УДК 519.7:004,8

В.И. Гриценко, М.И. Шлезингер

ВЗАИМОСВЯЗЬ ПРОБЛЕМ РАСПОЗНАВАНИЯ ОБРАЗОВ, МАШИННОГО МЫШЛЕНИЯ И ОБУЧЕНИЯ*

Ключевые слова: распознавание образов, машинное обучение, выполнимость ограничений, гиббсовы поля, субмодулярная минимизация.

Введение

Наука об искусственном интеллекте подвергается сейчас серьезному испытанию, которое, следуя знаменитой заметке К. Шеннона [1], можно назвать бандвагоном. Бандвагон, в буквальном смысле этого слова, — это музыкальная группа на платформе, проезжавшая в былые времена по американскому городу, прославляя успех победителя на прошедших выборах. К. Шеннон использовал это слово для обозначения чрезмерно восторженной рекламы научных достижений, адресованной главным образом за пределы научного сообщества. Бандвагон по своей сути — вненаучное явление, но он влияет на науку и в силу этого заслуживает серьезного научного анализа. Мы выполним этот анализ в определенных ограниченных рамках и уклонимся от вопроса, насколько возникший бандвагон плодотворен или разрушителен для науки об искусственном интеллекте в целом, и рассмотрим лишь, как он влияет на распознавание образов, которое традиционно считается одним из разделов искусственного интеллекта.

За последние 10–15 лет в распознавании образов действительно произошел значительный прогресс. Существенно изменилось научное содержание проблемы распознавания образов, ставшего самостоятельной дисциплиной в системе компьютерных наук, а не просто областью применения известных математических методов оптимизации и статистических решений. Наряду с наукой о распознавании возникла индустрия распознающих систем, и их разработка в значительной степени переместилась из исследовательских лабораторий в коммерческие предприятия. Наряду со спросом на единичные распознающие системы, предназначенные для эксплуатации в одной или нескольких крупных организациях (например, анализ космических снимков), появился массовый спрос на персональные распознающие системы, предназначенные для использования в повседневной жизни (например, системы визуального контроля доступа в помещение). Коммерциализация распознавания образов формирует новые условия его развития, в которых, наряду с научными результатами, значительную роль играют методы конкурентной борьбы, рекламы и другие средства успешной рыночной активности.

* Работа выполнена по заданию Отделения информатики НАН Украины (государственный регистрационный номер 0118U002410).

Возрастающая популярность распознавания образов вызывает у исследователей этого направления не только удовлетворение, но и вполне оправданную тревогу, когда эта популярность приобретает явное сходство с бандвагоном. Как это уже происходило прежде, распознавание образов опять приобретает репутацию волшебной палочки-выручалочки, позволяющей решать любую сложную задачу без каких-либо интеллектуальных усилий по ее исследованию, а иногда даже без ее точной формулировки. Основным действующим лицом этого прекрасного мифа является некая система (биологически инспирированная, генетическая, эволюционная, нейронная или с иным экзотическим названием), действующая следующим образом. Сначала система обучается на совокупности $[(x_i, k_i) \mid i = 1, \ldots, n]$ примеров «правильного» распознавания. Каждый пример (x_i, k_i) указывает результат k_i, который должна выдать распознающая система для входных данных x_i. В результате этого обучения распознающая система обретает способность выдавать «правильный» результат и для таких входных данных x, которые не входили в предъявленную совокупность примеров. Система такого вида позволяет решать разнообразные задачи распознавания, избавляя разработчика от конструирования алгоритмов их решения и возлагая на него лишь формирование достаточного количества обучающих примеров.

Этот миф вполне успешен в роли рекламной афиши и в этом качестве вызывает лишь здоровый скептицизм исследователей, а не желание его оспаривать. Однако он заслуживает самого серьезного рассмотрения, когда проникает в науку в виде как бы само собой разумеющегося тезиса, что научная проблематика распознавания полностью исчерпывается проблематикой обучения, а термины «Pattern Recognition» и «Machine Learning» — это просто синонимы. Такое представление проблемы недопустимо сужает ее, исключая из нее обширный класс проблем собственно распознавания, выходящих за рамки машинного обучения. Отождествление распознавания образов и машинного обучения характерно для ранних стадий исследований в середине прошлого века, когда задачи распознавания исследовались только для отдельных хорошо изученных к тому времени статистических моделей. Для каждой такой модели алгоритм распознавания следовал непосредственно из ее определения и допускал вычислительно эффективную реализацию. В силу этого на первый план выходила проблема построения статистической модели на основе тех или иных эмпирических данных, называемых обучающей выборкой. Однако решение реальных, а не модельных задач, привело к существенному изменению приоритетов. Ключевую роль в этом прогрессе сыграли работы [2–5] по распознаванию речевых сигналов и текстовых строк. В этих работах показано, что определенные задачи оптимизации необходимо решать не только при построении алгоритма распознавания, но и для каждого отдельного объекта, предъявленного для распознавания. Это положило начало нового на то время направления в распознавании образов, которое получило название структурного распознавания в отличие от классического, рамки которого сформировались до 1968 года.

Полвека, прошедшие после опубликования работ [2, 4], были периодом существенного развития их идейного содержания и формирования науки о распознавании как составной части машинного мышления. Поэтому возрождение мифов о всесильности обучения, которое может справиться со всем многообразием задач распознавания, — не что иное как возврат к состоянию полувековой давности. Цель данной статьи — не критический анализ известных методов обучения, хотя и будут рассмотрены некоторые их особенности, которые можно понимать как их изъяны. Цель статьи состоит в обзоре основных идей современного распознавания, которые выходят за рамки проблемы обучения. Мы стремимся показать, насколько сужается научная проблематика распознавания образов при его отождествлении с машинным обучением.

В разд. 1 «Структурное распознавание и мышление» рассмотрены характерные черты, присущие как структурному распознаванию, так и мышлению определенного типа, названному образным; сформулированы понятия, служащие основой для последующей формализации информационных процессов, реализующих как структурное распознавание, так и машинное мышление.

В разд. 2 «Математический аппарат структурного распознавания» приведены основные понятия классической проблемы распознавания совместимости ограничений — одной из признанных парадигм машинного мышления; показано, как следует их модифицировать применительно к реальным задачам распознавания, и сформулирована обобщенная задач разметки, частным случаем которой является классическая задача совместимости ограничений; приведен краткий обзор наиболее известных полиномиально-разрешимых подклассов задач разметки.

В разд. 3 «Гиббсовы поля и задачи разметки» на примере гиббсовой модели распознаваемых объектов показаны фундаментальные проблемы, лежащие за пределами проблематики машинного обучения, и показано, как эти проблемы сводятся к задачам разметки.

В разд. 4 «Распознавание и обучение» сформулирован общий формат процедур обучаемого распознавания в рамках общей теории статистических решений; сформулирован подкласс этих процедур, в который входят известные в настоящее время процедуры.

В разд. 5 «Риск-ориентированные стратегии» сформулирована риск-ориентированная постановка задачи обучаемого распознавания и указан класс процедур, решающих задачу в этой постановке; показано, что ряд наиболее известных процедур не входит в этот класс, что можно рассматривать как их существенный изъян.

1. Структурное распознавание и мышление

В данном разделе и разделе 2 формулируется система понятий, которые служат основой для формализации определенного класса задач распознавания, известных как структурное распознавание. Представленная далее система понятий охватывает не весь круг проблем, которые уместно относить к распознаванию образов, а ту его часть, в которой процессы распознавания можно интерпретировать как мыслительные процессы определенного типа. Укажем характерную черту таких мыслительных процессов, которая отличает их от вычислений. Это различие состоит в том, что вычислительное и мыслительное устройства решают задачи на основании существенно различных форматов представления исходных данных задачи.

Задача считается сформулированной для ее решения на вычислительном устройстве, если задана тройка $\langle X, Y, f\rangle$, где X — множество возможных

ISSN 0572-2691

входных данных задачи, Y — множество результатов ее решения и $f: X \to Y$ — алгоритм вычисления результата $f(x) \in Y$ для входных даных $x \in X$. В тройке $\langle X, Y, f \rangle$ отсутствует то, что в обыденной практике понимается как условие задачи, подлежащей решению. Дополним тройку $\langle X, Y, f \rangle$ функцией $\varphi: X \times Y \to$ $\to \{0, 1\}$, которая формализует понятие условия задач и для каждой пары $x \in X$, $y \in Y$ указывает, является ли y правильным решением задачи при исходных данных X. Отметим при этом, что проверка условия $\varphi(x^*, y^*) = 1$ для заданной пары (x^*, y^*), как правило, существенно проще реализации алгоритма $f: X \to Y$ для заданных входных данных x^*. Поэтому, если для решаемой задачи известен не только алгоритм f решения задачи, но и условие φ решаемой задачи, то полученный результат y^*, независимо от того, как он был получен, можно окончательно проверить, подставив его в условие задачи и убедившись, что $\varphi(x^*, y^*) = 1$. Вычислительное устройство, которому известен не только алгоритм $f: X \to Y$, но и условие $\varphi: X \times Y \to \{0, 1\}$ задачи, приобретает способность самоконтроля получаемых результатов, что в свою очередь позволяет смягчить требования к алгоритму f. Алгоритм может иметь эвристический характер и не гарантировать решение произвольной поданной на его вход задачи, потому что самоконтроль $\varphi: X \times Y \to \{0, 1\}$ позволяет блокировать ошибочные результаты.

В отличие от вычислительного устройства, мыслительное устройство решает задачу не на основании тройки $\langle X, Y, f: X \to Y \rangle$, а на основании тройки $\langle X, Y, \varphi: X \times Y \to \{0,1\} \rangle$, где φ — условие задачи. В отличие от вычисления, которое состоит в получении значения $y^* = f(x^*)$ для исходных данных x^* с помощью заданного алгоритма f, мышление состоит в поиске такого значения y^*, которое обеспечивает равенство $\varphi(x^*, y^*) = 1$ для заданных x^* и $\varphi: X \times Y \to \{0, 1\}$. Эту характеристическую черту иногда выражают в виде лаконичного противопоставления «НЕ КАК, А ЧТО», которое обозначает, что в исходных данных задачи указывается лишь, что должно быть получено в результате решения, а не как следует получать этот результат.

Приведенное понимание различия между вычислением и мышлением не претендует на какую-либо однозначность или строгость, и поэтому приведем пример, иллюстрирующий это различие. Пусть X — множество 16-разрядных двоичных чисел, и некое устройство располагает памятью для хранения троек $(x_1, x_2, x_3) \in X^3$, таких что число x_3 есть сумма чисел x_1 и x_2. Вычислительное устройство действует на основании заданного алгоритма $f: X \times X \to X$ суммирования чисел и для заданных значений x_1^*, x_2^* входных переменных x_1, x_2 вычисляет значение $f(x_1^*, x_2^*)$ выходной переменой x_3. Подобно тому, как это принято в работах [6, 7], мы считаем, что в мыслительном устройстве отсутствует априорное разделение данных на входные и выходные, а есть общая память для хранения тройки (x_1, x_2, x_3) чисел и априори заданное условие $x_3 = x_1 + x_2$, которому должно удовлетворять содержимое памяти. Работа мыслительного устройства состоит в том, что при заданном содержимом того или иного подмножества памяти устройство доопределяет содержимое остальной памяти так, чтобы выполнялось равенство $x_3 = x_1 + x_2$. В частности, если определены значения переменных x_1, x_2, устройство работает как сумматор, а если определены значения переменных x_1, x_3, устройство определяет значение переменной x_2 как разность $x_3 - x_1$. Более того, если значение каждой из переменных (x_1, x_2, x_3) определено частично, т.е. лишь отдельные разряды двоичного представления чисел, устройство доопределяет значения всех остальных разрядов, исходя из требования $x_3 = x_1 + x_2$. Задача в этом случае становится сходной с математическими головоломками, решаемыми ради развлечения (например, судоку) или для тестирования мыслительных способностей (типа РЕШИ+ЕСЛИ=СИЛЕН или TWO+TWO=FOUR).

Решение задач с помощью однозначно заданного алгоритма совсем не характерно для мыслительной деятельности человека. Для многих людей такая работа очень утомительна, а для некоторых — просто невыполнима. С другой стороны, для очень многих людей решение задач на основе заданных требований к решению, например, решение головоломок, служит увлекательным интеллектуальным развлечением, пусть и не всегда легким. Наконец, интеллектуальные способности человека тестируются не по его способности дисциплинированно действовать в соответствии с заданным алгоритмом, а по умению находить требуемый результат. Все это может послужить доводом, веским для одних и менее веским для других, что рассматриваемый нами класс информационных технологий адекватно отражает одну из многих граней мыслительной деятельности человека. Мы назвали этот тип мыслительных процессов образным мышлением [8] и согласны с любой критикой этого названия, потому что не придаем этому принципиального значения.

Эти неформальные соображения можно конкретизировать в виде следующей, все еще предварительной формулировки того класса задач структурного распознавания и образного мышления, о которых идет речь в данной статье. Точная формулировка этого класса задач представлена в разд. 2.

Пусть X — конечное множество, элементы $x \in X$ которого назовем метками, а T — конечное множество, называемое памятью. Элементы $t \in T$ памяти назовем ячейками, а подмножества $s \subseteq T$ — окошками. Заполнение памяти T символами из X будем представлять в виде функции $\bar{x}: T \to X$, которую назовем разметкой. Обозначим $x(t)$ метку, которую определяет разметка $\bar{x}$ для ячейки $t \in T$, и $x(s)$ — сужение разметки $\bar{x}$ на окошко $s \subset T$. Сужение $x(s)$ назовем фрагментом разметки $\bar{x}$ в окошке $s \subset T$. Пусть $\varphi: X^T \to \{0, 1\}$ — условие задачи, которое для каждой разметки $\bar{x}: T \to X$ указывает ее допустимость, $\varphi(\bar{x}) = 1$, или недопустимость, $\varphi(\bar{x}) = 0$. Исследуемые нами задачи распознавания состоят не в вычислении значения $\varphi(\bar{x}^*)$ для заданной разметки $\bar{x}^* \in X^T$, а в определении, существует ли допустимое расширение заданного фрагмента $x: s \to X$ в окошке $s \subset T$ на всю память T, т.е. существует ли разметка $y: (T \setminus s) \to X$, такая что $\varphi(x, y) = 1$.

 ISSN 0572-2691

Определение 1. Исходными данными задачи структурного распознавания являются конечные множества T и X и функция $\varphi: X^T \to \{0,1\}$, а само распознавание заключается в вычислении числа

$$\psi(x^*) = \bigvee_{y \in X^{T \setminus s}} \varphi(x^*, y) \tag{1}$$

для заданных $s \subset T$ и $x^*: s \to X$ и, если $\psi(x^*) = 1$, — построении разметки $y^*: (T \setminus s) \to X$, такой что $\varphi(x^*, y^*) = 1$.

Отметим, что вычисление значения функции ψ для заданной разметки x^* окошка s может оказаться (и, как правило, оказывается) существенно более сложной вычислительной процедурой, чем проверка условия $\varphi(x^*, y^*) = 1$ задачи для заданной пары (x^*, y^*).

Введенная система понятий позволяет проследить, как возникают мифы, что проблема распознавания образов полностью исчерпывается проблемой машинного обучения, и как эти мифы разрушаются при первых же контактах с реальными задачами. Ошибочное представление о всесильности процедур обучения возникает в результате непонимания фундаментального различия между однозначным определением функции или множества и эффективным вычислением значения этой функции в заданной точке или распознаванием принадлежности заданной точки заданному множеству. На ранних этапах классического распознавания это различие не принималось во внимание, потому что использовались только легковычислимые функции (линейные, квадратичные, пороговые и т.п.) и множества (линейные подпространства, сферы, эллипсы и т.п.), которые были хорошо усвоены в прикладной информатике того периода. Алгоритмы вычисления значения таких функций в заданной точке или принадлежности заданной точки таким множествам тривиальным образом следуют из их определения. В современном распознавании потребовались другие функции и другие множества. Основная проблема современного распознавания, в отличие от классического, состоит не только и не столько в том, что какие-то функции или множества не известны, а в том, что они трудновычислимы. Даже если эти функции известны и однозначно определены, вычисление значения этих функций в заданной точке наталкивается на препятствия фундаментального характера.

Трудновычислимые функции не являются чем-то экзотическим для распознавания. Скорее наоборот, редкостью являются задачи распознавания с легковычислимыми функциями. Определение (1) непосредственно показывает, как любую задачу с легковычислимыми функциями буквально со всех сторон «окружают» вполне реалистичные задачи с трудновычислимыми функциями. Пусть распознаваемый объект характеризуется скрытым состоянием $k \in \{1, 2\}$ и совокупностью $\bar{x} = (x(t) \in X \mid t \in T)$ сигналов (признаков), излучаемых объектом. Совокупность $\bar{x}$ зависит от состояния k объекта, и эта зависимость представлена функциями $\varphi^k: X^T \to \{0,1\}$, $k \in \{1, 2\}$, такими, что $\varphi^k(\bar{x}) = 1$ тогда и только тогда, когда объект находится в состоянии k. Если функции $\varphi^k: X^T \to \{0,1\}$ легко вычислимы и известны значения $x(t)$, $t \in T$, всех сигналов, то отсутствуют какие-либо препятствия для принятия решения о состоянии объекта на основании этих данных. Однако эти препятствия немедленно возникают в такой вполне реалистичной ситуации, когда не все сигналы доступны для наблюдения, а только сигналы $x^0 = (x(t) \in X \mid t \in s)$ в окошке $s \subset T.$ В этом случае решение о состоянии объекта должно приниматься на основании значений

$$\psi^k(x^0) = \bigvee_{y \in X^{T \setminus s}} \varphi^k(x^0, y), \quad k \in \{1, 2\}, \tag{2}$$

а не значений $\varphi^k(\bar{x})$. Вычисление значений $\psi^k(x^0)$, $k \in \{1, 2\}$, может представлять трудность фундаментального характера, несмотря на то, что они вполне однозначно определены выражением (2). Известно, что для некоторых, даже очень легко вычислимых функций φ^k вычисления по формуле (2) образуют NP-полный класс проблем.

Таким образом, мы говорим о трудных задачах распознавания, в которых вычисление значений функций φ^k, $k \in K,$ для заданного набора $\bar{x}^*$ признаков вызывает серьезные вычислительные трудности, и легких задачах, где такие трудности не возникают. На ранних этапах классического распознавания исследовались только легкие задачи распознавания, что и породило ошибочное отождествление проблематик распознавания образов и машинного обучения. Это ошибочное представление, вполне объяснимое для науки о распознавании на самых ранних этапах ее зарождения, недопустимо сужает проблематику современного распознавания, которое имеет дело только с трудными задачами распознавания.

2. Математический аппарат структурного распознавания

Как было сказано, определение 1 задачи является предварительным, потому что очерчивает чересчур широкий класс информационных процессов, в который входят структурное распознавание и образное мышление. Общеизвестной конкретизацией этого определения является проблема выполнимости ограничений [9, 10] — одна из основных парадигм машинного мышления. Применение этой теории к задачам распознавания образов приводит к расширению математической проблематики этой науки (см. [11] и выпуски трудов конференции EMMCVPR). Как следствие, расширяется и понятие машинного мышления, на формализацию которого эта наука ориентируется. Таким образом, здесь наблюдается тот особо ценный момент взаимодействия фундаментальной и прикладной наук, когда это взаимодействие оказывается плодотворным не только в области применения теории, но и для самой применяемой теории. В процессе соприкосновения теоретических схем с жесткими прикладными требованиями они неизбежно уточняются, модифицируются и в таком обогащенном виде возвращаются в фундаментальную науку.

Основные понятия классической теории непротиворечивости ограничений приведены в подразделе 2.1. В подразделе 2.2 сформулирована общая задачи структурного распознавания, которая представлена как обобщение классической проблемы непротиворечивости ограничений, а в подразделе 2.3 — ее известные на сегодня полиномиально разрешимые подклассы.

2.1. Проблема выполнимости ограничений. Основные понятия классической проблемы удовлетворения ограничений сформулированы в [9], а также в необозримом множестве других учебников и монографий. Привязка этих понятий

ISSN 0572-2691

к контексту данной статьи приводит к следующей их формулировке. Как и прежде, X и T — это два конечных множества, элементы которых называются, соответственно, метками и ячейками, $\bar{x}: T \to X$ — функция, называемая разметкой, $x(t)$ — метка ячейки $t \in T$, а $x(s)$ — фрагмент разметки $\bar{x}$ в окошке $s \subset T$. Пусть, дополнительно к этому, для множества T задано подмножество $\mathbb{S} = \{s_1, s_2, \ldots, s_n\}$ окошек, называемое его структурой. В частности, структурой может быть определенное подмножество пар или троек ячеек, или если T — поле зрения изображения, то структурой может быть определенное подмножество прямоугольных участков поля зрения. Число $ord(\mathbb{S}) = \max\limits_{s \in \mathbb{S}} |s|$ называется порядком структуры $\mathbb{S}$.

Пусть для каждого окошка $s \in \mathbb{S}$ определена функция $\varphi_s : X^s \to \{0,1\}$, которая для каждого фрагмента $u \in X^s$ на окошке s указывает, допустимо ли размещение фрагмента u в окошке s. Совокупность $(\varphi_s,\ s \in \mathbb{S})$ функций определяет функцию

$$\varphi : X^T \to \{0,1\},\ \ \varphi : \bar{x} \mapsto \bigwedge_{s \in \mathbb{S}} \varphi_s(x(s)), \tag{3}$$

которая для каждой разметки $\bar{x}: T \to X$ всей памяти T указывает, допустимо ли размещение разметки $\bar{x}$ в памяти T. Определенная указанным способом функция φ является условием задачи выполнимости ограничений. Сама же задача выполнимости состоит не в проверке, является ли допустимой та или иная заданная разметка, а в определении, существует ли такая допустимая разметка.

Определение 2. Задача выполнимости ограничений состоит в том, что для заданных конечных множеств X и T, структуры $\mathbb{S} \subset 2^T$ и функций $\varphi_s : X^s \to \{0,1\}$. $s \in \mathbb{S}$, требуется вычислить значение

$$\Phi = \bigvee_{\bar{x} \in X^T} \bigwedge_{s \in \mathbb{S}} \varphi_s(x(s)) \in \{0,1\} \tag{4}$$

и, если $\Phi = 1$, построить разметку $\bar{x}^* \in X^T$, такую что $\varphi_s(x^*(s)) = 1$ для каждого $s \in \mathbb{S}$.

Определение 2 является конкретизацией ранее сформулированного определения 1 в том смысле, что выражение (3) конкретизирует вид функции $\varphi : X^T \to \{0,1\}$ — условия задачи. Определение 2 охватывает обширное множество задач, которое образует NP-полный класс. Тем не менее, только некоторые модельные или, как говорят, игрушечные задачи распознавания формулируются непосредственно в терминах сформулированной классической задачи выполнимости ограничений. Для решения реальных задач требуются существенные модификации классической задачи, которые описаны далее в подразделе 2.2. Чтобы отличать классическую задачу от последующих ее модификаций, назовем ее $(\vee, \wedge)$-задачей в соответствии с операциями, которые используются в формулировке (4).

2.2. Обобщенная задача разметки. Модификации классической задачи выполнимости состоят в расширении класса функций φ_s, $s \in \mathbb{S}$, и φ, определяющих допустимость фрагментов $x(s) : s \to X$ и разметок $\bar{x}: T \to X$. В классической задаче эти функции принимают значения 0 или 1, а в рассматриваемых далее задачах это функции с иными областями значений. Наиболее естественным является следующее, рассмотренное в работах [12–14], обобщение задачи (4) на случай размытых логических функций $\vee$ и $\wedge$.

Пусть W — ограниченное полностью упорядоченное множество, например, интервал $[0,1]$, и для каждого окошка $s \in \mathbb{S}$ задана функция $\varphi_s : X^s \to W$, которая для каждого фрагмента $u \in X^s$ в окошке s указывает степень $\varphi_s(u)$ его допустимости. Совокупность φ_s, $s \in \mathbb{S}$, функций определяет функцию $\varphi : X^T \to W$, которая для каждой разметки $\bar{x} : T \to X$ указывает степень $\varphi(\bar{x})$ ее допустимости в соответствии с формулой $\varphi(\bar{x}) = \min\limits_{s \in \mathbb{S}} \varphi_s(x(s))$, где $x(s)$ — сужение разметки $\bar{x}$ на s. Задача выполнимости системы $(\varphi_s, s \in \mathbb{S})$ размытых ограничений состоит не в определении степени допустимости $\varphi(\bar{x}^*)$ заданной разметки $\bar{x}^*$, а в определении степени допустимости наиболее допустимой разметки.

Определение 3. Задача выполнимости размытых ограничений состоит в том, что для заданных конечных множеств X и T, полностью упорядоченного множества W, структуры $\mathbb{S} \subset 2^T$ и функций $\varphi_s : X^s \to W$, $s \in \mathbb{S}$, требуется вычислить значение

$$\Phi = \max_{\bar{x} \in X^T} \min_{s \in \mathbb{S}} \varphi_s(x(s)) \in W \tag{5}$$

и построить разметку $\bar{x}^* \in X^T$, такую что $\varphi_s(x^*(s)) \geq \Phi$ для каждого $s \in \mathbb{S}$.

Назовем эту задачу $(\max, \min)$-задачей разметки и отметим, что хотя она формально и обобщает классическую $(\vee, \wedge)$-задачу, это обобщение не очень существенное. Решение любой $(\max, \min)$-задачи сводится к решению определенной совокупности $(\vee, \wedge)$-задач, количество которых не превышает $\log(|\mathbb{S}|) + ord \times$ $\times \log |X|$, где ord — порядок структуры $\mathbb{S}$. Более существенным обобщением классических $(\vee, \wedge)$-задач являются следующие $(\min, +)$- и $(+, \times)$-задачи [15, 16].

Пусть W — полностью упорядоченное ограниченное снизу множество, замкнутое относительно сложения, например множество неотрицательных целых чисел. Пусть для каждого окошка $s \in \mathbb{S}$ и каждого фрагмента $u \in X^s$ указан штраф $\varphi_s(u)$ за размещение фрагмента u в окошке s. Эти штрафы определяют штраф $\varphi(\bar{x}) = \sum\limits_{s \in \mathbb{S}} \varphi_s(x(s))$ за размещение разметки $\bar{x}$ в памяти T.

Определение 4. $(\min, +)$-задача разметки состоит в том, что для заданных конечных множеств X и T, полностью упорядоченного множества W, структуры $\mathbb{S} \subset 2^T$ и функций $\varphi_s : X^s \to W$, $s \in \mathbb{S}$, требуется вычислить значение

$$\Phi = \min_{\bar{x} \in X^T} \sum_{s \in \mathbb{S}} \varphi_s(x(s)) \in W \tag{6}$$

и построить разметку $\bar{x}^* = \arg\min\limits_{\bar{x} \in X^T} \varphi(\bar{x})$.

В разд. 3 показано, как одна из задач структурного распознавания, а именно отыскание наиболее вероятной реализации скрытой части гиббсова поля сводится к решению $(\min, +)$-задачи.

Пусть W — множество чисел, для которых определены как сложение, так и умножение.

Определение 5. $(+,\times)$-задача состоит в том, что для заданных конечных множеств X и T, структуры $\mathbb{S} \subset 2^T$ и функций $\varphi_s : X^s \to W,\ s \in \mathbb{S}$, требуется вычислить число

$$\Phi = \sum_{\bar{x} \in X^T} \prod_{s \in \mathbb{S}} \varphi_s(x(s)) \in W. \quad (7)$$

Как видно, в последней задаче вообще не идет речь об отыскании какой-то одной разметки, а требуется вычислить определенную интегральную характеристику множества всех возможных разметок. В разд. 3 показано, как распознавание объектов в рамках гиббсовой статистической модели сводится к решению $(+,\times)$-задач.

Задачи (4)–(7) отличаются друг от друга лишь множеством W и двухместными операциями $\oplus$ и $\otimes$ на этом множестве. В каждой из этих задач тройка $\langle W, \oplus, \otimes \rangle$ образует конструкцию, называемую коммутативным полукольцом. Это значит, что для всех $a, b, c \in W$ выполняются равенства

$$a \oplus b = b \oplus a \ \text{ и } \ a \otimes b = b \otimes a,$$

$$a \oplus (b \oplus c) = (a \oplus b) \oplus c \ \text{ и } \ a \otimes (b \otimes c) = (a \otimes b) \otimes c,$$

$$a \otimes (b \oplus c) = (a \otimes b) \oplus (a \otimes c)$$

и существует элемент $0 \in W,$ такой что $a \oplus 0 = a$ для всех $a \in W.$ Следовательно, сформулированные четыре задачи допускают следующую общую формулировку [17].

Определение 6. Обобщенная $(\oplus, \otimes)$-задача разметки состоит в том, что для заданных конечных множеств X и $T,$ структуры $\mathbb{S} \subset 2^T,$ полукольца $\langle W, \oplus, \otimes \rangle$ и функций $\varphi_s : X^s \to W,\ s \in \mathbb{S}$, требуется вычислить значение

$$\Phi = \bigoplus_{\bar{x} \in X^T} \bigotimes_{s \in \mathbb{S}} \varphi_s(x(s))$$

и найти разметку $\bar{x}^* \in X^T,$ такую что $\bigotimes_{s \in \mathbb{S}} \varphi_s(x^*(s)) = \Phi,$ если такая существует.

Множество $(\oplus, \otimes)$-задач включает в себя не только задачи (4)–(7), но и обширное множество других задач. Заслуживает краткого упоминания еще один частный случай $(\oplus, \otimes)$-задачи, который является естественным и практически важным обобщением задач (4)–(6). Обобщение классической задачи (4) совместимости, которое можно назвать задачей d-совместимости, заключается в поиске d разметок, удовлетворяющих условию задачи, а не одной такой разметки. Обобщение задач (5) и (6), которое назовем d -оптимизацией, состоит в отыскании не одной наилучшей разметки, а подмножества $D \subset X^T,$ $|D| = d,$ разметок, наилучшего в том смысле, что любая разметка $\bar{x} \in D$ не хуже любой разметки $\bar{y} \notin D.$ В работах [18–21] показано, каким образом, выбирая подходящие для этой цели полукольца $\langle W, \oplus, \otimes \rangle$, задачи d -выполнимости и d -оптимизации сводятся к решению $(\oplus, \otimes)$-задач.

2.3. Полиномиально разрешимые классы задач разметки. Как было сказано во введении, цель данной статьи состоит в обзоре современных проблем распознавания, которые не укладываются в рамки проблематики машинного обучения, а не методов решения этих проблем. Частично эта цель уже достигается формулировкой общей $(\oplus, \otimes)$-задачи, которая формализует достаточно широкий круг проблем, включающий структурное распознавание. В разд. 3 будет показано, как задачи распознавания в определенном классе статистических моделей сводятся к формальным задачам (4)–(7), и будет продолжено движение к достижению цели статьи. Тем не менее, уместно по крайней мере упомянуть основные известные методы решения задач разметки, которые применяются в современном распознавании.

Множество всех возможных $(\oplus, \otimes)$-задач образует NP-полный класс проблем, равно как и сформулированное в подразделе 2.1 множество $(\vee, \wedge)$-задач и определенные в подразделе 2.2 множества $(\max, \min)$- , $(\min, +)$- и $(+, \times)$-задач. В силу этого неизвестны эффективные алгоритмы решения этих задач в общем виде, однако известны достаточно обширные их полиномиально разрешимые подклассы.

Возможно, наиболее известным и применяемым на практике является решение $(\oplus, \otimes)$-задачи разметки для произвольного полукольца $\langle W, \oplus, \otimes \rangle$, но для ограниченного класса структур, называемых ациклическими. Если $\mathbb{S}$ — структура второго порядка, то вместе с множеством T она образует неориентированный граф, множеством вершин которого является T, а вершины $t \in T$ и $r \in T$ соединены ребром тогда и только тогда, когда $\{t, r\} \in \mathbb{S}$. Структуру $\mathbb{S}$ множества T назовем ациклической, если пара $(T, \mathbb{S})$ образует граф, не содержащий циклов. Задачи разметки на ациклических структурах рассмотрены в работах [21, 22], в которых описан единообразный алгоритм их решения для любого заданного полукольца. Алгоритм состоит в последовательном исключении ячеек из T с помощью процедуры динамического программирования. Вычислительная сложность этого алгоритма имеет тот же порядок, что и выполнение $|T| \times |X|^2$ операций $\oplus$ и $\otimes$.

Указанный результат достаточно естественно ассимилировался в практику распознавания образов, потому что его частные случаи широко применялись в распознавании еще до его формулировки в общем виде. В частности, основная идея работ [2, 4], положивших начало структурному распознаванию, состояла в том, что распознавание текстовых изображений и речевых сигналов сводится к решению $(\min, +)$- или $(+, \times)$-задач на ациклических структурах частного вида, а именно, «одномерных» цепочках символов.

На структурах произвольного вида, в отличие от ациклических, уже не достигается единообразное решение задач для произвольных полуколец $\langle W, \oplus, \otimes \rangle$. Задачи на разных полукольцах решаются с помощью различных, хотя в чем-то и сходных алгоритмов. Полиномиально разрешимые подклассы $(\vee, \wedge)$-задач исследованы в работах по классической проблеме совместимости ограничений [23–26]. Результаты этих исследований достаточно просто переносятся на класс $(\max, \min)$-задач, потому что, как уже было сказано, решение любой $(\max, \min)$-задачи сводится к решению некоторой последовательности $(\vee, \wedge)$-задач.

 ISSN 0572-2691

Прорыв в теории и практике распознавания оказала работа [27], в которой показано, что определенные задачи сегментации изображений сводятся к так называемым субмодулярным $(\min,+)$-задачам [28], множество которых образует полиномиально разрешимый подкласс. В терминах данной статьи этот класс задач определяется следующим образом.

Пусть V — некоторое полностью упорядоченное множество и для каждой ячейки $t \in T$ определена функция $p_t : X \to V,$ называемая нумерацией, не обязательно обратимая. Утверждение $p_t(u) = \max\{p_t(x),\, p_t(y)\}$ для ячейки $t \in T,$ нумерации $p_t : X \to V$ и меток $x, y \in X$ будем выражать кратко в виде $u = (x,\, y)_t^{\uparrow},$ а утверждение $p_t(v) = \min\{p_t(x),\, p_t(y)\}$ — в виде $v = (x, y)_t^{\downarrow}.$ Для окошка $s \subset T,$ $s \in \mathbb{S}$, нумераций $p_t,$ $t \in T,$ и фрагментов $x, y, u, v \in X^s$ выражение $u = (x,\, y)_s^{\uparrow}$ обозначает, что $u(t) = (x(t),\, y(t))_t^{\uparrow}$ для всех $t \in s,$ а выражение $v = (x,\, y)_s^{\downarrow}$ — что $v(t) = (x(t),\, y(t))_t^{\downarrow}$ для всех $t \in s.$.

Определение 7. $(\min,+)$-задача разметки называется субмодулярной, если существуют такие нумерации $p_t : X \to V,$ $t \in T,$ что для каждого окошка $s \in \mathbb{S}$ и каждой пары $x, y \in X^s$ фрагментов выполняется неравенство

$$\varphi_s((x,y)_s^{\uparrow}) + \varphi_s((x,y)_s^{\downarrow}) \le \varphi_s(x) + \varphi_s(y). \tag{8}$$

Прорывное значение работы [27] состояло в формулировке определенной задачи сегментации изображений как субмодулярной $(\min,+)$-задачи на структуре второго порядка. Эта задача, в свою очередь, сводилась к хорошо исследованной к тому времени задаче о минимальном сечении графа. Этот результат получил дальнейшее развитие в работах [29–33] и необозримо обширном множестве последующих работ.

Для применения методов, основанных на сведении задачи разметки к отысканию минимального сечения в графе, необходима не только априорная гарантия субмодулярности решаемой задачи, но и знание тех конкретных нумераций $p_t : X \to V,$ $t \in T,$ при которых выполняются неравенства (8). Возможно, поэтому в практике распознавания получил широкое распространение альтернативный подход к решению $(\min,+)$-задач [34–42], основанный на теоремах двойственности в линейном программировании и известный как эквивалентное преобразование $(\min,+)$-задач, или как их репараметризация, или еще под какими-то другими, трудно переводимыми на русский язык названиями. Для реализации этого альтернативного подхода не требуется знание нумераций $p_t : X \to V,$ $t \in T,$ а достаточна (но не необходима!) гарантия их существования. Область применимости этого подхода включает все возможные субмодулярные $(\min,+)$-задачи, но не только их.

Совокупность известных в настоящее время методов решения $(\min,+)$-задач образует одно из магистральных направлений в теории и практике современного распознавания, известного как EMMCVPR — Energy Minimization Method of Computer Vision and Pattern Recognition.

Повидимому, наименее исследован класс $(+,\times)$-задач. За исключением задач на ациклических структурах, сейчас неизвестен какой-либо другой полиномиально разрешимый подкласс этих задач, представлющий практический интерес. Более того, как показано в работе [43], трудно ожидать обнаружение такого подкласса в обозримом будущем. В настоящее время, при необходимости решать $(+,\times)$-задачи в том или ином приложении используют так называемое гиббсово самплирование [44], область применимости которого недостаточно четко очерчена.

3. Гиббсовы поля и задачи разметки

Покажем, как задачи распознавания в так называемой гиббсовой статистической модели сводятся к тем или иным задачам разметки. Гиббсова модель является частным случаем экспоненциальной модели, которая формулируется на основе следующих понятий [45]. Пусть разметка $\bar{x}: T \to X$ случайна и для каждой разметки $\bar{x} \in X^T$ существует ее вероятность $p(\bar{x})$, но распределение $p: X^T \to \mathbb{R}$ вероятностей на множестве разметок определено не явно, а через постулирование следующих его свойств. Обозначим $f: X^T \to \mathbb{R}^n$ вектор-функцию, называемую базисом экспоненциальной модели, и будем считать, что известно ее математическое ожидание $\sum\limits_{\bar{x}\in X^T} p(\bar{x}) \cdot f(\bar{x}) = \theta$, называемое параметром экспоненциальной модели.

Определение 8. Случайная разметка $\bar{x}: T \to X$ с распределением вероятностей $p: X^T \to \mathbb{R}$ называется экспоненциальной в базисе f с параметром θ, если

$$p = \operatorname*{arg\,max}_{p' \in P} \left[-\sum_{\bar{x}\in X^T} p'(\bar{x}) \log p'(\bar{x}) \right], \tag{9}$$

где P — множество функций $p: X^T \to \mathbb{R}$, таких что $\sum\limits_{\bar{x}\in X^T} p(\bar{x}) \cdot f(\bar{x}) = \theta$ и $\sum\limits_{\bar{x}\in X^T} p(\bar{x}) = 1$.

Энтропию в правой части (9) будем считать определенной не только для положительных, но и для нулевых вероятностей, полагая, что $p \cdot \log p = 0$ при $p = 0$.

Теорема 1. Если условия $\sum\limits_{\bar{x}\in X^T} p(\bar{x}) \cdot f(\bar{x}) = \theta$ и $\sum\limits_{\bar{x}\in X^T} p(\bar{x}) = 1$ непротиворечивы, то существует единственное распределение $p^*: X^T \to \mathbb{R}$ вероятностей, соответствующее определению 8. При этом существуют вектор $\Lambda \in \mathbb{R}^n$ и число $c \in \mathbb{R}$, такие что $p^*(\bar{x}) = c \cdot \exp(\Lambda \cdot f(\bar{x}))$ для всех $\bar{x} \in X^T$.

Доказательство. Энтропия $\left[-\sum\limits_{\bar{x}\in X^T} p(\bar{x}) \log p(\bar{x}) \right]$ — строго вогнутая функция вероятностей $p(\bar{x})$, которая не превышает число $\log\left|X^T\right|$, а P — выпуклое непустое множество. Из этого следует первое утверждение теоремы, что максимум энтропии на множестве P достигается в единственной точке p^*.

ISSN 0572-2691

Для точки p^* существуют такие вектор $\Lambda \in \mathbb{R}^n$ и число $\lambda_0 \in \mathbb{R}$, что тройка $(p^*, \Lambda, \lambda_0)$ яляется решением системы уравнений

$$F(p, \Lambda, \lambda_0) = -\sum_{\bar{x} \in X^T} p(\bar{x}) \log p(\bar{x}) + \left(\Lambda \cdot \sum_{\bar{x} \in X^T} p(\bar{x}) \cdot f(\bar{x}) \right) + \lambda_0 \cdot \sum_{\bar{x} \in X^T} p(\bar{x});$$

$$\partial F(p, \Lambda, \lambda_0) / \partial p(\bar{x}) = 0, \; \bar{x} \in X^T.$$

Это значит, что $-1 - \log p^*(\bar{x}) + (\Lambda \cdot f(\bar{x})) + \lambda_0 = 0$ для всех $\bar{x} \in X^T$ и $p^*(\bar{x}) = $ $= c \cdot \exp(\Lambda \cdot f(\bar{x}))$, где $c = \exp(\lambda_0 - 1)$.

Теорема доказана.□□

Очевидно, что распределение вероятностей $p^* : X^T \to \mathbb{R}$ экспоненциальной случайной разметки можно представить и в виде

$$p^*(\bar{x}) = \frac{\exp(\Lambda \cdot f(\bar{x}))}{\sum_{\bar{y} \in X^T} \exp(\Lambda \cdot f(\bar{y}))}. \tag{10}$$

Известно, что при заданном базисе $f : X^T \to \mathbb{R}^n$ экспоненциальной случайной разметки максимально правдоподобное оценивание ее параметра $\theta \in \mathbb{R}^n$ выполняется исключительно простым способом. Сформулируем и докажем этот результат в терминах данной статьи.

Пусть $f^* : X^T \to \mathbb{R}^n$ — заданный базис, $p_\theta : X^T \to \mathbb{R}$ — распределение вероятностей случайной разметки, экспоненциальной в базисе f^* с параметром θ. Для заданной выборки $(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_i, \ldots \bar{x}_m)$ разметок обозначим $\theta^* = \arg\max_{\theta \in \mathbb{R}^n} \prod_{i=1}^{m} p_\theta(\bar{x}_i) =$ $= \arg\max_{\theta \in \mathbb{R}^n} \sum_{i=1}^{m} \log p_\theta(\bar{x}_i)$ оценку наибольшего правдоподобия параметра θ.

Теорема 2. Пусть $p_\theta : X^T \to \mathbb{R}$ – распределение вероятностей случайной разметки, экспоненциальной в заданном базисе $f^* : X^T \to \mathbb{R}^n$, определенное с точностью до параметра θ, а $(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_i, \ldots, \bar{x}_m)$ — выборка случайных реализаций этой разметки. В таком случае

$$\theta^* = \arg\max_{\theta \in \mathbb{R}^n} \sum_{i=1}^{m} \log p_\theta(\bar{x}_i) = \frac{1}{m} \sum_{i=1}^{m} f^*(x_i).$$

Доказательство. Обозначим $P = \{p_\theta \mid \theta \in \mathbb{R}^n\}$ — семейство экспоненциальных в базисе f^* распределений вероятностей и отметим, что искомое значение $\theta^* = \arg\max_{\theta \in \mathbb{R}^n} \sum_{i=1}^{m} \log p_\theta(\bar{x}_i)$ — параметр $\sum_{\bar{x} \in X^T} p^*(\bar{x}) \cdot f^*(\bar{x})$ распределения вероятностей $p^* = \arg\max_{p \in P} \sum_{i=1}^{m} \log p(\bar{x}_i)$. В силу (10) вероятности $p^*(\bar{x})$, $\bar{x} \in X^T$, представимы в виде

$$p^*(\bar{x}) = \frac{\exp(\Lambda^* \cdot f^*(\bar{x}))}{\sum_{\bar{y}\in X^T} \exp(\Lambda^* \cdot f^*(\bar{y}))}, \quad \text{где} \quad \Lambda^* = \arg\max_{\Lambda\in\mathbb{R}^n} \sum_{i=1}^{m} \log \frac{\exp(\Lambda \cdot f^*(\bar{x}_i))}{\sum_{\bar{y}\in X^T} \exp(\Lambda \cdot f^*(\bar{y}))}.$$

Обозначим

$$F(\Lambda) = \sum_{i=1}^{m} \log \frac{\exp(\Lambda \cdot f^*(\bar{x}_i))}{\sum_{\bar{y}\in X^T} \exp(\Lambda \cdot f^*(\bar{y}))} = \sum_{i=1}^{m} (\Lambda \cdot f^*(\bar{x}_i)) - \sum_{i=1}^{m} \log \sum_{\bar{y}\in X^T} \exp(\Lambda \cdot f^*(\bar{y})) =$$

$$= \sum_{i=1}^{m} (\Lambda \cdot f^*(\bar{x}_i)) - m \cdot \log \sum_{\bar{y}\in X^T} \exp(\Lambda \cdot f^*(\bar{y}))$$

и $\dfrac{dF(\Lambda)}{d\Lambda}$ — градиент этой функции,

$$\frac{dF(\Lambda)}{d\Lambda} = \sum_{i=1}^{m} f^*(\bar{x}_i) - m \cdot \frac{\sum_{\bar{y}\in X^T} f^*(\bar{y}) \cdot \exp(\Lambda \cdot f^*(\bar{y}))}{\sum_{\bar{y}\in X^T} \exp(\Lambda \cdot f^*(\bar{y}))}.$$

Поскольку $\Lambda^* = \arg\max_{\Lambda\in\mathbb{R}^n} F(\Lambda)$, то в точке $\Lambda = \Lambda^*$ выполняется равенство $\dfrac{dF(\Lambda)}{d\Lambda} = 0$, или, более подробно,

$$\sum_{i=1}^{m} f^*(\bar{x}_i) - m \cdot \frac{\sum_{\bar{y}\in X^T} f^*(\bar{y}) \cdot \exp(\Lambda^* \cdot f^*(\bar{y}))}{\sum_{\bar{y}\in X^T} \exp(\Lambda^* \cdot f^*(\bar{y}))} =$$

$$= \sum_{i=1}^{m} f^*(\bar{x}_i) - m \cdot \sum_{\bar{y}\in X^T} f^*(\bar{y}) \cdot p^*(\bar{y}) = \sum_{i=1}^{m} f^*(\bar{x}_i) - m \cdot \theta^* = 0.$$

Теорема доказана.

Определим понятие гиббсовой случайной разметки. Пусть $\mathbb{S}$ — структура и для каждого окошка $s \in \mathbb{S}$ и каждого фрагмента $u \in X^s$ известна вероятность $p_s(u)$ появления этого фрагмента в этом окошке. Это значит, что функция $p: X^T \to \mathbb{R}$ является решением системы линейных уравнений

$$\sum_{\bar{x}\in X^T} p(\bar{x}) = 1, \quad \sum_{\bar{x}\in X(s,u)} p(\bar{x}) = p_s(u), \; s \in \mathbb{S}, \; u \in X^s, \qquad (11)$$

где $X(s,u)$ — множество всех разметок, содержащих фрагмент u в окошке s.

Определение 9. Случайная разметка $\bar{x}: T \to X$ с распределением вероятностей $p: X^T \to \mathbb{R}$ называется гиббсовым случайным полем на структуре $\mathbb{S}$ с параметрами $p_s(u)$, $s \in \mathbb{S}$, $u \in X^s$, если

 ISSN 0572-2691

$$p = \arg\max_{p' \in P} \left[-\sum_{\bar{x} \in X^T} p'(\bar{x}) \log p'(\bar{x}) \right],$$

где P — множество решений системы уравнений (11).

Как видно, гиббсово случайное поле — это частный случай экспоненциальной случайной разметки, в силу чего справедлива теорема, аналогичная теореме 1.

Теорема 3. Если разметка $\bar{x}: T \to X$ — случайное гиббсово поле на структуре $\mathbb{S}$, то существуют такие числа $g_s(u)$, $s \in \mathbb{S}$, $u \in X^s$, что

$$p(\bar{x}) = \prod_{s \in \mathbb{S}} g_s(x(s)) \tag{12}$$

для всех $\bar{x} \in X^T$.

Сомножители $g_s(u)$ для $s \in \mathbb{S}, u \in X^s$ в формуле (12) зависят от соответствующих параметров $p_s(u)$ гиббсового поля, но вычисление значений $g_s(u)$ по заданным значениям $p_s(u)$ является далеко не тривиальной проблемой. Если же сомножители $g_s(u)$ для конкретной гиббсовой модели тем или иным способом получены, то различные задачи распознавания имеют один и тот же формат. А именно, распознавание состоит в том, что по наблюдаемому фрагменту $x_1: T_1 \to X$ случайной разметки $\bar{x}: T \to X$ в окошке $T_1 \subset T$ следует принять разумное в определенном смысле решение о скрытом фрагменте $x_2: T_2 \to X$ этой же разметки в окошке $T_2 \subset T$. При этом не обязательно $T_1 \in \mathbb{S}$, $T_2 \in \mathbb{S}$, $T_1 \cup T_2 = T$.

Покажем одну из возможных интерпретаций введенных абстрактных понятий. Будем понимать ячейки $t \in T$ как элементарные объекты, из которых состоит составной объект T. В частности, t — это пиксель в поле зрения T, на котором представлено изображение. Каждый элементарный объект $t \in T$ характеризуется сигналом $x(t)$, который он излучает, и скрытым состоянием $y(t)$. Сигналы принимают значения из конечного множества X, а состояния — из конечного множества Y. Разметка в данном случае — это функция $T \to X \times Y$, т.е. пара $(\bar{x}, \bar{y})$, где $\bar{x}$ — наблюдаемая часть разметки, а $\bar{y}$ — ее скрытая часть. Пусть скрытая разметка $\bar{y}: T \to Y$ — это гиббсово случайное поле на заданной структуре $\mathbb{S}$ и для каждой пары $s \in \mathbb{S}$ и $u \in Y^s$ известны сомножители $g_s(u)$, позволяющие вычислить вероятность $p_Y(\bar{y})$ каждой скрытой разметки $\bar{y} \in Y^T$ по формуле $p_Y(\bar{y}) = \prod_{s \in \mathbb{S}} g_s(y(s))$. Пусть $\bar{x}: T \to X$ — случайное наблюдение, зависящее от скрытой разметки $\bar{y}: T \to Y$, и эта зависимость имеет простейший вид $q(\bar{x} \mid \bar{y}) = \prod_{t \in T} q_t(x(t) \mid y(t))$. Если числа $g_s(u)$ заданы для каждой пары $s \in \mathbb{S}$, $u \in Y^s$, а условные вероятности $q_t(x \mid y)$ — для каждой тройки $t \in T$, $x \in X$, $y \in Y$, то вероятность пары $(\bar{x}, \bar{y})$ вычисляется по явной формуле

$$p(\bar{x}, \bar{y}) = \prod_{s \in \mathbb{S}} g_s(y(s)) \times \prod_{t \in T} q_t(x(t) \mid y(t)). \qquad (13)$$

Однако распознавание заключается не в вычислении вероятностей $p(\bar{x}, \bar{y})$ по формуле (13), а в решении существенно более сложных задач. Простейшая из них состоит в классификации множества X^T наблюдений на те, которые могут с ненулевой вероятностью появиться на выходе составного объекта, и все прочие. Это значит, что для заданной наблюдаемой части $\bar{x}^*$ разметки следует определить, существует ли такая скрытая ее часть $\bar{y}^*$, для которой выполняется условие $p(\bar{x}^*, \bar{y}^*) > 0.$ Ответ на этот вопрос сводится к решению $(\vee, \wedge)$-задачи разметки

$$\Phi(\bar{x}^*) = \bigvee_{\bar{y} \in Y^T} \left[\bigwedge_{s \in \mathbb{S}} g'_s(y(s)) \wedge \bigwedge_{t \in T} q'_t(x^*(t) \mid y(t)) \right] \qquad (14)$$

для бинарных переменных $g'_s(u)$ и $q'_t(x \mid y)$, которые принимают значения 0 или 1 в зависимости от того, являются ли сомножители $g_s(u)$ и $q_t(x \mid y)$ нулевыми или положительными. Усиление этой задачи состоит в отыскании скрытой разметки

$$\bar{y}^* = \operatorname*{argmax}_{\bar{y} \in Y^T} \prod_{s \in \mathbb{S}} g_s(y(s)) \times \prod_{t \in T} q_t(x^*(t) \mid y(t)) \qquad (15)$$

с наибольшей апостериорной вероятностью при условии наблюдения $\bar{x}^*$. Как видно, это частный случай $(\oplus, \otimes)$-задачи, сформулированный для полукольца $(\max, \times)$, который вполне очевидным образом сводится к $(\min, +)$-задаче.

Введем в рассмотрение дополнительный параметр k — состояние составного объекта в целом, которое принимает значения из множества K, состоящего из малого количества элементов, например $K = \{A, B\}$. Состояние k является случайным, и для каждого $k \in K$ известна априорная вероятность $\alpha(k)$ пребывания составного объекта в этом состоянии. Разметка $(\bar{x}, \bar{y})$ — случайное гиббсовое поле, зависящее от состояния k, так что

$$p^k(\bar{x}, \bar{y}) = \prod_{s \in \mathbb{S}} g_s^k(y(s)) \times \prod_{t \in T} q_t^k(x(t) \mid y(t)), \; k \in K, \; \bar{x} \in X^T, \; \bar{y} \in Y^T,$$

а совместная вероятность тройки $(k, \bar{x}, \bar{y})$ выражается явно как

$$p(k, \bar{x}, \bar{y}) = \alpha(k) \times \prod_{s \in \mathbb{S}} g_s^k(y(s)) \times \prod_{t \in T} q_t^k(x(t) \mid y(t)), \; k \in K, \; \bar{x} \in X^T, \; \bar{y} \in Y^T.$$

Задача распознавания состоит не в вычислении этой вероятности, а в том, что по наблюдению $\bar{x}^*$ следует принять решение о состоянии k составного объекта. Для этого необходимо располагать вероятностями

$$p(k, \bar{x}^*) = \alpha(k) \times \sum_{\bar{y} \in Y^T} \prod_{s \in \mathbb{S}} g_s^k(y(s)) \times \prod_{t \in T} q_t^k(x^*(t) \mid y(t)), \; k \in K, \qquad (16)$$

т.е. решить $(+, \times)$-задачу разметки для каждого состояния $k \in K$.

Приведенные примеры иллюстрируют те трудные вопросы распознавания, которые лежат вне проблематики машинного обучения. Статистическая модель распознаваемого объекта в приведенных примерах полностью и однозначно определена, в силу чего отсутствует необходимость в какой-либо дополнительной обучающей информации для ее уточнения. Фундаментальная сложность рассмотренных задач состоит не в том, что какие-то функции или множества не полностью известны, а в том, что они трудновычислимы в том смысле, как это указано в конце разд. 1. Они не принадлежат тем классам множеств и функций, на которых строится классическое распознавание. Распознаваемый объект — это не точка в конечномерном линейном пространстве, множество объектов — это не полупространство, не сфера, не выпуклое множество, а функции в правых частях выражений (14)–(16) — это не линейные, не квадратичные, не пороговые функции. Это совсем другие функции и множества.

Представленная схема решения задач распознавания совсем не похожа на те палочки-выручалочки, которые рекламируются на разнообразных бандвагонах. Даже если ограничиться рассмотрением гиббсовых моделей и считать, что статистическая модель объекта однозначно задана структурой $\mathbb{S}$ и вероятностями $p_s(u)$, то для построения алгоритма распознавания необходимо решить две серьезные проблемы: вычислить сомножители $g_s(u)$ на основании известных вероятностей $p_s(u)$; решить ту или иную задачу разметки. Эти проблемы выходят за рамки современной проблематики машинного обучения, и общие методы их решения неизвестны. Их должен решать разработчик на основании знания как математического, так и прикладного содержания конкретной решаемой задачи.

Рассмотрим все же ситуацию, когда для заданной структуры $\mathbb{S}$ гиббсово поле определено не полностью, а с точностью до массива $\theta = (p_s(u) \mid s \in \mathbb{S}, u \in X^s)$ вероятностей, и эта неполнота знаний восполняется обучающей выборкой $(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n)$ разметок. Один из базовых известных методов обучения состоит в наиболее правдоподобном оценивании вероятностей $p_s(u), s \in \mathbb{S}, u \in X^s$,

$$\theta^* = (p_s^*(u) \mid s \in \mathbb{S}, u \in X^s) = \arg\max_{\theta} \prod_{i=1}^{n} p_\theta(\bar{x}_i), \tag{17}$$

и последующем применении полученных оценок так, будто они равны действительным вероятностям. В рамках рассматриваемых гиббсовых моделей эти два этапа совершенно неравноценны. После того, как для имеющейся обучающей выборки уже получены оценки (17), необходимо еще преобразовать вероятности $p_s^*(u)$ в сомножители $g_s^*(u)$ и решить ту или иную задачу разметки, например задачу (14), (15) или (16). Каждая из этих задач далеко не тривиальна. Решение же оптимизационной задачи (17), т.е. собственно обучения, не представляет особого труда, так как достигается следующим простым способом.

Для каждого окошка $s \in \mathbb{S}$ и каждого фрагмента $u \in X^s$ обозначим $m_s(u)$ количество разметок в выборке $(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_m)$, сужение которых на s равно u.

Теорема 4. Пусть $\mathbb{S}$ — структура, $\theta = (p_s(u) \mid s \in \mathbb{S}, u \in X^s)$ — массив вероятностей, а $p_\theta : X^T \to \mathbb{R}$ — функция, обеспечивающая максимум энтропии $\left[-\sum_{\bar{x} \in X^T} p(\bar{x}) \log p(\bar{x}) \right]$ на множестве функций $p : X^T \to \mathbb{R}$, для которых

$$\sum_{\bar{x} \in X(s,u)} p(\bar{x}) = p_s(u), \; s \in \mathbb{S}, \; u \in X^s, \text{ и } \sum_{\bar{x} \in X^T} p(\bar{x}) = 1.$$

Пусть $(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_m)$ — обучающая выборка. В таком случае

$$\theta^* = \arg\max_\theta \prod_{i=1}^{m} p_\theta(\bar{x}_i) = \left(\frac{m_s(u)}{m} \mid s \in \mathbb{S}, u \in X^s \right). \tag{18}$$

Теорема является частным случаем теоремы 2.

Таким образом, насколько малопродуманной является расхожая фраза типа «Если статистическая модель распознаваемого объекта тем или иным образом определена, то дальше уже делать нечего». Трудности, связанные с неполным знанием гиббсовой модели объекта и вытекающей отсюда необходимостью обучения, исчезающе малы по сравнению с трудностями, с которыми нужно справиться, даже если эта модель полностью известна. Уточнение модели на основании обучающей выборки состоит в простейших вычислениях по формуле (18), а распознавание при полностью известной модели — в решении задач (14), (15) или (16) фундаментальной сложности.

4. Распознавание и обучение

Рассмотрим взаимосвязь обучения и распознавания в рамках статистической модели, более общей, чем рассмотренная в разд. 3 гиббсова модель. Обозначим x доступный для наблюдения сигнал, который излучает распознаваемый объект, y — скрытое состояние объекта, θ — параметр, называемый далее моделью и определяющий, как сигнал и состояние зависят друг от друга. Обозначим X, Y и Θ конечные множества значений, которые принимают сигнал, состояние и модель. Пара $(x, y) \in X \times Y$ является случайной, и для каждой пары значений $x^* \in X$, $y^* \in Y$ существует их совместная вероятность $p^*_{XY}(x^*, y^*)$. Распределение вероятностей $p^*_{XY} : X \times Y \to \mathbb{R}$ неизвестно, потому что зависит от модели θ, которая является фиксированной, но неизвестной характеристикой объекта. Известно, однако, множество Θ моделей, которому эта фиксированая модель принадлежит. Таким образом, знания о распознаваемом объекте выражаются функцией $p_{XY} : X \times Y \times \Theta \to \mathbb{R}$ трех переменных, значение которой $p_{XY}(x, y; \theta)$ обозначает совместную вероятность сигнала x и состояния y в модели θ. Пусть $z = (x_1, y_1; x_2, y_2; \ldots x_i, y_i; \ldots x_n, y_n)$ — случайная обучающая выборка, а $Z = (X \times Y)^n$ — множество всех возможных обучающих выборок длины n. Случайная выборка z зависит от модели θ, например, так, что $p_Z(z; \theta) = \prod_{i=1}^{n} p_{XY}(x_i, y_i; \theta)$.

 ISSN 0572-2691

Обозначим $q_{XZ} : X \times Z \to Y$ функцию, называемую стратегией, а Q_{XZ} — множество всех возможных функций вида $X \times Z \to Y$. Значение $q_{XZ}(x_0, z) = y_0$ этой функции обозначает, что на основании обучающей выборки z и текущего сигнала x_0, излучаемого объектом, принимается решение y_0 о текущем состоянии объекта. Качество $R(q_{XZ}, \theta)$ этой стратегии, называемое риском, определяется как математическое ожидание потерь $w(y, y')$, имеющих место при решении y' для объекта, находящегося в состоянии y,

$$R(q_{XZ}, \theta) = \sum_{z \in Z} p_Z(z; \theta) \sum_{x \in X} \sum_{y \in Y} p_{XY}(x, y; \theta)\, w(y, q_{XZ}(x, z)).$$

Обозначим $q_X : X \to Y$ стратегию принятия решений о текущем состоянии объекта на основании наблюдения только сигнала, и Q_X — множество таких стратегий. Отметим, что для любой модели $\theta \in \Theta$ выполняется равенство

$$\min_{q_{XZ} \in Q_{XZ}} R(q_{XZ}, \theta) = \min_{q_X \in Q_X} R(q_X, \theta). \tag{19}$$

Равенство (19) выражает тот очевидный факт, что если модель известна, то отпадает небходимость в какой-либо обучающей выборке.

Определение 10. Исходными данными задачи обучаемого распознавания являются множества X, Y, Θ, Z и функции $p_{XY} : X \times Y \times \Theta \to \mathbb{R}$, $p_Z : Z \times \Theta \to \mathbb{R}$, $w : Y \times Y \to \mathbb{R}$, а ее решением — стратегия $q_{XZ} : X \times Z \to Y$.

Формат процедуры, решающей ту или иную задачу обучаемого распознавания, обозначим

$$\langle X, Y, \Theta, Z, p_{XY}, p_Z, w \rangle \mapsto q_{XZ}. \tag{20}$$

Определение 10, равно как и формат (20), является не математической формулировкой задачи, а лишь ее спецификацией, перечислением данных, о которых идет речь в задаче. В определении 10 и формате (20) отсутствует то, что в разд. 1 названо условием задачи. Для того чтобы определение 10 стало формулировкой задачи, его следовало бы дополнить требованиями к искомой стратегии, которые позволяли бы обосновать или опровергать стратегии, предположительно решающие задачу, и выводить искомую стратегию из исходных данных. Тем не менее, в настоящее время известны многие сотни процедур в формате (20), каждая из которых следует из того или иного интуитивно понимаемого, но не сформулированного явно требования к искомой стратегии. Все эти процедуры образуют определенный подкласс процедур формата (20), специфика которого состоит в представлении процедуры (20) в виде двух этапов. На первом этапе на основании исходных данных и обучающей выборки z вычисляется оценка θ^* модели, т.е. реализуется процедура формата

$$\langle X, Y, \Theta, Z, p_{XY}, p_Z, z \rangle \mapsto \theta^*. \tag{21}$$

На втором этапе реализуется процедура формата

$$\langle X, Y, p_{XY}, w, \theta^* \rangle \mapsto q_X^*, \tag{22}$$

которая строит стратегию $q_X^* : X \to Y$ распознавания так, будто модель θ для распознаваемого объекта известна и равна модели θ^*, полученной на первом этапе. Процедуры обучения указанного частного формата приняты, как нечто само собою разумеющееся, и поэтому назовем их каноническими.

Указанный формат известных процедур обучения позволяет более сдержанно оценить наиболее восторженную часть мифов об обучаемом распознавании, что обучаемая система строит стратегию распознавания, даже если разработчик системы ее не знает. Разработчик действительно не знает стратегию, пригодную в данной конкретной прикладной ситуации, потому что он не знает статистическую модель объекта, который ему придется распознавать. Но для применения известных процедур обучения он должен знать значительно больше, а именно, знать стратегию распознавания для каждой модели из заданного класса, чтобы ее выполнить на втором этапе (22) процедуры обучения. Обучение не является средством решения задач, алгоритм решения которых неизвестен, а средством, которое лишь определяет, какая именно задача из заданного класса подлежит решению, с последующим применением для ее решения известного алгоритма, общего для этого класса задач.

В настоящее время, каждый из двух этапов (21) и (22) в отдельности хорошо формализован и образует область обширных математических исследований. Требования к этапу (22) естественным образом формулируются как минимизация байесовского риска,

$$q_X^* = \arg\min_{q_X \in Q_X} R(q_X, \theta^*) = \arg\min_{q_X \in Q_X} \sum_{x \in X} \sum_{y \in Y} p_{XY}(x, y; \theta) \cdot w(y, q_{XZ}(x, z)).$$

Одной из двух наиболее известных реализаций этапа (21) является максимально правдоподобное оценивание модели,

$$\theta^* = \arg\max_{\theta \in \Theta} \log p_Z(z, \theta) = \arg\max_{\theta \in \Theta} \sum_{i=1}^{n} \log p_{XK}(x_i, k_i; \theta).$$

Другая реализация этапа (21), известная как минимизация эмпирического риска, заключается в вычислении

$$\theta^* = \arg\min_{\theta \in \Theta} \sum_{i=1}^{n} w(x_i, q^*(\theta)(x_i)) \text{ при } q^*(\theta) = \arg\min_{q_X \in Q_X} R(q_X, \theta). \tag{23}$$

Оба метода, как и другие, менее распространенные реализации этапа (21), состоятельны в том смысле, что при неограниченном росте длины обучающей выборки риск стратегии, получаемой на втором этапе, становится неотличимым от риска, который достигался бы при известной модели. Естественно, что при ограниченной длине выборки и уменьшении ее длины этот риск увеличивается. Менее естественным и недостаточно исследованным является так называемый эффект коротких выборок, который проявляется при обработке выборок достаточно малой длины. Заключается этот эффект в том, что риск статегии, полученной для такой выборки с помощью канонической процедуры обучения, может оказаться больше, чем риск стратегии, которая такую выборку не использует вообще. Иными словами, вопреки тому, что любая выборка зависит от модели объекта и, следовательно, содержит в себе информацию, пусть и малую, о требуемой стратегии распознавния, в определенных ситуациях эту информацию лучше игнорировать, чем использовать ее с помощью канонических процедур обучения. Примеры, иллюстрирующие этот эффект, приведены в [46].

Эффект коротких выборок свидетельствует о существенном изъяне канонических процедур обучения, который следует из того, что они приняты как само собой разумеющиеся, а не выведены из каких-либо разумных требований к риску стратегии, получаемой в результате обучения. Однако окончательное решение о пригодности той или иной процедуры обучения принимается, как правило, на основаниии качества распознавания, которое достигается после обучения. Таким образом, получается, что алгоритмы обучения разрабатываются на основании одних требований, а проверяются на соответствие совсем другим требованиям. В разд. 5 сформулированы риск-ориентированные требования к обучаемому распознаванию и показано, какого вида стратегии следуют из этих требований.

5. Риск-ориентированные стратегии

Определим понятие обучающей информации как обобщение обучающей выборки, рассматриваемой в предыдущих разделах. Пусть z — случайная величина, принимающая значения из конечного множества Z и зависящая от модели θ, и для каждой пары $z \in Z$, $\theta \in \Theta$ известна вероятность $p_Z(z; \theta)$. Таким образом, знание об источнике обучающей информации z представляется тройкой $\langle Z, \Theta, p_Z : Z \times \Theta \to \mathbb{R} \rangle$, а знание о распознаваемом объекте, как и прежде, — четверкой $\langle X, Y, \Theta, p_{XY} : X \times Y \times \Theta \to \mathbb{R} \rangle$.

Определение 11. Случайная величина $\langle Z, \Theta, p_Z : Z \times \Theta \to \mathbb{R} \rangle$ называется обучающей информацией для распознаваемого объекта $\langle X, Y, \Theta, p_{XY} : X \times Y \times \Theta \to \mathbb{R} \rangle$, если для любых $x \in X$, $y \in Y$, $z \in Z$, $\theta \in \Theta$ выполняется равенство $p_{XYZ}(x, y, z; \theta) = \; = p_Z(z; \theta) \cdot p_{XY}(x, y; \theta)$.

Обучающая выборка $z = (x_1, y_1; \dots x_n, y_n)$ длины n является частным случаем обучающей информации при $Z = (X \times Y)^n$ и $p_Z(z; \theta) = \prod_{i=1}^{n} p_{XY}(x_i, y_i; \theta)$. Другим примером обучающей информации является обучающая выборка $z = (x_1, \dots x_n)$, используемая при обучении без учителя. В этом случае $Z = X^n$ и $p_Z(z; \theta) = \prod_{i=1}^{n} \sum_{y \in Y} p_{XY}(x_i, y; \theta)$. Понятие обучающей информации, равно как и описанные далее методы ее использования, охватывает обучающие выборки любой длины, вплоть до нулевой, а не только сколь-угодно длинные. Обучающая информация может быть не только обучающей выборкой, а иметь самую разнообразную природу. Это может быть, например, заключение эксперта о модели или даже отсутствие какой-либо обучающей информации.

Обобщим понятие стратегии на случай так называемых рандомизированных стратегий, которые представляются функцией $q_{XZ} : Y \times X \times Z \to \mathbb{R}$ трех переменных, значение $q_{XZ}(y \mid x^*, z^*)$ которой имеет следующий смысл. На основании обучающей информации z^* и наблюдения текущего сигнала x^* распознающая система генерирует случайное решение y^* о текущем состоянии объекта в соответствии с распределением условных вероятностей $q_{XZ}(y \mid x^*, z^*)$. Приведем в соответствие с этими обобщенными понятиями ранее сформулированное определение 10 формата задачи обучаемого распознавания. Пусть $w: Y \times Y \to \mathbb{R}$ — функция потерь.

Определение 12. Исходными данными задачи обучаемого распознавания являются множества X, Y, Θ, Z и функции $p_{XY}: X \times Y \times \Theta \to \mathbb{R}$, $p_Z: Z \times \Theta \to \mathbb{R}$, $w: Y \times Y \to \mathbb{R}$, а ее решением — стратегия $q_{XZ}: Y \times X \times Z \to \mathbb{R}$.

Математическое ожидание потерь при применении стратегии q_{XZ} для распознавания объектов в модели θ есть риск

$$R(q_{XZ}, \theta) = \sum_{z \in Z} p_Z(z; \theta) \sum_{x \in X} \sum_{y \in Y} p_{XY}(x, y; \theta) \sum_{y' \in Y} q_{XZ}(y' \mid x, z)\, w(y, y'). \qquad (24)$$

Как видно, риск стратегии зависит не только от самой стратегии, но и от модели распознаваемого объекта. Стратегия сама по себе, рассматриваемая отдельно от модели, характеризуется не одним числом-риском, а функцией риска, определенной на множестве Θ моделей, т.е. $|\Theta|$ числами. Поэтому в общем случае невозможно отдать предпочтение какой-то одной из двух заданных стратегий: q^1_{XZ} и q^2_{XZ}. Однако стратегию q^1_{XZ} можно безусловно предпочесть стратегии q^2_{XZ}, если $R(q^1_{XZ}, \theta) < R(q^2_{XZ}, \theta)$ для всех $\theta \in \Theta$.

Определение 13. Стратегия q^0_{XZ} называется неприемлемой, если существует стратегия q^*_{XZ}, такая что $R(q^*_{XZ}, \theta) < R(q^0_{XZ}, \theta)$ для всех $\theta \in \Theta$.

Для того чтобы сформулировать разумные требования к процедурам (20) обучаемого распознавания, прежде всего нужно исключить из рассмотрения все неприемлемые стратегии и вывести общую форму всех оставшихся. Обозначим T множество всех неотрицательно определенных функций $\tau: \Theta \to \mathbb{R}$, таких что $\sum_{\theta \in \Theta} \tau(\theta) = 1$. Назовем такие функции весовыми.

Определение 14. Стратегия q^*_{XZ} называется байесовской, если существует такая весовая функция $\tau \in T$, что $q^*_{XZ} = \operatorname*{arg\,min}_{q_{XZ} \in Q_{XZ}} \sum_{\theta \in \Theta} \tau(\theta) R(q_{XZ}, \theta)$.

Теорема 5. Любая стратегия $q_{XZ} \in Q_{XZ}$ является либо байесовской, либо неприемлемой, и не существует стратегии, которая была бы байесовской и неприемлемой одновременно.

Теорема доказана в работе [46], и ее доказательство основано на известной лемме Неймана о седловой точке.

Теорема 5 позволяет уточнить определение 12 формата задачи обучаемого распознавания, требуя, чтобы решением задачи была не любая стратегия $q_{XZ}: Y \times X \times Z \to \mathbb{R}$, а только приемлемая. Множество приемлемых стратегий слишком обширно, и поэтому здесь необходимы дальнейшие конкретизации этого формата, примеры которых приведены в [46]. Однако из теоремы 5 следуют определенные важные выводы, даже если не прибегать к таким конкретизациям.

Из теоремы 5 следует, что любая приемлемая стратегия обучаемого распознавания должна иметь вид

$$q^*_{XZ} = \arg\min_{q_{XZ} \in Q_{XZ}} \sum_{\theta\in\Theta} \tau(\theta) R(q_{XZ}, \theta)$$

при некоторых весах $\tau(\theta)$, не зависящих от обучающей информации, а не вид канонической процедуры, которая строит стратегию $q^*_X = \arg\min_{q_X \in Q_X} R(q_X, \theta^*)$ для некоторой модели θ^*, зависящей от обучающей информации. В частном случае, если штрафы определены как

$$w(k, k') = 0 \text{ при } k = k' \text{ и } w(k, k') = 1 \text{ при } k \neq k',$$

то для текущего сигнала x_0 и обучающей информации z решение y^*_0 о текущем состоянии объекта должно равняться

$$y^*_0 = \arg\max_{y_0 \in Y} \sum_{\theta\in\Theta} \tau(\theta) \cdot p_Z(z; \theta) \cdot p_{XY}(x_0, y_0; \theta). \quad (25)$$

Если к тому же обучающая информация — это выборка $z = (x_1, y_1; \ldots x_n, y_n)$, такая что $p_Z(z; \theta) = \prod_{i=1}^{n} p_{XY}(x_i, y_i; \theta)$, то это решение должно быть

$$y^*_0 = \arg\max_{y_0 \in Y} \sum_{\theta\in\Theta} \tau(\theta) \prod_{i=0}^{n} p_{XY}(x_i, y_i; \theta),$$

$$\text{а не } y^*_0 = \arg\max_{y_0 \in Y} p_{XY}(x_0, y_0; \theta^*), \text{ где } \theta^* = \arg\max_{\theta\in\Theta} \prod_{i=1}^{n} p_{XY}(x_i, y_i; \theta),$$

которое принимает каноническая процедура, основанная на максимально правдоюподобном оценивании модели.

Теорема 5 вносит определенную ясность в нередкие дискуссии о том, как следует реализовать первый этап (21) канонических процедур обучения: на основе наиболее правдоподобного оценивания модели, или минимизации эмпирического риска (23), или на основе каких-то иных, ранее не известных подходов. Теорема 5 дает на эти вопросы довольно радикальный ответ. Если исходить из того, что результатом обучения является стратегия распознавания, качество которой представляет функция риска (24), то на основании обучающей информации z не следует выбирать никакую модель, а принимать решение о текущем состоянии объекта с учетом всех моделей заданного класса. Как видно из формулы (25), все модели влияют на это решение, а влияние каждой отдельной модели определяется двумя сомножителями. Сомножитель $p_Z(z; \theta)$ характеризует имеющийся источник обучающей информации и не зависит от каких-либо требований к качеству принимаемого решения, а сомножитель $\tau(\theta)$ не зависит от имеющейся обучающей информации.

Понятие риск-ориентированных стратегий обучаемого распознавания вносит ясность в проблему коротких выборок, в результате чего эта проблема перестает иметь самостоятельное значение. Фактически проблема коротких выборок состоит в отказе от распознавания, если имеющаяся в распоряжении обучающая выборка не обеспечивает требуемое качество последующего распознавания. Для краткости назовем такие выборки плохими. Однако не все короткие выборки плохие, более того, любая обучающая выборка может оказаться хорошей для распознавания одних сигналов и плохой для распознавания других. Поэтому вопрос о пригодности предъявленной обучающей выборки z для построения достаточно хорошей стратегии распознавания следует формулировать более конкретно: пригодны ли выборка z и текущий сигнала x_0 для принятия достаточно хорошего решения y_0 о текущем состоянии объекта. Вопрос в такой более конкретной форме является не более чем упражнением по элементарному курсу статистического распознавания. Ответ на этот вопрос достигается тем, что множество Y' решений о состоянии распознаваемого объекта принимается равным не множеству Y состояний, а равным $Y' = Y \cup \{\#\}$, где $\#$ обозначает отказ от принятия решения. Функция потерь в этом случае имеет формат $w: Y \times (Y \cup \{\#\}) \to \mathbb{R}$.

Обеспечивая ясность по указанным вопросам, определение 12 порождает новые, более глубокие вопросы фундаментального характера. Определение 12 допускает два различные понимания взаимосвязи обучаемого распознавания в формате

$$\langle X, Y, \Theta, Z, p_{XY}, p_Z, w \rangle \mapsto q_{XZ} \tag{26}$$

и распознавания без обучения в формате

$$\langle X, Y, \Theta, p_{XY}, w \rangle \mapsto q_X. \tag{27}$$

Очевидно, что формат (27) можно понимать как экзотический, предельный частный случай формата (26) при $|Z| = 1$. Справедливо и обратное, менее очевидное утверждение, что формат $\langle X', Y, \Theta, Z, p'_{X'Y}, p'_Z, w \rangle \mapsto q_{X'Z}$ обучаемого распознавания можно представить в формате $\langle X, Y, \Theta, p_{XY}, w \rangle \mapsto q_X$ распознавания без обучения. Для этого в качестве множества X следует принять множество пар (x', z), $x' \in X'$, $z \in Z$, а в качестве распределения вероятностей $p_{XY}: X \times Y \times \Theta \to \mathbb{R}$ — функцию, принимающую значения $p_{XY}(x, y; \theta) = p'_{X'Y}(x', y; \theta) \cdot p'_Z(z; \theta)$. Иными словами, любая задача распознавания без обучения состоит в принятии решения о состоянии объекта на основании каких-то исходных данных x, но в некоторых задачах эти исходные данные состоят из двух частей, $x = (x^*, z^*)$. Компонента x^*, как и данные x в целом, зависит от состояния y и модели θ, а компонента z^* непосредственно зависит только от модели и при фиксированной модели не зависит от состояния, т.е. $p(z^* \mid y; \theta) = p(z^*; \theta)$. Множество этих задач и образует множество задач формата (26).

Это значит, что в рамках теории статистических решений проблема машинного обучения теряет свое самостоятельное значение, потому что поглощается более общей проблемой распознавания без обучения. Этот, скажем прямо, парадоксальный вывод порождает вопрос фундаментального характера: адекватна ли статистическая теория решений базовым понятиям распознавания образов и машинного обучения, коль скоро она приводит к такому парадоксальному выводу. Для вопроса такого класса противопоказаны скоропалительные ответы. Для ответа на него надо будет либо постепенно привыкать к полученному парадоксальному выводу, либо подвергать тщательной ревизии уже сформировавшиеся основы

 ISSN 0572-2691

науки о распознавании. В любой науке на определенных этапах ее развития возникают подобные ситуации, и справляться с ними нужно, следуя известной мудрой (и парадоксальной!) рекомендации: смело идти с теми, кто ищет истину, и бежать без оглядки от тех, кто эту истину уже нашел.

Заключение

За последние несколько лет наука о машинном обучении заняла выдающееся место как в популярной, так и в научной литературе, обрастая все более заманчивыми ожиданиями. Ученые различных специальностей, привлеченные открывшимися перед ними заманчивыми перспективами, связывают возможность решения своих частных задач на основе идей машинного обучения. В результате всего этого значение машинного обучения было, вероятно, преувеличено и расширено до пределов, превышающих его действительные возможности.

Сложившиеся веками стандарты научного исследования служат надежным барьером для подобных преувеличений. Современное распознавание образов, рассматриваемое как научная дисциплина в системе компьютерных наук, должно удовлетворять этим требованиям. Подобно тому, как наука о случайных процессах не должна быть случайной, а наука о размытых множествах — размытой, наука об образном мышлении должна опираться не на образные метафоры, а на точные определения, из которых выводятся однозначно понимаемые утверждения. Представленное в статье исследование, удовлетворяющее этим методологическим требованиям, приводит к следующему выводу о взаимосвязи проблематик распознавания образов, машинного обучения и машинного мышления.

Существует обширный класс задач распознавания, выходящих за рамки сложившейся на сегодня проблематики машинного обучения. В частности, это задачи, известные под общим названием «структурное распознавание». Сложившееся в последние годы ходячее представление, что проблема распознавания полностью исчерпывается проблемами обучения, непозволительно сужает научную проблематику распознавания. Как структурное распознавание, так и машинное обучение формализуют мыслительные процессы определенных подклассов, но это различные подклассы. Вместе с тем, между структурным распознаванием и машинным обучением нет никакого барьера, который препятствовал бы их совместному использованию как отдельных звеньев единой технологической цепочки.

В.І. Гриценко, М.І. Шлезінгер

ВЗАЄМОЗВ'ЯЗОК ПРОБЛЕМ РОЗПІЗНАВАННЯ ОБРАЗІВ, МАШИННОГО МИСЛЕННЯ І НАВЧАННЯ

Виконано огляд досліджень з структурного розпізнавання — одного з напрямків сучасного розпізнавання образів. Показано, що основна проблематика структурного розпізнавання виходить за рамки проблем машинного навчання, і таким чином вноситься певна стриманість у поширене уявлення, що проблеми розпізнавання повністю поглинаються проблемами навчання. Разом з цим показано, що базові поняття і задачі структурного розпізнавання творять основу для формалізації певного класу процесів мислення, що відрізняються від навчання і названі образним мисленням. Математичною основою робіт, залучених до огляду, є класична теорія несуперечності обмежень — одна з загальновизнаних парадигм машинного мислення. В роботі показано, як застосування цієї теорії до реальних задач розпізнавання викликає необхідність її подальших узагальнень, що в свою чергу розширює поняття машинного мислення, на формалізацію якого ця теорія на самому початку була спрямована. Формулюється узагальнена задача структурного розпізнавання і образного мислення, окремим випадком якої є класична задача несуперечності обмежень. Для гіббсової статистичної моделі об'єктів, що розпізнаються, показано, як розпізнавання цих об'єктів зводиться до розв'язку того чи іншого окремого випадку узагальненої задачі структурного розпізнавання. Для більш загальних статистичних моделей, не обов'язково гіббсових, виконано аналіз відомих процедур навчання при їх використанні з навчальною інформацією обмеженого об'єму. Досліджено недолік цих процедур, відомий як ефект коротких вибірок, показано його причини і шляхи його подолання.

Ключові слова: розпізнавання образів, машинне навчання, несуперечність обмежень, гіббсові поля, субмодулярна мінімізація.

V.I. Gritsenko, M.I. Schlesinger

INTERRELATION OF PATTERN RECOGNITION, MACHINE THINKING AND LEARNING PROBLEMS

The paper reviews the state-of-the-art in structural recognition, a research area in modern pattern recognition theory. The paper shows that basic problems of structural recognition go beyond machine learning theory and in such way slightly moderates the revived idea that the pattern recognition problem is entirely exhausted with machine learning. At the same time, the paper shows that main concepts and problems of structural recognition form a base for appropriate formalization of particular type of thought processes, which differ from learning and are called imaginative thinking. The main idea of this formalization relies on classical theory of Constraint Satisfaction Problem, one of the acknowledged paradigms of machine thinking. However the binding of this theory to real recognition tasks forces to generalize the theory itself and in such way to specify and refine the concept of machine thinking, for formalization of which the theory was intended. A generalized problem of structural recognition and imaginative thinking is formulated in the paper, classical Constraint Satisfaction Problem being its special case as well as its stochastic and optimization modifications, appropriate with regard to realistic recognition problems. For the Gibbs' statistical model of recognized object, it is shown how recognition of such objects is reduced to such or other special case of generalized structural recognition problem. For more general statistical models, not necessarily Gibbs' models, the application of known learning procedures to fixed size learning samples is analyzed. The flaw known as the short sample effect is explored, its deep-rooted causes are determined as well as a way to overcome them.

Keywords: Pattern recognition, machine learning, constraint satisfaction problem, Gibbs fields, submodular minimization.

1. Shannon C. The Bandwagon, Trans. IRE. 1956. **3**. IT-2, № 1.
2. Шеннон К. Работы по теории информации и кибернетике. М. : Изд-во иностр. лит., 1963. С. 667–668.
3. Винцюк Т.К. Распознавание устной речи методами динамического программирования. *Кибернетика.* 1968. №1. С. 81–88.
4. Винцюк Т.К. Анализ, распознавание и интерпретация речевых сигналов. Киев : Наук. думка, 1987. 262 с.

5. Ковалевский В.А. Оптимальный алгоритм распознавания некоторых последовательностей изображений. *Кибернетика.*1967. № 4. С. 75–80.
6. Ковалевский В.А. Методы оптимальных решений в распознавании изображений. М. : Наука, 1976. 328 с.
7. Нариньяни А.С. Недопределенность в системах представления и обработки знаний. *Изв. АН СССР. Техн. Кибернетика*. 1986. № 5. С. 3–28.
8. Нариньяни А.С. Телерман В.В., Ушаков Д.М., Швецов И.Е. Программирование в ограничениях и недоопределенные модели. *Информационные технологии*. 1998. № 5.
9. Гриценко В.И., Шлезингер М.И. Формальные модели, задачи и алгоритмы образного мышления. *Автоматика-2011/AUTOMATICS-2011, Матеріали 18-ї міжнародної конференції з автоматичного управління* (Львів 28-30 вересня, 2011). С. 110–113 .
10. Rossi F., van Beek P., Walsh T, Handbook of Constraint Programming, Foundations of Artificial Intelligence, Elsevier. 975 p.
11. Щербина О.А. Удовлетворение ограничений и программирование в ограничениях, *Интеллектуальные системы.* 2011. **15**, № 1–4. С. 53–170.
12. Pelillo M., Hancock E. Energy minimization methods in computer vision and pattern recognition. *11th International Conference*. *EMMCVPR 2017.* Venice, Italy, October 30 - November 1, 2017, *Lecture Notes in Computer Science*. 10746, Springer 2018.
13. Ruttkay Zs. Fuzzy constraint satisfaction. *In Proc. 3rd IEEE International Conference on Fuzzy Systems.* 1994. P. 1263–1268.
14. Cooper M.C. Reduction operations in fuzzy or valued constraint satisfaction. *Fuzzy Sets and Systems*. 2003. **134**. P. 311–342. www.elsevier.com/locate/fss
15. Водолазский Е.В., Флах Б., Шлезингер М.И. Минимаксные задачи дискретной оптимизации, инвариантные относительно мажоритарных операторов. *ЖВММФ.* 2014. **54**, № 8. С. 1368–1378.
16. Cohen D.A., Cooper M.C., Peter G. Jeavons P.G., Krokhin A.A. The complexity of soft constraint satisfaction. *Artificial Intelligence*. 2006. **170**. P. 983–1016. www.elsevier.com/-locate/artint
17. Jeavons, P.G., Krokhin, A.A., Zivny, S. The complexity of valued constraint satisfaction. *Bulletin of EATCS*. 2014.
18. Bistarelli S, Montanari U., Rossi F., Schiex T., Verfaillie G., Fargier H. Semiring-Based CSPs and Valued CSPs. *Frameworks, Properties, and Comparison, Constraints*. 1999. **4**, № 3. P. 199–240.
19. Шлезингер М.И. Математические средства обработки изображений. Киев : Наук. думка, 1989. 197 с.
20. Rollon E., Flerova N., Dechter R. Inference schemes for m best solutions for soft csps. *Proceedings of Workshop on Preferences and Soft Constraints*. 2011.
21. Flerova N., Marinescu R., Dechter R. Searching for the M Best Solutions in Graphical Models. *Journal of Artificial Intelligence Research*. 2016. **55**. P. 889–952.
22. Шлезингер М.И., Флах Б., Водолазский Е.В. Поиск заданного количества решений системы размытых ограничений. *Кибернетика и системный анализ*. 2018. № 1. С. 67–83.
23. Шлезингер М.И. Распознавание образов как реализация определенного подкласса процессов мышления. *Управляющие системы и машины.* 2017. № 2. С. 20–37.
24. Jeavons P., Cohen M., Cooper M. Constrarints, Consistency and Closure. *Artificial Intelligence*. 1998. **101**(1-2). P. 251–265.
25. Bulatov A. Mal'tsev constraints are tractable. technical report PRG-RR-02-05, Oxford University, 2002.
26. Bulatov A. Tractable conservative Constraint Satisfaction problems, LICS 2003.
27. Cohen D.A., Jeavons P.G. The Complexity of Constraint Languages. *Handbook of Constraint Programming. Elsevier.* 2006. P. 245–280.
28. Ishikawa H., Geiger D. Segmentation by grouping junctions. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition.* 1998. P. 125–131.
29. Lovasz L. Submodular functions and convexity. A. Bachem, M. Grotschel, and B. Korte, editors, Mathematical Programming — The State of the Art. New York : Springer-Verlag, 1983. P. 235–257.
30. Boykov Y., Kolmogorov V. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *PAMI*. 2004. **26**(9).

31. Boykov Y., Veksler O., Zabih R. Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI)*. 2001. **23**(11). P. 1222–1239.
32. Kolmogorov V., Zabih R. What energy functions can be minimized via graph cuts? *Eur. Conf. on Comp. Vision (ECCV).* Springer-Verlag, 2002. P. 65–81. См. также PAMI, 26(2):147159, February 2004.
33. Osokin A., Vetrov D. Submodular relaxation for Inference in Markov random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 2015. **37**, N 7. P. 1347–1359.
34. Osokin A., Vetrov D., Kolmogorov V. Submodular decomposition framework for inference in associative Markov networks with global constraints. *IEEE Conference on Computer Vision and Pattern Recognition. IEEE*. 2011. P. 1889–1896
35. Шлезингер М.И. Синтаксический анализ двумерных зрительных сигналов в условиях помех. *Кибернетика*. 1976. № 4. С. 113–130.
36. Коваль В.К., Шлезингер М.И. Двумерное программирование в задачах анализа изображений. *Автоматика и телемеханика*. 1976. № 8. С. 149–168.
37. Schlesinger M.I., Flach B. Some solvable subclasses of structural recognition problems. *Czech Pattern Recognition Workshop*. 2000. P. 55–62.
38. Wainwright M.J., Jaakkola T.S., Willsky A.S. MAP estimation via agreement on (hyper)trees: Message-passing and linear-programming approaches. *IEEE Transactions on Information Theory*. 2005. **51**(11):3697–3717.
39. Kolmogorov V. Convergent Tree-reweighted message passing for energy minimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. **28**, N 10, P. 1568–1583.
40. Шлезингер М.И., Гигиняк В.В. Решение (MAX,+)-задач структурного распознавания с помощью их эквивалентных преобразований. *Управляющие системы и машины.* 2007. Ч. 1 в № 1. С. 3–15. Ч. 2: № 2. С. 5–17.
41. Werner T. A linear programming approach to max-sum problem: A Review. *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*. 2007. **29**(7), P. 1165–1179.
42. Werner T. Revisiting the linear programming relaxation approach to Gibbs energy minimization and weighted constraint satisfaction. *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*. 2010. **32**(8). P. 1474–1488.
43. Komodakis N., Paragios N., Tziritas G. Mrf energy minimization and beyond via dual decomposition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2011. **33**(3). P. 531–552.
44. Bulatov A., Grohe M. The complexity of partition functions. *Elsevier, Theoretical Computer Science*. 2005. **348**, N 2–3. P. 148–186.
45. Lynch S. Introduction to applied bayesian statistics and estimation for social scientists. Springer, Berlin ... Tokyo. 2007. P. 385.
46. Wainwright M., Jordan M. Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*. 2008. **1**, N 1–2. P. 1–305.
47. Schlesinger M., Vodolazskiy E. Minimax deviation strategies for machine learning and recognition with short learning samples. *An Article.* 2017 (arXiv preprint arXiv:1707.04849).

Получено 17.03.2020