

Курс лекций «Введение в ИИ.». Часть II.
Распознавание образов
Лекция 13. Алгоритмы распознавания образов

О.Г. Чанышев

Содержание

1	Распознавание образов	1
2	Решающие правила, опирающиеся на прецеденты	2
2.1	Алгоритм ближайшего соседа	2
2.2	Метод потенциальных функций	2
2.3	Минимизация набора прецедентов	3
2.3.1	Алгоритм STOPL	3
2.3.2	Метод дробящихся эталонов - алгоритм ДРЭТ	3
3	Логические решающие правила. Алгоритм CORAL	3
4	О выборе информативного множества признаков	5
4.1	Методы последовательного сокращения и добавления признаков DEL и ADD	6
4.2	Метод случайного поиска с адаптацией (алгоритм СПА)	6
5	Меры близости между предикатами	7
5.1	Расстояния r	8
5.2	«Консенсус»	8
5.3	Энтропийная мера или категоричность суждений	8
5.4	Расстояние между знаниями	9

1 Распознавание образов

В предшествующей лекции мы рассмотрели алгоритмы класса *FOREL* - алгоритмы кластеризации. Итог работы алгоритмов можно зафиксировать по разному в зависимости от назначения. В настоящей лекции мы рассмотрим некоторые алгоритмы классификации, или задачи распознавания образов.

Напомним, что по принятой нами классификации задач анализа данных, задача таксономии - это задача типа 1.3.Н и заключается в предсказании всех элементов нового столбца в шкале наименований. Т.е., для каждого объекта должен быть указан номер его таксона.

Наиболее распространенный способ представления результатов решения задач кластеризации состоит в переформировании исходной таблицы путем собирания в отдельные слои всех

m_i строк (объектов), входящих в один и тот же таксон, создавая, таким образом, базу для алгоритмов распознавания образов или классифицирующие таксоны.

Для более краткого представления основного содержания такой таблицы можно записать, например, средние значения и дисперсию свойств объектов каждого таксона. Можно сохранить по одному или несколько типичных представителей (прецедентов) из каждого таксона (например, объекты, являющиеся центрами таксонов). Можно в пространстве характеристик описать границы, разделяющие таксоны.

Описание такого рода представляет собой обобщенный образ класса объектов. Если предъявить новый объект, не участвующий ранее в таксономии, и потребовать отнести его к одному из классов, то нужно проанализировать характеристики предъявляемого объекта, сравнить их с характеристиками объектов, представляющих классифицирующие таксоны, определить степень сходства и отнести объект к одному из классов, для которого сходство максимально.

Такая процедура получила наименование процедуры распознавания образов и соответствует задаче типа 1.1.Н, когда требуется распознать один элемент в столбце свойств, измеренных в шкале наименований.

На вход алгоритма распознавания подается таблица данных, содержащая m объектов $A = \bigcup_i^m a_i$ с соответствующими характеристиками $X = \bigcup_j^n x_j \cup z$. Характеристика z измерена в качественной шкале. Такая таблица называется обучающей выборкой.

Процесс распознавания состоит из двух основных этапов - обучения и принятия решения (контроля). На первом требуется обнаружить закономерную связь между значениями x -характеристик и z . Эта закономерность выражается в виде *решающего правила*, при помощи которого принимается решение о принадлежности нового объекта классу.

2 Решающие правила, опирающиеся на прецеденты

Как мы уже неоднократно говорили ранее, все реальные задачи распознавания отличаются отсутствием знаний о генеральной совокупности объектов.

В таком случае стратегии распознавателей можно разделить на два направления. Идея первого заключается в стремлении максимально приблизить реальную ситуацию к идеальному случаю и затем спокойно пользоваться строгими аналитическими методами построения решающих правил. Делается предположение о том, что имеющаяся конечная выборка хорошо отражает свойства генеральной совокупности (*гипотеза о представительности выборки*) и что генеральная совокупность подчиняется одному или смеси из нескольких простых законов распределения (*гипотеза о типе распределения*). Умом этот подход можно понять, но все равно, как и в любые эмпирические гипотезы, в него можно только верить. В защиту такого подхода говорит тот факт, что с ростом объема обучающей выборки точность распознавания обычно увеличивается.

Второй подход опирается на гипотезу локальной компактности, из которой следует, что в локальной окрестности имеющихся реализаций образа обучающей выборки могут появиться только реализации того же образа. И чем ближе классифицируемый образ к реализациям обучающей выборки, с тем большей вероятностью он к ней относится.

2.1 Алгоритм ближайшего соседа

Тогда можно предложить следующую процедуру в качестве решающего правила: оставить в памяти машины все реализации обучающей выборки и классифицируемую точку (образ) отне-

сти к тому классифицирующему образу, чья реализация оказалась ближайшей. Это - *правило ближайшего соседа*.

Учитывая, что результаты реальных измерений свойств могут быть «зашумлены», можно использовать правило *k ближайших соседей*: если больше половины из *k* ближайших соседей принадлежат образу *i*, то и объект (точка) *q* относится к образу *i*.

2.2 Метод потенциальных функций

Иногда в «голосовании» принимают участие все реализации обучающей выборки, но с разными весами, зависящими от расстояния. Одним из первых алгоритмов такого рода является алгоритм потенциальных функций. Его смысл заключается в том, что точки каждого класса как бы излучают потенциал, величина которого убывает с расстоянием. Характер убывания может быть самым различным:

$$\frac{1}{r}, \frac{1}{r^2+a}, e^{-\alpha r^2} \dots$$

В точке *q* определяется «притяжение» к каждому из классов.

Если окажется, что какая-то точка распознается с ошибкой, то можно изменить картину потенциального поля. Доказана сходимость алгоритма к оптимальному при увеличении обучающей выборки и конечность числа шагов при не слишком сложной («вычурной») картине потенциального поля.

2.3 Минимизация набора прецедентов

Может оказаться, что памяти ЭВМ недостаточно для хранения всех точек обучающей выборки.

В таком случае можно оставить только точки, удовлетворяющие условию: расстояние от любой точки обучающей выборки до ближайшей «своей» точки меньше, чем до ближайшей чужой. Контрольные реализации распознаются по методу ближайшего соседа.

2.3.1 Алгоритм STOPL

Очевидная сложность заключается в комбинаторном характере задачи - вечный поиск компромисса между требуемой скоростью и затратами памяти. Для сокращения перебора выбирают точки пограничные точки наибольшего риска.

Пусть r_{in} - расстояние до своей ближайшей точки, r_{out} - до чужой. Тогда отношение $W = \frac{r_{in}}{r_{out}}$ характеризует величину риска быть опознаной в качестве чужого образа. Среди точек каждого образа выбираются в качестве прецедентов по одной точке с максимальным значением W .

После этого распознаются все точки обучающей выборки с опорой на прецеденты по методу ближайшего соседа. Среди опознанных неправильно вновь выбирается точка с максимальным значением W и ею пополняется список прецедентов.

Процесс повторяется до тех пор, пока все точки обучающей выборки не будут распознаваться правильно.

2.3.2 Метод дробящихся эталонов - алгоритм ДРЭТ

Стремимся опять же к безошибочному распознаванию обучающей выборки, но для выбора прецедентов используем метод покрытий обучающей выборки каждого образа простыми фигурами (которые можно усложнять при необходимости).

Выберем в качестве покрывающих фигур гиперсферы. Для каждого образа строится гиперсфера минимального радиуса, покрывающая все его точки (реализации). Значения радиусов гиперсфер и расстояния между центрами позволяет определить непересекающиеся гиперсферы. Это - *эталон*ы первого поколения. Если сферы пересекаются, но пересечения пусты, то такие гиперсферы (центры и радиусы) также относятся к эталонам первого поколения. При этом область пересечения считается принадлежащей гиперсфере меньшего радиуса. Эталонны второго поколения строятся только для пересечений, содержащих точки двух или более образов. Если эталонны второго поколения также пересекаются, то процесс продолжается до полной надежности распознавания обучающих выборок. Контрольные образцы классифицируются по попаданию в эталонны гиперсферы. Если контрольная точка не попадает в гиперсферу, то определяется ближайшине и далее используется метод ближайшего соседа.

Для сокращения радиусов сильно «разряженных» гиперсфер по критерию отношения числа точек к радиусу, можно их «таксономировать», используя, скажем, алгоритм FOREL. Не следует только увлекаться.

3 Логические решающие правила. Алгоритм CORAL

В задачах распознавания образов зачастую требуется, чтобы компактные точки в n -мерном пространстве признаков были компактными и несовпадающими в проекциях на координатные оси (гипотеза локальной компактности). Конечно, это не всегда так. Например, все трехмерные геометрические фигуры (сферы, пирамиды, параллелепипеды) могут быть синими. Но здесь мы сталкиваемся с проблемой ситуативной информативности признаков. Очевидно, фигуры с геометрической точки зрения мы не будем классифицировать по цвету.

Обычно, проекции на координатные оси пересекаются, но могут выглядеть по разному и это дает надежду, что комбинация несовпадающих проекций на несколько осей позволит построить эффективное решающее правило за счет сокращения размерности признакового пространства.

Решающие правила, учитывающие разницу в проекциях разных образов имеют вид «Если-условие-то-следствие» и получили наименование логических решающих правил (ЛРП).

Алгоритм CORAL

Выделим подмножество значений $X_{jv} \in X_j$ признака X_j . Для сильных признаков - это интервал значений, для шкал порядка - ряд соседних порядковых позиций, для шкал наименований - одно или несколько имен.

Обозначив факт, что некоторое значение признака X_j объекта a_i принадлежит подмножеству X_{jv} как $J(a_i, X_{jv})$, факт попадания объекта a в область v , образованную границами подмножеств X_{jv} , т.е. в гиперпараллелепипед $\prod_j^{n^{prime}} X_{jv}$, запишем в виде логического высказывания:

$$S(a, X) = \&_{j}^{n'} J(a, X_{jv})$$

$n' \leq n$, где n - размерность признакового пространства. Число n' называется длиной высказывания.

Логической закономерностью называется высказывание, удовлетворяющее двум условиям:

$$P_{ws} = \frac{m_{ws}}{m_w} \leq \alpha \tag{1}$$

$$P_{ws^-} = \frac{m_{ws^-}}{m_w^-} \geq \beta \tag{2}$$

где

w - индекс объектов своего образа,

w^- - индекс всех чужих объектов,

m_w - число всех своих объектов,

m_{w^-} - число чужих объектов,

m_{ws} - число своих, удовлетворяющих высказыванию S ,

m_{ws^-} - число чужих, удовлетворяющих тому же высказыванию S ,

α β - некоторые величины в диапазоне от 0 до 1.

Желательно, чтобы высказывание S отбирало побольше своих объектов и поменьше чужих, т.е. чтобы α было как можно большим, а β - как можно меньшим.

Набор закономерностей называется *покрывающим* для образа w , если для любой его реализации выполняется хотя бы одна закономерность из этого набора. Желательно, чтобы число закономерностей в наборе было минимальным.

Поиск закономерностей начинается с больших значений α (например, $\alpha = 1$) и малых значений $\beta \approx 0.02$.

Просматриваются все подмножества значений первого случайно выбранного признака и находится высказывание S , удовлетворяющее условиям 1 и 2. Если таковое не находится, то процесс поиска продолжается при более низком пороге α вплоть до $\alpha = 0.5$. Если и в этом случае нет результата, то увеличивается β в предельно допустимой доле «чужих среди своих». Если условия 1 и 2 не выполняются и при $\alpha = \beta = 0.5$., то делается переход к рассмотрению второго признака, случайным образом выбранным из оставшихся. Если условия 1 и 2 на каком-то шаге выполняются, то объекты своего образа, удовлетворяющие высказыванию S , из дальнейшего рассмотрения исключаются. Для оставшихся объектов образа w длина высказывания увеличивается на единицу. Процесс продолжается до получения покрывающего набора закономерностей для всех объектов образа w . Аналогично строятся покрывающие наборы и для всех других распознаваемых образов.

Можно потребовать, чтобы алгоритм делал для каждого образа не по одному, а по несколько покрывающих наборов, что соответствует высказыванию Р. Фейнмана о том, что мы можем говорить о понимании явления, если в состоянии объяснить его несколькими способами. С этой целью после получения первого покрытия исключаем из рассмотрения первый признак, включенный в это покрытие, и процесс поиска закономерностей начинается с другого случайно выбираемого признака.

Распознавание контрольного объекта q с помощью покрывающих наборов закономерностей сводится к проверке того, каким высказываниям удовлетворяют его характеристики. Если таких высказываний одно или несколько и все они находятся в списке образа w , то объект q распознается в качестве реализации образа w . Если же объект q удовлетворяет закономерностям нескольких образов, то решение принимается в пользу того образа, которому принадлежит закономерность с наибольшим значением величины P_{ws} .

Анализ общего списка закономерностей может показать, что некоторые признаки из исходной системы X в них отсутствуют. Это означает, что они оказались неинформативными и в процессе принятия решения на них можно не обращать внимания. Для каждого i -го образа подмножество информативных признаков может оказаться разным. А это значит, что при проверке гипотезы о принадлежности объекта к тому или иному образу, нужно анализировать не все пространство признаков, а его информативное подпространство, что хорошо согласуется с интуитивными методами неформального распознавания. (Одних людей мы распознаем по росту, других - по походке, третьих - цвету волос и т.д.)

4 О выборе информативного множества признаков

В главе 3 «Базовые гипотезы» Николай Григорьевич затрагивает одну из важнейших и интереснейших тем (которую он развивает в отдельной главе (6)) - проблему выбора множества информативных признаков.

В условиях, когда законы распределения неизвестны и трудно оценить представительность выборки, не может быть указаний на число признаков, чтобы гарантированно подтверждалась гипотеза компактности.

Представительность выборки и информативность признаков - понятия условные. Обучающая выборка представительна, если при заданном наборе признаков и заданном типе решающих правил удастся построить правило, распознающее объекты с заданной точностью. Система признаков информативна, если при заданной выборке и заданном типе решающих правил удастся сделать то же самое. Оценка информативности того или иного признака производится по количеству ошибок распознавания.

Возможны и другие способы оценки информативности. Из гипотезы компактности следует, что для хорошего распознавания образов желательно, чтобы расстояния между «своими» точками в среднем были меньше расстояний между чужими и расстояния до чужих дочек были больше расстояний до чужих. Если гиперсферы, каждая из которых включает точки только одного образа, пересекаются, то желательно, чтобы их диаметры различались как можно больше.

Компактность (плотность) образа W_i , представленного в обучающей выборке m_i точками $1, 2, \dots, t, \dots, l, \dots, m_i$ можно характеризовать средней длиной ребер полного графа:

$$W_i = \frac{1}{C_{m_i}^2} \sum_{t,l=1}^{m_i} r(t, l)$$

То же мы можем сказать и образе W_j , представленного m_j точками.

Разнесенность образов в пространстве можно охарактеризовать через среднее расстояние между всеми парами точек из разных образов:

$$W(i, j) = \frac{1}{m_i m_j} \sum_{i=1}^{m_i} \sum_{j=1}^{m_j} r(i, j)$$

Тогда, информативность пространства признаков тем больше, чем больше величина

$$J = \frac{W(i, j)}{W_i + W_j}$$

Этот критерий информативности будет хорошо работать для независимых признаков. В противном случае нам нужно будет перебрать и проверить на информативность C_g^m комбинаций для выбора n признаков из g . Регулярно встречаются случаи необходимости перебора 10^{15} . Очевидно, в таких случаях речь о нахождении оптимального решения не идет и разрабатываются различные эвристические алгоритмы.

4.1 Методы последовательного сокращения и добавления признаков DEL и ADD

Допустим, требуется выбрать 25 признаков из 50. Оценим ошибку распознавания (α_0) при использовании всех 50 признаков. Затем, последовательно будем исключать признаки по одному,

каждый раз определяя ошибку распознавания в 49 мерном пространстве. Получим ряд ошибок $(\alpha_{11}, \alpha_{12}, \dots, \alpha_{1i}, \alpha_{50})$, из которого выберем наименьшую α_{1j} и уберем из рассмотрения j -ый признак.

Эту процедуру будем повторять до тех пор, пока не останется 25 признаков.

Количество проверяемых множеств признаков:

$$L = \sum_{j=0}^{g-n} (g - j)$$

и в нашем случае оно равно 900, что на 12 порядков меньше объема полного перебора.

Если предшествующий метода - это метод «сверху вниз», то метод последовательного добавления - ADD - метод «снизу вверх». Начинаем с одномерных признаковов подпространств. Распознаем контрольную последовательность каждым из признаков по отдельности, выбираем признак, дающий минимальную ошибку, к нему последовательно добавляем каждый из $(g - 1)$, выбираем наилучший и т.д.

4.2 Метод случайного поиска с адаптацией (алгоритм СПА)

Единичный отрезок разбивается на g участков одинаковой длины $(1/g)$. Каждому участку сопоставляется свой признак: первому - первый, второму - второй и т.д. Запускается датчик случайных чисел с равномерным распределением в диапазоне 0..1. После n шагов работы выбирается n признаков. По числу ошибок оценивается качество распознавания. Такая процедура продлевается r раз. В итоге получаем список оценок $L = (\alpha_1, \dots, \alpha_r)$. Теперь можно упорядочить список L по возрастанию α , т.е. по убыванию качества распознавания, и ввести систему поощрений и наказаний. Участки, соответствующие признакам, дающим лучший результат, увеличиваются, «худшие» - уменьшаются, но так, чтобы суммарная длина по-прежнему была равна 1.

Испытываем r новых признаковов подсистем, но теперь вероятность попадания на «лучшие» участки выше, чем на плохие.

Продолжаем процесс адаптации таким образом, что длина участков признаков, регулярно попадающих в самые информативные подсистемы, увеличивается на величину $h < 1/g$, а для самых неинформативных длины их участков уменьшаются.

После некоторого количества циклов поиска и адаптации, процесс стабилизируется.

Алгоритм СПА был протестирован для систем, в которых возможен полный перебор сочетаний признаков. Результаты: на одних и тех же примерах алгоритм СПА дает лучшие результаты, чем DEL и ADD.

5 Меры близости между предикатами

В результате исследований в области экспертных систем Н.Г Загоруйко и В.М. Бушуев ввели метрику в пространстве знаний.

Знания, используемые в экспертных системах, часто представлены в виде продукций типа $\forall(X_1, X_2, \dots, X_n)(X_1 \& X_2 \& \dots \& X_n) \Rightarrow A$

При этом значения переменных могут задаваться по-разному:

$$X_1 = 7, X_2 = (2..6), X_10 > 0$$

и.т.п. При этом величина отличия значений предикатов роли не играет, нам же требуется оперировать такими понятиями, как степень похожести (близости), аналогичности, т.е. понятиями, которыми оперирует человек, рассуждая по аналогии. Таким образом, для расширения логических возможностей экспертных систем нужно научиться измерять степень похожести знаний, т.е. ввести метрику в пространстве знаний для измерения расстояний между знаниями. Это сделано следующим образом.

Можно считать, что каждый предикат отражает знание эксперта о распределении возможных значений данной характеристики. Например, утверждение $X_5 = (a|b|c)$ равносильно утверждению, что X_5 с равной вероятностью $\frac{1}{3}$ может принимать одно из трех значений, утверждение $X_4 = (2, 3, 4, 5)$ равносильно утверждению о том, что предикат X_4 с вероятностью 0.25 принимает одно из 4 значений в диапазоне целых чисел 2..5. В таком случае, расстояние между одноименными предикатами можно определять через расстояние между двумя распределениями вероятности.

Предложенная мера для измерения этого расстояния $R = f(r, h, w)$ учитывает расстояние r ото всех элементов одного распределения до всех элементов другого, энтропийную меру h , по смыслу близкую к дисперсии распределений, и степень пересечения распределений w , характеризующую величину согласия («консенсуса»). Аргументы функции вычисляются следующим образом.

5.1 Расстояния r

Если функции плотности вероятности $f_1(x)$, $f_2(x)$ отражают мнения двух экспертов о значении предиката X на участке $x_{min}..x_{max}$, тогда различия двух распределений измеряются величиной:

$$r = \frac{1}{2} \int_{x_{min}}^{x_{max}} |f_1(x) - f_2(x)| dx$$

В дискретном случае разделим ось X , отображающую мнение экспертов о распределении значений предиката, на m частей (квантилей) так, чтобы в каждой части была заключена плотность вероятности, равная $\frac{1}{m}$. Тогда:

$$r = \frac{\sum_{i=1}^m |x_{1i} - x_{2i}|}{x_{max} - x_{min}}$$

Здесь x_{1i}, x_{2i} - правые границы i -х квантилей 1-го и 2-го экспертов.

5.2 «Консенсус»

«Консенсус» вычисляется следующим образом. Интервал разбивается на T равных участков и определяется вероятности P_{1t}, P_{2t} попадания оценок экспертов в t -тую градацию. Тогда:

$$w = \frac{1}{2} \sum_{t=1}^T |P_{1t} - P_{2t}|$$

5.3 Энтропийная мера или категоричность суждений

Расстояния между суждениями зависят и от категоричности (h) их оценок. При одинаковых r и w мера R считается тем большей, чем больше распределения, отражающие мнения двух

экспертов, отличаются от равномерного распределения по всему диапазону значений $x_{min}..x_{max}$. Величина h находится следующим образом:

$$h = \frac{1}{2}(h_1 + h_2)$$
$$h_1 = \frac{1}{2T} \sum_{t=1}^T |P_{1t} - 1|$$
$$h_2 = \frac{1}{2T} \sum_{t=1}^T |P_{2t} - 1|$$

Теперь общая мера расстояния между предикатами, отражающая знания двух экспертов о характеристике X принимается равной $r \times w \times h$. Проверка данной меры экспертным путем показала, что в экспериментах по упорядочиванию распределений использование приведенной меры сохраняет порядок, установленный экспертами.

5.4 Расстояние между знаниями

Зная расстояние между одноименными предикатами, можно построить меру расстояния и между двумя знаниями, выраженными в виде множества предикатов.

Даже если эксперты оперируют не полностью совпадающими наборами характеристик, можно положить, что распределение значений отсутствующих характеристик принимается равномерным в известном диапазоне. В таком случае, расстояние между двумя знаниями, представленными продукциями, состоящими из n предикатов, по аналогии с n мерным евклидовым пространством, находим по формуле:

$$R = \sqrt{R_1^2 + R_2^2 + \dots + R_n^2}$$

Метрика в пространстве знаний позволяет формулировать и решать на материале базы знаний те же задачи, которые обычно решаются на материале базы данных аналогичными методами.

Следующая лекция

[Лекция 14. Нейрокомпьютер. Введение](#)