А Н А Л И З Р А З Н О Т И П Н Ы Х Д А Н Н Ы Х
(Вычислительные системы)
1983 год Выпуск 99

УДК 519.712.3

МОДЕЛИ ПРОГНОЗА МИНИМАЛЬНОЙ СЛОЖНОСТИ

В.Г.Устюжанинов

Введение

Решение большинства задач прогноза делится на два этапа. На первом этапе строится функция f (модель), связывающая целевую переменную x_0, значения которой требуется прогнозировать, с рядом независимых переменных $x_1,\ldots,x_n$. Второй этап (экзамен) состоит из собственно прогноза. Он заключается в вычислении значения функции f по заданным значениям независимых переменных. Объявляется, что прогноз значения целевой переменной равен вычисленному значению функции f. При этом возможны ошибки, так как истинное значение целевой переменной может отличаться от рассчитанного. Ясно, что успех прогноза полностью определяется тем, насколько удачно выбрана модель f на первом этапе. Таким образом, подзадача выбора модели является центральной в задаче прогноза. В предлагаемой работе развивается успешно применявшийся в задачах распознавания подход [1-3], основанный на стремлении к выбору модели минимальной сложности. Рассматриваются случаи статического и динамического прогноза. Основное внимание уделяется технике поиска моделей минимальной сложности. Вопрос о точности прогноза, получаемого на экзамене с помощью выделенной модели, в статье не затрагивается. Нет и математического обоснования избранного подхода, связывающего сложность модели с ее точностью на экзамене. Все это дело будущего.

§1. Статический прогноз

Имеются ряд объектов $d_1,\ldots,d_m$ (обучающая выборка) и совокупность числовых переменных $x_0,x_1,\ldots,x_n$. Задана таблица τ с

m строками и $n+1$ столбцами. Строке i таблицы τ сопоставляется объект d_i, а столбцу j – переменная x_j. На пересечении строки i со столбцом j находится значение τ_{ij} переменной x_j на объекте d_i. Задаются области значений $Z_0,\ldots,Z_n$ переменных $x_0,\ldots,x_n$. Фиксируется класс моделей F, т.е. класс функций $f: Z \to Z_0$, где $Z = Z_1 x \ldots x Z_n$ – прямое произведение множеств $Z_1,\ldots,Z_n$. Строится вектор $\tau_0 = (\tau_{10},\ldots,\tau_{m0})$, и рассчитывается вектор $\tilde{\tau}_0(f) = (\tilde{\tau}_{10},\ldots,\tilde{\tau}_{m0})$, в котором компонента $\tilde{\tau}_{i0} = f(\tau_{i1},\ldots,\tau_{in})$. Подчеркнем, что вошедшие в эти выражения числа $\tau_{i1},\ldots,\tau_{in}$ берутся из таблицы τ. На множестве m-мерных векторов $\theta = (\theta_1,\ldots,\theta_m)$ определяется метрика $\rho(\theta',\theta)$. Например, метрика вида

$$\rho(\theta',\theta) = \frac{1}{m}\sum_{i=1}^{m}(\theta_i - \theta_i')^2 \tag{1}$$

или

$$\rho(\theta',\theta) = \max_{1\le i\le m}|\theta_i - \theta_i'|,$$

или

$$\rho(\theta',\theta) = \begin{cases} 0, & \text{если } \theta = \theta', \\ 1, & \text{если } \theta \neq \theta'. \end{cases} \tag{2}$$

Качество модели f, ее точность, измеряется величиной $\rho(\tau_0,\tilde{\tau}_0(f))$. Выбирается число $\varepsilon \ge 0$, ограничивающее точность модели.

Задача синтеза модели формулируется как задача выбора из класса F функции f, удовлетворяющей требованию $\rho(\tau_0,\tilde{\tau}_0(f)) \le \varepsilon$.

Объединим в класс $F_\varepsilon(\tau)$ функции $f \in F$, отвечающие выписанному неравенству. Назовем их ε-моделями. Семейство моделей $F_\varepsilon(\tau)$ может быть весьма многочисленным. На выбор конкретной модели из этого семейства в значительной мере влияет гипотеза о характере связи между целевой переменной x_0 и независимыми переменными $x_1,\ldots,x_n$. Существует много подходов, развитых в рамках различных гипотез. Среди них наиболее глубоко исследованы статистический и детерминистский. Приходится констатировать, что на сегодняшний день ни один из подходов не гарантирует адекватности выбранной модели истинной закономерности и, таким образом, не избавляет от ошибок прогноза на экзамене [I]. Это связано с тем, что любой подход обосновывается системой предположений о характере закономерности и далеко не все из них поддаются проверке при решении конкретной задачи. В этой ситуации, с точки зрения практика, все нынешние подходы, независимо от глубины их разработки, в равной степени эвристичны. Разумеется, последнее относится и к одному из детерминистских подходов, излагаемому ниже, в котором задача выбора модели ставится как задача наилучшего приближения элементами класса F неизвестной функции, заданной набором значений в точках $d_1,\dots,d_m$ пространства Z.

Для выбора среди функций класса $F_\varepsilon(\tau)$ той единственной функции f, которая будет использоваться для прогноза, предлагается п р и н ц и п м и н и м а л ь н о й с л о ж н о с т и. Согласно этому принципу, каждой функции $f \in F$ сопоставляется число $s(f)$ - ее сложность, и среди функций из класса $F_\varepsilon(\tau)$ наиболее предпочтительными в качестве моделей прогноза считаются те функции, у которых значение $s(f)$ минимально.

Попытаемся аргументировать выбор такого критерия. Если положить Z_0 конечным, метрику $\rho(\theta,\theta')$ выбрать в форме (2) и ε приравнять нулю, то получим задачу классификации, или, как ее еще называют, задачу распознавания образов. Если, более того, потребовать, чтобы Z_0 состояло из двух элементов 0 и I, то объекты $d_1,\dots,d_m$ по значению целевой переменной можно разбить на два класса, а функции $f \in F_\varepsilon(\tau)$ интерпретировать как отделители объектов первого класса от объектов второго класса. Опыт решения задач распознавания подсказывает, что наиболее успешно ведут себя на экзамене те отделители, которые экстремальны по своим характеристикам. Это такие отделители, которые используют минимальное число действующих переменных (признаков), или такие, которым соответствует минимальное число гиперсфер [I] или гиперкубов [2] в пространстве Z, или такие, как тестовые алгоритмы, входящие в класс алгоритмов голосования [3]. В явном виде это эмпирическое наблюдение было сформулировано в виде принципа простоты в [I, с.I82].

В настоящей работе делается попытка спроецировать принцип простоты на задачи численного прогноза в случае, когда областью изменения переменных служит вещественная ось. Конкретно это выражается в обсуждении методов синтеза ε-моделей минимальной сложности.

§2. Линейные модели. Статика

Важное место в задачах прогноза занимают модели вида

$$f = \sum_{i=1}^{n} a_i x_i + a_0 . \tag{3}$$

Варьируя параметры a_i в пределах интервала $(-\infty,\infty)$, будем получать различные функции f. Совокупность этих функций образует класс F.

Зафиксируем функцию $f \in F$. Параметр a_i в записи функции f назовем действующим, если $a_i \neq 0$. Подсчитаем количество действующих параметров функции f. Обозначим их число через $s(f)$. Назовем его сложностью функции f.

Для оценки качества произвольной функции f выберем метрику в форме (I). Выделим в F подкласс $F_\varepsilon(\tau)$.

Задачу поиска ε-модели минимальной сложности запишем в виде:

$$s(f) \to \min_{f \in F_\varepsilon(\tau)} . \tag{4}$$

Редуцируем эту задачу к задаче булевого программирования.

Рассмотрим семейство двоичных наборов $\alpha = (\alpha_0,\dots,\alpha_n)$, у которых компоненты $\alpha_i \in \{0,1\}$ $(i = 0,\dots,n)$. Сопоставим фиксированному набору α число

$$\varphi(\alpha) = \min_{\substack{-\infty<a_k<\infty \\ k=0,\dots,n}} \frac{1}{m} \sum_{i=1}^{m} \left\{ \tau_{i0} - \sum_{j=1}^{n} \alpha_j a_j \tau_{ij} - \alpha_0 a_0 \right\}^2 . \tag{5}$$

Пусть f есть некоторая функция из F. Поставим в соответствие функции f набор α. Положим $\alpha_i = 1$, если в записи (3) функции f параметр $a_i \neq 0$ и $\alpha_i = 0$, если $a_i = 0$. Ясно, что сложность функции f равна $\sum_{i=0}^{n} \alpha_i$. Обозначим эту сумму значком $|\alpha|$. Назовем ее нормой набора α.

Запишем задачу булевого программирования:

$$\sum_{i=0}^{n} \alpha_i \to \min ,$$
$$\varphi(\alpha) \le \varepsilon, \quad \alpha_i \in \{0,1\}. \tag{6}$$

Легко показать, что задача (6) эквивалентна задаче (4), т.е. решение задачи (6) дает решение задачи (4), и наоборот.

Можно предложить тривиальный переборный алгоритм решения задачи (6), а именно: перебирать все двоичные наборы α и на каждом фиксированном двоичном наборе α методом наименьших квадратов [4] вычислять значение выражения $\varphi(\alpha)$. Затем среди наборов, для которых $\varphi(\alpha) \le \varepsilon$, выбрать набор α^* с минимальной нормой. На этом решение задачи (6) заканчивается. Чтобы по набору α^* построить решение задачи (4), надо набор α^* подставить в формулу (5) и найти значения a_i^* $(i=0,\dots,n)$, на которых в выражении (5) достигается минимум. Функция $f^*(x_1,\dots,x_n) = \sum\limits_{i=1}^{n} \alpha_i^* a_i^* x_i + \alpha_0^* a_0^*$ и будет решением задачи (4).

Трудность реализации этого метода заключена в колоссальности числа просматриваемых наборов. Всего их 2^{n+1}. Даже если положить время вычисления функции $\varphi(\alpha)$ на наборе α равным 0,01 сек, то и в этом случае уже при $n \ge 34$ общее время, затраченное на решение, составит годы. К сожалению, сомнительно, чтобы существовал метод точного решения задачи (6) с трудоемкостью, существенно меньшей трудоемкости переборного алгоритма. Это связано с тем, что задача (6) относится к числу так называемых универсально-переборных проблем [5], для которых не найдено эффективных методов решения. Таким образом, единственный выход из создавшегося положения - это отказ от точного решения задачи в пользу приближенного. Переходим к его описанию.

Будем считать, что набор $\alpha=(\alpha_0,\dots,\alpha_n)$ предшествует набору $\beta=(\beta_0,\dots,\beta_n)$, если для всех $i=0,\dots,n$ выполняется неравенство $\alpha_i \le \beta_i$. Отношение предшествования будем обозначать значком $\alpha \preccurlyeq \beta$. Набор α назовем непосредственно предшествующим набору β, если $|\alpha| = |\beta| + 1$ и $\alpha \preccurlyeq \beta$. Обозначим через $H(\beta)$ множество наборов, непосредственно предшествующих набору β. Заметим, что если $\alpha \preccurlyeq \beta$, то $\varphi(\alpha) \ge \varphi(\beta)$. Набор β назовем тупиковым, если $\varphi(\beta) \le \varepsilon$, а для всех $\alpha \in H(\beta)$ справедливо неравенство $\varphi(\alpha) > \varepsilon$.

Тупиковые наборы играют важную роль в решении задачи (6). В частности, точное решение задачи (6), как в этом нетрудно убедиться, достигается на тупиковом наборе. Отсюда потребность - уметь разыскивать тупиковые наборы. Способов их поиска можно придумать много. Рассмотрим один из них.

П р о ц е д у р а "Г р а д и е н т". На вход процедуры подается исходный набор β. Процедура либо устанавливает, что $\varphi(\beta) > \varepsilon$, либо, если $\varphi(\beta) \le \varepsilon$, находит тупиковый набор, подчиненный исходному набору β. Из сказанного ясно, что β может быть произвольным двоичным набором. Но если на момент обращения к процедуре мы не располагаем ни одним тупиковым набором и ищем тупиковый набор в первый раз, то исходный набор β лучше взять состоящим из одних единиц.

1. Проверяется условие $\varphi(\beta) \le \varepsilon$. Если оно выполняется, то переходим к исполнению следующего пункта.

2. Среди наборов $\alpha \in H(\beta)$ разыскивается набор α^*, на котором достигается $\min\limits_{\alpha \in H(\beta)} \varphi(\alpha)$. Отметим, что число наборов в множестве $H(\beta)$ равно $|\beta| \le n+1$. Поэтому упомянутый здесь минимум может быть найден простым перебором по всем наборам $\alpha \in H(\beta)$.

3. Проверяем условие $\varphi(\alpha^*) \le \varepsilon$. Если оно выполняется, то полагаем $\beta = \alpha^*$ и переходим к пункту 2. Если нет, то текущий набор β есть тупиковый.

Найденный с помощью процедуры "Градиент" тупиковый набор β вполне может служить в качестве приближенного решения задачи (6). Но совсем не обязательно, хотя и не исключается, чтобы он был точным решением.

Пусть γ есть точное решение задачи (6). Расхождение между точным решением и приближенным, т.е. число $|\beta| - |\gamma|$, может оказаться весьма значительным. Поэтому разумно многократно повторить процедуру градиентного спуска в различные тупиковые наборы $\beta^1, \ldots, \beta^q$ и затем среди них выбрать тот, норма которого минимальна. Возникает вопрос, как организовать многократный градиентный спуск, чтобы дважды не вычислять функцию φ на одном и том же наборе. Опишем один из таких способов.

Пусть $\beta^1, \ldots, \beta^q$ – найденные тупиковые наборы и надо разыскать очередной тупиковый набор β^{q+1}. Обозначим через $Q(\beta^i)$ множество наборов $\{\alpha / \beta^i \nless \alpha\}$. Образуем множество $Q = \bigcup\limits_{i=1}^{q} Q(\beta^i)$.

Оно состоит из наборов α, каждому из которых предшествует по крайней мере один из наборов $\beta^1, \ldots, \beta^q$. Отметим, что значение функции φ на наборах $\beta^1, \ldots, \beta^q$ не больше ε. Тем более на всех $\alpha \in Q$ $\varphi(\alpha) \le \varepsilon$. Наша задача заключается в подборе исходного набора для процедуры "Градиент", не принадлежащего множеству Q. Точнее, нам нужен такой алгоритм, который по наборам $\beta^1,\ldots,\beta^q$ выдавал бы требуемое $\beta \notin Q$. Алгоритм прост. Изложим его.

П р о ц е д у р а "П о к р ы т и е". Построим бинарную таблицу

$$B = \begin{Vmatrix} \beta_0^1 & \beta_1^1 & \ldots & \beta_n^1 \\ \beta_0^2 & \beta_1^2 & \ldots & \beta_n^2 \\ \cdot & \cdot & \cdot & \cdot \\ \beta_0^q & \beta_1^q & & \beta_n^q \end{Vmatrix}.$$

Набор столбцов $\nu_1,\ldots,\nu_s$ таблицы B назовем покрытием, если в образованной ими подтаблице в каждой строке есть по крайней мере одна единица. Другими словами, для всех $j=1,\ldots,q$ $\sum_{i=1}^{s} \beta_{\nu_i}^j \geq 1$.

Сопоставим покрытию $\nu_1,\ldots,\nu_s$ двоичный набор $\beta = (\beta_1,\ldots,\beta_n)$, у которого $\beta_i = 0$, если $i \in \{\nu_1,\ldots,\nu_s\}$, и $\beta_i = 1$, если $i \notin \{\nu_1,\ldots,\nu_s\}$.

Легко показать, что построенный таким способом набор $\beta \notin Q$.

Вообще, описанное построение устанавливает взаимно-однозначное соответствие между покрытиями таблицы B и наборами из множества $\bar{Q}$. Используя это соответствие, можно конструировать бесчисленные стратегии обхода тупиковых наборов. Но на этом мы останавливаться не будем.

§3. Нелинейные модели. Статика

Пусть задан набор элементарных функций $\eta_1(x_1,\ldots,x_n),\ldots,\eta_r(x_1,\ldots,x_n)$. Образуем из них обобщенный многочлен

$$f(x_1,\ldots,x_n) = \sum_{j=1}^{r} a_j\, \eta_j(x_1,\ldots,x_n) + a_0, \tag{7}$$

в котором значения параметров a_i суть произвольные числа из интервала $(-\infty, \infty)$.

Обобщенные многочлены играют ключевую роль в задачах приближения функций [6], в частности, функций, заданных набором значений в конечном числе точек, т.е. в нашем случае. Естественно желание использовать их в качестве моделей.

Пусть, как и прежде, имеется таблица τ. Для нее требуется построить модель f, которая с достаточной точностью прогнозирует целевую переменную, используя для этой цели независимые переменные. В качестве класса моделей F выберем класс обобщенных многочленов. Включим в F все многочлены, которые представляют произвольные линейные комбинации элементарных функций $\eta_1(x_1,\dots,x_n),\dots$ $\dots, \eta_r(x_1,\dots,x_n)$. Для оценки качества произвольной модели $f \in F$ будем пользоваться метрикой (I). Зафиксируем функцию $f \in F$. Выпишем список $a_0,\dots,a_n$ значений параметров, которые участвуют в записи функции f в форме (7). Подсчитаем количество отличных от нуля элементов этого списка. Полученное число назовем сложностью модели f. Обозначим его через $s(f)$.

Поставим задачу поиска ε-модели $f \in F$ минимальной сложности. Её можно решить, используя технику синтеза линейной модели минимальной сложности, изложенную в §2. Для этого надо свести нелинейную задачу к линейной.

Введем r переменных $y_1,\dots,y_r$. Положим $y_j = \eta_j(x_1,\dots, x_n)$ $(j = 1,\dots,r)$. Зафиксируем функцию $f = \sum_{j=1}^{r} a_j\eta_j + a_0$. Сопоставим ей функцию

$$\phi(y_1,\dots,y_r) = \sum_{j=1}^{r} a_j y_j + a_0 . \qquad (8)$$

Варьируя функцию f в пределах класса F, получим класс Ψ, состоящий из функций ϕ.

По таблице τ построим новую таблицу ξ из m строк и $r+1$ столбцов. Строке i таблицы ξ сопоставим объект обучающей выборки d_i, а столбцу j - переменную y_j. На пересечении строки i и столбца $j \geq 1$ разместим величину $\xi_{ij} = \eta_j(\tau_{i1},\dots,\tau_{in})$. Столбец $j = 0$ таблицы ξ образуем из элементов $\tau_{10},\dots,\tau_{m0}$ таблицы τ.

Качество прогноза произвольной модели $\phi \in \Psi$ на таблице ξ, т.е. на обучающей выборке, будем измерять метрикой (I).

Сложностью модели ϕ назовем число отличных от нуля значений параметров a_i в линейной форме (8). Обозначим ее $s(\phi)$. Ясно, что если функции ϕ сопоставлена функция f, то $s(\phi) = s(f)$.

Потребуем найти ε-модель $\phi \in \Psi$ минимальной сложности. Мы пришли к линейной задаче, эквивалентность исходной задаче которой очевидна.

§4. Динамический прогноз

Потребности практики не ограничиваются только задачами статического прогноза. Часто необходимо моделировать поведение процессов во времени. К обсуждению такого типа моделей мы и переходим.

Пусть имеются объект d и совокупность описывающих его переменных $x_0, x_1, \ldots, x_s, x_{s+1}, \ldots, x_n$. Эта совокупность разбивается на три группы: целевую переменную x_0, переменные управления $x_{s+1}, \ldots$ $\ldots, x_n$ и остаток $x_1, \ldots, x_s$. Объединение x_0 с $x_1, \ldots, x_s$ образует множество фазовых переменных.

Управление объектом d осуществляется через переменные управления. Считается, что в любой момент времени им могут быть назначены произвольные значения из некоторой фиксированной области. В отличие от последних фазовые переменные приобретают свои значения опосредованно через механизмы, связывающие их с переменными управления. Основной задачей моделирования является прогноз поведения во времени целевой переменной. Поэтому она и выделена в отдельную группу.

Особенность динамической задачи прогноза заключается в том, что влияние переменных управления на целевую переменную носит двоякий характер - непосредственное влияние и опосредованное через переменные $x_1, \ldots, x_s$. С этим связана необходимость учета переменных $x_1, \ldots, x_s$. Впрочем, совершенно не исключается, что список переменных $x_1, \ldots, x_s$ может оказаться пустым.

Для построения модели прогноза необходима обучающая выборка. Она организуется путем t-кратного измерения переменных $x_0, \ldots$ $\ldots, x_n$ на объекте d в моменты времени $1, 2, \ldots, t$. Данные измерений собираются в таблицу τ с t строками и $n+1$ столбцами. Строке l сопоставляется измерение в момент l, а столбцу j соответствует переменная x_j. На пересечении строки l со столбцом j размещается число τ_{lj} - результат измерения переменной x_j в момент времени l.

Ограничимся рассмотрением класса моделей прогноза вида:

$$x_i^{l+1} = \sum_{j \in V} (a_{ij}^0 x_j^l + a_{ij}^{-1} x_j^{l-1} + \quad + a_{ij}^{-p} x_j^{l-p}) +$$

$$+ \sum_{k \in W} (b_{ik}^1 x_k^{l+1} + b_{ik}^0 x_k^l + \ldots + b_{ik}^{-p} x_k^{l-p}) + c_i. \qquad (9)$$

В этой записи $V \subseteq \{0,\dots,s\}$ есть множество номеров фазовых переменных, учитываемых в данной конкретной модели. В связи с тем, что главной целью создания модели является прогноз поведения целевой переменной, множество V должно включать ноль. Индекс i пробегает V. Таким образом, модель (9) представляет систему уравнений для совокупности переменных x_i с $i \in V$. Благодаря тому, что ноль принадлежит V, система (9) содержит уравнение для x_0. Учет переменных управления осуществляется с помощью множества $W \subseteq \{s+1,\dots,n\}$. Если W пусто, то управлять процессом, описываемым (9), невозможно, т.е. он автономен. Число p назовем глубиной модели. Оно целое и может быть любым из интервала $[0, t-2]$. Индекс l отсчитывает моменты времени $p+1, p+2, p+3, \dots$, которые выбраны таким образом, чтобы для всех них выполнялось неравенство $l-p \geq 1$. Иначе запись правой части системы (9) была бы бессмысленной. Символ x_k^{l-q} $(q=-1,0,1,\dots,p)$ означает, что при расчете x_i в момент времени $l+1$ по формуле (9) на место x_k^{l-q} следует поставить значение переменной x_k в момент $l-q$. Разумеется, до этого необходимо либо вычислить значение x_k^{l-q} в момент времени $l-q$, либо задать его. Если x_k есть фазовая переменная, то для моментов времени $1, 2, \dots, p+1$ ее значение задается, а для моментов $p+2, p+3, \dots$ и т.д. оно вычисляется по формуле (9). В правую часть входят только те фазовые переменные, уравнения для которых содержатся в системе (9). Это дает возможность рассчитать значения x_k в любые моменты времени, начиная с $p+2$. Что же касается переменных управления, то их значения должны быть назначены во все моменты времени прогнозируемого периода. Значки a_{ij}^{-q} и b_{ik}^{-q} следует понимать как коэффициенты a_{ij}^{-q} и b_{ik}^{-q}, стоящие перед x_j^{l-q} и x_k^{l-q}, а не как записи a_{ij} и b_{ik} в степени $-q$. Параметры a_{ij}^{-q}, b_{ij}^{-q}, c_i принимают значения из интервала $(-\infty, \infty)$.

Выберем глубину p. Класс моделей M_p получается путем варьирования множеств $V \subseteq \{0,\dots,s\}$, $W \subseteq \{s+1,\dots,n\}$ и произвольного выбора a_{ij}^{-q}, b_{ik}^{-q}, c_i из интервала $(-\infty, \infty)$.

Пусть $\mathcal{M}$ - некоторая модель из M_p со своими значениями параметров $a_{ij}^{-q}, b_{ik}^{-q}, c_i$ и множествами V и W. Сопоставим i-му уравнению модели $\mathcal{M}$ вектор $\tilde{\tau}_i = (\tilde{\tau}_{p+2,i}, \dots, \tilde{\tau}_{t,i})$. Компоненты вектора подсчитаем по формуле:

$$\tilde{\tau}_{l+1,i} = \sum_{j\in V} (a^0_{ij}\, \tau_{l,j} + \ldots + a^{-p}_{ij}\, \tau_{l-p,j}) +$$

$$+ \sum_{k\in W} (b^1_{ik}\, \tau_{l+1,k} + \ldots + b^{-p}_{ik}\, \tau_{l-p,k}) + c_i ,$$

где l пробегает значения от $p+1$ до $t-1$ и в правую часть подставляются элементы $\tau_{l-q,i}$ из таблицы τ. Используя таблицу τ, сформируем вектор $\tilde{\tau}_i = (\tilde{\tau}_{p+2,i}, \ldots, \tilde{\tau}_{t,i})$. Качеством i-го уравнения модели $\mathfrak{M}$ назовем величину:

$$\chi_i(\mathfrak{M}) = \frac{1}{t-p-1} \sum_{l=p+2}^{t} (\tau_{l,i} - \hat{\tau}_{l,i})^2 .$$

Качество модели в целом можно оценивать по-разному. Для конкретности за меру качества модели примем $\chi(\mathfrak{M}) = \max\limits_{i\in V} \chi_i(\mathfrak{M})$. Отметим, что другой выбор меры качества приводит к непринципиальным изменениям излагаемой ниже процедуры "Модель". Назовем $\mathfrak{M}$ ε-моделью, если $\chi(\mathfrak{M}) \leq \varepsilon$. Выделим в M_p подкласс ε-моделей. Обозначим его $M_{\varepsilon,p}(\tau)$. Введем понятие сложности модели $\mathfrak{M}$. В отличие от переменной x_k назовем x_k^{l-q} аргументом. Аргумент x_k^{l-q} в уравнении i модели $\mathfrak{M}$ будем считать действующим, если стоящий перед ним в i-м уравнении модели $\mathfrak{M}$ коэффициент отличен от нуля. Обозначим через $Q_i(\mathfrak{M})$ множество действующих аргументов i-го уравнения. Образуем множество $Q(\mathfrak{M}) = \bigcup\limits_{i\in V} Q_i(\mathfrak{M})$. Значком $\|Q(\mathfrak{M})\|$ обозначим число элементов множества $Q(\mathfrak{M})$. За сложность модели $\mathfrak{M}$ примем величину $s(\mathfrak{M}) = \|Q(\mathfrak{M})\|$.

Задачу поиска модели минимальной сложности сформулируем в виде

$$s(\mathfrak{M}) \to \min_{\mathfrak{M}\in M_{\varepsilon,p}(\tau)} . \tag{10}$$

§5. Синтез линейных динамических моделей

Так же, как и в статическом случае, сведем задачу (10) к задаче булевого программирования.

Запишем двоичный набор: $\beta = (\beta_0^l, \ldots, \beta_0^{l-p}, \beta_1^l, \ldots, \beta_1^{l-p}, \ldots$ $\ldots, \beta_s^l, \ldots, \beta_s^{l-p}, \beta_{s+1}^{l+1}, \ldots, \beta_{s+1}^{l-p}, \ldots, \beta_n^{l+1}, \ldots, \beta_n^{l-p})$, компоненты которого принимают значения из множества $\{0, 1\}$, Набор β назовем правильным, если среди компонент $\beta_0^l,\ldots,\beta_0^{l-p}$ есть хоть одна отличная от нуля. В дальнейшем будем иметь дело только с правильными наборами.

Построим по набору β множество V. Составим его из тех номеров $i \in \{0,\ldots,s\}$, для которых сумма $\beta_i^l + \ldots + \beta_i^{l-p} > 0$. Из правильности β вытекает, что ноль будет принадлежать V. Сопоставим набору β класс моделей $N(\beta)$. Включим в $N(\beta)$ все модели, которые записываются в виде системы

$$x_i^{l+1} = \sum_{j=0}^{s} \sum_{q=0}^{p} \beta_j^{l-q} a_{ij}^{-q} x_j^{l-q} + \sum_{k=s+1}^{n} \sum_{q=-1}^{p} \beta_k^{l-q} b_{ik}^{-q} x_k^{l-q} + c_i, \quad (11)$$

где индекс $i \in V$ и значения параметров a_{ij}^{-q}, b_{ik}^{-q}, c_i берутся из интервала $(-\infty, \infty)$.

Пусть $K(\alpha) = \{\beta / \beta \preccurlyeq \alpha\}$. Образуем множество моделей $M(\alpha) = \bigcup_{\beta \in K(\alpha)} N(\beta)$. Зададим на множестве всех двоичных наборов α функцию $\varphi(\alpha)$. Положим $\varphi(\alpha) = 0$, если в классе моделей $M(\alpha)$ содержится ε-модель, и $\varphi(\alpha) = 1$, если такой модели в классе $M(\alpha)$ нет.

Пусть, как и прежде, $|\alpha|$ есть число отличных от нуля компонент в наборе α. Заметим, что число действующих аргументов у любой модели $\mathcal{M} \in M(\alpha)$ не превосходит величины $|\alpha|$. Таким образом, у всех моделей $\mathcal{M} \in M(\alpha)$ сложность $s(\mathcal{M}) \le |\alpha|$. Поэтому если устремить $|\alpha|$ к минимуму при одновременном требовании $\varphi(\alpha) = 0$, то в результате будет выделена модель, являющаяся решением задачи (10). Представим сказанное в сжатом виде:

$$\begin{gathered} |\alpha| \to \min, \\ \varphi(\alpha) = 0. \end{gathered} \quad (12)$$

Решать задачу (12) можно тем же методом, что и задачу (6). Действительно, с точки зрения изложенного в §2 метода, содержательная подоплека функции φ и способ ее вычисления не имеют ровным счетом никакого значения. Единственное, что требуется от функции φ, это:

1) возможность ее вычисления на любом наборе α;

2) монотонность, т.е. чтобы при $\beta \preccurlyeq \alpha$ выполнялось неравенство $\varphi(\beta) \ge \varphi(\alpha)$.

Выполнение монотонности для функции φ вытекает из включения $M(\beta) \subseteq M(\alpha)$, справедливого для $\beta \preccurlyeq \alpha$.

Принципиальная возможность вычисления функции φ на произвольном наборе α не вызывает сомнения. Остается сформулировать сам способ вычисления.

Пусть зафиксирован набор β и по нему построены множество V и класс моделей $N(\beta)$, представимых в виде системы уравнений (II). Введем величину предельного качества i-го уравнения этой системы. Обозначим ее $\xi_i(\beta)$.

Расчет $\xi_i(\beta)$ осуществим по формулам:

$$\xi_i(\beta) = \min_{a_{ij}^{-q},\, b_{ik}^{-q},\, c_i} \rho(\tau_i, \tilde{\tau}_i),$$

$$\rho(\tau_i, \tilde{\tau}_i) = \frac{1}{t-p-1} \sum_{l=p+2}^{t} (\tau_{l,i} - \tilde{\tau}_{l,i})^2,$$

зависящим от векторов τ_i и $\tilde{\tau}_i$.

Вектор $\tau_i = (\tau_{p+2,i}, \ldots, \tau_{t,i})$ составляется из элементов таблицы τ, введенной в §4. Компоненты вектора $\tilde{\tau}_i = (\tilde{\tau}_{p+2,i}, \ldots, \tilde{\tau}_{t,i})$ удовлетворяют равенству

$$\tilde{\tau}_{l+1\ i} = \sum_{j=0}^{s} \sum_{q=0}^{p} \beta_j^{l-q}\, a_{ij}^{-q}\, \tau_{l-q,j} + \sum_{q=s+1}^{p} \sum_{k=s+1}^{n} \beta_k^{l-q}\, b_{ik}^{-q}\, \tau_{l-q,k} + c_i,$$

в котором l пробегает значения от $p+1$ до $t-1$, а величины $\tau_{l-q,k}$, $\tau_{l-q,j}$ берутся из таблицы τ.

Минимизацию по a_{ij}^{-q}, b_{ik}^{-q}, c_i в равенстве для $\xi_i(\beta)$ можно провести по методу наименьших квадратов [4].

Способ вычисления функции φ сформулируем в виде процедуры "Модель".

П р о ц е д у р а "М о д е л ь". Пусть задан правильный исходный двоичный набор α и требуется установить, есть в классе моделей $M(\alpha)$ ε-модель или ее там нет.

I. Присваиваем текущему набору β значение исходного набора α.

2. По набору β строим множество V. Используя набор β и множество V, подсчитаем предельные качества всех уравнений системы (II). Получим множество чисел $\{\xi_i(\beta)/i \in V\}$.

Процесс работы процедуры организован так, что текущий набор всегда будет правильным. Вследствие этого построенное по β множество V всегда будет содержать ноль и среди чисел $\xi_i(\beta)$ будет присутствовать $\xi_0(\beta)$.

3. Сравним величину $\xi_0(\beta)$ с ε. Если $\xi_0(\beta) > \varepsilon$, то в $M(\alpha)$ нет ε-моделей и, значит, $\varphi(\alpha) = 1$.

4. Если же $\xi_0(\beta) \leq \varepsilon$, то из V выделим подмножество $R = \{i / \xi_i(\beta) > \varepsilon\}$.

5. Если $R = \emptyset$, то все уравнения системы (II) обладают ε-качеством, а сама система представляет ε-модель. Таким образом, в $M(\alpha)$ содержится ε-модель, и поэтому $\varphi(\alpha) = 0$.

6. Если же $R \neq \emptyset$, то присвоим нулевые значения всем компонентам набора β, у которых нижний индекс $i \in R$. Получаем новый текущий набор β и переходим к исполнению пункта 2.

Доказательство правильности работы процедуры опускается.

Литература

1. ЗАГОРУЙКО Н.Г. Методы распознавания и их применения.-М.: Сов. радио, 1972.

2. УСТЮЖАНИНОВ В.Г. Исследование алгоритмов, решающих дискретные задачи распознавания образов.Диссертация на соискание ученой степени кандидата физ.-мат.наук. М., 1980.

3. ЖУРАВЛЁВ Ю.И. Об алгебраическом подходе к решению задач распознавания и классификации. -Проблемы кибернетики,1978, вып.33, с. 5-68.

4. ДЕМИДЕНКО Е.З. Линейная и нелинейная регрессия. М.: Финансы и статистика, 1981.

5. ГЭРИ М., ДЖОНСОН Д. Вычислительные машины и трудно решаемые задачи. -М.: Мир, 1982. - 416 с.

6. БЕРЕЗИН И.С., ЖИДКОВ Н.П. Методы вычислений. Т.I.-М.:Наука, 1966. - 500 с.

Поступила в ред.-изд.отд.
19 июля 1982 года