

УДК 004.8:004.93

В.В. Сторож

Донецкий государственный институт искусственного интеллекта, Украина

Ограничения классических подходов к распознаванию образов

В статье анализируются наиболее общие и принципиальные проблемы распознавания образов. Подчеркивается общность большинства из этих проблем с проблемами кибернетики и искусственного интеллекта. Их происхождение связано с попытками описать очень сложный объект с помощью слишком упрощенного и неадекватного понятийного аппарата. Проанализированы с позиций распознавания образов основные компоненты этого понятийного аппарата: теория информации, классическая схема распознавания, учет априорной информации, математика и вычислительные методы и ряд других составляющих. Показана недостаточность указанных компонентов для реализации распознавания образов в большинстве практически интересных случаев. Отмечены, а в ряде случаев и обоснованы меры по преодолению ограничений, присущих указанным компонентам.

Введение

Проблема распознавания образов является одной из центральных в концепции искусственного интеллекта. В соответствии с важностью проблемы значительные усилия научного сообщества на протяжении примерно полувека затрачены на ее решение. Получены определенные результаты, особенно в прикладных областях и с использованием алгоритмов, ориентированных на решение сравнительно узкого круга задач. В то же время эти результаты значительно скромнее ожидаемых, особенно заметно отсутствие прогресса на концептуальном, общеметодологическом уровне. Застой приобрел столь затяжной характер, что отсутствие прогресса в построении общей теории распознавания образов отмечается в работах, разнесенных во времени на 26 (!) лет [1, с. 10; 2, с. 88]. Для сравнения – основы квантовой механики (включая коренное изменение взглядов на природу) были созданы всего за шесть лет (1925 – 1930 гг.). Возникает естественный вопрос: в чем причина застоя? Не те исследователи? Не там ищут? Или предмет сложнее квантовой механики?

Прежде чем отвечать на этот вопрос, заметим, что в настоящее время кризисное состояние наблюдается не только в распознавании образов и искусственном интеллекте, а также и в родительской по отношению к искусственному интеллекту науке – кибернетике. Причем кризис в кибернетике настолько серьезен, что на Западе это понятие исчезло из названий курсов в учебных заведениях, а в отечественной печати идет дискуссия, есть ли такая наука [3]?

Наличие затяжного кризиса в ряде родственных наук и научных направлений наталкивает на мысль, что этот кризис имеет общее происхождение. В таком случае, причины кризиса не в отсутствии каких-то удачных разработок или алгоритмов в рамках конкретного научного направления, а гораздо глубже, на общефилософском и методологическом уровнях.

Настоящая работа посвящена анализу причин кризиса в распознавании образов и возможных путей его преодоления. Анализ по необходимости проводится с привлечением результатов других наук и научных направлений: философии, математики, нейрофизиологии, синергетики и т.д.

«Классические» подходы к распознаванию образов

Под классическими в данной статье понимаются подходы или методы распознавания образов, удовлетворяющие двум условиям:

- эти методы должны использоваться в «классической» схеме распознавания образов;
- это наиболее распространенные на протяжении последних десятилетий методы распознавания образов.

Под «классической» понимается задача распознавания образов в следующей постановке (см. предисловие Ю.И. Журавлева к работе [4]): «объекты распознавания заданы набором признаков, известно некоторое число эталонных объектов, и их описания составляют исходную (обучающую) информацию. На основе этой информации синтезируется алгоритм, определяющий для вновь поступивших объектов, к какому (или к каким) из конечного числа классов они принадлежат». Эта же схема распознавания приводится и в справочнике по искусственному интеллекту [5, с. 150].

Чтобы выделить основные «классические» подходы к распознаванию образов, рассмотрим вначале кратко их перечень, или классификацию. Существует несколько различающихся классификаций этих методов. Так, в [2] выделяются нейросети, геометрический подход, статистические и лингвистические методы. В [5, с. 153-154] выделяются методы разделяющих поверхностей, статистические, методы исчисления высказываний, алгебраические методы, методы вычисления оценок (голосования) и др. В [6] выделяются методы сравнения с эталоном, статистические методы, синтаксические или структурные и неросетевые методы.

Наибольшее распространение из перечисленных получили статистические методы. Это положение обосновано в работе [6] путем подсчета количества статей, использующих тот или иной метод распознавания образов. На примере журналов IEEE Transactions on Pattern Analysis and Machine Intelligence за 1979 – 1999 гг. видно, что примерно 85 % статей посвящены рассмотрению того или иного статистического метода. Геометрический подход или метод разделяющих поверхностей тесно примыкает к статистическим методам, так что его, как правило, относят к разновидности статистических методов [1], [6]. Широкое распространение в распознавании образов получили также алгебраические, синтаксические, структурные или структурно-лингвистические методы и ряд других родственных им методов, но важно подчеркнуть, что они имеют много общего со статистическими методами, поскольку используют, во-первых, одну и ту же классическую схему распознавания, а во-вторых, «классическую» дискретную математику.

В дальнейшем под классическими методами распознавания образов мы будем понимать перечисленные в предыдущем абзаце методы.

Несколько слов о других методах, перечисленных в классификационных списках, но не внесенных в данной работе в классические.

Методы вычисления оценок (голосования, согласования оценок, коллектив решающих правил и др.) очень интересны, но они преимущественно используются не в качестве самостоятельных методов, а как дополнение к уже существующим [7, с. 85-153]. Помимо этого, представляется, что методы согласования оценок содержат в себе очень большой (и еще не до конца использованный) потенциал, и к ним неприменима часть последующих рассуждений об ограничениях классических методов.

Нейросетевые методы представляют собой, скорее, инструмент, чем некоторую единую идеологию. Во всяком случае, в их работе может быть реализована самая различная идеология, например, какая-либо разновидность статистического метода, имеющая все ограничения классических методов, или некоторый механизм, имитирующий работу мозга в процессе распознавания и не имеющий ограничений классических методов. Таким образом, в нейросетях возможна реализация как классических, так и неклассических механизмов распознавания, при этом сами классические механизмы относятся к одному из перечисленных выше классов. В связи с этим в данной статье ограничения нейросетевых методов не рассматриваются.

В заключение данного раздела несколько слов о соотношении термина «распознавание образов» с родственными ему терминами «распознавание изображений» и «Data Mining» (извлечение знаний). Распознавание изображений является частью распознавания образов (в общем случае с более простыми и ясными алгоритмами), и все далее полученные в работе выводы относительно распознавания образов в равной мере применимы и к распознаванию изображений.

Относительно Data Mining. Ряд специалистов в области информационных технологий склонны видеть в Data Mining продолжение идеологии, используемой в распознавании образов, и на этом основании отождествляют эти термины. Но, на наш взгляд, отождествление этих понятий неправомерно, во всяком случае, с позиций рассмотрения ограничений этих идеологий. Можно привести следующие доводы в пользу этой точки зрения.

Во-первых, Data Mining включает в себя гораздо более широкий набор методов, чем распознавание образов. Так, в [8] помимо статистических методов к Data Mining относят эволюционное программирование, генетические алгоритмы, алгоритмы ограниченного перебора, метод группового учета аргументов и ряд других новейших методов переработки информации. Важно также, что в Data Mining набор этих методов не «канонизирован», как в распознавании образов, а может быть без проблем расширен в любой момент.

Во-вторых, существенно различаются и задачи рассматриваемых идеологий. Так, если в распознавании образов основной упор делается на выбор одного из разработанных ранее вариантов (образа, решения и т.д.), то в Data Mining основной целью является нахождение новых, неизвестных ранее закономерностей.

В связи с вышесказанным, рассмотренные далее ограничения в распознавании образов в общем случае не применимы к идеологии Data Mining (хотя в некоторых частных случаях это может быть и не так).

Ограничения классических методов

Перед рассмотрением недостатков и ограничений традиционных методов распознавания необходимо отметить практически полное отсутствие работ общего характера на эту тему. Имеется множество публикаций, посвященных развитию конкретных методов, сотни школ и групп, которые ссылаются лишь на свои собственные работы и, по-видимому, плохо представляют, что делается в других группах и школах. Очень мало обзорных работ. Да и в философии наиболее модными темами являются квантовая физика, Вселенная и т.п., но нам неизвестно ни одной работы, в которой на общем, близком к философскому или общеметодологическому уровню рассматривались бы вопросы распознавания образов (исключение составляет разве что работа [2], в значительной мере посвященная рассмотрению вопросов распознавания на общеметодологическом уровне, но и в ней ограниченность классических методов распознавания декларируется, но не исследуется). Таким образом, первая крупная проблема в области распознавания образов относится, скорее, не собственно к теории (и практике) распознавания, а к организации исследований и заключается в дроблении на множество слабосвязанных направлений и отсутствии специалистов, занимающихся проблемами распознавания на общем уровне, позволяющем связать работы отдельных направлений в единую целостную картину.

В связи с вышесказанным перечень ограничений классических методов невозможно было получить на основании анализа литературных данных, и он, безусловно, носит следы субъективизма и личных предпочтений автора. Поэтому этот перечень нужно рассматривать как предварительный, как предмет для диалога и дискуссий.

Наряду с рассмотрением собственно ограничений в работе приведены и соображения по их преодолению. Естественно, они не могут иметь вид законченных теорий и представляют собой, скорее, примеры, почерпнутые из доступных литературных данных, показывающие направление и возможность развития в данном направлении, но ни в коей мере не исчерпывающие направления целиком.

Как представляется, основные ограничения классических методов распознавания образов можно выразить следующими положениями:

Неполнота классической схемы распознавания образов, отсутствие в ней механизмов развития и встраивания ее в схему познания.

Недостаточность и противоречивость существующих подходов к содержанию понятия информации, отсутствие в них возможности информационного представления передачи информации между уровнями в процессе распознавания.

Неполнота, неточность и противоречивость понятия «образ» и составляющих его компонентов (признак и др.). Неадекватность классической схемы при использовании понятий «образ» и «класс».

Недостаточное использование априорной информации.

Недостаточность дискретной математики и общепринятой парадигмы вычислений.

Недостаточный учет многоуровневости процесса обработки информации.

Невозможность создания полноценной теории распознавания в рамках одного, пусть и самого перспективного научного направления.

1. Неполнота классической схемы распознавания образов

Классическая схема распознавания, использующая понятия признаков, их комбинирования и последующей классификации, использует несколько неявных, нигде не декларируемых допущений, сводящихся к следующему:

- в каждый данный момент воспринимается лишь один образ, при этом уровень информации, связанной с фоном, либо незначителен, либо мы имеем простые инструменты для отделения той части информации, которая связана с фоном, от информации о признаках;
- задача субъекта восприятия (человека или робота) сводится к отнесению объекта к одному из классов и ни к чему более.

В действительности ни одно из этих допущений не выполняется. Мы (или робот) воспринимаем окружающий мир, содержащий множество объектов. В зависимости от задач одни из этих объектов мы выделяем, другие воспринимаем как фон, и задача распознавания в этом случае значительно сложнее классической схемы: а) мы должны выделить из шумов информацию о признаках множества распознаваемых объектов, при этом заранее не очевидно, что в сигнале считать шумом, а что полезной информацией; т.е. какие признаки выделять из шума. Очевидно, эта задача имеет множество возможных решений и мы должны выбрать нужное; б) мы должны каким-либо образом отнести признаки к какому-то объекту, причем и без численных оценок очевидно, что комбинаторным путем эту задачу не решить, и здесь необходимо использование некоторых принципов, резко ограничивающих число возможных вариантов; в) мы должны далее эти объекты распознать, выделить из них интересующие нас и соотнести их с нашими проблемами.

По поводу второго допущения: распознавание образов не является независимой процедурой, оно встроено в схему восприятия и мышления. И в этой схеме распознавание составляет только часть задачи. Вторая часть задачи – на фоне некоторой относительно инвариантной части выделить особенности образа или ситуации, характеризующие его здесь и сейчас, и использовать эти конкретные особенности в мышлении. Например, когда мы встречаем старого знакомого, то мы не просто относим его к одному из классов, а определяем его настроение, узнаем о проблемах и т.д. И в конце встречи мы не относим полученный нами новый образ к какому-то существующему классу – это невозможно из-за чисто комбинаторных соображений, а конструируем новый образ, состоящий из двух частей: некоторого инварианта I и уникальной переменной части Δ , которую мы составляем из элементов наших впечатлений и языка (по сути, понятий, которые тоже можно представить как некоторый класс образов). Таким образом, результат восприятия V по отношению к отдельному выделенному нами образу более правильно выразить формулой:

$$V = I + \Delta.$$

Подчеркнем принципиальную важность представления образов в указанной форме. Работа мышления и интеллекта в значительной мере определяется именно уникальной, а не инвариантной частью воспринимаемых образов, поэтому, не учитывая ее, мы, по сути, выбрасываем наиболее важные

слагаемые мышления и интеллекта. Становится очевидным, что только по одной этой причине (неучете уникальной составляющей образа) классическая схема распознавания не может быть использована для моделирования работы интеллекта.

Неучет уникальной составляющей образа приводит к так называемому алгоритмическому или автоматному подходу, который эквивалентен работе нервной системы на уровне насекомого. Представляется, что большинство специалистов, занимающихся распознаванием образов согласятся, что разработанные в настоящее время искусственные распознающие системы в лучшем случае достигают уровня насекомого.

Заметим, что и кибернетика как наука, и распознавание образов не смогли преодолеть автоматный рубеж, в результате чего вместо кибернетики появились информационные технологии, Data Mining и ряд других новых дисциплин, а наука о распознавании образов переживает кризис. (К большому сожалению, эти новые дисциплины не смогли полностью компенсировать упадок кибернетики, поскольку не используют должным образом лежащий в основании кибернетики системный принцип – и это веский аргумент в пользу обновленной кибернетики).

2. Недостаточность и противоречивость понятия информации

Хорошо известно, что процесс распознавания у человека и животных – это сложный и многоэтапный процесс, сопровождающийся неоднократным перекодированием, сжатием, обобщением и изменением представления (языка и набора признаков) информации, относящейся к образу, или, точнее, образам вместе с фоном и шумом (экспериментальные данные по этому вопросу достаточно хорошо изложены в [9], а их анализ проведен, в частности, в работе [10]). Кажется естественным, что описание всех этих процедур должно осуществляться на основе понятия информации. Но позволяет ли современное понятие информации это сделать? Можно однозначно и категорически утверждать – нет, не позволяет. Рассмотрим кратко причины этого. Не ругал понятие информации разве что самый ленивый, в литературе существует множество различающихся формулировок этого понятия, но мы отметим лишь самые известные, которые принято использовать в серьезных исследованиях. В 1948 г. К. Шеннон определил количество *передаваемой* информации как логарифм числа возможных состояний сигнала [11] (комбинаторный подход). Введенное понятие получило широкую известность и было модифицировано в нескольких направлениях. Во-первых, была показана тесная связь понимаемой таким образом информации с энтропией. Во-вторых, комбинаторный подход был дополнен вероятностной трактовкой (при этом формулы для количества информации сохранили практически прежний вид). И в-третьих, было введено понятие условной вероятности как разности логарифмов состояний или выборов системы до и после получения информации. В дальнейшем А.Н. Колмогоровым [12] была предложена алгоритмическая трактовка количества информации как наиболее краткого описания программы перевода объекта из одного состояния в другое. Все эти подходы к определению понятия информации объединяют две общие черты: 1) предполагается, что как сообщение в целом, так и все его составные части имеют информационную ценность (причем одинаковую) для воспринимающей системы, т.е. каждая часть сигнала приводит к уменьшению числа возможных состояний системы; 2) предполагается, что все

состояния или выборы воспринимающей системы равноправны. Реально ни одно из этих предположений не выполняется, в связи с чем на практике мы, в основном, используем другое понимание информации, основывающееся не на длине сообщения, а на его важности для воспринимающего субъекта. В связи с этим было предложено большое число подходов к информации, основывающихся на ее важности (см. напр. обзор состояния этого вопроса в [13]). Но ни один из этих подходов не получил в настоящее время широкого признания. По-видимому, это связано с тем, что понятия ценности, важности и аналогичные невозможно ввести без субъективизма – зависимости от субъекта, задачи, сложности информации, этапа ее обработки и т.д. При этом мы получаем многокритериальную задачу. Но хорошо известно, что задачи такого типа в общем случае имеют множество возможных решений (множество Парето). В таком случае вряд ли в обозримом будущем мы будем иметь ценностную трактовку информации, более четко определенную, чем, например, понятия свободы, справедливости или моральности.

Таким образом, ни одна из существующих трактовок информации не позволяет использовать это понятие для описания процесса передачи информации между уровнями и представлениями (языками) в процессе распознавания образов. В связи с этим, возможно, необходима разработка понятий, заменяющих понятие информации, но более узких по содержанию и позволяющих более однозначно и адекватно описывать информационные преобразования в процессе распознавания. Помимо этого, необходимо использовать и развивать теорию перевода. При этом акцент должен быть сделан на понимании сущности перевода между языками различного (а не одинакового) уровня и различными представлениями.

3. Неадекватность классической схемы при использовании понятия «образ» и его компонентов

В процессе распознавания по классической схеме в отношении распознаваемого образа используется несколько допущений (их можно назвать также принципами, подходами, схемами и т.п.), в частности, следующие:

- на пути между исходной «сырой» информацией и конечным распознанным образом вводится некоторый промежуточный этап упорядочения исходной информации, называемый признаком или набором признаков;
- распознавание осуществляется по схеме: получение изображения, выделение из него признаков, получение образа как комбинации признаков, процесс отнесения образа к одному из классов (назовем это распознаванием снизу вверх);
- во всех случаях как набор распознанных признаков, так и получаемый из него образ может быть распознан единственным правильным способом.

Насколько обоснованы эти допущения?

Рассмотрим первое допущение. Как следует из литературы, «задача выбора признаков... понимается как преобразование (отображение) исходного пространства в пространство признаков меньшей размерности без потери информации» [14, с. 8], т.е. с помощью признаков на первом этапе существенно сокращается как описание образа, так и последующее оперирование информацией относительно образа. На втором этапе признаки используются для построения образа в виде набора признаков, и далее происходит собственно распознавание. Если второй этап любовно, тщательно и детально описан многочисленными авторами, то по поводу первого пишут мало и в основном в терминах целей, а не

алгоритмов. Неудовлетворительное состояние этого вопроса отмечается, в частности, в работе [2, с. 152]: «Хорошо мы научились решать задачи, в которых еще до обучения так или иначе уже построено пространство признаков или задан словарь языка... И чем удачнее (человеком, а не машиной) выбраны признаки, тем компактнее образы, а значит, тем успешнее решается задача обучения... Но если признаки заранее не заданы или выбираются случайно и неудачно, то все методы теряют работоспособность, их надежность и качество катастрофически снижаются... Итак, мы еще не научились главному – автоматизации процесса построения таких пространств изображений объектов, в которых объекты были бы компактными».

Чтобы понять причины недостаточного внимания к проблеме выделения признаков, рассмотрим распознавание образов на понятийном уровне у человека. Очевидно, что процесс распознавания в этом случае состоит, во-первых, более чем из двух уровней, а во-вторых, на всех вышестоящих уровнях мы работаем, по сути, не с признаками, а с понятиями, которые более правильно представлять как образы, а не как признаки. Например, когда мы распознаем человека, согласно классической схеме мы сначала распознаем признаки. В частности, пусть одним из этих признаков будут очки. Но образ очков сам по себе достаточно сложен, и в другой задаче распознавания очки могут выступать в качестве самостоятельного распознаваемого образа, содержащего подобразы, например, линзы и т.д. Следовательно, по крайней мере на уровне понятийного мышления схема распознавания выглядит не как цепочка признак – образ, а как образ 1-го уровня – образ 2-го уровня ... – конечный образ. Таким образом, введение понятия признак – это некий интеллектуальный фокус (по крайней мере на уровне понятийного мышления), с помощью которого мы из рассмотрения процесса распознавания выбрасываем собственно распознавание и оставляем более понятную задачу манипулирования образами различных уровней (более понятна она в том случае, когда мы выбрасываем из этой схемы выходы на мышление, т.е. переменную часть образа, см. раздел 1).

Эти рассуждения, по-видимому, менее применимы к более простым случаям распознавания, например, речи или несложных изображений. Связано это с тем, что в этом случае в качестве признаков или промежуточного этапа на пути к признакам используются достаточно несложные инварианты (непроизводные элементы по Д. Марру [15]), для которых разработаны алгоритмы получения их из исходной «сырой» информации.

Теперь рассмотрим последовательность распознавания: исходная информация – признак – образ. Как указано выше, следуя логике существующего подхода к распознаванию, для сложных случаев ее более правильно выражать цепочкой: исходная информация – образ 1-го уровня... – конечный образ. Но если исходить из экспериментальных данных, то у человека, по крайней мере, при зрительном восприятии, работает совершенно другая схема. В этой схеме наше распознавание сцены в целом базируется прежде всего на некотором целостном понимании, которое в психологии принято называть гештальтом. «Главное эмпирическое существо открытого гештальтизмом феномена целостности перцептивного образа... заключается в доминировании целостной структуры перцепта над восприятием его отдельных элементов» [16, гл. 9]. Может последовать возражение, что эти выводы относятся к процессу мышления –

стадии, следующей за распознаванием, когда мы уточняем содержание сцены. Однако имеются веские нейрофизиологические данные, свидетельствующие, что и сам процесс распознавания образов осуществляется по гештальтной схеме. Так, в работе [17] было показано, что разные пространственно-частотные компоненты сложного изображения у живых существ выделяются и описываются в зрительной коре независимо работающими каналами. В дальнейшем в работах В.Д. Глезера [18] и других исследователей была предложена и экспериментально обоснована схема с использованием фурье-преобразования. Низкочастотному фурье-преобразованию соответствуют крупномасштабные детали изображения, высокочастотной части – мелкие детали. Экспериментально было доказано, что крупномасштабные детали человек воспринимает быстрее, чем мелкомасштабные (см. детальный обзор и анализ этого вопроса в [18, гл. 3]). Соответственно и распознавание начинается с крупных деталей, описывающих сцену в целом, а не отдельные ее признаки. Этот вывод является простым следствием из законов передачи информации: малое разрешение требует меньше информации, чем высокое, а в связи с невысокой скоростью передачи информации у человека (примерно 10^2 импульсов/сек на одно нервное волокно) или большим ее объемом, информация о деталях изображения воспринимается распознающей системой (мозгом) заметно позднее, чем очертания сцены в целом.

Таким образом, распознавание образов у человека правильнее было бы описывать следующей схемой: исходная информация – образ в целом – уточнение деталей (признаков) образа (распознавание сверху вниз). В более сложных случаях, учитывая, что человек в состоянии оперировать одновременно с 7 ± 2 объектами, схема распознавания, по-видимому, усложняется и выглядит примерно следующим образом: исходная информация – одновременное выделение и предварительное распознавание нескольких образов – одновременное уточнение деталей каждого из образов и распознавание сложного образа или придание смысла сцене (сочетание элементов распознавания сверху вниз и снизу вверх).

С учетом состояния искусственных и естественных распознающих систем представляется, что нет и не может быть единой правильной схемы распознавания образов. Возможно использование самых различных схем, причем это определяется предметной областью, задачами исследования и имеющимся инструментарием. Основанием к выбору таких схем должна быть отсутствующая в настоящее время (см. раздел 2) общая теория преобразования и передачи информации между представлениями и языками различного уровня и происхождения.

В заключение этой темы добавим, что экспериментальные данные [17-19] показывают, что в достаточно сложных распознающих системах, в частности у человека и высших животных, используется одновременно несколько распознающих схем.

Рассмотрим теперь допущение о единственности распознаваемого образа или образов. Очевидно, что это допущение существенно опирается на допущение о единственности классификации. Не оспаривая этого допущения для некоторых простых случаев, заметим, что в сложных случаях оно, как правило, не выполняется [20, гл. 2]. Например, специалистам хорошо известно, что при описании социально-экономических процессов в зависимости от поставленных

целей возможны самые различные системы классификации. По-видимому, этот вывод можно распространить на все сложные системы и на все неполные и неточные описания, т.е. за исключением специальных случаев, практически на все, что мы используем для описания и представления информации на понятийном уровне. В таких случаях задача распознавания дополняется этапом выбора классификационной системы. Истины ради следует заметить, что даже у такой совершенной распознающей системы, как человек, процесс классификации сложной информации осуществляется не в процессе распознавания, а в процессе анализа информации, т.е. мышления. Здесь мы еще раз убеждаемся, что грань между распознаванием и мышлением очень тонкая и мы не можем рассматривать распознавание без мышления. Более того, сам процесс мышления, оказывается, также может быть встроен в процесс распознавания.

Таким образом, отметим в заключение данного раздела, что приведенные в начале раздела допущения классической схемы распознавания либо применимы только для наиболее простых частных случаев (использование единственной классификации), либо работают с точностью до наоборот (схема распознавания снизу вверх), либо приводят к подмене задачи распознавания (введение понятия «признак» без процедур его получения).

4. Недостаточное использование априорной информации

Чтобы более наглядно оценить важность априорной информации, представим, что мы пытаемся распознать образ без всяких предварительных сведений о нем. Нетрудно заметить, что задача в такой постановке эквивалентна задаче установления контактов (понимания) с инопланетной цивилизацией на основании радиосигналов. Несколько десятилетий назад такая идея (о возможности понимания или распознавания) горячо обсуждалась в обществе, но однозначный ответ так и не был получен. Это и не удивительно, учитывая низкий уровень и кризис научного направления, связанного с распознаванием. Но даже и в обозримом будущем, при условии устранения всех перечисленных выше проблем, трудно предположить, что задача распознавания сообщения инопланетной цивилизации по радиосигналу (но не в процессе диалога) осуществима в полном объеме без учета какой-либо дополнительной информации.

На самом деле, априорная информация, конечно же, используется в процессе распознавания (иначе наука о распознавании не существовала бы ни в каком виде). В частности, предварительный выбор признаков может быть осуществлен только с использованием этой дополнительной, обычно плохо формализованной информации о предметной области, и, следовательно, выбранные заранее признаки содержат в себе априорную информацию об образе [1, с. 23]. Да и в схемах распознавания, используемом инструментарии, форме получаемых результатов и т.п. содержится некоторая априорная информация. Второй метод использования априорной информации – это обучение на примерах. Для решения ряда задач обучения часто бывает достаточно.

Проблема, однако, заключается в том, что нет общей теории (метода, способа) построения оптимальных признаков на основе априорной информации. Да и как уже указывалось ранее, реальный процесс распознавания у живых существ имеет многоуровневый характер, а теории распознавания в этом случае вообще не существует. В ряде случаев дополнительная информация о

распознаваемом образе вообще может быть представлена отрывочно, неполно и неточно. Для всех этих случаев нет теории (процедур, механизмов), позволяющих достаточно простым образом получить максимальный эффект от разбросанной и представленной в различной форме дополнительной информации о распознаваемом образе.

Каким образом это сделать? Немедленных и простых рецептов здесь, по-видимому, не существует. Необходимо развитие теории информации, необходимо более глубокое понимание роли и механизмов проявления контекста, внешних критериев и т.п. понятий [10], [21].

5. Недостаточность существующей математики и парадигмы вычислений

Одной из ключевых проблем распознавания образов является необходимость проведения большого объема вычислений, которые далеко не всегда удается реализовать и на супермощных компьютерах. Например, в статистических методах распознавания число решающих правил, которые необходимо оценить в процессе распознавания даже для наиболее простых линейных бинарных правил составляет величину порядка 2^m , где m – размерность пространства [1, с. 91], равная числу признаков. Учитывая, что число признаков во многих задачах превышает десятки и сотни, мы получаем задачи, которые рассчитать традиционными методами невозможно (проклятие размерности).

Существует несколько направлений, в рамках которых пытаются решить вычислительные проблемы. Условно их можно разделить на три класса:

- нейросетевые методы;
- новые методы в рамках традиционной идеологии;
- нетрадиционные методы (мозг и квантовые компьютеры).

Развиваемые как в рамках традиционной идеологии, так и в нейросетях методы явно или неявно используют доказанную А.Н. Колмогоровым теорему существования о представимости функции многих переменных суперпозицией конечного числа функций одной переменной (при выполнении ряда ограничений) [22]. С точки зрения задачи распознавания эта теорема важна тем, что показывает теоретическую возможность резкого снижения размерности пространства, в результате чего объем вычислений растет с увеличением количества признаков не экспоненциально, а полиномиальным образом.

Одним из широко известных методов, существенно использующим принципы, заложенные в теореме Колмогорова, является, на наш взгляд, метод группового учета аргументов (МГУА), предложенный А.Г. Ивахненко ([23], [24], а также другие работы на сайте www.inf.kiev.ua/GMDH-home). Вообще говоря, этот метод не сводится только к указанной теореме, и его вторая существенная положительная черта – наличие механизмов построения модели оптимальной сложности с учетом короткой выборки или зашумленных исходных данных. Сейчас этот метод бурно развивается целой группой исследователей из различных стран, создан коммерческий продукт на его основе (KnowledgeMiner), немаловажно также, что он может быть реализован как с использованием традиционных последовательных алгоритмов, так и с помощью нейросетей. Таким образом, он имеет все предпосылки стать одним из ведущих при использовании систем распознавания и вообще систем получения новых знаний. В то же время необходимо отметить и недостаточный анализ теоретических

предпосылок, лежащих в его основе, что затрудняет его развитие и затемняет трактовку полученных результатов.

Представляется очень интересным также предложенный В.И. Васильевым метод редукции или метод предельных упрощений [2, гл. 8], [25], [26], основанный на предельно возможном снижении размерности признакового пространства, в результате чего резко снижаются требования как к объему выборки, так и к количеству вычислений. В статистике эта идеология (снижение размерности пространства) хорошо известна, детально разработана и широко используется [27, гл. 12-17]. Известна и ограниченность этой идеологии, связанная с тем, что мы не можем выбросить важные признаки. В предельном случае одинаковой важности всех признаков мы вообще не можем сократить размерность пространства признаков. Поэтому представляется, что основная область применимости теории редукции – не собственно процесс распознавания в классической схеме, а стадия выбора признаков – процесс конструирования признакового пространства. А поскольку этот процесс должен основываться на мышлении, то теория редукции тесно примыкает к мышлению, языку, способам представления информации и т.д. и, очевидно, не может развиваться в отрыве от этих понятий.

Классическая математика основана на использовании принципа последовательных вычислений. При этом неявно используется допущение о том, что мы можем процесс преобразования информации о любой предметной области представить в виде последовательной цепочки вычислений (назовем его принципом локальности). Но это не единственно возможный в настоящее время принцип обработки информации. Принцип параллельной обработки, используемый, в частности, в нейросетях, также хорошо известен и широко используется, он позволяет ускорить процессы обработки информации в n раз, где n – число параллельных каналов. Но есть еще один принцип – принцип совместной обработки информации, когда производится взаимосогласованная обработка информации в ряде алгоритмов или элементов вычислительного устройства (нелокальная обработка информации). Этот принцип реализован, в частности, в квантовых компьютерах [28], [29] – новом, перспективном и быстро развивающемся направлении. Возможно также, что по этому же принципу хотя бы в некотором отношении работает и наш мозг.

Понятие «нейросети» объединяет в себе сотни различных методов с различной идеологией и аппаратной реализацией, в связи с чем обзор этих методов с единых позиций, как уже указывалось в предисловии, крайне затруднен. По этой причине мы лишь отметим, что наиболее перспективными с точки зрения распознавания образов считаются сети Кохонена [6], концепция когнитрона или неокогнитрона [30], [31] и теория инвариантов (см. подборку материалов на сайте www.kyb.tuebingen.mpg.de). Нельзя также не отметить попытки преодоления колмогоровского барьера сложности с помощью нейросетей [32] (там же и обзор литературы по этому вопросу). Если эта попытка удастся в полном объеме, то эффект от этого (по крайней мере, в распознавании образов) будет сравним с использованием квантовых компьютеров.

Работа квантового компьютера основана на регулируемом изменении избранных состояний в квантовой системе, состоящей из n квантовых частиц – кубитов. Такая квантовая система имеет ряд замечательных особенностей. Во-первых, квантовые состояния кубитов взаимно влияют друг на друга, приводя

к появлению 2^n базовых состояний. Во-вторых, если мы переводим один из кубитов в другое состояние, то все остальные состояния очень быстро (со скоростью интерференционных процессов) переходят в новое состояние, с учетом изменения состояния исходного кубита. Такой процесс с помощью некоторых дополнительных алгоритмов можно представить как процесс вычисления. И, в-третьих, что немаловажно, мы можем снимать информацию о новом состоянии, т.е. получать результаты расчетов [29]. Если рассматривать этот процесс в рамках традиционных понятий, то это не просто параллельные вычисления (в параллельных вычислениях отдельные процессы независимы друг от друга по крайней мере на протяжении некоторых циклов), а взаимосогласованная обработка информации, когда каждый кубит все время «чувствует» состояния других кубитов, и все время согласованно с ними изменяет свое собственное. В результате этого скорость расчетов растет экспоненциально с увеличением числа элементов в квантовом компьютере. В настоящее время на практике реализованы лишь отдельные элементы квантовых компьютеров, но специалисты не сомневаются в его принципиальной реализации, разногласия в основном касаются сроков реализации – 5 или 10 лет. Ожидаемый прирост производительности квантовых компьютеров по некоторым оценкам достигает 10-ти и более порядков раз, что, несомненно, позволит решать многие в настоящее время неосуществимые задачи распознавания.

В связи с квантовыми компьютерами нельзя также не упомянуть и о возможностях человеческого мозга. Как известно, человек в состоянии держать в своем сознании 7 ± 2 объектов. При этом, если мы оперируем одним из объектов, то и в других происходят более или менее согласованные изменения (связи, трактовка признаков, общее значение данного объекта для ситуации в целом и т.д.), – все это очень напоминает работу квантового компьютера. Вследствие этого можно высказать предположение, что мозг, по крайней мере в отношении некоторых процессов обработки информации, проводит взаимосогласованную одновременную обработку информации. И эта нелокальная обработка реализуется не с использованием квантовых компьютеров, а путем использования сложной структуры межнейронных связей и передачи импульсов между нейронами. Возможно, именно использование этого принципа объясняет колоссальные возможности человеческого мозга, несмотря на то, что количество элементов в современных компьютерах (во всяком случае, в многопроцессорных системах) уже вплотную приближается к числу нейронов в мозгу у человека.

6. Недостаточный учет многоуровневости процесса обработки информации

Специалисты в области нейрофизиологии экспериментально показали, что у человека и животных процесс обработки информации носит многоуровневый и многоэтапный характер [9]. При этом даже на предварительной стадии выделения информации, когда речь о распознавании еще не идет, выделяются 3 – 4 самостоятельных уровня. В общем же количество уровней в зависимости от типа обрабатываемой информации вполне может составлять 6 – 10. Причем каждый из уровней имеет свою «аппаратную» часть (разделы мозга), свой способ представления информации и свой язык.

Экспериментально также показано [9], что, во всяком случае, на начальной стадии обработка информации на каждом уровне осуществляется

простыми и понятными для нас способами, которые можно достаточно просто реализовать на аппаратном уровне, либо программно. В то же время искусственные системы распознавания, как правило, имеют лишь три уровня (исходная информация – признаки – образ) или, в лучшем случае, четыре (исходная информация – производные элементы – признаки – образ). Учитывая превосходство живых существ в распознавании образов над искусственными системами, можно сделать вывод о большей эффективности многоуровневого представления информации.

7. Невозможность создания полноценной теории распознавания в рамках одного научного направления

Несмотря на перечисление ограничений классического подхода, автор вовсе не считает, что в распознавании образов нет достижений. Напротив, за последние десятилетия был развит ряд мощных междисциплинарных направлений, требующих значительных интеллектуальных и практических усилий в нейрофизиологии, теоретической и прикладной математике, физике, искусственных нейросетях и т.д. К числу этих направлений можно отнести статистическую теорию распознавания образов, создание ряда мощных вычислительных методов, обнаружение на основании экспериментальных данных ряда важных принципов обработки информации у человека, развитие идеологии семантических сетей, теория и практика нейросетей... Список можно продолжить. Возникает замечание: что-то из этого списка оказалось не совсем адекватным, но давайте, определим самый подходящий метод или идеологию и, в конце концов, решим проблему распознавания.

Представляется, что это неверный подход. Как уже указывалось в предисловии, есть веские основания считать, что ключевые проблемы распознавания образов, кибернетики и искусственного интеллекта имеют одно и то же происхождение: мы пытаемся описать слишком сложный предмет с помощью слишком простого понятийного аппарата. В частности, к предмету, работу которого нужно объяснить, относится и сам субъект познания – человеческий мозг. Задача эта сложная, значительно сложнее квантовой механики и, возможно, самая сложная из тех, с какими до этих пор сталкивалось человечество. Сложность ее состоит, прежде всего, в необходимости коренного, и, что не менее важно, взаимосогласованного преобразования целого ряда научных дисциплин. Теория информации должна описывать преобразование информации на всех уровнях и этапах ее переработки – зрительной, лингвистической и т.д., должны быть установлены законы обобщения, принципы оптимального представления информации и т.д. В математике должна быть создана в полном объеме теория преобразования сложных многомерных данных, представленных в неполной и неточной форме, а также любыми видами описания – формальными, на естественных языках и т.д. Нейросети должны прийти до уровня, моделирующего наиболее важные принципы работы человеческого мозга и квантовых компьютеров.

Выводы

С общих – методологических, философских и познавательных – позиций проанализированы проблемы и ограничения распознавания образов. Отмечено,

что большинство этих проблем имеет общее происхождение с проблемами кибернетики и искусственного интеллекта.

Первопричинами застоя указанных научных направлений является, с одной стороны, попытка объяснить работу интеллекта в рамках автоматного, алгоритмического подхода, а с другой – необычайная сложность объекта исследования.

Основные ограничения классической схемы распознавания образов можно свести к следующим положениям:

1. Распознавание образов не является самостоятельной процедурой, а включено в схему мышления. Не учитывая мышления и связанной с ним индивидуальной компоненты, мы в принципе не можем ни понять процесс распознавания, ни использовать его для построения искусственного интеллекта (а также для объяснения работы естественного).
2. Классическая схема распознавания не является ни бесспорной, ни единственной. Понятие «признак» является внутренне противоречивым и недостаточно обоснованным.
3. Существующего понятия информации недостаточно для адекватного описания преобразования информации в процессе распознавания.
4. Не существует последовательной теории учета априорной информации. Особенно плохо обстоит дело с учетом отрывочной, неполной и неточной информации.
5. Существующие математические методы и основанная на них обработка информации не позволяют провести вычисления для большей части практически интересных случаев. Решение этой проблемы может быть получено тремя взаимодополняющими способами: в рамках классической математики на основе теоремы Колмогорова о возможности представления сложной функции более простыми; с помощью нейросетей, использующих наиболее важные принципы работы мозга; с помощью квантовых компьютеров и аналогичных им устройств, осуществляющих нелокальную обработку информации.
6. Оптимальная реализация процесса распознавания должна содержать значительно больше уровней, чем существующая классическая схема, а само строение уровней и процедуры преобразования информации между ними должны быть гораздо более простыми и понятными.
7. Распознавание образов тесно связано с целым комплексом дисциплин: нейрофизиологией, теорией информации, математикой, нейросетями, искусственным и естественным интеллектом и рядом других научных дисциплин и направлений. В связи с этим крупные прорывы в распознавании образов следует ожидать только на пути комплексного и взаимосогласованного развития этих дисциплин.

Литература

1. Вапник В.Н., Червоненкис А.Я. Теория распознавания образов (статистические проблемы обучения). – М.: Наука, 1974. – 416 с.
2. Васильев И.В., Шевченко А.И. Искусственный интеллект. Формирование и распознавание образов. – Донецк: ДонГИИИ, 2000. – 360 с.

3. Дидук Н.Н., Коваль В.Н. Существует ли наука кибернетика? (О роли кибернетики в естествознании) // Проблемы управления и информатики. – 2001. – № 3. – С. 132-154.
4. Ту Дж., Гонсалес Р. Принципы распознавания образов. – М.: Мир, 1978. – 411 с.
5. Искусственный интеллект: – В 3 кн. Справочник / Под ред. Д.А. Поспелова. – М.: Радио и связь, 1990. – Кн. 2.: Модели и методы. – 304 с.
6. Jain A.K., Duin R.P.W. Mao Jianchang Statistical Pattern Recognition: A Review // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2000. – Vol. 22, № 1. – P. 4-38.
7. Селекция и распознавание на основе локационной информации / А.Л. Горелик, Ю.Л. Барабаш, О.В. Кривошеев, С.С. Эпштейн. – М.: Радио и связь, 1990. – 240 с.
8. Дюк В., Самойленко А. Data Mining: учебный курс. – СПб.: Питер, 2001. – 368 с.
9. Хьюбел Д. Глаз, мозг, зрение. – М.: Мир, 1990. – 239 с.
10. Сторож В.В. Обработка информации у человека. Ч.1. Уровни и базисные элементы // Искусственный интеллект. – 2001. – № 4.
11. Shannon C.E. A Mathematical Theory of Communication // The Bell System Technical Journal. – 1948, July, October. – P. 379-423, 623-656.
12. Колмогоров А.Н. Три подхода к определению понятия «количество информации» // Проблемы передачи информации. – Т. 1, № 1. – С. 3-11.
13. Экономическая кибернетика: Учеб. пособие / Ю.Г. Лысенко, В.Л. Петренко, В.А. Забродский и др. – Донецк: ДонГУ, 1999. – 397 с.
14. Фукунага К. Введение в статистическую теорию распознавания образов. – М.: Наука, 1979. – 368 с.
15. Марр Д. Зрение. Информационный подход к изучению представления и обработки зрительных образов. – М.: Радио и связь, 1987. – 400 с.
16. Веккер Л.М. Психика и реальность. Единая теория психических процессов. – М.: Смысл, 1998.
17. Campbell F.W., Robson J.G. Application of Fourier analysis to the visibility of gratings // J. Physiol. – 1968. – Vol. 197. – P. 551-566.
18. Глезер В.Д. Зрение и мышление. – Л., 1985. – 248 с.
19. Невская А.А., Леушина Л.И. Асимметрия полушарий головного мозга и опознание зрительных образов. – Л.: Наука, 1990. – 152 с.
20. Куссуль Э.М. Ассоциативные нейрноподобные структуры. – Киев: Наук. думка, 1990. – 144 с.
21. Сторож В.В. Работа мозга. Нейролингвистический подход // Искусственный интеллект. – 2001. – № 3. – С. 426-438.
22. Колмогоров А.Н. О представлении непрерывных функций нескольких переменных в виде суперпозиции непрерывных функций одного переменного и сложения // Доклады АН СССР – 1957. – Т. 114, Вып. 5. – С. 953-956.
23. Ивахненко А.Г. Непрерывность и дискретность. – Киев: Наук. думка, 1990. – 224 с.
24. Ivakhnenko A.G., Ivakhnenko G.A. Problems of Further Development of the Group Method of Data Handling Algorithms. Part 1 // Pattern Recognition and Image Analysis. – Vol. 10, № 2. – P. 187-194.
25. Васильев В.И. Теория редукции в проблемах экстраполяции // Проблемы управления и информатики. – 1996. – № 1, 2. – С. 239-251.
26. Васильев В.И., Ланге Т.И., Шевченко А.И. Выявление и моделирование закономерностей сходства, равенства и порядка // Искусственный интеллект. – 2001. – № 3. – С. 26-39.
27. Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики. – М.: ЮНИТИ, 1998. – 1022 с.
28. Килин С.Я. Квантовая информация // УФН. – 1999. – Т. 169, № 5. – С. 507-527.
29. Валиев В.А. Квантовая информатика: компьютеры, связь и криптография // Вестник РАН. – 2000. – Т. 70, № 8. – С. 688-695.
30. Fukushima K. Cognitron: A self-organizing multilayered neural network // Biological Cybernetics. – 1975. – Vol. 20. – P. 121-136.
31. Fukushima K., Miyake S. Neocognitron: A new algorithm for pattern recognition tolerant of deformations and shifts in position // Pattern recognition. – 1982. – Vol. 15, № 6. – P. 455-469.
32. Радченко А.Н. Многомерность без проклятий // Мат-лы конф. «Интеллектуальные системы в информационных технологиях», Псков, 19 – 23 июня 2000 г.

Статья поступила 06.08.02.