

УДК 004.934

В.Г. Прохоров

РАСПОЗНАВАНИЕ ГРАФИЧЕСКИХ ОБРАЗОВ ТЕКСТОВЫХ СИМВОЛОВ, ПРЕДСТАВЛЕННЫХ В ВИДЕ ХАРАКТЕРИСТИЧЕСКИХ ВЕКТОРОВ

Предложен алгоритм распознавания, который сочетает в себе преимущества алгоритмов теории решений, нейронных сетей и основан на представлении графических образов символов в виде набора характеристик. Предложен набор характеристик, иллюстрирующих распознавание с помощью данного алгоритма и отмечены некоторые аспекты предобработки характеристик. Приведены экспериментальные данные, иллюстрирующие эффективность предложенного алгоритма.

Введение

Одной из актуальных проблем современных информационных технологий в настоящее время остается распознавание образов, в частности графически представленных символов. Для решения данной задачи используется ряд алгоритмов, которые можно условно разделить на три категории:

- алгоритмы, основанные на сравнении шаблонов;
- алгоритмы, основанные на методах теории решений;
- алгоритмы, использующие нейронные сети.

Все вышеперечисленные группы алгоритмов, имеют, в целом, одинаковую структуру и содержат такие блоки:

- блок обучения системы на наборе образов, подобным тем, которые будут распознаваться системой в будущем. По завершению обучения, его результаты сохраняются;
- блок распознавания, который используя информацию, полученную на этапе обучения, относит незнакомый образ к одному из классов с помощью определенного математического аппарата (классификатора).

Следует заметить, что шаблонные методы, в силу своей негибкости (привязанности к шрифту, масштабу, наклону символов) не подходят для общего решения данной задачи, однако в силу своей простоты, небольшой ресурсоемкости, малым вычислительным затратам они, с оп-

ределенными модификациями, используются достаточно часто при решении отдельных задач распознавания [1].

Метод нейронных сетей, использующий некоторые принципы искусственного интеллекта, широко используется в большинстве современных систем распознавания символов. Среди его преимуществ стоит упомянуть универсальность, а также незначительные требования к проектированию алгоритмов в силу особенностей нейросетей. Недостатками данного метода являются значительный объем вычислительных ресурсов, необходимых для организации процесса обучения, а также плохая предсказуемость результатов распознавания. Наибольшим преимуществом таких методов является высокая точность распознавания.

Алгоритмы, основанные на методах теории решений, во много раз упрощают процесс обучения, но их точность значительно уступает точности нейросетевых алгоритмов. Важным преимуществом данных алгоритмов является представление символов в векторной форме (в виде набора характеристик), в то время как большинство нейросетевых алгоритмов оперируют с растровым изображением. При использовании некоторых классификаторов теории решений, возможно добиться высокой степени предсказуемости результатов распознавания.

В настоящей работе рассмотрен класс нейросетевых алгоритмов распознавания,

оснований на використанні теорії рішень для прискорення процесу навчання без суттєвої втрати точності розпізнавання.

Далі будуть детально розглянуті деякі блоки вищеописаних алгоритмів, які будуть використані для побудови альтернативного алгоритму розпізнавання.

1. Розпізнавання символів при використанні методу представлення в формі характеристических векторів теорії рішень

Суть розпізнавання в теорії рішень полягає в аналізі символу, розрахунок його характеристик і встановленні відповідності між оброблюваним символом і одним із еталонних (завідомо обчисленими характеристиками) за рахунок порівняння їх характеристик. Слід зауважити, що мова йде не про однозначне порівняння символів, а лише про знаходження еталону, характеристики якого найбільш точно відповідають характеристикам даного символу (стопроцентне співпадіння характеристик неможливо в зв'язі з різницею шрифтів і почерків, а також в зв'язі з іншими факторами).

Для розуміння нижчеописаних алгоритмів необхідно, перш за все, чітко визначити основні поняття теорії розпізнавання образів. Під образом розуміється деяке упорядковане множинство дескрипторів (або ознак), які описують визначений об'єкт. Класом образів (або просто класом) називається множинство образів, які мають певні загальні властивості. В подальшому, будемо позначати класи символів w_1, w_2, \dots, w_N , де N – кількість класів. Під машинним розпізнаванням образів розуміються методи, які дозволяють відносити образи до певних або інших класів – автоматично або з мінімальним участям людини [2].

Одним із найбільш часто використовуваних способів представлення ознак об'єкта є представлення його в формі векторів ознак, які позначаються

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix}, \quad (1)$$

де кожен із компонентів x_i представляє i -ий дескриптор, а n – загальна кількість дескрипторів, пов'язаних з певним образом. Образи представляються в формі векторів-столбців виду (1), або в еквівалентній формі $x = (x_1, x_2, \dots, x_n)^T$, де T – оператор транспонування.

1.1. Основна задача теорії рішень.

Підхід до задачі розпізнавання образів з позиції теорії рішень оснований на використанні дискримінантних (або вирішувальних) функцій. Нехай $x = (x_1, x_2, \dots, x_n)^T$ – n -вимірний вектор ознак об'єкта. Основна задача розпізнавання в теорії рішень може бути сформульована наступним чином. Припустимо, що існують N класів образів w_1, w_2, \dots, w_N . Потрібно знайти N дискримінантних функцій $d_1(x), d_2(x), \dots, d_N(x)$, таких, що якщо образ x належить класу w_i , то

$$d_i(x) > d_j(x), \quad j = 0, 1, 2, \dots, N; j \neq i. \quad (2)$$

Іншими словами, незнайомий образ x відносять до i -го класу, якщо при підстановці x в усі дискримінантні функції найбільше числове значення дає функція $d_i(x)$. В разі неоднозначності рішення приймається довільний образ.

Розділяюча поверхня між класами w_i і w_j визначається множинством значень x , для яких $d_i(x) = d_j(x)$, або, такими x , для яких

$$d_i(x) - d_j(x) = 0. \quad (3)$$

Общепринятая практика состоит в том, чтобы описать разделяющую поверхность между двумя классами единой функцией $d_{ij} = d_i(x) - d_j(x) = 0$. Тогда $d_{ij}(x) > 0$ для образов из класса w_i , и $d_{ij}(x) < 0$ для образов из класса w_j . Решение задачи распознавания символов можно рассматривать как поиск различных подходов, используемых для нахождения дискриминантных функций, которые удовлетворяют неравенство (3).

Рассмотрим одну из наиболее часто используемых математических моделей алгоритма распознавания теории решений. Символ представляется в виде вектора, состоящего из характеристик графического образа этого символа.

$X = (x_1, x_2, \dots, x_n)$, здесь x_1 – первая характеристика вектора X , x_2 – вторая характеристика вектора X и т.д.

Рассмотрим задачу сравнения двух символов, представленных в форме векторов $X = (x_1, x_2, \dots, x_n)$ и $Y = (y_1, y_2, \dots, y_n)$. В качестве критерия близости используем евклидово расстояние.

$$D(X, Y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}. \quad (4)$$

С учетом того, что одни характеристики при сравнении имеют большее значение, чем другие, формула расстояния приобретает вид

$$D(X, Y) = \sqrt{\sum_{i=1}^n w_i (x_i - y_i)^2}, \quad (5)$$

где w_i – весовой коэффициент важности i -ой характеристики. Поскольку при распознавании нас интересует вычисление минимального расстояния между векторами, большее значение весового коэффициента w_i соответствует более важному признаку и наоборот.

Вышерассмотренные формулы реализуют классификатор по минимуму расстояния, один из наиболее логичных и интуитивно понятных классификаторов теории решений. Среди других, часто используемых классификаторов, стоит упомянуть байесовский классификатор [3, 4], работа которого основана на вероятностной оцен-

ке принадлежности образа к тому или иному классу.

Далее рассмотрены некоторые характеристики, используемые для распознавания символов.

1.2. Инвариантные моменты распределения точек. Моменты различных порядков используются в таких задачах обработки изображений как зрение роботов, обнаружение и распознавание летательных аппаратов и судов по снимкам, анализ сцен и распознавание символов. В последнем случае в качестве признаков используют значения статистических моментов совокупности "черных" точек относительно некоторого выбранного центра.

Наиболее общеупотребительными в приложениях такого рода являются построчные, центральные и нормированные моменты.

Для цифрового изображения, хранящегося в двумерном массиве, построчные моменты являются функциями координат каждой точки изображения следующего вида:

$$m_{pq} = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} x^p y^q f(x, y), \quad (6)$$

где $p, q = 0, 1, \dots, \infty$; M и N – размеры изображения по горизонтали и вертикали, $f(x, y)$ – яркость пикселя в точке (x, y) на изображении.

Центральные моменты – функция расстояния точки от центра тяжести символа:

$$\mu = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} (x - \bar{x})^p (y - \bar{y})^q f(x, y), \quad (7)$$

где (\bar{x}, \bar{y}) – координаты центра тяжести.

Вследствии, нормированные центральные моменты получаются в результате деления центральных моментов на моменты нулевого порядка:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^\gamma}, \quad \text{где } \gamma = \frac{p+q}{2}. \quad (8)$$

Отдельный вид моментов – инвариантный по отношению к операциям масштабирования, параллельного переноса, поворота. Такие моменты называют инвариантными. Они вычисляются через нормированные центральные моменты второго и третьего порядков. Использование инвариантных моментов позволяет распознавать слегка повернутые символы, а также символы разных размеров [5, 6]. Инвариантные моменты вычисляются по формулам

$$\phi_1 = \eta_{20} + \eta_{02}, \quad (9)$$

$$\phi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2, \quad (10)$$

$$\phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2, \quad (11)$$

$$\phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2. \quad (12)$$

Всего в расчетах используется семь инвариантных моментов, значения которых логарифмируются.

На практике применение инвариантных моментов сопряжено с небольшими различиями значений (которых по идее не должно быть) при небольших углах поворота. Данный эффект связан с дискретностью рассматриваемого изображения [7, 8].

1.3. Центр тяжести рисунка. Вычисления центра тяжести рисунка одна из базовых операций применяемых при анализе статистического распределения точек рисунка.

Координаты (\bar{x}, \bar{y}) центра тяжести рисунка определяются по формуле:

$$\bar{x} = \frac{\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} xf(x, y)}{\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y)}, \quad (13)$$

$$\bar{y} = \frac{\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} yf(x, y)}{\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y)},$$

где M и N — размеры изображения по горизонтали и вертикали, $f(x, y)$ — яркость пикселя в точке (x, y) на изображении.

С практической точки зрения, вычисление центра тяжести – «улучшенный вариант» вычисления средней яркости изображения, которая подробно описана в [7]. Центр тяжести характеризует распределение черных точек по всему рисунку.

1.4. Структурные характеристики символа. Структурные признаки обычно используются для выделения общей структуры образа. Они описывают геометрические и топологические свойства символа.

Одними из наиболее используемых признаков являются штрихи и пробелы, применяемые для определения таких характерных особенностей изображения: концевых точек, пересечения отрезков, замкнутых циклов, а также их положения относительно рамки, объемлющей символ.

Пусть матрица, содержащая скелетизированный символ, разделена на ряд областей, каждой из которых присвоено буквы А, В, С и т.д. Символ рассматривается как набор штрихов. При этом штрих является прямой линией (l) или кривой (c), соединяет некоторые две точки в начертании символа. Штрих является кривой, если его точки удовлетворяют следующему выражению:

$$ABS \left| \frac{\sum_{i=1}^n ax_i + by_i + c / \sqrt{a^2 + b^2}}{n} \right| > 0.69, \quad (14)$$

в противном случае – это прямая. В данной формуле (x_i, y_i) – точка, принадлежащей штриху; $ax + by + c = 0$ – уравнение прямой, проходящей через концы штриха, коэффициент 0.69 получен опытным путем.

При введенных обозначениях признаки символа могут быть записаны, например, в виде 'A l C' и 'AcD', что означает наличие прямой, проходящей из области 'A' в область 'C', и кривой, проходящей из области 'A' в область 'D' соответственно. Достоинство структурных признаков по сравнению с другими методами определяется устойчивостью к сдвигу, масштабированию и повороту символа на небольшой угол, а также к возможным дисторсиям и различным стиливым вариациям шрифтов.

Недостатком структурних признаков является повышенное требование к алгоритмам скелетизации, а также ресурсоемкость. Поскольку данные характеристики невозможно записать в строгой аналитической форме, применение таких признаков при распознавании требует использования дополнительных алгоритмов, которые находят соответствие между эталонами символов и распознаваемыми символами. Один из таких алгоритмов – метод маски.

1.5. Метод маски (пересечений). На изображение накладывается маска, которая состоит из прямых, которые выходят из начала координат (верхний левый угол рисунка) под углами 0° , 45° , 90° , 135° . Затем подсчитывается количество пересечений скелета символа с прямыми маски. Этот метод инвариантен к дисторсии и небольшим стилистическим вариациям символов, а также обладает достаточно высокой скоростью и не требует значительных вычислительных затрат.

В данной работе, в качестве начала координат используются все углы изображения с целью избегания возможных ошибок вызванных, например, скелетизацией или нестандартным написанием символов.

2. Распознавание символов с помощью нейронных сетей. Геометрическая интерпретация обучения и распознавания

Распознавание символов – один из классических примеров использования нейронных сетей, нейросетевым алгоритмам распознавания посвящено значительное число статей и публикаций. Поэтому, вместо объемного описания стандартного алгоритма обратного распространения ошибки, автор считает уместным обратить внимание на некоторые аспекты использования нейронных сетей.

Независимо от того, какой классификатор используется для распознавания, геометрическое представление символа одинаково. И в случае использования нейронных сетей или использования классических методов теории решений, символ представляется в виде точки n -мерного

пространства, где n – количество признаков (характеристик), используемых для описания. В классическом варианте использования нейронных сетей для распознавания символов, признаками являются значения всех пикселей графического образа символа (предполагается, что графический образ распознаваемого символа – изображение в оттенках серого, а значения пикселей лежат в диапазоне $[0;1]$). Данный метод описания имеет два существенных преимущества для дальнейшего обучения и распознавания:

- полнота информации о распознаваемом образе;
- однородность информации.

Рассмотрим суть вышеупомянутых преимуществ. Полнота информации гарантирует, что на вход нейронной сети поступает все изображение и таким образом нейронная сеть будет располагать всей доступной информацией о символе для обучения или распознавания. Однородность информации гарантирует то, что во всех измерениях n -мерного пространства единица измерения одна и та же, а именно – пиксел. За счет этого, на каждый из входов нейросети будут поступать числа, которые лежат в одном диапазоне. Это существенно упрощает процесс обучения нейронной сети.

Данный способ имеет существенный недостаток, а именно – слишком большой объем хранимой информации об образе. Так, при обучении образа размером $N \times N$ пикселей (как правило, при распознавании или обучении графические образы символов имеют квадратную форму), данный образ хранится в виде N^2 характеристик. Нетрудно посчитать, что хранение в памяти компьютера базы символов, состоящей из нескольких десятков тысяч образов символов, требует несколько сот мегабайт оперативной памяти.

Но наибольший недостаток такого описания состоит в очень большом количестве вычислений, необходимых для обучения сети. Причины такой ресурсоемкости с вычислительной точки зрения: чем больше число связей, тем больше эпох необходимо пройти для их настройки [9]. Для полного понимания проблемы следует

рассмотреть геометрическую трактовку процесса распознавания. Заметим, что в контексте данной задачи, под нейросетью понимаем многослойный персиптрон без обратной связи.

При этом, входом каждого узла нейросети (кроме узлов входного слоя) является взвешенная сумма всех узлов предшествующего слоя. Пусть слой K предшествующий слою J сети (заметим, что на рис. 1 не предполагается алфавитной упорядоченности), создает на входе активирующего элемента каждого узла слоя J сигнал, обозначаемый I_j :

$$I_j = \sum_{k=1}^{N_k} w_{jk} O_k . \quad (15)$$

Для $j = 1, 2, \dots, N_j$, где N_j – число узлов в слое J , N_k – число узлов в слое K , w_{jk} – веса, модифицирующие выходные сигналы O_k узлов слоя K на входе узлов слоя J . Эти выходные сигналы слоя K имеют значения

$$O_k = h_k(I_k), \quad k = 1, 2, \dots, N, \quad (16)$$

где $h(I)$ – активирующая функция нейросети, тип которой не играет роли в дальнейшем.

Вышеупомянутый классификатор по минимуму расстояния ищет минимальное расстояние между поступившим образом и эталонными образами. Нейросеть, с другой стороны, занимается построением разделяющих поверхностей. Структура нейросети определяет характер разделяющих поверхностей. Однослойная сеть реализует разделяющую поверхность в форме гиперплоскости, которая делит на две части все n -мерное пространство признаков объектов. Двухслойная нейросеть строит произвольные выпуклые поверхности, образованные перескающими гиперплоскостями. Трехслойная сеть реализует разделяющие поверхности произвольной сложности. В силу вышеописанного факта, именно трехслойные нейросети чаще всего применяются для распознавания символов при реализации классических алгоритмов распознавания. (Нейронные сети с четырьмя и больше слоями применяются в отдельных случаях, обычно компромисс

между количеством слоев и количеством узлов в слое достигается методом проб и ошибок). Сопоставив рис. 1 и формулу (15) видно, что разделяющая классы поверхность формируется множеством гиперплоскостей, так как веса в (15) – коэффициенты гиперплоскости порядка N_k . Сложность такой поверхности определяется числом узлов сети: чем больше узлов, тем большее количество гиперплоскостей формируют данную поверхность (разделяющую поверхность можно рассматривать как некую суперпозицию гиперплоскостей).

Порядок гиперплоскости определяется количеством характеристик и равен N^2 . Учитывая то, что число узлов скрытого слоя задает сложность поверхности, можно сделать вывод, что это число в основном связано с размером обучающей выборки, а общее число вычислений, необходимых для работы нейросети зависит от числа узлов во всех слоях. При этом, число узлов выходного слоя нельзя уменьшить в принципе, поскольку это противоречит самому смыслу задачи распознавания (фактически, в этом случае будет уменьшено число классов). Число нейронов скрытого слоя (или нескольких скрытых слоев) можно уменьшить за счет уменьшения размера обучающей выборки, что не всегда приемлемо с точки зрения ожидаемой эффективности распознавания. Число входных узлов можно уменьшить только одним способом: уменьшением числа признаков данного символа. При этом уменьшение разрешения графического образа не является приемлемым путем, так как при уменьшении числа точек происходит потеря информации об образе [10, 11]. В качестве решения данной проблемы, можно использовать представление символа в виде вектора признаков теории решений.

2.1. Особенности векторного представления символов. Геометрическая интерпретация обучения и распознавания. Наибольшим достоинством представления символов в виде вектора признаков является компактность записи и небольшой объем информации, необходимой для

описания отдельного символа. Это позволяет уменьшить число вычислений при обучении и распознавании. Как и при растровом подходе, символ представляется в виде точки n -мерного пространства, только в этом случае размерность пространства возможно уменьшить в зависимости от количества выбранных характеристик. При правильно выбранной системе характеристик, можно получить важную смысловую информацию о символе, недоступную при растровом подходе, например, количество линий, формирующих остов символа, конечные точки и точки перегиба и прочие. Такая информация не теряет своей актуальности при изменении масштаба, незначительном повороте и прочих трансформациях.

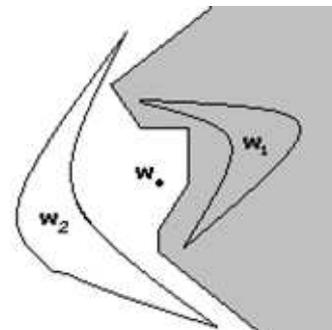
К недостаткам данного описания можно отнести две особенности:

- возможная потеря полноты информации при формировании характеристик;
- неоднородность информации.

В отличие от нейросетевого подхода, который использует растровое описание, эффективность распознавания символов при векторном описании напрямую зависит от выбранной системы признаков, тем самым предъявляя повышенные требования к алгоритмам, а не вычислительным мощностям системы, на которой будет происходить обучение. Важно понимать, что если признаки выбраны неудачно, распознавание будет неэффективным. Очевидно, что выбранные для описания характеристики должны пытаться однозначно описать разные классы символов.

Значительной проблемой может стать неоднородность информации, описывающей образ символа. В отличие от растрового представления, информация хранящаяся в векторе признаков является однородной только при использовании одинаковых по своей сути характеристик. Частично, данная проблема решается за счет нормирования значений характеристик, однако на практике это зачастую приводит к «выпаданию» определенных символов за границы диапазона значений.

Обучение и распознавание символов при использовании теории решений реализуется следующим образом. В процессе обучения вычисляются характеристики символов, принадлежащих обучающей выборке и формируются векторы с характеристиками, которые часто называют эталонами. Во время распознавания среди эталонов ищется тот, характеристики которого минимально отдалены от характеристик неизвестного символа с точки зрения Евклидовой метрики. Заметим, что если, по сравнению с нейросетями, процесс обучения в теории решений является простым с вычислительной точки зрения, то процесс распознавания требует больших вычислительных ресурсов, так как требуется провести сравнение неизвестного символа со всеми эталонами, следовательно при равном объеме обучающей выборки нейросеть всегда будет распознавать быстрее. Преимущество нейросетей в скорости показано на рисунке.



Рисунок

Здесь рассмотрено распознавание неизвестного символа, представленного на рисунке точкой w . В случае использования нейросети, распознавание происходит за счет подстановки координат точки w в уравнение разделяющей поверхности. При использовании методов теории решений необходимо вычислить расстояние от точки w до всех точек областей w_1 и w_2 , а затем отнести w к той области, к которой принадлежит ближайший к ней эталон. Очевидно, что количество вычислений во втором случае значительно выше.

3. Побудова модифікованого алгоритма. Порівняння з алгоритмами, заснованими на теорії рішень і нейронними алгоритмами

Для перевірки ефективності алгоритма розроблено три основні програми розпізнавання символів: **Cunning Eye Solution** (заснована на класифікаторі по мінімуму відстані), **Cunning Eye Neuro** (заснована на класическому нейронному підході), **Cunning Eye Neuro Plus** (заснована на синтезі вищеописаних алгоритмів розпізнавання). Для аналізу ефективності роботи алгоритмів програми **Cunning Eye Neuro** і **Cunning Eye Neuro Plus** працювали з різним кількістю входних даних.

Вищеописані системи навчалися на одній і тій же вибірці, потім тестувалися на єдиній для всіх перевіркової вибірці. Як такі вибірки взяті бази рукописних цифр MNIST. Ця база є загальноприйнятою базою серед розробників OCR систем. Розмір навчальної вибірки – 6000 символів, розмір тестової – 1000 символів.

Як основні параметри оцінки роботи системи оцінювалися такі характеристики:

- швидкість навчання;
- точність розпізнавання символів з навчальної вибірки;
- точність розпізнавання символів з тестової вибірки.

Коротко розглянемо використовувані в експериментах системи розпізнавання символів.

Cunning Eye Solution. Ця система реалізує розпізнавання символів з використанням класифікатора по мінімуму відстані. Як ознаки використовуються характеристики, описані в розділах 1.2 – 1.5. Для оцінки інваріантності використовуваних характеристик при розпізнаванні використовувалися їх різні комбінації.

Cunning Eye Neuro. Ця система проводить розпізнавання символів з допомогою класических трішарових нейронних мереж з архітектурами 784-50-10, 196-50-10, 361-50-10. В першому випадку, цифри

з дозволом 28x28 пікселів без змін входять в нейронну мережу. В інших випадках цифри масштабувалися і зводилися до розмірів 14x14 і 19x19 пікселів. Кількість нейронів прихованого шару було визначено методом проб і помилок.

Cunning Eye Neuro Plus. Ця система проводить розпізнавання символів з допомогою класическої трішарової нейронної мережі з архітектурами 13-15-10, 9-15-10, 7-15-10. В першому випадку, на вход подаються (в нормованому для однорідності вигляді) 7 інваріантних моментів, координати центру мас символу, кількість перетинів з маскою, всього 13 характеристик. В другому випадку на вход нейронної мережі подаються 7 інваріантних моментів, координати центру мас символу, т.е. дані про перетини з маскою не використовуються. В третьому випадку на вход нейронної мережі подаються тільки 7 інваріантних моментів, інші характеристики не використовуються. Кількість нейронів прихованого шару визначено методом проб і помилок.

Проаналізуємо результати експериментів, записані в таблиці. Необхідно відзначити, що система **Cunning Eye Neuro** не змогла навчитися розпізнаванню 63 символів, а **Cunning Eye Neuro Plus** – 123. В цих випадках як час навчання бралося час, за який система навчалася розпізнаванню максимальної кількості символів.

Що стосується безпосередніх результатів експерименту, то варто відзначити неефективність системи, побудованої за методами теорії рішень. В даному випадку так важливою властивістю як швидке навчання, нівелюється недостатньою точністю розпізнавання.

Класический алгоритм розпізнавання на основі нейронних мереж продемонстрував високу точність розпізнавання, хорошу здатність до узагальнення і низьку швидкість навчання.

Алгоритм, що синтезує різні підходи до розпізнавання продемонстрував дуже високу, порівняно з класическим алгоритмом, швидкість обу-

Таблиця. Результати експериментів

Назва системи	Точність розпізнавання обучаючої виборки, %	Точність розпізнавання перевірочної виборки, %	Час навчання, год; хв; сек
Cunning Eye Solution (всі характеристики)	83.32	61.45	0:0:26
Cunning Eye Solution (моменти+центр мас)	71.88	55.67	0:0:24
Cunning Eye Solution (моменти)	66.7	50.93	0:0:24
Cunning Eye Neuro (784-50-10)	98.57	86.48	5:14:35
Cunning Eye Neuro (361-50-10)	96.78	79.13	4:06:39
Cunning Eye Neuro (196-50-10)	95.56	75.19	3:29:06
Cunning Eye Neuro Plus (13-15-10)	94.84	69.63	2:04:01
Cunning Eye Neuro Plus (9-15-10)	83.12	61.84	1:41:19
Cunning Eye Neuro Plus (7-15-10)	75.17	56.18	1:49:59

чення. Невисока точність розпізнавання перевіряючої вибірки свідчить про поганий здатність до узагальнення, однак точність розпізнавання навчаючої вибірки досить висока, тому при належному навчанні така система здатна ефективно вирішувати задачу розпізнавання. Необхідно відзначити, що точність розпізнавання такої системи залежить від вибраних характеристик, тому додавання нових інваріантних ознак також підвищить точність розпізнавання.

Висновки

Розглянутий алгоритм розпізнавання символів дозволяє усунювати окремі недоліки алгоритмів теорії рішень і класических нейронних алгоритмів. При побудові алгоритму першочерговою задачею було суттєве

збільшення швидкості навчання, яке досягнуто за рахунок невеликого зменшення точності розпізнавання.

Як було згадано раніше, основним фактором, що впливає на точність розпізнавання символів в запропонованому алгоритмі, є система вибраних характеристик. При використанні запропонованого набору характеристик була отримана дуже висока швидкість навчання нейронної мережі, а також достатня точність розпізнавання. При цьому, за рахунок введення більшої кількості інваріантних характеристик можливо збільшення як точності розпізнавання, так і швидкості навчання. В цьому випадку, запропонований алгоритм можна покращити, додавши як характеристики ознаки, обчислювані методом поліноміальної регресії, запропоновані в [11]. Відзначимо, що в оптимально вибраній системі характеристик, підвищення швидкості навчання можливо без суттєвих втрат у точності розпізнавання.

1. *Ватолин Д.С.* Алгоритмы сжатия изображений.– МГУ, 2002. – С. 30 – 32.
2. *Гонсалес Р., Вудс Р.* Цифровая обработка изображений. – М.: Техносфера, 2005. – С. 1110 – 1148.
3. *Форсайт Д., Понс Д.* Компьютерное зрение. Современный подход. – М.: Вильямс, 2004. – С. 603 – 610.
4. *Dr. Maria Isabel Ribeiro.* “Gaussian Probability Density Functions: Properties and Error Characterization”, www.imageprocessingplace.com
5. *Wood J.* “Invariant pattern recognition: A review“. Pattern Recognition. – 1996 – Vol. 29(1). – P. 1 – 17.
6. *Jamie Shutler.* “Statistical moments”, 2002. http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/SHUTLER3/node5.html. – P. 1
7. *Бондаренко А.В., Галактионов В.А., Желтов С.Ю.* Исследование подходов к построению систем автоматического считывания символьной информации. – М.: Изд-е ИПМ им. М.В. Келдыша РАН, 2003. – С. 5 – 10.
8. *Alexander G. Mamistvalov.* “N-Dimensional Moment Invariants and Conceptual Mathematical Theory of Recognition n-Dimensional Solids”. IEEE Press, – 1998. – Vol. 20. – P. 1 – 9.
9. *Саймон Хайкин.* «Нейронные сети. Полный курс». Издание второе (исправленное). Прэнтис Холл, 2006. – С 239 – 298.
10. *George N. Sazaklis* “Geometric methods for optical character recognition”, New York State University. – 1997. – P. 1 – 19.
11. *Sameer Singh.* “Shape Detection Using Gradient Features for Handwritten Character Recognition” School of Computing, University of Plymouth, 2001. – P. 1 – 13.

Получено 03.05.2007

Об авторе:

Прохоров Валерий Георгиевич,
аспирант Института программных
систем НАН Украины,
e-mail: makumazan84@yahoo.com