

УДК 519.863

В.И. Васильев

Международный научно-учебный центр информационных технологий и систем, Киев

Взаимодополняемость метода группового учета аргументов (МГУА) и метода предельных упрощений (МПУ)

Рассматриваются и анализируются два индуктивных метода восстановления функции: метод группового учета аргументов (МГУА) и метод предельных упрощений (МПУ). Указываются их преимущества перед другими методами, и показывается возможность их совмещения с целью улучшения их экстраполяционных свойств.

Все индуктивные методы, основанные на неполной индукции, а к ним, безусловно, относятся и МГУА [1] и МПУ [2], отличаются тем, что в них общие выводы делаются на основе частных фактов, а это может привести как к верным, так и к ошибочным решениям. Причина такой неопределенности состоит в том, что частные факты, на которых основываются общие выводы, не всегда достаточно хорошо характеризуют изучаемое явление. Вместе с тем получаемые общие выводы должны объяснять не только частные факты, но и все изучаемое явление в целом, т.е. эти общие выводы не должны изменяться при практически бесконечном расширении числа экспериментов. Поэтому качество индуктивного вывода должно определяться не только и не столько объяснением отдельных, полученных в процессе экспериментов фактов, сколько от экстраполяционных способностей этих выводов, их способностью к экспансии в область явления, не охваченную экспериментами.

Ограниченность информации накладывает допустимые пределы сложности модели. Чем больше известных фактов, тем выше может быть предельная сложность синтезируемой модели и, наоборот, чем беднее фактический материал, тем беднее по сложности может быть построенная модель. Чем сложнее модель, тем больше у нее возможностей в объяснении ограниченного числа экспериментальных фактов, т.е. больше различных сильно отличающихся вариантов, одинаково хорошо объясняющих эмпирические данные. Всякий раз, когда модель выбирается из слишком сложного класса, всё в большей мере не хватает эмпирических данных для её однозначного объяснения, т.е. модель оказывается сложнее того, что несут в себе накопленные факты, и эти факты просто не в состоянии воссоздать такую модель.

В задачах восстановления многомерных зависимостей всякие огрубления (упрощения) модели приводят к сглаживанию тех или иных

деталей. Напротив, чрезмерное усложнение без учета объема экспериментальных данных приводит к непомерной свободе поведения аппроксимирующей функции в области, не охваченной экспериментом, в то время как более простые модели в этой области ведут себя более «осторожно».

Переусложнение модели может привести к подлинной «катастрофе», т.е. к «развалу» решения [4], но при этом она может очень хорошо объяснять эмпирические факты, но вне этих фактов в рамках изучаемого явления становится абсолютно «слепой», т.е. даёт такие показатели, которые генерируются не изучаемым явлением, а самой этой моделью в результате ее переусложнения. Здесь возникает центральная проблема всех индуктивных методов, состоящая в правильном соотношении сложности синтезируемой модели с количеством имеющихся фактов.

Цель статьи – рассмотреть индуктивные методы применительно к решению задачи восстановления функций, состоящей в обнаружении и моделировании некоторой закономерности, связывающей основные характеристики изучаемого явления в функциональную зависимость вида

$$y = F(X). \quad (1)$$

Сделать это нужно на основании ограниченного числа экспериментов (короткой выборки) для удобства сведенной в таблицу:

$y \setminus X$	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 этой таблице X – вектор, x_i ($i = 1, m$) – его составляющие, определяемые из эксперимента и характеризующие каждое значение скаляра y_j ($j = 1, l$). В процессе обучения должна быть построена модель зависимости (1), хорошо аппроксимирующая ее не только в точках y_j , но и в любой точке пространства, определяемого координатами x_i .

Проблемы, возникающие при восстановлении функциональных зависимостей в условиях коротких выборок, являются типичными проблемами, возникающими при применении индуктивных методов, и существенно отличаются от классических проблем восстановления по выборкам большого объема. Особенность состоит в том, что при ограничении объема выборки качество восстановления зависит не только от качества аппроксимации в точках y_j , но еще и от таких факторов, как сложность аппроксимирующей функции и размерности пространства m . Эта особенность заставляет сосредоточить внимание на правильном соотношении сложности приближающей функции с объемом обучающей выборки, так как имеющейся информации может не хватить даже для

восстановления функции только в точках y_j . Тем более этой информации может не хватить для удовлетворительного восстановления в любой точке ее существования.

Среди немногих методов, в которых особое внимание уделяется поиску такого соотношения, выделяются метод группового учета аргументов (МГУА) [1] и метод предельных упрощений (МПУ) [2], [3]. Поэтому здесь будут рассмотрены эти два метода с точки зрения возможностей их взаимного дополнения.

Особенности метода группового учета аргументов (МГУА)

В основу метода [1] положено несколько основных принципов: неокончателности промежуточных решений, внешнего дополнения, самоотбора промежуточных решений, единственности окончательного решения. Эти принципы дают возможность решать задачи восстановления функций при ограниченном числе эмпирических данных. Одновременно с оптимизацией коэффициентов аппроксимирующего полинома происходит и оптимизация сложности, что позволяет не заботиться о выборе структуры этого полинома. Специальный критерий (внешнее дополнение) выступает преградой переусложнения модели, отбрасывая все, что не в состоянии воссоздать короткая обучающая выборка (отбрасывание «мусора»). В результате остается только то, что может быть надежно «подтверждено» конкретной выборкой, а отбрасываются «фантазии» модели.

Задача решается в несколько этапов. Вначале из всех независимых переменных (аргументов) образуются всевозможные группы комбинаций. В практических алгоритмах в каждую комбинацию входит только два аргумента (попарный учет аргументов). Относительно каждой пары (группы) составляется частное описание, т.е. некоторое простое уравнение не выше второго порядка, аргументами которого является выбранная пара аргументов. Вид частного описания одинаков для всех групп в течение всего процесса обучения. Вся выборка разделяется на две: обучающую и проверочную. Тем самым порождается внешнее дополнение (проверочная выборка), которое играет роль сита, отсеивающего все чрезмерно сложные, не имеющие права на существование в рамках ограниченной информации. Коэффициенты частных описаний определяются по данным обучающей выборки. В результате получается множество решений, поскольку частное уравнение каждой пары рассматривается как некоторая упрощенная модель восстанавливаемой функции.

Из полученного набора упрощенных моделей отбирается часть в некотором смысле лучших. При этом проявляются почти все основные принципы МГУА. Принцип неокончателности решений проявляется в том,

что ни одна из полученных на первом этапе моделей не принимается за истину и только часть из этих решений пропускается для дальнейшего усложнения модели. Отбор лучших решений осуществляется на основе принципа внешнего дополнения. Каждая из частных моделей проверяется на проверочной выборке, не участвующей в определении коэффициентов уравнений. Для дальнейшего усложнения модели допускаются только несколько промежуточных моделей, ведущих себя «осторожно» на данных проверочной выборки, т.е. «приобретенный опыт» этих моделей не противоречит новым данным. Такая проверка является реализацией принципа внешнего дополнения, а отбор лучших моделей осуществляет принцип самоотбора.

Прошедшие самоотбор частные описания формируют множество новых переменных, которые являются исходными аргументами для следующего этапа, состоящего в том, что из промежуточных переменных образуются группы (пары), которые формируют более сложные модели. Коэффициенты этих моделей определяются по обучающей выборке, а качество этих моделей проверяется по проверочной выборке, т.е. проходят фильтр внешнего дополнения, не допускающего переусложнения моделей и т.д. Часто возникает вопрос: почему бы не использовать обе выборки для более точного определения коэффициентов промежуточных моделей? Ответ на этот вопрос определяет всю глубину самой идеи МГУА. Дело в том, что структура МГУА позволяет сколь угодно точное определение коэффициентов результирующего полинома, т.е. полного совпадения модели с изучаемым процессом в точках интерполяции или в точках обучающей последовательности. Но среди этих «точных» моделей могут оказаться переусложненные, ведущие себя на новых данных непредсказуемо, т.е. «точные», но переусложненные модели могут дать фантастические отклонения на проверочной выборке. Внешнее дополнение (проверка на новых данных) не допускает усложнений, ухудшающих результат на проверочной выборке. Синтезируемая модель как бы защищена от излишних переусложнений.

Рассмотрим процесс синтеза модели оптимальной сложности более подробно. Представим полином, аппроксимирующий функцию (1), в общем виде

$$y = F(x_1, \dots, x_m). \quad (2)$$

В качестве такого полинома очень часто используется полином Колмогорова-Габора [1]:

$$y = \alpha_0 + \sum_{i=1}^m \alpha_i x_i + \sum_{i=1}^m \sum_{j=1}^m \alpha_{ij} x_i x_j + \sum_{i=1}^m \sum_{j=1}^m \sum_{k=1}^m \alpha_{ijk} x_i x_j x_k + \dots, \quad (3)$$

так как с помощью такого полинома можно добиться достаточно точной аппроксимации любой дифференцируемой функции F .

Эта сложная зависимость заменяется множеством простых функций:

$$y_1 = f(x_1, x_2); y_2 = f(x_1, x_3); \dots; y_s = f(x_{m-1}, x_m), \quad (4)$$

где $s = C_m^2$, причем функция f всюду одинакова.

Очень часто в качестве функции f выбираются простые зависимости.

$$y(x_i, x_j) = a_0 + a_1 x_i + a_2 x_j + a_3 x_i x_j \quad (5)$$

или

$$y(x_i, x_j) = a_0 + a_1 x_i + a_2 x_j + a_3 x_i x_j + a_4 x_i^2 + a_5 x_j^2, \quad (6)$$

связывающие только две переменные. Коэффициенты этих зависимостей можно определить по МНК, имея соответственно 4 или 6 точек наблюдений в обучающей последовательности. Среди моделей первого ряда выбираются несколько, например S^* наилучших, показавших хорошие результаты на проверочной выборке. Среди отобранных моделей остаются только те, которые «впитали» в себя нечто большее, чем хорошая аппроксимация в узлах интерполяции; они «угадывают» поведение функции (1) в области, не охваченной экспериментом. Во втором ряду алгоритма полученные на обучающей выборке значения y_i , соответствующие отобранным моделям, рассматриваются в качестве аргументов нового ряда:

$$z_1 = f(y_1, y_2); z_2 = f(y_1, y_3); \dots; z_{S^*} = f(y_{S^*-1}, y_{S^*}). \quad (7)$$

Здесь функции f остаются прежними и соответствуют соотношениям (5) или (6), но аргументами этих функций выступают переменные y_i . Коэффициенты новых моделей (7) находятся по МНК на точках той же обучающей последовательности. Новые модели проверяются на точках проверочной последовательности, и среди них выбирается S^* наилучших, которые используются в качестве аргументов следующего третьего ряда и т.д.

Сложность полиномов возрастает от ряда к ряду. Так, например, во втором ряду будут получены полиномы, содержащие нелинейные члены вида $x_1^2, (x_1 x_3), (x_2 x_3), (x_1^2 x_3), x_1^2 x_2 x_3$ и т.д. Коэффициенты этих полиномов могут быть определены по тем же точкам обучающей последовательности и не требуют дополнительной информации, хотя их сложность все время возрастает. При этом число определяемых коэффициентов значительно превосходит число точек обучающей последовательности. Если бы не было внешнего дополнения, т.е. проверочной выборки, то алгоритмы МГУА благодаря сложности синтезируемых полиномов могли бы абсолютно точно аппроксимировать функцию (1) во всех точках обучающей выборки, но при этом не осталось бы никаких гарантий удовлетворительного поведения восстанавливаемой функции на новых точках.

Возможность точной аппроксимации функции (1) в узлах интерполяции благодаря неограниченному возрастанию сложности аппроксимирующего полинома без пополнения обучающей выборки является важной и исключительной особенностью МГУА, дающей ему неоспоримые преимущества перед другими методами. Однако именно эта особенность при отсутствии внешнего дополнения влечет за собой непредсказуемое поведение аппроксимирующего полинома в области, не охваченной экспериментом. Внешнее дополнение не допускает переусложнения и останавливает алгоритм при ухудшении поведения модели на новых данных. Алгоритм останавливается сразу же по достижении единственного минимума отклонений, полученных на проверочной выборке. Тем самым выбирается модель оптимальной сложности, устанавливающая компромисс между сложностью и объемом информации, используемой при синтезе модели. Как только сложность превысит возможности обучающей информации, процесс синтеза модели прекращается.

Особенности метода предельных упрощений (МПУ)

Всякая индукция включает в себя элементы дедукции и с ней непрерывно связана. «Общее существует лишь в отдельном, через отдельное... Чтобы понять, нужно эмпирически начать понимание, изучение, от эмпирии подниматься к общему» [5]. И наоборот. Дедукция является обращением индукции. Такая диалектическая взаимосвязь индукции и дедукции приводит к мысли о том, что всякие индуктивные методы должны содержать в себе элемент дедукции, т.е. опираться на некоторые утверждения, проверять на них правильность индуктивного вывода на всех этапах формирования. Именно эти соображения были приняты во внимание при создании МПУ. В этом методе внешним дополнением является дедуктивное утверждение (теорема), не позволяющее переусложнять результирующую модель, построенную методами индукции. Здесь, так же, как и в МГУА, внешнее дополнение является фильтром, отбрасывающим «мусор» и оставляющим только то, что формирует анализируемую закономерность (1).

Задача восстановления многомерной функции (1) в МПУ сначала сводится к стандартной задаче обучения распознаванию образов (ОРО), а затем решение задачи ОРО обеспечивает решение задачи восстановления с заданным качеством и надежностью. Пусть задана обучающая выборка пар $x_i y_i, \dots, x_n y_n$, где x_v – вектор, а y_v – скаляр, определяющий значение функции $F(x)$ в точке, соответствующей вектору x_v . Требуется восстановить многомерную непрерывную функцию F так, чтобы для любого x_v выполнялось неравенство

$$|y_v - F(x_v, \alpha)| \leq \xi \delta, \quad (8)$$

где α – параметры восстанавливаемой функции.

Поставим, как и в [3], каждому x_v в соответствие два значения y :

$$y_{v1} = y_v + \xi; \quad y_{v1} = y_v - \xi. \quad (9)$$

При этом обучающая выборка увеличится вдвое и разделится на два подмножества V_1 и V_2 , одно из которых содержит элементы (x_{v1}, y_{v1}) , а другое (x_{v2}, y_{v2}) . Подмножества V_1 и V_2 можно рассматривать, как заведомо разделимые образы, и если удастся разделить эти образы, то тем самым удастся восстановить функцию $F(x, \alpha)$, гарантирующую выполнение (8) для всей обучающей выборки. Нарушение соотношения (8) будет происходить с частотой ошибочного распознавания образов V_1 и V_2 . Для разделения образов можно использовать любой алгоритм обучения распознаванию образов, в том числе и альфа-процедуру [3], которая, в свою очередь, является одной из реализации метода предельных упрощений. Если альфа-процедура приведет к успеху, то тем самым будет синтезировано пространство малой размерности, в котором решающая функция окажется линейной по параметрам, и при этом ни на одной из точек обучающей выборки не будет нарушено неравенство (8). В качестве решающей, а значит и аппроксимирующей, функции $F(x, \alpha)$ используется полином (3), в котором m – размерность вектора $X = x(x_1, \dots, x_m)$, а α – настраиваемые параметры.

С геометрической точки зрения полином (3) представляет собой гиперплоскость в спрямляющем пространстве обобщенных координат $x_i; x_i x_j; x_i x_j x_k$ и т.д. Именно в этом смысле решающая функция будет линейной по параметрам. Каждое слагаемое полинома можно рассматривать как исходное свойство для альфа-процедуры. В конце концов, процедура выберет только те слагаемые, которые не будут противоречить неравенству (8), т.е. процедура как бы пропускает полином (8) через сито, оставив только то, что определяет зависимость (1), а весь «мусор» будет отсеян. Полученный полином разделит выборку на два образа V_1 и V_2 и будет линейным по настраиваемым параметрам, но существенно нелинейным по исходным переменным x_i .

В процессе обучения часть членов полинома отбрасывается, а часть используется в качестве аргументов функции $F(x, \alpha)$. Такая сортировка членов полинома осуществляется с помощью внешнего дополнения, представляющего некоторое дедуктивное утверждение, основанное на предпосылках теории эмпирического риска. Основное содержание этой теории состоит в указании тех условий, при выполнении которых эмпирический риск (т.е. риск, вычисляемый по эмпирическим данным)

совпадает или почти совпадает со средним риском, вычисленным по бесконечным выборкам. В качестве пар первого шага альфа-процедуры принимаются все пары $(y_i, x_i)_{i=1, \overline{m}}$ [6].

Одно из центральных утверждений теории эмпирического риска, используемой в рамках проблемы ОРО, содержится в теореме [6], которая гласит о том, что если из N решающих правил выбирается одно, безошибочно разделяющее случайную и независимую выборку длины l , то с вероятностью $(l - \eta)$ можно утверждать, что вероятность ошибочной классификации с помощью этого правила не превысит величины

$$\varepsilon = \frac{\ln N - \ln \eta}{l}. \quad (10)$$

Главный смысл этой теоремы состоит в том, что она указывает вероятность ошибки распознавания для любого объекта, в том числе и не содержащегося в случайной и независимой обучающей выборке длины l , без указания каких-либо вероятностных характеристик множеств V_1 и V_2 . Используются только эмпирические данные выборки (факт безошибочного разделения выборки длины l), а вывод делается для любой случайной и независимой выборки, в том числе и для самой неблагоприятной, т.е. рассматривается даже самое неблагоприятное расположение точек выборки длины l .

Эта теорема легла в основу теории редукции, на которой базируется метод предельных упрощений. С первого взгляда утверждение теоремы не оставляет шансов на приемлемое решение задачи ОРО, так как число N , т.е. число способов, которыми можно разделить любую выборку длины l на два образа выбранным классом решающих правил, оказалось очень большим, а поэтому слишком большой оказывалась вероятность ошибки. Для снижения вероятности ошибки нужно увеличивать длину обучающей выборки l до фантастических размеров. Так, для линейного (самого простого) решающего правила при размерности пространства $m = 20$ и при $\varepsilon = 0,1$ и $(l - \eta) = 0,9$ обучающая выборка должна содержать не менее 4000 объектов. Для более сложных правил величина l быстро возрастает.

Со временем было замечено, что приемлемых значений вероятности ошибки можно достичь не только повышением l . Это может быть достигнуто за счет снижения размерности пространства и предельного, вплоть до линейного, упрощения решающего правила, так как число N очень сильно зависит и от этих параметров. В этом и состоит главная идея теории редукции, которая дает способы отбрасывания несущественной информации (отбрасывания «мусора») и выделения только того, что способствует объединению объектов в образы или вносит различие между ними. Задача синтеза сложнейших разделяющих поверхностей в исходном «засоренном» пространстве огромной размерности заменяется задачей синтеза такого пространства малой размерности ($n_0 \ll m$), в котором

образы, представленные обучающей выборкой, легко разделяются простым (чаще всего – линейным) решающим правилом. При этом гарантируется заранее заданная вероятность правильного распознавания новых объектов с заранее заданной надежностью. Другими словами, прежде всего приводится в соответствие сложность индуцируемых конструкций с имеющимся объемом информации. Исходная задача редуцируется в более простую в такой степени, чтобы ее сложность не превышала возможностей эмпирических данных, т.е. сложность должна быть не выше той, которая может уверенно проявиться в рамках имеющейся информации.

Основой редукции при индуктивном синтезе пространства является дедуктивное утверждение приведенной выше теоремы. Расширение пространства ограничено дедуктивным утверждением, которое не допускает никаких усложнений без существенного улучшения результата. Дедуктивное утверждение теоремы как бы взвешивает каждое усложнение и определяет, чего оно стоит. Осуществляется последовательный синтез пространства, в котором возможно линейное разделение. При этом дедуктивное утверждение проверяется очень просто на обучающей последовательности. Доказано [7], что если организовать последовательный синтез пространства, в котором, в конце концов, наступит линейное разделение образов, заданных на обучающей выборке длиной l , то $\ln N < n_0 \ln m$, где m – число первичных свойств, из которых выбираются или формируются по признаков.

В методе предельных упрощений заранее указывается размерность синтезируемого пространства n_0 , превышение которой приводит к потере гарантий достижения заданных ε и η . Такая размерность вычисляется согласно теореме

$$n_0 = \frac{\varepsilon l + \ln \eta}{\ln m} \quad (11)$$

Для того чтобы размерность синтезируемого пространства не превышала n_0 , вводится понятие разделяющей силы каждого признака x_i :

$$F(x_i) = \frac{\omega_i - \omega_{i-1}}{l}, \quad (12)$$

где ω_i, ω_{i-1} – число объектов обучающей выборки, правильно классифицируемых до и после появления признака x_i . Показано [5], что если разделяющая сила каждого признака, используемого для синтеза пространства выше максимального допустимой:

$$F_{\min}(x_i) = \frac{1}{n_0}, \quad (13)$$

то при $n \leq n_0$ обучающая выборка длины l может быть безошибочно разделена линейным решающим правилом. Минимально допустимая

решающая сила указывает ту часть «нагрузки» в процессе разделения выборки, которую должен взять на себя каждый признак. Более точное определение разделяющей силы приведено в [8].

Разновидностью МПУ является альфа-процедура [3], основная особенность которой состоит в том, что каждое из m свойств (таблица 1) подвергается особой проверке при помощи критерия внешнего дополнения, определяемого дедуктивным утверждением теоремы. При этом, прежде чем строить индуктивную конструкцию, каждое свойство проверяется на соответствие теореме (дедуктивному утверждению), и если практика не противоречит требованиям теории, то продолжается усложнение индуктивного вывода. К синтезу пространства допускаются только такие признаки, которые гарантируют линейное разделение образов, заданных на обучающей выборке длины l .

Пусть задана обучающая выборка длины l и на ней образы V_1^* и V_2^* . Рассмотрим непрерывный признак v_i . На оси, соответствующей этому признаку, все объекты обучающей выборки будут 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 расположатся ниже этого порога, а часть – выше). Полученные классы эквивалентности V_{1i}, V_{2i} в той или иной мере будут совпадать с образами V_1^* и V_2^* . Если выбрать некоторый критерий совпадения (например, мера совпадения множеств V_{1i} и V_1^* , V_{2i} и V_2^*), 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. Часто в качестве критерия выбирается функционал эмпирического риска.

Пусть выбран критерий, представляющий собой число объектов, правильно классифицируемых порогом x_i^0 . Тогда разделяющая сила признака x_i вычисляется согласно соотношению:

$$F(x_i) = \frac{\omega_i}{1}, \quad (14)$$

где ω_i – число объектов, правильно классифицируемых оптимальным порогом x_i^0 . На первом шаге альфа-процедуры выбирается некоторое свойство, разделяющая сила которого больше максимально допустимой (13), т.е. больше $\frac{1}{n_0}$, и это свойство объявляется признаком x_1 . На этом этапе уже действует дедуктивный критерий внешнего дополнения, так как признаком объявляется только такое свойство, которое в совокупности с себе подобными приведет к линейному разделению образов в пространстве,

размерность которого не превышает n_0 , а значит, в этом пространстве существует плоскость, разделяющая образы с заданными значениями ε и η .

После этого проверяется одно из оставшихся свойств, например x_j . Для этого свойства для каждого объекта обучающей выборки вычисляются числа $x_j = \rho_j \cos(\beta_j + \alpha_{sj})$, где $\rho_j = \sqrt{x_i^2 + x_j^2}$, $\beta_j = \arctg \frac{x_j}{x_i}$; α_{sj} – переменный параметр ($\alpha_{sj} = \overline{0^\circ, 180^\circ}$); x_i – свойство, объявленное ранее признаком.

Выбираем такое α_{sj} , для которого $\alpha_{sj} = \arg \max_s \omega_{js}$, где ω_{js} – число правильно классифицируемых объектов при оптимальном пороге, устанавливаемом на каждом направлении \tilde{x}_j , определяемом углом α_{si} .

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

Если разделяющая сила свойства x_j , вычисленная согласно (12), окажется выше минимально допустимой (13), то свойство x_j объявляется признаком x_2 . Здесь опять вступает в силу дедуктивный критерий внешнего дополнения, который отбрасывает мусор и оставляет то, что приводит к успеху.

В некоторых алгоритмах альфа-процедуры на первом этапе рассматриваются не отдельные свойства x_i , а последовательно все C_m^2 пар, и для каждой пары ищется оптимальное направление \tilde{x}_{ij} . После этого выбирается новое свойство x_k , и для него в плоскости $\tilde{x}_k - \tilde{x}_{ij}$ вычисляется

величина $\tilde{x}_k = p_k \cos(\beta_k + \alpha_{sk})$, где $p_k = \sqrt{\tilde{x}_{ij}^2 + x_k^2}$; $\beta_k = \frac{x_k}{\tilde{x}_{ij}}$. Если

разделяющая сила свойства x_k выше минимально допустимой, то это свойство объявляется признаком x_3 и т.д. Альфа-процедура повторяется до тех пор, пока на одном из базовых направлений \tilde{x} не произойдет полное разделение образов. Размерность этого пространства не должна превышать n_0 . Плоскость, перпендикулярная базовому направлению и проходящая через оптимальный порог, принимается в качестве решающего правила, синтезируемого альфа-процедурой.

Если проводить аналогию между МГУА и МПУ, то в МГУА внешнее дополнение формируется на основе индуктивного вывода, и на каждом шаге сравниваются две индуктивных конструкции, построенные на двух

различных выборках: на обучающей и проверочной. И если эти конструкции близки, то усложнение продолжается. В МПУ сравниваются две модели: индуктивная и дедуктивная. Усложнение допускается, если разделяющая сила такого усложнения, легко вычисляемая по обучающей выборке, превышает минимально допустимую разделяющую силу, вычисленную согласно теореме, т.е. по теоретическим соображениям, и играющую роль внешнего дополнения. В том и другом методе усложнение продолжается до тех пор, пока оно дает положительный результат, т.е. пока сравниваемые модели не противоречат друг другу.

В МПУ усложнение модели продолжается до тех пор, пока не будет достигнут предельно возможный вариант – безошибочное разделение выборки, либо уже никакое усложнение не может дать улучшения без нарушения соответствия между дедуктивной и индуктивной моделями. То же самое осуществляется и в МГУА: усложнение допускается до тех пор, пока оно дает положительный результат на проверочной выборке, т.е. до тех пор, пока две индуктивные модели не противоречат друг другу.

Возможности совмещения МГУА и МПУ

Оба метода основаны на применении принципа внешнего дополнения для устранения некорректности в постановке задачи. В МГУА – это дополнительная проверка получаемых промежуточных моделей на проверочной выборке, а в МПУ индуктивная модель строится с учетом ее соответствия с теоретическими рекомендациями, полученными в результате дедуктивных соображений. В том и другом случае применяются меры против излишнего переусложнения модели, но в МГУА это осуществляется путем дополнительного контроля экстраполяционных свойств, а в МПУ – путем обеспечения условий, при которых эти экстраполяционные свойства гарантируются теорией. Трудно отдать полное предпочтение тому или другому методу. Ясно одно, что в некоторых задачах удобнее применять МГУА, а в других МПУ. Это позволяет надеяться на то, что совмещение этих методов заметно улучшит решение задач восстановления зависимостей в условиях коротких выборок путем последовательного применения сразу двух критериев внешнего дополнения.

Можно в качестве исходных данных альфа-процедуры использовать промежуточные полиномы МГУА, уже прошедшие дополнительную проверку на проверочной выборке. Это дает явное преимущество при усложнении результирующего полинома. В обычной альфа-процедуре (МПУ) после проверки всех линейных членов полинома (3) в качестве дополнительных переменных используются все нелинейные члены этого полинома, состоящие из двух аргументов (x_i, x_j) , а в комбинированном

методе для этой цели можно использовать модели, полученные в первом ряду МГУА, прошедшие самоотбор по какому-либо критерию внешнего дополнения (например, по критерию среднеквадратичного отклонения на проверочной выборке). В этом случае альфа-процедура осуществит проверку только тех моделей $y(x_i, x_j)$, вычисленных по (5) либо по (6), которые прошли порог самоотбора по критерию внешнего дополнения МГУА. В результате каждая промежуточная модель, т.е. каждое значение y пройдет двойную проверку: по способности работать на новых данных (проверочная выборка) и по соответствию теории, т.е. по способности взять на себя нагрузку, определяемую минимально допустимой разделяющей силой. Такая двойная проверка позволит из полного полинома (3) выбрать только те члены, которые, во-первых, сильнее всего приближают окончательное решение, а во-вторых, не вносят чрезмерных усложнений, ухудшающих работу результирующей модели на новых данных. Таким образом, в этом случае МГУА играет роль отбора нелинейных переменных для альфа-процедуры. Линейные переменные отбираются обычным алгоритмом альфа-процедуры, а если среди них не найдется достаточного количества переменных, обладающих разделяющей силой большей минимально-допустимой, то начинает работать МГУА, который на первом ряду генерирует, а затем проверяет по внешнему критерию промежуточные полиномы второй степени, часть из которых используются в качестве входных в альфа-процедуре. Если же и это не приводит к полному линейному разделению обучающей выборки, то используются полиномы более высоких степеней, прошедшие проверку на проверочной выборке.

Следует особо отметить, что МГУА в этой схеме работает автономно, без какой-либо зависимости от МПУ. Каждый из двух методов выдает свой окончательный результат. Предпочтение можно отдать модели, которая получена раньше, а можно право выбора оставить за конструктором.

Можно поменять местами МПУ и МГУА, используя МПУ в качестве предварительной проверки переменных, используемых в последствии в МГУА. В этом случае, согласно МПУ, строится линейная модель по общей схеме. Если эта модель окажется неполной, т.е. не даст безошибочного разделения выборки, то среди всех x_i ($i = 1, m$) выбираются только m^* , разделяющая сила которых максимальна, и эти переменные используются в первом ряду МГУА для построения квадратичных частных полиномов. Коэффициенты построенных по соотношениям (5) и (6) промежуточных полиномов определяются по МНК, а сами полиномы проверяются на проверочной выборке. Из построенных полиномов выбирается m^* наилучших, и из них формируется $C_{m^*}^2$ пар, каждая из которых проверяется и ранжируется алгоритмом альфа-процедуры по дедуктивному критерию минимально допустимой разделяющей силы. На этом этапе можно для каждой пары вычислять абсолютную разделяющую силу, а можно

учитывать «работу», совершенную уже отобранными линейными членами. Так или иначе, все пары, построенные по переменным y вычисленным алгоритмом МГУА ранжируются по дедуктивному критерию разделяющей силы. Каждой паре будет соответствовать обобщенная переменная \tilde{x}_{ij} [4]. Лучшие пары, т.е. соответствующие им переменные \tilde{x}_{ij} используются для формирования пар нового ряда МГУА. Далее работает обычная схема МГУА, т.е. всем парам приписывается свой полином согласно (5) или (6), и находятся по обучающей последовательности его коэффициенты. Все полиномы проверяются на проверочной выборке и m^* лучших из них формируются в пары, каждая из которых проверяется по критерию разделяющей силы. Среди всех пар выбирается m^* наилучших и соответствующие им обобщенные переменные \tilde{X} используются для формирования новых пар для алгоритма МГУА. Такая схема совместной работы двух методов предполагает параллельную работу двух алгоритмов – алгоритма МГУА и алгоритма МПУ. Остановка алгоритма МПУ происходит тогда, когда на каком-то этапе либо произойдет безошибочное разделение выборки, либо алгоритм МГУА приведет к ухудшению внешнего критерия. Как и в предыдущей комбинированной схеме правило остановки можно оставить за конструктором, который выберет момент остановки либо в момент остановки МПУ, либо – МГУА.

Таким образом, рассмотрены методы восстановления зависимостей в условиях коротких выборок. Даны рекомендации применимости каждого из методов к конкретным моделям.

Литература

1. Ивахненко А.Г., Степашко В.С. Помехоустойчивость моделирования. – Киев: Наукова думка, 1985. – 215 с.
2. Васильев В.И. Теория редукции в проблемах экстраполяции // Проблемы управления и информатики. – 1996. – №1-2.
3. Васильев В.И. Индукция и редукция в проблемах экстраполяции // Кибернетика и вычислительная техника. – 1998. – выб. 116. – С. 65-81.
4. Вапник В.Н. Алгоритмы и программы восстановления зависимостей. – М.: Наука, 1984. – 815 с.
5. Ленин В.И. Философские тетради. – М.: Политиздат, 1947. – 314 с.
6. Вапник В.Н., Червоненкис А.Я. Теория распознавания образов. – М.: Наука, 1974. - 416 с.
7. Vasilyev V.I. The Reduction Principle in Pattern Recognition learning (PRI) Problem // Pattern Recognition and Image Analysis? Interperiodice. – 1991. №1. – P. 25-52.
8. Васильев В.И. Индукция, дедукция и редукция в проблемах обнаружения закономерностей // Искусственный интеллект. – Донецк: ИПИИ, 1998. – №1.

Материал поступил в редакцию 22.11.00.