

ЭМПИРИЧЕСКОЕ ОБОБЩЕНИЕ И РАСПОЗНАВАНИЕ: КЛАССЫ ЗАДАЧ, КЛАССЫ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ И ПРИМЕНИМОСТЬ ТЕОРИЙ. ЧАСТЬ II

© Донской В. И.

ТАВРИЧЕСКИЙ НАЦИОНАЛЬНЫЙ УНИВЕРСИТЕТ ИМ. В.И. ВЕРНАДСКОГО

ФАКУЛЬТЕТ МАТЕМАТИКИ И ИНФОРМАТИКИ

ПР-Т ВЕРНАДСКОГО, 4, Г. СИМФЕРОПОЛЬ, 95007, УКРАИНА

E-MAIL: *donskoy@tnu.crimea.ua*

Abstract. Classification of pattern recognition problems is offered. This classification is founded on the basic properties of pattern recognition problems. It is shown, a choice of methods of decisions must be coordinated with features of classes of pattern recognition problems.

1. КОРРЕКТНЫЕ АЛГОРИТМЫ И ЗАДАЧИ РАСПОЗНАВАНИЯ ИЗ КЛАССА $D/ - /R_1T/AS \vee LS/-$

В первой части статьи [1] рассматривались задачи обучения распознаванию в детерминистской постановке и предполагалось, что выбор решающего правила $f : \mathbf{X} \rightarrow \{0, 1\}$ осуществляется из некоторого семейства S и отождествляется с нахождением точного решения системы

$$\begin{cases} f(X_1) = \alpha_1; \\ f(X_2) = \alpha_2; \\ \dots\dots\dots \\ f(X_l) = \alpha_l; \\ f \in S, \end{cases} \quad (1)$$

Здесь $\omega : \mathbf{X} \rightarrow \{0, 1\}$ — отображение объектов генеральной совокупности \mathbf{X} в значения их основного свойства: $\alpha_q = \omega(X_q)$; $X_q \in \mathbf{X}$; $\alpha_q \in \{0, 1\}$; Отображение ω порождает столбец значений этого основного свойства в правой части системы (1) и отражает в таблице обучения $\{(X_q, \tilde{\omega}(X_q)), q = 1, \dots, l\}$ закономерность, в соответствии с которой объекты из генеральной совокупности обладают или не обладают основным свойством. Следует подчеркнуть, что в теории машинного обучения и распознавания представляют интерес именно те задачи, в которых закономерность ω существует и отличается от случайной функции, равномерно распределяющей значения $\{0; 1\}$ на множестве \mathbf{X} . В этом смысле закономерность — это неслучайность, и в правой части системы (1) должны содержаться не какие угодно столбцы, а именно те, которые связаны с объектами выборки некоторой закономерностью.

Если извлеченная из генеральной совокупности в соответствии с моделью R_1T выборка является безошибочной, то выбор в процессе обучения правила распознавания, которое допускает ошибки на этой выборке, представляется бессмысленным. Поэтому в этом случае требуется найти точное решение f^* функциональной системы (1) в некотором классе решений S . Это решение f^* называется *точной настройкой на выборку*. Если система имеет более чем одно решение $f^* \in S$, то при достаточно широком семействе S выбранное правило может давать большую ошибку

$$Er(f^*) = \int_{X \in \mathbf{X}} |f^*(X) - f_0(X)| dP(X),$$

определяемую отличием выбранного решения f^* от существующего истинного правила классификации f_0 по вероятностной мере $P(\cdot)$ на генеральной совокупности \mathbf{X} . Мера $P(\cdot)$, если она существует, оценивает вероятность появления объектов $X \in \mathbf{X}$. Если такая мера не существует, не имеет смысла для некоторых задач, то ошибка точного на выборке решения f^* в некоторых случаях может иметь место почти всюду на \mathbf{X} . Поэтому для того, чтобы выполнялось равенство $f^* = f_0$ при условии, что выборка — безошибочная, решение системы (1) должно быть единственным. Действительно, если существуют два решения этой системы f_1^* и f_2^* — функции, совпадающие на выборке \tilde{X}_l , — то их продолжения на \mathbf{X} могут различаться почти всюду (например: $f_2^*(X) = f_1^*(X)$, если $X \in \tilde{X}_l$, и $f_2^*(X) = \bar{f}_1^*(X)$, если $X \in \mathbf{X} \setminus \tilde{X}_l$; \bar{y} — инверсия значения $y \in \{0, 1\}$).

Здесь и в дальнейшем *продолжением* выборки будем называть любую последовательность объектов из \mathbf{X} , не содержащую элементов этой выборки.

Представим теперь, что в обучающей информации появились ошибки, которые привели к изменению только столбца $\tilde{\alpha} = (\alpha_1, \dots, \alpha_l)^T$ в системе (1) и превращению его в столбец с ошибками $\tilde{\alpha}^E$. Пусть система

$$\left\{ \begin{array}{l} f(X_1) = \alpha_1^E; \\ f(X_2) = \alpha_2^E; \\ \dots\dots\dots \\ f(X_l) = \alpha_l^E; \\ f \in S, \end{array} \right. \quad (2)$$

также имеет решение f^E . Но тогда $f^E \neq f_0$, и в таком случае представляется абсурдной точная настройка алгоритма обучения на выборку. Это приводит к следующему выводу: для любой выборки \tilde{X}_l , в которой отражена некоторая закономерность, должны существовать такие двоичные столбцы $\tilde{\alpha}^E$, что по таблице обучения $(\tilde{X}_l, \tilde{\alpha}^E)$

точная настройка является невозможной. В связи с этим представляется важной теорема 2 из первой части статьи, которую можно усилить до необходимого и достаточного условия существования таких столбцов:

Теорема 1. Пусть в задаче обучения распознаванию решающее правило выбирается из семейства S , и обучающая информация представляется в виде $(\tilde{X}_l, \tilde{\alpha})$. Тогда для любой выборки \tilde{X}_l булевский набор $\tilde{\alpha}$ такой, что точная настройка невозможна, найдется если и только если $VCD(S) < l$.

Как и в первой части этой статьи [1], ёмкость класса S решающих функций (размерность Вапника-Червоненкиса [2]) обозначается $VCD(S)$.

Будем говорить, что найдено точное решение задачи, если реализован выбор $f^* = f_o$. В этом случае решение f^* системы часто называют *корректным алгоритмом*. Напомним, что в работах Ю. И. Журавлева корректными называются алгоритмы, безошибочно распознающие элементы любой заданной контрольной выборки [3]. Целесообразно уточнить используемые ниже понятия, связанные с корректностью алгоритмов.

Определение 1. Решающее правило (алгоритм) называется *корректным* (на выборке), если оно является точным решением системы (1). Решающее правило (алгоритм) называется *абсолютно точным решением* задачи обучения распознаванию, если позволяет безошибочно определить основное свойство $\alpha = f_o(X)$ для любого объекта X генеральной совокупности \mathbf{X} .

Следствие 1. Корректный на выборке длины l алгоритм, выбранный из семейства S , имеющего ёмкость $VCD(S) \geq l$, может давать ошибку почти всюду на генеральной совокупности объектов \mathbf{X} .

Нужно подчеркнуть, что алгоритмы всюду в этой статье рассматриваются с точностью до классов функциональной эквивалентности. Иначе говоря, одним и тем же считаются все алгоритмы (машины Тьюринга), которые для одной и той же начальной информации (слова на ленте) всегда выдают один и тот же результат.

Теорема 2. Для того, чтобы выбор корректного на выборке алгоритма из заданного семейства S в задаче обучения распознаванию из класса $D / - / R_1 T \vee ST / - / -$ всегда обеспечивал получение абсолютно точного решения, необходимо и достаточно, чтобы в семействе S для любой выборки существовал единственный с точностью до функциональной эквивалентности корректный на этой выборке алгоритм.

Доказательство. Необходимость. Действительно, если для какой-нибудь выборки в семействе S не существует корректного на ней алгоритма, то некоторое число объектов этой выборки всегда классифицируется неверно. Если же корректность на выборках достигается всегда, но хотя бы для одной выборки — не единственным алгоритмом из S , а хотя бы двумя неэквивалентными алгоритмами, то их продолжения на множестве последовательностей из X не будут совпадать. Тогда хотя-бы один из них будет давать ошибки на своем продолжении.

Достаточность. Если выбор корректного на выборке алгоритма из заданного семейства S не всегда обеспечивает получение абсолютно точного решения, то для некоторой выборки существует корректный алгоритм, не являющийся абсолютно точным решением. Зафиксируем эту выборку. Постановка задачи предполагает существование абсолютно точного решения. Это точное решение — некоторое правило f_o — также будет корректным на зафиксированной (безошибочной в соответствии с рассматриваемой моделью) выборке. Тогда корректное на ней решение — не единственно. \square

В задачах обучения распознаванию предполагается использование только конечных обучающих выборок любой длины. Очевидно, что если в семействе S для любой конечной выборки длины l существует единственный с точностью до функциональной эквивалентности корректный на этой выборке алгоритм, то $VCD(S) < l$.

Следствие 2. *Для того, чтобы выбор корректного на выборке длины l алгоритма из заданного семейства S в задаче обучения распознаванию из класса $D/ - /R_1T \vee ST/ - /-$ всегда обеспечивал получение абсолютно точного решения, необходимо выполнение условия $VCD(S) < l$.*

Но условие $VCD(S) < l$ не является достаточным. Это сразу же видно из случая, когда при его выполнении семейство S не содержит в себе истинного решающего правила f_o . Очевидно также, что условие $f_o \in S$ является необходимым для осуществления возможности нахождения абсолютно точного решения.

2. ОБУЧЕНИЕ ИЛИ НАСТРОЙКА?

Детерминистская постановка задач обучения распознаванию предполагает существование точного решения f_o (истинного решающего правила), согласно которому каждой выборке \tilde{X}_l должен сопоставляться единственный булевый вектор $\tilde{\alpha}^*$, в котором $\alpha_i^* = f_o(X_i), i = 1, \dots, l$. Пары $\langle \tilde{X}_l, \tilde{\alpha}^* \rangle$ для каждой выборки \tilde{X}_l определяют закономерность (регулярность), выделяющую вектор $\tilde{\alpha}^*$ из всех остальных $2^l - 1$ соответствий $\langle \tilde{X}_l, \tilde{\alpha} \rangle$. Выше установлено, что для устранения неоднозначности и

обеспечения возможности получить точное решение задачи обучения распознаванию необходимо и достаточно чтобы существовала возможность нахождения точного решения системы (1), и это решение должно быть единственным. При этом обязательно должно выполняться емкостное ограничение $VCD(S) < l$, поскольку в противном случае условие единственности корректного алгоритма для любой выборки длины l немедленно нарушается.

Теперь можно рассмотреть вопрос об *отличии обучения от настройки*. Этот вопрос представляется важнейшим в теории обучения распознаванию.

Если не ограничивать емкость используемого для решения задачи обучения распознаванию семейства S , то всегда можно добиться точной настройки на непротиворечивую начальную информацию (достаточно представить абсурдно-тривиальный пример использования решающего правила как суммы характеристических точек выборки любой конечной длины). В таком случае ни о каком обучении говорить не приходится.

Будем называть *обучением «снизу-вверх»* такой последовательный процесс построения по заданной обучающей выборке решающего правила \hat{f}_0 , на каждом этапе которого происходит минимальное необходимое усложнение \hat{f}_0 , обеспечивающее уменьшение ошибок решающего правила на обучающей выборке. При обучении происходит поэтапное усложнение решающего правила и, соответственно, расширение семейства S , которому оно принадлежит. Выбор начального приближения и способы поэтапного усложнения решающего правила определяют алгоритм обучения.

Для обоснования алгоритма обучения «снизу-вверх» целесообразно приводить теорему о возможности расширения (в процессе выполнения именно этого алгоритма) семейства правил, которому принадлежит вычисляемое решающее правило, до некоторого семейства S_0 , содержащего истинное решающее правило f_0 и имеющего ёмкость $VCD(S_0) < l$, где l — длина обучающей выборки.

Обучением «сверху-вниз» будем называть последовательный процесс нахождения решающего правила \hat{f}_0 , принадлежащего некоторому подклассу минимальной сложности S' из выбранного изначально некоторым способом семейства S , направленный на достижение наибольшей точности правила \hat{f}_0 на заданной обучающей выборке.

Для обоснования алгоритма обучения по методу «сверху-вниз» целесообразно приводить теорему об адекватности изначально заданного семейства — наличию истинного решающего правила в начальном семействе S — и сохранении этого свойства при поэтапном сужении начального семейства.

Комбинированным обучением будем называть процесс построения решающего правила, сочетающий оба метода обучения — «снизу-вверх» и «сверху-вниз». Такой процесс аналогичен поиску с возвратом.

Процесс нахождения решения системы (1), отличающийся от обучения, будем называть *настройкой*.

3. ОСОБЕННОСТИ КЛАССА $D/-/R_2F \vee SF/-/-$

Если выборка содержит ошибки, то можно считать, что их появление связано с «искажением» правильной выборки, или, говоря иначе, с переходом от правильной — к ошибочной выборке. Будем обозначать такой переход следующим образом: $\langle \tilde{X}_l, \tilde{\alpha} \rangle \xrightarrow{Er} \langle \tilde{X}_l^*, \tilde{\alpha}^{Er} \rangle$. Переход $\tilde{X}_l \xrightarrow{Er} \tilde{X}_l^*$ можно рассматривать как изменение набора точек выборочного пространства в пределах допустимого множества и говорить о безошибочной паре $\langle \tilde{X}_l^*, \tilde{\alpha}^* \rangle$, переходящей в пару $\langle \tilde{X}_l^*, \tilde{\alpha}^{Er} \rangle$. Обозначим $\Delta(Er) = \|\tilde{\alpha}^* - \tilde{\alpha}^{Er}\|$ — число ошибок в векторе $\tilde{\alpha}^{Er}$. Если условие теоремы 1 не выполняется, то из семейства S может быть выбрано решающее правило f^{Er} , точно настроенное на выборку $(\tilde{X}_l^*, \tilde{\alpha}^{Er})$ (эмпирическая ошибка при этом — нулевая: $\hat{\varepsilon}^* = 0$), но истинная ошибка этого правила $\varepsilon = \varepsilon(f^{ER}) \geq \Delta(Er)$.

Если выполняется условие теоремы 2, — точная настройка на ошибочную выборку (пару $\langle \tilde{X}_l^*, \tilde{\alpha}^{Er} \rangle$) будет невозможна. Настроиться точно можно будет только на безошибочную выборку.

Пусть \mathfrak{X} — такое подмножество элементов обучающей выборки, что их удаление из этой выборки позволяет осуществить точную настройку, но удаление никакого собственного подмножества $\mathfrak{X}' \subset \mathfrak{X}$ из выборки уже не позволяет настроиться точно. Будем называть такой набор детерминированной помехой. Из её определения усматривается переборный алгоритм фильтрации (удаления) \mathfrak{X} из обучающей выборки по критерию возможности точной настройки.

Пусть Δ_{max} — наибольшее число ошибок $\Delta(Er)$, порождаемых переходом $\langle \tilde{X}_l, \tilde{\alpha} \rangle \xrightarrow{Er} \langle \tilde{X}_l^*, \tilde{\alpha}^{Er} \rangle$, является изначально заданным параметром задачи. Процесс обучения может состоять из следующих последовательно решаемых подзадач:

1° Выбор адекватного начального семейства S , которое должно содержать истинное решающее правило f_o , для реализации алгоритма фильтрации.

2° Выполнение алгоритма фильтрации (глубина перебора определяется параметром Δ_{max}).

3° Собственно обучение одним из описанных выше методов.

Таким образом, для решения задач из класса $D/-/R_2F \vee SF/-/-$ целесообразно применять корректные алгоритмы.

4. ОСОБЕННОСТИ КЛАССА $ND/D_k/ - / - / -$

Класс недетерминированных задач обучения распознаванию характеризуется тем, что не имеется никакой информации о законах, определяющих существование и появление той или иной выборки \tilde{X}_l — последовательности элементов из \mathbf{X} . Более того, неизвестно: существует ли точное решение f_o или нет.

Полагается, что недетерминированные и стохастические задачи обучения распознаванию принципиально различаются. Информация о существовании вероятностных распределений или более — об их типах — в стохастических задачах в некоторых случаях может дать возможность в явном виде выписать статистически оптимальное решающее правило. В стохастических задачах речь идет о решениях, получаемых с точностью, определяемой заданием вероятностных мер.

Рассмотрим следующую задачу. Пусть \mathbf{X}^{MT} — множество шифров машин Тьюринга. $X(M) \in \mathbf{X}^{MT}$ — шифр машины M — является натуральным числом, которое, в частности, может быть представлено двоичной строкой конечной длины. Машина Тьюринга M называется самоприменимой, если, начав работу над словом $p = X(M)$, являющимся шифром этой машины M , она остановится, выполнив конечное число шагов.

Обозначим решающую функцию, которую должен найти алгоритм обучения, следующим образом:

$$f_o^{sa}(X) = \begin{cases} 1, & \text{если машина } M \text{ самоприменима;} \\ 0, & \text{если машина } M \text{ несамоприменима.} \end{cases}$$

Известно, что данная функция f_o^{sa} не является вычислимой — не существует алгоритма (строго определенного тезисом Черча-Тьюринга), правильно вычисляющего для любого входа $X(M)$ значение $f_o^{sa}(X(M))$. Тем не менее, можно сконструировать обучающую выборку $\mathfrak{M}^{sa} = \mathfrak{M}_1 \cup \mathfrak{M}_o$, состоящую из $|\mathfrak{M}_1| = m_1$ примеров шифров самоприменимых машин Тьюринга и $|\mathfrak{M}_o| = m_o$ примеров несамоприменимых машин. Что же будет, если выборку \mathfrak{M}^{sa} предоставить как начальную информацию — таблицу обучения — для построения алгоритма распознавания свойства самоприменимости? Такого алгоритма в принципе не существует. Тем не менее, алгоритм обучения, выбранный из подходящего для данной задачи семейства и имеющий достаточную ёмкость, может дать в качестве решения частичную функцию \hat{f}_o^{sa} , которая безошибочно классифицирует все примеры выборки \mathfrak{M}^{sa} .

Приведенный пример принадлежит классу недетерминированных задач: неизвестно, существует ли вообще правильное решение (в данном примере — не существует алгоритмического правильного решения), и неизвестно, существует ли какой-нибудь

закон появления объектов генеральной совокупности. Но предикат, определяющий основное свойство (здесь — свойство самоприменимости) и отражающий соответствующую закономерность — существует. Данный пример доказывает следующую теорему.

Теорема 3. *Существуют недетермированные задачи обучения распознаванию, для которых абсолютно точное решающее правило не является вычислимым.*

Для решения задач из класса $ND/D_k/-/-/-$ целесообразно использовать алгоритмы, извлекающие закономерность, которая имеет как можно меньшую колмогоровскую сложность. Действительно, недетерминированность предполагает полное отсутствие сведений о распределении объектов генеральной совокупности и вследствие этого допускает подход к выбору решения, которое можно обосновать как неслучайное.

Для задач из класса $ND/D_k/R_1T \vee ST/-/-$ целесообразно применение алгоритмов наименьшей колмогоровской сложности [5].

5. КЛАССИФИКАЦИЯ ДЛИН ВЫБОРОК

Практика применения машинного обучения показала, что одна и та же длина обучающей выборки может в некоторых случаях оказаться достаточной для получения требуемой точности распознавания, а в других случаях — быть слишком короткой. Так, взяв тривиальный пример детерминированной задачи с заведомо линейным но неизвестным решающим правилом-предикатом $f_o(x_1, x_2) = \ll ax_1 + bx_2 = c \gg$ для случая двух переменных-признаков и точную выборку, в которой две точки будут принадлежать классу "лежащих на прямой можно абсолютно точно решить задачу восстановления линейного предиката.

Для обучения многослойных нейронных сетей требуются большие выборки, поскольку нейросетевые семейства решающих правил имеют большую емкость.

Классы длин выборок для рассмотренных выше задач обучения распознаванию должны определяться ситуативно, в зависимости от емкости семейств, из которых в процессе обучения извлекается решающее правило.

Ниже приведена таблица, согласно которой следует определять значение параметра $SLen$ в стандартных кодах $STD/VAR/SFM/Slen/ADI$ задач обучения распознаванию; l — длина обучающей выборки.

Параметр $SLen$	Значение параметра	Определяющее условие
SS	Малая выборка	$l < VCD(S)$
AS	Средняя выборка	$VCD(S) \leq l < 1.5VCD(S)$
LS	Большая выборка	$l \geq 1.5VCD(S)$

6. КЛАСС $S_k/C/ - / - /SI$

Параметр $ADI = SI$ (специальная информация) в стохастических параметрических задачах чаще всего определяет типы используемых вероятностных распределений и, возможно, специфические характеристики параметров (в приведенном ниже примере — равенство ковариационных матриц классов).

Рассмотрим пример k -параметрической стохастической задачи обучения распознаванию объектов двух классов, которая хорошо изучена в теории статистических решений.

Пусть согласно дополнительной информации условные вероятности появления в выборке объектов каждого из двух классов $J \in \{0, 1\}$ имеют многомерное нормальное распределение

$$p(X|J) = \frac{1}{(2\pi)^{n/2} |\Sigma_J|^{1/2}} \exp\left\{-\frac{1}{2}(X - M_J)^T \Sigma_J^{-1} (X - M_J)\right\},$$

где $X \in \mathbb{R}^n$; M_J и Σ_J — математическое ожидание и ковариационные матрицы двух классов $J = 1, 2$. Пусть также известны так называемые априорные вероятности появления объектов каждого из классов: p_0 и p_1 . Известно, что оптимальная (минимизирующая средний риск ошибки) дискриминантная функция в случае равных ковариационных матриц $\Sigma_0 = \Sigma_1 = \Sigma$ является линейной и имеет вид [4]:

$$g(X) = X^T \Sigma^{-1} (M_0 - M_1) - \frac{1}{2} M_0^T \Sigma^{-1} M_0 + \frac{1}{2} M_1^T \Sigma^{-1} M_1 + \ln\left(\frac{p_0}{p_1}\right). \quad (3)$$

Соответствующая решающая функция f_g имеет вид

$$f_g(X) = \begin{cases} 0, & g(X) < 0; \\ 1, & g(X) \geq 0. \end{cases}$$

Решение приведенной задачи, когда задана только обучающая выборка $(\tilde{X}_l, \tilde{\alpha})$ длины l , состоит в нахождении по этой выборке статистических оценок $\hat{p}_0, \hat{p}_1, \hat{M}_0, \hat{M}_1, \hat{\Sigma}$ и вычислении $\hat{g}(X)$ по формуле (3).

В рассматриваемой задаче, очевидно, не существует абсолютно точного решения f_o , но при точно заданных векторах математических ожиданий, априорных вероятностей и ковариационной матрице — соответствующая статистическая задача принятия решений имеет точное вероятностное решение f_g (с точностью до заданной вероятностной меры — распределений и априорных вероятностей). В постановке обучения распознаванию наилучшее решение $f_{\hat{g}}$ является известным и требует только вычисления статистических оценок $\hat{p}_0, \hat{p}_1, \hat{M}_0, \hat{M}_1, \hat{\Sigma}$.

Никакой корректный алгоритм для решения приведенной статистической задачи, разумеется, не подходит. Действительно, разделяющая поверхность $\hat{g}(X)$, соответствующая наилучшему решающему правилу, является линейной, в то же время классы пересекаются, и выборка, вообще говоря, может оказаться не делимой линейно. Тогда корректный алгоритм построит нелинейное правило распознавания, заведомо худшее, чем $f_{\hat{g}}$.

Обобщим этот вывод на случай произвольной стохастической задачи обучения распознаванию с двумя пересекающимися классами. Среди всевозможных решающих правил для такой задачи обязательно существует правило, минимизирующее вероятность ошибки или заданную функцию потерь (взвешенную функцию ошибки). Будем обозначать такое наилучшее правило f_o^S , а соответствующую ему дискриминантную функцию обозначим g_o^S . Очевидно, что любой корректный алгоритм, примененный к рассматриваемой задаче, определит решающее правило, вообще говоря, отличающееся от f_o^S , поскольку точки обучающей выборки могут быть расположены «по разные стороны» дискриминантной функции g_o^S произвольным образом. Следовательно, корректные алгоритмы для решения таких задач не подходят.

Для стохастических параметрических задач распознавания ёмкость класса, которому принадлежит дискриминантная функция, вообще говоря, не имеет значения; важно лишь то, чтобы эта функция минимизировала средний риск ошибки. Эта функция уже определена стохастическими параметрами задачи, и её не требуется *отыскивать* ни в каком классе.

7. СТОХАСТИЧЕСКИЕ НЕПАРАМЕТРИЧЕСКИЕ ЗАДАЧИ ОБУЧЕНИЯ РАСПОЗНАВАНИЮ ($STD = S$)

С непараметрическими задачами распознавания дело обстоит иначе. Вероятностные распределения предполагаются существующими, но они неизвестны; их восстановление по обучающей выборке как правило приводит к не менее сложным задачам, чем задача обучения распознаванию в классической постановке.

Для задач рассматриваемого класса всегда можно полагать существование некоторой решающей функции f_o^S (дискриминантной функции g_o^S) наилучшей в статистическом смысле. Эта функция является неизвестной, и алгоритм обучения, конечно, должен находить её наилучшее приближение. Понятно, что такой алгоритм вовсе не обязан быть корректным на выборке. Но должен ли он давать на этой выборке минимальную эмпирическую ошибку, т.е. иметь на ней как можно более близкую к точной настройку?

Учитывая результаты рассмотрения параметрических стохастических задач, можно предположить, что для рассматриваемого класса задач обучения перенастройка (выбор корректного или с очень малой эмпирической ошибкой алгоритма) может привести к большим ошибкам классификации объектов, не принадлежащих обучающей выборке. По-видимому, это связано с тем, что неизвестная дискриминантная функция g_o^S (если она бейесовская, минимизирующая средний риск, т.е. статистически оптимальная) должна быть полиномом невысокой степени для неизвестных, но существующих многоэкстремальных (и, тем более, одноэкстремальных) вероятностных распределений.

Представляется целесообразным пытаться искать решающее правило как можно более близкое к бейесовскому классификатору — по максимуму апостериорной условной вероятности класса. Для некоторых семейств моделей распознавания доказаны теоремы о качестве приближения отыскиваемых решающих правил к бейесовскому. Именно такие модели наиболее пригодны для работы с непараметрическими стохастическими задачами обучения распознаванию.

Нужно обратить внимание на то, что в ряде случаев, оценивая эмпирический риск, на самом деле осуществляют оценку наихудшего правила из используемого семейства S . Подчеркнём — в определённом выбранном для решения задачи семействе правил S . Так в работе [2] эмпирический риск оценивается как равномерное по всему семейству S отклонение частоты ошибки решающего правила A на выборке от вероятности ошибки *этого же самого правила* A :

$$\sup_{A \in S} |\nu(A) - P(A)|.$$

А на самом деле необходим выбор правила A , наиболее близкого к неизвестному наилучшему правилу f_o^S , которое для задач рассматриваемого класса является статистически оптимальным.

ЗАКЛЮЧЕНИЕ

В статье предложено выделить среди задач обучения распознаванию по эмпирической информации такие классы однотипных задач, для решения которых можно осуществить обоснованный выбор метода решения. Целесообразность такого подхода объясняется тем, что достаточно богатый набор моделей и методов решения задач распознавания образов, созданный более чем за полвека, не всегда правильно используется.

Построенная классификация задач дополнена указаниями методов обучения распознаванию, наиболее подходящих для каждого конкретного класса задач.

Направления дальнейшей работы связаны с уточнением классификатора задач и дальнейшим углублённым исследованием подходов к обоснованию применимости или неприменимости различных математических моделей обучения к задачам из введенных в работе типовых классов.

СПИСОК ЛИТЕРАТУРЫ

1. *Донской В. И.* Эмпирическое обобщение и распознавание: классы задач, классы математических моделей и применимость теорий. Часть I / В. И. Донской // Таврический вестник информатики и математики. — 2010. — №1. — С.15 — 23.
2. *Валник В. Н.* Теория распознавания образов / В. Н. Валник, А. Я. Червоненкис. — М.: Наука, 1974. — 416 с.
3. *Журавлев Ю.И.* Об алгебраическом подходе к решению задач распознавания или классификации / Ю. И. Журавлев // Проблемы кибернетики. — Вып.33.— М.: Наука, 1978.— С. 5—68.
4. *Нильсон Н. А.* Обучающиеся машины / Н. Нильсон. — М.:Мир, 1967. — 180 с.
5. *Donskoy V. I.* The Estimations Based on the Kolmogorov Complexity and Machine Learning from Examples / V. I. Donskoy // Proceedings of the Fifth International Conference "Neural Networks and Artificial Intelligence" (ICNNAI'2008). — Minsk: INNS. — 2008. — P. 292 — 297.

Статья поступила в редакцию 10.12.2011