

Б.М. Гавриш¹, к.т.н., ст. викл, УАД, О.В. Тимченко^{1, 2}, д.т.н, професор,
О.О. Тимченко¹, аспірант

ДОСЛІДЖЕННЯ ОСОБЛИВОСТЕЙ СТРУКТУРНИХ ТА НЕСТРУКТУРНИХ МЕТОДІВ РОЗПІЗНАВАННЯ ТЕКСТОВИХ ЗОБРАЖЕНЬ

Анотація. Аналізується особливості зображень текстових документів та досліджуються методи структурного та не структурного розпізнавання текстових зображень. Зроблено висновок про доцільність використання в методах розпізнавання текстових зображень структурних методів.

Ключові слова: системи розпізнавання текстів, структурні та неструктурні методи розпізнавання, нейронні мережі.

Abstract. The peculiarities of images of text documents are analyzed and methods of structural and non-structural recognition of text images are investigated. A conclusion is made on the expediency of using structural methods for the recognition of textual images of methods.

Key words: text recognition systems, structural and non-structural methods of recognition, neural networks.

Вступ

Завдання розпізнавання людина вирішує неперервно. При цьому інформація, що надходить з органів зору, обробляється мозком, який забезпечує прийняття рішення про об'єкт розпізнавання. З розвитком інформаційних технологій стало можливим полегшити, прискорити та підвищити якість розпізнавання. Системи розпізнавання літер працюють разом зі сканерами – пристроями, які використовуються для введення до комп'ютера друкованих зображень і текстів. При введенні друкованого тексту сканер формує лише графічне зображення; для того щоб створити текстовий документ, з яким може працювати текстовий редактор, необхідно впізнати на цьому зображенні окремі літери. Існує цілий ряд програмних продуктів, які практично автоматизували процеси розпізнавання, зокрема зображень текстів. Проте забезпечити задовільний результат у випадку спотворень зображень текстів різного типу (геометричних, шумів та подібних) вдається не завжди. Слід зазначити, що здатність людини читати друкований текст низької якості перевершує можливості комп'ютерних систем [1].

Можна виділити наступні методи розпізнавання зображень текстів.

У методі перебору проводиться порівняння виділеного елемента з базою даних, де для кожного виду об'єктів представлені об'єкти під різними кутами,

¹, Українська академія друкарства

² Uniwersytet Warmińsko-Mazurski w Olsztynie

масштабами, зміщеннями, деформаціями і т. д. Для букв слід визначити шриффт, властивості шрифту і т. д.

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

Новітні методи використовують штучні нейронні мережі. Хоча метод вимагає великої кількості прикладів при навчанні, але має 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.

Метою роботи є розгляд особливостей методів структурного та неструктурного розпізнавання зображень текстових документів.

Алгоритми розпізнавання текстових зображень

Процедури навчання розпізнаванню добре вивчені в рамках неструктурного розпізнавання. Очевидно, що навчання є найсуттєвішою частиною технології розпізнавання в цілому. Процедури власне розпізнавання, тобто функціонування вже навченої системи, є досить простими. З іншого боку процедури структурного розпізнавання є досить складними навіть тоді, коли йдеться про функціонування вже навченої системи розпізнавання. Проблеми навчання в структурному розпізнаванні значно складніші і менш досліджені [2-4].

Алгоритми розпізнавання зображень документів визначаються двома типами параметрів: еталонними зображеннями елементарних частин вхідних зображень та правилами побудови зображення в цілому з цих частин.

В загальному формальний опис задачі розпізнавання формулюється наступним чином. Нехай $D = \{d^{(1)}, \dots, d^{(N)}\}$ – множина об'єктів, $C = \{c_1, \dots, c_M\}$ – множина категорій, Φ – цільова функція, яка по парі $\langle d^{(j)}, c_{(i)} \rangle$, визначає, чи відноситься об'єкт $d^{(j)}$ до категорії $c_{(i)}$. Задача розпізнавання полягає у побудові функції $\bar{\Phi}$, яка максимально наближена до Φ .

Роботи, що мають на меті автоматичне налаштування правил граматики, зазвичай орієнтуються на розв'язання задач на точний збіг, мають справу зі скелетизованим зображенням, оперують з визначеним набором примітивів (відрізки, точки) і описують відносно прості об'єкти, що складаються з невеликої кількості примітивів [5]. Побудова правил полягає у визначенні взаємного розташування та типів примітивів на пред'явленому зразку і перетворення отриманих даних у певний визначений формат правил.

Більш широкі властивості відкриває метод побудови еталонних

елементарних частин зображень (шаблонів символів), якими можуть бути апроксимовані відповідні частини вхідного зображення:

1. задано певну кількість прикладів зображення кожного з символів (класи символів). Необхідно побудувати алгоритм, який для кожного з пред'явлених прикладів вказуватиме клас, до якого той належить. Цей алгоритм на стадії розпізнавання використовується як класифікатор зображень вхідних символів.

2. Задано певну кількість прикладів зображення певного символу. Необхідно побудувати шаблон цього символу так, щоб мінімізувати значення заданого критерію схожості даного символу на пред'явлені приклади [6].

Перший спосіб - вирішення задачі на точний збіг, коли до початку розпізнавання зображення сегментується, а зображення окремих символів розпізнаються за допомогою класифікатора, отриманого на етапі навчання.

Другий метод - вирішення задачі на найкращий збіг, тобто наскільки схожим є пред'явлений фрагмент зображення на кожен із можливих символів.

Недолік обох підходів у визначенні еталонів, які не можуть охопити всі можливі варіанти.

Вказаний недолік, очевидно, є результатом недосконалості постановки задачі розпізнавання. Така постановка не враховує факт, що на вхід алгоритму розпізнавання подається зображення документа в цілому, а не його фрагменти. Не враховується також кореляція між різними частинами зображення. Для врахування цього задачу настройки слід було б ставити так, щоб до навчальної вибірки входило вхідне зображення загалом, а не окремі його частини. Інакше кажучи, задачу настройки слід формулювати так: задано навчальну вибірку, яка містить одне або кілька зображень та відповідні їм дерева виводу. Задано також алгоритм розпізнавання, визначений з точністю до еталонного вигляду символів. Задача полягає в побудові таких еталонів символів, щоб результатом розпізнавання кожного із зображень було відповідне йому, вказане в навчальній вибірці, дерево виводу.

Текстові зображення – один з чи не найпопулярніших об'єктів розпізнавання. Інтерес до цієї задачі виник ще на зорі обчислювальної техніки і зберігся дотепер, щоправда, постановки прикладних задач значно ускладнилися: здебільшого дослідників цікавить вже розпізнавання лише рукописних текстів або ж друкованих, але отриманих у складних умовах поганого освітлення, малої роздільної здатності тощо [1].

З точки зору структурного розпізнавання, текстові зображення є виразним і водночас досить простим прикладом складноструктурованого об'єкта. Відносна простота текстових зображень робить їх хорошим полігоном для випробування нових підходів до розпізнавання зображень, особливо ж різноманітних документів.

Водночас відомо, що задача знаходження апостеріорі найімовірнішого стану об'єкта (в нашому випадку – найімовірнішої послідовності еталонів літер) є окремим випадком байєсівської теорії прийняття рішень. З точки зору цієї теорії, знаходження найімовірнішої послідовності еталонів відповідає

дельтоподібній функції штрафу, згідно з якою правильна послідовність не штрафується, а всі інші штрафуються абсолютно однаково, незалежно від того, скільки літер в цих послідовностях розпізнано неправильно.

Функцію якості виводу можна задати безпосередньо, або скористатись з рекомендацій байєсівської теорії: відшукати апостеріорі найімовірнішу сукупність прихованих параметрів зображення, яке розпізнається. Ця рекомендація визначає визначення відповідної функції втрат.

Структурні методи розпізнавання зображень документів

Структурні (лінгвістичні) системи для побудови алгоритмів розпізнавання використовують речення, кожне з яких описує структуру (будову) об'єкта з непохідних ("атомарних") елементів. Ці речення складають спеціальну мову. Класифікація об'єкта виконується шляхом порівняння речення невідомого об'єкта з еталонними реченнями класів. Відносно розпізнавання текстів структурні методи розпізнавання текстових символів зберігають інформацію не про по-точкове написання символу, а про його топологію.

Структуровані зображення, зокрема тексти, таблиці тощо, створюються і читаються згідно з певними правилами, набір яких в [7] прийнято розуміти як певну граматику. Очевидно, й алгоритми розпізнавання зображень текстів повинні базуватись на використанні правил цієї граматики.

Використовують наступні формалізми:

- котрі базуються на використанні понять та методів теорії графів. Зображення при цьому представляється у вигляді деякого графа, а задачі розпізнавання визначають знаходження ізоморфізму еталонного та вхідного графів.

- формалізми, які базуються на апараті і методах теорії формальних мов та граматик. Зображення при цьому розглядається як слово у деякій формальній мові, яка задається за допомогою конструкцій, що є узагальненнями класичних контекстно-вільних мов та граматик Хомського. Розпізнавання ж полягає у відшуванні найкращого в певному значенні виводу зображення у заданій граматиці.

Використання теорії графів, приводить до алгоритмів з експоненційних часом їх. Прийнятний час розпізнавання можна досягнути у випадках:

- малого розміру графів, тобто вказівка для кожного зображення відповідного йому елемента зі скінченої і порівняно невеликої множини;
- графів спеціального вигляду: графи-ланцюги, графи-цикли, графи-дерева, плоскі графи.

Для формулювання задачі на найкращий збіг кожному правилу приписується штраф за його використання. Розпізнавання вхідного зображення полягає у відшуванні його виводу з мінімальним штрафом [8, 9].

Для задач, що описуються регулярними мовами та граматиками, використовується алгоритм, який полягає в упорядкуванні всіх фрагментів поля зору, які тільки можуть зустрітись у виводі, за їхніми розмірами. В

множину цих фрагментів входять як фрагменти, що відповідають елементарним частинам зображення (наприклад, зображенням літер в друкованому тексті) так і більші фрагменти, що позначають вже збудований фрагмент зображення і складаються з багатьох елементарних частин (наприклад, фрагмент зображення текстового рядка). Для кожного фрагмента, починаючи з найменших, вирішується питання, чи може він бути побудований з менших фрагментів (для яких це питання вже отримало відповідь). Після цього знаходиться штраф за найдешевшу побудову цього фрагмента. Так само, будується все зображення, яке також входить до множини фрагментів.

На першому етапі розпізнавання про певні фрагменти поля зору, а саме про ті, які відповідають елементарним, неподільним частинам зображення, приймається рішення щодо їх найменувань. Очевидно, що ці фрагменти будуть термінальними. Будь-який інший фрагмент може складатися в лише з термінальних фрагментів. Алгоритм розпізнавання будується так, щоб уникнути перегляду тих фрагментів, які не можуть бути виведені з термінальних. Для всіх пар фрагментів, що задовольняють правила граматики, генерується фрагмент – результат їх об'єднання, який додається до згаданого переліку. Ускладнення алгоритму порівняно з алгоритмом розв'язання задачі на точний збіг полягає у пошуку пар фрагментів, які можуть бути об'єднані, та у відповіді на запитання на кожному кроці алгоритму, чи є вже в поточному переліку щойно утворений фрагмент.

Висновок. Методи структурного розпізнавання зображень текстів повинні ґрунтуватись на відповідно створеній граматиці, що включає упорядковані фрагменти елементарних частин зображення (термінальні елементи) та правил їх використання.

1. *Тимченко О.В.* Алгоритми та функції інформаційної системи розпізнавання символів на основі методів поліпшення зображень / Тимченко О.В., Кульчицька І.О., Тимченко О.О. // Моделювання та інформаційні технології. Зб. наук. пр. ІПМЕ НАН України. – Вип.69. – К.: 2013. – С.167-173.
2. *Анисимов, Б. В.* Распознавание и цифровая обработка изображений: Учебное пособие для вузов / Б. В. Анисимов, В. Д. Курганов, В. К. Злобин. – М.: Высшая школа, 1983. – 295 с.
3. *Савчинський Б. Д.* Контекстно-вільні граматичні конструкції для розпізнавання зображень текстових та графічних документів. Дис. ... канд. техн. наук за спец. 05.13.23 - системи та засоби штучного інтелекту. Київ, 2007.
4. *Шлезингер М., Главач В.* Десять лекцій по статистическому и структурному распознаванию. - Киев: Наукова Думка, 2004. - 545 С.
5. *Myers R., Wilson R. C., Hancock E. R.* Bayesian graph edit distance // IEEE Trans. on PAMI. June 2000. Vol. 22, no. 6. Pp. 628-635.
6. *Корес G. E.* Multilevel character templates for document image decoding // IS&T/SPIE 1997 Intl. Symposium on Electronic Imaging: Science & Technology. San Jose, CA: February 8-14 1997.

7. *De Berg M., Van Kreveld M., Overmars M., Schwarzkopf O.* Computational geometry. Algorithms and applications. - Berlin, Heidelberg: Springer-Verlag, 2000. P. 367.
8. *Chou P. A., Kopec G. E.* A stochastic attribute grammar model of document production and its use in document image decoding // Document Recognition II / Ed. by L. Vincent, H. Baird. Vol. 2422 of SPIE Proc. 1995. Pp. 66–73.
9. *Kopec G. E., Chou P. A., Maltz D. A.* Markov source model for printed music decoding // Journal of Electronic Imaging.- January 1996.- Vol. 5, no. 1.- Pp. 7–14.

Поступила 3.04.2017г.