Навчання з неповною інформацією від учителя при розпізнаванні текстових зображень

Олефіренко С.А.

Міжнародний науково-навчальний центр
інформаційних технологій та систем, Київ, Україна

Анотація

Запропоновано постановку та алгоритм розв'язку задачі оцінки параметрів алгоритму розпізнавання текстового рядка на основі навчальної вибірки, що містить лише приклади зображень текстових рядків та відповідних їм послідовностей літер і не містить сегментації зображень на зображення окремих літер. Таке формулювання задачі навчання дозволяє значно спростити та прискорити побудову навчальної вибірки, звівши її до простого набору тексту, що відповідає навчальним зображенням.

1. Вступ

Задача розпізнавання тексту, незважаючи на свою популярність та уявну вивченість, все ще містить значний простір для наукових досліджень. Одночасно розпізнавання текстового рядка є яскравим і, напевно, одним з найпростіших прикладів в структурному розпізнаванні зображень. Саме з цієї причини ця задача є зручним полігоном для впровадження та експериментальної перевірки нових методів в структурному розпізнаванні.

Однією з галузей структурного розпізнавання, яка, на наш погляд, потребує значних досліджень, є навчання та оцінка параметрів розпізнаючих алгоритмів на основі навчальної вибірки. Як відомо, навчальна вибірка містить певну кількість зображень та відповідних їм результатів розпізнавання. У випадку розпізнавання зображення текстового рядка результатом є не лише послідовність літер, що йому відповідає, але й сегментація цього рядка на зображення окремих літер. При цьому до сегментації висуваються зазвичай доволі жорсткі вимоги: однакові літери в межах різних сегментів, що їх містять, повинні бути однаковим чином відцентровані. Це означає, що координати відповідних пікселів однакових літер на різних сегментах мають бути однаковими. Побудова такої сегментації потребує значних зусиль та часу вчителя.

В даній роботі ми пропонуємо постановку та алгоритм розв'язку задачі оцінки параметрів алгоритму розпізнавання текстового рядка на основі навчальної вибірки, що містить лише приклади зображень текстових рядків та відповідних їм послідовностей літер і не містить сегментації зображень на зображення окремих літер. Таке формулювання задачі навчання дозволяє значно спростити та прискорити побудову навчальної вибірки, звівши її до простого набору тексту, що відповідає навчальним зображенням.

Робота складається з шести розділів. Наступний присвячено основним означенням та постановці задачі, третій — її розв'язанню, четвертий та п'ятий відповідно — експериментальній перевірці алгоритмів та можливим застосуванням. В шостому розділі наведено висновки.

2. Означення та постановка задачі навчання

Введемо основні позначення, що будуть використовуватись в роботі.

Позначатимемо вхідні зображення через x. Всі вони мають однакову висоту H, але, взагалі кажучи, різну ширину. Вважатимемо, що незашумлене зображення, яке відповідає заданому текстовому рядку, є горизонтальною послідовністю еталонних зображень літер рядка, причому ці зображення не перекриваються, а можливі проміжки між ними заповнюються кольором фону. Вхідне зображення відрізняється від незашумленого зображення лише гаусівським шумом визначеної дисперсії σ, що накладається в кожному пікселі незалежно від інших.

Скінчену множину A_0 літер тексту називатимемо алфавітом. Послідовність елементів алфавіту $\bar{k} = (k_1, k_2, \ldots, k_L)$, $k_l \in A_0, l = \overline{1, L}$ називатимемо текстовим рядком. За допомогою $L_{\bar{k}}$ позначатимемо надалі довжину рядка $\bar{k}$.

Для формального опису проміжків між зображеннями літер введемо додатковий елемент алфавіту. Назвемо його *вставкою* і позначатимемо κ. Множину $A_0 \bigcup \{\kappa\}$, що складається з алфавіту A_0 та вставки κ, будемо позначати A, а її елементи

$a \in A$ називатимемо *символами*.

З кожним символом $a \in A$ пов'яжемо його еталонне зображення e_a висоти H та ширини $d(a)$, що залежить від літери. Ширини $d(a), a \in A$ еталонних зображень всіх символів вважатимемо фіксованим і відомими. Еталон вставки має ширину $d(\kappa) = 1$. Множину еталонних зображень позначатимемо E.

Називатимемо сегментом поіменований прямокутний фрагмент, який містить зображення певного символу. При цьому висота фрагмента збігається з висотою H вхідного зображення, а ширина — з шириною відповідного символу. Таким чином сегмент s визначається координатою свого лівого краю i та символом $a \in A$, зображення якого він містить.

Сегментацією зображення ширини W назвемо послідовність сегментів $\bar{s} = (s_1, \ldots, s_N)$ довільної довжини N, які покривають все зображення і розташовані впритул один до одного:

$$\begin{cases} i(s_1) = 0; \\ i(s_{n+1}) = i(s_n) + d(a(s_n)), n = \overline{1, N-1}; \\ i(s_N) + d(a(s_N)) = W. \end{cases}$$

Множину всіх сегментацій позначимо $\bar{S}$.

Задача розпізнавання зображення x полягає в пошуку найімовірнішої сегментації $\bar{s}^*$:

$$\bar{s}^* = \arg\max_{\bar{s}} p(x, \bar{s}; E) = \arg\max_{\bar{s}} p(\bar{s}) \cdot p(x|\, \bar{s}; E).$$

Умовна ймовірність $p(x|\, \bar{s}; E)$ визначається моделлю гаусового адитивного шуму і є відомою. Апріорні ймовірності $p(\bar{s})$ та еталонні зображення E є невідомими параметрами задачі, що й оцінюються в задачі навчання. При заданих параметрах ця задача розв'язується [1] за допомогою алгоритму динамічного програмування.

Перш ніж перейти до формулювання задачі навчання, відзначимо зв'язок між сегментаціями та послідовностями літер алфавіту. Довільній сегментації $\bar{s} = (s_1, \ldots, s_N)$ відповідає послідовність символів $\bar{a}(\bar{s}) = (a_1, \ldots, a_N | a_n = a(s_n), n = \overline{1, N})$. В свою чергу, цій послідовності символів відповідає послідовність літер, що отримується з неї видаленням всіх вставок. Таким чином, будь-яка послідовність літер $\bar{k} = (k_1, \ldots, k_N), k_n \in A_0$ пов'язана з множиною $\bar{S}(\bar{k})$ всіх сегментацій таких, що відповідні їм текстові рядки після видалення всіх вставок збігаються з $\bar{k}$.

Перейдемо до формулювання задачі навчання. Нехай

$$D = \begin{pmatrix} x^1 & x^2 & \ldots & x^M \\ \bar{k}^1 & \bar{k}^2 & \ldots & \bar{k}^M \end{pmatrix} -$$

навчальна вибірка, що складається з M вхідних зображень і M відповідних їм текстових рядків.

Ймовірність $p(x, \bar{k}; E)$ пари $(x, \bar{k})$ зображення x та текстового рядка $\bar{k}$, рівна сумарній ймовірності $\sum_{\bar{s} \in \bar{S}(\bar{k})} p(x, \bar{s}; E)$ усіх сегментацій, що відповідають рядку $\bar{k}$. Ймовірність вибірки $p(D, E)$ таким чином приймає вигляд

$$p(D) = \prod_{m=1}^{M} \sum_{\bar{s} \in \bar{S}(\bar{k}^m)} p(\bar{s}) \cdot p(x^m|\, \bar{s}; E).$$

Задача 1 *Задача навчання алгоритму розпізнавання полягає у знаходженні таких еталонів E^* та апріорних ймовірностей сегментацій $p^*(\bar{s})$, які максимізують ймовірність вибірки D:*

$$(E^*, p^*(\bar{s})) = \arg\max_{(E, p(\bar{s}))} \prod_{m=1}^{M} \sum_{\bar{s} \in \bar{S}(\bar{k}^m)} p(\bar{s}) \cdot p(x^m|\, \bar{s}; E). \tag{1}$$

3. Розв'язок задачі навчання

Нам невідомий алгоритм точного розв'язку задачі 1. В роботі для пошуку розв'язку було використано алгоритм самонавчання, описаний в [1]. Як відомо, цей алгоритм, взагалі кажучи, забезпечує пошук лише локального екстремуму. Водночас, з практичної точки зору це не становить значної проблеми, оскільки якість оцінки параметрів може бути легко проконтрольована візуально, а на основі результатів розпізнавання зображень з вибірки може бути оцінено якість розпізнавання зображень, що не увійшли до неї.

Водночас, алгоритм самонавчання не може бути використаний для вирішення задачі 1 безпосередньо, оскільки потребує експоненційних за розмірами вхідних зображень часу та пам'яті. В роботі запропоновано модифікацію цього алгоритму, що розв'язує ту саму задачу, але може бути ефективно реалізована.

4. Результати експериментів

Експерименти проводилися наступним чином: на вхід алгоритму подавалась вибірка, що складалась з зображень та відповідних їм текстових рядків, а також розміри еталонів всіх літер алфавіту. Експериментальну вибірку було розбито на дві частини: навчальну та тестову. Перша частина використовувалась для побудови еталонів символів, друга ж частина вибірки слугувала для оцінки якості розпізнавання за допомогою побудованих еталонів.

На рис. 1 наведено зразок тестових зображень, створених штучним шляхом із застосуванням моделі, описаної у [4]. Рис. 2 містить зображення еталонів символів, отриманих при навчанні. На

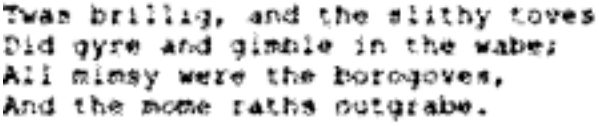

Рис. 1: *Приклад 1. Зразок тестових зображень.*

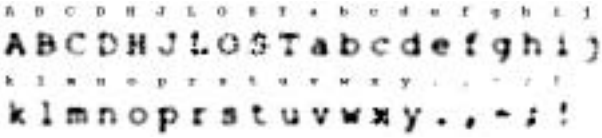

Рис. 2: *Приклад 1. Побудовані еталони.*

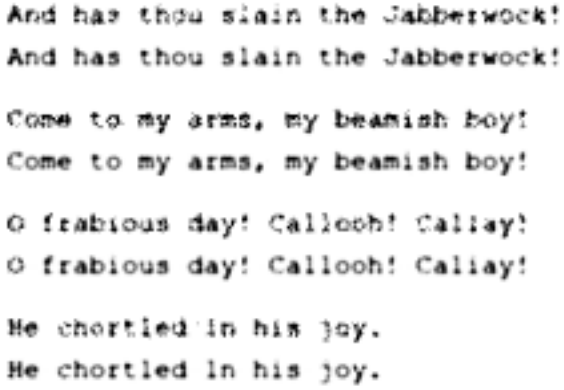

Рис. 3: *Приклад 1. Результати розпізнавання.*

рис. 3 наведено пари зображень текстових рядків, що демонструють результат розпізнавання текстових зображень. В кожній парі перше зображення — це вихідний рядок тестової вибірки, а друге — зображення рядка, склеєного з еталонів літер текстового рядка, що є результатом розпізнавання цього зображення. Помилка при розпізнаванні в цьому прикладі склала близько 1%.

Інший приклад, який ми наведемо, є реальним зображенням з сильною деградацією. Результати розпізнавання наведено на рис. 4 та рис. 5. Помилка розпізнавання тестової вибірки склала 3%, причому близько половини помилок виникло із-за того, що необхідні літери не були представлені в навчальній вибрці.

Взагалі помилка при розпізнаванні коливалась від $0,8-1\%$ для зображень, характер шуму яких відносно добре описується прийнятою в роботі моделлю, до $3-5\%$ для сильно зашумлених зображень або зображень, характер шуму яких значно відрізняється від моделі. Також слід відзначити швидкість збіжності алгоритму: потрібно від 3 до 5 ітерацій для того, щоб отримати результат, який вже не покращується, у тому сенсі, що функція правдоподібності мало змінюється. Тим не менш,

Рис. 4: *Приклад 2. Результати розпізнавання.*

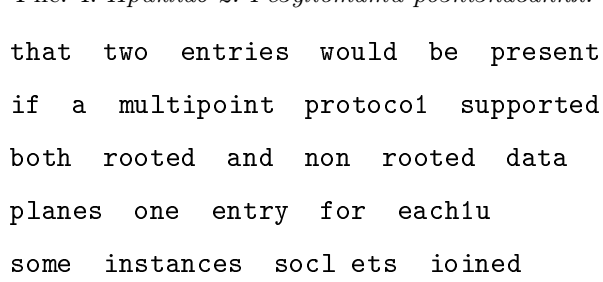

Рис. 5: *Приклад 2. Розпізнаний текст.*

отриманий результат може виявитись локальним максимумом функції правдоподібності, а не глобальним, що відображається у неправильній побудові декількох еталонів і, відповідно, неправильному розпізнаванні навчальної вибірки. У такому випадку алгоритм можна виштовхнути з локального максимуму реініціалізацією неправильно побудованих еталонів та апріорних ймовірностей сегментів.

5. Застосування алгоритму як допоміжного для алгоритмів настройки параметрів

Описаний підхід може використовуватись не лише як самостійний метод статистичного оцінювання еталонів, але й як допоміжний метод для інших алгоритмів навчання, наприклад алгоритмів настройки (див. [3]).

Алгоритми настройки не так сильно, як описаний нами, залежать від характеру зашумленості зображень, тому при значному відхиленні характеру зашумлення від прийнятої в роботі гаусо-

Рис. 6: *Приклад 3. Вхідне зображення.*

вої моделі забезпечують кращі результати. Недоліком алгоритмів настройки є вимога наявності в навчальній вибірці точної сегментації кожного з навчальних зображень.

Водночас на основі описаного нами підходу навіть для сильнозашумлених зображень можуть бути отримані сегментації достатньої точності. Побудова сегментації складається з двох етапів: спочатку на основі текстового рядка та вхідного зображення будуються еталони літер так, як це описано в даній роботі. На другому етапі на основі отриманих еталонів знаходиться найімовірніша сегментація серед тих, що відповідають введеному на першому етапі текстовому рядку. Ця сегментація і приймається в якості вхідної для наступного алгоритму настройки.

Приведемо приклад побудови таких сегментацій для зашумленого зображення (наведеного на рис. 6), характер шуму якого значно відрізняється від прийнятої в роботі моделі. На рис. 7 зображено результати розпізнавання, де перший рядок кожної групи — це вхідне зображення, другий — результат звичайного розпізнавання, третій — найімовірніша сегментація серед тих, що відповідають текстовому рядку з навчальної вибірки. Видно, що незважаючи на значну кількість помилок при розпізнаванні, побудовані сегментації є досить точними.

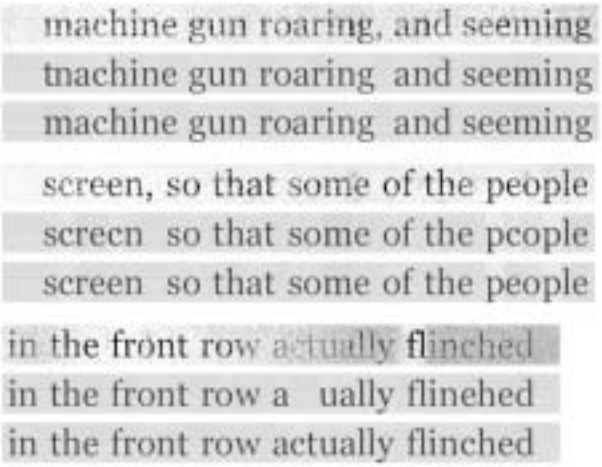

Рис. 7: *Приклад 3. Побудова точних сегментацій.*

6. Висновки

Підхід до оцінювання еталонів літер для задачі розпізнавання тексту, запропонований в роботі, дозволяє суттєво зменшити обсяг ручної роботи при підготовці навчальної вибірки, оскільки замість точної сегментації навчальних зображень на окремі літери на вхід алгоритму подаються лише текстові рядки, що відповідають цим зображенням.

Перспективним напрямком подальших досліджень в цій галузі є алгоритми навчання, в ході роботи яких автоматично налаштовуються не лише кольори (градації сірого) еталонних зображень літер, але й розміри цих зображень.

7. Література

[1] Шлезингер М., Главач В. Десять лекций по статистическому и структурному распознаванию.— Киев: Наукова думка, 2004.— С. 545.

[2] Ковалевский В.А. Методы оптимальных решений в распознавании изображений.— Москва: Наука, 1976.— С. 328.

[3] Савчинський Б.Д., Камоцький О.В. Настройка алгориту розпізнавання тексту // *Управляющие системы и машины.*— 2005.— №2.— С. 17–24.

[4] Pratek Sarkaar, Henry S. Baird, and Xiaohu Zhang. Training on severely text-line images. In *IAPR 7th International Conference on Document Analysis and Recognition (ICDAR03)*, pages 38–43, Edinburgh, Scotland, August, 2003.