

МЕТОДИКА ЛОКАЛИЗАЦИИ ИЗОБРАЖЕНИЯ ЛИЦА ДЛЯ СИСТЕМ ВИДЕОКОНТРОЛЯ НА ОСНОВЕ НЕЙРОННОЙ СЕТИ

А.А. Макаренко, В.Т. Калайда

Томский государственный университет систем управления и радиоэлектроники
E-mail: makaleks@mail.ru

Предлагается метод и алгоритм локализации лица человека для автоматизированных систем распознавания и видеоконтроля на базе сверточных нейронных сетей. Преимущество использования свёрточных нейронных сетей в том, что они обеспечивают устойчивость к изменениям масштаба, смещениям, поворотам, смене ракурса и другим искажениям входного изображения. Обосновывается топология используемой нейронной сети и методика ее обучения.

Нейронные сети успешно применяют в решении многих проблем распознавания образов [1–3]: распознавание символов, распознавание объектов, и многих других. Проблема обнаружения образа лица очень трудна из-за большого разнообразия искажений, таких как различное выражение лица, условия съемки и т. д. Преимущество использования нейронных сетей для обнаружения лица — обучаемость системы для выделения ключевых характеристик лица из обучающих выборок.

В настоящее время наиболее часто в задачах распознавания и идентификации изображений используют классические нейросетевые архитектуры (многослойный перцептрон, сети с радиально-базисной функцией и др.), но, как показывает анализ данных работ, применение классических нейросетевых архитектур к данной задаче является неэффективным по следующим причинам:

- к данной задаче обычно применяется ансамбль нейронных сетей (2–3 нейронные сети, обученные с различными начальными значениями синоптических коэффициентов и порядком предъявления образов), что отрицательно сказывается на вычислительной сложности решения задачи и соответственно на времени выполнения;
- как правило, классические нейросетевые архитектуры используются в совокупности с вспомогательными методами выделения сюжетной части изображения (сегментация по цвету кожи, выделение контуров и т. д.), которые требуют качественной и кропотливой предобработки обучающих и рабочих данных, что не является эффективным;
- нейросетевые архитектуры являются крайне чувствительными к влиянию различных внешних факторов (изменения условий съемки, присутствие индивидуальных особенностей на изображении, изменение ориентации).

Дополнительно возникают трудности применения традиционных нейронных сетей к реальным задачам распознавания и классификации изображений.

Во-первых, как правило, изображения имеют большую размерность, соответственно вырастает размер нейронной сети (количество нейронов и т. п.). Большое количество параметров увеличивает

вместимость системы и соответственно требует большей обучающей выборки, что увеличивает время и вычислительную сложность процесса обучения.

Во-вторых, недостаток полностью связной архитектуры — то, что топология ввода полностью игнорируется. Входные переменные могут быть представлены в любом порядке, не затрагивая цель обучения. Напротив, изображения имеют строгую 2-мерную местную структуру: переменные (пиксели), которые являются пространственно соседними, чрезвычайно зависимы.

От данных недостатков свободны так называемые свёрточные нейронные сети. Свёрточные нейронные сети обеспечивают частичную устойчивость к изменениям масштаба, смещениям, поворотам, смене ракурса и другим искажениям. Свёрточные нейронные сети объединяют три архитектурных идеи, для обеспечения инвариантности к изменению масштаба, повороту, сдвигу и пространственным искажениям:

- локальные рецепторные поля (обеспечивают локальную двумерную связность нейронов);
- общие веса (обеспечивают детектирование некоторых черт в любом месте изображения и уменьшают общее число весовых коэффициентов);
- иерархическая организация с пространственными подвыборками.

Топология нейронной сети, используемой в работе, изображена на рис. 1.

Свёрточная нейронная сеть является многослойной. Используются слои двух типов: свёрточные и подвыборочные. Свёрточные и подвыборочные слои чередуются друг с другом. В свою очередь, каждый из этих слоёв состоит из набора плоскостей, причём нейроны одной плоскости имеют одинаковые веса (так называемые общие веса), ведущие ко всем локальным участкам предыдущего слоя (как в зрительной коре человека). Изображение предыдущего слоя сканируется небольшим окном и пропускается сквозь набор весов, а результат отображается на соответствующий нейрон текущего слоя. Таким образом, набор плоскостей представляет собой карты характеристик, и каждая плоскость находит «свои» участки изображения в любом месте предыдущего слоя.

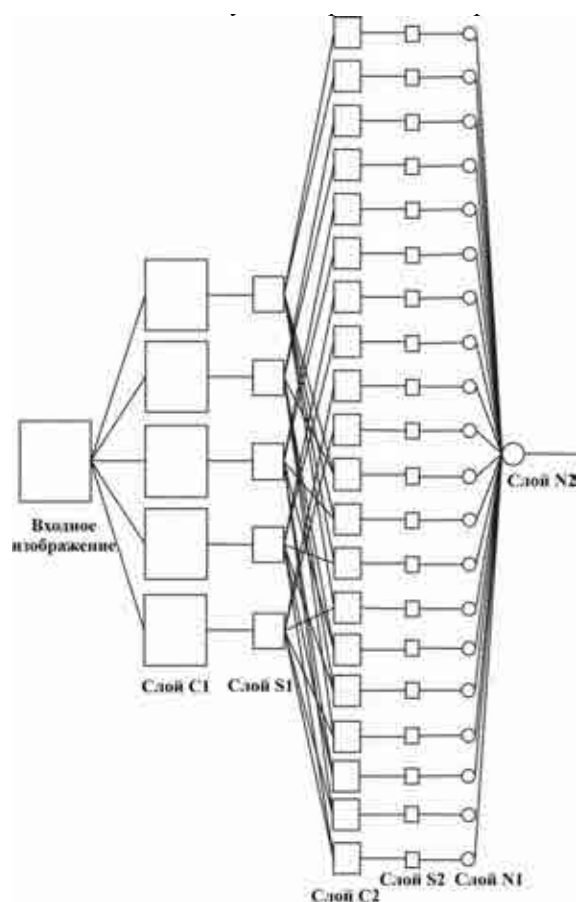


Рис. 1. Схема используемой нейронной сети

Используемая в работе нейронная сеть состоит из шести слоев. Входными данными нейронной сети являются полутоновые изображения размером 32×36 пикселей, которые классифицируются как лицо или «нелицо». Так как задача, решаемая нейронной сетью, – классификация, то для ее решения достаточно одного выхода. Выходное значение нейронной сети находится в интервале $[-1; 1]$, что соответственно означает отсутствие или присутствие лица на классифицируемом изображении.

Входной слой размером 32×36 нейронов не несет какой-либо функциональной нагрузки и служит лишь для подачи входного образа в нейронную сеть. Следом за входным слоем находится сверточный слой C1. Каждый нейрон в плоскости сверточного слоя получает свои входы от некоторой области предыдущего слоя (локальное рецептивное поле), то есть входное изображение предыдущего слоя как бы сканируется небольшим окном и пропускается сквозь набор весов, а результат отображается на соответствующий нейрон сверточного слоя.

Процесс функционирования нейрона сверточного слоя задается выражением:

$$y_k^{(i,j)} = b_k + \sum_{s=1}^K \sum_{t=1}^K w_{k,s,t} x^{((i-1)+s, (j-1)+t)},$$

где $y_k^{(i,j)}$ – нейрон k -ой плоскости сверточного слоя, b_k – нейронное смещение k -ой плоскости, K – раз-

мер рецептивной области нейрона, $w_{k,s,t}$ – элемент матрицы синаптических коэффициентов, x – выходы нейронов предыдущего слоя.

Слой C1 состоит из 5 сверточных плоскостей и выполняет свертывание входного изображения с помощью синаптической маски размером 5×5 , таким образом, слой C1 осуществляет 5 свёрток входного изображения.

Размер сверточной плоскости определяется в соответствии со следующими выражениями:

$$w_c = w_e - K + 1,$$

$$h_c = h_e - K + 1,$$

где w_c, h_c – ширина и высота сверточной плоскости соответственно, w, h – ширина и высота плоскости предыдущего слоя, K – ширина (высота) окна сканирования.

Исходя из выражений для расчета размера сверточной плоскости имеем, что размер плоскости сверточного слоя C1 – 28×32 нейрона. Нейроны в слое организованы в плоскости, в пределах которых все нейроны имеют один и тот же набор синаптических коэффициентов. Набор выходных сигналов в такой плоскости называют картой характеристик. Полный сверточный слой составлен из нескольких карт характеристик с различными наборами синапсов так, чтобы множественные характеристики могли быть извлечены в каждом местоположении. Таким образом, набор плоскостей представляет собой карты характеристик, и каждая плоскость находит «свои» участки изображения в любом месте предыдущего слоя.

Как указано выше каждая плоскость слоя C1 имеет собственную синаптическую маску и нейронное смещение, рецептивные области нейронов пересекаются, нейроны извлекают одни и те же особенности входного изображения, независимо от их точного местоположения. Таким образом, слой C1 имеет всего лишь 130 настраиваемых параметров (синапсов).

Следующий за слоем C1 подвыборочный слой S1 состоит из 5 карт характеристик и обеспечивает локальное усреднение и подвыборку. Этот слой также состоит из плоскостей количество плоскостей такое же, как и в предыдущем слое. Рецепторная область каждого нейрона – 2×2 область в соответствующей карте особенностей предыдущего слоя. Каждый нейрон вычисляет среднее его четырех входов, умножает на синаптический коэффициент, добавляет нейронное смещение и передает результат через активационную функцию. Процесс функционирования нейрона подвыборочного слоя задается следующим соотношением:

$$y_k^{(i,j)} = b_k + \frac{1}{4} w_k \sum_{s=1}^2 \sum_{t=1}^2 x^{(((i,j))+s, ((i,j)))}.$$

Затем полученный результат подвыборки передается через активационную функцию. После операции подвыборки, точное местоположения и специфические признаки каждой особенности изо-

бражения становятся менее важными, что дает нейронной сети довольно большую степень инвариантности.

Смежные нейроны в подвыборочном слое имеют непересекающиеся рецептивные области. Следовательно, карта особенности слоя подвыборки имеет половину числа рядов и колонок карты особенности в предыдущем слое. В качестве активационной функции используется гиперболический

$$\text{тангенс } y = 1,7159 \tanh\left(\frac{2}{3}x\right) [4].$$

Каждая плоскость слоя $S1$ связана лишь с одной плоскостью слоя $C1$. Размер каждой плоскости слоя $S1$ – 14×16 нейронов, что вдвое меньше чем размер плоскости предыдущего слоя. Каждая плоскость слоя $S1$ имеет единственный синаптический коэффициент и нейронное смещение, что дает в итоге 10 настраиваемых параметров.

Свёрточный слой $C2$ состоит из 20 плоскостей, слои $S1$ и $C2$ перекрестно связаны. Плоскости слоя $C2$ формируются следующим образом: каждая из 5 плоскостей слоя $S1$ свёрнута 2 различными синаптическими масками 3×3 , обеспечивая 10 плоскостей в $C2$, другие 10 плоскостей $C2$ получены, суммируя результаты 2 свёртываний на каждой возможной паре плоскостей слоя $S1$. Таким образом, сети добавляется способность объединять различные виды характеристик, чтобы составлять новые менее зависящие от искажений входного изображения.

Размер плоскости слоя $C2$ – 12×14 нейронов. Таким образом, данный слой имеет 290 синаптических коэффициентов. Слой $S2$ состоит из 20 плоскостей, размер каждой 6×7 нейронов. Каждая плоскость слоя $S2$ имеет единственный синаптический коэффициент и нейронное смещение, что дает в итоге 40 настраиваемых параметров.

Слои $M1$ и $N2$ содержат простые нейроны. Роль этих слоев состоит в обеспечении классификации, после того, как выполнены извлечение особенностей и сокращение размерности входа. В слое $M1$ находится 20 нейронов (по одному на каждую плоскость слоя $S2$), каждый нейрон полностью связан с каждым нейроном только одной плоскости слоя $S2$, он выполняет взвешенное суммирование своих 42 входов, добавляет нейронное смещение и пропускает результат через активационную функцию. Таким образом, данный слой содержит 860 синаптических коэффициентов.

Единственный нейрон слоя $N2$ полностью связан со всеми нейронами слоя $M1$. Роль этого нейрона в вычислении окончательного результата классификации. Выход этого нейрона используется для классификации входного образа на лица и не лица.

Использование принципа объединения весов дает эффект уменьшения количества настраиваемых параметров нейронной сети. Данная нейронная сеть имеет 1351 синаптический коэффициент.

Способность к обучению является фундаментальным свойством мозга. В контексте искусственных нейронных сетей процесс обучения может рассматриваться как настройка архитектуры сети и весов связей для эффективного выполнения специальной задачи. Процесс функционирования нейронной сети зависит от величин синаптических связей, поэтому, задавшись определенной структурой нейронной сети, отвечающей какой-либо задаче, необходимо найти оптимальные значения всех переменных коэффициентов (некоторые синаптические связи могут быть постоянными). Этот этап называется обучением нейронной сети, и от того, насколько качественно он будет выполнен, зависит способность сети решать поставленные перед ней проблемы во время эксплуатации. В основе всех алгоритмов обучения положен единый принцип – минимизация эмпирической ошибки. Функция ошибки, оценивающая данную конфигурацию сети, задается извне в зависимости от того, какую цель преследует обучение. Но далее сеть начинает постепенно модифицировать свою конфигурацию – состояние всех своих синаптических весов таким образом, чтобы минимизировать эту ошибку.

Для обучения описанной нейронной сети был использован алгоритм обратного распространения ошибки (*back propagation*). Метод был предложен в 1986 г. Румельхартом, Макклеландом и Вильямсом [5].

Обучение сети начинается с предъявления образа и вычисления соответствующей реакции. Сравнение с желаемой реакцией дает возможность изменять веса связей таким образом, чтобы сеть на следующем шаге могла выдавать более точный результат. Обучающее правило обеспечивает настройку весов связей. Информация о выходах сети является исходной для нейронов предыдущих слоев. Эти нейроны могут настраивать веса своих связей для уменьшения погрешности на следующем шаге.

Когда ненастроенной сети предъявляется входной образ, она выдает некоторый случайный выход. Функция ошибки представляет собой разность между текущим выходом сети и идеальным выходом, который необходимо получить. Для успешного обучения сети требуется приблизить выход сети к желаемому выходу, т. е. последовательно уменьшать величину функции ошибки. Это достигается настройкой межнейронных связей. Каждый нейрон в сети имеет свои веса, которые настраиваются, чтобы уменьшить величину функции ошибки. Выражение для коррекции синаптических коэффициентов имеет вид:

$$w_{ij}(t+1) = w_{ij}(t) + \eta \frac{dE}{dw_{ij}}, \quad (1)$$

где $w_{ij}(t)$ и $w_{ij}(t+1)$ – вес связи между i -м и j -м нейронами на текущем и последующем шаге обучения, $\frac{dE}{dw_{ij}}$ – производная функции ошибки, η – параметр скорости обучения.

В основе алгоритма обратного распространения ошибки лежит методика, позволяющая быстро вычислять вектор частных производных (градиент) сложной функции многих переменных, если структура этой функции известна. В качестве такой функции в алгоритме рассматривается функция ошибки сети и учитывается тот факт, что структура функции ошибки сети полностью определяется архитектурой нейронной сети, которая считается известной.

Метод обратного распространения ошибки может быть очень медленным особенно для многослойных сетей, где поверхность целевой функции является неквадратичной, невыпуклой и высоко размерной с множеством локальных минимумов и/или плоских областей.

Чтобы вычислить градиент на каждой итерации, ур. (1) должно быть рассчитано для всего обучающего набора данных. Эта процедура называется пакетным обучением; при таком способе обучения вся обучающая выборка рассматривается до обновления синаптических коэффициентов. Альтернативно, можно использовать стохастическое обучение (онлайн), где отдельный пример $\{Z_i, D_i\}$ выбран (например, случайно) из учебного набора на каждой итерации. Поскольку эта оценка градиента является зашумленной, синаптические коэффициенты, возможно, не перемещаются точно по градиенту на каждой итерации. Как далее будет показано, этот «шум» на каждой итерации может быть полезным. Стохастическое обучение – более предпочтительный метод для алгоритма обратного распространения ошибки: результат достигается обычно намного быстрее, чем при пакетном методе.

Стохастическое обучение часто приводит к лучшим решениям из-за шума в обновлениях весовых коэффициентов. Нелинейные сети обычно имеют множество локальных минимумов различной глубины. Цель обучения состоит в том, чтобы определить местонахождение одного из этих минимумов. Пакетное обучение обнаружит минимум любого бассейна, куда первоначально помещены синаптические коэффициенты. В стохастическом обучении шумовая составляющая в обновлениях может привести к весам, вскакивающим в бассейн другого, возможно более глубокого локального минимума. Результаты экспериментов [6] говорят о том, что затраты на одномерную оптимизацию шага не дают практической пользы, обучение по суммарному градиенту (с оптимизацией шага или без неё) всегда проигрывает стохастическому обучению.

Нелинейные активационные функции – то, что придает нейронным сетям их нелинейные свойства. В данной работе в качестве активационной функции используется гиперболический тангенс. Это обусловлено следующими причинами:

- симметричные активационные функции, типа гиперболического тангенса, обеспечивают более быструю сходимость, чем стандартная логистическая функция;

- данная функция имеет простую и непрерывную первую производную.

В задачах классификации целевые выходы типично являются бинарными (например, ± 1), т. е. целевые выходы устанавливаются по асимптотам активационной функции. Такой подход имеет несколько недостатков.

1. Результат обучения может быть нестабилен. Процесс обучения будет настраивать выход нейронной сети как можно близко к целевым значениям, которые могут быть достигнуты только асимптотически. В результате синаптические коэффициенты (выходного слоя и скрытых слоев) стремятся к большим и большим значениям, в которых производная активационной функции близка к нулю. В результате коррекция синаптических коэффициентов может стать незначительной.
2. В случае если выходы насыщенные, сеть не дает уверенности в правильности классификации. Когда входной образ находится на границе решения, результат классификации сомнителен. Большие значения весовых коэффициентов имеют тенденцию смещать выходное значение к хвостам активационной функции независимо от класса. Таким образом, сеть может предсказать неправильный класс, не давая уверенности в результате. Решение этих проблем состоит в том, чтобы заставить целевые выходы быть в пределах диапазона активационной функции. Установка целевых значений на грани максимума второй производной активационной функции – лучший способ использовать в своих интересах нелинейность, не насыщая активационную функцию. По этой причине используется указанная функция активации. Она имеет максимальную вторую производную в точках ± 1 , которые соответствуют целевым значениям, используемым в задачах классификации.

Начальная инициализация синапсов нейронной сети имеет огромное влияние на количество итераций обучения. От того, насколько удачно выбраны начальные значения синаптических коэффициентов зависит, как долго сеть за счет обучения и подстройки будет искать их оптимальные величины, и найдет ли она их. Начальные значения синаптических коэффициентов могут существенно влиять на процесс обучения. Синаптические коэффициенты должны быть выбраны случайно, но таким способом, чтобы активационная функция прежде всего активизировалась в своей линейной области. Как показано в работе [4] для достижения данного эффекта необходимо согласование между нормализацией входных значений нейронной сети, выбором активационной функции и выбором начальных значений весовых коэффициентов. Таким образом, для выбранной активационной функции и выбранного метода нормализации входных данных в данной работе синаптические коэффициенты инициализировались следующим

образом: значения весовых коэффициентов были выбраны случайным образом из нормального распределения с нулевым средним и стандартным отклонением

$$\sigma_w = \sqrt{m},$$

где m – число связей, входящих в нейрон.

В выражении (1) параметр η по сути является мерой точности обучения сети. Чем он больше, тем более грубым будет следующее уменьшение суммарной ошибки сети. Чем он меньше, тем больше времени сеть будет тратить на обучение и тем более возможно ее попадание в окрестность локального минимума. Интуитивно понятно, что выбор индивидуальной скорости обучения для каждого синаптического коэффициента может ускорить сходимость и улучшить качество решения (в зависимости от формы поверхности функции ошибки, некоторые веса могут требовать небольшой скорости обучения, чтобы избежать расхождения, в то время как другие могут требовать большого значения скорости обучения, чтобы ускорить сходжение алгоритма). Поэтому в данной работе каждому синаптическому коэффициенту дана индивидуальная скорость обучения. Для автоматической подстройки скорости обучения на каждой итерации обучения использовался алгоритм, предложенный в работе [7].

Как входами, так и выходами могут быть совершенно разнородные величины. Очевидно, что результаты нейросетевого моделирования не должны зависеть от единиц измерения этих величин. А именно, чтобы сеть трактовала их значения единообразно, все входные и выходные величин должны быть приведены к единому масштабу. Скорость сходимости обычно быстрее, если среднее значение каждой входной переменной по учебному набору близко к нулю. Поэтому, нужно отмасштабировать входные значения так, чтобы среднее число по учебному набору было близко к нулю [4]. Таким образом, в данной работе все входные значения из интервала $[0; 255]$ (т. е. цвет пиксела от черного до белого) были предварительно отмасштабированы в интервал $[-1; 1]$.

В качестве учебного набора используется большая коллекция изображений, полученных из различных источников. Эта коллекция эффективно охватывает изменчивость и богатство естественных данных, чтобы обучить данную систему для работы в реальных условиях. В данной работе, используется окно размером 32×36 пикселей, содержащее лицо и некоторую часть фона, таким образом, добавляя к входному окну некоторую вспомогательную информацию: границу лица и некоторую часть фона. В процессе извлечения лиц не выполнялось какой-либо нормализации изображений типа выравнивания гистограммы или коррекции яркости. Кроме того, примеры лиц не нормализовывались так, чтобы глаза, рот и другие части лиц всегда оставались примерно на том же самом положении. Кроме того, как упомянуто ранее, сверточная ней-

ронная сеть является устойчивой к изменению масштаба и положения, таким образом, для повышения этой надежности необходимо давать нейронной сети ненормализованные примеры. Кроме того, для создания большого количества примеров и увеличения инвариантности к небольшим вращениям и изменениям в интенсивности, к выше упомянутому набору применялся ряд преобразований, включая отражение, вращение до $\pm 20^\circ$. Некоторые из изображений обучающей выборки представлены на рис. 2.



Рис. 2. Изображения из лицевой обучающей выборки

Сбор представительного набора нелицевых образов более труден, т. к. фактически, любое изображение может принадлежать к классу нелицевых образов. Практическое решение этой проблемы состоит в стратегии самонастройки [8], при которой система многократно переобучается на ряде изображений пейзажа, которые не содержат лиц. Перед началом процедуры самонастройки, был построен начальный учебный набор. Большинство из этих изображений содержит части лиц, поскольку как было замечено в ранних экспериментах, этот вид изображений является серьезным источником ложных сигналов. Некоторые из этих изображений показаны на рис. 3.



Рис. 3. Изображения из стартовой нелицевой обучающей выборки

Процесс обнаружения лица состоит из 3 этапов (рис. 4):

1. Входное изображение последовательно масштабируется, получается пирамида изображений, затем каждое изображение сканируется нейронной сетью, происходит выделение участков-кандидатов.
2. Все участки-кандидаты приводятся к масштабу входного изображения, и затем смежные участки-кандидаты группируются в кластеры, устраняя избыточность.
3. Далее происходит дальнейшая проверка кластеров, выносится решение о принадлежности или не принадлежности каждого участка к классу лиц.

Нами было произведено тестирование данной программной системы на наборе из 120 изображений. На данном тестовом наборе получена точность обнаружения равная 97 %. Исходя из этого, считаем, что предложенная нами топология сверточной нейронной сети обеспечивает необходи-



Рис. 4. Процесс обнаружения лица

мую надежность и инвариантность к искажениям и зашумлениям входного сигнала, что позволяет создать на ее базе систему выделения сюжетной части изображения. Модификации процесса обучения обеспечивают качественный процесс обучения нейронной сети, улучшают ее обобщающие и клас-

сифицирующие способности, позволяют использовать данную программную систему для решения практических задач в системах видеонаблюдения и контроля доступа.

Работа выполнена при поддержке РФФИ, проект № 06-08-00751.

СПИСОК ЛИТЕРАТУРЫ

1. Moody J., Darken C.J. Fast Learning In Networks of Locally Tuned Processing Units // *Neural Computation*. – 1989. – № 1. – P. 281–284.
2. Haddadnia J., Faez K. Human face Recognition Using Radial Basis Function Neural Network // *Proc. of 3rd International Conf. on Human and Computer*. – Aizu, Japan, 6–9 Sept. 2000. – P. 137–142.
3. Lin S.H., Kung S.Y., Lin L.J. Face Recognition/Detection by Probabilistic Decision-Based Neural Network // *IEEE Trans. Neural Networks*. – 1997. – V. 8. – № 1. – P. 114–132.
4. LeCun Y. Generalization and network design strategies // *Proc. of the Intern. Conf. Connectionism in Perspective*. – University of Zurich, 10–13 October 1988. – P. 143–155.
5. Rumelhart D.E., Hinton G.E., Williams R.J. Learning internal representations by error propagation // *Parallel distributed processing*. – 1986. – V. 1. – P. 318–362.
6. Царегородцев В.Г. Общая неэффективность использования суммарного градиента выборки при обучении нейронной сети // *Нейроинформатика и ее приложения: Матер. XII Всеросс. семинара*. – Красноярск, 2004. – С. 145–151.
7. Murata N., Miller K.R., Ziehe A., Amari S. Adaptive on-line learning in changing environments // *Advances in Neural Information Processing Systems*. – 1997. – V. 9. – P. 599–623.
8. Sung K.K. Learning and Example Selection for Object and Pattern Detection: PhD thesis. – MIT AI Lab, 1996. – 150 p. (<ftp://publications.ai.mit.edu/ai-publications/1500-1999/AITR-1572.ps.Z>)