

Структурное распознавание образов на примере задачи о распознавании структуры рукописной таблицы

Ганиев Ибраим Шукриевич

КРЫМСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ ИМЕНИ В.И. ВЕРНАДСКОГО

ТАВРИЧЕСКАЯ АКАДЕМИЯ

ФАКУЛЬТЕТ МАТЕМАТИКИ И ИНФОРМАТИКИ

КАФЕДРА ИНФОРМАТИКИ (ГРУППА 601)

e-mail: olologin@gmail.com

В статье объясняются общие принципы построения систем структурного распознавания образов на примере задачи распознавания структуры рукописной таблицы

ВВЕДЕНИЕ

Существует целый класс задач распознавания образов не поддающийся решению при помощи обычного статистического подхода, стандартных нейросетей или SVM. Примерами таких задач могут послужить распознавание химической структуры, математической формулы, таблицы. Это лишь некоторые задачи структурного распознавания, но уже можно заметить: образ состоит из иерархии более простых образов, и сама иерархия несёт в себе много информации об объекте. В случае с математической или химической формулой важно не только распознать символ, но и верно распознать структуру, где этот символ находится, является ли он степенью, знаменателем или числителем дроби и т. д. В подобных задачах отсутствие информации о структуре делает все остальные данные бесполезными. Для подобных задач выделяют отдельный подход в распознавании — *структурное распознавание образов*. Вопрос: чем класс задач структурного распознавания отличается от задач обычного распознавания образов?

- (1) Каждый образ представляет собой иерархию взаимоотношений между более простыми образами (непроизводными элементами), при этом эта иерархия важна для получения результата, т.е. важно положение более простых образов в структуре образа.
- (2) Множество всевозможных классов образов в задаче настолько велико что сложно получить даже репрезентативную выборку из этого множества.

ПОСТАНОВКА ЗАДАЧИ РАСПОЗНАВАНИЯ СТРУКТУРЫ РУКОПИСНОЙ ТАБЛИЦЫ

Сформулируем *задачу распознавания структуры рукописной таблицы*.
Дано: цветное растровое изображение произвольной нарисованной от руки таблицы, которая может содержать объединённые ячейки, неровные, прерывистые, косые линии. *Требуется* — представить таблицу в виде описания на формальном языке, к примеру на языке html или L^AT_EX.

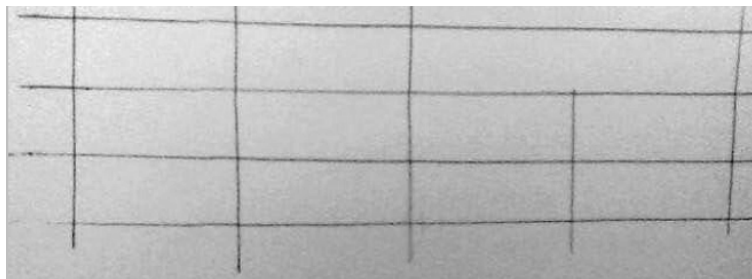


Рис. 1. Пример изображения на входе

Данная задача является задачей структурного распознавания. Рассмотрим некоторые определения данного подхода.

Определение 1. *Подклассом образов понимается некоторая категория, определяемая рядом свойств, общих для всех её элементов [1].*

Определение 2. *Непроизводный элемент — подобраз являющийся составным элементом образа, структурой в непроизводном элементе мы можем пренебречь [1].*

Определение 3. *Операции композиции — всевозможные взаимоотношения между непроизводными элементами, для химической формулы это может*

быть наличие химической связи между группами производных элементов, для математического выражения это могут быть различные операции между производными элементами, например взаимная ориентация производных элементов на плоскости [2].

Определение 4. *Язык описания образа — язык обеспечивающий структурное описание образа в терминах множества производных элементов и операций композиции этих элементов. Как правило является грамматикой, Но в некоторых случаях вместо грамматики в данном блоке могут использоваться графы зависимостей [2].*

В нашем случае производными элементами будут являться разные типы пересечения пар отрезков (Т-образные, Г-образные, Х-образные), операция композиции — наличие связи между соседними производными элементами, в зависимости от их ориентаций и типов пересечения.

ОБЩЕЕ УСТРОЙСТВО СИСТЕМЫ СТРУКТУРНОГО РАСПОЗНАВАНИЯ ОБРАЗОВ

Любую систему структурного распознавания образов можно разделить на три основных стадии следующие друг за другом, а именно: стадия предобработки, стадия представления объекта, стадия структурного анализа [1].

Предобработка осуществляет предобработку образа объекта, сюда относятся кодирование и аппроксимация, функции фильтрации и удаления шумов. В случае нашей задачи этот блок занимается переводом цветного изображения в градации серого [3] и последующей адаптивной бинаризацией [3].

Представление объекта занимается сегментацией объекта, поиском производных элементов, обнаружением структуры путём композиции производных элементов, их классификацией в зависимости от образа производного элемента и его положения в структуре в соответствии с заданной грамматикой. В нашей задаче этот блок будет заниматься поиском множества различных отрезков с помощью преобразования Хаффа [3], нахождением пересечений между ними, а также созданием структуры объединения соседние производимые элементы последовательно, учитывая их ориентацию и тип. Таким образом мы получим планарный граф вершинами которого будут точки пересечения, а наличие отрезка между соседними на фото пересечениями будет означать наличие ребра между соответствующими вершинами графа.

Структурный анализ объекта занимается проверкой производных элементов и операций композиции между этими элементами, согласно ранее

заданному языку описания образа. Также этот блок занимается построением ответа системы. Т.е. в случае нашей задачи он будет заниматься обходом построенного ранее планарного графа связей между производными элементами, учитывая горизонтальным или вертикальным является определённый отрезок соединяющий производные элементы на изображении, а также тип производного элемента. Это позволит нам получить с каждой строки и столбца, объединив их в таблицу мы можем описать её на любом формальном языке.

Вывод

Рассмотренный обобщённый подход предлагает вариант декомпозиции задачи структурного распознавания на структуру и неприводимые объекты, в результате чего ранее нерешаемые статистическими подходами задачи могут быть сведены к простым разделённым задачам статистической классификации и поиску структуры. С помощью этого подхода удаётся решить задачу распознавания структуры рукописной таблицы. Подобные задачи сегодня не редкость, и помимо обычного распознавания достаточно абстрактных понятий как то математические выражения, химические формулы и таблицы, подобными способами строят современные системы позволяющие создавать вполне корректные описания объектов на сцене (фото) и связей между ними (Как, например, «Человек сидит в кресле и держит стакан») [4].

СПИСОК ЛИТЕРАТУРЫ

- [1] Фу К. *Структурные методы в распознавании образов*. – М.: Мир, 1977. –1152 с.
- [2] Новикова Н. М. *Структурное распознавание образов. Учебно-методическое пособие для вузов* — Издательско-полиграфический центр Воронежского государственного университета, 2008.
- [3] Р. Гонсалес, Р. Вудс. *Цифровая обработка изображений* — М.: Техносфера, 2005. – 1072 с.
- [4] *CaptionBot*. Microsoft. URL: <https://www.captionbot.ai/>