

Адаптивная нейросетевая система визуализации изображений, распознавания и классификации микрообъектов

О. И. Джуманов

Поставлена задача построения адаптивных систем обработки нестационарной по природе информации для решения задач распознавания изображений микрообъектов и прогнозирования динамических процессов на основе теории и практики применения нейронных сетей (НС). Разработаны модели адаптивного обучения на основе различных архитектур НС. Реализованы программные модули обучения и логико-информационная модель функционирования программной системы визуализации изображений, распознавания и классификации пыльцевых зёрен.

1. ВВЕДЕНИЕ

В настоящее время при решении задач построения и обработки баз данных и знаний различного назначения специалисты ориентируются на концепции технологии Data Mining [1]. Технология Data Mining является мультидисциплинарной дисциплиной, которая позволяет извлекать данные и увидеть скрытые правила и закономерности в неопределённом наборе данных, получить знания для построения адекватных моделей эффективного предвидения и управления при недостаточных априорных сведениях и слабой формализуемости процесса. Эту технологию целесообразно применить в системах обработки данных для визуализации изображений и распознавания микрообъектов, отпечатков пальцев, текстов, графиков технико-экономических показателей, параметров производства и др. С помощью методов Data Mining могут быть эффективно решены задачи разработки методов, моделей, алгоритмов и инструментальных средств баз данных, машинного обучения, моделирования и алгоритмизации.

Решение указанных задач, в свою очередь, сводится к моделированию динамических процессов, так как рассматриваемые в них объекты являются источниками нестационарной по природе непрерывной информации. Такие динамические задачи представляют важную и, к сожалению, недостаточно исследованную область знаний. Трудность решения таких задач связана с проблемами адекватного описания динамических процессов, которым свойственны нелинейность, сложная зависимость входных и выходных факторов, недостаточная база априорных сведений и за-

частую их отсутствие. Кроме того, при визуализации, распознавании, прогнозировании система обработки данных должна обладать способностью адаптации к изменяющимся условиям в процессе её эксплуатации в реальном режиме времени, особенно когда имеет место нестационарность, слабая формализуемость и недостаточность априорных сведений.

В связи с этим, удобной компьютерной технологией для построения адаптивной системы обработки данных нестационарных и слабоформализуемых процессов представляется использование аппарата искусственных нейронных сетей (НС), успешно применяющегося в настоящее время в системах искусственного интеллекта, распознавании и классификации образов, аппроксимации функций многих переменных, а также для прогнозирования, оптимизации и управления объектами различной природы [2]. Теория аппарата искусственных НС в настоящее время является хорошо развитой [3–5].

Существуют программные средства, в которых аппарат НС включён как специализированные пакеты. Среди известных программ такого рода MatLab и Statistica. Существует также множество программ-оболочек, позволяющих быстро конструировать модели на ИНС, содержащих встроенные алгоритмы обучения.

Однако продолжение исследований с целью разработки новых информационных технологий на основе НС, позволяющих интеллектуализировать процессы обработки непрерывной информации на изложенных концепциях, является обоснованным и представляет большой теоретический и практический интерес [6–8].

Следует отметить, что главное преимущество НС, используемое при обработке нестационарной и слабоформализуемой информации, – это способность к обучению и самообучению, позволяющая обнаружить в случайных процессах порядок и некоторую вероятностную структуру. Для НС строго математически доказана их способность представлять любую вещественную непрерывную векторную функцию любого вещественного непрерывного векторного аргумента. В отличие от статистических методов,

НС способны не только выполнять заранее запрограммированную последовательность действия на заранее определённом наборе данных, но и анализировать вновь поступающую информацию, находить в ней закономерности, адаптироваться на основе предыдущих значений, что даёт возможность учитывать динамику поведения представляемых для обработки нестационарных процессов.

Предлагаемая методология создания системы обработки данных непрерывной информации опирается на следующие три основных свойства НС:

- а) способность адаптации к свойствам динамических данных за счёт выбора структуры сети, активационных функций нейронов, коэффициентов синаптических связей между нейронами, устранения незначимых связей;
- б) аппроксимационные возможности, выраженные в том, что обученная сеть способна сглаживать случайные выбросы в динамических рядах, работать с неполной слабо формализованной или, напротив, избыточной информацией;
- в) предсказательные свойства, выраженные в том, что обученная модель НС позволяет осуществлять предсказание изменения вектора выходных координат в широком диапазоне изменения входных переменных, при этом имеется возможность наблюдения за всеми элементами внутренней структуры модели.

При построении систем обработки данных на основе НС нами ставятся следующие теоретические задачи:

- провести моделирование предметной области, представленной обобщёнными характеристиками и специфическими свойствами объектов;
- доказать возможность построения адекватных моделей решения поставленных задач, которая обеспечивается за счёт адаптивного обучения НС;
- доказать, используя адекватную модель НС, возможность построения новой структуры систем обработки данных, обеспечивающей прозрачность как в отношении структуры, так и в смысле значений коэффициентов синаптических связей;
- осуществить оптимизацию НС на основе выбора в качестве начальной некоторой избыточной структуры НС, в которой количество входных нейронов равно количеству объектов, количество выходных нейронов – количеству свойств и характеристик объекта, а в промежуточных слоях учитываются природа и характер связей нейронов;
- учёт возможности эволюционной трансформации НС путём дополнения новыми знаниями и адаптации модели к обучаемым данным до тех пор, пока она не будет полностью настроена;
- доказать для реализации адаптивной системы возможность использования модуля первоначальной настройки, обеспечивающего возврат к определённой модели, характерной исследуемому объекту;
- доказать возможность выполнения объективных сравнений предсказательных и аппроксимационных способностей различных моделей НС.

2. МОДЕЛИРОВАНИЕ ПРЕДМЕТНОЙ ОБЛАСТИ

На концептуальном уровне в задачах распознавания изображений микрообъектов, параметров технологических процессов, результатов стендовых испытаний предложим

представление информации в виде последовательности отсчётов координатной сетки [9]

$$x_{11}, x_{12}, \dots, x_{1n}, x_{21}, \dots, x_{2n}, \dots, x_{mn}$$

где x_{ij} – элемент матрицы отсчётов, расположенный на пересечении i -й строки и j -го столбца; m – число строк матрицы, соответствующее числу строк в кадре; n – число столбцов матрицы, соответствующее числу элементов в строке. На рис. 1-а) и 1-б) в качестве характерного примера проиллюстрированы принцип дискретизации изображения микрообъекта (пыльцевого зерна) и процедуры определения разброса проквантованных координат изображения микрообъекта, представляемого для распознавания.

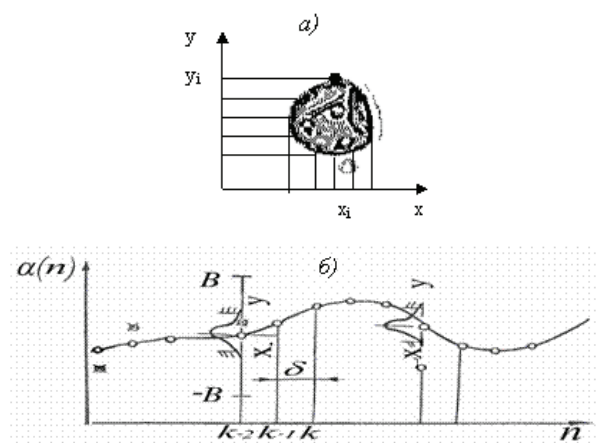


Рис. 1.

а). Представление изображения микрообъекта в декартовой системе

б). Принцип ограничения разброса координаты изображения

Как видно из рисунка 1-б), входные данные, вводимые для решения задач распознавания изображений микрообъектов, представляют информацию о динамическом процессе, которые обусловлены влиянием помех, шума или ошибок передачи, что является причинами нестационарности и слабой формализуемости. В связи с этим, для достижения наилучшего качества выхода динамического процесса представляется целесообразной разработка новых способов, правил и процедур, направленных на получение методики использования НС.

Изложим результаты исследований, направленных на определение математической модели формального нейрона.

3. МАТЕМАТИЧЕСКАЯ МОДЕЛЬ НЕЙРОНА

Следует отметить, что несмотря на большое разнообразие вариантов НС, все они имеют общие черты. Каждая НС состоит из однотипных элементов, имитирующих работу нейронов. Искусственный нейрон характеризуется своим текущим состоянием, которое может быть возбуждённым или заторможенным. Он имеет несколько входов (дендриты) и один выход (аксон). Также он обладает группой синапсов – однонаправленных входных связей, соединённых с выходами других нейронов. По аксону сигнал возбуждения или торможения поступает на синапсы следующих нейронов. Каждому входу ставится в соответствие некоторый весовой коэффициент w , характеризующий величину

синаптической связи и оценивающий степень влияния сигнала с этого входа на сигнал на выходе [10].

Основные отличия проявляются в типах моделей нейронов. Существуют логические, непрерывные и импульсные модели нейрона [11]. Исследуемые нами проблемные задачи в основном связаны с аппроксимацией непрерывных функций, в связи с этим рассмотрим непрерывный тип моделей искусственного нейрона. Он характеризуется своим весовым коэффициентом (синаптическим весом), который может принимать как положительные, так и отрицательные значения. Взвешенные входные сигналы подаются на линейный сумматор, после чего результат их сложения поступает на блок активационной функции. Модель нейрона также включает в себя сдвиг b , который добавляется к входному сигналу блока активационной функции.

На рисунке 2 изображена функциональная схема такой модели.

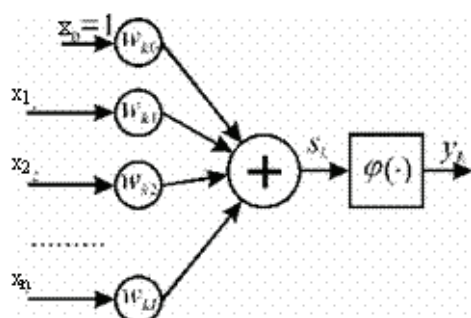


Рис. 2. Непрерывная модель нейрона

Если весовые коэффициенты синапсов одного слоя нейронов свести в матрицу W , состоящую из n строк и m столбцов и элементы w_{ij} которой задают величину синаптической связи i -го входа с j -ым нейроном, то процесс, происходящий в нейронной сети, можно записать в матричной форме:

$$Y = F(XW)$$

где X – входной вектор, Y – выходной вектор.

Математически модель нейрона описывается следующими зависимостями:

$$v_k = \sum_{j=1}^l w_{jk} x_j; \quad (1)$$

$$s_k = v_k + b_k; \quad (2)$$

$$y_k = \sigma(s_k), \quad (3)$$

где $x_1, x_2, \dots, x_l$ – входные сигналы нейрона; $w_{1k}, w_{2k}, \dots, w_{lk}$ – синаптические веса k -го нейрона; v_k – выход линейного сумматора; b_k – сдвиг; $\sigma(s_k)$ – активационная функция; y_k – выходной сигнал нейрона.

Для упрощённого представления записей (1-3) введём дополнительный входной сигнал $x_0 = 1$ и вес $w_{0k} = b_k$. Тогда (2) примет вид:

$$s_k = \sum_{j=0}^l w_{jk} x_j, \quad (4)$$

Конкретная процедура получения зависимости (3) определяется видом функции $\sigma(\cdot)$ и видом функционала ошибки и зависит от применяемой архитектуры НС.

Дальнейшее исследование построения методики использования НС в системе обработки данных нестационарной природы связано с решением следующих задач: разработка принципа организации работы НС, выбор архитектуры НС, определение правил обучения и коррекции весовых коэффициентов, выбор активационной функции, оценка качества обучения, построение алгоритмов обучения, реализация моделей НС в системах визуализации изображений и распознавания микрообъектов.

4. ОРГАНИЗАЦИЯ СТРУКТУРЫ НС

В качестве начального варианта структуры нейронной сети предлагается конструкция многослойной НС с заданным количеством входных нейронов, соединённых с нейронами скрытых слоёв нелинейной функцией, и один выходной нейрон.

Для распознавания изображений микрообъектов предложим организовать работу НС на основе трёх слоёв: входного, признакового и эффекторного [12]. Основное условие работы модели многослойной НС заключается в том, чтобы каждый нейрон последующего слоя был связан со всеми нейронами предыдущего слоя, причём важно определение активационной функции между слоями. Нелинейность активационной функции является ключевым моментом при увеличении мощности НС.

На рисунке 3 проиллюстрирована структура организации работы такой НС. Следует отметить, что рассматриваемая модель НС представляет собой один из вариантов самоорганизующейся системы обучения. Опишем функции каждого слоя, задачи, возникающие в них, и решения.

Слой 1 – Входной слой. Этот слой получает переменные x_1, x_n непосредственно от входов нейросети. Основной задачей входного слоя является формирование активного нейрона или нейронных ансамблей для каждого класса входных векторов, которые передаются на последующие уровни НС.

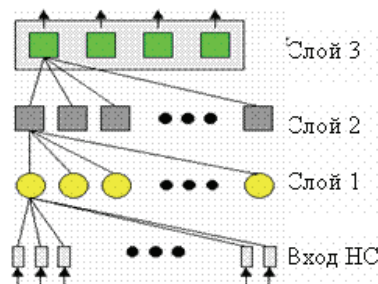


Рис. 3. Структура НС

Веса активных нейронов устанавливаются по правилу

$$w_n = w_c + dw(x - w_c)a, \quad (5)$$

где w_n – новое значение веса; w_c – старое значение; dw – скорость поворота, причём $dw < 1$; x – нормированный входной вектор; a – активность нейрона.

Геометрически это правило иллюстрирует рис. 4.

Здесь важным параметром является скорость поворота dw , и его постепенное уменьшение позволяет произвести статистическое усреднение входных векторов, на которые реагирует данный нейрон, что характерно для рассматриваемой задачи распознавания изображений микрообъекта.

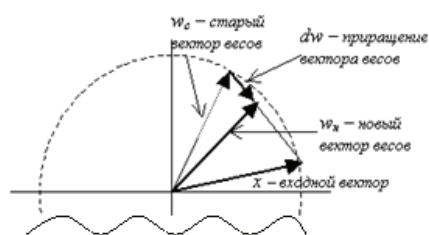


Рис. 4. Коррекция весов нейрона

Слой 2 – Признаковый слой. В данном слое алгоритмом обучения НС вырабатываются признаки распознаваемого объекта, соответствующие ссылающейся базе данных, с которыми ассоциируется соответствующий входной нейрон. Для нейрона этого слоя вычисляется суммарный взвешенный сигнал на входах, и при помощи функции активации он передаётся на выход. Предлагается модификация весов связей при обучении по следующему правилу:

$$w_{ijn} = w_{ijc} + \delta(y_i - w_{ijc})x_i, \quad (6)$$

где w_{ijn} , w_{ijc} – веса связей до и после модификации; через δ обозначено приращение dw , которое идентифицирует скорость обучения, $\delta < 1$; y_i – выход нейрона; x_i – вход нейрона.

По этому правилу модифицируются связи только от активных в данный момент нейронов входного слоя и вектор весов связей стремится к выходному вектору. Выходы y признакового слоя, как правило, являются бинарными, т.е. нейрон может быть активен, тогда $y_i = 1$, или неактивен, тогда $y_i = 0$.

Входной слой совместно с признаковым слоем позволяют сопоставить каждому классу входных сигналов определённый образец микрообъекта.

Слой 3 – Эффекторный слой. Этот слой определяет эффективность обработки изображений, которые в НС представляются в виде сигналов. Слой получает сигналы от признакового слоя и на выходе слоя формируется вектор эффекторов, активность которых задаёт параметры НС. Здесь каждому нейрону признакового слоя, а следовательно, и каждому образцу микрообъекта сопоставляется некоторый вектор эффекторов. Моделирование этого слоя требует особого внимания, и обучение НС в данном слое в зависимости от заданной архитектуры является самостоятельным вопросом исследования. Исследование начнём с математических принципов обучения НС.

5. МАТЕМАТИЧЕСКИЕ ПРИНЦИПЫ ОБУЧЕНИЯ НС

Выше было отмечено, что основным элементом НС является формальный нейрон, который представляет собой многовыходной нелинейный преобразователь с адаптивным взвешиванием входных сигналов. Все нейроны сети соединены между собой синаптическими связями в определённую архитектуру.

Задача обучения НС заключается в «запоминании» образов, предъявляемых на вход, и в выдаче на их основе желаемого множества выходов. Оно осуществляется путём подстройки весовых коэффициентов одновременно при предъявлении входных векторов.

Существующие приёмы обучения НС могут быть классифицированы на «алгоритмы обучения с учителем» и «алгоритмы обучения без учителя». В созданном варианте

системы обработки данных нестационарной природы нами реализованы процедуры обучения без учителя. Можно показать, что такая модель является более эффективной в адаптивных системах, чем обучение с учителем.

В случае обучения без учителя выходы нейронной сети формируются самостоятельно, а весовые коэффициенты изменяются по алгоритму, учитывающему только входные и производные от них сигналы. Обучающая выборка состоит только из входных векторов. Алгоритм обучения должен подстраивать веса сети так, чтобы получались согласованные выходные векторы, то есть, чтобы предъявление достаточно близких входных векторов давало одинаковые выходы. В связи с этим нами выработаны процедуры выделения статистических свойств обучающей выборки, позволяющие группировать сходные векторы в классы на основе учёта свойств стационарности, квазистационарности и нестационарности. Получая на входе вектор из данного класса, сеть даёт вектор на выходе с качественной фильтрацией случайных явлений.

Согласно предложенной структуре организации НС выходной сигнал нейрона эффекторного слоя запишем в виде

$$y_j = \Gamma_j \sigma(\gamma_j W_j^T x), \quad (7)$$

где Γ_j – коэффициент усиления, определяющий максимальные и минимальные значения выходного сигнала; $\sigma(\cdot)$ – функция активации; γ_j – параметр, задающий «крутизну» функции активации;

$W_j^T = (\theta_j, w_{j1}, w_{j2}, \dots, w_{jn})^T$ – настраиваемые синаптические веса j -го нейрона сети; θ_j – сигнал смещения; $x = x_1, x_2, \dots, x_n$ – входные сигналы.

В [13] доказано, что в НС с одним скрытым слоем выходной нейрон может с любой наперёд заданной точностью аппроксимировать произвольную функцию. В связи с этим, выходной сигнал рассматриваемой сети, состоящей из n нейронов, перепишем в виде

$$y = \sum_{j=1}^n \Gamma_j \sigma(\gamma_j W_j^T x) \approx f(x),$$

т.е. он аппроксимирует любую непрерывную функцию $f(x)$, обеспечивая условие близости для всех возможных входов x , принадлежащих некоторому гиперкубу.

Отметим, что аппроксимирующие свойства конкретного нейрона в значительной степени зависят от выбора формы функции $\sigma(\cdot)$, которая должна отвечать следующим традиционным требованиям:

- это должна быть сколь угодно сложная функция, построенная из простых элементов;
- выбор конкретной структуры функции должен сводить задачу к подбору конечного множества параметров;
- процедура подбора параметров должна быть устойчива к ошибкам наблюдений и вычислительным погрешностям.

Активационная функция может иметь различный вид. Наиболее часто для этой цели используются: линейная, пороговая функции, гиперболический тангенс и нелинейная функция с насыщением, которую часто называют логистической функцией или сигмоидом.

Сигмоидальная функция дифференцируема на всей оси абсцисс. Центральная область этой функции имеет боль-

шой коэффициент усиления и поэтому позволяет решить проблему обработки слабых сигналов, в то время как области с падающим усилением на положительном и отрицательном концах подходят для больших возбуждений. Кроме того, при обработке непрерывной информации на основе НС важным условием является то, чтобы в результате нейрон функционировал с большим усилением в широком диапазоне уровня входного сигнала.

Гиперболический тангенс также часто используется в качестве активационной функции. В отличие от сигмоидальной функции, она принимает значения различных знаков, что оказывается выгодным для ряда сетей. Нами в работе предложено использовать эту функцию в качестве математической модели активации эффекторов.

В случае сигмоидальной функции, $\sigma(\cdot)$ определяется как

$$0 < \sigma(\gamma u) = (1 + e^{-\gamma u})^{-1} < 1. \quad (8)$$

Она определена на множестве всех действительных чисел и принимает только положительные значения. Следует отметить, что активационная функция сигмоида является униполярной непрерывной.

В случае гиперболического тангенса функция активации является биполярной функцией и задаётся в виде

$$-1 < \tanh(\gamma u) = \frac{1 - e^{-2\gamma u}}{1 + e^{-2\gamma u}} < 1.$$

Она связана с униполярной сигмоидой соотношением

$$\sigma(\gamma u) = \frac{1}{2} \left(\tanh \left(\frac{\gamma u}{2} \right) + 1 \right).$$

Наряду с этим, в качестве возможных функций активации нейрона (7), задавая ограничения на квадрате $-1 \leq u_j \leq 1$, $-1 < \gamma_j < 1$, мы исследовали следующие функции:

$$\sigma^1(\gamma u) = \tanh(\gamma u) = \frac{1 - e^{-2\gamma u}}{1 + e^{-2\gamma u}}, \quad \Gamma < \frac{1}{\tanh \gamma}; \quad (9)$$

$$\sigma^2(\gamma u) = \frac{\gamma u}{\sqrt{1 + \gamma^2 u^2}}, \quad \Gamma^2 < \frac{\sqrt{1 + \gamma^2}}{\gamma}; \quad (10)$$

$$\sigma^3(\gamma u) = \sin \left(\frac{\pi}{2} \gamma u \right), \quad \Gamma^3 < \frac{1}{\sin \left(\frac{\pi}{2} \gamma \right)}; \quad (11)$$

$$\sigma^4(\gamma u) = \frac{2}{\pi} \arctg(\gamma u), \quad \Gamma^4 < \frac{\pi}{2 \arctg \gamma}; \quad (12)$$

$$\sigma^5(\gamma u) = \gamma u - \frac{\gamma^3}{3} u^3, \quad \Gamma^5 < \frac{3}{3\gamma - \gamma^3}. \quad (13)$$

В качестве примеров построены графики зависимостей данных для функций (9)–(13). На рис. 5 проиллюстрированы кривые, соответствующие функциям:

$$\sigma^1(\gamma u) \text{ с } \gamma = 3,5; \sigma^2(\gamma u) \text{ с } \gamma = 4;$$

$$\sigma^3(\gamma u) \text{ с } \gamma = 1; \sigma^4(\gamma u) \text{ с } \gamma = 6;$$

$$\sigma^5(\gamma u) \text{ с } \gamma = 1.$$

Определено, что конкретный вид функции активации зависит от параметра γ_j и соответствующим выбором этого параметра можно добиться их идентичности.

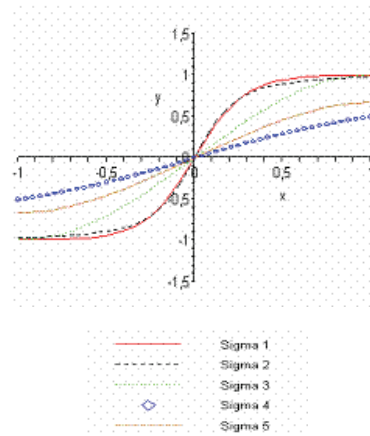


Рис. 5. Графики функций активации

Это позволяет утверждать, что представленные функции активации упрощают процесс обучения и дают возможность ввести унифицированную схему обучения формального нейрона для широкого класса задач обработки непрерывной по природе информации, в частности, для задачи распознавания изображений микрообъектов.

При реализации системы нами исследованы и разработаны алгоритмы обучения по следующим методам: с обратным распространением ошибки, покоординатный спуск, случайный поиск, метод Ньютона, статистический градиентный метод, комбинации этих методов, которые подробно изложены в литературе, посвящённой теории нейронных сетей.

Исходя из структуры функционирования НС можно заметить, что наряду с определением принципов обучения и выбором активационной функции, также немаловажно определение вида выходного нейрона НС. В связи с этим изложим результаты исследований моделей построения НС для различных архитектур.

6. МОДЕЛИ ПОСТРОЕНИЯ НС ДЛЯ РАЗЛИЧНЫХ АРХИТЕКТУР

6.1. Многослойная НС прямого распространения

Однослойная НС прямого распространения описывается рекуррентными соотношениями

$$y_i = W_i x_i, \quad (14)$$

$$x_i = \varphi_i(y_{i-1}), \quad (15)$$

где x_i — вход сети; $y_i = f(x, w)$ — выход, W_i — матрица весов, φ_i — активационная функция, которая действует покоординатно.

При обучении такой сети необходимо обращать внимание на поиск вектора W , являющегося совокупностью матриц W_i , для которого суммарная ошибка по всем примерам (опытам) минимальна

$$E(w) = \sum_{n=1}^N e_n = \min, \quad (16)$$

где e_n — ошибка для n -го примера.

Формулы (14), (15) позволяют легко вычислить градиент обычного функционала ошибки по весам сети с помощью метода наименьших квадратов и применить также для обучения НС один из методов нелинейной оптимизации, которая основана на вычислении производных второго порядка.

Заменив выражение (14) выражением

$$y_l = \sum_{i=0}^l W_{l,i} x_i, \quad (17)$$

где $W_{l,i}$ – матрица весов l -го слоя, получаем модель многослойной НС прямого распространения, применение которой эффективно для экстраполяции и для моделирования процессов со скачками.

Нами рассмотрены процедуры их модификации, работающие на уровне отдельных связей. В связи с этим, предложено использовать НС прямого распространения с частичной структурой связей.

6.2. Модель НС с частичной структурой связи

Особенностью такой модели НС является следующее: НС представляется потоковым графом, вершины которого соответствуют нейронам, а также входам и выходам сети, а дуги – связям. Пусть V – множество вершин сети, тогда множество дуг D является подмножеством $V \times V$, т.е. каждой дуге соответствует упорядоченная пара вершин (v_i, v_j) – из первой дуга исходит, а во вторую – входит. Каждой вершине v сопоставим активационную функцию φ_v , каждой дуге (u, v) – вес $W_{u,v}$.

Следует отметить, что если $u < v$, то выше рассмотренная НС прямого распространения не имеет циклов, т.е. её вершины всегда можно пронумеровать так, что для каждой дуги и следовательно, V можно считать множеством натуральных чисел.

Теперь обозначим вход нейрона, соответствующего вершине с номером v , через y_v , а его выход – через x_v . Тогда формулы (14), (15) для НС прямого распространения с частичной структурой связей приобретают вид

$$y_v = \sum_{u:(u,v \in D)} w_{u,v} x_u, \quad (18)$$

т.е. суммирование производится по всем входящим в нейрон дугам вида

$$x_v = \varphi_v(y_v). \quad (19)$$

В нередких случаях поставленные вопросы исследования сводятся к решениям задач, связанных с кластеризацией на основе НС Кохонена, которая также представляет интерес для задачи распознавания микрообъектов.

6.3. Модель НС Кохонена

Модель Кохонена относится к самоорганизующейся системе, которая строит свою работу на основе карты признаков, причём самоорганизующаяся карта признаков проходит два этапа обучения. Следует отметить, что в отличие от описанных ранее сетей прямого распространения, сети Кохонена не требуют задания выходов сети на обучающем множестве, т.е. обучение проходит только по входам сети.

В решённой нами задаче распознавания микрообъектов на первом этапе обучения элементы упорядочены так, что-

бы отражать пространство входных элементов, а на втором происходит уточнение их позиций. Входные векторы выбираются на основе однородного распределения в некоторой сфере в зависимости от формы объекта, и начинается обучение карты. В ходе обучения строятся изображения карты путём использования принципа соответствия. Элементы соединяются линиями, чтобы показать их относительное размещение. Конечным результатом обучения является карта, покрывающая всё входное пространство и являющаяся достаточно регулярной, т.е. элементы оказываются распределёнными почти равномерно.

Построены различные карты изображений для пыльцевых зёрен. Например, для пыльцы *Alnus* в качестве входных данных было принято 250 точек, взятых из единичной окружности, и проведено её обучение, которое начиналось со случайного набора весовых значений, задающих размещение кластерных элементов в центре входного пространства. Доказано, что качество результата обучения зависит от количества экспериментальных данных, условий предварительной обработки, а также от выбора параметров обучения.

Кроме того, определено, что в случае решения задач прогнозирования, сети Кохонена можно также реализовать путём построения нелинейной регрессии. Простейший способ сделать это – подавать на выход сети не номер выигравшего нейрона, а линейные комбинации выходов слоя Кохонена. Тогда придём к сети Гроссберга.

6.4. Модель НС Гроссберга

Обычно сеть Гроссберга обучается намного быстрее, чем сеть Кохонена, и удачно используется при аппроксимации нестационарного процесса. Однако её недостаток состоит в том, что для получения требуемой точности при обучении НС требуется намного больше нейронов.

Если представить выход сети в виде

$$y = \sum_{j=1}^m w_j \varphi_j(x - c_j), \quad (20)$$

где c_j – некоторые центры, вокруг которых происходит аппроксимация, тогда получаем НС радиальных базисных функций (RBF-сети). Отметим, что применение такой НС послужило удачным инструментом при распознавании микрообъектов на основе интерполяции координат изображений.

В связи с тем, что предметная область исследований представляет нестационарный процесс, целесообразно рассмотреть также динамических НС.

6.5. Многослойные НС с временными задержками

Для такой сети формулы (14) и (15) заменяются на соотношения

$$y_l(n) = \sum_{t=0}^{p_l} W_l(t) x_l(n-t), \quad (21)$$

$$x_l(n) = \varphi_l(y_{l-1}(n)). \quad (22)$$

Здесь на вход нейрона могут подаваться линейные комбинации выходов всех предыдущих слоёв, а не только непосредственно предшествующего.

Для распознавания изображений микрообъектов нами в НС с временными задержками каждому временному отсчёту сопоставлен свой экземпляр множества вершин V . Множество дуг D принято подмножеством $V \times U \times V \times U$, при этом нейроны можно пронумеровать так, чтобы связи шли от нейрона с меньшим номером к нейрону с большим номером, т.е. $u < v$ для каждой дуги. Дуге соответствует упорядоченная пара вершин $u(n-t)$, $v(n)$, $t = 0, 1, 2, \dots$. Каждой вершине $v(n)$ сопоставлена активационная функция φ_v , каждой дуге $u(n-t)$, $v(n)$ – вес $w_{u,v}(t)$.

Обозначив $y_v(n)$ вход соответствующего нейрона, а $x_v(n)$ – его выход, формулы (18), (10) переписываются в виде

$$y_v(n) = \sum_{u:(u,v) \in D} w_{u,v}(t)x_u(n-t), \quad (23)$$

$$x_v(n) = \varphi_v(y_v(n)), \quad (24)$$

т.е. суммирование производится по всем входящим в нейрон дугам.

6.6. Рекуррентные НС

Для рекуррентных сетей, т.е. сетей с обратными связями, формулы (23) – (24) остаются такими же, причём если при $t = 0$ будет существовать как $w_{u,v} \neq 0$, так и $w_{v,u} \neq 0$, тогда возможно вычисление выхода сети при известных входах только с помощью решения системы уравнений (23) – (24).

Для использования такой сети в задачах распознавания микрообъектов мы полагали, что нейроны можно пронумеровать таким образом, чтобы $w_{u,v} \neq 0$ только при $u < v$. Обозначив через $x(n)$ – вектор входов сети, $z(n)$ – вектор внутренних состояний, которые не являются ни входами, ни выходами, $y(n)$ – вектор выходов сети, искомые формулы можно записать в виде:

$$z(n) = \sum_{t=0}^T W_{zx}(t)x(n-t) + \sum_{t=0}^T W_{zz}(t)\varphi(z(n-t)) + \sum_{t=0}^T W_{zy}(t)y(n-t). \quad (25)$$

$$y(n) = \sum_{t=0}^T W_{yx}(t)x(n-t) + \sum_{t=0}^T W_{yz}(t)\varphi(z(n-t)). \quad (26)$$

В качестве частных случаев рекуррентных НС исследовано применение следующих моделей.

6.6.1. НС Хопфилда

Для получения выходного нейрона равенства (25) и (26) перепишем в виде

$$z(n) = Wy(n-1),$$

$$y(n) = \varphi(z(n)).$$

НС Хопфилда эффективно применяются для решения задачи визуализации изображений и распознавания микрообъектов. Однако стандартное применение сетей Хопфилда в случае обработки информации непрерывной природы для решения задач прогнозирования не является достаточно эффективным, потому что оно основано на методе гра-

диентного спуска, который при сложных нестационарных процессах сходится медленно.

6.6.2. НС Хемминга

НС Хемминга является модификацией сетей Хопфилда. Эта архитектура получается, если на входе сети Хопфилда добавить слой нейронов, которые вычисляют меру рассогласования входного вектора с набором обучающих эталонов.

$$x_i^{(n)} = dW_i(y_i^{(n)}).$$

6.6.3. Двухнаправленная ассоциативная память

НС этой архитектуры позволяют модифицировать сеть Хопфилда путём усиления связи между входным и выходным векторами. Для этого установим связи между двумя векторами x и y следующим образом:

$$z_1(n) = Wx(n-1); y(n) = \varphi(z_1(n));$$

$$z_2(n) = W^r y(n); x(n) = \varphi(z_2(n)).$$

6.6.4. Сеть Джордана.

Этот вид сетей получается из многослойной НС, если на его вход помимо входного вектора подать также выходной слой с задержкой на один или несколько тактов. Для задачи распознавания изображений микрообъектов функционирование такой сети опишем соотношениями

$$y_0(n) = W_0 x(n) + W(1)y(n-1),$$

$$y_l(n) = W_l \varphi_l(y_{l-1}(n)),$$

$$y(n) = W_L \varphi_L(y_{L-1}(n)).$$

6.6.5. Сеть Элмана

Так же как и сеть Джордана, эта сеть получается путём введения обратных связей. В данном случае связи идут не от выхода сети, а от выходов внутренних нейронов. Исходя из классического варианта сети нами получены формулы выходного нейрона одного слоя НС следующего вида

$$y_0(n+1) = W_0 x(n) + W(1)\varphi(y_0(n)),$$

$$y(n) = W_l \varphi_l(y_0(n)).$$

Таким образом, во всех полученных моделях выходные нейроны НС позволяют эффективно аппроксимировать функции непрерывного характера, а это значит, что они могут служить базой при построении систем обработки непрерывной по природе информации, в частности, для решения задач распознавания изображений микрообъектов, прогнозирования нестационарных процессов в многоагрегатных технологических системах, а также при построении систем адаптивной контроля точности передаваемой информации.

Однако реализация любой из изложенных моделей НС связана с выбором критериев обучения НС, которые, в свою очередь, зависят от характера решаемой задачи.

7. КРИТЕРИИ КАЧЕСТВА ОБУЧЕНИЯ НС

Исходя из специфики решаемой задачи можно проводить исследования на основе различных критериев оценки качества, которые позволили бы ограничить пространство

допустимых решений и оптимизировать процессы обучения НС. Так, например, при решении ряда задач распознавания объектов динамического процесса широко используется критерий, учитывающий количество ошибок на обучающей выборке, либо критерий «количество (процент) угаданных образцов» [14].

В связи с тем, что в настоящем исследовании наше внимание сосредоточено на решении задач распознавания объектов динамических процессов на основе обработки непрерывной по природе информации, то определено, что в качестве эффективного критерия, оптимизирующего процесс обучения без учителя, целесообразно выбрать построение исследований на основе среднеквадратической ошибки, которую можно представить в виде:

$$E = \sqrt{\frac{\sum_{i=1}^M (d_i - y_i)^2}{M}},$$

где y_i – значение i -го выхода НС; d_i – целевое значение i -го выхода; M – число образцов в обучающем множестве.

Основным условием применения данного критерия качества обучения принято ограничение пространства поиска при обучении за счёт наилучшей идентификации и фильтрации обучающей выборки. Нами подтверждено, что целевую функцию ошибки НС можно представить в виде функционала ошибки, а минимизацию величины E осуществить с помощью градиентного метода. При этом изменение весов происходит в направлении, обратном к направлению наибольшей крутизны для этой функции.

Согласно методу наименьших квадратов, минимизируемой целевой функцией ошибки НС является величина:

$$E(w) = \frac{1}{2} \sum_{j,p} (y_{j,p}^{(N)} - d_{j,p})^2, \quad (27)$$

где $y_{j,p}^{(N)}$ – реальное выходное состояние нейрона j выходного слоя N нейронной сети при подаче на её входы p -го образа; $d_{j,p}$ – идеальное (желаемое) выходное состояние этого нейрона. Суммирование ведётся по всем нейронам выходного слоя и по всем обрабатываемым сетью образам.

Одним из механизмов минимизации целевой функции (27) является подстройка весовых коэффициентов, которая в общем случае осуществляется следующим образом:

$$\Delta w_{ij}^{(n)} = \eta \frac{\partial E}{\partial w_{ij}}, \quad (28)$$

здесь w_{ij} – весовой коэффициент синаптической связи, соединяющей i -ый нейрон слоя $n-1$ с j -ым нейроном слоя n , η – коэффициент скорости обучения, $0 < \eta < 1$. Множитель в правой части формулы (28) представляется в виде

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial y_j} \cdot \frac{\partial y_j}{\partial s_j} \cdot \frac{\partial s_j}{\partial w_{ij}}, \quad (29)$$

где под y_j подразумевается выход нейрона j , а под s_j – взвешенная сумма его входных сигналов, то есть аргумент активационной функции.

Нами определено, что применение функции единичного скачка и других активационных функций с неоднородностями характерно при дискретном процессе и не подхо-

дят для решаемой задачи распознавания объектов динамического непрерывного процесса. В связи с этим, более подходящими являются гладкие функции, такие как классический сигмоид с экспонентой или гиперболический тангенс. Тогда второй множитель определяется как

$$\frac{dy}{ds} = 1 - s^2. \quad (30)$$

Третий множитель формулы (29), очевидно, равен выходу нейрона предыдущего слоя $y_i^{(n-1)}$.

Первый множитель в (29) легко раскладывается в ряд следующим образом:

$$\frac{dE}{dy_j} = \sum_k \frac{dE}{dy_k} \cdot \frac{dy_k}{ds_k} \cdot \frac{\partial s_k}{\partial y_j} = \sum_k \frac{\partial E}{\partial y_k} \cdot \frac{dy_k}{ds_k} \cdot w_{jk}^{(n+1)}. \quad (31)$$

Здесь суммирование по k выполняется среди нейронов слоя $n+1$.

Введя новую переменную

$$\partial_j^{(n)} = \frac{\partial E}{\partial y_j} \cdot \frac{dy_j}{ds_j}, \quad (32)$$

мы получим рекурсивную формулу для расчётов величин $\partial_j^{(n)}$ слоя n из величин $w_{jk}^{(n+1)}$ более старшего слоя $n+1$.

$$\partial_j^{(n)} = \left[\sum_k \partial_k^{(n+1)} w_{jk}^{(n+1)} \right] \frac{dy_j}{ds_j}. \quad (33)$$

Для выходного же слоя

$$\partial_i^{(N)} = (y_i^{(N)} - d_i) \frac{dy_i}{ds_i}. \quad (34)$$

Теперь мы можем записать (28) в виде:

$$\Delta w_{ij}^{(n)} = -\eta \partial_j^{(n)} y_i^{(n-1)}. \quad (35)$$

Для придания процессу коррекции весов некоторой инерционности, сглаживающей резкие скачки при перемещении по поверхности целевой функции, формула (35) дополняется значением изменения веса на предыдущей итерации

$$\Delta w_{ij}^{(n)}(t) = \eta (\mu \Delta w_{ij}^{(n)}(t-1) + (1 - \mu) \partial_j^{(n)} y_i^{(n-1)}), \quad (36)$$

где μ – коэффициент инерционности, t – номер текущей итерации.

Отметим, что полученная методика распространяется на все виды архитектур НС для расчёта и коррекции выходных нейронов.

Далее изложим результаты исследований для построения системы обработки информации при распознавании изображений микрообъектов на примере пылевых зёрен.

8. ПРОГРАММНО-РЕАЛИЗУЕМЫЕ МОДЕЛИ НС

На рис. 6 проиллюстрирована модель функционирования НС в виде блок-схемы, на основе которой нами построена программная система визуализации изображений, распознавания и классификации пылевых зёрен.

Для лучшего обозрения работы НС изложим последовательность действий подсистемы реализации обучения на основе алгоритма с обратным распространением ошибки.

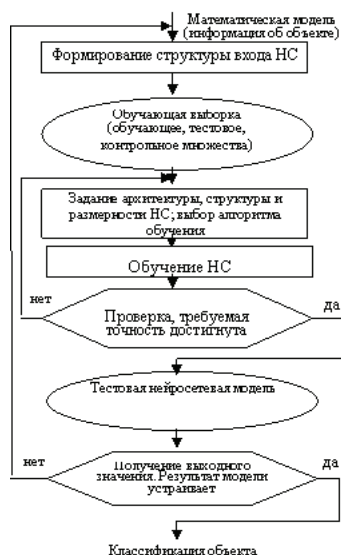


Рис. 6. Модель функционирования НС

Реализация алгоритма обучения состоит из пяти программно-реализуемых модулей.

В модуле 1 на входы сети подаются данные одного из возможных образов и рассчитываются промежуточные значения выходов. В дальнейшем все тренировочные образы предъявляются попеременно в случайном порядке.

Взвешенная сумма входных сигналов рассчитывается по формуле:

$$s_j^{(n)} = \sum_{i=0}^M y_i^{(n-1)} w_{ij}^{(n)},$$

где M – число нейронов в слое $n-1$ с учётом нейрона с постоянным выходным состоянием (+1), задающего смещение; $y_i^{(n-1)} = x_{ij}^{(n)} - i$ -ый вход нейрона j слоя n и

$$y_j^{(n)} = \sigma(s_j^{(n)}).$$

В модуле 2 рассчитываются весовые коэффициенты $w_{ij}^{(n)}$ для выходного слоя по формуле (8). Изменения весов $\Delta w_{ij}^{(n)}$ слоя n рассчитываются по формуле (36).

В модуле 3 по формулам (33) и (36) рассчитываются соответственно $\delta_j^{(n)}$ и $\Delta w_{ij}^{(n)}$ для всех остальных слоёв, $n = n, n-1, \dots, 1$.

В модуле 4 корректируются все веса в НС по формуле

$$w_{ij}^{(n)}(t) = w_{ij}^{(n)}(t-1) + \Delta w_{ij}^{(n)}(t).$$

В модуле 5 проверяется ошибка сети, и если она существенна и не устраивает эксперта НС, то осуществляется итерация путём перехода на шаг 1. Если же ошибка удовлетворяет требованиям эксперта, то работа сети прекращается.

В этом модуле для расширения возможностей эксперта реализованы процедуры проверки по следующим видам ошибок обучения и тестирования:

- средняя ошибка обучения, которая рассчитывается путём деления суммы ошибок распознавания примеров на общее число примеров;
- средняя ошибка тестирования, которая рассчитывается аналогичным образом, только в качестве примеров используются данные, не входящие в обучающее множество;

- максимальная ошибка обучения, которая является максимальной ошибкой, допущенной НС при распознавании обучающих примеров;
- максимальная ошибка тестирования, которая является максимальной ошибкой, допущенной НС при распознавании тестовых примеров.

Теперь изложим основные моменты организации функционирования программной системы визуализации изображений, распознавания и классификации пылевых зёрен, ориентированной на заданную предметную область.

Следует отметить, что при разработке программной системы визуализации изображений и распознавания микрообъектов модели предметной области сформированы нами по принципу «смены сцен». При таком подходе память оперирует не с отдельными отображениями деталей объекта, а с их функциональными единицами – свойствами. Объект наделён следующими свойствами: форма, цвет, компоновка, способность к движению в пространстве как отдельных его деталей, так и всего объекта в целом; способность к вступлению во взаимодействие с другими объектами; способность к выдаче некоторых сигналов во внешнее, по отношению к самому объекту, пространство.

В процессе реконструкции сцен система как бы прокручивает, «запускает» объекты по заданным правилам, и таким образом воспроизводится (моделируется) реальный образец. Правомомерность этого подхода хорошо стыкуется с широко распространённым в настоящее время внедрением объектно-ориентированного программирования, главными достоинствами которого являются: инкапсуляция, наследование, полиморфизм. Подобный подход к пониманию процессов построения «внутренней» модели микрообъекта кажется более реалистичным, однако, и он позволяет избавиться от специальных способов записи «правил» для объектов, необходимости сложных математических преобразований для их анимации и других.

Информационно-логическая модель работы программной системы осуществлена в четырёх этапах.

Этап I. Работа с данными.

Составить базу данных из примеров, характерных для данной предметной области.

Разбить всю совокупность данных на три подмножества: обучающее, тестовое и подтверждающее.

Этап II. Предварительная обработка.

Выбрать систему признаков объектов предметной области, характерных для данной задачи, и преобразовать данные соответствующим образом для подачи на вход сети; нормировки; стандартизации и т.д. В результате получается линейно отделяемое пространство множества образцов.

Адаптировать систему кодирования изображений, используя методы поэлементного, позиционного, структурного кодирования.

Этап III. Конструирование, обучение и оценка качества сети.

Выбрать архитектуру сети: количество слоёв, число нейронов в слоях, длину обучающих выборок.

Выбрать функцию активации нейронов.

Адаптировать алгоритм обучения сети.

Оценить качество работы сети на основе выбранного критерия и оптимизировать архитектуру (уменьшение весов, прореживание пространства признаков).

Выбрать вариант сети, который обеспечивает наилучшую способность к обобщению. Оценить качество работы по тестовому множеству.

Этап IV. Распознавание и классификация при вводе искусственных помех.

Выяснить степень влияния различных факторов (свойств нестационарности обучающей последовательности и процедур фильтрации) на принимаемое решение. Убедиться, что сеть даёт требуемую точность классификации.

При неудовлетворительных результатах распознавания и классификации вернуться на этап 2, изменить способ представления образов.

При удовлетворительном распознавании и неудовлетворительной классификации вернуться на этап 1 для дополнения и изменения базы данных, т.е. задать пыльце новое имя и образовать новый класс объектов.

9. ЗАКЛЮЧЕНИЕ

Модели и алгоритмы построения адаптивных систем обработки нестационарной по природе информации в соответствии с концепцией Data Mining на основе теории и практики применения НС могут быть использованы при решении широкого спектра задач актуальных предметных областей, связанных с распознаванием изображений микрообъектов, отпечатков пальцев, элементов текстовой информации, а также представляют интерес при решении задач медицинской диагностики, прогнозирования динамических процессов, контроля точности передачи информации.

Исследованные и выработанные теоретико-методологические положения моделирования искусственного нейрона, процессов адаптивного обучения распространяются на различные архитектуры НС. Полученные формулы служат для определения вида и оценки выходного нейрона при различных условиях корректировки активационных функций и моделях обучения.

Алгоритм обучения НС с обратным распространением ошибки служит только примером. На практике реализация программных модулей обучения и логико-информационной модели функционирования программной системы визуализации изображений, распознавания и классификации пыльцевых зёрен расширяется алгоритмами обучения по следующим методам: покоординатный спуск, случайный поиск, метод Ньютона, статистический градиентный метод, а также их комбинации.

ЛИТЕРАТУРА

1. Венкатеш Ганти, Йоханнес Герке, Раджу Рамакришнан. Добыча данных в сверхбольших базах данных. // Открытые системы, 1999, №9–10, с. 45–53.
2. Ф. Уоссермен «Нейрокомпьютерная техника: Теория и практика» -Перевод на русский язык Ю. А. Зуев, В. А. Точенов, 1992. -307 с.
3. Галушкин А. И. Основы нейроуправления // Информационные технологии (Приложение). – М.: Новые технологии, 2002. – № 10.
4. Медведев В. С., Потёмкин В. Г. Нейронные сети. Matlab 6. – М.: Диалог МИФИ, 2002. – 496 с.
5. Осовский С. Нейронные сети для обработки информации. – М.: Финансы и статистика, 2002. – 304 с.
6. Джуманов О.И. Нейросетевая система адаптивной обработки информации нестационарной природы при управлении вузом // В сб. мат. X межд. науч.-метод. конф. вузов и факультетов телекоммуникаций, 16–19 июня 2008 г., Ростов-на-Дону, Россия, с.81-85.
7. Джуманов О.И. Основные подходы к созданию интеллектуальной системы адаптивной обработки и анализа информации для прогнозирования нестационарных процессов // журнал Естественные и технические науки, №3 (35) 2008 г., изд-во «Компания Спутник+», ISSN 1684-2626, Москва, 2008, с. 318–325.
8. Джуманов О.И., Ахатов А.Р. Разработка правил обучения нейросистем для модификации алгоритмов распознавания изображений элементов текста// Узбекский журнал «Проблемы информатики и энергетики», №4.2007, Изд. «Фан» АН РУз, Ташкент, 2007, с. 28–31.
9. Джуманов О.И. Структурное кодирование информации в системах визуализации изображений //«Химическая технология, контроль и управление», № 3, ТГТУ, Ташкент, 2006, с. 47–51.
10. Э.М.Куссуль «Ассоциативные нейроподобные структуры» – Киев: Наукова думка, 1990. – 240 с.
11. Н.М. Амосов и др. «Нейрокомпьютеры и интеллектуальные роботы» – Киев: Наукова думка, 1991. – 237 с.
12. Джуманов О.И. Системы распознавания микрообъектов на основе нейронных сетей // В сб. «Вопросы кибернетики», РИСО АН РУз, № 177, Ташкент, 2008 г, с. 122–132.
13. Djumanov O.I., Ulugbek Kh. Narzullaev, Akmal R. Akhatov, Construction of Computer System for Microobjects Recognition Based on Neural Networks//10th International Workshop, CASC 2007, Bonn, Germany, September 16–20, 2007. Lecture Notes in Computer Science, Springer Berlin / Heidelberg, 0302-9743 (Print) 1611-3349 (Online), Volume 4770/2007, 10.1007/978-3-540-75187-8, 2007, p. 328–338
14. Banarse D. S. and A. W. G. Duller. Deformation invariant pattern classification for recognising hand gestures//In Proceedings of the 1996 IEEE International Conference on Neural Networks, 1996, p. 1812–1816.

Джуманов Олимжан Исраилович

к.т.н., доцент кафедры информационных технологий Самаркандского государственного университета, (703004, Узбекистан, Самарканд, Университетский бульвар, 15) тел. (8 366) 264-5510, e-mail: akmalar@rambler.ru