

ОПТИМАЛЬНЫЙ АЛГОРИТМ РАСПОЗНАВАНИЯ ТЕКСТОВ

S. B. Mamytov, L.M. Musabekova

OPTIMAL ALGORITHMS FOR THE RECOGNITION OF THE TEXT

УДК: 621:550.3/21

В настоящее время существует ряд методов распознавания речи, текста. Они отличаются друг друга особенностями. Текущем работе изложен один из этих современных методов распознавания текстов, виды и направления распознавания.

Currently, there are several methods speech recognition, text. They differ from each other features. This paper, we report one of these modern methods of image recognition text, type and direction of recognizers.

Понятие клеточного автомата с метками введено автором данной работы. Причиной его создания являлась необходимость достижения заданных характеристик клеточных автоматов при разработке алгоритмов процесса распознавания текста. Как будет показано далее, это понятие позволяет эффективно реализовать алгоритмы выделения признаков символов.

В процессе работы над диссертацией - анализа литературы - обнаружено понятие памяти в теории тьюринговых машин. Оно определяет, что головка тьюринговой машины может хранить некоторое конечное число данных, которые используются правилами автомата машины для изменения ленты машины или передвижения ее головки. Данное понятие при введении некоторых ограничений и перехода на теорию клеточных автоматов может быть также сведено к клеточным автоматам с метками.

Принцип клеточной системы с метками заключается в ассоциации каждой клетки поля с одним или несколькими метками.

Примером клеточного автомата с метками может служить набор: поле, на котором определены только клетки черного и белого цветов, метка, которая может присутствовать или отсутствовать в клетках поля, и правила, учитывающие наличие меток у соседних клеток. В случае определения одной метки для клеточного автомата, количество состояний каждой клетки увеличивается в два раза относительно случая без меток. В приведенном примере количество состояний каждой клетки равно четырем.

Правило клеточного автомата в такой системе будут выглядеть так, как представлено на рис 1.

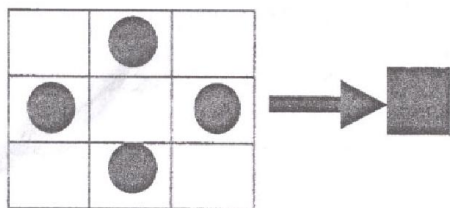


Рис. 1. Пример правила клеточного автомата с метками

В данном примере используется правило, основанное на восьми соседях. Без меток количество вариантов подобных правил (в случае определения только двух цветов клеток) будет $2^{10} = 1024$ (девять клеток условий и одна - результата). Добавляя метку, получится $(2 \cdot 2)^{10} = 1\ 024 \cdot 2 = 2\ 048$ вариантов правил. Удобнее, в данном случае, разделять двоеточием нумерацию цветов и меток. Тогда для правила, представленного на рисунке 10, номером может служить 1:340 (или в двоичном виде 0000000001:0101010100, где каждая цифра определяет состояние клетки, клетки нумеруются слева направо сверху вниз), где первое число определяет правило преобразования цвета, а второе число - правило изменения метки.

Правило для поля, на котором определено более одной метки, нумеруется аналогичным образом: для правила с двумя метками - x: y: z, где первое число определяет правило преобразования цвета, второе число - преобразование одной метки, третье число - второй метки.

Клеточный автомат с метками позволяет получать сложные результаты без изменения структуры используемой системы клеток. Например, для заданного изображения, данный автомат позволяет определить некоторые свойства данного изображения и выделить его характерные признаки без изменения цвета.

Клеточный автомат с метками - это набор $\{G, M, Z, N, f\}$. Описание каждого элемента представлено ниже.

1. G - Конечное дискретное метрическое множество, гарантирующее конечность расстояний между клетками.
2. M - конечное множество меток, определенное для каждой клетки.
3. Z - Конечный набор состояний клеток.
4. N - конечное множество, определяющее окрестность клетки таким образом, что каждый элемент множества позволяет определить соседа для каждой клетки. $|N|$ - количество соседних клеток, которые влияют на состояние данной.
5. f - Правила клеточного автомата, соответствующие математической функции переходов:

$$Z \times C \times z^M \times C^{|N|} \rightarrow Z \times C, \text{ где } C \in M \cup \bar{M}.$$

Ниже показано, что клеточный автомат с метками удовлетворяет требованиям классического клеточного автомата.

По определению клеточный автомат задается четырьмя элементами $\{G^*, Z^*, N^*, f^*\}$. Элементы G и N клеточного автомата с метками соответствуют классическим элементам G^* и N^* .

Введем $Z^\wedge = Z \times C$ таким образом, что $|Z^\wedge| = |Z| \times |C|$. Множество Z^\wedge будет соответствовать Z^* классического определения клеточного автомата, так как оно является конечным и определяет все варианты состояний клетки автомата.

Аналогично, $f^\wedge = Z^\wedge \times Z^\wedge^{|N|} \rightarrow Z^\wedge$ будет соответствовать f^* , так как учитываются все состояния на первом и втором временном слое.

Таким образом, получен набор $\{G, Z^\wedge, N, f^\wedge\}$ клеточного автомата, в котором выполняются все свойства: локальность правил (обеспеченное N), однородность системы на основе метрики G , конечность множества состояний клетки Z^\wedge , одновременность изменений для

всех клеток, гарантируемое набором правил f^\wedge .

На основе моделирующей программы разработан пример клеточного автомата с метками «шахматная доска».

Задача автомата «шахматная доска» заключается в отрисовке чередующихся черно-белых точек на белом изображении.

Для выполнения задачи клеточному автомату с метками «шахматная доска» достаточно восьми правил: 0:257, 0:65, 0:321, 1:52, 513:52, 256:52, 64:52, 320:52 (нерассмотренные состояния девяти исходных клеток не изменяют состояние результирующей клетки). Эти правила изображены на рис. 1.

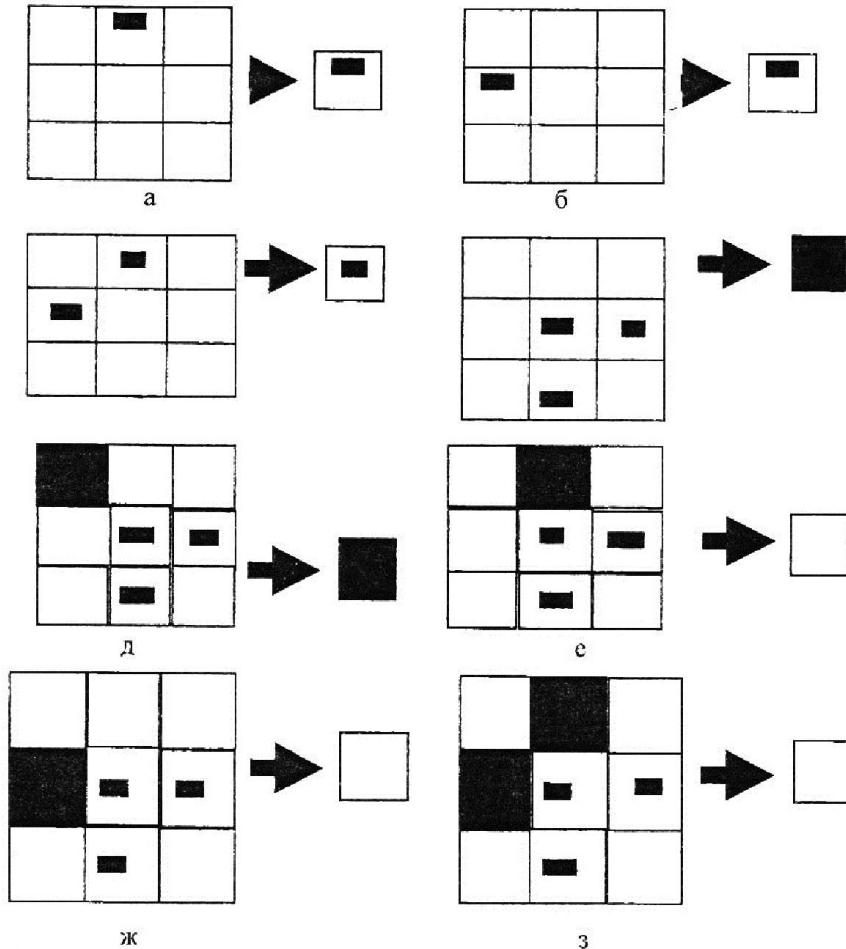


Рис. 11. Правила клеточного автомата «шахматная доска», правила:

а – 0:257, б – 0:65, в – 0:321, г – 1:52, д – 513:52, е – 256:52, ж – 64:52, з – 320:52

Для начала работы автомата необходимо поставить метку на стартовой точке поля. Результат работы автомата на шагах 1, 3, 5, 10, 20 изображены на рис. 2.

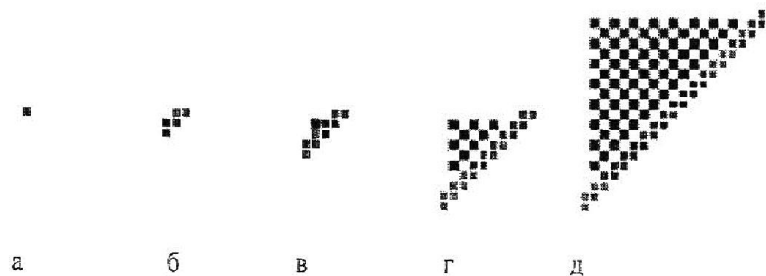


Рис. 2. Результат работы клеточного автомата с метками «шахматная доска» на шагах: а – первом, б – третьем, в – пятом, г – десятом, д – двадцатом

Данный пример иллюстрирует применение клеточных автоматов с метками, но не затрагивает практическую полезность, которую можно обеспечить с их использованием: возможность определения характеристик, не изменяя цвета обрабатываемого изображения.

Литература:

1. Р. Дуда, П. Харт. Распознавание образов и анализ сцен,- Москва: Мир, 1976. -407с.
2. Вапник В.Н., Червоненкис А.Я. Теория распознавания образов,-Наука, 1974. -417с
3. В. Т. Фисенко, Т. Ю. Фисенко. Компьютерная обработка и распознавание изображений (учебное пособие). - Санкт-Петербург, 2008. - 102с.
4. Островский С. Нейронные сети для обработки информации. -Москва, 2000. -220с.
5. Верхаген К. и др. Распознавание образов. -Радиосвязь, 1985.
6. Дж. Ту, Р. Гонсалес. Принципы распознавания образов- Москва: Мир, 1978.

Рецензент: д.т.н., профессор Ескендилов Ш.З.
