УДК 004.032.26

Е.В. Бодянский, Е.А. Винокурова, П.П. Мулеса, И.Г. Перова

ДИАГНОСТИРУЮЩАЯ НЕЙРО-ФАЗЗИ-СИСТЕМА И ЕЕ АДАПТИВНОЕ ОБУЧЕНИЕ В ЗАДАЧАХ ИНТЕЛЛЕКТУАЛЬНОЙ ОБРАБОТКИ ДАННЫХ МЕДИКО-БИОЛОГИЧЕКИХ ИССЛЕДОВАНИЙ

Аннотация. Интеллектуальный анализ данных является мощным инструментарием при решении широкого спектра задач, и прежде всего диагностирования, классификации, распознавания образов и т.п. В статье предложена диагностирующая нейро-фаззи система для ситуаций, когда размерность входных сигналов имеет один порядок с объемом обучающей выборки, а сами данные поступают на вход системы в последовательном on-line режиме.

Ключевые слова: нейро-фаззи-система, адаптивное обучение, интеллектуальная обработка данных, медико-биологическая диагностика

Введение

Для решения широкого круга задач интеллектуального анализа данных (Data Mining), связанных прежде всего, с диагностикой, классификацией, распознаванием образов и т.п., все чаще используется искусственные нейронные сети (ANN) [1], благодаря их универсальным аппроксимирующим возможностям и способности обучения по имеющимся экспериментальным данным. И хотя для решения подобных задач наибольшее распространение получил классический многослойный персептрон (MLP), нельзя не отметить такие его основные недостатки, как достаточно большой объем обучающей выборки, низкая скорость сходимости процесса обратного распространения ошибок, необходимость использования большого количества эпох обучения. И если сугубо вычислительные проблемы можно обойти, необходимость наличия репрезентативной обучающей выборки существенно осложняет использование этой нейронной сети при решении многих практических задач. Особенно остро эта проблема возникает в медико-биологических исследованиях, где выборка данных короткая и при этом объект описывается множеством разноплановых характеристик [2, 3].

В данной ситуации более предпочтительными представляются радиально-базисные нейронные сети (RBFN) [1], чей выходной сигнал линейно зависит от настраиваемых синаптических весов. Это обстоятельство позволяет использовать для обучения этих сетей широкий арсенал хорошо известных подходов, от стандартного метода наименьших квадратов до популярных алгоритмов линейной адаптивной идентификации [4]. И хотя специфика задач диагностики-классификации ограничивает возможности использования традиционного квадратичного критерия обучения, применение специализированного критерия Дж. Шинка [5], ориентированного на задачи распознавания образов с двоичным обучающим сигналом, позволило синтезировать достаточно простую и эффективную диагностирующую радиально-базисную нейронную сеть [6, 7].

Несмотря на все свои достоинства, радиально-базисные сети все же не являются панацеей, поскольку их возможности ограничиваются, так называемым, «проклятием размерности», ведущему к экспоненциальному росту числа настраиваемых синаптических весов в зависимости от размерности пространства входных сигналов-образов.

Для преодоления этой проблемы чаще всего используется процедура предварительной установки центров радиально-базисных функций с помощью того или иного метода кластеризации. Таким образом, контролируемое обучение сети дополняется самообучением ее центров, что делает такое обучение излишне громоздким. В [8] для решения задач обработки текстовых документов в рамках проблемы Text Mining, характеризующейся высокой размерностью входных сигналов, была введена иерархическая радиально-базисная нейронная сеть с многослойной архитектурой, в каждом узле использующая обычную RNFN, на вход которой подается только некоторая часть признаков, что позволило преодолеть проблему «проклятия размерности». Главное, что удалось добиться в данной ситуации – это возможность работы в условиях, когда размерности входных образов соизмеримы с объемом обучающей выборки. Вместе с тем, нельзя не отметить громоздкость этой системы, невозможность работы в последовательном on-line режиме, высокий уровень субъективизма при разбиении входного образа на множество подвекторов-входов для каждого узла сети.

Так или иначе, проблема обработки данных с высокой размерностью вектора признаков в условиях, когда объем обучающей выборки соизмерим с этой размерностью, представляет несомненный интерес и особенно для задач классификации и диагностики для Text Mining, Web Mining, медико-биологических приложений.

Весьма перспективным здесь представляется использование нейро-фаззи-систем (NFS) [9], которые наряду с аппроксимирующими возможностями и способностью и обучению, позволяет обеспечить лингвистическую интерпретацию результатов. Нельзя также не отметить и их эквивалентность с точки зрения получаемых результатов радиально-базисным нейронным сетям [10], что позволяет использовать для их обучения те же алгоритмы, что и в RBFN. Синтезу такой диагностирующей нейро-фаззи системы для ситуаций, когда размерность входных сигналов имеет один порядок с объемом обучающей выборки, а сами данные поступают на вход системы в последовательном on-line режиме, посвящена настоящая статья.

Архитектура диагностирующей нейро-фаззи системы

Архитектура рассматриваемой NFS приведена на рис. 1 и состоит из шести последовательно соединенных слоев. На входной (нулевой, рецепторный) слой NFS подается $(n\times 1)$-мерный вектор входных сигналов-образов $x(k)=(x_1(k),x_2(k),\dots x_n(k))^T$, где $k=1,2,\dots,N$ - номер наблюдения в исходном массиве данных. При этом предполагается, что все компоненты $x_i(k)$ предварительно преобразованы так, что

$$0\le x_i(k)\le 1,\ \forall i=1,2,\dots,n,$$

а двоичные входные признаки принимают значения 0 или 1.

Первый скрытый слой содержит nh функций принадлежности $\mu_{li}(x_i(k))$, $i=1,2,\dots,n$; $l=1,2,\dots,h$ и выполняет фаззификацию входных переменных, при этом чем большее число h, тем лучше аппроксимирующие свойства NFS, хотя для двоичных признаков достаточно иметь и $h=2$.

Второй скрытый слой производит агрегирование уровней принадлежности, вычисленные в первом слое, и состоит из h блоков умножения Π.

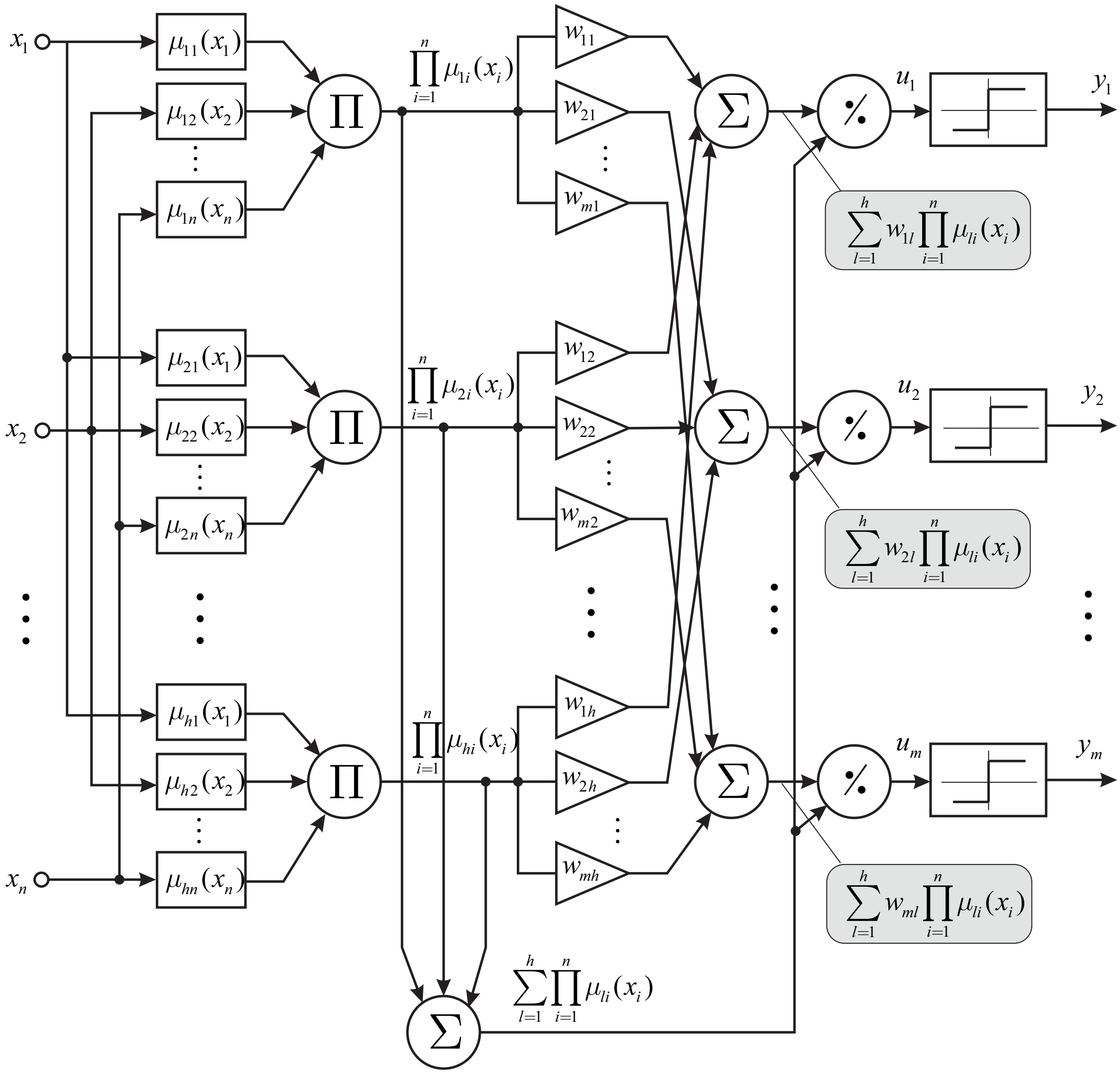

Рисунок 1 – Диагностирующая нейро-фаззи система

Третий скрытый слой – это слой синаптических весов w_{jl}, $j = 1, 2, \ldots, m$, подлежащих обучению в процессе обучения. Рассматриваемая NFS содержит mh настраиваемых весов, где m-количество возможных классов-диагнозов, по одному на каждый выход системы. Понятно, что $mh \ll e^n$, т.е. количество весов NFS существенно меньше числа весов RBFN.

 ISSN 1562-9945

Четвертый скрытый слой образован $m+1$ сумматорами Σ, вычисляющих суммы сигналов на выходах второго и третьего скрытых слоев.

В пятом скрытом слое, образованном m блоками деления $\bullet/\bullet$, производится нормализация выходных сигналов четвертого слоя.

И, наконец, выходной (шестой) слой содержит m нелинейных активационных функций, при этом в задачах диагностики целесообразно использование простейших сигнум-функций, принимающих значение +1 в случае правильного диагноза и -1 - в противном случае. Таким образом выходные сигналы системы $y_j(k)$ могут принимать только два значения ± 1.

Таким образом, при подаче на вход NFS векторного сигнала $x(k)$, элементы первого слоя вычисляют уровни принадлежности $\mu_{li}(x_i(k))$, при этом обычно в качестве функций принадлежности используются колоколообразные (ядерные) конструкции с нестрого локальным рецепторным полем, что позволяет избежать возникновение «дыр» в фаззифицированном пространстве [11]. Чаще всего – это традиционные гауссианы

$$\mu_{li}(x_i(k)) = \exp\left(-\frac{(x_i(k)-c_{li})^2}{2\sigma_i^2}\right), \tag{1}$$

где c_{li} - параметр центра (в простейшем случае центры равномерно размещаются на интервале [0,1] с шагом $(h-1)^{-1}$), σ_i - параметр ширины, выбираемый эмпирически или настраиваемый с помощью обратного распространения ошибок [12]. На рис. 2 показан вид таких функций принадлежности.

Понятно, что для двоичных переменных $x_i(k)$ достаточно всего двух треугольных функций принадлежности

$$\begin{cases} \mu_{1i}(x_i(k)) = 1 - x_i(k), \\ \mu_{2i}(x_i(k)) = x_i(k), \end{cases} \tag{2}$$

приведенных на рис. 3.

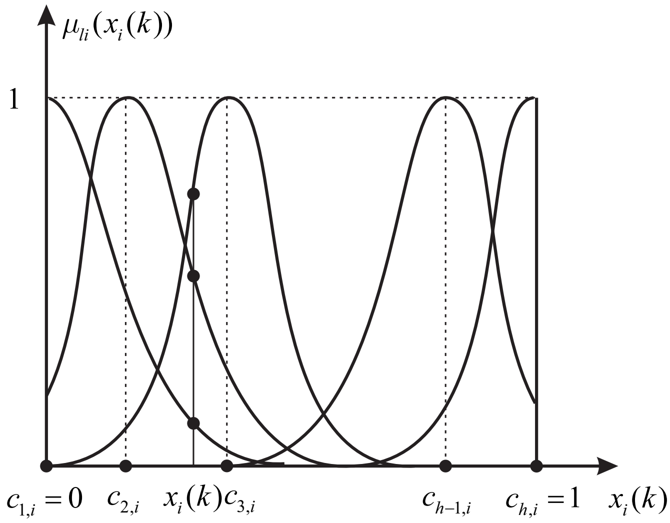

Рисунок 2 – Колоколообразные функции принадлежности

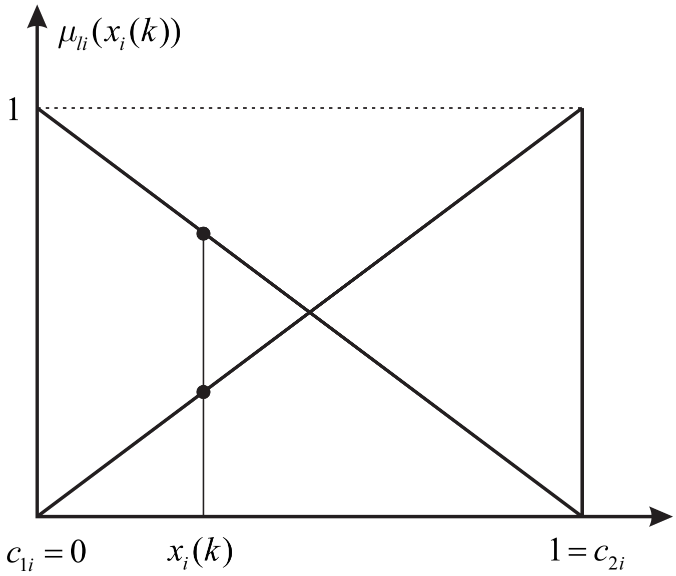

Рисунок 3 – Функции принадлежности для двоичных переменных

Заметим также, что функции принадлежности (2) в ряде случаев с успехом могут быть использованы и для признаков, принимающих произвольное количество значений (см. рис. 3), а число настраиваемых синаптических весов принимает минимально возможное значение $2m$.

На выходах второго скрытого слоя появляются агрегированные значения $\prod_{i=1}^{n}\mu_{li}(x_i(k))$, при этом несложно заметить, что, если параметры ширины σ_i одинаковы для всех признаков, т.е. $\sigma_i=\sigma$, то

ISSN 1562-9945

$$\prod_{i=1}^{n}\mu_{li}(x_i(k))=\prod_{i=1}^{n}\exp\left(-\frac{(x_i(k)-c_{li})^2}{2\sigma^2}\right)=\exp\left(-\frac{\|x_i(k)-c_{li}\|^2}{2\sigma^2}\right),$$

(здесь $c_l=(c_{l_1},c_{l_2},\ldots,c_{l_n})^T$), т.е. реализуется нелинейное преобразование аналогичное RBFN.

Выходами третьего скрытого слоя являются сигналы $w_{jl}\prod_{i=1}^{n}\mu_{li}(x_i(k))$, четвертого $\sum_{l=1}^{h}w_{jl}\prod_{i=1}^{n}\mu_{li}(x_i(k))$ и $\sum_{l=1}^{h}\prod_{i=1}^{n}\mu_{li}(x_i(k))$, пятого

$$u_j(k)=\frac{\sum_{l=1}^{h}w_{jl}\prod_{i=1}^{n}\mu_{li}(x_i(k))}{\sum_{l=1}^{h}\prod_{i=1}^{n}\mu_{li}(x_i(k))}=\sum_{l=1}^{h}w_{jl}\frac{\prod_{i=1}^{n}\mu_{li}(x_i(k))}{\sum_{l=1}^{h}\prod_{i=1}^{n}\mu_{li}(x_i(k))}=$$

$$=\sum_{l=1}^{h}w_{jl}\varphi_l(x(k))=w_j^T\varphi(x(k))$$

(здесь $\varphi_l(x(k))=\prod_{i=1}^{n}\mu_{li}(x_i(k))\left(\sum_{l=1}^{h}\prod_{i=1}^{n}\mu_{li}(x_i(k))\right)^{-1}$,

$w_j=(w_{j1},w_{j2},\ldots,w_{jh})^T$, $\varphi(x(k))=(\varphi_1(x(k)),\varphi_2(x(k)),\ldots,\varphi_h(x(k)))^T$) и, наконец, шестого

$$y_j(k)=\operatorname{sign}u_j(k).$$

Несложно видеть, что рассматриваемая NFS есть модификация системы Л.Ванга-Дж. Менделя [13, 14], ориентированная на решения задач on-line диагностики-классификации.

Обучение диагностирующей нейро-фаззи системы

Для обучения синаптических весов введенной системы будем использовать метод обучения, основанный на специализированном критерии [5], который предназначен для решения задач распознавания образов, классификации, диагностики и т.п.

Введем в рассмотрение m ошибок обучения

$$e_j(k)=d_j(k)-y_j(k)=d_j(k)-\operatorname{sign}u_j(k)$$

и m критериев, основанных на этих ошибках

$$\begin{aligned} E_j(k) &= e_j(k)u_j(k) = d_j(k)u_j(k) - \left|u_j(k)\right| = \\ &= \left(d_j(k) - \operatorname{sign} w_j^T \varphi(x(k))\right) \cdot w_j^T \varphi(x(k)), \end{aligned} \tag{3}$$

где $d_j(k) \in \{-1,1\}$ - обучающий сигнал, принимающий значения 1, если входной вектор $x(k)$ относится к $j-$ му диагнозу, и -1 в противном случае.

Для настройки синаптических весов используем стандартную градиентную процедуру минимизации критерия (3)

$$w_{jl}(k+1) = w_{jl}(k) - \eta(k)\frac{\partial E_j(k)}{\partial w_{jl}}$$

(здесь $\eta(k)$ - параметр шага обучения), которая в векторной форме может быть переписана в виде

$$\begin{aligned} & w_j(k+1) = w_j(k) + \eta(k)e_j(k)\varphi(x(k)) = \\ & = w_j(k) + \eta(k)\left(d_j(k) - \operatorname{sign}\, w_j^T(k)\varphi(x(k))\right) \cdot \varphi(x(k)), \\ & j = 1, 2, \ldots, m. \end{aligned} \tag{4}$$

Вводя далее общий критерий для всех выходов системы

$$E(k) = \sum_{j=1}^{m} E_j(k) = \sum_{j=1}^{m} e_j(k)u_j(k),$$

можно записать алгоритм обучения всех синаптических весов системы в виде

$$W(k+1) = W(k) + \eta(k)\left(d(k) - \operatorname{sign}\, W(k)\varphi(x(k))\right) \cdot \varphi^T(x(k)), \tag{5}$$

где $\operatorname{sign}(u_1(k), u_2(k), \ldots, u_m(k))^T = (\operatorname{sign} u_1(k), \operatorname{sign} u_2(k), \ldots, \operatorname{sign} u_m(k))^T$, $d(k) = (d_1(k), d_2(k), \ldots, d_m(k))^T$,

ISSN 1562-9945

$$W(k)=\begin{pmatrix} w_1^T(k) \\ w_2^T(k) \\ \vdots \\ w_m^T(k) \end{pmatrix}-(m\times h)$$ - матрица настраиваемых синаптических весов.

Известно, что градиентные алгоритмы типа (3)-(5) обеспечивают сходимость в достаточно широком диапазоне варьирования параметра шага $\eta(k)$ [15], однако при этом скорость сходимости может оказаться недостаточной. Увеличить эту скорость можно, воспользовавшись квазиньютоновскими процедурами обучения [16], например,

$$w_j(k+1)=w_j(k)+\left(\varphi(x(k))\varphi^T(x(k))+\eta I\right)^{-1}e_j(k)\varphi(x(k)), \quad (6)$$

где $\eta>0$ - регуляризующий параметр, $I-(\mathrm{h}\times\mathrm{h})$ - единичная матрица.

Воспользовавшись далее леммой обращения матриц, несложно показать, что [7]

$$\left(\varphi(x(k))\varphi^T(x(k))+\eta I\right)^{-1}\varphi(x(k))=\frac{\varphi(x(k))}{\eta+\|\varphi(x(k))\|^2},$$

и переписать (6) компактной форме

$$w_j(k+1)=w_j(k)+\frac{e_j(k)\varphi(x(k))}{\eta+\|\varphi(x(k))\|^2}, \quad (7)$$

или

$$W(k+1)=W(k)+\frac{d(k)-signW(k)\varphi(x(k))}{\eta+\|\varphi(x(k))\|^2}\varphi^T(x(k)), \quad (8)$$

что при $\eta=0$ является многомерным вариантом оптимального по быстродействию алгоритма обучения, введенного в [17].

Выводы

Предложенная диагностирующая нейро-фаззи система и адаптивный алгоритм ее обучения предназначены для решения задач распознавания образов, классификации, диагностики и т.п. в условиях когда объем обучающей выборки соизмерим с размерностью входных образов, а сами эти образы поступают на обработку последовательно в on-line режиме. Особенностью предлагаемой системы являются существенно меньшее количество настраиваемых параметров по сравнению с искусственными нейронными сетями, решающими тут же задачу.

Система характеризуется простотой численной реализации, высоким быстродействием при обучении, возможностью обработки информации, задаваемой в различных шкалах (количественной, ранговой, бинарной), что за частую возникает в задачах интеллектуальной обработки медицинских данных.

ЛИТЕРАТУРА

1. Haykin S. Neural Networks. A Comprehensive Foundation / S. Haykin. - Upper Saddle River, NJ: Prentice Hall. - 1999. - 842 p.
2. Rizzo R. Computational Intelligence Methods for Bioinformatics and Biostatistics / Rizzo R. - In Lecture Notes in Bioinformatics (7th International Meeting, CIBIB 2010, Palermo, Italy, September 16-18, 2010). – Springer. – 2011. - 301 p.|
3. Kountchev R. Advances in Intelligent Analysis of Medical Data and Decision Support Systems (Studies in Computational Intelligence) / Kountchev R. and etc - Springer. – 2013. - 246 p.
4. Ljung L. System Identification: Theory for the User / L. Ljung. - PTR Prentice Hall, Upper Saddle River, N.J., 1999. - 672 p.
5. Shynk J.J. Performance surfaces of a single-layer perceptron / Shynk J.J. // IEEE Trans. on Neural Networks. – 1990. – 1. – P. 268-274.
6. Бодянский Е.В. Диагностика и прогнозирование временных рядов с помощью многослойной радиально-базисной нейронной сети / Бодянский Е.В., Кучеренко Е.И., Чапланов А.П. // Труды 8 Всеросийской конф. с междунар. Участием «Нейрокомпьютеры и их применение». – Москва, 2002. – С. 209-213.
7. Бодянский Е.В. Нейро-фаззи сети Петри в задачах моделирования сложных систем / Бодянский Е.В., Кучеренко Е.И., Михалев А.И. – Днепропетровск: Системные технологии, 2005. – 311 с.

8. Бодянский Е.В. Семантическое аннотирование текстовых документов на основе иерархической радиально-базисной нейронной сети / Бодянский Е.В., Шубкина О.В. // Восточно-европейский журнал передовых технологий. – 2010. - Вып. 9 (90). – С. 70-74.
9. Jang J.-S.R. Neuro-Fuzzy and Soft Computing / J.-S.R. Jang, C.-T. Sun, E. Mizutani. - Prentice Hall, Upper Saddle River, NJ. - 1997. - 640 p.
10. Jang J.S.R. Functional equivalence between radial basis function networks and fuzzy inference systems / J.S.R. Jang, C.T. Sun // IEEE Trans. on Neural Networks. - 1993. - 4. - P. 156-159.
11. Friedman J. The Elements of Statistical Learning. Data Mining, Inference and Prediction / Friedman J., Hastie T., Tibshirani R. – Berlin: Springer, 2003. – 552 p.
12. Осовский С. Нейронные сети для обработки информации / С. Осовский. -М.: Финансы и статистика, 2002. - 344 с.
13. Wang L.X. Fuzzy basis functions, universal approximation, and orthogonal least squares learning / L.X. Wang, J.M. Mendel // IEEE Trans. on Neural Network. - 1992. - 3. - P. 807-814.
14. Wang L.-X. Adaptive Fuzzy Systems and Control: Design and Stability Analysis / L.-X. Wang. - New Jersey: Prentice Hall. - 1994. - 256 p.
15. Деревицкий Д.П. Прикладная теория дискретных адаптивных систем управления / Деревицкий Д.П., Фрадков А.Л. – М.: Наука, 1981. – 216 с.
16. Shepherd A.J. Second-Order Methods for Neural Networks / A.J. Shepherd. - London: Springer-Verlag. - 1997. - 145 p.
17. Цыкин Я.З. Основы теории обучающихся систем / Цыкин Я.З. – М.: Наука, 1984. – 320 с.