УДК 004.8

МУРГА Н.А.

ПРИМЕНЕНИЕ НЕЧЁТКИХ НЕЙРОННЫХ СЕТЕЙ К ЗАДАЧЕ АНАЛИЗА ПОВЕДЕНИЯ КОТИРОВОК ВАЛЮТ

В статті приводяться результати порівняльного аналізу застосувань радиальної базисної нейронної мережі та нечіткої нейронної мережі, запропонованої автором у [1], до задачі прогнозувань денних котирувань валютних пар EUR/GBP, EUR/USD, USD/JPY, USD/CHF. В роботі також приводиться опис нейронної мережі [1] та метод її навчання.

Results of comparative analysis of application of radial-basis neural network and fuzzy neural network, was proposed by author in [1], for problem of forecasting daily currency EUR/GBP, EUR/USD, USD/JPY, USD/CHF quotes is proposed in the paper. The neural network [1] and its training method are also shown in the paper.

Введение

В работе [1] была представлена нечёткая нейросетевая архитектура, которой в качестве нечёткого логического вывода используется нечёткий вывод Такаги-Сугено ([2,3]). В консеквенте нечётких правил указанной нечёткой нейронной сети в качестве функциональных зависимостей входов сети от выходов используются тригонометрические полиномы ([2,3]). Обучение сети из работы [1] происходит не с целью минимизации единого для всей сети критерия качества, а для каждого правила нечёткой системы формируется своё подмножество множества точек обучающей выборки и каждое правило обучается отдельно, с целью оптимизации своего критерия качества. В работе [1] исследовалось применение сети к задаче прогноза значений котировок валютных пар на основании их предыстории. В качестве котировок были взяты котировки EUR/GBP, EUR/USD, USD/JPY, USD/CHF из [4].

Далее, с целью сокращения названия, метод обучения, предложенный в [1], и нейросетевая архитектура, рассматриваемая в [1], будет называться – Нечёткая нейросетевая архитектура с независимым обучением каждого правила – ННСА НОКП – и в ходе повествования будет использоваться эта аббревиатура.

Данная работа посвящена сравнительному анализу применения радиальной базисной нейронной сети и ННСА НОКП к той же задаче, что и в работе [1] при той же выборке данных [4]. Следует отметить, что описание радиальной базисной сети в данной работе приводиться не будет вследствие ограничений на объём статьи, однако можно найти её описание в работе [5].

Структура работы построена следующим образом. В первой части статьи, которая носит название «Нечёткая нейросетевая архитектура с независимым обучением каждого правила логического вывода», приводится общая методика обучения нейронных сетей, после чего описывается модификация общей методики, а также предлагаются к рассмотрению способы оценки качества работы обученных нейронных сетей. В этой же части описывается ННСА НОКП. Во второй части, которая носит название «Результаты применения ННСА НОКП и радиальной базисной сети к анализу котировок валют» представлен сравнительный анализ применения к вышеуказанной задаче ННСА НОКП и радиальной базисной нейронной сети. В данной части обосновывается адекватность проведения такого сравнительного анализа и излагаются результаты сравнения сетей по ряду параметров. Несмотря на то, что в статье рассматривалась только задача прогнозирования значений котировок валютных пар, целевая аудитория данной

работы – это специалисты в области анализа временных рядов, специалисты в области интеллектуальных систем, а также аналитики финансовых рынков.

Нечёткая нейросетевая архитектура с независимым обучением каждого правила логического вывода

Пусть будет выбран для определённости нечёткий вывод Такаги-Сугено ([2,3]). Рассуждения для данной системы нечёткого логического вывода легко переносятся на другие типы систем нечёткого логического вывода.

С целью оценки качества функционирования нейронной сети и качества её обучения, вся выборка данных, на которых будет функционировать сеть, разбивается на обучающую выборку и проверочную. На обучающей выборке происходит обучение сети на проверочной, соответственно, проверка способности сети обрабатывать вновь поступающие данные. При этом качество оценивается по определённому критерию. Обычно качество функционирования нейронной сети оценивается с помощью критерия ***RMSE***. Для дальнейших выкладок необходимо ввести слідующие обозначения: n – количество точек выборки данных, y_i – выход модели, который должен получаться при i-х входных значениях и это значение взято из выборки; $\overline{y}_i$ – значение, которое выдаёт система обработки данных выборки (в данном случае – нечёткая система) при тех же входных значениях. Согласно введённым обозначениям критерий ***RMSE*** определяется формулой (1):

$$\boldsymbol{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \overline{y}_i\right)^2}, \qquad (1)$$

Однако данный критерий является абсолютным критерием, а это означает, что он является чувствительным к масштабу данных. Критерием, который не чувствителен к масштабу данных, является относительны критерий, например ***MAPE***, который определяется формулой (2):

$$\boldsymbol{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|y_i - \overline{y}_i\right|}{y_i}. \qquad (2)$$

Обучение сети происходит с целью минимизации другого, упрощённого для выполнения расчётов критерия, например, приведённого в формуле (3):

$$E = \frac{1}{2}\sum_{i=1}^{n}\left(y_i - \overline{y}_i\right)^2. \qquad (3)$$

Приведённый критерий удобен тем, что функция E дифференцируема и при взятии производной от неё по определённому аргументу коэффициент будет равен 1.

Классический подход к обучению нейронных сетей подразумевает обучение сети в целом по всей обучающей выборке. Таким образом, ищется усреднённая по всем точкам выборки данных закономерность. Следует, однако, отметить, что динамика процесса в одной части пространства состояний может подчиняться одной закономерности, а в другой части (или частях) – другой. Способом разрешения данной проблемы, собственно, который и предлагался в работе [1], является разбиение обучающей выборки данных на подвыборки и поиск своих закономерностей в каждой из подвыборок.

Формально это описывается следующим образом. Пусть A – множество точек выборки данных. A_k, где $k = \overrightarrow{1,K}$ – подмножества множества A, которые удовлетворяют следующим условиям: $A = \bigcup_{k=1}^{K} A_k$ и $A_i \cap A_j = \varnothing$ при $i \neq j$. E_{A_k} – значение критерия (3) для подмножества A_k множества точек выборки. Подход предполагает обучение сети путём $\min E_{A_k}, \forall k = \overrightarrow{1,K}$ в отличие от классического подхода $\min E_A$.

Кроме того, данный метод настройки параметров нейронной сети позволяет получить следующее свойство: каждое частичное описание в консеквенте каждого нечёткого правила имеет свой собственный порядок (имеется в виду – порядок функции) и данный порядок настрашивается в ходе обучения сети.

Схема ННСА НОКП представлена на рис.1.

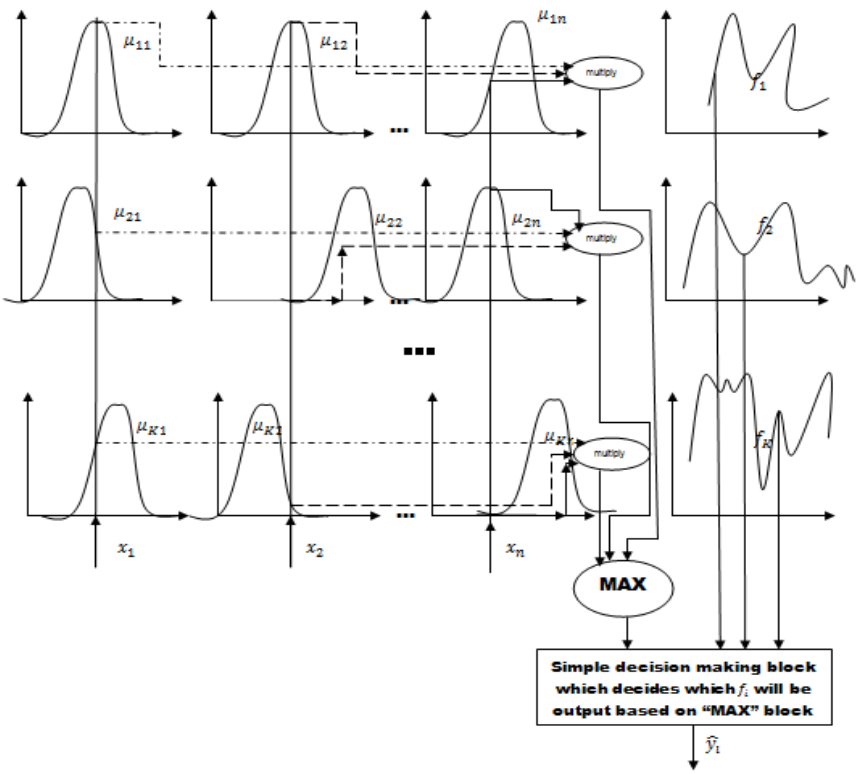

Рис. 1. Система нечёткого логического вывода Такаги-Сугено

Прежде чем переходить к объяснению рисунка, необходимо пояснить некоторые символы, которые в нём используются.

Выборка данных представляет собой матрицу размерности $L \times (n+1)$, где n – число входов сети (размерность пространства состояний), а L – количество точек выборки данных. Данная матрица имеет следующий вид:

$$A = \begin{pmatrix} x_{11} & \dots & x_{1n} & y_1 \\ \dots & \dots & \dots & \dots \\ x_{L1} & \dots & x_{Ln} & y_L \end{pmatrix}. \qquad (4)$$

Если будет подразумеваться некоторая (неважно какая) точка i выборки данных, то вместо $x_{i1}, x_{i2}, \dots, x_{in}, y_i$ будет записано $x_1, x_2, \dots, x_n, y$.

μ_{kj} – это функции принадлежности определённого признака j терму k_j.

В данной работе будут использованы функции принадлежности гауссовского вида, то есть μ_{kj} описывается выражением (5):

$$\mu_{kj}(x_j) = e^{-\frac{(x_j - c_{kj})^2}{2\sigma_{kj}^2}}. \qquad (5)$$

В качестве функций $f_k(x_1, \dots, x_n), k = \overrightarrow{1, K}$ будут использованы тригонометрические полиномы, которые описаны в [1] и [3], и восстанавливаться данные функциональные зависимости будут с помощью МНК по процедуре, описанной в [1]. Следует ещё раз отметить, что каждое частичное описание $f_{kj}(x_j)$, такое что $f_k(x_1, \dots, x_n) = \sum_{j=1}^{n} f_{kj}(x_j)$, имеет собственный порядок зависимости.

Таким образом, обучение сети состоит в нахождении таких c_{kj}, σ_{kj} и $f_k(x_1, \dots, x_n)$, которые бы наилучшим образом описывали динамику рассматриваемого процесса.

Процесс обучения предлагается строить следующим образом.

Первоначально происходит кластеризация той части выборки данных, которая соответствует входам сети. Для этого применяется сначала метод разностного группирования ([2,3]) для первоначально-

го задания центров кластеров, а после – нечёткий метод k-средних ([2,3]). Так определяются параметры c_{kj} функций принадлежностей, которые описываются формулой (5). Также нечёткий метод k-средних позволяет получить степень принадлежности каждой точки выборки $x_{i1}, x_{i2},..., x_{in}$ каждому полученному правилу. По максимальной степени принадлежности правилу происходит разбивка точек выборки на множества A_k, где $k = \overrightarrow{1,K}$, которые были описаны ранее.

Значения σ_{kj}, как отмечается в работах [2] и [3], влияют на качество функционирования сети незначительно. В текущей работе для каждого терма k_j каждого правила они будут выбраны свои, но одинаковые – это будет сделано для ускорения процесса обучения сети.

Однако для полноты выкладок можно предложить следующий метод нахождения σ_{kj}.

Так как множества A_k, $k = \overrightarrow{1,K}$, уже сформированы, то для каждого правила построение σ_{kj} и $f_k(x_1,..., x_n)$ будет происходить отдельно, основываясь на выборке данных, представленной множеством A_k.

На рис. 1 показано, что для определения k, а соответственно f_k, которое будет вычислять значение $\overline{y}_i$, выполняется следующая процедура:

$$k = \arg\max_S \prod_{j=1}^{n} \mu_{sj} \qquad (6)$$

Таким образом, очевидно, что для каждого правила k при $(x_1,..., x_n) \in A_k$ значение $\prod_{j=1}^{n} \mu_{sj}$ должно быть как можно большим, а в противном случае – наименьшим.

Таким образом, для нахождения оптимальных σ_{kj}, задавшись некоторым $\varepsilon \to 0, \varepsilon > 0$ для каждого правила, в среднеквадратическом смысле необходимо найти решение системы (7) относительно σ_{kj}.

$$\begin{cases} \prod_{j=1}^{n} \mu_{kj}(x_{ij}) = 1, & if\ (x_{i1}, x_{i2},..., x_{in}) \in A_k, \\ \prod_{j=1}^{n} \mu_{kj}(x_{ij}) = \varepsilon, & else. \end{cases} \qquad (7)$$

Если взять натуральный логарифм от обеих частей равенств системы, то систему можно переписать в следующем виде:

$$\begin{cases} \sum_{j=1}^{n} \ln(\mu_{kj}(x_{ij})) = 0, & if\ (x_{i1}, x_{i2},..., x_{in}) \in A_k, \\ \sum_{j=1}^{n} \ln(\mu_{kj}(x_{ij})) = \ln \varepsilon, & else. \end{cases} \qquad (8)$$

С учётом формулы (5), выражение (8) можно переписать в виде:

$$\begin{cases} \sum_{j=1}^{n} \ln\left(e^{-\frac{(x_{ij} - c_{kj})^2}{2\sigma_{kj}^2}} \right) = 0, & if\ (x_{i1}, x_{i2},..., x_{in}) \in A_k, \\ \sum_{j=1}^{n} \ln\left(e^{-\frac{(x_{ij} - c_{kj})^2}{2\sigma_{kj}^2}} \right) = \ln \varepsilon, & else. \end{cases} \qquad (9)$$

С учётом того, что $\ln e^a = a$, и введя обозначения (10), (11) и (12) :

$$h_{kij} = -\frac{\left(x_{ij} - c_{kj}\right)^2}{2}, \quad (10) \qquad p_{kj} = \frac{1}{\sigma_{kj}^2}, \quad (11)$$

$$\begin{cases} g_{ki} = 0, & if \ \left(x_{i1}, x_{i2}, ..., x_{in}\right) \in A_k, \\ g_{ki} = \ln \varepsilon, & else, \end{cases} \quad (12)$$

получается:

$$\sum_{j=1}^{n} h_{kij} p_{kj} = g_{ki}, \ i = \overrightarrow{1, L}, \ j = \overrightarrow{1, n}. \quad (13)$$

Если ввести обозначения (14)-(16):

$$H_k = \left\| h_{kij} \right\|_{i=\overrightarrow{1,L}}^{j=\overrightarrow{1,n}}, \quad (14)$$

$$p_k = \left\| p_{kj} \right\|_{j=\overrightarrow{1,n}}, \quad (15)$$

$$g_k = \left\| g_{ki} \right\|_{i=\overrightarrow{1,L}}, \quad (16)$$

то выражение (13) можно переписать в матричном виде (17):

$$H_k p_k = g_k. \quad (17)$$

Таким образом, для нахождения σ_{kj} необходимо решить задачу квадратичного программирования (18)-(19) для каждого правила $k = \overrightarrow{1, K}$:

$$\min_{p_{kj}} \left(H_k p_k - g_k\right)^T \left(H_k p_k - g_k\right), \quad (18)$$

$$p_{kj} \geq 0, \ j = \overrightarrow{1, n}, \quad (19)$$

Следует отметить, что p_{kj} может равняться нулю, так как если $x_{ij} = x_{mj}$, $\forall i, m$, то $\sigma_{kj} \to \infty$, что означает $p_{kj} \to 0, \ \forall k$.

Методика решения оптимизационных задач квадратичного программирования детально описана в [6].

Методика восстановление зависимостей $f_k\left(x_1, ..., x_n\right)$ приводится в данной работе не будет, так как она описана в [1]. Только необходимо отметить, что тригонометрические полиномы являются квазилинейными функциями, а, следовательно, их коэффициенты могут быть восстановлены с помощью метода наименьших квадратов. Также необходимо ещё раз отметить, что каждая $f_k\left(x_1, ..., x_n\right)$ восстана - вливается на своём множестве A_k

На этом обучение сети завершается.

При поступлении на вход сети очередного $\left(x_{i1}, x_{i2}, ..., x_{in}\right)$ вычисляется $k = \arg\max_S \prod_{j=1}^{n} \mu_{sj}$ и $\overline{y}_i = f_k\left(x_{i1}, x_{i2}, ..., x_{in}\right)$.

Оценка качества работы сети производится на полной обучающей выборке A и проверочной выборке по значениям критериев (1) и (2).

Результаты применения ННСА НОКП и радиальной базисной сети к анализу котировок валют

Для оценки применимости вышеприведённых нейронных сетей и для сравнительного анализа разных показателей их работы, радиальная базисная сеть и ННСА НОКП применялись к задаче анализа значений валютных котировок на основании их предыстории.

Эксперимент был поставлен следующим образом (аналогичен постановке эксперимента в работе [1]). Параметрами эксперимента было соотношение обучающей выборки и проверочной, а также число входов сетей. Соотношение обучающей выборки принимало значение 10:90, 20:80, 30:70, 40:60, 50:50, 60:40, 70:30, 80:20, 90:10 – в соотношении на первом месте стоит процент объёма обучающей выборки от объёма всей выборки данных, а на втором – процент проверочной выборки от всей выборки данных. Входами сетей было значение котировки валюты для 2, 3 и 4 периодов времени до прогнозного значения, а выходом сети было значение котировки валюты, которое необходимо

спрогнозировать. Эксперимент ставился для котировок валютных пар EUR/GBP, EUR/USD, USD/JPY, USD/CHF, которые были взяты из [4].

Анализ значений критериев (1) и (2) на обучающих выборках показал, что качество обучения сетей было приблизительно одинаково, о чём свидетельствует «близкость» значений критериев для обоих сетей при одинаковых параметрах эксперимента. Это даёт основание для проведения данного сравнительного анализа.

Исходя из значений тех же критериев на проверочных выборках, было продемонстрировано одинаковое их поведение при сравнении обоих сетей и значения были очень близки друг к другу при одинаковых параметрах эксперимента – сети имеют одинаковую прогностическую способность.

Следует отметить ещё раз факт, приведённый в работе [1].

Значения всех критериев для обеих выборок (обучающей и проверочной) для обоих сетей, при объёме обучающей выборки 10-20% от всей выборки данных, значительно отличаются от значений при других объёмах обучающей выборки. Как отмечается в работе [1], и как подтверждают результаты применения радиальной базисной сети, это вызвано тем, что при столь малых объёмах обучающих выборок растёт влияние каждой отдельной точки выборки на процесс обучения. Следовательно, выше (по сравнению со случаями других соотношений обучающей и проверочной выборки) влияние случайных шумов, не связанных с природой процесса на процесс обучения.

Исследования динамики рассматриваемых процессов, которые опираются на анализ поведения значений критериев качества (1) и (2) на обучающих и проверочных выборках, показывают незначительный прирост качества работы сетей при увеличении размерности пространства входных данных. Отсюда следует вывод, сделанный уже в работе [1]. При значении размерности входного пространства для нейронных сетей около 3 и 4, значения критериев начинают входить в «насыщение», и значения критериев качества позволяют сделать заключение, что размерность 4 можно считать «точкой насыщения». Однако данный факт ещё необходимо проверить в последующих исследованиях с усеченным диапазоном изменения параметра эксперимента «объём обучающей выборки».

В ходе проведённых опытов были отмечены «всплески» значений критериев качества на проверочной выборке при объёме обучающей выборки - 90%. Это объясняется (как отмечается также в работе [1]) тем же малым объёмом проверочной выборки, а, следовательно, большим влиянием каждой отдельной точки выборки на значение критерия. Таким образом, для оценки работоспособности сети «всреднем» данное соотношение выборок не является адекватным.

Анализ обеих сетей позволяет сделать вывод об их относительно стабильном поведении при соотношении обучающей и проверочной выборок 50:50 – 80:20.

Если анализировать скорость обучения сети, то радиальная базисная сеть была обучена значительно быстрее ННСА НОКП. Таким образом, по скорости обучения радиальная базисная сеть превосходит ННСА НОКП.

Однако применение сетей в реальном времени подразумевает их переобучение при поступлении новой информации при заданных ограничениях на время переобучения (минутные котировки, часовые, дневные…). Система прогнозирования, которая не удовлетворяет установленному ограничению, непригодна для применения в реальном времени. При поступлении новой точки выборки весовой вектор радиальной базисной сети ([5]) должен быть найден заново. Для ННСА НОКП определяется лишь принадлежность точки ваборки данных одному из кластеров, которые задаются, де-факто, правилами нечёткой системы. Точка принадлежит тому кластеру, для которого получается максимальная степень принадлежности антецеденте нечёткого правила, которое соответствует данному кластеру. После происходит процесс «доучивания» рекуррентным методом коэффициентов соответствую-

щей функциональной зависимости в консеквенте нечёткого правила. Выходит, что по параметру «способность к доучиванию» ННСА НОКП превосходит радиальную базисную сеть.

Если анализировать объём занимаемой сетями памяти, то с ростом обучающей выборки будет расти и объём памяти, занимаемой радиальной базисной сетью. Если же допустить конечное реальное количество кластеров точек обучающей выборки и при задании ограничения на рост порядка частичных описаний,

ННСА НОКП будет занимать ограниченный объём памяти вычислительной системы. По этому параметру ННСА НОКП превосходит радиальную базисную сеть.

Заключение

Работа была посвящена сравнительному анализу ННСА НОКП и радиальной базисной нейронной сети применительно к задаче прогнозирования котировок валютных пар на основании их предыстории. В качестве выборок данных были взяты котировки EUR/GBP, EUR/USD, USD/JPY, USD/CHF с [4].

Применение сетей к данной задаче позволило сделать ряд общих выводов. Первое, с увеличением размерности пространства данных, которые поступают на вход сети, прирост качества функционирования незначителен. Это позволяет допустить существования периода «насыщения» в зависимости «размерность пространства входных данных – качество функционирования прогнозирующей модели» для данной задачи. Исходя из значений критериев для обоих сетей, был сделан вывод, что при размерности 4 входного пространства прогнозирующая модель способна адекватно оценивать процесс. Второе, объём обучающей выборки влияет на качество работы сети как на обучающей, так и на проверочной выборке. Третье,– вывод подобен предыдущему – для оценки прогнозирующей способности сети, объём проверочной выборки не должен быть мал. Анализ динамики значений критериев качества показал, что соотношение обучающей и проверочной выборки 90%:10% от исходной выборки является неприемлемым для оценки прогнозирующей способности сети.

Сравнительный анализ применения обоих рассматриваемых в данной работе сетей дал следующие результаты. Подобный анализ – адекватен, так как соответствующие значения критериев на обучающих выборках были относительно близки друг к другу для обеих сетей – сети были одинаково хорошо обучены. На проверочных выборках соответствующие значения критериев были близки для обеих сетей. Таким образом, по качеству функционирования сети примерно идентичны. Что касается времени обучения, то ННСА НОКП уступает по этому параметру радиальной базисной нейронной сети. В то же время, при «дообучении» радиальной базисной нейронной сети нужно переобучать весь весовой вектор, что при больших размерностях может оказаться весьма проблематичным, когда для ННСА НОКП требуется переобучить лишь одно правило рекурентным методом при заданной стартовой («нулевой») точке. Так что, по этому параметру ННСА НОКП превосходит радиальную базисную нейронную сеть. Третьим фактом сравнительного анализа был следующий результат.

В то время, когда рост объёма поступающих данных сопровождается ростом объёма памяти вычислительной системы, которую занимает радиальная базисная нейронная сеть, рост объёма памяти, занимаемой ННСА НОКП, – ограничен. Данное ограничение связано с заданием верхнего значения порядка частичных описаний и количеством реально существующих кластеров выборки данных.

Безусловно, требуется проводить дальнейшие сравнительные анализы ННСА

НОКП с другими нейронными сетями. В последующих работах необходимо экспериментально исследовать полезность метода (18)-(19) для обучения параметра σ_{kj} для ННСА НОКП.

Необходимо расширение множества исследуемых выборок данных на другие типы финансовых рынков.

Литература

1. Мурга Н.А. Нейросетевая архитектура на частичных обучениях //New Trends in Classification and Data Mining. ITHEA (Sofia, Bulgaria). – 2010. – № 16 – С. 170-184.
2. Зайченко Ю.П. Основи проектування інтелектуальних систем. – К.: Видавничий Дім «Слово», 2004. – 352 с.
3. Зайченко Ю.П. Нечёткие модели и методы в интеллектуальных системах. – К.: «Издательский Дом «Слово»», 2008 – 344 с.
4. Дневные котировки валют за период с 25.03.2009 по 24.03.2010. http://www.finam.ru/analysis/export/default.asp.
5. Хайкин С. Нейронные сети: полный курс, 2-е издание.: Пер. с англ. – М.: Издательский дом “Вильямс”, 2006. – 1104 с.
6. Зайченко Ю.П. Дослідження операцій. Сьоме видання, перероблене та доповнене. – К.: Видавничий Дім «Слово», 2006. – 816 с.