УДК 004.62

НЕЙРОМЕРЕЖЕВІ МЕТОДИ РОЗПІЗНАВАННЯ ЗОБРАЖЕНЬ ТЕКСТІВ

О. В. Тимченко¹, Б. М. Гавриш², Б. В. Дурняк¹

¹Українська академія друкарства,
вул. Під Голоском, 19, Львів, 79020, Україна

²Національний університет «Львівська політехніка»,
вул. С. Бандери, 12, Львів, 79000, Україна

Проведено системний аналіз та порівняння методів розпізнавання текстових зображень. Показано можливість застосування різних методів для реалізації завдання розпізнавання текстів. Розглянуто методологію розпізнавання текстових зображень на основі нейронних мереж, методи та алгоритми побудови нейро-нечіткої системи розпізнавання текстових зображень, зокрема методи покращення якості текстових зображень та зменшення шуму за допомогою лінійної та нелінійної фільтрації. Показано особливості бінаризації таких зображень, нечіткої обробки зображень для виділення меж та сегментації символів, а також можливість реалізації граматики для структурного розпізнавання текстових зображень. Розглянуто методи навчання нейронної мережі та проведено її тестування.

Ключові слова: розпізнавання зображень текстів, система розпізнавання, обробка нечітких і зашумлених зображень, нейронні мережі.

Постановка проблеми. З розвитком інформаційних технологій стало можливим полегшити, прискорити та підвищити якість розпізнавання друкованого або рукописного текстів. Першим елементом системи розпізнавання літер є сканери або цифровий фотоапарат, що вводять в комп'ютер зображення текстів. Для створення текстового документа, необхідно розпізнати на цьому зображенні окремі символи. Існує ціла низка програмних продуктів (англ. optical character recognition, OCR), які практично автоматизували процеси розпізнавання текстів. Проте забезпечити задовільний результат у випадку спотворень зображень текстів різного типу (геометричних, шумів та подібних) вдається не завжди [1].

Аналіз останніх досліджень та публікацій. Всі методи розпізнавання об'єктів діляться на два види: методи, які засновані на теорії рішень, і структурні методи. Перші засновані на обчисленні за допомогою кількісних величин, таких як довжина, текстура тощо. Другі орієнтовані на образи, для опису яких більше підходять якісні величини, наприклад реляційні. Також в розпізнаванні об'єктів важливу роль відіграє навчання на основі відомої вибірки [1, 2].

Образ — деяка упорядкована сукупність ознак. Класом образів є об'єкти з однаковими властивостями. Класифікатором або вирішальним правилом називається

правило зарахування образу до одного з класів на підставі його вектора ознак. На практиці широке застосування мають три форми подання ознак: вектор ознак (для кількісних величин), символічний рядок і дерева ознак (для структурних величин) [3].

Методи, засновані на зіставленні, є наборами векторів ознак кожного класу об'єктів. Новий образ буде зарахований до того класу, який виявиться найбільш близьким у межах заздалегідь заданій метрики. Очевидно, що найпростіший підхід полягає у пошуку мінімальної відстані, що обчислюється за допомогою евклідових норм між векторами ознак невідомого об'єкта і векторами прототипу. Висновок про належність об'єкта до певного класу відбувається за найменшим з цих відстаней.

Мета статті. Основною метою дослідження є розроблення системи нейронно-чіткого розпізнавання текстів на основі структурних методів.

Виклад основного матеріалу дослідження.

Основні методи розпізнавання. Проблема розпізнавання — знаходження таких ознак, які чітко розділяють простір на класи та виключення ознак, які не містять ніякої корисної інформації для класифікації, а тільки створюють надмірне навантаження на обчислювальні ресурси.

Далі необхідно виявити оптимальну вирішальну процедуру. Оскільки на цьому етапі ми маємо n класів і вимірювальні вектори для кожного з них, то задачу можна звести до побудови границь, що розділяють класи.

Рис. 1. Границі, що розділяють класи в евклідовому просторі

Порівняння об'єктів можна зробити на основі їх подання у вигляді векторів вимірювань. Схожість векторів ознак двох об'єктів може бути описано за допомогою евклідової відстані:
$$\|x_1 - x_2\| = \sqrt{\sum_{j=1,d} (x_{1,j} - x_{2,j})^2}.$$

Виділяють 3 групи методів розпізнавання образів:

- Порівняння із зразком. До цієї групи належать структурні методи і методи, які використовують відстань (класифікації по найближчому середньому і по відстані до найближчого сусіда).
- Статистичні методи. Прикладом цієї групи слугує байєсівський метод прийняття рішення. Статистичні методи засновані на обчисленні ймовірності.
- Нейронні мережі. Окремий клас методів розпізнавання. На відміну від інших методів, нейронні мережі здатні навчатися вже у процесі розпізнавання.

Структурний метод застосовується для об'єктів, які можна структурно розділити на складові. Так само важливо, щоб при розпізнаванні система знайшла такі ознаки, які точно дадуть змогу сказати, що об'єкт належить до цього класу і до ніякого іншого (рис. 2). Друга назва цього методу — синтаксичний, так як він має на увазі використання мови опису образів, яка структурно описує кожен елемент і піделементи, структурно розділяючи образ на підобрази. Метод буде корисний для розпізнавання складних образів, що складаються з багатьох образів нижчого, простого рівня.

Рис. 2. Ілюстрація структурного методу розпізнавання

В цьому випадку коло складається з двох дуг, а прямокутник з чотирьох відрізком прямих. Ця класифікація дає змогу однозначно визначити належність об'єкта до одного з класів.

Байєсівський підхід до прийняття рішення. Цей метод заснований на теоремі Байєса і визначенні апіорних ймовірностей, тобто ймовірність результатів або належність об'єкта певного класу змінюється після отримання нових експертних оцінок (підтвердження наявності нових ознак).

Поява того чи іншого способу є випадковою подією і ймовірність цієї події можна описати за допомогою закону розподілу ймовірностей багатовимірної випадкової величини ξ в тій чи іншій формі. Знаючи елементи навчальної вибірки, можна відновити імовірнісні характеристики цього середовища.

Байєсівський класифікатор на основі спостережуваних ознак зараховує об'єкт до класу, до якого цей об'єкт належить з найбільшою ймовірністю. Припустимо, що ми розглядаємо деяку випадкову величину X , яка має щільність ймовірності $p(x, w)$ з параметром w . Нам потрібно отримати дані про інший випадкової величини w , що має деякий розподіл ймовірності $\tau(w)$. Нехай в результаті спостережень отримані статистичні дані x . З визначення умовної ймовірності випливає:

$$\Pr(w|x) = \frac{\Pr(w, x)}{\Pr(x)} \quad \text{і} \quad \Pr(x, w) = \Pr(x|w) \Pr(w).$$

Підставивши другий вираз в перший, отримаємо формулу Байєса:

$$\Pr(w|x) = \frac{\Pr(x|w) \Pr(w)}{\Pr(x)}.$$

Для завдання розпізнавання є m класів, тобто m можливих змінних $(w_1, \dots, w_m)$, формула Байєса набуде вигляду:

$$\Pr(w_i|x) = \frac{\Pr(x|w_i) \Pr(w_i)}{\Pr(x)} = \frac{\Pr(x|w_i) \Pr(w_i)}{\sum_{k=1,m} \Pr(w_k) \Pr(x|w_k)}.$$

Розподіл $\Pr(w_i)$ — апостеріорний розподіл значень w_i , за умови отримання статистичних даних. Тут w_i — гіпотеза, а x — свідчення, що підтримує гіпотезу.

На практиці необхідно забезпечити обчислення $\Pr(x|w_i)$. Емпіричний підхід передбачає розбиття діапазону x на інтервали, потім підрахунок частот появи x серед значень кожного інтервалу і побудова гістограми на основі цих вимірів.

Наведемо приклад використання теореми Байєса для розпізнавання образів.

Припустимо, що є вибірка елементів. Яка ймовірність того, що m -й елемент, який не належить нашій вибірці, буде відповідати зразку? У теорії ймовірностей цю передумову можна записати, використовуючи теорему Байєса:

$$p(Y=1|X=x_m) = \frac{p(X=x_m|Y=1)p(Y=1)}{p(X=x_m)}.$$

У цій формулі X — передумова, а Y — наслідок. Цей вираз є функцією правдоподібності.

Рішення класифікації залежить від тих параметрів, які ми знаємо. Наприклад, ми можемо знати тільки значення апіорної ймовірності, а інші значення оцінити неможливо. Якщо ми зможемо підрахувати значення функції правдоподібності, то й зможемо знайти значення шуканої ймовірності за формулою Байєса. Рішення тут буде таким — поставити мітку того класу, для якого ймовірність — максимальна.

Недолік цього методу полягає в тому, що в складних завданнях розмірність ознак може бути дуже високою і для оцінки ймовірності багатовимірної випадкової величини може не вистачити числа спостережень зі списку з історичними даними.

Нейронні мережі дають змогу вирішувати найбільш широке коло завдань розпізнавання (наприклад, зашумлених або рукописних текстів) і представляють структуру з декількох шарів — штучних нейронів (обчислювальних елементів) і зв'язків між ними. Структура імітує структуру і властивості організації нервової системи живих організмів (рис. 3). Мережа отримує на вхід набір сигналів і на виході видає відповідну відповідь (вихідні сигнали), які описують рішення деякої задачі. Нижче наведено схему нейромережі.

Рис. 3. Найпростіша структура нейронної мережі

На кожен нейрон вхідного шару поступає сигнал, який множиться на відповідний йому. Функція активації є або пороговим значенням, що передає на вихід одиничний сигнал, або сигмоїдальною функцією, яка перетворює суму всіх сигналів в число, що є в діапазоні від 0 до 1.

Отже, на виході мережі отримуємо розподіл усіх значень, що подібне на результат методу Байєса.

Порівняно з лінійними методами, нейромережі дають змогу ефективно будувати нелінійні залежності, точніше описують набори даних. Що ж стосується байсівського класифікатора, який будує квадратичну розділяючу поверхню, нейронна мережа може побудувати поверхню більш високого порядку.

Загальна характеристика систем розпізнавання текстів.

Система OCR (наприклад, FineReader, OmniPage, ReadIris чи інші) отримує цифрове зображення сканованого або сфотографованого документа і формує текст, який містить це зображення, в одному з форматів електронних текстових документів (рис. 1 [2]).

Рис. 4. Узагальнена структурна схема OCR-системи

Блок зчитування вхідних даних (сканування, фотографування) проводить перетворення світлового сигналу в електричний. Блок попередньої обробки переводить зображення у чорно-біле з метою значно зменшити розміри матриці ознак і зменшення витрат на обробку інформації. Також відбувається зменшення рівня шумів, ліквідації бахроми, дрібних прогалин (порожнеч) всередині ліній і, якщо буде потреба, операції потоншення ліній. Блок виділення ознак може мати різну складність залежно від використаних методів і природи зображення, яке розпізнається. Кожен, що піддається попередній обробці об'єкт має бути представлений у вигляді структури мовного типу, наприклад ланцюжка. Цей процес складається з двох підпроцесів: із сегментації і з виділення непохідних елементів. Тому при розпізнаванні друкованого тексту, в якому символи відокремлені один від іншого, сегментація дає результати, які дають змогу в результаті розпізнавання отримати текст, практично ідентичний оригінальному. Блок класифікації формує ознаки елементів зображення еталонним зразкам. Оцінка ефективності класифікації залежить від якості виконання попередніх етапів і оцінюється кількістю ознак, за якими буде встановлена ця відповідність. Блок навчання модифікує представлення елементів у такий спосіб, щоб якнайшвидше класифікувати елементи зображення на наступ-

них етапах. Він надає можливість збереження системою методу, стилю та інших результатів класифікації, які після успішного завершення процесу виділення ознак можуть бути використані для більш точного і швидкого розпізнавання зображень текстів [3, 4].

Більшість таких операцій систем розпізнавання символів через нечіткість та розмитість параметрів доцільно виконувати методами нечіткої логіки.

Отже, під час розробки методів та алгоритмів побудови нейронечіткої системи розпізнавання текстових зображень необхідно:

- зменшити шум текстових зображень;
- здійснити бінаризацію зображень;
- виділити межі та провести сегментацію образів;
- реалізувати граматики для структурного розпізнавання текстових зображень;
- провести перевірку запропонованих методів.

Зменшення шумів цифрового зображення. Для зменшення шуму цифрового зображення необхідне використання лінійних та нелінійних методів фільтрації, які зберігають різкі зміни та краї об'єктів за одночасного усунення шуму. Тому найчастіше за шум на зображенні вважають імпульсні шуми, гаусівський та змішаний імпульсний та гаусівський шуми [5].

Фільтр Гауса (лінійний) усереднює по закону Гауса пікселі в околі точки зображення $(x, y) \in Z^2$: $G(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right)$. Цей фільтр є сепарабельним,

фільтрація може виконуватись по рядках і стовпцях матриці зображення, оскільки розкладається на два незалежних фільтри по різних координатах.

Суть медіанного фільтра (нелінійного) полягає у виборі медіани із набору пікселів околу: $Im_{i,j} = \text{med}[Im_{i+s,j+t}, (s,t) \in W]; (i,j) \in Z^2$. Спочатку усі значення пікселів околу (s,t) відносно пікселя зображення $(i,j) \in Z^2$ сортуються у певному порядку (зростання) та вибирається медіанне значення, яким замінюється центральний піксель околу. Медіанний фільтр найкраще справляється із імпульсним шумом.

Нечітке опрацювання зображень — це множина різних нечітких підходів, якими є розуміння, представлення, опрацювання зображень, їх сегментація і класифікація. У процесі розпізнавання величезне значення має процес попереднього нечіткого опрацювання зображень, оскільки саме від нього залежить якість даних, які надходять на входи нейронної мережі. Алгоритм попереднього нечіткого опрацювання можна представити у вигляді такої послідовності кроків (рис. 5): отримання зображення; перетворення отриманого кольорового зображення в зображення в градаціях сірого; нечітка обробка зображень (фазифікація, система нечіткого виводу, дефазифікація) [6].

Для перетворення зображення у відтінки сірого для кожної точки зображення виділяються інтенсивності червоної R , зеленої G і синьої B складової кольору, а потім здійснюється перетворення кольору за формулою:

$$Y = 0.255R + 0.255G + 0.255B, \quad (1)$$

де Y — нове значення кольору. Отримуємо відтінки сірого між 0 і 1.

Рис. 5. Алгоритм попереднього нечіткого опрацювання зображень для їх сегментації

Бінаризація зображення. При бінаризації зображення значення пікселя умовно вважають рівним нулю або одиниці, водночас пікселі з нульовим значенням (білого кольору) називають заднім планом, а пікселі зі значенням рівним одиниці (чорного кольору) називають переднім планом. Проте бінарне зображення, отримане в результаті такого перетворення, спотворюється відповідно до оригіналу, що характеризується появою розривів і розмитостей на об'єктах, виникненням зашумлення зображення в однорідних областях, а також до втрати цілісності структури об'єктів.

Існує велика кількість методів бінаризації. Суть такого перетворення растрових зображень полягає в порівняльному аналізі яскравості поточного пікселя $P(x, y)$ з деяким пороговим значенням $PT(x, y)$: якщо $P(x, y) > PT(x, y)$, то піксель на бінарному зображенні — білий, інакше — чорний.

Пороговою поверхнею PT є матриця, розмірність якої відповідає розмірності вихідного зображення [7].

У розробленому алгоритмі вихідне півтонове зображення, задане в цифровому вигляді, поділяється на квадратні матриці S розміром $h \times h$ з елементами $S(x, y) \in [0, 1]$. Кожна матриця S перетворюється в бінарну матрицю r того ж розміру. Перетворення $S \rightarrow r$ здійснюється за умови рівності яскравості матриці (1):

$$Ent \sum_{x,y} S(x, y) = \sum_{x,y} r(x, y) = b, \quad (2)$$

де яскравістю будемо називати суму елементів кожної матриці. Яскравість бінарної матриці дорівнює числу одиниць b . Алгоритм працює у такий спосіб. При кожному з $k = 1, b$ обходів елементів матриці S в ній знаходиться позиція елемента з максимальним значенням. На місці цієї позиції в S записується нуль, а в бінарній матриці r записується одиниця [7, 8].

Алгоритм нечіткого опрацювання для виділення меж і сегментації зображень. Нечітке опрацювання зображень (НОЗ) (рис. 3) складається з трьох основних етапів: фазифікації зображень F , системи нечіткого виведення M на значеннях приналежності та дефазифікація зображень D . Головне нечітке опрацювання зоб-

ражень знаходиться в середньому кроці (система нечіткого виведення M). Після передачі даних зображення з сірого рівня у фазифікації система нечіткого виведення визначається значенням приналежності. Фазифікація — кодування даних зображень, дефазифікація — декодування результатів, які дають змогу опрацювати зображення нечіткими методами.

Рис. 6. Алгоритм нечіткого опрацювання зображень для виділення меж

Зображення X — розміру $M \times N$ з L рівнями сірого $g=0,1,2,\dots,L-1$ може бути визначено як масив, який вказує значення приналежності кожного пікселя по заздалегідь вибраних властивостях зображення (наприклад, яскравість тощо).

$$X = \bigcup_{m=1}^M \bigcup_{n=1}^N \frac{\mu_{mn}}{g_{mn}}, \mu_{mn} \in [0.1], \quad (3)$$

де μ_{mn} і g_{mn} — призначення пікселя mn в нечіткій множині. Визначення значень приналежності залежить від конкретних вимог застосування і від відповідної бази знань.

Конкретний вибір методів дефазифікації здійснюють залежно від бажаної поведінки нечіткого виводу. Найчастіше використовується «центр ваги» функції приналежності нечіткої множини.

Вихід системи Y для вхідних системи X описується залежністю:

$$Y = D(M(F(X))). \quad (4)$$

Методика структурного розпізнавання текстових зображень і побудови відповідної граматики. У методі порівняння проводиться порівняння виділеного елемента з базою даних, де для кожного представлені об'єкти під різними кутами,

масштабами, зміщеннями, деформаціями тощо. Для букв у базі даних потрібно зазначити також шрифт, властивості шрифту тощо.

Другий підхід — здійснюється аналіз характеристик образу. У разі оптичного розпізнавання тексту це може бути визначення геометричних характеристик окремих символів.

Новітні методи розпізнавання використовують штучні нейронні мережі. Хоча методи вимагають великої кількості прикладів під час навчання, але мають більш високу ефективність і продуктивність. У цих методах зображення символу приводиться до деякого стандартного розміру, наприклад 16x16 пікселів. Значення яскравості у вузлах нормалізованого растра використовуються як вхідні параметри нейронної мережі. Число вихідних параметрів нейронної мережі дорівнює числу розпізнаваних символів. Результатом розпізнавання є символ, якому відповідає найбільше зі значень вихідного вектора нейронної мережі.

Структурний опис:

- зрозумілий для людини, що вирішує задачу розпізнавання об'єкта;
- прийнятний для комп'ютерної реалізації;
- вільний від трудомісткості обчислень і втрат інформації.

Структурні ознаки, які використовують в структурному описі, це непохідні (тобто елементарні унікальні) елементи (символи), примітиви зображення об'єкта розпізнавання. Структурні методи зберігають інформацію не про поточкове написання символу, а про його топологію. Тобто еталон містить інформацію про взаємне розміщення окремих складових частин символу [9, 10]. Зрозуміло, що при цьому стає неважливим розмір розпізнаваної букви і навіть шрифт, яким вона надрукована.

Отриманий образ піддається процедурі скелетизації (зменшення) [11]. Контур скелетного представлення описується у вигляді послідовного набору особливих точок і ланцюгового коду обходу контура за годинниковою стрілкою, що складається з точки прив'язки, числа кодів і масиву напрямків з чергової точки на наступну точку. Загалом для опису довільного об'єкта непохідні елементи послідовуються в ланцюжки (речення) за своїм, характерним тільки для цього об'єкта, набором правил.

У результаті зв'язків з непохідних елементів (структурних ознак) утвориться об'єкт, аналогічно до того, як речення мови будуються шляхом з'єднання слів, які своєю чергою складаються з букв. У цьому структурні методи мають аналогію із синтаксисом природної мови. Звідси структурні ознаки носять ще назву лінгвістичних чи синтаксичних [12, 13].

Структурні або лінгвістичні системи для розпізнавання використовують речення, кожне з яких описує структуру (будову) об'єкта з непохідних (елементарних) елементів. Ці речення становлять спеціальну мову. Класифікація об'єкта виконується шляхом порівняння речення невідомого об'єкта з еталонними реченнями класів.

Характеристики зображення текстів і створення відповідної граматики. Особливістю цих зображень є те [14], що вони складаються з великої кількості

взаємозалежних частин. Тому аналіз таких зображень доцільно проводити за допомогою методів структурного розпізнавання. Результатом такого розпізнавання є не зарахування зображення символу до прототипу, а перелік символів і відношень між ними.

Структурне розпізнавання символів в умовах випадкових шумів зводиться до пошуку оптимального зображення символів, які розпізнаються. Функція якості полягає у відшукуванні найімовірнішої сукупності прихованих параметрів зображення. Наприклад, під час розпізнавання рядка тексту — це вимога знайти найімовірнішу послідовність літер рівнозначна вимозі мінімізації кількості неправильно розпізнаних символів.

Зображення, які містять тексти, таблиці, рисунки, створюються і читаються згідно з певними правилами, набір яких можна формалізувати як певну граматику. Очевидно, що алгоритми розпізнавання таких зображень повинні базуватись на використанні правил цієї граматики.

Методи створення таких граматик:

- використання теорії графів. Зображення тексту представлено у вигляді розміченого графу. Задачі розпізнавання ставляться як задачі знаходження ізоморфізму еталонного та вхідного графів, або ізоморфізму їх підграфів;
- методи теорії формальних мов та граматик. Зображення розглядається як слово у деякій формальній мові, яка задається за допомогою конструкцій, що є узагальненнями граматик Хомського. Розпізнавання полягає у відшукуванні найкращого в певному значенні виводу зображення у заданій граматиці [15].

Для цих граматик характерним є розгляд зображення як об'єкта, що складається за певними правилами з великої кількості елементарних частин. Ці частини і правила можуть значно відрізнятись між собою (наприклад, розпізнавання нот [16] або розпізнавання математичних формул [17]).

Нехай T — певна прямокутна підмножина двовимірної цілочисельної решітки: $T = \{(i, j) \mid 0 \leq i < H, 0 \leq j < W\}$. Множину T називатимемо полем зору, числа H та W — його висотою та шириною відповідно, а елементи поля зору називатимемо пікселями. Кольори пікселів вибираються з певної скінченної множини Y . Відображення $x: T \rightarrow Y$ називатимемо зображенням, величина $x(t)$ визначає колір зображення x в пікселі $t \in T$.

Двовимірною контекстно-вільною граматикою буде $G = \langle E, V, P, \varepsilon \rangle$, де E — множина термінальних зображень, що містить однопіксельні зображення (зображення, визначені на полі зору розміром 1 піксель) всіх кольорів з множини Y ; V — множина нетермінальних імен (метасимволи), що присвоюються частинам зображення у процесі його генерування за допомогою граматики; $\varepsilon \in V$ — аксіома, що використовується для іменування всього, повністю згенерованого зображення; P — множина правил виводу, містить правила трьох типів: правила горизонтальної конкатенації, вертикальної конкатенації (об'єднання блоків зображення) та правила заміни. Кожну окрему множину цих правил позначатимемо відповідно P_h, P_v, P_s .

Правила заміни P_s мають вигляд $v \rightarrow e$, де $v \in V$ — деяке нетермінальне ім'я, а $e \in E$ — термінальне зображення.

Множина правил горизонтальної конкатенації P_h містить трійки нетермінальних імен вигляду $v \rightarrow v_l | v_r$. Наявність такого правила вказує на те, що будь-який прямокутний фрагмент зображення може отримати ім'я v , якщо його можна розбити вертикальною лінією на такі два прямокутні фрагменти, що лівий вже має ім'я v_l , а правий — v_r . Символ $|$ використовується для розділення пари нетермінальних імен у правилах горизонтальної конкатенації, а не для альтернативного запису кількох правил, як це прийнято у формальних граматиках.

Аналогічно множина правил вертикальної P_v конкатенації містить трійки нетермінальних імен вигляду $v \rightarrow \begin{smallmatrix} v_t \\ v_b \end{smallmatrix}$.

Наявність такого правила вказує на те, що будь-який прямокутний фрагмент зображення може отримати ім'я v , якщо його можна розбити горизонтальною лінією на такі два прямокутні фрагменти, що верхній вже має ім'я v_t , а нижній — v_b .

Мова граматики G складається з таких зображень, яким може бути присвоєне ім'я ϵ . Послідовність правил, застосованих до зображення x , результатом чого є присвоєння імені ϵ всьому зображенню, називається виводом зображення x в граматиці G [18, 19].

Навчання нейронної мережі. Одним з найбільш поширених алгоритмів навчання багатошарових нейронних мереж є алгоритм зворотного поширення помилки. У ньому обчислюється вектор градієнта поверхні помилок. Потім просуваємося на деяку величину в напрямку вектора (він буде вказувати напрямок найшвидшого спуску), де значення помилки буде вже менше. Таке послідовне просування поступово призведе до мінімізації помилки.

Введемо такі позначення. Матрицю вагових коефіцієнтів від входів до прихованого шару позначимо W , а матрицю ваг, що з'єднують прихований і вихідний шар — V . Для індексів прийемо такі позначення: входи будемо нумерувати тільки індексом i , елементи прихованого шару — індексом j , а виходи — індексом k . Число входів мережі — n , число нейронів у прихованому шарі — m , число нейронів у вихідному шарі — p . Нехай мережа навчається на вибірці (X', D') , $t = \overline{1, T}$. Тоді алгоритм навчання багатошарового персептрона буде виглядати так [20].

Крок 1. Ініціалізація мережі. Ваговим коефіцієнтам привласнюються малі випадкові значення, наприклад, з діапазону $(-0.3, 0.3)$; задаються ϵ — параметр точності навчання, α — параметр швидкості навчання (зазвичай ≈ 0.1 і може ще зменшуватися в процесі навчання), N_{max} — максимально допустима кількість ітерацій.

Крок 2. Обчислення поточного вихідного сигналу. На вхід мережі подається один з образів навчальної вибірки, і визначаються значення виходів усіх нейронів нейромережі.

Крок 3. Налаштування синоптичних ваг. Розрахувати зміну ваг для вихідного шару нейронної мережі за формулами:

$$V_{jk}^{N+1} = V_{jk}^N - \alpha \frac{\partial E}{\partial V_{jk}}, \text{ де } \frac{\partial E}{\partial V_{jk}} = \delta_k y_j^c, \delta_k = (y_k - d_k) y_k (1 - y_k) \quad (5)$$

Розрахувати зміну ваг для прихованого шару за формулами:

$$W_{ij}^{N+1} = W_{ij}^N - \alpha \frac{\partial E}{\partial W_{ij}}, \quad (6)$$

$$\text{де } \frac{\partial E}{\partial W_{ij}} = \left(\sum_{k=1}^p \delta_k V_{jk}^{N+1} \right) y_j^c (1 - y_j^c) X_i.$$

Крок 4. Кроки 2–3 повторюються для всіх навчальних векторів. Навчання завершується після досягнення для кожного з навчальних образів значення функції помилки, яке не перевищує ϵ або після максимально допустимого числа ітерацій.

Вважаємо, що задана безконтекстна структура $G = \langle E, V, P, \epsilon \rangle$. Ми припускаємо, що зображення, подані на розпізнавання, складаються з елементарних, нецілісних частин. Для зображень текстових сторінок такими є зображення текстових літер та пробілів між рядками тексту та між ними. Кожна з цих неподільних частин має свій досконалий образ. У разі текстових сторінок вони мають зображення кожної літери, а також пробіли. Набір таких зображень позначається E набором опорних зображень.

Набір сегментів терміналів V має певну особливість. Сегменти з однаковими назвами мають однаковий розмір і відрізняються один від одного лише зсувом, тобто паралельним перенесенням усіх пікселів цього відрізка на один і той самий вектор. Сегменти з однаковою назвою відповідають одному еталонному зображенню з множини E .

І навпаки, кожен відрізок $e \in E$ відповідає групі сегментів набору терміналів V ($e \subset V$). На штучних зразках моделювалася ситуація з невідомими еталонними зображеннями. У цьому випадку, коли зображення спотворюється за допомогою адитивного гаусового шуму, вирішення завдання найбільш ймовірної оцінки опорних зображень кінцевих сегментів полягає у знаходженні середнього зображення літери, доступні в навчальній добірці зображень текстових рядків.

Для налаштування програм було вибрано рядки тексту без шуму. Для того щоб з цих рядків було проведено сегментацію та еталони, було встановлено так, щоб результат розпізнавання цих рядків збігався з введеною сегментацією. Отже, для навчання в мережі використовується приблизно $25 \times 32 = 800$ еталонів (букви латинського алфавіту). Для перегляду алгоритму використано зашумлені реальні зображення символів — інші рядки, які не приймають учасників у процесі налаштування еталонів.

Сегментація була введена для кожного рядка тексту, а стандарти налаштовані таким чином, що результат розпізнавання цих рядків збігався з введеною сегментацією. Всі інші лінії, які не брали участі у процесі встановлення стандартів, були використані для перевірки отриманих стандартів. Результат розпізнавання показаний на рис. 5, де верхній рядок — зашумлене зображення, нижній — результат розпізнавання.

capable by the mere power of his voice
capable by the mere power of his voice

Рис. 5. Результат розпізнавання зашумленого зображення

Нижче наведено огляд роботи розробленої нейро-нечіткої системи (з вікном 5×5 пікселів) з комерційним продуктом ABBYY FineReader 11 Corporate Edition (зображення тексту сторінки становить 702 символи, гаусів шум-0,03 від чорного рівень). Для практичної реалізації вибрано програмне середовище Matlab Simulink із вбудованими нечіткими логічними елементами Fuzzy Logic Toolbox.

Розроблена система		ABBYY FineReader 11	
Час роботи	% помилок	Час роботи	% помилок
40 s	16 %	< 2 s	2

Відсоток помилок — це відношення кількості неправильно розпізнаних символів до їх загальної кількості. Недостатньо висока якість розпізнавання розробленої системи, порівняно з комерційним продуктом, пояснюється невеликими розмірами вікна. Збільшення кількості внутрішніх шарів нейронної мережі дає змогу підвищити гнучкість нейромережевої системи OCR і дещо покращити кількість правильно розпізнаних символів, але за значно більший час.

Висновки. У завданні розпізнавання текстових зображень особливо актуальним є застосування загальних методів розпізнавання образів. Найбільш гнучким методом, який дає змогу вирішити цілу низку завдань для розпізнавання текстів, є застосування нейронних мереж. Розглянуто основні етапи обробки цифрових зображень для задач сегментації символів та подальшого розпізнавання текстів. Детально розглянуто алгоритм попередньої обробки з використанням нечіткої логіки та процес бінаризації зображення. Алгоритм нечіткої обробки побудований для визначення меж символів на зображенні. Розглядається завдання побудови граматики для реалізації структурного розпізнавання текстових зображень.

Час і просторова складність у випадку двовимірних безконтекстових граматик дуже високі ($O(N^2W^2(N+W))$ та $O(N^2W^2)$). Час і просторова складність цих алгоритмів визначаються насамперед кількістю фрагментів, які переглядаються у процесі їх роботи.

Розглянуті методи та алгоритми дають змогу наблизити методи розпізнавання символів до тих, якими користуються люди, оскільки, незважаючи на великі досягнення в цій галузі, не існує систем, які б могли зрівнятися у розпізнаванні тексту з людиною.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Muzzle point pattern recognition system using image pre-processing techniques / Kumar S., Chandrakar S., Panigrahi A., Singh S. K. 2017 Fourth International Conference on Image Information Processing (ICIIP), Shimla, India, 2017. Pp. 1–6.
2. Cloud based color coding scheme violator plate detection through character recognition using image processing / Balbin J. R., Sejera M. P., Martinez C. O. A., Cataquis N. A. M., Ontog L. M. H., Toribio J. K. 2017 7th IEEE International Conference on Control System, Computing and Engineering (ICCSCE), Penang, 2017. Pp. 253–257.
3. Antoshchuk S. The automatized systems with the visual information processing design. Proceedings of the International Conference Modern Problems of Radio Engineering, Telecommunications and Computer Science, 2004. Lviv-Slavsko, Ukraine, 2004. Pp. 268.
4. Forward scattering bistatic radar imaging method and practice data processing / Cao Y., Zhang T., Zhang S., Luo B. *Journal of Systems Engineering and Electronics*. April 2011. Vol. 22. № 2. Pp. 206–211.
5. Słomiński S. Potential resource of mistakes existing while using the modern methods of measurement and calculation in the glare evaluation. 2016 IEEE Lighting Conference of the Visegrad Countries (Lumen V4). Karpacz, 2016. Pp. 1–5.

6. Kalaykov I., Tolt G. Fast fuzzy signal and image processing hardware. 2002 Annual Meeting of the North American Fuzzy Information Processing Society Proceedings. NAFIPS-FLINT 2002 (Cat. No. 02TH8622), 2002. Pp. 7–12.
7. Dorosinskiy L., Myasnikov F. Radarsignal classification algorithms synthesis and analysis. 2017 IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC). Chongqing, 2017. Pp. 119–122.
8. Fuzzy image processing in quality control application / Marino P., Pastoriza V., Santamarfa M., Martinez E. Sixth International Conference on Computational Intelligence and Multimedia Applications (ICCIMA'05), 2005. Pp. 55–60.
9. Chacon M. I., Aguilar L., Delgado A. Definition and applications of a fuzzy image processing scheme. Proceedings of 2002 IEEE 10th Digital Signal Processing Workshop, 2002 and the 2nd Signal Processing Education Workshop, 2002. Pp. 102–107.
10. Edge-Detection Method for Image Processing Based on Generalized Type-2 Fuzzy Logic / Melin P., Gonzalez C. I., Castro J. R., Mendoza O. and Castillo O. *IEEE Transactions on Fuzzy Systems*. Dec. 2014. Vol. 22. № 6. Pp. 1515–1525.
11. Chellappa R. et al. Towards the design of an end-to-end automated system for image and video-based recognition. 2016 Information Theory and Applications Workshop (ITA), La Jolla, CA, 2016. Pp. 1–7.
12. Peleshko D., Rak T., Izonin I. Image Superresolution via Divergence Matrix and Automatic Detection of Crossover. *International Journal of Intelligent Systems and Applications (IJISA)*. 2016. Vol. 8. No. 12. Pp. 1–8. DOI: 10.5815/ijisa.2016.12.01.
13. Ito N., Hagiwara M. Image description generation without image processing using fuzzy inference. 2012 IEEE International Conference on Fuzzy Systems, Brisbane, QLD, 2012. Pp. 1–8.
14. Vivona L. et al. Unsupervised clustering method for pattern recognition in IIF images. 2016 International Image Processing, Applications and Systems (IPAS), Hammamet, 2016. Pp. 1–6.
15. A novel night vision image color fusion method based on scene recognition / Qu Z., Xiao G., Xu N., Diao Z., Jia-Zhou H. 2016 19th International Conference on Information Fusion (FUSION), Heidelberg, 2016. Pp. 1236–1243.
16. Calvo-Zaragoza J., Toselli A. H., Vidal E. Handwritten Music Recognition for Mensural Notation: Formulation, Data and Baseline Results. 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), Kyoto, 2017. Pp. 1081–1086.
17. Tapia E., Rojas R. Recognition of on-line handwritten mathematical formulas in the E-chalk system. Seventh International Conference on Document Analysis and Recognition, 2003. Proceedings, 2003. Pp. 980–984.
18. Rousset F., Ducros N., Peyrin F. A Semi Nonnegative Matrix Factorization Method for Pattern Generalization in Single-Pixel Imaging. *IEEE Transactions on Computational Imaging*. Vol. PP. no. 99. Pp. 1–1.
19. Alvaro F., S'anchez J. A., Benedi J. M. Recognition of Printed Mathematical Expressions Using Two-Dimensional Stochastic Context-Free Grammars. 2011 International Conference on Document Analysis and Recognition, Beijing, 2011. Pp. 1225–1229.
20. Two-dimensional Picture Grammar models / Subramanian K. G., Geethalakshmi M., Nagar A. K., Lee S. K. 2008 Second UKSIM European Symposium on Computer Modeling and Simulation, Liverpool, 2008. Pp. 263–267.

REFERENCES

1. Kumar, S., Chandrakar, S., Panigrahi, A., & Singh, S. K. (2017). Muzzle point pattern recognition system using image pre-processing techniques. 2017 Fourth International Conference on Image Information Processing (ICIIP), Shimla, India, 1–6 (in English).
2. Balbin, J. R., Sejera, M. P., Martinez, C. O. A., Cataquis, N. A. M., Ontog, L. M. H., & Toribio, J. K. (2017). Cloud based color coding scheme violator plate detection through character recognition using image processing. 2017 7th IEEE International Conference on Control System, Computing and Engineering (ICCSCE), Penang, 253–257 (in English).
3. Antoshchuk, S. (2004). The automatized systems with the visual information processing design. Proceedings of the International Conference Modern Problems of Radio Engineering, Telecommunications and Computer Science, 2004. Lviv-Slavsko, Ukraine, 268 (in English).
4. Cao, Y., Zhang, T., Zhang, S., & Luo, B. (April 2011). Forward scattering bistatic radar imaging method and practice data processing: Journal of Systems Engineering and Electronics, 22, 2, 206–211 (in English).
5. Słomiński, S. (2016). Potential resource of mistakes existing while using the modern methods of measurement and calculation in the glare evaluation. 2016 IEEE Lighting Conference of the Visegrad Countries (Lumen V4). Karpacz, 1–5 (in English).
6. Kalaykov, I., & Tolt, G. (2002). Fast fuzzy signal and image processing hardware. 2002 Annual Meeting of the North American Fuzzy Information Processing Society Proceedings. NAFIPS-FLINT 2002 (Cat. No. 02TH8622), 7–12 (in English).
7. Dorosinskiy, L., & Myasnikov, F. (2017). Radarsignal classification algorithms synthesis and analysis. 2017 IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC). Chongqing, 119–122 (in English).
8. Marino, P., Pastoriza, V., Santamarfa, M., & Martinez, E. (2005). Fuzzy image processing in quality control application. Sixth International Conference on Computational Intelligence and Multimedia Applications (ICCIMA'05), 55–60 (in English).
9. Chacon, M. I., Aguilar, L., & Delgado, A. (2002). Definition and applications of a fuzzy image processing scheme. Proceedings of 2002 IEEE 10th Digital Signal Processing Workshop, 2002 and the 2nd Signal Processing Education Workshop, 102–107 (in English).
10. Melin, P., Gonzalez, C. I., Castro, J. R., Mendoza, O., & Castillo, O. (Dec. 2014). Edge-Detection Method for Image Processing Based on Generalized Type-2 Fuzzy Logic: IEEE Transactions on Fuzzy Systems, 22, 6, 1515–1525 (in English).
11. Chellappa, R. et al. (2016). Towards the design of an end-to-end automated system for image and video-based recognition. 2016 Information Theory and Applications Workshop (ITA), La Jolla, CA, 1–7 (in English).
12. Peleshko, D., Rak, T., & Izonin, I. (2016). Image Superresolution via Divergence Matrix and Automatic Detection of Crossover: International Journal of Intelligent Systems and Applications (IJISA), 8, 12, 1–8. DOI: 10.5815/ijisa.2016.12.01 (in English).
13. Ito, N., & Hagiwara, M. (2012). Image description generation without image processing using fuzzy inference. 2012 IEEE International Conference on Fuzzy Systems, Brisbane, QLD, 1–8 (in English).
14. Vivona, L. et al. (2016). Unsupervised clustering method for pattern recognition in IIF images. 2016 International Image Processing, Applications and Systems (IPAS), Hammamet, 1–6 (in English).

15. Qu, Z., Xiao, G., Xu, N., Diao, Z., & Jia-Zhou, H. (2016). A novel night vision image color fusion method based on scene recognition. 2016 19th International Conference on Information Fusion (FUSION), Heidelberg, 1236–1243 (in English).
16. Calvo-Zaragoza, J., Toselli, A. H., & Vidal, E. (2017). Handwritten Music Recognition for Mensural Notation: Formulation, Data and Baseline Results. 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), Kyoto, 1081–1086 (in English).
17. Tapia, E., & Rojas, R. (2003). Recognition of on-line handwritten mathematical formulas in the E-chalk system. Seventh International Conference on Document Analysis and Recognition, 2003. Proceedings, 980–984 (in English).
18. Rousset, F., Ducros, N., & Peyrin, F. A Semi Nonnegative Matrix Factorization Method for Pattern Generalization in Single-Pixel Imaging: IEEE Transactions on Computational Imaging, PP, 99, 1–1 (in English).
19. Alvaro, F., S´nchez, J. A., & Benedi, J. M. (2011). Recognition of Printed Mathematical Expressions Using Two-Dimensional Stochastic Context-Free Grammars. 2011 International Conference on Document Analysis and Recognition, Beijing, 1225–1229 (in English).
20. Subramanian, K. G., Geethalakshmi, M., Nagar, A. K., & Lee, S. K. (2008). Two-dimensional Picture Grammar models. 2008 Second UKSIM European Symposium on Computer Modeling and Simulation, Liverpool, 263–267 (in English).

doi: 10.32403/0554-4866-2021-1-81-72-88

NEURAL NETWORK METHODS OF TEXT IMAGE RECOGNITION

O. V. Tymchenko¹, B. M. Havrysh², B. V. Durnyak¹

¹*Ukrainian Academy of Printing,
19, Pid Holoskom St., Lviv, 79020, Ukraine*

²*Lviv Polytechnic National University,
12, St. Bandera St., Lviv, 79000, Ukraine
o_tymch@ukr.net*

The paper considers the methodology of improving the quality of recognition of text images, in particular distorted by noise and with geometric distortions due to poor scanning or photography. It has been shown that the problem of text document recognition is closely related to the problem of pattern recognition, so that methods of comparison with a sample, statistical and based on neural networks can be applied to text recognition. The comparison of methods shows the expediency of using structural recognition methods with the use of neural networks.

Methods and algorithms for constructing a fuzzy text image recognition system are considered, a generalized structural scheme of a text image recognition system and functions of the main nodes are presented. Methods of pre-processing of text images in order

to improve the image quality, convert the initial color image into images in grayscale and reduce noise using linear (Gaussian filter) and nonlinear median filtering. An algorithm for preliminary fuzzy image processing for their segmentation is presented. Features of binarization of such images by means of a threshold surface that allows reducing the influence of unevenness of illumination are shown. A fuzzy processing algorithm has been created to highlight boundaries and character segmentation.

For structural recognition of text images the methods of creation of the corresponding grammar with application of the theory of graphs and methods of the theory of formal languages and grammars are considered. The used methods of horizontal and vertical concatenation of image elements are shown. Recognition consists in finding the best in a certain sense output of an image in a given grammar. Methods of training a neural network according to the algorithm of error backpropagation are considered and its testing is carried out. Segmentation was introduced for each line of text, and the methods were configured so that the result of line recognition coincided with the entered segmentation.

The result of noise image recognition is presented. A comparison of the work of the developed neuro-fuzzy system (with a window of 5x5 pixels) with the commercial product ABBYY FineReader 11 Corporate Edition (the image of the text of the page is 702 characters, Gaussian noise - 0.03 from the black level). For practical implementation, the Matlab Simulink software environment with built-in fuzzy logic elements Fuzzy Logic Toolbox was chosen. It is concluded that increasing the number of inner layers of the neural network allows to improve the number of correctly recognized symbols, but for a much longer time. The time and spatial complexity of the developed algorithms are determined by the number of fragments that are viewed in the process of their work.

Keywords: *recognition system, fuzzy and noisy image processing, text image recognition, structural recognition, neural networks.*

Стаття надійшла до редакції 18.05.2021.

Received 18.05.2021.