

НЕЙРОСЕТЕВОЙ ПОДХОД К ПРОГНОЗИРОВАНИЮ ЦЕН НА ПОДЕРЖАННЫЕ АВТОМОБИЛИ

Скобцов Ю.А., Хмелевой С.В.

Донецкий национальный технический университет,
кафедра автоматизированных систем управления

E-mail: hmelevoy_sergey@ukr.net

Abstract

Skobtsov Y.A., Hmelevoi S.V. Neural network approach to prognostication of prices of the used cars. The used cars prices prognostication method by means of neural network is described. The input dates, net structure and training parameters are chosen. Prognostication results got as the result of neural-net model applying are analysed.

Введение

Прогнозирование в области экономики всегда было и будет одной из важнейших областей применения научных знаний. Одной из малоисследованных областей в настоящее время в стране является прогнозирование цен на подержанные автомобили.

Целью этой работы является создание метода прогнозирования цен на подержанные транспортные средства (ТС) на основе нейронных сетей.

Входные данные обучающей выборки

На основании данных, полученных в средствах массовой информации, были выбраны следующие факторы, которые значимо влияют на цену подержанных автомобилей:

1. Курс иностранных валют

Так как в выборке представлены практически исключительно русские автомобили, то из этой группы факторов уместно будет представить только: V1: курс рубля к гривне. Этот фактор взят из [8].

2. Инфляция

К этой группе относятся следующие факторы ([3],[4],[5]):

V2: Индекс инфляции (индекс потребительских цен),

V3: Индекс инфляции на непродовольственные товары.

3. Факторы развития экономики:

V4: Индексы цен производителей транспортного оборудования на Украине,

V5: Индексы цен машиностроения на Украине,

V6: Индексы цен производителей промышленной продукции,

V7: Изменение объемов производственного производства,

V8: Изменение номинального валового внутреннего продукта,

V9: Изменение реального валового внутреннего продукта,

V10: Экспорт товаров,

V11: Импорт товаров,

Эти показатели взяты из [3],[4],[5].

4. Таможенные пошлины

В настоящее время таможенные пошлины на Украине состоят из двух частей:

- Собственно таможенные пошлины. Их размер определен в законе Украины № 2371-III от 05.04.01.
- Акциз. Размер определен законом Украины №2134-III от 17.12.2000.

После анализа дополнений и изменений к этим документам на сайте Верховной рады Украины [9], выяснено, что со времени принятия вышеуказанных законов изменений в части налогообложения легковых автомобилей не произошло. Так как в базе данных автотранспортных средств хранится информация за 2002 – 2003 гг., можно сделать вывод, что за это время таможенные пошлины не изменялись, соответственно, влияния этот параметр на прогноз оказывать не будет.

5. Доходы населения

В этой группе приведено два фактора (из [3],[4],[5]):

V12: Среднемесячная номинальная заработная плата,

V13: Задолженность по выплата зарплат.

6. Сезонный фактор

Цены на автомобили подвержены сезонным колебаниям. В [6] приведены сезонные колебания на спрос на автомобили. Поскольку цена напрямую зависит от спроса, можно выделить фактор

V14: Спрос на автомобили.

Информация из [7] по сезонным колебаниям цен на автомобили не противоречит вышеуказанному источнику.

7. Предыдущие значения временного ряда цен

Также был добавлен такой фактор, как V15: **Предыдущая цена автомобиля.**

Значения этого факторы были взяты из базы данных автомобилей.

Выходные данные обучающей выборки

Имеется база данных по подержанным транспортным средствам. В ней хранится информация о ценах на автомобили в пределах Украины за последние полтора года, а также первичная информация об этих автомобилях.

Цены на автомобили определялись следующим образом: из средств массовой информации (газеты, журналы, Internet-сайты) отбирались объявления по продаже транспортных средств. Далее эта информация заносилась в базу данных. Затем отбирались все объявления для одной модели автомобиля, и с помощью регрессионного анализа строилась сглаженная зависимость цены от года выпуска. Этот процесс повторялся приблизительно раз в квартал. Таким образом, усредненные цены на подержанные транспортные средства в базе данных представлены за определенный период, например, за январь-апрель 2003г.

Из базы данных были взяты цены производителей стран СНГ, в основном Российских. Автомобилей этих производителей имеется на нашем рынке подержанных транспортных средств больше всего, поэтому выборка для этих ТС достаточно большая и статистика по ним получается с большой долей достоверности. Из этих цен была сделана выборка по отдельным годам выпуска транспортного средства (2-3 цены по разным годам для одного автомобиля). Эта информация фактически является выходной составляющей обучающей выборки.

Нейросетевой подход к прогнозированию.

Для решения задачи прогнозирования с использованием нейронных сетей использовалась обобщенная схема, представленная на рис.1.

В соответствии с этой схемой построен весь дальнейший процесс работы с сетью.

Предварительная обработка данных

Приведение данных к относительным значениям

Входные факторы представлялись в самом разном виде, – в процентах к предыдущему месяцу (V3, V14), в процентах нарастающей суммой к аналогичному периоду предыдущего года (V2, V4, V5, V6, V7, V9) или в натуральных измерителях (V1, V8, V10, V11, V12, V13). Так как использование относительных изменений значений временных рядов позволяет

принципиально повысить чувствительность нейросети к входным данным, а также для приведения всех данных в единый базис, все значения факторов (V1-V13) приводились сначала к процентному изменению за месяц. Поскольку в базе данных автомобилей значения цен рассчитывались не строго поквартально, данные входных факторов затем группировались в зависимости от периода представления цен в базе данных.

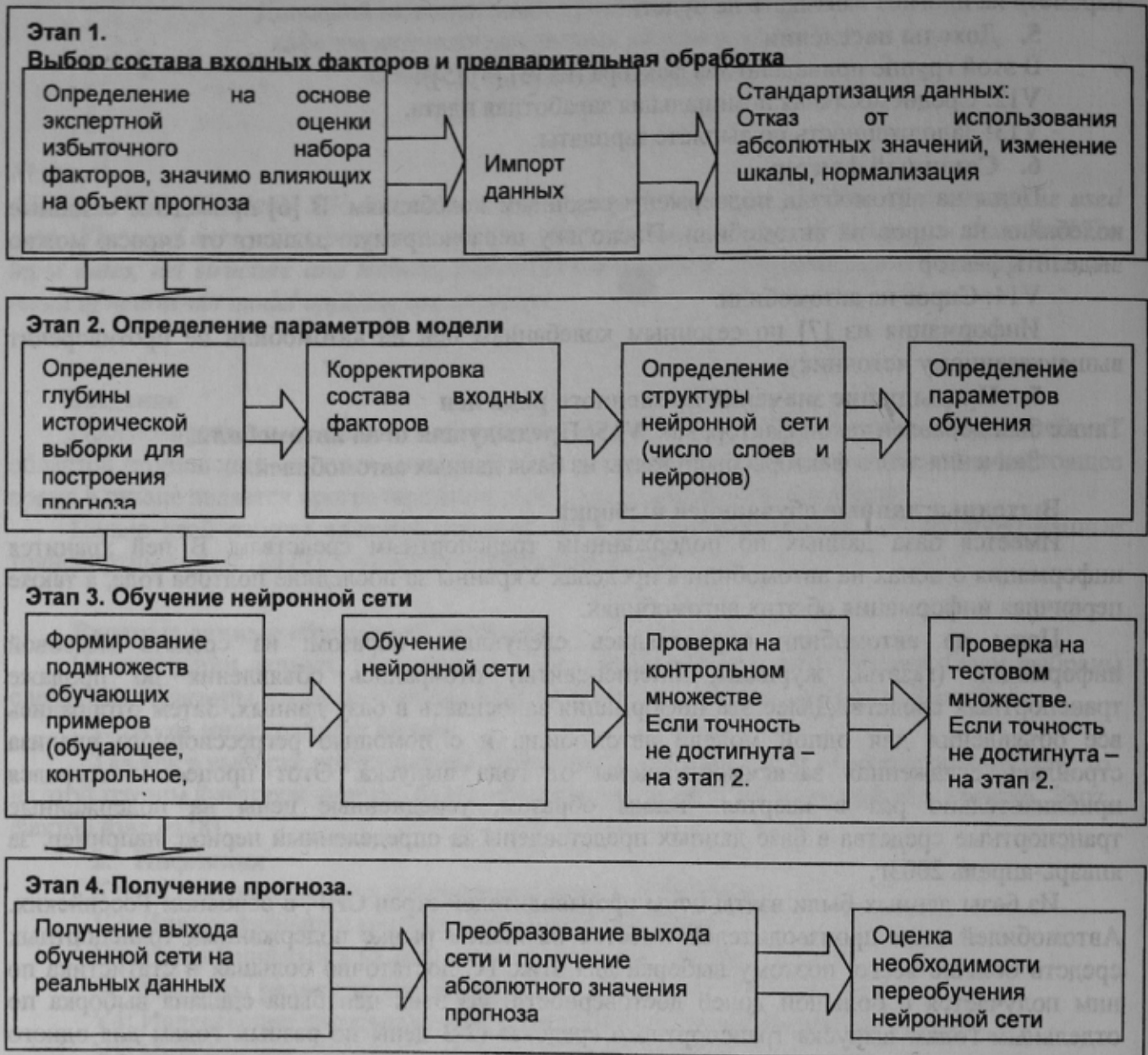


Рис.1. Обобщенная схема для решения задачи прогнозирования с использованием нейронных

Приведение фактора V15 (предыдущая цена автомобиля) к относительным значениям нежелательно, так как в процессе прогнозирования нейронной сетью строится зависимость цены ТС непосредственно от диапазона цен.

Нормировка данных

Нормировка данных позволяет привести разнородные данные к сопоставимым диапазонам, а также приблизить их распределение к равномерному, что повышает информативность обучающего примера. Из возможных вариантов нормировки был выбран следующий:

Линейная нормировка: сводит все значения к единому диапазону [-1,1] [2].

Таким образом, все значения входных факторов, как приведенные к относительной величине изменения с предыдущего периода, так и не приведенные, были нормированы. На вход нейронной сети подавались только значения в диапазоне $[-1,1]$.

Определение параметров нейронной сети

Определение глубины исторической выборки

При определении глубины исторической выборки, необходимой для построения прогноза, критическим фактором служила глубина выборки, из базы данных автомобилей. За полтора года было получено шесть измерений цен каждого автомобиля. Этого явно недостаточно для построения обучающей выборки по одному автомобилю. Поэтому для построения обучающей выборки использовались данные по нескольким транспортным средствам производства стран СНГ.

Корректировка состава входных факторов

Для корректировки (сокращения) состава входных факторов использовался корреляционный анализ. Был произведен анализ входных данных между собой, а также анализ выходных и входных данных.

В результате корреляционного анализа входных данных были выбраны факторы, наиболее связанные между собой.

Таблица 1. Факторы, наиболее коррелированные между собой.

Фактор 1	Фактор 2	Коефф. корреляции
V7: Изменение объемов производственного производства	V13: Задолженность по выплате зарплаты	-0,85
V8: Изменение номинального валового внутреннего продукта	V13: Задолженность по выплате зарплаты	-0,88
V9: Изменение реального валового внутреннего продукта	V13: Задолженность по выплате зарплаты	0,92

На основании этой информации было принято решение об изъятии фактора V13 из множества входных факторов.

Также был произведен корреляционный анализ выходных и входных данных. В результате этого анализа выяснены следующие особенности:

- чрезвычайно большой разброс для различных автомобилей. Для многих факторов максимальное значение коэффициента корреляции приближается к 1, а минимальный – к -1
- для многих факторов среднее по всем коэффициентам корреляции стремится к 0

Все это говорит о весьма слабой общей для всех автомобилей тенденции и больших расхождениях. Проведение корреляционного анализа для выходных данных также свидетельствует об этом.

В результате анализа связей между входными и выходными данными было принято решение об исключении из множества факторов фактор V10: экспорт товаров в связи с тем, что среднее значение коэффициента корреляции последнего равняется 0.003, а максимальное и минимальное значения = 0,99 и $-0,99$.

Выбор вида сети и параметров обучения

Согласно [12] для решения задач прогнозирования больше всего подходят следующие виды нейронных сетей: многослойный перцептрон (MLP-сеть) и радиально-базисная сеть (RBF-сеть).

Но в большинстве случаев при решении одной и той же задачи размер сети RBF будет превышать размер MLP [10]. Сеть RBF имеет в своем составе один скрытый уровень, а сеть MLP в общем случае – несколько. Однако в [11] показано, что любая функция может быть

аппроксимирована трехслойной сетью MLP с одним скрытым слоем с любой степенью точности. При этом применение большего количества уровней является нецелесообразным, так как при этом время обучения увеличивается, а точность прогноза уменьшается. Таким образом, задача выбора числа слоев для обоих типов нейросетей не стоит: как сеть RBF, так и сеть MLP должны иметь один скрытый слой.

Сети RBF позволяют использовать лишь гауссову функцию активации, либо ее модификации. В тоже время MLP обладает способностью работать с различными функциями, что дает более широкие возможности для экспериментирования в процессе построения нейросетевой модели.

Кроме того, сеть RBF не обладает способностью к экстраполяции данных при увеличении ширины диапазона значений входных данных, так как отклик сети быстро затухает при удалении от сформированных на обучающей выборке центров классов.

Недостатком сети MLP по сравнению с RBF является более высокая сложность обучения ввиду необходимости совместного обучения нескольких слоев нейронов.

Таким образом, в качестве базовой сети была выбрана MLP-сеть с 1 скрытым и 1 выходным слоем. Количество нейронов в выходном слое равняется количеству выходов – 1.

В качестве инструментальной среды использовался пакет Matlab6.0 (neural network toolbox). Был проведен эксперимент по выбору оптимального числа нейронов. Количество нейронов промежуточного слоя первоначально было равно 20. Их количество уменьшалось, пока величина погрешности не начала увеличиваться. В результате количество нейронов промежуточного слоя оказалось равным 2!!! Архитектура сети представлена на рис.2. Данный результат можно объяснить только одним: в обучающей выборке все значения факторов, кроме предыдущей цены на автомобиль, представлены одинаково для каждого автомобиля. Таким образом, обучающая выборка не обладает достаточным разнообразием и для получения результата не требуется нейросети большой вычислительной мощности. Активационная функция скрытого слоя – гиперболический тангенс, выходного слоя – линейная. Для обучения использовался метод обратного распространения ошибки *traingd*.

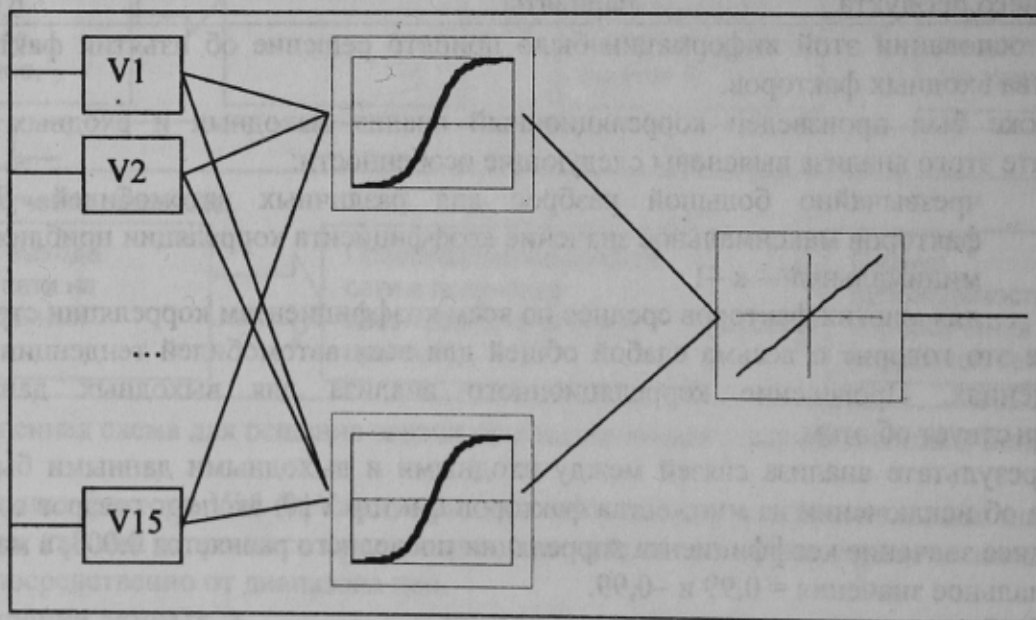


Рис.2. Структура нейронной сети.

Обучение нейронной сети и получение прогноза

Так как для каждого автомобиля имеется в наличии не более 6 точек из временного ряда его цены, невозможно формирование полноценных обучающих примеров на основании только одного автомобиля. Поэтому обучающие примеры формировались на основе цен на

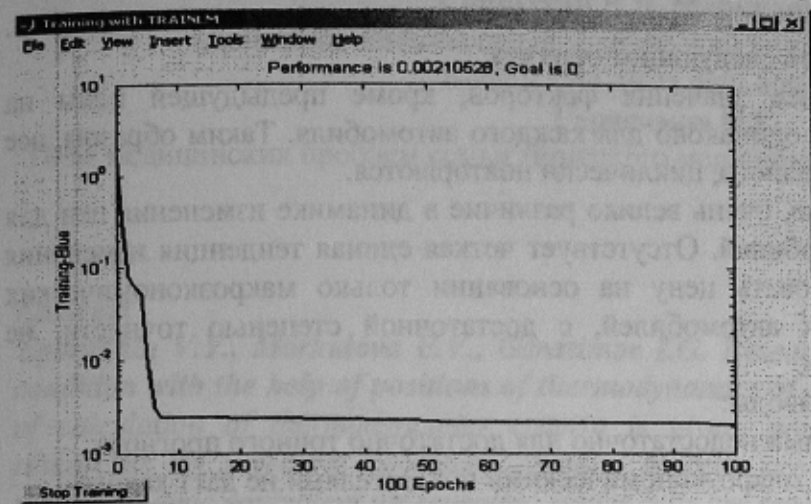


Рис.3 График обучения нейронной сети

несколько моделей автомобилей. Первоначально было отобрано 81 моделей автомобилей. Для каждого из них представлено 5 точек временного ряда. Таким образом, первоначальная длина общей обучающей последовательности равнялась $81 \cdot 5 = 405$ точек. Они были разделены на обучающую и тестовую последовательность в пропорции 2:1. Для тестовой последовательности считался средний процент отклонения прогноза от действительного значения а также максимальная величина ошибки прогноза. Были получены следующие результаты: средняя величина ошибки – порядка 5.5 – 6 %, максимальная ошибка – порядка 30% (так как первоначально веса нейронной сети иницировались случайным образом, после каждого запуска программы были получены различные результаты). После сокращения обучающей последовательности вдвое (величина обучающей выборки – 140 точек, проверочной - 65), за счет тех моделей, цены на которые максимально коррелировались со всеми остальными моделями, средняя величина ошибки не изменилась, а максимальная – уменьшилась до величины порядка 15%.

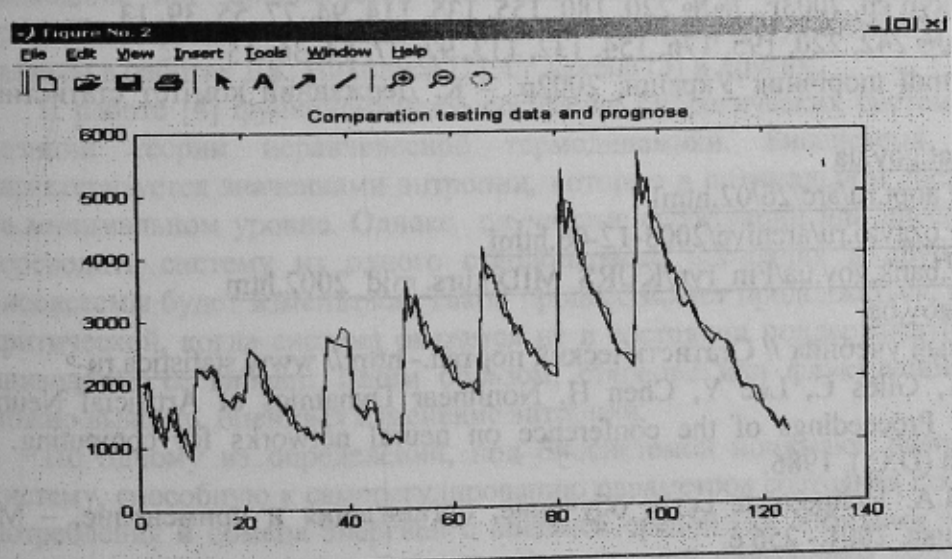


Рис.4 Сравнение тестовых и прогнозируемых выходных цен

Выводы

Полученные результаты можно трактовать неоднозначно. С одной стороны, нейронная сеть показывает результаты лучшие, чем классические методы прогнозирования, основанные на регрессионном анализе. С другой стороны, величина средней погрешности 6%, и 15% - максимальной погрешности от величины цены автомобиля является достаточно высокой. Данные результаты можно объяснить следующим образом:

1. В обучающей выборке все значения факторов, кроме предыдущей цены на автомобиль, представлены одинаково для каждого автомобиля. Таким образом, все значения факторов, кроме одного, циклически повторяются.
2. Как уже упоминалось выше, очень велико различие в динамике изменения цен для различных моделей автомобилей. Отсутствует четкая единая тенденция изменения цен. Поэтому спрогнозировать цену на основании только макроэкономических данных, единых для всех автомобилей, с достаточной степенью точности, не представляется возможным.

В результате можно сделать выводы:

1. Вышепредставленных данных недостаточно для достаточно точного прогноза.
2. Любое дополнение данных макроэкономическими показателями не даст какого-либо значительного эффекта, так как нет единой тенденции для всех автомобилей из данной выборки.
3. Возможно неточное определение самих цен автомобилей в базе данных.

Вопрос о развитии методики прогнозирования применительно к данному предмету исследования, таким образом, остается открытым. Необходим поиск дополнительных факторов, влияющих на цены автомобилей индивидуально, а также повышение точности определения исходных данных на цены автомобилей.

Литература

1. Д.-Э. Бэстенс, В.-М. Ван Ден Берг, Д. Вуд «Нейронные сети и финансовые рынки. Принятие решений в торговых операциях». – М.: Научное издательство ТВП, 1997. 235с.
2. Медведев В.С., Потемкин В.Г. Нейронные сети. – М.: ДИАЛОГ-МИФИ, 2002. – 496с.
3. Урядовий кур'єр, 2003г., №№ 220, 180, 155, 135, 114, 94, 77, 55, 39, 13
2002г., №№ 242, 220, 195, 176, 154, 132, 113, 92, 77, 54, 36, 15
4. Статистичний щорічник України, 2002р. – К. Державний комітет статистики України.
5. www.ukrstat.gov.ua
6. <http://www.aopt.ru/arc/26/02.html>
7. <http://www.ozavto.ru/archive/2003-12-06.html>
8. http://www.bank.gov.ua/Fin_ryn/KURS_MID/kurs_mid_2002.htm
9. www.rada.gov.ua
10. Он-лайнный учебник // Статистический портал.- [http:// www.statistica.ru](http://www.statistica.ru)
11. Maxwell T, Giles C, Lee Y, Chen H. Nonlinear Dynamics of Artificial Neural Systems // Proceedings of the conference on neural networks for computing. – Washington (D.C.), 1986.
12. Головкин В.А. Нейронные сети: обучение, организация и применение. – М.: Радиотехника, 2001.- 256 с.