УДК 629.736

С.Н. Эш, Еш С.М., Esh S.

Индуктивные методы обнаружения закономерностей
Індуктивні методи виявлення закономірностей
Inductive methods conformities are display

Рассматривается общий подход к решению задач обнаружения закономерностей, основанный на принципе редукции. Обсуждаются задачи обнаружения и моделирования закономерностей сходства, равенства и порядка.

Розглядається загальний підхід до вирішення задач виявлення закономірностей, основою якого є принцип редукції. Обговорюються задачі виявлення і моделювання закономірностей схожості, рівності і порядку.

Consider general approach of solution a problem display conformities on these grounds is principe reduction. Problems under discussion display and modeling conformities similarity, equality and order.

Практически новая информация о мире базируется на анализе реальных фактов, получаемых либо в результате наблюдений за естественными процессами, происходящими в природе, либо в результате специально организованных экспериментов. Опыт позволяет обнаружить связи и закономерности, определяющие сущность происходящих процессов. Факты, способствующие приобретению опыта, бывают крайне малочисленными, но зато они объективно отражают реальность и содержат в себе все закономерности реального мира. Малочисленность фактов влечет за собой основные трудности обнаружения закономерностей, так как в основе большинства методов лежит индуктивный принцип, не всегда позволяющий получить надежные результаты. Изучение условий, при которых индуктивные методы дают надежные результаты, - предмет научного подхода к проблеме индуктивного вывода. Поэтому чрезвычайно важной становится задача, в которой по анализу ограниченного материала, представленного в виде протоколов и таблиц, можно обнаружить закономерности, скрытые в этих протоколах.

Здесь рассматривается только ограниченный класс закономерностей: закономерности равенства, сходства и порядка. Закономерности равенства устанавливают функциональную зависимость некоторого целевого параметра от одного или многих измеряемых величин. Закономерности сходства указывают правила, по которым отдельные объекты можно отнести к определенному классу, объединяющему в своих пределах сходные или похожие объекты. И наконец, закономерности порядка указывают, в каком порядке объекты следуют один за другим по какому-то не совсем определенному (нечеткому) свойству, например свойству «красивый». Во всех задачах предполагается, что все эмпирические данные можно свести определенного вида протокол или таблицу (табл. 1), и на основании анализа этой таблицы восстановить некоторую функцию вида:

$$y = F(X), \qquad (1)$$

где y — целевой параметр, а $X = X(x_1...,x_m)$ — многомерный вектор, составляющие которого в той или иной мере определяют целевой параметр y. В закономерностях сходства F — это индикаторная функция, определяющая принадлежность любого объекта к определенному множеству; в закономерностях равенства F — непрерывная функция, отображающая вектор X в определенные значения скаляра ***y;*** в закономерностях порядка F —функция принадлежности, устанавливающая на множестве объектов X определенный порядок.

Таблица 1

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

1. Принцип редукции. Все эти задачи решаются на базе одного и того же подхода, основанного на принципе редукции [1], состоящего в том, что в результате анализа эмпирических данных выбираются только такие аргументы x_i, которые формируют наиболее простые закономерности, устанавливающие связь целевого параметра ***y*** с определенными составляющими вектора X. При этом отбрасывается все то, что мешает проявлению закономерности. Принцип редукции дает возможность влиять не только на процедуру построения функции F, но и на процесс синтеза пространства, в котором эта функция выражена наиболее просто и рельефно. Вместо построения хитроумных и сложных функций в случайном пространстве размерности $\boldsymbol{m}$ выбираются только такие $\boldsymbol{n_o} \ll \boldsymbol{m}$, составляющие $\boldsymbol{x_i}$, в пространстве которых простая зависимость F проявляется наиболее ярко.

В основе принципа редукции лежит особенность, определяющая зависимость расположения точек в пространстве от реальных координат этого пространства. Такая зависимость указывает на то, что любые закономерности вида (1) существуют только в пространстве с определенными координатами. Так, закон Ома можно обнаружить только в определенных координатах, т.е. напряжение (U) связывается в отчетливую зависимость в координатах тока(I) и сопротивления (Я), но не цвета провода и яркости его изоляции. Если сравнивать животных по весу, то у льва и коровы он примерно одинаков. Но если сравнение вести с учетом величины клыков, то волк и лев попадают в один класс, существенно отличающийся от коровы. Поэтому вначале необходимо правильно указать координаты пространства, а затем пытаться восстанавливать в нем закономерность, т.е. задача восстановления самой закономерности отодвигается на второй план, как более простая и более изученная. Зависимость взаимного расположения точек от выбора координат пространства позволяет с помощью фиксации различных свойств объектов управлять расположением точек, соответствующих различным объектам. Решению задачи индукции должна предшествовать редукция (упрощения), позволяющая надеяться на получения успешного результата в условиях ограниченной выборки.

Для решения перечисленных задач обнаружения закономерностей в той или иной степени используется одна и та же процедура (альфа-процедура), основанная на теории редукции [1]. Поскольку теория редукции первоначально возникла и развивалась в рамках проблемы обучения распознаванию образов (проблема ОРО), а затем была перенесена на широкий круг задач экстраполяции, остановимся подробнее на ключевых моментах этой теории в рамках проблемы ОРО, чтобы использовать обсуждаемые результаты для решения других задач обнаружения закономерностей.

Теоретической базой теории редукции является фундаментальная теорема Вапника-Червоненкиса [2]. Утверждение этой теоремы состоит в том, что если из N решающих правил выбирается одно, безошибочно разделяющее случайную и независимую выборку длиной ℓ, то с вероятностью (1 - η]) можно утверждать, что вероятность ошибочной классификации с помощью этого правила не превысит величины

$$\varepsilon \leq \frac{\ln N - \ln \eta}{l} \qquad (2)$$

Если в этом неравенстве в качестве N взять число способов, которыми любую выборку длиной ℓ можно разделить на два подмножества данным классом решающих правил, то теорема полностью характеризует выбранный алгоритм. При заданных ε и η можно указать минимально допустимую длину обучающей выборки. Если же заданы ℓ и η, то можно найти гарантированную вероятность правильного распознавания выбранным классом решающих правил при условии, что обучающая выборка разделена безошибочно.

В этой теореме величина N существенно зависит от размерности пространства, в котором строится решающее правило, и от сложности выбираемого класса решающих правил. Доказано [3], что если организовать последовательный синтез пространства, в котором в конце концов наступит линейное разделение образов, то $\ln N < n_0 \ln m$, где m — число первичных свойств, из которых выбирается n_0 признаков.

Теория редукции [1] указывает процедуры синтеза пространства размерности n_0, в котором достигается безошибочное линейное разделение образов. Такие процедуры объединяются общим названием — метод предельных упрощений (МПУ).

Величина

$$n_0 = \frac{\varepsilon \iota + \ln \eta}{\ln m} \tag{3}$$

определяет размерность синтезируемого пространства, превышение которой приводит к потере гарантии достижения заданных ε и η. Для того чтобы размерность пространства не превышала n_0, вводится понятие разделяющей силы каждого признака x_i:

$$F(x_i) = \frac{\omega_{i-}\omega_{i-1}}{\iota}, \tag{4}$$

где ω_i $\omega_{i-\iota}$ - число объектов обучающей выборки, правильно классифицируемых до и после появления x_i. В [4] показано, что если разделяющая сила каждого используемого признака выше минимально допустимой:

$$F_{min}\,(x_i) = \frac{\iota}{n_0}, \tag{5}$$

то при $n \leq n_0$ обучающая выборка длиной ℓ будет безошибочно разделена линейным решающим правилом.

Альфа-процедура. Разновидностью МПУ является альфа-процедура [4]. Пусть задана обучающая выборка длиной ℓ и на ней — образы V_1^* и V_2^*. Рассмотрим непрерывный признак x_i На его оси все объекты обучающей выборки представлены точками. Если на этой оси выбрать некоторый порог x_i^0, то вся выборка относительно этого порога разделится на два класса эквивалентности (часть точек ниже этого порога, а часть — выше). Полученные классы V_{1i} и V_{2i} в той или иной мере совпадают с образами V_1^* и V_2^*. Если выбрать некоторый критерий совпадения (например, эмпирический риск), то можно найти порог x_i^0, оптимизирующий выбранный критерий.

Выбор критерия зависит от конкретной задачи и не оказывает существенного влияния на структуру алгоритмов метода.

Пусть выбран критерий $\iota(\omega)$, где ω — число правильных классификаций обучающей выборки. Тогда разделяющая сила признака x_i вычисляется согласно соотношению

$$F(x_i) = \frac{\omega_i}{\iota} \qquad (6)$$

где ω_i — мощность множества объектов, правильно классифицируемых при оптимально выбранном пороге x_i^0. На первом шаге процедуры выбирается некоторое свойство, разделяющая сила которого максимальна и превышает минимально допустимую, и это свойство объявляется признаком x_ι. Затем для нового свойства x_i для каждого объекта обучающей выборки вычисляются числа

$$\widetilde{x_\iota} = p_i \cos(\beta_i + \alpha_{si}), \qquad (7)$$

где $\rho_i = \sqrt{x_1^2 + x_i^2}$; $\beta = arctg\frac{x_i}{x_\iota}$; α_{si} – переменный параметр $\left(\alpha_{si} = \overline{0^0 180^0}\right)$.

Выбирается такое α_{si}, для которого $\alpha_{si} = \arg\max \omega_{is}$, где ω_{is} – число правильно классифицируемых объектов при оптимальном пороге, устанавливаемом на каждом направлении, определяемом углом α_{si}.

При этом все объекты обучающей выборки проецируются на направления $\widetilde{x_\iota}$, а сами эти направления определяются величинами углов α_{si}. Каждому α_{si} соответствует свое направление $\widetilde{x_\iota}$ и на каждом таком направлении устанавливается оптимальный порог $\widetilde{x_\iota^0}$. В результате выбирается такое направление $\widetilde{\boldsymbol{x}_\iota^*}$, определяемое углом α_{si}^*, на котором оптимальный порог дает наилучшее разделение в смысле выбранного критерия

Если

$$\frac{\omega_{i-}\,\omega_{i-l}}{\iota} > \frac{l}{n_0}$$

то x_i; объявляется признаком x_2 с базовым направлением $\widetilde{x_2}$. Затем выбирается новое свойство x_i и для него в плоскости $\widetilde{x_\iota}$ - x_2 вычисляются величины $\widetilde{x_k} = \rho_k \cos(\beta_k + \alpha_{sk})$, где

$$\rho_k = \sqrt{x_2^2 + x_k^2}; \quad \beta_k = arctg\left(\frac{x_k}{x_i}\right).$$

Если неравенство $\frac{\omega_3-\omega_2}{\iota} > \frac{I}{n_0}$— удовлетворяется для оптимального α^*_{sk}

I n_0

то свойство $x_к$ объявляется признаком x_3 .Такая процедура (альфа-процедура) повторяется до тех пор, пока на одном из базовых направлений не произойдет полное разделение обучающей выборки. Плоскость, перпендикулярная этому базовому направлению и проходящая через его оптимальный порог, принимается в качестве решающего правила, синтезируемого альфа-процедурой. Описанная альфа-процедура хорошо представлена графически в [4].

Применяя МПУ, в том числе и альфа-процедуру, строим пространства, в которых сходные (похожие) объекты разделяются линейным решающим правилом. Тем самым сначала обнаруживается, а затем восстанавливается закономерность сходства, в результате обнаруживаются правила, по которым отдельные объекты относятся к определенному классу, объединяющему в своих пределах сходные или похожие объекты.

Обнаружение закономерностей равенства. Ранее рассматривалась задача обнаружения закономерностей сходства, как задача восстановления индикаторной функции (1). Теперь представим себе, что (1) — непрерывная многомерная функция. Пусть задана табл. 1, в каждой строчке которой располагается одна реализация функции (1), т.е. задана выборка пар y_1, X_1; y_2, X_2; … ; y_1, X_ι, где $X_\nu = X(x_1,\ldots,x_m)$, $a\ y_\nu$— значение функции *F* в точке, соответствующей $\boldsymbol{X_\nu}$. Требуется восстановить многомерную непрерывную функцию (1) так, чтобы для любого $\boldsymbol{X_\nu}$ выполнялось неравенство

$$y_\nu - F(\alpha, X_\nu) \le \xi \qquad (8)$$

Поставим каждому X_ν в соответствию два значения $\boldsymbol{y}$:

$$y_{\nu 1} = y_\nu + \xi; \qquad y_{\nu 1} = y_\nu - \xi \qquad (9)$$

При этом табл. 1 увеличивается вдвое, так что в одной ее части находятся элементы $y_{\nu 1} \epsilon V_1$ а в другой $y_{\nu 2} \epsilon V_2$. Подмножества $\boldsymbol{V_1}$ и $\boldsymbol{V_2}$ можно рассматривать как образы в задачах ОРО, и если удастся разделить эти образы безошибочно, то тем самым мы сможем восстановить функцию (1), гарантирующую выполнение (8) для всей таблицы. Нарушение соотношения (8) происходит с частотой ошибочного распознавания образов V_1 и V_2, а значит, восстановленная функция определяет закономерность равенства с точностью, определяемой величиной $\pm\xi$.

Для разделения образов используем альфа-процедуру, и если она приведет к успеху, то будет синтезировано пространство малой размерности, в котором восстанавливаемая функция окажется линейной по параметрам, и при этом ни в одной из точек обучающей выборки не будет нарушено неравенство (8). В качестве аппроксимирующей функции используем полином

$$y = \alpha_0 + \sum_{i=1}^{m} \alpha_i x_i + \sum_{i=l}^{m} \sum_{j=l}^{m} \alpha_{ij} x_i x_j + \sum_{i=l}^{m} \sum_{j=l}^{m} \sum_{k=l}^{m} \alpha_{ijk} x_i x_j x_k + \cdots, \quad (10)$$

где m — размерность вектора $\boldsymbol{X} = \boldsymbol{X}(\boldsymbol{x_l}, \dots, \boldsymbol{x_m})$, $\boldsymbol{\alpha}$ — настраиваемые коэффициенты. С геометрической точки зрения полином (10) представляет собой гиперповерхность в спрямляющем пространстве обобщенных координат $x_i, x_i x_j$ и т.д. Именно в этом смысле восстанавливаемая функция считается линейной по параметрам. Каждое слагаемое этого полинома можно рассматривать как исходное свойство для альфа-процедуры. В конце концов процедура выберет только те слагаемые, которые не противоречат неравенству (8), т.е. процедура как бы пропустит полином (10) через сито, оставив только то, что определяет зависимость (1), а весь «мусор» отсеит. Полученный полином линейный по параметрам, но существенно нелинейный по исходным переменным x_i.

Представим теперь, что обучающая выборка состоит не из пар, как ранее (см. табл. 1), а только из векторов $(X_1, \dots, X_l)$ составляющими которых выступают переменные x_i. Эту обучающую выборку представим в виде табл. 2. Используя ее, составим относительно каждого столбца гипотетические зависимости:

$$\begin{aligned} &x_{11} = f_1(x_{12}, x_{13}, \dots, x_{lm}); \dots\ x_{lm} = f_m(x_{11}, x_{12}, \dots, x_{(m-l)}); \\ &x_{21} = f_1(x_{22}, x_{23}, \dots, x_{2m}) = f_m\big(x_{21}, x_{22}, \dots, x_{2(m-1)}\big); \qquad (11) \\ &\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ &x_{l1} = f_1\big(x_{l_1}, x_{l_2}, \dots, x_{l_m}\big); \dots x_{l_m} = f_m\big(x_{l_2} x_{l_2}, \dots, x_{l(m-1)}\big). \end{aligned}$$

Каждый столбец этой новой таблицы можно рассматривать как обучающую таблицу, составленную относительно различных столбцов. Теперь сформулируем задачу так. Задана исходная выборка векторов $X_1, \dots, X_l$ (см. табл. 2). В соответствии с (8) зададим конкретное ξ = const. Требуется ответить на вопрос: существует ли в таблице закономерность в рамках выбранного ξ. Другими словами, существуют ли такие f, связывающие столбцы таблицы так, чтобы восстановленные зависимости вкладывались в коридор $\pm\xi$. Если такие зависимости существуют, их нужно восстановить. Для этого следует m раз применить стандартную альфа-процедуру при одинаковых ξ. Если удастся восстановить одну или несколько зависимостей удовлетворяющих (8), то эти зависимости будут функциональными закономерностями равенства, скрытыми в исходной таблице эмпирических данных.

Таблица 2

x_1	x_2	x_3	…	x_m
$\boldsymbol{x_{11}}$	x_{12}	x_{13}	…	x_{1m}
$\boldsymbol{x_{21}}$	x_{22}	x_{23}	…	x_{2m}
•	•	•	•	•
•	•	•	•	•
•	•	•	…	•
x_{l1}	x_{l2}	x_{l3}	…	x_{im}

Обнаружение закономерностей порядка. При решении задач распознавания образов наиболее важная, а часто и наиболее трудная задача — установить принципов численной оценки некоторых качественных параметров, присущих объектам распознавания. Большинство понятий, отражающих сущность объектов окружающего мира, порождаются в результате обобщения представлений об объектах по существенным, специфическим признакам. Однако признаки, определяющие понятия, чаще всего никем не определены, а поэтому множества объектов, объединяемые понятиями, не имеют четких границ. В этом случае понятия представлены нечеткими множествами, на элементах которых заданы не числа, а функции принадлежности, указывающие степень уверенности в принадлежности каждого элемента некоторому множеству и устанавливающие приоритеты между элементами множества.

Пусть наблюдаются $\boldsymbol{l}$ объектов $(X_1,...,X_l) \in \boldsymbol{V}$. На этих объектах установлен строгий порядок [y] относительно некоторого размытого понятия P, т.е. $y_1 < y_2 < \cdots < y_l$. Величина $\boldsymbol{y}$ характеризирует все объекты $\boldsymbol{X} \in \boldsymbol{V}$ обучающей выборки с точки зрения размытого понятия (например, «умный», «дурак»). Это значит, что если объекту X_i соответствует величина y_i, а объекту X_j — величина $\boldsymbol{y_{i+1}}$ и т.д., то объекты расположены по мере увеличения $\boldsymbol{y}$. Среди объектов обучающей выборки указан максимальный (минимальный) элемент X_0. Если известен порядок всех объектов обучающей выборки, то можно указать порядок для каждой пары в отдельности (например, $y_i > y_j$). Требуется восстановить такую функцию (1) из класса F, чтобы величина y, вычисленная по этой функции для каждого объекта $X_1,\dots,X_l$, соответствовала заданному порядку [y]. Согласно основному принципу редукции нужно подобрать такие n_0 свойств $(n_0 \ll m)$, в пространстве которых можно установить порядок, не отличающийся от порядка, установленного на обучающейся выборке. Другими словами, порядок [p], т.е. система неравенств $P_1 < P_2 < \cdots < P_l$ должен совпадать с порядком [y]. В [5] показано, что расположение точек в пространстве зависит от выбора координат этого пространства, поэтому задачу можно свести к задаче синтеза такого пространства, в котором порядок каждой пары не должен противоречить порядку, заданному на обучающей выборке.

О мере упорядочивания набора $[p]$ относительно набора $[y]$ можно судить [6] по величине $R(p \to y) = \frac{\omega}{l}$, где ω — число элементов множества $X \in V$, для которых номер индекса в наборе $[p]$ совпадает с номером в наборе $[y]$. Такое число, названное в [6] упорядочивающей силой, можно вычислить подсчетом пар, для которых порядок, заданный учителем, совпадает с порядком полученным в процессе синтеза пространства размерности n_0. При $R(y \to p) = 1$ порядок элементов набора [p] полностью совпадает с порядком элементов в наборе $[y]$.

Предположим, что для каждого объекта $X \in V$ известны значения свойств

$x_i(i = \overline{1, m})$, измеряемых на этих объектах. Задача состоит в том, чтобы выбрать такое подпространство свойств, в котором существует функция, принадлежащая классу F, для которой с вероятностью, не меньшей (1-η), можно утверждать [6], что вероятность ошибочного указания порядкового индекса для любого объекта не превысит значение ε.

Пусть $X_0 \in V$ — максимальный (минимальный) объект множества V. Класс функций F определим как расстояния от объекта X_0 до любой точки пространства. Функцию $F(\alpha, X) \in f$ обозначим

$$p = d_n^{\alpha}(X_0, X_v) = \sqrt{\sum_{i=1}^{n}(x_{io} - x_{iv})^2}. \qquad (12)$$

Здесь x_i — одно из свойств, измеряемых на объектах; α — индекс подпространства размерности $n < m$, являющегося одним из подпротранств, определяемых совокупностью, и признаков, выбираемых из m свойств.

Требуется найти подпространство таких свойств, размерность которого не превосходит предельно допустимого значения n_0 порядок $[p]$ объектов обучающей выборки в этом подпространстве должен совпадать с порядком [p] объектов, заданным учителем. Здесь n_0 определяется согласно теореме (2) в предположении [6] что $\ln N \leq n_0 \ln m$.

Для каждого свойства x_i объектов обучающей выборки в соответствии с (12) формируется вектор расстояния по этому свойству до объекта X_0. Для каждого свойства x_i вычисляется его упорядочивающая сила $R(x_i) = \omega_i / l$ [6]. В качестве первой координаты x_1синтезируемого пространства выбирается свойство, упорядочивающая сила которого удовлетворяет соотношение $R(x_i.) \geq \frac{1}{n_0}$ Затем проверяется новое свойство x_{i+1} и в пространстве, определяемом уже отобранным **СВОЙСТВОМ** x_1 и свойством x_{i+1}, формируется вектор расстояний от объекта X_0 **ДО** всех остальных объектов. Затем вычисляется упорядочивающая сила пары x_1, x_{i+1}, и если упорядочивающая сила пары больше минимально допустимой, то x_{i+1} объявляется координатой x_2 синтезируемого пространства. Если n_0 выбирается согласно (3) (с учетом, что $\ln N < n_0 \ln m$) то алгоритм приводит к синтезу такого пространства размерности $n_0 < m$, в котором расстояние (12) определяет место каждого объекта в порядке [y] и при этом с вероятностью, не меньшей (1 – η|), можно утверждать, что вероятность ошибочного указания места каждого объекта в порядке $[y]$ не превышает величины ε.

1. Васильев, В.И. Теория редукции в проблемах экстраполяции / В.И. Васильев // Проблемы управления и информатики. — 1996. — № 1-2. — С. 239-251.
2. Вапник, В.Н. Теория распознавания образов / В.Н. Вапник, А.Я. Червоненкис. — М.: Наука, 1974. — 416 с.
3. Vasilyev V.L The reduction principle in pattern recognition learning (PRL) Problem I I Pattern recognition and Image Analysis. — 1991. — N 1. — P. 23-52.
4. Васильев, В.И. Индукция и редукция в проблемах экстраполяции / В.И. Васильев // Кибернетика и вычислительная техника. — 1998. — Вып. 116. —С. 65-81.
5. Васильев, В.И. Обнаружение и моделирование закономерностей сходства, равенства и порядка / В.И. Васильев, Т.И. Ланге, А.И. Шевченко // Искусственный интеллект. — 2001. -— № 3. — С.56-62.
6. Васильев, В.И. Синтез пространств для восстановления функций принадлежности в задачах распознавания нечетких множеств / В.И. Васильев, И.И. Сушко // Автоматика. — 1994. — № 1-2.— С. 76-82.
7. Васильев, В.И. Самоорганизация коллективов решающих правил в задачах распознавания / В.И. Васильев, И.И. Суровцев, С.Н. Эш // Кибернетика и вычислительная техника. — 2001. — Вып. 133. — С. 8-18.