УДК 681.513

В.И. Васильев

Обнаружение закономерностей сходства, равенства и порядка

Показана не только возможность применения основных принципов теории редукции для решения широкого круга задач экстраполяции, но и идентичность решения таких задач в рамках методов, использующих эти принципы.

A possibility of application of main principles of the reduction theory for solving a wide circle of extrapolation problems is shown. The identity of the solution of such problems within the framework of the methods using these principles is also shown.

Показано не тільки можливість застосування основних принципів теорії редукції для розв'язання широкого кола задач екстраполяції, але й ідентичність розв'язання цих задач у межах методів, що використовують ці принципи.

Введение. Теория редукции [1, 2] интенсивно развивалась в рамках проблемы обучения распознаванию образов (ОРО). Затем оказалось, что ее можно с успехом распространить на широкий класс задач экстраполяции, таких как идентификация объектов, задачи косвенных измерений, диагностики и прогноза.

Постановка задачи

В каждой из перечисленных задач требуется отследить некоторую закономерность

$$Y = F(X), \tag{1}$$

причем сделать это можно на основании ограниченного числа экспериментов (короткой выборки), для удобства сведенных в таблицу, где X – вектор, а $x_i (i = 1, m)$ – его составляющие, определяемые из эксперимента и характеризующие каждое значение скаляра y_j.

y/X	x_1	x_2	…	x_m
y_1	x_{11}	x_{12}	…	x_{1m}
y_2	x_{21}	x_{22}	…	x_{2m}
…	…	…	…	…
y_l	x_{l1}	x_{l2}	…	x_{lm}

В задачах ОРО требуется обнаружить некоторую индикаторную функцию F, которая принимала бы значение нуля на всех объектах X, принадлежащих одному образу, и единицы – на всех X другого образа. В задачах восстановления функций и прогноза обнаруженная закономерность должна выполняться для любых X, т.е. для любого X нужно «угадать» значение y. В задачах обнаружения скрытых закономерностей следует ответить на вопрос: существует ли зависимость какого-либо столбца x_j в таблице от всех остальных столбцов этой таблицы. Если она существует, то ее необходимо не только указать, но и восстановить, а в задачах идентификации «угадать» реакцию объекта на любое возмущение. Для упрощения задачи часто задаются структурой функций F. Теория редукции позволяет решить задачу без предварительного указания такой структуры. Алгоритмы сами выбирают структуру восстанавливаемых функций.

Редукция в проблеме ОРО (закономерность сходства)

Рассмотрим задачу ОРО в наиболее полной постановке. Пусть некоторое множество V объектов X состоит из M подмножеств V_j, причем $\cap V_j = 0$. Для простоты примем $M = 2$, тем более, что в дальнейшем не появится причин увеличивать M. Существует некоторое распределение $P(X)$, но о нем ничего не известно, кроме того, что оно существует. Зато задана случайная и независимая выборка, порожденная этим распределением и состоящая из ℓ объектов $X \in V = V_1 \cup V_2$. О каждом объекте известна его принадлежность к одному из подмножеств $V_j(j-1,2)$, т.е. задана выборка пар (X_v, V_j), причем $|\ell| << |V|$. Каждый объект представлен вектором $X_v(x_1,..., x_m)$, составляющие которого определяются некоторыми свойствами объектов. Требуется, используя только обучающую выборку, обнаружить, а затем синтезировать решающее правило, указывающее на принадлежность любого объекта из множества V с вероятностью ошибки, не превышающей ε, достигаемой с вероятностью $(1 - \eta)$.

Теория редукции базируется на фундаментальной теореме [3]. Если из N решающих правил выбирается одно, безошибочно разделяющее случайную и независимую обучающую выборку длиной ℓ, то с вероятностью $(1 - \eta)$ можно утверждать, что на новых данных вероятность ошибочной классификации с помощью этого правила не превысит величину

$$\varepsilon = \frac{\ln N - \ln \eta}{\ell}. \qquad (2)$$

Заметим, что вывод теоремы не связан с распределением $P(X)$, а величина N определяется числом всевозможных решающих правил, с помощью которых можно разделить выборку длиной ℓ. Эта теорема вызвала сомнение в работоспособности обычных методов ОРО, основанных на синтезе сложных решающих правил в пространствах большой размерности. Все портило число N, оказавшееся очень большим. Для бинарного пространства размерности m и для самого простого, т.е. линейного решающего правила, справедливо неравенство $\ln N \le m^2$. Если же рассматривать более сложные правила, например, кусочно-линейные, то $\ln N \le (km)^2$, где k – число используемых гиперплоскостей. Так при $m = 20$, $\eta = 0{,}1$, $k = 1$ (линейное правило) для обеспечения $\varepsilon \le 0{,}1$ потребуется выборка длиной $\ell \approx 4000$.

При всех последствиях, вытекающих из приведенной теоремы, в ней содержится и ключ к выходу из создавшегося тупика, состоящий в том, что приемлемых значений ε и $(1 - \eta)$ можно достичь не только увеличением ℓ, но и предельным сокращением размерности пространства и упрощением (вплоть до линейного) решающего правила.

Теория редукции указывает пути выбора или синтеза таких признаков, которые позволили бы в пространстве малой размерности n $(n \ll m)$ безошибочно разделить выборку ($\ell \approx 100$) линейным решающим правилом $(k = 1)$. Задача синтеза сложных решающих поверхностей в исходном «засоренном» пространстве огромной размерности заменяется задачей синтеза такого пространства признаков малой размерности, в котором образы легко разделяются плоскостью. Процедуры обучения распознаванию образов, основанные на теории редукции, объединяются общим названием – *метод предельного упрощения* (МПУ). Один из первых алгоритмов МПУ был разработан для бинарных пространств [1, 4]. Более универсальная процедура МПУ – альфа-процедура [1, 2], действующая в непрерывных (или смешанных) пространствах. Обе процедуры сами по себе обеспечивают только линейное разделение обучающей выборки. Для обеспечения заданных значений ε и $(1 - \eta)$ теория редукции устанавливает дополнительные ограничения на выбираемые признаки. С этой целью используется теорема (2). Показано, что если организовать последовательный синтез пространства, в котором в конце концов наступает линейное разделение образов, заданных на обучающей выборке длины ℓ, то величина N характеризуется неравенством

$$\ln N \le n_0 \ln m, \qquad (3)$$

где m – число первичных свойств, из которых выбирается или формируется n_0 признаков (n_0 – размерность синтезируемого пространства [4]). Поэтому используя (2) можно указать такую размерность синтезируемого пространства n_0, при которой линейное разделение случайной и независимой выборки длины ℓ гарантирует требуемые значения ε и $(1 - \eta)$.

Редукция в задачах обнаружения и восстановления функциональных закономерностей (закономерности равенства)

Особенность алгоритмов, основанных на теории редукции, состоит в том, что любая задача экстраполяции сводится к задаче ОРО, а затем решается одним из алгоритмов МПУ, что в конце концов приводит к решению исходной задачи.

Ранее при решении задачи ОРО рассматривалась индикаторная функция вида (1). Представим себе, что (1) – непрерывная многомерная функция. Пусть задана последовательность пар: $X_1 y_1, ..., X_2 y_2$, где X_v – вектор, y_v – скаляр, определяющий значения функции *F*. В точке, соответствующей X_v, требуется восстановить многомерную непрерывную функцию *F* так, чтобы для любого *X* выполнялось неравенство

$$\left| y_v - F(X_v) \right| < \text{o}. \quad (4)$$

Поставим каждому X_v в соответствие два значения *y*:

$$y_{v1} = y_v + \text{e}, \; y_{v2} = y_v - \text{e}. \quad (5)$$

При этом обучающая выборка увеличится вдвое и образует два подмножества V_1 и V_2, одно из которых содержит элементы (X_v, y_{v1}), а другое – (X_v, y_{v2}). Подмножества V_1 и V_2 можно рассматривать как образы. Если удастся безошибочно разделить эти образы, то тем самым можно восстановить функцию, обеспечивающую выполнение (4) для всей обучающей выборки. Представим теперь, что обучающая выборка состоит не из пар, как раньше (см. таблицу), а из векторов $(X_1, X_2, \ldots, X_\ell)$, составляющими которых являются переменные x_i. В этом случае для каждого столбца таблицы можно составить функцию $X_i = f(x_1, x_v, \ldots, x_m)$, $\{x_v \neq x_i\}$, организовать альфа-процедуру, в результате чего можно будет ответить на вопрос: существуют ли в таблице столбцы, зависящие от других столбцов в рамках выбранного значения ξ. Если такая зависимость существует, то она будет восстановлена с точностью, определяемой выбранным ξ.

Если же соотношение (1) определяет функцию одного аргумента:

$$Y = F(t), \quad (6)$$

то можно искать приближение функции $y(t)$ в виде

$$y(t) = \sum_{i=1}^{m} \beta \varphi_i(t), \quad (7)$$

где $\varphi_i(t)$ – полная система функций на интервале $0 \le t \le T$.

Пусть задана случайная и независимая выборка (y_v, t). Тогда всем значениям t_v будет соответствовать конкретное значение $y(t_v)$ и *m* точек $\varphi_i(t)$, каждая из которых соответствует номеру *i*-й составляющей, т.е. каждому значению $y(t_v)$ будет соответствовать *m* точек на каждой из составляющих $\varphi_i(t)$ $i = 1, m$). Таким образом, можно получить аналогичную таблицу, в каждой строке которой будет реализация функции:

$$y(t_V) = F\{\varphi_1(t_v), ..., \varphi_m(t_v)\}. \quad (8)$$

В этом случае задача приближения функции (7) преобразуется в задачу восстановления многомерной функции (8). При этом безошибочное разделение образов, порожденных подстановкой (5), равносильно построению приближения (7), не отличающегося от функции (6) больше чем на (±ξ). В качестве $\varphi_1(t_v)$ могут рассматриваться производные от функции $y(t)$, и тогда задача преобразуется в задачу идентификации динамического объекта, т.е. в задачу восстановления дифференциального уравнения, одним из порождений которого является функция (6) [5].

Редукция в задачах интерпретации размытых понятий (закономерности порядка)

Пусть функция принадлежности к какому-либо понятию представима соотношением (1), где *F* – непрерывная или индикаторная функция; *X* – скаляр или вектор, составляющие которого являются числовыми характеристиками элементов, определяющих понятие; *y* – функция принадлежности, принимающая значения из множества действительных чисел. Для интерпретации понятия достаточно определить составляющие вектора *X* и восстановить функцию *F*. В случае если функция *F* индикаторная, то задача легко сводится к задаче ОРО, решаемой в рамках теории редукции, т.е. синтезируется пространство, в котором понятия становятся линейно разделимыми. Если же *F* – непрерывная функция, то функцию принадлежности можно восстановить следующим образом [1]: пусть наблюдаются ℓ объектов, описываемых лингвистическими переменными $Y = (y_1, y_2, ..., y_\ell)$, на которых установлен некоторый порядок, например $y_1 < y_2 < \ < y_\ell$. Задана обучающая последовательность, состоящая из ℓ объектов, каждый из которых описывается вектором $X(x_1, x_2, \dots, x_m)$, где x_i – измеряемые свойства объектов.

Таким образом, существуют два множества $y_1, y_2, ..., y_\ell$ и $x_1, x_2, ..., x_m$. На основе этих данных требуется установить на множестве X порядок, существующий на множестве Y. Используя для этого принципы редукции, т.е. из m измеряемых свойств объектов выбираются такие $n < m$ свойства, в пространстве которых на множестве X можно установить порядок, существующий на множестве Y, установленными лингвистическими переменными y. Таким образом, необходимо обнаружить закономерность, порождающую порядок на множестве Y. При этом будут решены две задачи: обнаружен порядок на множестве X, порождаемый порядком Y и появится возможность указать свойства объектов x_i, определяющие, например, различие между понятиями «красивый» и «уродливый».

Итак, задача заключается в создании пространства свойств объектов размерности $n < m$, в котором может быть построена функция принадлежности $\rho = F(x)$, определяющая место каждого x_i в порядке X такое же, какое занимает объект с соответствующим индексом в порядке Y.

Мерой упорядоченности множества X относительно множества Y в [1] предлагается считать величину $R(x) = w/\ell$, названную упорядочивающей силой свойств x. Здесь w число элементов множества X, место которых в наборе совпадает с номером его в наборе Y. При $R(x) = 1$ порядок на множестве X совпадает с порядком на множестве Y, а при $R(x) = 0$ – противоположен порядку на множестве Y.

Пусть объект X_0 – эталон понятия P, в качестве которого может быть выбрано начало координат. Для каждого свойства x_i обучающей выборки вычисляется расстояние от объекта X_0 до всех остальных объектов. Эти расстояния могут выполнять роль функции ρ при установлении порядка на объектах X. Вычисляется значение $R(x_i)$ упорядочивающей силы свойств. В качестве первой координаты пространства признаков x_1 выбира-ется свойство x_i, упорядочивающая сила которого удовлетворяет соотношению $R(x_i) \geq 1/n_0$. Затем выбирается новое свойство x_{i+1}, и в пространстве, определяемом признаком x_1 и свойством x_{i+1}, опять формируется вектор расстояний от объекта X_0 до всех остальных объектов и вычисляется значение $R(x_1, x_i)$ упорядочивающей силы пары x_1, x_i. Если упорядочивающая сила пары x_1, x_i удовлетворяет соотношению $R(x_1, \ x_i) \geq 2/n_0$, то свойство x_{i+1} объявляется признаком понятия P и выбирается в качестве второй координаты x_2 пространства признаков.

Однако если $R(x_1, \dots, x_i) \geq i/n_0$, то в пространстве размерности $n \leq n_0$ произойдет полное совпадение порядка объектов обучающей выборки относительно эталонного объекта x_0. Здесь $R(x_1, \dots, x_i)$ – упорядочивающая сила набора, состоящая из $x_1, \dots, x_i$ признаков.

Если выбирать n_0 исходя из соотношения $n_0 \leq \frac{\varepsilon \ell + \ln \eta}{\ln m}$ с надежностью $1 - \eta$ можно гарантировать, что вероятность неправильного указания степени принадлежности нового объекта (не принадлежавшего обучающей последовательности) не превысит ε.

Заключение. Итак, содержание статьи раскрывает предположение о том, что основные принципы теории редукции могут быть использованы для решения задач, лежащих за пределами проблемы обучения распознаванию образов, и показывает не только возможность применения этих принципов для решения широкого круга задач экстраполяции, но и идентичность их решения в рамках методов, использующих эти принципы. Следовательно, задачи ОРО и задачи восстановления функций (обнаружения закономерностей равенства) могут решаться практически одинаковыми алгоритмами, что позволяет использовать единый подход как в формулировании задач, так и в процессе их решения.

1. *Васильев В.И.* Теория редукции в проблемах экстраполяции // Проблемы управления и информатики. – 1996. – № 1, 2. – С. 238–252.
2. *Васильев В.И.*, *Суровцев И.В.* Индуктивные методы обнаружения закономерностей, основанные на теории редукции // УСиМ. – 1998. – № 5. – С. 3–13.
3. *Вапник В.Н.*, *Червоненкис А.Я.* Теория распознавания образов. – М: Наука, 1974. – 416 с.
4. *Vasilyev V.I.* The Reduction Principle in Pattern Recognition Learning (PRL) Problem // Pattern Recognition and Image Analysis/ – Interperlodlka. – 1991. – № 1. – P. 23–32.
5. *Васильев В.И.*, *Горелов Ю.И.* Индуктивные принципы восстановления динамических зависимостей // Проблемы управления и информатики. – 1998. – № 6. – С. 96–106.

Поступила 16.12.2008

Тел. для справок: (044) 526-4187 (Киев)