

Лекция 6.

Тема: Общая характеристика задач распознавания образов и их типы.

Под образом понимается структурированное описание изучаемого объекта или явления, представленное вектором *признаков*, каждый элемент которого представляет числовое значение одного из *признаков*, характеризующих соответствующий *объект*. Общая структура системы распознавания и этапы в процессе ее разработки показаны на рис. 1.

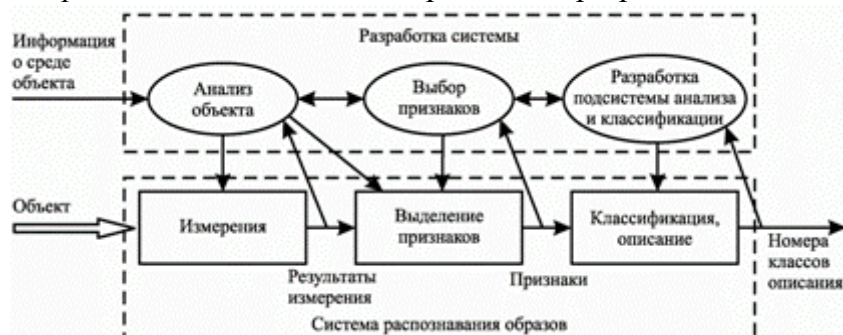


Рис. 1. Структура системы распознавания

Суть задачи распознавания - установить, обладают ли изучаемые объекты фиксированным конечным набором *признаков*, позволяющим отнести их к определенному классу.

Задачи распознавания имеют следующие *характерные черты*.

1. Это *информационные задачи*, состоящие из двух этапов: а) приведение исходных данных к виду, удобному для распознавания; б) собственно *распознавание* (указание принадлежности объекта определенному классу).

2. В этих задачах можно *вводить понятие аналогии или подобия объектов и формулировать понятие близости объектов* в качестве основания для зачисления объектов в один и тот же класс или разные классы.

3. В этих задачах можно *оперировать набором прецедентов-примеров*, классификация которых известна и которые в виде формализованных описаний могут быть предъявлены алгоритму *распознавания* для настройки на задачу в процессе обучения.

4. Для этих задач *трудно строить формальные теории и применять классические математические методы* (часто недоступна информация для точной математической модели или выигрыш от использования модели и математических методов не соизмерим с затратами).

5. В этих задачах *возможна "плохая" информация* (информация с пропусками, разнородная, косвенная, нечеткая, неоднозначная, вероятностная).

Целесообразно выделить следующие *типы задач распознавания*.

1. Задача *распознавания* - отнесение предъявленного объекта по его описанию к одному из заданных классов (*обучение с учителем*).

2. Задача автоматической классификации - *разбиение множества объектов (ситуаций)* по их описаниям на систему непересекающихся классов (*таксономия, кластерный анализ, обучение без учителя*).

3. Задача выбора информативного набора *признаков* при распознавании.

4. Задача приведения исходных данных к виду, удобному для распознавания.

5. *Динамическое распознавание* и динамическая классификация - задачи 1 и 2 для динамических объектов.

6. Задача прогнозирования - это задачи 5, в которых решение должно относиться к некоторому моменту в будущем.

Основы теории анализа и распознавания изображений.

Пусть дано множество M объектов; на этом множестве существует разбиение на конечное число подмножеств (классов) $\Omega_i, i = \{1, m\}, M = \cup \Omega_i (i = 1..m)$. Объекты

ω задаются значениями некоторых признаков $x_j, j = \{1, N\}$. Описание объекта $I(\omega) = (x_1(\omega), \dots, x_N(\omega))$ называют стандартным, если $x_j(\omega)$ принимает значение из множества допустимых значений.

Пусть задана таблица обучения (таблица 1). Задача распознавания состоит в том, чтобы для заданного объекта ω и набора классов $\Omega_1, \dots, \Omega_m$ по обучающей информации в таблице обучения $I_0(\Omega_1 \dots \Omega_m)$ о классах и описанию $I(\omega)$ вычислить предикаты:

$$P_i(\omega \in \Omega_i) = \{1(\omega \in \Omega_i), 0(\omega \in \Omega_i), (\omega \in \Omega_i)\},$$

где $i = \{1, m\}$, Δ - неизвестно.

Таблица 1. Таблица обучения

Объект	Признаки и их значения			Класс
	x_1	x_j	x_n	
ω_1	α_{11}	α_{1j}	α_{1n}	Ω_1
ω_{r_1}	$\alpha_{r_1 1}$	$\alpha_{r_1 j}$	$\alpha_{r_1 n}$	
ω_{r_k}	$\alpha_{r_k 1}$	$\alpha_{r_k j}$	$\alpha_{r_k n}$	Ω_m
ω_{r_m}	$\alpha_{r_m 1}$	$\alpha_{r_m j}$	$\alpha_{r_m n}$	

Рассмотрим алгоритмы распознавания, основанные на вычислении оценок. В их основе лежит принцип прецедентности (в аналогичных ситуациях следует действовать аналогично).

Пусть задан полный набор признаков x_1, \dots, x_N . Выделим систему подмножеств множества признаков S_1, \dots, S_k . Удалим произвольный набор признаков из строк $\omega_1, \omega_2, \dots, \omega_{r_m}$ и обозначим полученные строки через $S\omega_1, S\omega_2, \dots, S\omega_{r_m}, S\omega'$.

Правило близости, позволяющее оценить похожесть строк $S\omega'$ и $S\omega_r$ состоит в следующем. Пусть "усеченные" строки содержат q первых символов, то есть $S\omega_r = (a_1, \dots, a_q)$ и $S\omega' = (b_1, \dots, b_q)$. Заданы пороги $\varepsilon_1 \dots \varepsilon_q, \delta$. Строки $S\omega_r$ и $S\omega'$ считаются похожими, если выполняется не менее чем δ неравенств вида

$$|a_j - b_j| \leq \varepsilon_j, j = 1, 2, \dots, q.$$

Величины $\varepsilon_1 \dots \varepsilon_q, \delta$ входят в качестве параметров в модель класса алгоритмов на основе оценок.

Пусть $\Gamma_i(\omega')$ - оценка объекта ω' по классу Ω_i .

Описания объектов $\{\omega'\}$, предъявленные для распознавания, переводятся в числовую матрицу оценок. Решение о том, к какому классу отнести объект, выносится на основе вычисления степени сходства распознавания объекта (строки) со строками, принадлежность которых к заданным классам известна.

Проиллюстрируем описанный алгоритм распознавания на примере. Задано 10 классов объектов (рис. 2а). Требуется определить признаки таблицы обучения, пороги и построить оценки близости для классов объектов, показанных на рис. 2б. Предлагаются следующие признаки таблицы обучения:

- x_1 - количество вертикальных линий минимального размера;
- x_2 - количество горизонтальных линий;
- x_3 - количество наклонных линий;

x_4 - количество горизонтальных линий снизу объекта.

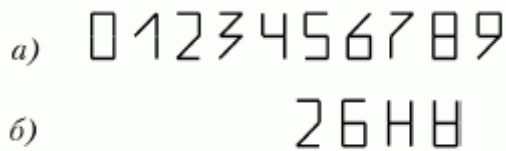


Рис. 2. Пример задачи по распознаванию

На рис. 4.3 приведена *таблица обучения и пороги*
 $\varepsilon_1 = 1, \varepsilon_2 = 1, \varepsilon_3 = 1, \varepsilon_4 = 1, \delta = 1$.

Из этой таблицы видно, что неразличимость символов 6 и 9 привела к необходимости ввода еще одного признака x_4 .

	X_1	X_2	X_3	X_4	
	4	2	0		0
	2	0	1		1
	1	2	1		2
	0	2	2		3
	3	1	0		4
	2	3	0		5
✓	2	2	1	1	6
	1	1	1		7
	4	3	0		8
✓	2	2	1	0	9
	$\varepsilon_1=1$	$\varepsilon_2=1$	$\varepsilon_3=1$	$\varepsilon_4=1$	$\delta=1$

Рис. 3. Таблица обучения для задачи по распознаванию

Теперь может быть построена *таблица распознавания* для объектов на рис. 5.2б.

Объект	x_1	x_2	x_3	x_4	Результат распознавания
Объект 1	1	2	1		Цифра 2
Объект 2	3	3	0	1	Цифра 8 или 5
Объект 3	4	1	0		
Объект 4	4	2	0	1	

Читателю предлагается самостоятельно ответить на вопрос: что будет, если увеличить пороги $\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4, \delta$? Как изменится качество *распознавания* в данной задаче?

Отметим важную мысль, высказанную А. Шамисом: качество *распознавания* во многом зависит от того, насколько удачно создан *алфавит признаков*, придуманный разработчиками системы. Поэтому *признаки* должны быть инвариантны к ориентации, размеру и вариациям формы объектов.

Распознавание по методу аналогий.

Этот метод очень хорошо знаком студентам (*знание* решения аналогичной задачи помогает в решении текущей задачи).

Рассмотрим *этот метод* на примере задачи П. Уинстона по поиску геометрических аналогий, представленном на рис. 4. Среди фигур второго ряда требуется выбрать

$X \in \{1, 2, 3, 4, 5\}$ такое, что **A** так соотносится с **B**, как **C** соотносится с **X**, и такое, которое лучше всего при этом подходит. Для решения задачи необходимо понять, в чем

разница между фигурами **A** и **B** (наличие/отсутствие жирной точки), и после этого ясно, что лучше всего для **C** подходит **X=3**.

Решение таких задач предполагает описание изображения и преобразования (отношения между фигурами на изображениях), а также описание изменения отдельных фигур, составление правил и оценка изменений.

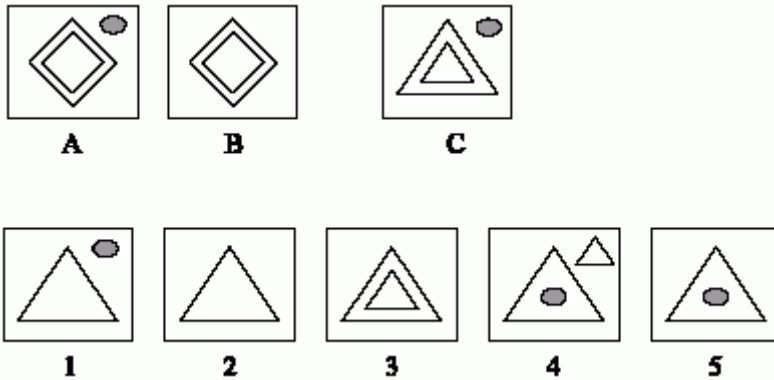


Рис. 4. Задача поиска геометрических аналогий

В качестве примера запишем три правила, показывающие, каким образом одно изображение (исходное) становится результирующим (рис. 5).

Правило 1 (исходное изображение): **k** выше **m**, **k** выше **n**, **n** внутри **m**

Правило 2 (результир. изображение): **n** слева **m**

Правило 3 (масштабирование, повороты):

k исчезло

m изменение масштаба 1:1, вращение 0^0

n изменение масштаба 1:2, вращение 0^0

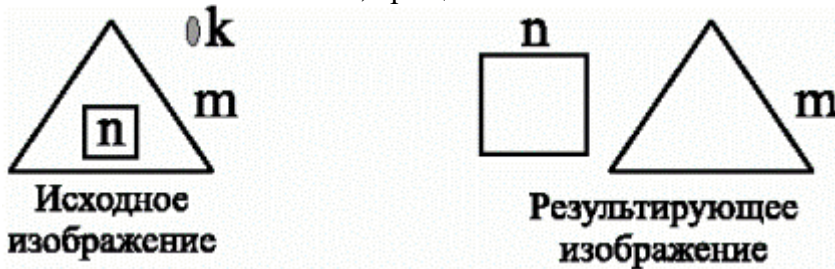


Рис. 5. Правила преобразования

Отметим важные моменты при таких преобразованиях. В исходном и результирующем изображениях допускаются отношения **ВЫШЕ**, **ВНУТРИ**, **СЛЕВА**, В результате преобразования изображение может стать **МЕНЬШЕ**, **БОЛЬШЕ**, испытать **ПОВОРОТ** или **ВРАЩЕНИЕ**, **ОТРАЖЕНИЕ**, **УДАЛЕНИЕ**, **ДОБАВЛЕНИЕ**. Написание правил лучше всего начинать с проведения диагональных линий через центры фигур. Лишние отношения (**СПРАВА ОТ** и **СЛЕВА ОТ**, **ВЫШЕ** и **НИЖЕ**, **ИЗНУТРИ** и **СНАРУЖИ**.) использовать не рекомендуется.

Теперь задачи распознавания мы можем решать достаточно просто, записав для отношений правила 1, 2, 3 и проведя сопоставление, например так, как это сделано для следующей задачи: найти **X** такое, что **A => B**, как **C => X** (рис. 6).

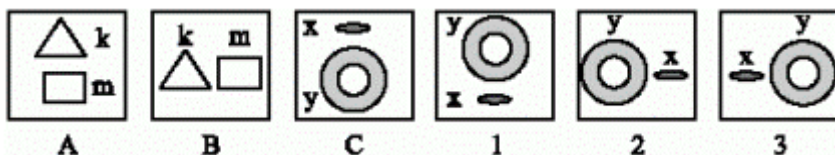


Рис. 6. Пример задачи распознавания по аналогии

	Правило 1	Правило 2	Правило 3	Результат
A => B	k выше m	k слева m	k, m масштаб 1:1 поворот 0 ⁰	
C => 1	x выше y	y выше x	x, y масштаб 1:1 поворот 0 ⁰	
C => 2	x выше y	y слева x	x, y масштаб 1:1 поворот 0 ⁰	
C => 3	x выше y	x слева y	x, y масштаб 1:1 поворот 0 ⁰	
				Сопоставление успешно

Дополнительно следует отметить, что разные виды преобразований могут иметь различные веса, например, исчезновению фигуры целесообразно назначить больший *вес*, чем преобразованию масштаба; а вращение фигуры может иметь меньший *вес*, чем отражение.

Методы распознавания по аналогии могут быть эффективнее, если используется обучение. Различают *обучение с учителем*, обучение по образцу (эталону) и др. виды обучения. Суть идеи такова. Программе распознавания предъявляется *объект*, например, арка. Программа создает внутреннюю модель:

(арка
 (компонент1 (назначение (опора))
 (тип (брусок)))
 (компонент2 (назначение (опора))
 (тип (брусок)))
 (компонент3 (назначение (перекладина))
 (тип (брусок))
 (поддерживается (компонент1), (компонент2))))

После этого предъявляется другой *объект* и говорится, что это тоже арка. Программа вынуждена дополнить свою внутреннюю модель:

(арка
 (компонент1 (назначение (опора))
 (тип (брусок)))
 (компонент2 (назначение (опора))
 (тип (брусок)))
 (компонент3 (назначение (перекладина))
 (тип (брусок) или (клин))
 (поддерживается (компонент1), (компонент2))))

После такого обучения *система распознавания* будет узнавать в качестве арки как первый, так и второй *объект*.

Актуальные задачи распознавания.

Среди множества интересных задач по распознаванию (*распознавание* отпечатков пальцев, *распознавание* по радужной оболочке глаза, *распознавание* машиностроительных чертежей и т. д.) следует выделить **задачу определения реальных координат заготовки и определения шероховатости обрабатываемой поверхности**, рассмотренную в лекции 10. Другой актуальной задачей является *распознавание машинописных и рукописных текстов* в силу ее повседневной необходимости. Практическое значение задачи *машинного чтения печатных и рукописных текстов* определяется необходимостью представления, хранения и использования в электронном виде огромного количества накопленной и вновь создающейся текстовой информации. Кроме того, большое значение имеет оперативный ввод в информационные и управляющие системы информации с машиночитаемых бланков, содержащих как напечатанные, так и рукописные тексты. В связи с этим рассмотрим принципы и подход к *распознаванию в задаче машинного чтения печатных и рукописных текстов*.

Для решения данной задачи используются следующие основные принципы.

1. *Принцип целостности* - распознаваемый объект рассматривается как единое целое, состоящее из структурных частей, связанных между собой пространственными отношениями.

2. *Принцип двунаправленности* - создание модели ведется от изображения к модели и от модели к изображению.

3. *Принцип предвидения* заключается в формировании гипотезы о содержании изображения. Гипотеза возникает при взаимодействии процесса "сверху-вниз", разворачивающегося на основе модели среды, модели текущей ситуации и текущего результата восприятия, и процесса "снизу-вверх", основанного на непосредственном грубом *признаковом* восприятии.

4. *Принцип целенаправленности*, включающий *сегментацию* изображения и совместную интерпретацию его частей.

5. *Принцип "не навреди"* - ничего не делать до распознавания и вне распознавания, то есть без "понимания".

6. *Принцип максимального использования модели* проблемной среды.

Указанные принципы реализованы в пакете программ "Графит", в программах FineReader-рукопись и FormReader - для распознавания рукописных символов и, частично, в программе FineReader для *распознавания печатных текстов*. Входящая в FormReader *программа чтения рукописных текстов* была выпущена в 1998 году одновременно с системой АBBYY FineReader 4.0. Эта *программа* может читать все рукописные строчные и заглавные символы, допускает ограниченные соприкосновения символов между собой и с графическими линиями и обеспечивает поддержку 10 языков. Основное применение программы - *распознавание* и ввод информации с машиночитаемых бланков.

В системе АBBYY FormReader при *распознавании рукописных текстов* используются структурный, растровый, *признаковый*, дифференциальный и лингвистический уровни распознавания. Для более подробного освоения подходов к *распознаванию машинописных и рукописных текстов* в системе АBBYY FormReader читателю рекомендуется непосредственно ознакомиться с работой А. Шамиса, при этом *знание* основ машинной графики на уровне подразумевается.

Завершая этот раздел лекции, отметим особенности задачи зрительного восприятия роботов по сравнению с традиционными задачами *распознавания образов* и машинной обработки изображений:

- необходимость построения комплексного описания среды на основе учета значительной априорной информации (модели проблемной среды) в отличие от традиционной задачи выделения фиксированных *признаков* или измерения отдельных параметров;
- необходимость анализа трехмерных сцен не только в плане анализа трехмерных объектов по их плоским проекциям, но и в плане определения объемных пространственных отношений;
- необходимость анализа изображений, включающих одновременно несколько произвольно расположенных объектов (в общем случае произвольной формы) в отличие от традиционной задачи, когда для распознавания предъявляется, как правило, один объект;
- необходимость анализировать реальную динамическую среду, а не статические изображения;
- отсутствие постоянной фиксированной задачи и необходимость оперативно решать возникающие по ходу дела задачи;
- необходимость следить за изменениями в среде, которые могут порождать новые оперативные задачи;
- необходимость организации системного процесса взаимодействия в реальном времени нескольких подсистем робота ("глаз-мозг", "глаз-мозг-рука").