

УДК 004.4

ОБЗОР МЕТОДОВ РАСПОЗНАВАНИЯ РУКОПИСНОГО ТЕКСТА

Гайдуков Н.П., Савкова Е.О.

Донецкий национальный технический университет
кафедра Автоматизированных систем управления

E-mail: gaydukov.nikita@gmail.com

Аннотация

Гайдуков Н.П., Савкова Е.О. Обзор методов распознавания рукописного текста. Рассмотрены существующие методы распознавания текстов. В результате анализа, были выделены преимущества и недостатки методов, а также был выбран оптимальный, который будет использоваться в дипломной работе.

Введение

Распознавание рукописного ввода — это способность компьютера получать и интерпретировать интеллектуальный рукописный ввод. Распознавание текста может производиться «оффлайновым» методом из уже написанного на бумаге текста или «онлайновым» методом считыванием движений кончика ручки, к примеру по поверхности специального компьютерного экрана.

Оффлайновый вид распознавания успешно применяется в сферах деятельности, где необходимо обрабатывать большое количество рукописных документов, к примеру, в страховых компаниях. Качество распознавания можно повысить, используя структурированные документы (формы). Кроме того, можно улучшить качество, уменьшив диапазон возможных вводимых символов. Оффлайновое распознавание считается более сложным по сравнению с онлайн-овым.

Точное распознавание латинских символов в печатном тексте в настоящее время возможно только если доступны чёткие изображения, такие как сканированные печатные документы. Точность при такой постановке задачи превышает 99%, абсолютная точность может быть достигнута только путем последующего редактирования человеком. Проблемы распознавания рукописного «печатного» и стандартного рукописного текста, а также печатных текстов других форматов (особенно с очень большим числом символов) в настоящее время являются предметом активных исследований.

Анализ подходов к проектированию систем оптического чтения текстов

Широко исследуемой проблемой является распознавание рукописного текста. На данный момент достигнутая точность даже ниже, чем для рукописного «печатного» текста. Более высокие показатели могут быть достигнуты только с использованием контекстной и грамматической информации. Например, в процессе распознавания искать целые слова в словаре легче, чем пытаться проанализировать отдельные символы из текста. Знание грамматики языка может также помочь определить, является ли слово глаголом или существительным. Формы отдельных рукописных символов иногда могут не содержать достаточно информации, чтобы точно (более 98 %) распознать весь рукописный текст.

Методы автоматического распознавания образов и их реализация в системах оптического чтения текстов (OCR-системах - Optical Character Recognition) - одна из самых плодотворных технологий ИИ.

В приведенной трактовке OCR понимается как автоматическое распознавание с помощью специальных программ изображений символов печатного или рукописного текста

(например, введенного в компьютер с помощью сканера) и преобразование его в формат, пригодный для обработки текстовыми процессорами, редакторами текстов и т. д.

Сокращение OCR иногда расшифровывают как Optical Character Reader. В этом случае под OCR понимают устройство оптического распознавания символов или автоматического чтения текста. В настоящее время такие устройства при промышленном использовании обрабатывают до 100 тыс. документов в сутки. Промышленное использование предполагает ввод документов хорошего и среднего качества. Это соответствует задачам обработки бланков переписи населения, налоговых деклараций и т. п.

Отметим следующие особенности предметной области, существенные с точки зрения OCR-систем:

- шрифтовое и размерное разнообразие символов;
- искажения в изображениях символов (разрывы образов символов, например, при увеличении изображения; слипание соседних символов и др.);
- перекосы при сканировании;
- посторонние включения в изображениях;
- сочетание фрагментов текста на разных языках;
- большое разнообразие классов символов, которые могут быть распознаны только при наличии дополнительной контекстной информации.

Автоматическое чтение печатных и рукописных текстов является частным случаем автоматического визуального восприятия сложных изображений. Многочисленные исследования показали, что для полного решения этой задачи необходимо интеллектуальное распознавание, т. е. «распознавание с пониманием». Однако в настоящее время в технически реализуемых OCR-системах рассматриваемая проблема значительно упрощена и сведена к задаче классификации по признакам простых объектов. Эта задача описывается хорошо разработанным математическим аппаратом пороговых отделителей - разделяющими плоскостями.

В лучших OCR-системах используется технология распознавания, свойственная человеку. У человека распознавание образа является многоступенчатым.

Методы распознавания

Система распознавания реализуется как классификатор. Существуют несколько методов реализации классификаторов:

- шаблонные (растровые);
- признаковые;
- структурные.

В классификаторе первого типа с помощью критерия сравнения определяется, какой из шаблонов выбрать из базы. Самый простой критерий - минимум точек, отличающих шаблон от исследуемого изображения.

Достоинства: К достоинствам шаблонного классификатора относятся хорошее распознавание дефектных символов («разорванных» или «склеенных»), простота и высокая скорость распознавания.

Недостатки: Недостатком является необходимость настройки системы на типы и размеры шрифтов.

Наиболее распространены признаковые классификаторы. Анализ в них проводится только по набору чисел или признаков, вычисляемых по изображению. Таким образом, происходит распознавание не самого символа, а набора его признаков, т. е. производных данных от исследуемого символа.

Достоинства: Этот метод позволяет распознавать различные начертания символов, т.е. различные подчерки шрифты и т.д.

Недостатки: Этот метод неизбежно вызывает некоторую потерю информации.

Структурные классификаторы переводят шейп символа в его топологическое представление, отражающее информацию о взаимном расположении структурных элементов символа. Эти данные могут быть представлены в графовой форме.

Достоинства: Такой способ обеспечивает инвариантность относительно типов и размеров шрифтов.

Недостатки: Недостатками являются трудность распознавания дефектных символов и медленная работа.

Структурно-пятенный метод

Однако наиболее оптимальным среди всех методов является так называемый структурно-пятенный эталон и его фонтанное (от англ. font - шрифт) представление. Оно имеет вид набора пятен с попарными отношениями между ними. Подобную структуру можно сравнить со множеством шаров, нанизанных на резиновые шнуры, которые можно растягивать. Данное представление нечувствительно к различным начертаниям и дефектам символов.

Алгоритм основан на сочетании шаблонного и структурного методов распознавания образов. При анализе образца выделяются ключевые точки объекта — так называемые «пятна».

В качестве пятен, например, могут выступать:

- концы линий;
- узлы, где сходятся несколько линий;
- места изломов линий;
- места пересечения линий;
- крайние точки.

После выделения «пятен» определяются связи между ними — отрезок, дуга. Таким образом, итоговое описание представляет собой граф, который и служит объектом поиска в библиотеке «структурно-пятенных эталонов».

При поиске устанавливается соответствие между ключевыми точками образца и эталона, после чего определяется степень деформации связей, необходимая чтобы привести искомым объект к сравниваемому эталонному образцу. Меньшая степень необходимой деформации предполагает большую вероятность правильного распознавания символа.

Сравнение эталонов

Рассмотрим строчный образ (слово). В данном случае можно выделить два критерия, на основе которых можно строить меру близости:

- совпадение букв,
- монотонность (совпадение порядка букв).

Пусть r_1, r_2, \dots, r_I — эталон, t_1, t_2, \dots, t_J — пробный образ, причем $I \neq J$.

Построим соответствие между эталоном и пробным образом по следующему правилу: каждому символу в первом слове должен соответствовать хотя бы один символ во втором слове и каждому символу во втором слове должен соответствовать хотя бы один символ в первом слове, (но соответствие между символами не взаимнооднозначное, в частности, поскольку $I \neq J$). Введем меру следующим образом:

$$\rho(r_i, t_j) = \begin{cases} 1, & r_i \neq t_j \\ 0, & r_i = t_j \end{cases}$$

В качестве меры сходства двух слов принимаем соответствие, при котором суммарный вес всех дуг (изображенных на рисунках) минимален:

$$v(\bar{r}, \bar{t}) = \min_S \mu(S), \text{ где } \mu(S) = \sum_{(i,j) \in S} \rho(r_i, t_j)$$

Через $v(\bar{r}, \bar{F})$ далее будем обозначать меру близости двух \bar{r} и \bar{F} .

Соответствие S должно быть двудольным графом без изолированных вершин с непересекающимися ребрами. Рассмотрим задачу сравнения цепочек упорядоченных символов. В данной задаче могут возникать следующие ошибки:

- неправильно определенный символ (кошка – корка),
- ошибка вставки (кошка – кошрка),
- ошибка потери (кошка – кшка).

Редакторским расстоянием называется минимальное общее число изменений, вставок и потерь, требуемое для изменения образа A в образ B :

$$D(A, B) = \min_j [C(j) + I(j) + R(j)],$$

где минимизация происходит по всем возможным комбинациям символьных преобразований таких, чтобы получить B из A .

Пусть

$$d(i, j | i-1, j-1) = \begin{cases} 1, & \text{при } x(i) \neq y(i) \\ 0, & \text{при } x(i) = y(i) \end{cases}$$

Тогда

$$d(i, j | i-1, j) = d(i, j | i, j-1) = 1$$

Данные можно представить в виде таблицы, в которой столбцы – это символы образа, строки – символы эталона. Количество точек в матрице есть $I \times J$.

Таблица 1. Таблица данных слова.

	t_1	t_2	t_3	...	t_i
r_1	○	○	○	...	○
r_2	○	○	○	...	○
r_3	○	○	○	...	○
...
r_i	○	○	○	...	○

По данной таблице можно построить граф по следующему правилу. Если отображается точка (r_2, t_1) , то далее выбираем (r_2, t_2) , (r_1, t_2) или (r_3, t_1) (т.е. возможны три варианта). Соответствие слов реализуется в виде маршрута в графе. Этот маршрут обязательно начинается с точки (r_1, t_1) (иначе появится изолированная точка) и заканчивается в (r_i, t_j)

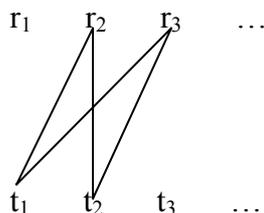


Рис. 1. Граф.

Таким образом, получили задачу выбора кратчайшего пути на графе от точки (r_1, t_1) до точки (r_i, t_j) где каждая вершина имеет свою стоимость: 0 или 1.

Заключение

Автоматическое зрительное восприятие на сегодняшний день не достигает совершенства человеческого восприятия текста. Главная причина этого заключается в неумении строить достаточно полные и семантически выразительные компьютерные модели предметной области.

Проанализировав существующие методы распознавания текстов, можно сделать вывод, что лучше всего использовать метод структурно-пятенного шаблона, так как он объединяет в себе достоинства многих методов и благодаря этому является достаточно гибким чтобы применить его при распознавании рукописного текста.

Список литературы

1. *Абраменко А.* Принципы распознавания//Компьютер-пресс. 1997.—№ 12.
2. *Загоруйко Н.Г.* Методы распознавания и их применение. — М.: Сов. радио,1972. —208 с.
3. *Линдсней П., Норман Д.* Переработка информации у человека: Пер. с англ. — М.:Мир, 1974. —550 с.
4. *Шамис А.Л.* Принципы интеллектуализации автоматического распознавания
5. изображений и их реализация в системах оптического распознавания символов //Новости искусственного интеллекта. 2000.
6. Cognitive Forms — система массового ввода структурированных документов [Электронный ресурс] / В.В. Арлазаров, В.В. Постников, Д.Л. Шоломов.
7. Уинстон П. Искусственный интеллект / П. Уинстон. — М.: Мир,1980. — 519 с.
8. Хант Э. Искусственный интеллект / Э. Хант. — М.: Мир, 1978. —558 с.
9. Логический подход к искусственному интеллекту / под ред. Г.П. Гаврилова — М.: Мир, 1998. — 493 с.
10. Эндрю А. Искусственный интеллект / Под ред. Поспелова Д.А. —М.: Мир, 1985. — 265 с.
11. Дарья Калинкина, Дмитрий Ватолин Проблема подавления шума на изображениях и видео и различные подходы к ее решению // *Компьютерная графика и мультимедиа* : журнал. — 2005. — № 3(2).
12. Распознавание рукописных текстов, Ижевск 2006, А. В. Кучуганов, Г. В. Лапинская
13. ЭКСПЕРТНЫЕ СИСТЕМЫ, Учебное пособие для вузов, Воронеж 2007, А. Ю. Телков
14. G.A.Carpenter and S. Grossberg, Pattern Recognition by SelfOrganizing Neural Networks, MIT Press, Cambridge, Mass., 1991.
15. "The First Census Optical Character Recognition System Conference", R.A.Wilkinson et al., eds., . Tech. Report, NISTIR 4912, US Deop. Commerse, NIST, Gaithersburg, Md., 1992.
16. J. Hertz, A. Krogh, and R.G. Palmer, Introduction to the Theory of Neural Computation, Addison-Wesley, Reading, Mass., 1991.
17. S. Haykin, Neural Networks: A Comprehensive Foundation, MacMillan College Publishing Co., New York, 1994.
18. G.A.Carpenter and S. Grossberg, Pattern Recognition by SelfOrganizing Neural Networks, MIT Press, Cambridge, Mass., 1991.
19. "The First Census Optical Character Recognition System Conference", R.A.Wilkinson et al., eds., . Tech. Report, NISTIR 4912, US Deop. Commerse, NIST, Gaithersburg, Md., 1992.