

УДК 004.93

Мельник К.В., Мельник В.М., Коптюк Ю.Ю.
Луцький національний технічний університет

ДОСЛІДЖЕННЯ МЕТОДІВ РОЗПІЗНАВАННЯ ЗОБРАЖЕНЬ НА ОСНОВІ НЕЙРОННИХ МЕРЕЖ

Мельник К.В., Мельник В.М., Коптюк Ю.Ю. Дослідження методів розпізнавання зображень на основі нейронних мереж. У роботі наведено огляд існуючих підходів до вирішення задачі розпізнавання зображень. Описані основні методи, що використовуються в задачі розпізнавання зображень, визначено етапи процесу розпізнавання зображень. Розкрито особливості використання, переваги та недоліки зазначених методів. Зроблено висновок щодо необхідності подальшого розроблення алгоритмів розпізнавання на основі зазначених методів, що були б простими в реалізації, ефективними, мали низькі обчислювальні витрати при навчанні та високу якість класифікації в реальних завданнях.

Ключові слова: розпізнавання зображень, метод опорних векторів, штучні нейронні мережі, багатослойний перцептрон.

Мельник Е.В., Мельник В.М., Коптюк Ю.Ю. Исследование методов распознавания изображений на основе нейронных сетей. В работе предоставлен обзор существующих подходов к решению задачи распознавания изображений. Описаны основные методы, используемые в задаче распознавания изображений, определены этапы процесса распознавания изображений. Раскрыты особенности использования, преимущества и недостатки указанных методов. Сделан вывод о необходимости дальнейшей разработки алгоритмов распознавания на основе указанных методов, которые были бы простыми в реализации, эффективными, имели низкие вычислительные затраты при обучении и высокое качество классификации в реальных задачах.

Ключевые слова: распознавание изображений, метод опорных векторов, искусственные нейронные сети, многослойный перцептрон.

Melnyk K.V., Melnyk V.M., Koptiuk Yu.Yu. Investigation of the images recognition methods based on neural networks. This paper provides an overview of existing approaches to solve the images recognition problem. There are described basic methods to use them for solving the problem of image recognition and the stages used in the image recognition process are also defined. The features of use are revealed with their advantages and disadvantages concerned to these methods. The conclusion about the need of the further recognition algorithms development based on these methods mentioned above is made that would be simple in implementation, effective, with low computational costs for training and gave high quality classification in real solved tasks.

Key words: image recognition, reference vectors method, artificial neural networks, multilayer perceptron.

Актуальність та постановка проблеми. На сучасному етапі розвитку та впровадження інформаційних технологій у різних галузях господарства на одне з перших місць виходить проблема оперативного одержання достовірної інформації про стан певних ресурсів, усебічного наукового обґрунтування з планування їх ефективного використання. Для ефективного виконання поставлених завдань у сучасних умовах потрібно збирати, аналізувати і класифікувати значні обсяги інформації з високою періодичністю та якістю, що сьогодні можна виконати тільки з використанням інформаційних технологій, що базуються на інтелектуальних системах, зокрема нейромережах.[1]

Створення штучних систем розпізнавання образів є складною теоретичною та технічною проблемою. Необхідність у такому розпізнаванні виникає у найрізноманітніших галузях: системах відеоспостереження, технічної діагностики, прогнозування і діагностиці в медицині, у плануванні, пошуку в геології, прогнозуванню у хімії та ін.

Аналіз останніх досліджень та публікацій. Застосування інтелектуальних технологій (ІТ) для розв'язання задач класифікації нині пов'язують з застосуванням обчислювального інтелекту, який також пов'язують з теорією нечітких множин та нейрокомп'ютигом. Одним з ефективних методів класифікації об'єктів є кластеризація, яка, зокрема, вміщує ієрархічні та нечіткі методи. Наявність кластерів, які найчастіше представлено у вигляді дерева, дає змогу сформулювати систему правил, на підставі якої з наявної множини даних беруть знання, тобто певні приховані закономірності, які притаманні об'єкту. Кожний з цих методів має свої недоліки та переваги, алгоритми досить часто мають евристичний характер, але, як показано в працях [1—3], у багатьох випадках задачу класифікації розв'язати вдається. Надзвичайно важливим є визначення системи ознак, які до того ж можуть бути зашумленими, або визначеними нечітко (у вигляді інтервалів) і т. ін.

Мета дослідження – огляд основних методів розпізнавання зображень прописних букв, розкрити їхні особливості використання, переваги та недоліки.

Виклад основного матеріалу дослідження та обґрунтування отриманих результатів.

Розпізнавання образів (об'єктів) – задача ідентифікації об'єкта або визначення будь-яких його властивостей по його зображенню (оптичне розпізнавання) або аудіо записи (акустичне розпізнавання) та інші характеристики.

Базою для вирішення завдань віднесення об'єктів до того чи іншого класу послужили результати класичної теорії статистичних рішень. В її рамках будувалися алгоритми, що забезпечують на основі експериментальних вимірювань параметрів (ознак), що характеризують цей об'єкт, а також деяких апріорних даних, що описують класи, визначення конкретного класу, до якого може бути віднесений об'єкт, що розпізнається.

Методи розпізнавання образів. Метод перебору. У цьому випадку проводиться порівняння з базою даних, де для кожного виду об'єктів представлені різноманітні модифікації відображення. Наприклад, для оптичного розпізнавання образів можна застосувати метод перебору виду об'єкта під різними кутами, масштабами, зміщеннями, деформаціями і т. д. Для букв потрібно перебирати шрифт, властивості шрифту. У разі розпізнавання звукових образів, відповідно, відбувається порівняння з деякими відомими шаблонами (наприклад, слово, вимовлене кількома людьми).

Другий підхід – проводиться більш глибокий аналіз характеристик образу. У разі оптичного розпізнавання це може бути визначення різних геометричних характеристик. Звуковий зразок в цьому випадку піддається частотному, амплітудному аналізу і т. д.

Наступний метод – використання штучних нейронних мереж (ШНМ). Цей метод вимагає або великої кількості прикладів завдання розпізнавання при навчанні, або спеціальної структури нейронної мережі, яка враховує специфіку даного завдання. Проте, його відрізняє більш висока ефективність і продуктивність.

Так як задачі розпізнавання об'єктів полягає у класифікації зображень на основі певних критеріїв, то важливим етапом є вибір оптимального класифікатора. Серед існуючих методів класифікації можна виділити наступні:

- ймовірнісний критерій якості класифікації;
- оптимальна стратегія статистичної класифікації;
- класифікатор Байеса;
- мінімакний класифікатор;
- класифікатор Неймана-Пірсона.

За результатами аналізу останніх досліджень та публікацій можна зробити висновок, що до найперспективніших підходів що до розв'язання задач розпізнавання рукописних символів у разі автоматизованого уведення у комп'ютер текстової інформації належить використання алгоритму логістичної регресії та ШНМ. В обох підходах важливим є машинне навчання.

Задача машинного навчання, а саме навчання з учителем (supervised learning), коли система навчається за допомогою наявної множини прикладів, описується так: метою є, з урахуванням навчального набору, навчити функцію $h: X \rightarrow Y$ так, що $h(x)$ є "добрим" показником для відповідних значень Y . З історичних причин ця функція h називається гіпотезою. Відображення цього процесу подано на рис. 1.

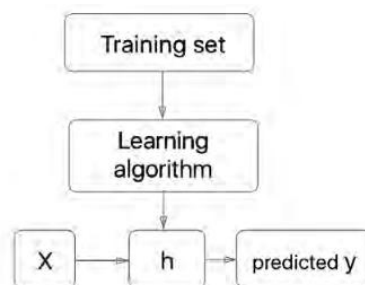


Рис 1. Узагальнена схема використання гіпотези

Нашим завданням було побудувати нейромережу, яка б уміла розпізнавати рукописні букви від a до $я$. Приклад набору зображень наведено на рис. 2. Для розпізнавання використано багат шаровий перцептрон та згорткову нейронну мережу, які навчали методом зворотного поширення похибки. Цей метод можна застосовувати для навчання одно - та багат шарових нейронних мереж, у яких немає прямих і зворотних зв'язків, що перестрибують через шари. Кожен нейрон попереднього шару зв'язаний тільки з нейронами наступного шару. Метод зворотного

поширення похибки полягає у послідовному поширенні похибки від вищих шарів до нижчих. Відповідно, корекцію ваг визначають за допомогою градієнтних методів. Для нейронів використовували функції активації SoftMax та ReLU (рис.3).

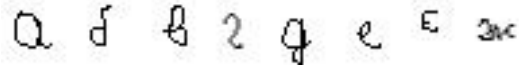


Рис. 2. Зразки для навчання

Softmax - це узагальнення логістичної функції для багатовимірного випадку. Функція перетворює вектор z розмірності K в вектор σ тієї ж розмірності, де кожна координата σ_i і σ_i отриманого вектора представлена дійсним числом в інтервалі $[0,1]$ і сума координат дорівнює 1. Координати σ_i обчислюються наступним чином:

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_{k=1}^K e^{z_k}}$$

Функція Softmax застосовується в машинному навчанні для задач класифікації, коли кількість можливих класів більше двох (для двох класів використовується логістична функція). Координати σ_i і σ_i , отриманого вектора, при цьому трактуються як ймовірності того, що об'єкт належить до класу i . Вектор-стовпець z при цьому розраховується наступним чином:

$$z = w^T x - \theta$$

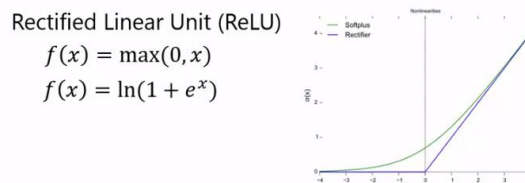


Рис. 3 Функція активації ReLU

де x - вектор-стовпець ознак об'єкта розмірності $M \times 1$; w^T - транспонована матриця вагових коефіцієнтів ознак, що має розмірність $K \times M$; θ - вектор-стовпець з граничними значеннями розмірності $K \times 1$, де K - кількість класів об'єктів, а M - кількість ознак об'єктів.

Часто Softmax використовується для останнього шару глибоких нейронних мереж для задач класифікації. Для навчання нейронної мережі при цьому в якості функції втрат використовується перехресна ентропія.

Корекцію ваг під час навчання нейромережі обчислюють за формулою

$$\Delta w^t = -\alpha \frac{\partial E}{\partial w} + M \Delta w^{t-1},$$

Де t – номер кроку навчання; Δw^t – значення корекції ваг на кроці t ; $\frac{\partial E}{\partial w}$ – градієнт похибки; α – швидкість навчання, M – момент навчання.

Нейромережі названого типу та процес їхнього навчання характеризують такими загальними глобальними параметрами: швидкість навчання α – відображає, на скільки ми враховуємо запропоновані алгоритмом поправки; момент навчання M – інертність алгоритму, міра врахування попередніх значень ваг кожного нейрона; очікувана похибка ϵ – значення похибки, у разі досягнення якого навчання мережі припиняється; кількість нейронів прихованого шару.

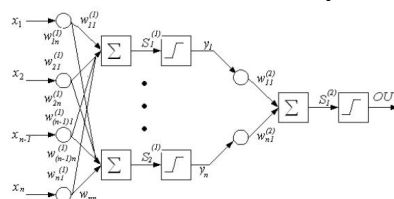


Рис. 4. Структура багатозарового персептрона

У процесі виконання поставленого завдання розроблено програмне забезпечення для розпізнавання рукописного тексту. Розроблялися та перевірялися дві нейронні мережі, на основі багат шарового перцептрона і згорткової мережі. Після закінчення навчання мережі, якість і швидкість роботи нейронних мереж порівнюється та вносяться певні корективи в алгоритм для покращення характеристик мережі. Середовищем створення нейронної мережі було обрано Jupiter Notebook з Python 3.5. Використовувалися вільні бібліотеки Keras і TensorFlow для написання коду програми.

Як бачимо з рис.5, а)-б), рельєф функції ефективності дуже складний, значно залежить від навчальної вибірки і навряд чи піддається аналітичному дослідженню, не говорячи про загальноприйняте "вгадування" параметрів. Зазначимо, що табулювання зі щораз меншим кроком виявляє нові вузькі ущелини та піки.

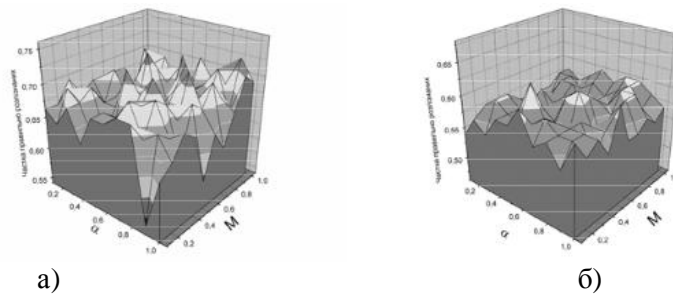


Рис. 5. Функція ефективності розпізнавання

Крім того, для тієї ж навчальної вибірки, але різних значень очікуваної похибки залежність ефективності розпізнавання від швидкості та моменту навчання теж буде дуже відрізнятися, що проілюстровано на рис. 6, а)-б), очікувана похибка для яких становить 0,01 та 0,02, відповідно.

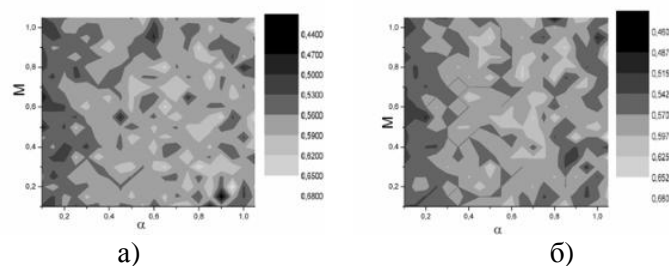


Рис. 6. Функція ефективності розпізнавання

Поліпшити якість навчання мережі можна шляхом зміни гіперпараметрів: кількість епох навчання, розмір міні-вибірки, кількість нейронів у вхідному шарі, кількість прихованих шарів

Дослідження оптимальних значень глобальних параметрів відбувалося для повнозв'язних і не повнозв'язних нейромереж, при цьому в останньому випадку для кожної окремої підмережі підбирали свій набір параметрів. Використовували простий генетичний алгоритм (ГА). Щоб отримати загальне уявлення про простір пошуку, виконували серію послідовних запусків. Оскільки нам потрібні оптимальні параметри для роботи в середовищі з великим впливом випадкових, малопередбачуваних факторів, зокрема не можливо сказати наперед, які варіанти написання будуть, то для роботи ГА використовували таку стохастичну процедуру:

1. Створюють популяцію різноманітних наборів параметрів.

2. Для кожного такого набору – особини в термінах ГА, навчальну вибірку випадковим способом розбивають на дві групи: група для навчання нейромережі та група для пробного розпізнавання навченою мережею. Процес повторюють декілька разів, визначають середню ефективність розпізнавання нейромережі для різних навчальних наборів.

3. Отриману на попередньому етапі ефективність розпізнавання нейромережею невідомих зразків використовують як пристосованість цієї особини. Виконують один крок простого ГА, генеруючи нову популяцію.

4. Переходять до кроку 2. Розбиття навчальної вибірки на дві випадкові групи є важливим, оскільки в іншому випадку – фіксованого розбиття, ми отримаємо мережу, оптимізовану під конкретну задачу – вчитися на одній множині так, щоб розпізнавати іншу.

Для поліпшення роботи мережі було проведено серію експериментів, в кожному з яких мінявся один з гіперпараметрів.

Кількість епох навчання. Оцінимо вплив кількості епох навчання на якість навчання мережі. Кількість епох задається в аргументі `epochs` методу `model.fit` :

`model.fit(X_train, Y_train, batch_size=200, epochs=XXX, validation_split=0.2, verbose=2)`

Навчали мережу протягом 50, 75, 100 і 125 епох. Вибиралась кількість епох, при якому найвища точність роботи мережі на тестових даних. Найкращий результат проявився при використанні 100 епох. Якість навчання сягала 96%. При 125 епохах запустився процес перенавчання нейронної мережі.

Розмір міні-вибірки. Оцінено вплив розміру міні-вибірки на якість навчання мережі. Розмір задається в аргументі `batch_size` методу `model.fit` :

`model.fit(X_train, Y_train, batch_size=XXX, epochs=100, validation_split=0.2, verbose=2)`

Використовувалися розміри міні-вибірки 50, 100, 200 і 400. При використанні максимального значення навчання мережі здійснювалося швидше, але робота ПК була сповільнена. Якість навчання залишалась при зміні параметрів сталою.

Зміна кількості нейронів у вхідному шарі. Кількість нейронів задається при створенні вхідного шару:

`model.add(Dense(XXX, input_dim=784, activation="relu", kernel_initializer="normal"))`

Використано значення 500, 700, 900, 1200. В даному випадку найоптимальніше по параметрах швидкість/якість підходить значення 900. При якому навчання здійснюється за короткий час та з точністю 96,7%

Додавання прихованого шару. Додано в мережу прихований шар:

**`model.add(Dense(800, input_dim=784, activation="relu", kernel_initializer="normal"))`
`model.add(Dense(600, activation="relu", kernel_initializer="normal"))` **`model.add(Dense(10, activation="softmax", kernel_initializer="normal"))`****

Якість навчання суттєво покращилася, хоча швидкість навпаки - сповільнилася.

Завдяки використанню такого методу вдалося досягнути ефективності розпізнавання 94 ± 5 % для багатошарового перцептронну. Для згорткової мережі результати ще кращі 97 ± 3 %, при чому кожна з підмереж досягає ефективності 95 – 99 % розпізнавання “свого” символу.

У процесі роботи з нейромережами експериментально з'ясовано, що: важко досягти якісного розпізнавання таких символів, як “о”, “з”, ”ж”, “ш”, “ф”, “я”; інші букви алфавіту розпізнавалися з невеликим набором даних для навчання.

Також продемонстровано наявність широкого простору для оптимізації за часом навчання та роботи нейромережі. Цей висновок ґрунтується на тому, що приблизно тієї ж ефективності розпізнавання можна досягнути при різних наборах параметрів та різних кількостях нейронів.

Висновки та перспективи подальшого дослідження. В результаті виконаного дослідження було проаналізовано переваги та недоліки методів розпізнавання рукописного тексту. Було побудовано декілька різних нейронних мереж та порівняно їхню продуктивність.

В майбутньому дані проведеного дослідження можна використати для розробки програм машинного розпізнавання тексту, полегшеного введення тексту за допомогою сенсорних технологій. Перспективним є дослідження продуктивності різних топологій нейронних мереж для виконання даного завдання.

1. traffic from January 2014 to September 2017, by month [Електронний ресурс] — Режим доступу: <https://www.statista.com/statistics/420391/spam-email-traffic-share/>

2. Vikas P. Deshpande. An Evaluation of Naive Bayesian Anti-Spam Filtering Techniques / Vikas P. Deshpande, Robert F. Erbacher, Chris Harris // Proceedings of the 2007 IEEE Workshop on Information Assurance United States Military Academy, West Point, 2007. — NY 20—22 June. — Режим доступу: <http://digital.cs.usu.edu/~erbacher/publications/Bayes-Vikas2.pdf>.

3. Graham P. A Plan for Spam / P. Graham, 2002. [Електронний ресурс] — Режим доступу: <http://www.paulgraham.com/spam.html>.

4. Мироненко А. Н. Алгоритм контентной фильтрации спама на базе совмещения метода опорных векторов и нейронных сетей : автореф. дис. на соискание науч. степени канд. техн. наук: спец. 05.13.19 «Методы и системы защиты информации, информационная безопасность» / А. Н. Мироненко. — СПб., 2012. — 18 с.

5. Терейковский І. Методология классификации листов электронной почты с использованием нейронных сетей. / І. Терейковский // Захист інформації – 2013, – Том 15, №2, – С. 115-122.

6. Кузьма К., Зівенко В. Аналіз методів фільтрації електронної пошти від спаму. / К. Кузьма, В. Зівенко // Науковий журнал – Геометричне моделювання та інформаційні технології № 1 (3), квітень 2017. – Миколаїв : МНУ імені В. О. Сухомлинського, 2017. — С. 84-89.

7. Dua, D. and Karra Taniskidou, E. (2017). UCI Machine Learning Repository [<http://archive.ics.uci.edu/ml>]. Irvine, CA: University of California, School of Information and Computer Science.