Методическая разработка

для проведения лекции

Занятие 10. Особенности обучения систем распознавания

Учебные вопросы занятия:
1. Постановка задачи и методы обучения
2. Обучение перцептрона
3. Построение эталонных описаний одномерных нормальных совокупностей

Заключительная часть

Введение

Важной особенностью современного этапа развития систем телекоммуникации и глобальных вычислительных сетей является рост объема, разнообразия и рассредоточения информационных потоков по многим сетям и каналам связи. Качественное выполнение задач обеспечения информационной безопасности в этих условиях требует решения, в первую очередь, проблемы автоматизации процесса классификации источников информации и идентификации информации, осуществляемой в условиях априорной неопределенности. При этом объективной реальностью является наличие проблем распознавания образов.

1. Постановка задачи и методы обучения

Распознавание представляет собой отнесение исследуемого объекта, задаваемого в виде совокупности наблюдений (или образа) к одному из классов. Источником информации о распознаваемых объектах является совокупность результатов независимых наблюдений, составляющих обучающую выборку $x_i^N = (x_1, x_2 .. x_N)$, причем x_i может быть одномерной, либо *p*-мерной случайной величиной

$$X = \begin{vmatrix} x_{11} & x_{12} & \dots & x_{1N} \\ x_{21} & x_{22} & \dots & x_{2N} \\ - & - & - & - \\ x_{p1} & x_{p2} & \dots & x_{pN} \end{vmatrix} \qquad (1)$$

где каждый столбец представляет собой *p*-мерный вектор наблюдений значений признаков $x_1, x_2 \dots x_p$.

Обучение системы является неотъемлемой составной частью распознающего процесса и имеет своей конечной целью формирование эталонных описаний классов, форма которых представляется способом их использования в решающих правилах. Иными словами, обучение служит для снятия априорной неопределенности о классах распознавания.

Если в результате предварительного анализа наблюдаемой совокупности выборочных значений оказывается возможным хотя бы с некоторым приближением установить вид закона их распределения, то априорная неопределенность относится лишь к параметрам этого распределения, так что целью обучения в этом случае становится получения оценок этих параметров. Подобная априорная неопределенность носит название параметрической, а методы распознавания, применяемые в этих условиях, именуются параметрическими.

Если в распознающую систему вводится классифицированные обучающие выборки и разделяющие функции (пороги), то такой вид параметрического обучения носит название «обучение с учителем». Другой вид параметрического обучения носит название – «обучение без учителя». При таком обучении в распознающую систему вводится только вид условных плотностей распределения, а оценка параметров и разделяющих функций (порогов) осуществляется самой системой.

В наиболее общем случае отсутствия априорных сведений не только о параметрах, но и самом виде закона распределения наблюдаемой совокупности выборочных значений, априорная неопределенность носит название непараметрической, а сами методы обучения именуются непараметрическими, а обучение становится самообучением. В основном для самообучения используют два подхода: детерминистский и статистический. Статистические алгоритмы выражают попытку аппроксимации плотностей распределения параметров различными методами (стохастическая аппроксимация по алгоритму Робинса-Монро, алгоритм корректирующих приращений, метод потенциальных функций и т.д.). Сами принципы распознавания основаны на байесовском классификаторе. Главной проблемой статистических алгоритмов является медленная сходимость операторных оценок $\widehat{\omega}_m\left(x/S_i\right)$ к априорной вероятности $\omega_m\left(x/S_i\right)$ при $N \to \infty$, т.е. достаточная точность оценивания обеспечивается при значительных объемах выборки. При произвольном N достоверность оценивания низкая, а полигональные и гистограммные оценки могут привести даже к отрицательному результату.

Детерминистский подход служит основой алгоритмов, которые конструируются независимо от каких-либо предположений о статистических свойствах классов образов.

Рассмотрим суть параметрического обучения.

Хотя с формальной точки зрения закон распределения выборочных значений может быть любым, на практике в параметрическом распознавании почти всегда используется нормальный закон. Дело в том, что если при распознавании одномерных совокупностей, их распределение всегда может быть описана одним законом (например, нормальным, биноминальных, пуассоновским и д.р.), то при распознавании многомерных совокупностей каждая компонента выборочных значений может иметь свой отличный от других компонент закон распределения, что в принципе не является аномалией. Но тогда многомерное совместное распределение выборочных значений должно описываться некоторым многомерным законом, включающим в себя компоненты с различными законами распределения. Как указано в ряде литературных источников (например «Многомерный статистический анализ и временные ряды», авторы М. Кендалл, А. Стьюарт) современный уровень знаний таков, что пока точному многомерному анализу поддаются лишь задачи, где рассматривается нормальный случай. Отсюда следует, что на сегодняшний день параметрические методы распознавания по существу являются методами распознавания нормально распределенных совокупностей и задачи обучения в этих случаях является оценивание параметров (средних значений, дисперсий, ковариационных матриц) нормальных плотностей вероятностей и разделяющих функций (порогов), используемых в решающем правиле:

$$\lambda_{ij} = \frac{\omega_n\left(x_1, x_2, \ldots, x_n / A_i\right)}{\omega_n\left(x_1, x_2, \ldots, x_n / A_j\right)} \geq \Theta_{ij}, \qquad (2)$$

где λ_{ij} – отношение правдоподобия,

$\omega_n(..)$ – условная функция правдоподобия,

Θ_{ij} – порог принятия решения.

Необходимым и достаточным условием применения параметрического обучения является подтверждение гипотезы о нормальности выборки.

В статистической литературе параметрическому обучению уделяется очень много внимания, поэтому мы поясним ее только в той мере, насколько это необходимо для рассмотрения методов оценок вектора средних значений и ковариационной матрицы характеризующих некоторую совокупность образов. Образы выбранные из совокупности с нормальным распределением проявляют тенденцию к образованию одного кластера, центр которого определяется вектором средних значений, а форма ковариационной матрицей.

Выборочный вектор средних вычисляется по формуле

$$\tilde{\vec{M}} = \frac{1}{N}\sum_{i=1}^{N} X_i^k = \begin{pmatrix} m_{k1} \\ : \\ m_{kp} \end{pmatrix} \qquad (3)$$

где p – размерность признакового пространства;

N – объем выборки;

k – номер класса.

Выборочная ковариационная матрица (несмещенная оценка) вычисляется по формуле

$$\tilde{C}_k = \frac{1}{N-1}\sum_{i=1}^{N}\left(x_i^k - \hat{a}_k\right)\left(x_i^k - \hat{a}_k\right)^T \qquad (4)$$

Вопрос: Дайте физическую интерпретацию этих понятий.

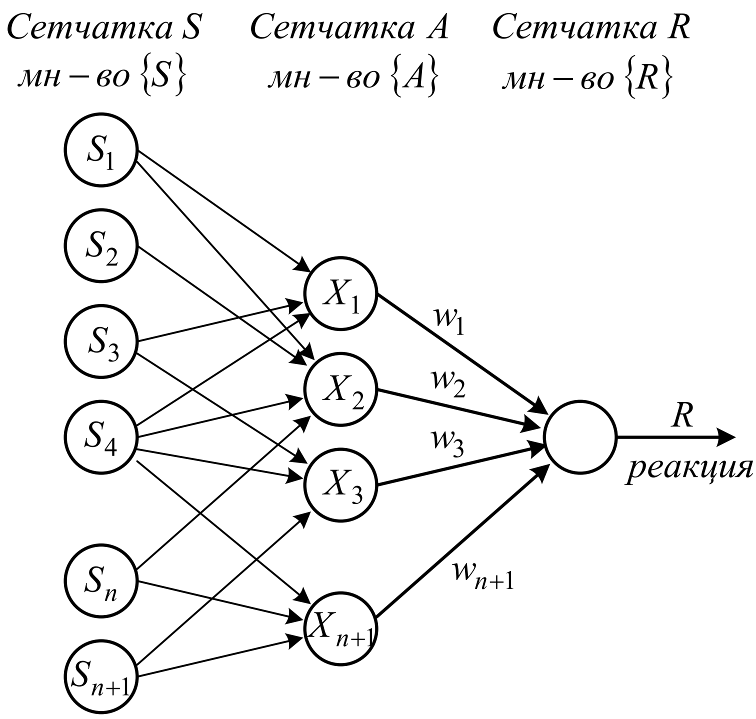
За исключением случаев, когда матрица $\tilde{C}$ оказывается вырожденной и, следовательно, перестает существовать обратная матрица $\tilde{C}^{-1}$, необходимо наложить условия $N > p$. Для одномерного случая:

$$\tilde{m}_k = \frac{1}{N}\sum_{i=1}^{N} x_i\,, \qquad \hat{\sigma}_{k^2} = \frac{1}{N-1}\sum_{i=1}^{N}\left(x_i - \tilde{m}_k\right)^2 \qquad (5)$$

2. Обучение перцептрона

Сущность детерминистского метода обучения рассмотрим на примере перцептронного подхода, предложенного Розенблаттом в 1957 году, и представляющего собой естественную и обладающую большими возможностями модель процесса обучения машины. В принципе алгоритм перцептрона является одним из представителей интерактивных процедур, которые можно легко строить с помощью градиента.

Основная модель перцептрона, обеспечивающего отнесения образа к одному из двух заданных классов, приведена на рисунке.



Вариант модели перцептрона

Устройство состоит из множества S сенсорных элементов, которые случайным образом соединены с ассоциативными элементами второго множества A. Каждый из элементов второго множества воспроизводят выходной сигнал в случае, если достаточное количество сенсорных элементов, соединенных с его входом, находится в возбужденном состоянии. Реакция всей системы пропорциональна сумме взятых с определенными весами реакций элементов ассоциативного множества, т.е.

$$R = \sum_{i-1}^{n+1} w_i x_i = W^T X \qquad (6)$$

где x_i - реакция i - го ассоциативного элемента;

w_i - соответствующий весовой элемент.

Если $R > 0$, значит предъявленный системе образ принадлежит классу ω_1, если $R < 0$, то наоборот классу ω_2.

Эту схему легко распространить на случай разделения на несколько классов, добавляя соответствующее число реагирующих элементов в R-сетчатки. Классификация проводится обычным способом: рассматриваются значению реакций $R_1, R_2, \ldots, R_n$ и образ причисляется классу ω_j, если $R_j > R_i$ для всех $j \neq i$.

Основную модель можно также легко распространить на случай нелинейных решающих функций введением соответствующих нелинейных преобразований между сетчатками A и R.

Обучение перцептрона сводится к простой схеме интерактивного определения вектора весов W. Пусть задано два обучающих множества представляющих классы ω_1 и ω_2 соответственно, пусть $W(1)$ - начальный вектор весов, который выбирается произвольно. В таком случае k-й шаг обучения выглядит следующим образом:

Если $X(k) \in \omega_1$ и $W^T(k)X(k) \leq 0$, то вектор весов заменяется вектором $W(k+1) = W(k) + cX(k)$, где c - корректирующее приращение.

Если $X(k) \in \omega_2$ и $W^T(k)X(k) \geq 0$, то $W(k)$ заменяется вектором $W(k+1) = W(k) - cX(k)$.

В противном случае $W(k)$ не изменяется, т.е.

$$W(k+1) = W(k) \qquad (7)$$

Очевидно, что алгоритм перцептрона является процедурой типа “подкрепление - наказание”, причем здесь подкреплением в сущности является отсутствием наказания. В чистом виде этот алгоритм применен в кодеке АДИКМ рекомендации МКТТ G.721 и носит название упрощенного градиентного метода. Сходимость алгоритма наступает при правильной классификации всех образов с помощью некоторого вектора весов. Если заданные классы линейно разделимы, то алгоритм перцептрона сходится на конечное число итераций. В тех ситуациях, когда разделимость отсутствует, эти алгоритмы зацикливаются и работают в таком режиме до тех пор, пока их выполнение не прервется извне. Однако можно избежать этого, если применить, например, НСКО – алгоритм или его еще называют алгоритм Хо-Кашьяпа.

Вывод: Если попытаться подвести некоторый итог проведенному краткому анализу методов обучения, то можно отметить, что все методы обучения можно классифицировать на параметрические и непараметрические. Значительное развитие получило первое направление. Непараметрическое распознавание, в том числе и обучение в основном базируется сейчас на статистическом подходе путем оперативной аппроксимации оценок плотностей вероятностей, но требует больших объемов обучающей выборки. Детерминистский подход при построении систем обнаружения компьютерных атак находится в начале своего развития, однако является перспективным, особенно при внедрении современных ЭВМ с большими вычислительными возможностями.

3. Построение эталонных описаний одномерных нормальных совокупностей.

При нормальном распределении признака построения эталонных описаний классов существенно упрощается, поскольку при наличии классифицированных обучающих выборок вместо трудоемких и сложных процедур формирования оценок условных плотностей вероятностей достаточно вычислить выборочное среднее и дисперсию по обучающим выборкам $\left\{x_i^{(k)}\right\}$

$$\tilde{m}_k = \frac{1}{N}\sum_{i=1}^{N} x_i^{(k)}, \tag{8}$$

$$\tilde{\sigma}_k^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(x_i^{(k)} - \tilde{m}_k\right)^2. \tag{9}$$

Таким образом, построение одномерных эталонных описаний нормальных совокупностей сводится к получению выборочных средних и дисперсий, являющихся несмещенными и состоятельными оценками соответственно среднего и дисперсии, и к подстановке их в известное выражение нормального закона распределения, которое и представляет собой эталонное описание класса.

б) Рассмотрим теперь построение эталонных описаний при *неклассифицированных* обучающих выборках, где оказывается необходимым применения методов самообучения.

Рассмотрим достаточно общий случай M равновероятных нормально распределенных классов $A_1, A_2, \ldots, A_M$ с неизвестными средними значениями $\tilde{m}_1, \tilde{m}_2, \ldots, \tilde{m}_k$ и неизвестной одинаковой дисперсией $\tilde{\sigma}^2$. Неклассифицированная обучающая выборка извлекается из суммарного M – модального распределения

$$\omega\left(x / \tilde{m}_1, \ldots, \tilde{m}_M, \tilde{\sigma}^2\right) = \frac{1}{\sqrt{2\pi}\tilde{\sigma}} \sum_{i=1}^{M} e^{-\frac{(x-\tilde{m}_i)^2}{2\sigma^2}} \tag{10}$$

Для нахождения выборочных средних и выборочной дисперсии целесообразно использовать метод *моментов Пирсона,* основанный на приравнивании неизвестных значений параметров распределения выборочным моментам и решении соответствующей системы уравнений.

Поскольку необходимо определить $M+1$ неизвестных $\tilde{m}_1, \ldots, \tilde{m}_M, \hat{\sigma}^2$, требуется составить систему из $(M+1)$ уравнений с указанными $(M+1)$ неизвестными:

$$\begin{cases} \dfrac{1}{M}\left(\tilde{m}_1 + \tilde{m}_2 + \ldots + \tilde{m}_M\right) = \dfrac{1}{N}\sum\limits_{i=1}^{N} x_i \\ \dfrac{1}{M}\left(\tilde{m}_1^2 + \tilde{m}_2^2 + \ldots + \tilde{m}_M^2\right) + \tilde{\sigma}^2 = \dfrac{1}{N}\sum\limits_{i=1}^{N} x_i^2 \\ \vdots \\ \dfrac{1}{M}\left(\tilde{m}_1^{M+1} + \tilde{m}_2^{M+1} + \ldots + \tilde{m}_M^{M+1}\right) + \hat{\sigma}^2 = \dfrac{1}{N}\sum\limits_{i=1}^{N} x_i^{M+1} \end{cases} \tag{11}$$

Эта система может быть решена различными известными способами. Например, можно последовательно исключить неизвестные из системы.

Так, разрешаем первое уравнение относительно $\tilde{m}_1$:

$$\tilde{m}_1 = \frac{M}{N}\sum_{i=1}^{N} x_i - \left(\tilde{m}_2 + \ldots + \tilde{m}_M\right) \tag{12}$$

И подставляя полученное таким образом значение во второе уравнение, исключаем первое неизвестное и т.д.

После повторения указанной операции M раз остается решить одно уравнение с одним неизвестным $\tilde{\sigma}^2$. Решая его и повторяя весь процесс в обратном порядке, находя последовательно после $\tilde{\sigma}^2$ неизвестные $\tilde{m}_M, \tilde{m}_{M-1}, \ldots, \tilde{m}_1$ решаем систему уравнений и определяем таким образом все

$M+1$ неизвестных. Ясно, что при большом количестве классов M в процессе решения системы уравнений может оказаться достаточно трудоемким.

В качестве примера рассмотрим простейший случай двух равновероятных нормально распределенных классов с известной дисперсией $\tilde{\sigma}^2$ и неизвестными средними значениями $\tilde{m}_1$ и $\tilde{m}_2$. В этом случае достаточно составить систему из двух уравнений с двумя неизвестными $\tilde{m}_1$ и $\tilde{m}_2$, то есть взять первые два уравнения выше описанной системы:

$$\frac{\tilde{m}_1+\tilde{m}_2}{2}=\frac{1}{N}\sum_{i=1}^{N}x_i=\bar{a}_1 \qquad (13)$$

$$\frac{\tilde{m}_1^2+\tilde{m}_2^2}{2}+\tilde{\sigma}^2=\frac{1}{N}\sum_{i=1}^{N}x_i^2=\bar{a}_2 \qquad (14)$$

где $\bar{a}_1$ – выборочное среднее, $\bar{a}_2$ – второй выборочный начальный момент суммарного бимодального распределения.

Решая первое уравнение системы относительно $\tilde{m}_2$ и подставляя полученное выражение во второе уравнение, а затем, повторяя весь процесс относительно $\tilde{m}_1$ получаем два симметричных квадратных уравнения для каждого из неизвестных $\tilde{m}_1$ и $\tilde{m}_2$

$$\begin{cases}\tilde{m}_1^2+2\bar{a}_1\tilde{m}_1+2\bar{a}_1^2+\tilde{\sigma}^2-\bar{a}_2=0\\ \tilde{m}_2^2+2\bar{a}_1\tilde{m}_2+2\bar{a}_1^2+\tilde{\sigma}^2-\bar{a}_2=0\end{cases} \qquad (15)$$

Корни полученных уравнений представляют собой значения неизвестных $\tilde{m}_1$ и $\tilde{m}_2$, например

$$\tilde{m}_1=\bar{a}_1\pm\sqrt{\bar{a}_2-\bar{a}_1^2-\tilde{\sigma}^2}\,;\ \tilde{m}_2=2\bar{a}_1-\tilde{m}_1. \qquad (16)$$

Заключительная часть.

Таким образом, построение эталонных описаний в рассматриваемом случае параметрического обучения сводится к выражению значений неизвестных параметров через выборочные моменты $\bar{a}_1,\bar{a}_2,...,\bar{a}_M$, которые легко вычисляются по обучающей неклассифицированной выборке, и подстановке их в суммарное M-модальное распределение.

Рекомендованная литература:

1. Горелик А.Л., Скрипкин В.А. Методы распознавания. – М.: Высшая школа, 1989.

2. Фомин Я.А., Тарловский Г.Р. Статистическая теория распознавания образов. – М.: Радио и связь, 1986.

3. Ту Дж., Гонсалес Р. Принципы распознавания образов. – М.: Мир, 1978.