

# Прогнозирование временных рядов с помощью искусственных нейронных сетей и регрессионных моделей на примере прогнозирования котировок валютных пар

Крючин О.В. ([kryuchov@gmail.com](mailto:kryuchov@gmail.com)), Козадаев А.С., Дудаков В.П.

Тамбовский государственный университет им. Г.Р. Державина

## Введение

Прогнозирование временных рядов является важной научно-технической проблемой, т.к. позволяет предсказать поведение различных факторов в экологических, экономических, социальных и иных системах. Таким образом, основной целью любого прогнозирования является создание некой «машины времени», которая позволяет заглянуть в будущее и оценить тенденции в изменениях того или иного фактора. Такая «машина времени» в большинстве случаев базируется на методах математического моделирования, в частности на построении модельной авторегрессии, скользящей по временному ряду, и позволяющей осуществлять экстраполирование на несколько шагов вперед [3, 4]. Качество прогноза в таком случае зависит от наличия предыстории изменяемого фактора, погрешностей измерения рассматриваемой величина, глубины памяти (т.е. числа одновременно учтенных членов временного ряда). Указанные алгоритмы широко используются в настоящее время в экологии, социальных науках и экономике [6].

В настоящее время активно развиваются системы искусственного интеллекта, базируемые на использовании аппарата искусственных нейронных сетей (ИНС). В частности, с их помощью решается широкий круг проблем: построение моделей объектов при их сильной зашумленности, недостатке информации, распознавание образов, кластеризация и т.д. [5, 13, 14]. Имеются попытки использования ИНС и в задачах прогнозирования.

Целью данной работы является сравнение прогнозирования временных рядов средствами ИНС и регрессионных моделей. В качестве временного ряда используются котировки валютных пар.

## Временные ряды

Для моделирования ВР в настоящий момент существует множество методов, что само по себе ставит проблему выбора модели ВР для ЭВМ. Чем больше методов решения задачи, тем сложнее выбрать один из них, поскольку каждый может иметь свои положительные качества (простота программирования, малая загрузка памяти, высокая скорость выполнения для ЭВМ, простота оценки погрешности, применимость к широкому кругу задач) [1]. Из практических соображений обычно для моделирования из всего существующего многообразия выбирается модель возможно более простой формы [12].

В задачах прогнозирования ВР, в отличие от анализа случайных выборок, данные рассматриваются как последовательность измерений, упорядоченных в неслучайные моменты времени, т.е. ВР включает два обязательных элемента – время и конкретное значение показателя (уровень ряда). Это отчасти позволяет идентифицировать и, более или менее, формально описать модель ряда, а соответственно использовать для задач анализа ВР методы математической статистики [3].

Для построения временного ряда должны быть выполнены следующие условия:

- периодизация развития образуется расчленением во времени на однородные этапы, в пределах которых показатель подчиняется общему закону развития;
- значения ВР должны быть сопоставимы по всем признакам по которым

осуществляется его формирование, например по территории, кругу охватываемых явлений, единицам измерения, ценам, методологии расчёта и т.д.;

- периоды должны соответствовать интенсивности процессов;
- ВР должен быть полным, т.е. не допускаются пропуски, если они неизбежны, ряд дополняют условно-расчетными значениями.

Только в случае выполнения перечисленных условий можно говорить, что рассматриваемая последовательность является ВР и для нее применимы методы анализа и прогнозирования ВР.

### Регрессионные методы в задачах моделирования временных рядов

При построении математических моделей физических объектов и систем практически невозможно использование детерминированных моделей из-за участия ряда неизвестных факторов, вносящих свое влияние на поведение системы в целом. Модели ВР, необходимые для получения оптимального прогнозирования и регулирования являются стохастическими.

Разделяют стационарные и нестационарные стохастические модели ВР. Первый ряд моделей основан на предположении, что процесс находится в равновесии относительно среднего уровня, однако такие модели имеют малое применение ввиду редкой встречаемости процессов, которые они представляют. Поэтому многие ВР описываются как нестационарные (не имеющие естественного среднего значения). Наиболее распространенным классом как стационарных, так и нестационарных моделей являются модели авторегрессии и проинтегрированного скользящего среднего (АРПСС) [3].

Для методов математической статистики характерно разделение ВР на составляющие (периодические, постоянные, случайные). Исходя из этого, математическая модель ряда в целом выглядит следующим образом:

$$S(t) = f(t) + ke(t) \quad (1)$$

где  $S(t)$  — исследуемый сигнал,  $f(t)$  — полезный сигнал,  $k$  — уровень шума,  $e(t)$  — шум.

Для анализа и прогнозирования ВР необходимо выявить его полезные составляющие и максимально исключить случайные (погрешность аппаратуры, ошибки измерений и вычислений, субъективные ошибки участников процесса измерений и анализа данных).

В случае, когда распределение экспериментальных данных не поддается описанию каких либо явных закономерностей может оказаться эффективным моделирование ВР случайным процессом.

В качестве базовой модели обычно принимается нормальный белый шум [2]  $a(t)$  — случайный процесс, значения которого в различные моменты времени статистически независимы, а значения в любой фиксированный момент времени распределены одинаково по нормальному закону:

$$p(a) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{a^2}{2\sigma^2}} \quad (2)$$

где  $\sigma^2$  — дисперсия.

Для полного описания нормального случайного процесса (т.е. процесса, для которого одномерное распределение в любой момент времени является нормальным) достаточно задать только его среднее значение, дисперсию и автокорреляционную функцию. Поскольку значения автокорреляционной функции при любой ненулевой задержке равны нулю, то для построения модели в виде нормального белого шума по данным нужно оценить только его дисперсию и среднее значение [2], в формуле (2) среднее значение принято равным нулю.

Однако, свойства исследуемого объекта могут не соответствовать гипотезе о том, что это нормальный белый шум (например в случае, когда автокорреляции могут быть отличны

от нуля для ненулевых задержек). В этом случае можно рассматривать процесс как нормальный белый шум, преобразованный линейным фильтром (при подаче на вход линейного фильтра нормального белого шума выходной сигнал может быть некоррелированным [2]).

Значения нормального белого шума, преобразованного линейным фильтром, для любого момента времени определяется выражением

$$v_n = a_n + \sum_{i=1}^{\infty} \psi_i a_{n-i} \tag{3}$$

где веса  $\psi_i$  должны удовлетворять условию  $\sum_{i=1}^{\infty} \psi_i \leq const$ , чтобы процесс  $v_n$  был стационарным (всевозможные конечномерные распределения не меняются при сдвиге времени).

Для построения модели вида (1.3) необходимо оценить по исходным данным дисперсию белого шума и веса  $\psi_i$ . Вычислить бесконечное количество весов не представляется возможным, однако, как правило, значения  $\psi_i$  быстро убывают со временем (номера  $i$ ) и на практике достаточно ограничиться моделью с конечным числом весов  $q$ . Таким образом можно получить модель скользящего среднего порядка  $q$ .

Модель скользящего среднего порядка  $q$  —  $MA(q)$  (от английского *Moving Average* — скользящее среднее):

$$v_n = a_n - \sum_{i=1}^q \theta_i a_{n-i} \tag{4}$$

Модель (4) содержит  $q + 1$  параметров ( $\theta_1, \theta_2, \dots, \theta_q$ ), значения которых нужно оценить по ВР.

Общее выражение (3) можно преобразовать последовательным исключением из него  $a_{n-1}, a_{n-2}$  и т.д., для чего сначала необходимо выразить значение шума  $a_{n-1}$  через  $v_1$  и предыдущие значения  $a$  по формуле  $a_{n-1} = v_{n-1} - \sum_{i=1}^{\infty} \psi_i a_{n-1-i}$ , затем подставить полученное выражение в (3), исключив таким образом,  $a_{n-1}$ , далее аналогично последовательно исключаются  $a_{n-1}$  и т.д., пока модель (3) не будет записана эквивалентно в виде:

$$v_n = a_n + \sum_{i=1}^{\infty} \pi_i v_{n-i} \tag{5}$$

где веса  $\pi_i$  выражены через  $\psi_i$ .

Для того, чтобы перейти к конечному числу весов, необходимо ввести параметр  $p$  и, таким образом, перейти к модели, которая называется процессом авторегрессии порядка  $p$  —  $AR(p)$  (от английского *AutoRegressive* — авторегрессионная):

$$v_n = a_n + \sum_{i=1}^p \varphi_i v_{n-i} \tag{6}$$

Модель (6) содержит  $p + 1$  параметров ( $\varphi_1, \varphi_2, \dots, \varphi_p$ ), значения которых нужно оценить по ВР (причем значения параметров должны удовлетворять определенным соотношениям [Бокс], чтобы процесс был стационарным).

В случае, когда наблюдаемый ВР генерируется процессом авторегрессии порядка 1, при попытке описать его процессом скользящего среднего потребуется модель (4) с бесконечным (или очень большим) числом параметров  $\theta_i$ . При оценке значений очень большого числа параметров теряется надежность. В данном случае это обязательно приведет к снижению эффективности модели. В случае же, когда ряд генерируется процессом

скользящего среднего порядка 1, то для его описания потребовался бы процесс авторегрессии очень высокого порядка [2]. Поэтому наиболее эффективный подход состоит в том, чтобы объединить модели (4) и (6). Таким образом, будет получена модель авторегрессии и скользящего среднего порядка  $(p, q)$  —  $ARMA(p, q)$  (от объединения *AutoRegressive* и *Moving Average*):

$$v_n = a_n + \sum_{i=1}^p \varphi_i v_{n-i} - \sum_{i=1}^q \Theta_i a_{n-i} \quad (7)$$

Модель (7) содержит  $p + q + 1$  параметров.

Если наблюдаемый ряд  $v(t)$  имеет признаки нестационарности, то стационарный процесс (7) не может быть адекватной моделью. Однако в таком случае может оказаться стационарной некоторая разность наблюдаемого процесса порядка  $d$ :  $w_n = \nabla^d v_n$ , где  $\nabla v_n = v_n - v_{n-1}$  — первая разность (аналог дифференцирования), а  $\nabla^d$  означает последовательное применение  $d$  раз оператора  $\nabla$  [2, 3, 4]. Таким образом, приходим к модели авторегрессии и проинтегрированного скользящего среднего порядка  $(p, d, q)$  —  $ARIMA(p, d, q)$  (от английского — *AutoRegressive Integrated Moving Average*).

Модель авторегрессии и проинтегрированного скользящего среднего с параметрами  $ARIMA(p, d, q)$  будет выглядеть следующим образом:

$$w_n = a_n + \mu + \sum_{i=1}^p \varphi_i w_{n-i} - \sum_{i=1}^q \Theta_i a_{n-i} \quad (8)$$

$$w_n = \nabla^d v_n \quad (9)$$

Общим недостатком рассмотренных моделей является сложность выбора типа модели и подбора (определения) ее параметров, что существенно увеличивает субъективный вклад участников процесса анализа и прогнозирования ВР. Таким образом, результат анализа и прогнозирования ВР зависит как от квалификации аналитика в предметной отрасли, так и от его квалификации в методах анализа.

### Искусственные нейронные сети в задачах моделирования временных рядов

ИНС являются совокупностью элементарных обработчиков — искусственных нейронов, связанных между собой синаптическими соединениями. Сеть обрабатывает входную информацию и в процессе обучения формирует выходные сигналы [5].

При обработке данных, используя аппарат ИНС, подразумевается применение его как самостоятельного компонента системы управления или принятия решений. Применение аппарата ИНС возможно для решения различных задач (аппроксимация и интерполяция; распознавание и классификация образов; ассоциация; сжатие информации; прогнозирование; управление), но в любом случае ИНС играет роль универсального аппроксиматора функции нескольких переменных [11].

Аппарат ИНС подразумевает минимальное участие аналитика в формировании модели, так как способность к обучению характерна для всех нейросетевых моделей, а алгоритмы обучения адаптируют (подстраивают) весовые коэффициенты в соответствии с структурой данных, представленных для обучения.

При моделировании изолированных ВР с помощью ИНС-моделей допускается трансформация исходных данных для увеличения количества входных нейронов и, соответственно, увеличения прогностической способности [6, 9].

Моделирование ВР с использованием аппарата ИНС заключается в формировании ИНС определенной структуры, описывающей поведение исследуемой системы в моменты времени, а прогнозирование заключается в предсказании будущего поведения системы по предыстории. Для обучения таких ИНС изменение весовых коэффициентов происходит на

основании изменения фактической погрешности прогнозирования на итерациях.

Определенной сложностью в построении ИНС-модели является формирование обучающей выборки, которая должна выполнять требования к полноте (выборка не должна содержать пропуски, должна содержать все допустимые примеры исследуемого диапазона) и непротиворечивости (выборка не должна содержать противоречивых примеров). Для проверки обучающей выборки на соответствие требованиям необходима оценка их качества, непосредственно определение понятий качества выборки, критериев оценки качества и разработки математического аппарата для их однозначной оценки [7]. Для повышения прогностической способности допустимы (иногда необходимы) дополнительные преобразования, такие как методы фильтрации и восстановления данных (например, вейвлет-преобразование) [10].

Существует множество нейросетевых структур, различающихся количеством и расположением нейронов и синаптических связей. Наиболее известной структурой является многослойный персептрон. Многослойный персептрон (*MLP*) — это полносвязная модель без обратных связей. Количество слоев и нейронов в них обычно обусловлено постановкой задачи и вычислительными способностями ЭВМ. Количество весовых коэффициентов (синаптических связей) в данной структуре можно вычислить по формуле (10).

$$l_w = \sum_{i=1}^{N_L-1} \hat{N}_i \hat{N}_{i+1} \quad (10)$$

где  $N_L$  — количество слоев в ИНС,  $\hat{N}_i$  — количество нейронов на  $i$ -ом слое.

На рисунке 1 изображен персептрон, имеющий один скрытый слой с 7 нейронами, входной — с 3 нейронами и выходной, содержащий 1 нейрон. Количество весовых коэффициентов, в данном случае равно 28.

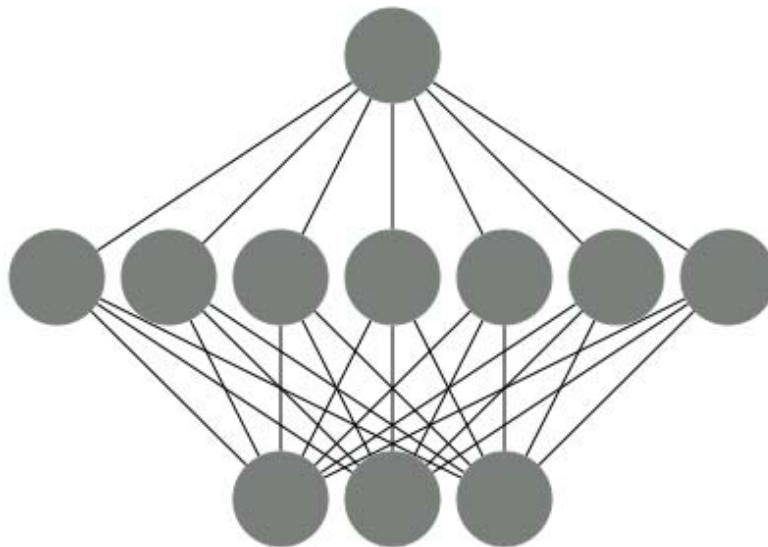


Рисунок 1. Многослойный персептрон, содержащий входной, выходной и скрытый слой.

В настоящее время существует множество методов подбора весовых коэффициентов. В данной работе использован один из градиентных методов *QuickProp*, разработанный Фальманом [15]. Особенностью градиентных алгоритмов является вычисление градиента по

весовым коэффициентам  $\nabla \varepsilon = \left( \frac{\partial \varepsilon(w)}{\partial w_0}, \frac{\partial \varepsilon(w)}{\partial w_1}, \dots, \frac{\partial \varepsilon(w)}{\partial w_{M-1}} \right)$  и изменение значений весовых

коэффициентов  $\vec{w}$  в противоположном направлении [8, 11]. В алгоритме *QuickProp* изменение  $i$ -ого весового коэффициента на  $I$ -ой итерации производится согласно правилу:

$$\Delta w_i^{(I)} = -s \times \left( \nabla \varepsilon_i^{(I-1)} + c_w \times w_i^{(I-1)} \right) + q_i^{(I)} \times \Delta w_i^{(I-1)} \quad (11)$$

где  $c_w$  - коэффициент минимизации значений весовых коэффициентов,  $q_i^{(t)}$  - коэффициент фактора момента. Особенность данного метода заключается в наличии двух слагаемых — минимизатора значений весовых коэффициентов ( $s \times c_w \times w_i^{(t-1)}$ ) и фактора момента ( $q_i^{(t)} \times \Delta w_i^{(t-1)}$ ). Коэффициент минимизации  $c_w$  обычно принимает значение  $10^{-4}$  и служит для ослабления весовых связей (вплоть до полного разрыва), а фактор момента необходим для адаптации алгоритма к текущим результатам обучения. Коэффициент  $q_i$  уникален для каждого весового коэффициента и вычисляется в два этапа. На первом определяется величина

$$\hat{q}_i = \frac{\nabla \varepsilon_i^{(t-1)}}{\nabla \varepsilon_i^{(t-2)} - \nabla \varepsilon_i^{(t-1)}} \quad (12),$$

а на втором коэффициент момента принимает значение минимальное значение из  $\hat{q}_i$  и  $q_{max}$ . В качестве значения  $q_{max}$  Фальманом предложено 1.75 [15].

Также существует модифицированная форма алгоритма, заключающаяся в уменьшении числа управляющих параметров без потери эффективности [16]. В модифицированном алгоритме формула (4) заменяется на следующую:

$$\Delta w_i^{(t)} = \begin{cases} q_i^{(t)} \times \Delta w_i^{(t-1)} & , \Delta w_i^{(t-1)} \neq 0 \\ s \times \nabla \varepsilon_i^{(t-1)} & , \Delta w_i^{(t-1)} = 0 \end{cases} \quad (13)$$

Таким образом, в случае, когда элемент вектора градиента  $\nabla \varepsilon_i^{(t-1)}$  принимает нулевое значение, на следующей итерации значение соответствующего ему весового коэффициента вычисляется классическим градиентным методом [8].

### Прогнозирование котировок валютной пары

В качестве временного ряда были взяты котировки валютной пары *EUR/USD* (отношение евро к доллару), полученные в марте 2010 года. Значения котировок были взяты с минутным интервалом. Для прогнозирования временного ряда была использована структура изображенная на рисунке 1. Данная ИНС имеет 12 входных нейронов, 25 скрытых и 1 выходной. При формировании обучающей выборки был использован коэффициент разряжения 15. Поскольку количество строк обучающей выборки должно быть много больше количества весовых коэффициентов, то было взято 500 обучающих примеров. Таким образом, обучающая выборка покрывает недельный период и выглядит следующим образом

$$X = \begin{pmatrix} x(t_{7340}) & x(t_{7355}) & \dots & x(t_{7485}) \\ x(t_{7325}) & x(t_{7340}) & \dots & x(t_{7460}) \\ \vdots & \vdots & \ddots & \vdots \\ x(t_{15}) & x(t_{30}) & \dots & x(t_{180}) \\ x(t_0) & x(t_{15}) & \dots & x(t_{165}) \end{pmatrix},$$

$$D = \begin{pmatrix} x(t_{7500}) \\ x(t_{7485}) \\ \vdots \\ x(t_{195}) \\ x(t_{180}) \end{pmatrix}$$

Используемый многослойный персептрон содержит активационные функции вида  $y = \frac{1}{1 + e^{0.5x}}$ , выходные значения которой расположены в диапазоне (0;1), поэтому обучающая выборка была пронормирована в этом же диапазоне [7].

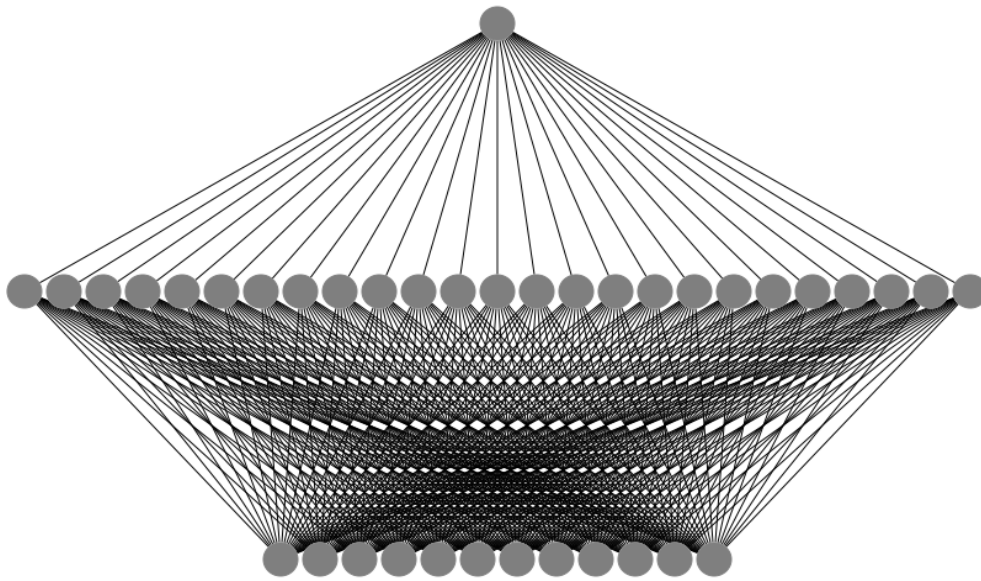


Рисунок 2. Структура многослойного персептрона, используемого для прогнозирования.

На рисунке 3 изображены временные ряды котировок валютной пары *EUR/USD* — реальный и спрогнозированный. Значения ошибки, вычисленные по формуле (14) составили 13% для ИНС и 17% для моделей, таких как *ARIMA*.

$$\varepsilon = \sum_{i=0}^{N-1} \left( \frac{|y_i - d_i|}{|d_i|} \right) \tag{14}$$

Здесь  $\bar{y}_i$  — спрогнозированные значения,  $\bar{d}_i$  — реальные значения.



Рисунок 3. Реальные (  $x$  ) и спрогнозированные (  $y$  ) значения котировок валютной пары *EUR/USD*.

### Выводы

Проведенные эксперименты показывают, что ИНС могут быть успешно использованы для прогнозирования временных рядов. Более того, в ряде случаев, они дают более точные результаты, чем традиционные модели прогнозирования.

### Литература

1. Бахвалов Н.С., Жидков Н.П., Кобельников Г.М. Численные методы – 4-е изд. – М.: БИНОМ. Лаборатория знаний, 2006. – 636 с.: ил.
2. Безрученко Б.П., Смирнов Д.А. Статистическое моделирование по временным рядам [Электронный ресурс] Саратов. отд-ние Ин-та радиотехники и электроники РАН. – Электрон. дан. – Саратов, 2000. – Режим доступа: <http://window.edu.ru/window/catalog/files/r29971/sgu025.pdf>, свободный. - Загл. с экрана.
3. Бокс Дж., Дженкинс Г. Анализ временных рядов. Прогноз и управление: М. Мир, 1974.
4. Бриллиндер Д. Временные ряды. -М.: Мир. 1980-536 с.
5. Горбань А., [gorban@cc.krascience.rssi.ru](mailto:gorban@cc.krascience.rssi.ru) Вычислительный центр СО РАН, Красноярск-36
6. Козадаев А.С., Арзамасцев А.А. Прогнозирование временных рядов с помощью аппарата искусственных нейронных сетей. Краткосрочный прогноз температуры воздуха // Вестн. Тамб. ун-та. Сер. Естеств. и техн. науки. – Тамбов, 2006. – Т. 11. – Вып. 3. – С. 299-304.
7. Козадаев А.С. Предварительная оценка качества обучающей выборки для искусственных нейронных сетей в задачах прогнозирования временных рядов // Вестн. Тамб. ун-та. Сер. Естеств. и техн. науки. – 2008. – Т. 13, вып. 1. – С. 99-100.
8. Крючин О.В. Разработка параллельных градиентных алгоритмов обучения искусственной нейронной сети // Электронный журнал "Исследовано в России", 096,



- стр. 1208-1221, 2009 г. // Режим доступа: <http://zhurnal.ape.relarn.ru/articles/2009/096.pdf>, свободный. – Загл. с экрана.
9. Метод скользящего окна // Глоссарий – BaseGroup Labs [Электронный ресурс]. – Электрон. дан. – Режим доступа: <http://www.basegroup.ru/glossary/definitions/windowing/>, свободный. – Загл. с экрана.
  10. Новиков Л.В. Основы вейвлет-анализа сигналов: учеб. пособие // СПб.: Модус+, 1999. – 152 с.: ил.
  11. Осовский С. Нейронные сети для обработки информации; пер. с пол. И.Д. Рудинского. – М.: Финансы и статистика, 2002. – 344 с.: ил.
  12. Спиди К., Браун Р., Гудвин Дж. Теория управления. – М.: Мир, 1973. – 248 с.
  13. Тюменцев Ю.В. Интеллектуальные автономные системы – вызов информационным технологиям.– Московский авиационный институт им. С.Орджоникидзе.
  14. Уоссермен Ф. Нейрокомпьютерная техника: Теория и практика / Перевод на русский язык, Ю. А. Зуев, В. А. Точенов, 1992.
  15. Fahlman S.E. An empirical study of learning speed in back-propagation networks. // Technical report. CMU-CS-88-162. Carnegie-Mellon University. 1988.
  16. Veith A.C., Holmes G.A., A modified quickprop algorithm // Neural Computation, 1991. - Vol.3 - Pp. 310-311.