:

$$T_{jimm} = -A d_{jm} - B d_{in} + G d_{jm} d_{in},$$

$$I_{ji} = -\left(A + B - \frac{G}{2}\right) - F r_{ji}, \quad i, j, m, n \in \overline{1, n}. \quad (24)$$

Используя другие выражения для определения параметров сети, решающей задачу о назначениях, можно получить семейство нейросетей для решения задачи сортировки данных.

### Литература

- 1.1. Галушкин А.И. Теория нейронных сетей. – М.: ИПРЖР, 2000. – 415с.
- 1.2. Щербаков М.А. Искусственные нейронные сети. Конспект лекций.- Пенза: ПГТУ, 1996.-45с.
- 1.3. Осовский С. Нейронные сети для обработки информации /Пер с польского.- М.: Финансы и статистика, 2002.-344с.
- 1.4. Калан Р. Основные концепции нейронных сетей /Пер с англ.- М.: "Вильямс", 2001. – 288с.
- 1.5. Терехов С.А. Лекции по теории и приложениям искусственных нейронных сетей. <http://alife.narod.ru/lectures>.
- 1.6. Головкин В.А. Нейронные сети: обучение, организация и применение. Кн.4. – М.:ИПРЖР, 2001.-256с.
- 1.7. Lectons: Prof. Y. P. ZAYCHENKO. [www.i2.com.ua](http://www.i2.com.ua).
- 1.8. Ежов А.А., Шумский С.А. Нейрокомпьютеринг и его применения в экономике и бизнесе (серия "Учебники экономико-аналитического института МИФИ" под ред. проф. В.В. Харитонов).-М.: МИФИ, 1998.- 224с.