

УДК 519.6



ВЫДЕЛЕНИЕ ИНФОРМАТИВНЫХ ПРИЗНАКОВ В ЗАДАЧАХ РАСПОЗНАВАНИЯ ОБРАЗОВ

В.И. Бритик¹, Е.А. Егорова²

¹ХНУРЭ, г. Харьков, Украина, valbri@inbox.ru

²ХНУРЭ, г. Харьков, Украина, evg.egorova@gmail.com

В статье представлены результаты исследований проблемы выделения информативных признаков в задачах распознавания образов, которые позволили выделить актуальность дальнейших исследований относительно возможности повышения эффективности выделения информативных признаков в задачах кластеризации посредством использования критериев качества кластеризации.

ЗАДАЧИ РАСПОЗНАВАНИЯ ОБРАЗОВ, ПОДЗАДАЧА ВЫДЕЛЕНИЯ ПРИЗНАКОВ РАСПОЗНАВАНИЯ, ИНФОРМАТИВНЫЕ ПРИЗНАКИ, МЕТОДЫ ВЫДЕЛЕНИЯ ИНФОРМАТИВНЫХ ПРИЗНАКОВ

Введение

Выделение информативных признаков (признаков распознавания) исследуемых объектов рассматривается многими исследователями как один из самых важных и сложных этапов решения задач распознавания [1, 2, 3, 4].

Философские и общеметодологические концепции, составляющие теоретический базис области распознавания образов, определяют и подтверждают значимость данной подзадачи в решении задач распознавания образов.

Сложность подзадачи выбора признаков распознавания обуславливается ее переборным характером. Достаточно большая размерность данных, исследуемых в задачах распознавания, и проблема комбинаторного взрыва¹ ограничивают возможности применения точных математических методов, реализующих общий подход к решению данной подзадачи – полный перебор [1].

Данная ситуация привела к появлению многочисленных математических методов выделения признаков, направленных на сокращение полного перебора. Однако для исследователей в области распознавания образов выделение признаков распознавания в многомерных задачах распознавания образов остается актуальной проблемой.

Целью данной работы является исследование проблемы выделения признаков распознавания в задачах распознавания образов.

Достижение поставленной цели предполагает решение следующих задач:

- исследовать типы и основные этапы решения задач распознавания образов;
- отследить взаимосвязь подзадачи выделения признаков с остальными основными этапами решения задач распознавания;
- провести анализ существующих методов отбора признаков распознавания.

¹ Уже при наличии порядка 20 потенциальных признаков прямой перебор всех комбинаций становится невозможным, поскольку приводит к комбинаторному взрыву.

1. Типы задач распознавания и основные этапы их решения

В зависимости от условий, в которых происходит процесс распознавания, и преследуемой цели в теории распознавания образов выделяются следующие типы задач распознавания².

1) Задача распознавания – отнесение предъявленного (нового) объекта по его описанию к одному из заданных классов.

2) Задача автоматической классификации (задача таксономии, кластеризации) – разбиение множества объектов по их описаниям на систему непересекающихся классов.

3) Динамическое распознавание и динамическая классификация – задачи 1 и 2 в постановке исследования объектов в динамике.

4) Задача прогнозирования – это задачи 3, в которых решение должно относиться к некоторому моменту в будущем [5].

Несмотря на существенные различия, которые прослеживаются в постановках приведенных выше задач распознавания, общие этапы их решения представляются в виде следующих подзадач [3]:

– первая подзадача связана с измерением характеристик, свойств объектов, подлежащих распознаванию, и представлением полученных в результате измерения числовых значений этих характеристик в некотором виде. Данная подзадача определяется как задача измерения и представления исходных данных;

– вторая подзадача связана с выделением характерных свойств (признаков распознавания) из полученных исходных данных и снижением размерности векторов измерений. Эту подзадачу часто определяют как задачу предварительной обработки и выбора признаков;

– третья подзадача состоит в отыскании решающих процедур, правил, необходимых для распознавания объектов.

² Следует отметить, что приведенные ниже задачи имеют также вероятностные постановки.

2. Задача представления и измерения исходных данных

Решение любой задачи распознавания одного из вышеприведенных типов начинается с анализа предметной области, которая заключается в формировании выборки объектов $X = \{x_1, x_2, \dots, x_L\}$, представителей этой предметной области, а также в исследовании и измерении их характеристик.

Любой объект материального мира x может быть описан с помощью бесконечного количества каких-либо характеристик или свойств $G = \{g_1, g_2, \dots, g_s, \dots\}$, то есть $(g_1(x), g_2(x), \dots, g_s(x), \dots)$. В идеале, на этапе анализа выборки объектов исследуемой предметной области набор этих характеристик должен содержать всю информацию об объектах, которая поддается измерению. На практике это условие невыполнимо и в процессе исследований приходится ограничиваться определенным объемом информации, обосновывая достаточность некоторого набора характеристик $G' = \{g_1, g_2, \dots, g_s\}$, $G' \subset G$, для решения поставленной задачи распознавания.

Далее решается задача измерения характеристик или свойств из сформированного набора $G' = \{g_1, g_2, \dots, g_s\}$, которая предполагает выбор соответствующих шкал измерения.

Существует пять основных типов шкал измерений: номинальная (наименований), порядковая, интервальная, относительная и шкала разностей. Измерение количественных характеристик, как правило, производится с помощью шкал интервалов, отношений или шкалы разностей [5]. Качественные характеристики объектов описываются либо номинальной шкалой, либо подвергаются предварительным преобразованиям и ранжируются порядковой или бальной шкалами.

Нередко тип шкал определяют сами значения свойств и характеристик исследуемых объектов: например, такие свойства физического происхождения как рост, вес, температура, значения интенсивности пикселей на полутоновых изображениях и прочие. Но подзадача выбора типа шкал перестает быть тривиальной в случаях, когда приходится работать с признаками, получаемыми искусственным путем, например, в процессе синтеза методами факторного или корреляционного анализов [5]. В таких случаях задача выбора шкал требует дополнительных исследований, поскольку любые преобразования, в том числе переход от одной шкалы к другой, вносят погрешности в измеряемые значения свойств объектов.

Следует отметить, что задача измерения не ограничивается выбором типа шкал и самого процесса измерения. Характерным свойством шкал измерения является такая величина как различимость. Различимость шкалы ρ — это нормированное расстояние между градациями шкалы, рассчитываемое по формуле:

$$\rho = \frac{1}{n-1}, \quad (1)$$

где n — количество градаций.

Свойство различимости не зависит от типа шкалы и, как следует из формулы (1), с ростом числа градаций на шкале ρ стремится к нулю. Условно множество значений различимости можно разделить на два класса. Шкалы в первом классе, различимость внутри которого достаточно велика, удобны для человека, поскольку измеряемые ими значения существенно отличаются друг от друга. Вторым классом характеризуется малыми значениями различимости. Шкалами этого класса человек в повседневной жизни практически не пользуется. Однако к ним приходится обращаться в исследованиях естественных или искусственных микрообъектов, и выбор значения ρ может значительно сказаться на результатах исследований, особенно если это значение превышает порог различимости человека [6]. Таким образом, учет свойства различимости шкал на этапе анализа объектов задачи распознавания является важным моментом: «закладываются» надежность и (или) разрешающая способность разрабатываемой системы распознавания.

Завершающей задачей анализа предметной области является формализация представления исходных данных. Как правило, исходные данные представляются посредством математических моделей, аналогичных моделям описания объектов (образов), которые используются системами распознавания в процессе их функционирования: вектор измерений (или вектор образа), матрица «объект — свойство», древовидные структуры или графы отношений и прочее. [3, 7, 8].

3. Задача предварительной обработки и выбор признаков

Дальнейшие исследования заключаются в выборе признаков распознавания из множества характеристик и упрощения (окончательного формирования) модели описания образов.

Согласно [1, 2, 3, 4] «признаки распознавания» представляют собой характеристики свойств объектов распознавания, которые отражают существенные особенности этих объектов и позволяют устанавливать их сходство или различия с другими объектами.

Таким образом, признаками распознавания является некоторое подмножество $F = \{f_1, f_2, \dots, f_N\}$ множества $G' = \{g_1, g_2, \dots, g_S\}$ ($N \leq S$) характеристик или свойств объектов, подлежащих распознаванию. Количественный и качественный состав подмножества F , называемый множеством признаков распознавания, обуславливается поставленной задачей распознавания. Это объясняется тем, что именно условия и преследуемая цель ре-

шаемой задачи «задают» существенность характеристикам объектов распознавания.

Учитывая факты бесконечного количества характеристик объектов и зависимости множества признаков от постановки задачи распознавания, а также разнообразие самих задач распознавания, можно утверждать о неограниченном числе потенциальных признаков распознавания в рамках общей постановки задачи распознавания.

Исходя из приведенных выше определения «признаков распознавания» и математических моделей описания образов, признаки распознавания являются средством описания объектов исследования, получившим в теории распознавания образов название «признакового описания». Специфика описания объектов признаками распознавания и их информационные особенности прослеживаются в их классификации по следующим критериям:

- по способу описания объектов признаки делятся на: количественные, качественные и логические;
- по характеру доставляемой информации выделяют: вероятностные, детерминированные и квазидетерминированные признаки;
- по источнику доставляемой информации выделяют: физические, математические и структурные признаки;
- по выполняемым функциям признаки делятся на: объединительные и разделительные.

Следует отметить, что одним из важных требований к признакам распознавания является их независимость от условий распознавания. Данное требование выражается в необходимости доказательства инвариантности значений используемых признаков к условиям распознавания.

Универсального формализованного метода формирования набора признаков распознавания F не существует.

Один из подходов к выбору признаков распознавания заключается в использовании знаний экспертов исследуемой предметной области. Данный подход предусматривает применение таких методов извлечения экспертных знаний как интервью, диалог, круглый стол и прочие. [9]. Но экспертный подход минимизирует или исключает возможность автоматизации процесса выбора признаков. Кроме того, выбор признаков экспертным путем поднимает проблему различимости шкал, описанную выше. Решение данной проблемы заключается в определении пороговых величин градаций шкал признаков, превышение которых свидетельствует о целесообразности отказа от мнения экспертов.

Второй подход формирования набора признаков предполагает применение различных математических методов и подходов: корреляционный; кластерный и факторный анализ; информационный подход; подход к выделению признаков; основанный на введении функционалов качества распознавания и прочие.

Нередко применение математических методов требует проведения предварительной обработки исходных данных – их нормализации, поскольку характеристики объектов имеют различные измерения и не сравнимы между собой.

Наиболее распространенные способы нормирования следующие [7, 10]:

$$z^1 = \frac{x - \bar{x}}{\sigma}, z^2 = \frac{x}{x'}, z^3 = \frac{x}{x_{\max}},$$

$$z^4 = \frac{x}{x_{\max}}, z^5 = \frac{x - \bar{x}}{x_{\max} - x_{\min}},$$

где x – исходное значение; z – нормированное значение; \bar{x}, σ – соответственно среднее и среднее квадратичное отклонение x ; x' – некоторое эталонное (нормативное) значение x ; x_{\max}, x_{\min} – наибольшее и наименьшее значения x . z^1 и z^5 определены для шкал отношений и интервалов, остальные способы – только для шкал отношений. Существуют и другие способы нормирования, производные от вышеперечисленных.

Предварительная обработка, приводя исходные данные к виду, удобному для распознавания, способствует уменьшению информативности этих данных, и важно, чтобы это уменьшение не сказало на качестве обрабатываемой информации.

При выборе признаков распознавания лучшим вариантом считается совместное использование методов извлечения экспертных знаний и математических методов обработки многомерных данных. Таким образом, формирование набора признаков распознавания, который учитывает трудности, связанные с реализацией процессов выделения (синтеза) или выбора признаков, и обеспечивает в то же время необходимое качество распознавания, представляет собой одну из наиболее трудных подзадач при построении систем распознавания.

4. Задача поиска решающих правил

Завершающим этапом решения задач распознавания является поиск решающих процедур или правил для распознавания объектов (образов). Универсального подхода нахождения вышеуказанных правил распознавания не существует. Выделяются следующие концепции их поиска: принцип сравнения с эталоном, принцип общности свойств, принцип кластеризации [3].

Использование концепции сравнения с эталоном требует хранения всех объектов, входящих в класс. Кроме этого, процедура сопоставления с эталоном не допускает существенных вариаций характеристик отдельных образов.

Основное допущение при использовании принципа общности свойств заключается в том, что образы, принадлежащие одному и тому же классу, обладают рядом общих свойств или призна-

ков, которые и отражают подобие этих образов. Таким образом, основная задача в данном случае заключается в выделении ряда общих свойств по конечной выборке образов. Использование данной концепции связано с необходимостью развития методов выбора признаков, поскольку достаточно трудно найти для некоторого класса полный набор различающих признаков.

Когда класс образов рассматривается как кластер, а процесс распознавания сводится к определению взаимного пространственного расположения отдельных кластеров, имеет место реализация принципа кластеризации.

Существуют три основных типа методологии, реализующих рассмотренные выше основные принципы либо их комбинации: эвристическая, математическая и лингвистическая (синтаксическая).

В основе эвристического подхода заложены интуиция и опыт человека, в нем используются принципы перечисления членов класса и общности свойств. Как правило, системы, реализующие эвристические методы, включают набор специфических процедур, разработанных применительно к конкретным задачам распознавания.

В основу математического подхода положены правила классификации, которые формулируются и выводятся в рамках определенного математического формализма с помощью принципов общности свойств и кластеризации. Математические методы, как правило, делят на два класса: детерминистские и вероятностные. Детерминистский подход базируется на математическом аппарате, не использующем в явном виде статистические свойства изучаемых классов образов. Вероятностный подход основывается на математических правилах классификации, которые формулируются и выводятся в терминах математической статистики.

Если описание образов производится с помощью непроектируемых элементов (подобразов) и их отношений, то применяется структурно-лингвистический или синтаксический подход с использованием принципа общности свойств. Образ описывается в данном случае с помощью иерархической структуры подобразов, а при решении задач распознавания образов применяется теория формальных языков [8].

В рамках каждой из указанных методологий разработано огромное количество методов, моделей и технологий, которые позволят строить решающие правила, удовлетворяющие требованиям той или иной задачи распознавания: вероятностные методы распознавания, метод потенциальных функций, методы кластерного анализа, генетические алгоритмы, нейросетевые технологии, нечеткие модели распознавания и прочие.

Рассмотренные выше подзадачи не являются изолированными этапами процесса распознава-

ния. Условия поставленной задачи распознавания, особенности предметной области, видение данной проблемы исследователем определяют некоторую концепцию, которая затем реализуется в средствах и методах, используемых на последующих этапах решения этой задачи распознавания. В то же время, результаты, полученные в процессе решения одной из подзадач, могут существенно влиять на выбор методов дальнейших исследований либо выступить в качестве дополнительной, уточняющей информации, использование которой позволяет упростить вычислительные схемы предыдущих этапов. Взаимосвязь подзадачи выделения признаков с остальными основными этапами решения задач распознавания представлена на рисунке.



1. Выборка объектов исследуемой предметной области.
- 1'. Объект (Объекты).
2. Характеристики и свойства объектов, средства и шкалы измерений.
3. Результаты измерений. Вектор (вектора) измерений.
4. Модель представления описания объектов.
5. Средства предварительной обработки данных.
- 5'. Информативные признаки.
6. Информативность признаков исследуемых объектов. Вектор (вектора) признаков.
7. Решающие правила классификации.
8. Информативность признаков исследуемых объектов. Класс (классы, идентификационные метки объектов).
- 9, 10. Уточняющая информация.

Рис. Блок-схема процесса распознавания образов как процедуры решения совокупности подзадач

5. Методы выделения признаков в задачах распознавания

Существующие на сегодняшний день многочисленные математические методы выделения признаков распознавания реализуют различные подходы, направленные на сокращение полного перебора потенциальных признаков распознавания: корреляция и кластеризация признаков; информационный подход; подход, основанный на введении функционалов качества распознавания; подход, базирующийся на минимизации энтропии и прочие.

Применение корреляционного анализа к исходным данным позволяет определить взаимосвязь характеристик объектов и выявить дублирующую информацию, использование которой способствует только усложнению алгоритмической схемы распознавания.

Установить наличие связей внутри совокупности характеристик исследуемых объектов и попытаться внести в нее структуру также с целью дальнейшего снижения размерности данной совокупности можно, применив методы кластерного анализа в пространстве измерений. Наибольшее распространение получили следующие методы кластерного анализа: метод ближайшего соседа, полносвязывающий метод, метод Кинга, метод k средних, ISODATA и пр. [8, 10].

Данные подходы отбора признаков применяются в детерминистских задачах распознавания образов как при известной, так и неизвестной правильной классификации исследуемого множества объектов. Следует также отметить, что кластеризацию признаков рекомендуется производить в качестве предварительной обработки в других методах выделения признаков, ссылаясь на следующие недостатки этих подходов [1]:

1) Информации о попарном сходстве между признаками в общем случае недостаточно для выделения оптимального набора признаков. В частности, набор попарно некоррелированных признаков может оказаться линейно зависимым.

2) Могут существовать кластеры, полностью состоящие из неинформативных признаков. В таких случаях, выбрав по одному из представителей от каждого кластера, все равно придется решать задачу отбора признаков.

Для выделения признаков в детерминистских задачах распознавания образов с известной правильной классификацией можно использовать подход, основанный на введении различных критериев качества распознавания, например функционал ошибки обобщения, критерий скользящего контроля, критерии непротиворечивости и регуляризации и прочие. С произвольными функционалами качества распознавания совместимы следующие методы отбора признаков: полный перебор; последовательное добавление признаков; поочередное добавление и удаление признаков; поиск в глубину (метод ветвей и границ); поиск в ширину (МГУА); генетические алгоритмы; случайный поиск с адаптацией [1].

Метод полного перебора характеризуется достаточно простой вычислительной схемой. Алгоритм, реализующий данный метод, осуществляет полный перебор наборов признаков в порядке возрастания сложности этих наборов (числа признаков в наборе). Для каждого значения сложности рассматриваются все возможные наборы. На втором

шаге из них выбирается лучший набор. На шаге 3 запоминается значение сложности, при котором был получен самый лучший набор. Если результат не улучшается на протяжении d итераций, считается, что минимум внешнего критерия пройден, и алгоритм заканчивает работу, выдавая наилучший набор. Таким образом, сокращение полного перебора происходит за счет того, что наборы с высокой сложностью вообще не рассматриваются.

Стратегию «жадного» наискорейшего спуска реализует метод последовательного добавления признаков, добавляя к набору по одному признаку, который приводит к наибольшему уменьшению внешнего критерия. Данный метод сокращает трудоемкость перебора с $O(2^n)$ операций до $O(n^2)$, но включает «лишние» признаки, поскольку однажды включенный признак остается в наборе до конца работы алгоритма.

В тех случаях, когда заранее известно, что информативных признаков намного больше шумовых, целесообразнее использовать метод последовательного удаления признаков. Работа алгоритма данного метода начинается с набора $F_0 = G'$, где множество выборочных характеристик и/или свойств объектов, подлежащих распознаванию, и далее из набора последовательно исключается по одному признаку так, чтобы значение критерия убывало как можно быстрее. Метод последовательного удаления признаков работает медленнее метода последовательного добавления, поскольку на начальных итерациях приходится обрабатывать почти все признаки.

Стратегии двух жадных алгоритмов, описанных выше, совмещает в себе метод поочередного добавления и удаления признаков. Процессы последовательных добавлений и удалений чередуются до тех пор, пока значение критерия качества распознавания в точках минимума функционала не перестанет уменьшаться, или пока состав признаков в оптимальном наборе не стабилизируется. Данный метод характеризуется достаточно большой трудоемкостью и сложностью реализации.

Метод ветвей и границ позволяет перейти к сокращенному перебору за счет следующих эвристик:

1) Оценивать перспективность ветви дерева признаков и отказаться от ее наращивания, если уже имеется лучшая ветвь. Набор не наращивается, если значение критерия качества распознавания оказывается хуже, чем на самом лучшем из уже проверенных наборов меньшей сложности.

2) Желательно как можно раньше построить наиболее удачную ветвь дерева признаков. Тогда значения критерия на самых лучших наборах сложности j , $j = 1, S$ будут близки к наилучшему значению критерия для набора сложности j , и первая эвристика будет отсекал большинство ветвей в самом начале алгоритма.

Качество решения задачи выбора признаков данным методом во многом определяется выбором значений его параметров и, следовательно, требует проведения дополнительных экспериментов.

Достаточно широкое применение для формирования набора признаков распознавания получили метод группового учета аргументов и генетические алгоритмы.

На каждой итерации алгоритма, реализующего метод группового учета аргументов (МГУА) [1, 2], строится не один набор, а множество из B_j наборов, называемое j -м рядом: $R_j = \{F_j^1, \dots, F_j^{B_j}\}$, $F_j^b \subseteq G'$, $|F_j^b| = j$, $b = \overline{1, B_j}$. Для перехода от текущего ряда R_j к следующему R_{j+1} от каждого набора $F \in R_j$ порождается $n-j$ новых наборов путем присоединения одного из признаков, не принадлежащих набору F . Из порожденных $B_j(n-j)$ наборов в следующий ряд отбирается не более B наборов, лучших по внешнему критерию. Таким образом, на каждой итерации сложность всех наборов увеличивается на единицу. Число B является параметром алгоритма и называется шириной поиска.

Генетический алгоритм (ГА) [1, 11] осуществляет поиск наилучшего набора признаков по принципам дарвиновской эволюции. Первое поколение наборов генерируется случайным образом. К этим наборам применяются операции скрещивания и мутации для порождения большого числа новых наборов. Затем производится искусственный отбор или селекция: во второе поколение отбираются только B наборов, лучших по заданному внешнему критерию Q . Ко второму поколению также применяются операции скрещивания, мутации и селекции, и порождается третье поколение. Эволюционный процесс переходит от поколения к поколению до тех пор, пока не наступит стагнация, то есть качество лучшего набора в поколении перестанет улучшаться. Анализ показывает, что генетический алгоритм от МГУА отличается, главным образом, правилом порождения следующей популяции, остальные различия – в терминологии.

Применение генетического алгоритма позволяет вводить различные эвристики. Можно по-разному вводить операции скрещивания и мутации, применять совокупности критериев качества, разрешать скрещиваться только лучшим индивидуумам, переносить или не переносить лучших индивидуумов в следующее поколение, увеличивать вероятность мутаций при наступлении стагнации, параллельно выращивать изолированные популяции и так далее.

Недостатком генетического алгоритма является его относительно медленная сходимость. Кроме этого, наряду с параметрами размера популяции B , максимального числа поколений и вероятнос-

ти мутации, ГА имеет и другие параметры, подбор которых является искусством и зависит от особенностей задачи.

Если в генетическом алгоритме отказаться от операции скрещивания, получается метод случайного поиска. В t -м поколении популяции из B наборов модифицируется каждый из них R раз случайным образом. Из полученных BR наборов отбираются B лучших по заданному внешнему критерию Q , и из них формируется $(t+1)$ -е поколение. Недостатком данного метода (аналогично ГА) является довольно медленная сходимость.

На ускорение сходимости направлен алгоритм случайного поиска с адаптацией [1]. Идея адаптации заключается в том, чтобы, генерируя наборы признаков случайным образом, увеличивать вероятность включения в них тех признаков, которые чаще входят в наилучшие наборы, а для признаков, входящих в наихудшие наборы, уменьшать вероятность их появления.

Трудоёмкость алгоритма СПА составляет $O(Rr(j^* + d))$ операций. К достоинствам данного метода относят простоту реализации и существенно меньшее, по сравнению с ГА, число параметров. СПА сходится гораздо быстрее, чем случайный поиск без адаптации, и во многих экспериментах находит решения лучшего качества, чем алгоритм поочередного добавления и удаления признаков.

Приведенные выше методы являются эвристиками и не могут гарантировать оптимальность получаемого результата, поэтому можно говорить лишь об удовлетворяющем качестве сформированного набора признаков. Недостатком данного подхода является значительная трудоёмкость его реализации.

Альтернативным средством выделения признаков в детерминистских задачах распознавания образов с известной классификацией является информационный подход, в рамках которого можно рассчитать количество информации в событии класса по отношению к каждому из событий признака или оценить информативность каждого признака отдельно с помощью мер Хартли, Шеннона, Макинтоша и прочих [3, 2]. Формирование набора признаков распознавания в рамках информационного подхода производится на основании принципов учета признаков, которые дают наибольшую информацию об исследуемых объектах, и пренебрежении признаками, обладающими малой информативностью. Информационный подход используется также для выделения информативных признаков в детерминистских задачах распознавания при неизвестной правильной классификации и в вероятностных задачах распознавания.

В рамках вероятностной задачи распознавания, когда известны существующие вероятностные связи между признаками объектов и классами, к ко-

торым они принадлежат, выбор информативных признаков распознавания можно также произвести, используя подходы минимизации энтропии или максимизации дивергенции. В [3] отмечается, что кроме вычислительных трудностей возможность применения на практике этих подходов ограничивает также то обстоятельство, что количество реальных задач распознавания, имеющих полную или частичную вероятностную информацию: плотности распределения, характеризующие совокупности объектов и/или условные плотности распределения значений признаков – невелика.

Известно, что в процессе решения конкретной задачи распознавания в рамках структурного и структурно-лингвистического подходов выбор признаков объектов обуславливается результатами многочисленных экспериментов, опытом и интуицией исследователя, спецификой данных, поставленными условиями и ограничениями решаемой задачи, а также общими требованиями к примитивам [8]. Но в то же время частичное уменьшение доли субъективизма в выборе примитивов объектов распознавания возможно посредством применения математических методов и подходов к отбору признаков.

Выводы

Проведенные исследования показали, что теоретический «багаж» области распознавания образов обладает широким набором средств для решения подзадачи выделения признаков распознавания. Тем не менее, для решения данной подзадачи в задачах автоматической классификации выбор достаточно ограничен: наибольшее применение получили кластерный, корреляционный и информационный подходы, в рамках которых выделение информативных признаков является «изолированным» процессом с точки зрения отсутствия обратных связей с этапом кластеризации [1].

Данная ситуация обуславливает актуальность дальнейших исследований относительно возможности повышения эффективности выделения информативных признаков в задачах автоматической классификации за счет внесения указанных выше обратных связей посредством критериев качества кластеризации. Потенциальным преимуществом данного подхода по сравнению с существующими альтернативными методами и подходами выделения признаков является учет кластеризации, что наиболее близко отображает сущность понятия «признак» («признак распознавания», «информативный признак», «значимый признак») в рамках теории распознавания образов.

Список литературы: 1. *Воронцов К.В.* Лекции по методам отбора и синтеза признаков и моделей. <http://ccas.ru/voron/download/FeaturesTransform.pdf> 2. *Ивахненко А.Г.* Самообучающиеся системы распознавания и автоматического управления. – К.: Техника, 1969. – 392 с. 3. *Дж. Ту, Р. Гонсалес.* Принципы распознавания образов. – М.: Мир, 1978. – 390 с. 4. *Белозерский Л.А.* Базовые понятия теории и практики современного распознавания // Искусственный интеллект. 2005. №2. С. 4–15. 5. *Афонин В.Л., Макушкин В.А.* Интеллектуальные робототехнические системы. Курс лекций. <http://www.intuit.ru/department/human/isrob/4/>. 6. *Крислов В.А., Юдин С.А.* Естественная и искусственная таксономия // Искусственный интеллект. – 2005. – №1. – С. 74–85. 7. *Браверман Э.М., Мучник И.Б.* Структурные методы обработки эмпирических данных. – М.: Наука, 1983. – 464 с. 8. *Классификация и кластер* / Под ред. Дж. Райзина. 1900. 9. *Андрейчиков А.В., Андрейчикова О.Н.* Интеллектуальные информационные системы. – М.: Финансы и статистика, 2004. – 424 с. 10. *Мандель И.Д.* Кластерный анализ. – М.: Финансы и статистика, 1988. – 176 с. 11. *Рутковская Д., Пилинский М., Рутковский Л.* Нейронные сети, генетические алгоритмы и нечеткие системы. – М.: Горячая линия, 2007. – 452 с.

Поступила в редакцию 7.04.2008