

ПАКЕТ ПРИКЛАДНЫХ ПРОГРАММ "СИНТЕЗ"
ДЛЯ ПОСТРОЕНИЯ И АНАЛИЗА МОДЕЛЕЙ ПРОГНОЗА

В.Г.Устюжанинов, Е.Н.Шемякина

§1. Области применения и концепции

Решение задачи прогноза делится на два этапа. На первом из них выбирается модель прогноза. На втором осуществляется собственно прогноз, т.е. вычисляется значение целевой переменной, для чего используется модель, полученная на первом этапе. Пакет "СИНТЕЗ" предназначен для выбора модели прогноза в условиях неуправляемой обучающей выборки малого объема, представленной таблицей данных или многомерным временным рядом.

Чтобы очертить круг идей, заложенных в пакет "СИНТЕЗ", необходимо точно сформулировать задачу выбора модели прогноза. Опишем ее так, чтобы в единой схеме объединить нужные нам постановки задач регрессионного анализа [1, гл.1] и теории интерполяции и аппроксимации функций [2, гл.2,4].

Пусть Q есть область евклидова пространства R^k , элементами которого являются векторы $\tilde{z} = (z_1, \dots, z_k)$. Заданы класс функций (закономерностей) Ψ и базисная система функций $\Phi = (\varphi_1, \dots, \varphi_n)$. Функции $\phi \in \Psi$ и $\varphi_j \in \Phi$ определены и непрерывны на Q и принимают значения из R^1 . На функции $\varphi_j \in \Phi$ натянуто n -мерное линейное пространство F обобщенных многочленов (моделей прогноза) $f(\tilde{z}) = \sum_{j=1}^n a_j \varphi_j(\tilde{z})$ с коэффициентами $a_j \in R^1$. В дальнейшем обобщенные многочлены будут именоваться просто многочленами. На множестве пар $(\phi, f) \in \Psi \times F$ определена функция расстояния $\rho: \Psi \times F \rightarrow R^1$. Фиксируется число $\epsilon > 0$. С каждой функцией $\phi \in \Psi$ ассоциируется множество многочленов $M(\phi) = \{f \mid f \in F, \rho(\phi, f) \leq \epsilon\}$. Классы Ψ, F и число

ϵ выбираются так, чтобы для любой функции $f \in \mathcal{F}$ множество $M(f) \neq \emptyset$. Вид функции ρ зависит от специфики решаемой задачи.

Так, в теории равномерного приближения [2, гл. 4] функция $\rho(f, \mathcal{F}) = \max_{\tilde{z} \in Q} |f(\tilde{z}) - f(\tilde{z})|$. Естественно, в этом случае область Q считается замкнутой и ограниченной.

Имеется "черный ящик" с одним входом и одним выходом. Если на вход "черного ящика" подать точку $\tilde{z} \in Q$, то выход примет значение в соответствии с уравнением $y(\tilde{z}) = \sigma(\tilde{z}) + \xi$, в котором величина ξ есть погрешность, вносимая "черным ящиком", а в отношении функции σ известно лишь то, что она принадлежит классу \mathcal{F} .

Задача выбора модели прогноза заключается в нахождении многочлена $g \in M(\sigma)$. Она может решаться с использованием различных алгоритмических схем. Нас интересует класс схем, который включает алгоритмы регрессионного анализа [1, гл. I] и теории аппроксимации [2, гл. 2, 4]. Опишем его.

Множество точек $h = \{\tilde{z}_1, \dots, \tilde{z}_m\}$ назовем обучающей выборкой. Число m есть объем выборки. Образует семейство выборок $H = \{h \mid \|h\| = m\}$. Конкретный алгоритм $\gamma = \langle \nu, \alpha \rangle$ задается фиксацией выборки $\nu \in H$ и функции $\alpha: \mathbb{R}^n \times \{\nu\} \rightarrow \mathbb{F}$. Функция α сопоставляет обучающей выборке ν и произвольному вектору $\tilde{y} = (y_1, \dots, y_m) \in \mathbb{R}^m$ некоторый многочлен $f \in \mathcal{F}$. Алгоритм состоит из двух этапов. На первом из них накапливается информация о функции σ . Для этого на вход "черного ящика" подаются точки выборки ν , и из наблюдаемых значений его выхода составляется вектор $\tilde{y} = (y(\tilde{z}_1), \dots, y(\tilde{z}_m))$. На втором этапе эта информация используется для выбора многочлена из \mathcal{F} . В качестве такового берется многочлен $g = \alpha(\tilde{y}, \nu)$. Объявляется, что он принадлежит множеству $M(\sigma)$. При этом алгоритм может ошибиться. Описанию условий безошибочной работы алгоритмов посвящена значительная часть теорий, излагаемых в [1 и 2, гл. 2, 4].

Необходимость использования для прогноза вместо закономерности σ модели g связана с тем, что далеко не всегда функция σ может быть выражена в формульном виде. С другой стороны, погрешность прогноза не должна быть большой. Поэтому в качестве модели прогноза подбирается многочлен g , который близок к закономерности σ по расстоянию ρ .

Пусть τ_{1j} есть значение базисной функции ϕ_j в точке \tilde{z}_1 . Назовем матрицей наблюдений матрицу $X = (\tau_{1j})_{m \times n}$. Введем вектор $\tilde{y} = (y(\tilde{z}_1), \dots, y(\tilde{z}_m))$. Составим систему уравнений для неизвестного вектора $\tilde{a} = (a_1, \dots, a_n)$:

$$\tilde{y} = X \tilde{a} . \quad (1)$$

Определим матрицу Грамма $G = X^T \cdot X$. Детерминант матрицы G обозначим через $\det G$. В регрессионном анализе [1] и в теории аппроксимации функций [2, гл.2,4] способы конструирования отображения $\alpha: R^n \times \{v\} \rightarrow F$ базируются на требовании $\det G \neq 0$. Однако на практике это требование часто не выполняется из-за того, что объем выборки m меньше числа базисных функций n . В результате аппараты регрессионного анализа и теории аппроксимации оказываются парализованными. Проиллюстрируем сказанное на примере задачи идентификации линейной регрессии [1, гл.1]. Она получается из общей схемы, если считать, что область $Q = R^n$, семейство базисных функций $\Phi = \{x_1, \dots, x_n\}$, класс обобщенных многочленов F состоит из линейных регрессий $f = \sum_{j=1}^n a_j x_j$, класс функций $V = F$, погрешность ξ случайна, и ее распределение не зависит от точки $\tilde{x} = (x_1, \dots, x_n) \in Q$, в которой производится измерение значения функции $\alpha(\tilde{x}) = \sum_{j=1}^n b_j x_j \in F$. Вид функции ρ определяется выбором геометрии доверительной области для вектора коэффициентов $\tilde{b} = (b_1, \dots, b_n)$. Вектор \tilde{b} назовем оптимальным вектором системы (1), или просто оптимальным, если на нем достигается минимум евклидовой нормы $\|\tilde{y} - X\tilde{a}\|$. Согласно [1, гл.1] в качестве оценки коэффициентов регрессии σ следует взять оптимальный вектор $\tilde{\sigma}$. Если $\det G \neq 0$, то такой вектор единственный. Если же $\det G = 0$, то множество оптимальных векторов имеет мощность континуум, и на вопрос, какой из них брать в качестве оценки, теория, развитая в [1], ответа не дает.

Число ненулевых компонент вектора $\tilde{\sigma}$ назовем его булевой нормой. В работе [3] для случая $\det G = 0$ было предложено в качестве оценки коэффициентов регрессии σ брать произвольный оптимальный вектор, имеющий минимальную булеву норму. Количество нефиктивных переменных, входящих в регрессию f , назовем ее сложностью. Понятно, что сложность регрессии f равна булевой норме соответствующего ей вектора коэффициентов $\tilde{\sigma}$. Выбор вектора $\tilde{\sigma}$ в качестве оценки вектора коэффициентов \tilde{b} тождествен выбору в роли модели прогноза регрессии минимальной сложности. Идея минимизации сложности моделей прогноза опирается на положительный опыт применения принципа простоты в построении логических отделителей в задачах распознавания образов [4,5]. В пакете "СИНТЕЗ" эта идея воплощена в

программах поиска регрессионных моделей минимальной сложности. По опыту эксплуатации пакета можно сказать, что идея простоты оправдала себя и в этом случае. Кроме того, в настоящее время получен ряд теоретических результатов, описывающих условия, в которых отбор моделей по критерию минимума сложности неизбежен. Один из таких результатов будет сейчас сформулирован.

Допустим, что вектор $\tilde{a} = (a_1, \dots, a_n) \in R^n$, класс P_k состоит из регрессий $f = \sum_{j=1}^n a_j x_j$, сложность которых не больше k , класс $\Psi = F = P_k$, "черный ящик" реализует функцию $y = \sigma(\tilde{x}) = \sum_{j=1}^n b_j x_j \in P_k$, множество $M(\sigma) = \{\sigma\}$. Требуется разыскать элемент множества $M(\sigma)$, или, что то же самое, идентифицировать регрессию σ , реализуемую "черным ящиком".

Пусть $\tilde{b}(\sigma) = (b_1, \dots, b_n)$ есть вектор коэффициентов регрессии σ , множество $h = \{\tilde{x}_1, \dots, \tilde{x}_m\}$ есть произвольная выборка точек в пространстве R^n , вектор $\tilde{y}(\sigma) = (\sigma(\tilde{x}_1), \dots, \sigma(\tilde{x}_m))$. Обозначим через $X(h)$ матрицу наблюдений, которая построена на основании выборки h .

ТЕОРЕМА I. Существует такая единая для всех $\sigma \in P_k$ выборка точек v^* объема $m = \min\{2k, n\}$, что для каждой регрессии $\sigma \in P_k$ оптимальный вектор с минимальной булевой нормой уравнения $\tilde{y}(\sigma) = X(v^*)\tilde{a}$ единствен и равен $\tilde{b}(\sigma)$.

Из теоремы вытекает алгоритм идентификации произвольной регрессии $\sigma \in P_k$. Он работает по следующим правилам. Фиксируется выборка точек $v^* = \{\tilde{x}_1, \dots, \tilde{x}_m\}$, $\tilde{x}_i \in R^n$, и строится матрица наблюдений $X(v^*)$. В точках выборки v^* измеряется значение неизвестной регрессии σ . Формируется вектор $\tilde{y}(\sigma)$, и составляется система уравнений $\tilde{y}(\sigma) = X(v^*) \cdot \tilde{a}$. Находится ее оптимальный вектор с минимальной булевой нормой. Согласно теореме, это будет вектор коэффициентов регрессии σ .

Пусть g^* есть регрессия, коэффициенты которой составляют оптимальный вектор с минимальной булевой нормой уравнения $\tilde{y} = X(v^*)\tilde{a}$. Ясно, что описанный сейчас алгоритм определяется фиксацией выборки точек v^* и функции $\alpha: R^n \times \{v^*\} \rightarrow P_k$ в виде $\alpha(\tilde{y}, v^*) = g^*$. Точки выборки v^* подбираются так, что $\det(X(v^*) \cdot X^T(v^*)) \neq 0$. Поэтому при $m = n$ решение системы $\tilde{y} = X(v^*)\tilde{a}$ единственно, и описанный алго-

ритм переходит в классическую схему идентификации многочлена $\sigma \in P_k$, используемую в регрессионном анализе [1] и теории интерполяции [2, гл.2].

Теорема 1 устанавливает величину объема выборки, достаточного для идентификации любой регрессии из класса P_k . Он равен $\min\{2k, n\}$ и может быть значительно меньше размерности задачи n . Так, при $k=1$ класс P_k состоит из всевозможных регрессий $f = cx_j$, у которых $j \in \{1, \dots, n\}$ и $c \in (-\infty, \infty)$. Для идентификации любой из них достаточно измерить их значение всего в двух фиксированных точках пространства R^n . В то же время размерность пространства n может измеряться сотнями. Для решения этой же задачи методами многомерной интерполяции требуется объем обучающей выборки (сетки) $m = n$. И это естественно, так как при общем подходе не учитывается сложность регрессий, составляющих класс P_k .

На практике замечено, что хорошие модели прогноза просты, а задачи переопределены, т.е. на стадии их постановки вводится много лишних базисных функций φ_j и переменных z_i . Теорема 1 утверждает, что для идентификации моделей прогноза в этой ситуации не требуется больших объемов обучающей выборки, если выборка управляема и известна верхняя граница сложности моделей.

Ответим на вопрос, можно ли идентифицировать произвольную регрессию из P_k , используя для этой цели выборку объема $m < \min\{2k, n\}$. Алгоритм $\gamma = \langle v, c \rangle$ назовем неразличающим в P_k , если в классе P_k найдется пара регрессий δ и σ со свойством $\alpha(\gamma(\delta), v) = \alpha(\gamma(\sigma), v)$.

ТЕОРЕМА 2. Все алгоритмы, использующие выборки объема $m < \min\{2k, n\}$, являются неразличающими в P_k .

Отсюда следует, что объем выборки, необходимый и достаточный для идентификации любой регрессии из P_k , равен $\min\{2k, n\}$.

Другая задача, решаемая пакетом "СИНТЕЗ", это анализ временных рядов. В качестве средства их описания были выбраны авторегрессионные многомерные модели произвольных порядков [6, гл.5]. Для оценки коэффициентов этих моделей в теории случайных процессов используется, в частности, метод максимального правдоподобия [6, §5.4], который сводится к нахождению решения системы линейных алгебраических уравнений типа (1). Из требования высокой статистической надежности оценок коэффициентов уравнений авторегрессии следует, что параметр m должен значительно превосходить величину

n . Но для этого нужно, чтобы отрезок наблюдаемого временного ряда, который используется для построения модели авторегрессии, был достаточно длинным. На практике же ситуация часто прямо противоположная. Отрезок временного ряда короткий. В этом случае параметр $m < n$, решение системы (1) не единственно, а, значит, существует множество авторегрессионных моделей, описывающих наблюдаемый отрезок временного ряда с нулевой погрешностью. Спрашивается, какую из этих моделей использовать для прогноза будущего поведения временного ряда. В работе [3] предлагается в этом случае выбирать ту модель авторегрессии, которая имеет минимальную сложность. Алгоритм выбора такой модели и реализован в виде программного модуля в пакете "СИНТЕЗ".

Решающим фактором в построении алгоритма поиска моделей прогноза является совокупность ограничений, накладываемых на функции из класса Ψ . Вспомним, алгоритм γ есть пара $\langle \nu, \alpha \rangle$, состоящая из выборки точек ν и функции $\alpha: R^m \times \{\nu\} \rightarrow F$. Учет ограничений, наложенных на класс Ψ , определяет состав точек выборки ν и конструкцию функции α . Так, в регрессионном анализе [1] и теории аппроксимации [2] главным требованием, предъявляемым к выборке, является невырожденность соответствующей ей матрицы Грамма. Это требование сравнительно просто обеспечивается в случае управляемой выборки, когда подбор очередной точки, входящей в ν , ничем не ограничен. На деле чаще всего выборка неуправляема, и гарантировать в этих условиях выполнение каких-либо предъявляемых к ней требований затруднительно. Это обстоятельство побудило практиков к созданию алгоритмов, в которых учет накладываемых на класс Ψ ограничений полностью концентрируется в конструкции функции α . Проще говоря, в этих алгоритмах на втором этапе их работы в качестве результата отбирается та модель $f \in F$, которая по мнению практика наиболее адекватна требованиям, предъявляемым к закономерности $\phi \in \Psi$. Примеры таких алгоритмов можно найти в [3, 7, 8]. Та же идея положена в основу пакета "СИНТЕЗ" [9]. Для облегчения отбора моделей в пакете предусмотрено следующее: экзамен моделей на контрольной выборке; вывод моделей на печать в виде формул; представление временных рядов в виде графиков траекторий в пространстве переменных; расчет траекторий процессов по заданным значениям переменных управления, что необходимо для анализа реакции моделей на управление и ее селекции по этому критерию; расчет спектральных характеристик моделей, что дает возможность отбирать модели по признакам устойчивости и частотным характеристикам порождаемых ими решений, а также амплитудный гармонический анализ этих решений.

После того, как модель окончательно выбрана, она может быть использована не только для прогноза, но и для целей планирования. В пакете "СИНТЕЗ" эта функция охвачена модулями расчета значения целевой переменной по заданным значениям переменных управления, а также модулем, решающим задачу оптимального управления.

§2. Функции пакета

Поиск линейной регрессионной модели. Дана таблица "объект-переменная" $T = (\tau_{ij})_{m \times (n+1)}$. Строка i таблицы T соответствует i -му объекту, столбец j - переменной x_j , элемент τ_{ij} есть значение переменной x_j на i -м объекте. Класс моделей F состоит из линейных регрессий $x_0 = f(\tilde{x}) = \sum_{j=1}^n a_j x_j + b$, в которых коэффициенты $a_j, b \in \mathbb{R}^1$. За сложность модели f принимается величина $|\{a_j \neq 0\}|$. Фиксируется число $\epsilon \in [0, 1]$. Модель f определяется как ϵ -модель, если $\|\sum_{j=1}^n a_j \tilde{\tau}_j + b \tilde{\zeta} - \tilde{\tau}_0\| \leq \epsilon^2 \|\tilde{\tau}_0\|$, где $\tilde{\tau}_k = (\tau_{1k}, \dots, \tau_{mk}) \in \mathbb{R}^m$ и все компоненты вектора $\tilde{\zeta}$ - единицы.

ЗАДАЧА 1. Даны $\epsilon \in [0, 1]$ и таблица T . Найти линейную регрессионную ϵ -модель минимальной сложности.

Поиск квадратичной регрессионной модели. Квадратичной регрессионной моделью f называется зависимость вида:

$$x_0 = \sum_{i,j=1}^n a_{ij} x_i x_j + \sum_{u=1}^n b_u x_u + c,$$

где коэффициенты $a_{ij}, b_u, c \in \mathbb{R}^1$. Сложность модели f есть величина $|\{a_{ij} \neq 0\} \cup \{b_u \neq 0\}|$. Модель f есть ϵ -модель, если

$$\|\sum_{i,j=1}^n a_{ij} \tilde{\tau}_i \tilde{\tau}_j + \sum_{u=1}^n b_u \tilde{\tau}_u + c \tilde{\zeta} - \tilde{\tau}_0\| \leq \epsilon^2 \|\tilde{\tau}_0\|.$$

ЗАДАЧА 2. Даны $\epsilon \in [0, 1]$ и таблица T . Найти квадратичную регрессионную ϵ -модель минимальной сложности.

Технологическая цепочка поиска модели квадратичной регрессии была введена в пакет с целью проверки способности программного обеспечения пакета к его наращиванию за счет присоединения новых проблемных модулей.

Поиск авторегрессионной модели. Имеются объекты d_1, \dots, d_n и совокупность описывающих их переменных. Эта совокупность разбивается на три группы: целевую переменную x_0 , переменные управления x_{s+1}, \dots, x_n и остаток. Объединение x_0 с x_1, \dots, x_s образует множество фазовых переменных.

Для того чтобы проиллюстрировать разницу между фазовыми переменными и переменными управления, приведем пример с самолетом. Так, углы поворота рулей и тяга двигателя будут относиться к переменным управления. В то время как координаты самолета в пространстве и составляющие вектора скорости будут фазовыми переменными. Целевой переменной может быть, например, высота полета самолета. Основной задачей моделирования является прогноз поведения во времени целевой переменной. Особенность динамической задачи прогноза заключается в том, что влияние переменных управления на целевую переменную носит двойкий характер - непосредственное влияние и опосредованное через переменные x_1, \dots, x_s .

Для построения модели прогноза используется набор таблиц T_1, \dots, T_s с m строками и $n+1$ столбцами. Строка с номером u таблицы T_1 описывает состояние объекта d_u в момент времени 1 . Под авторегрессионной моделью понимается система линейных разностных уравнений (см. [10, гл.5; 6, §5.2.4; II, с.19-22]) следующего вида:

$$x_1^{i+1} = \sum_{j \in V} (a_{ij}^0 x_j^i + a_{ij}^{-1} x_j^{i-1} + \dots + a_{ij}^{-p} x_j^{i-p}) + \sum_{k \in W} (b_{ik}^1 x_k^{i+1} + b_{ik}^0 x_k^i + \dots + b_{ik}^{-p} x_k^{i-p}) + c_i, \quad (2)$$

в которой $i \in V$. Множество $V = \{0, v_1, \dots, v_q\} \subseteq \{0, 1, \dots, s\}$ состоит из номеров фазовых переменных, учитываемых в данной конкретной модели. Учет переменных управления осуществляется с помощью множества $W \subseteq \{s+1, \dots, n\}$. Число p называется глубиной модели. Индекс l отсчитывает моменты времени $p+1, p+2, p+3, \dots$. Символ x_k^{l-q} ($q=-1, 0, 1, \dots, p$) означает, что при расчете x_1 в момент времени $l+1$ по формуле (2) на место x_k^{l-q} подставляется значение переменной x_k в момент времени $l-q$. Значки a_{ij}^{-q} и b_{ik}^{-q} понимаются как коэффициенты, стоящие перед x_j^{l-q} и x_k^{l-q} , а не как записи a_{ij} и b_{ik} в степени $-q$. Коэффициенты a_{ij}^{-q} , b_{ik}^{-q} , $c_i \in R^1$. Зафиксируем глубину p . Класс моделей F_p получается путем варьирования множеств $V \subseteq \{0, \dots, s\}$, $W \subseteq \{s+1, \dots, n\}$ и произвольного выбора значений коэффициентов a_{ij}^{-q} , b_{ik}^{-q} , c_i из R^1 . Пусть t - некоторая мо-

дель из F_p со своими значениями параметров a_{1j}^{-q} , b_{1k}^{-q} , c_1 и множествами V и W . Сопоставим i -му уравнению модели f вектор $\tilde{\theta}_i = (\theta_{p+2,i}(1), \dots, \theta_{t,i}(1), \dots, \theta_{p+2,i}(m), \dots, \theta_{t,i}(m))$. Компоненты вектора подсчитываем по формуле:

$$\begin{aligned} \theta_{i+1,i}(u) = & \sum_{j \in V} (a_{1j}^0 \tau_{1,j}(u) + \dots + a_{1j}^{-p} \tau_{1-p,j}(u)) + \\ & + \sum_{k \in W} (b_{1k}^{-1} \tau_{1+1,k}(u) + \dots + b_{1k}^{-p} \tau_{1-p,k}(u)) + c_1, \end{aligned}$$

где l пробегает значения от $p+1$ до $t-1$ и $\tau_{g,r}(u)$ есть элемент таблицы T_g , стоящий на пересечении u -й строки и r -го столбца. Сформируем вектор $\tilde{\tau}_i = (\tau_{p+2,i}(1), \dots, \tau_{t,i}(1), \dots, \tau_{p+2,i}(m), \dots, \tau_{t,i}(m))$. Качеством i -го уравнения модели f называется величина

$$\chi_i(f) = \frac{1}{m(t-p-1)} \sum_{u=1}^m \sum_{i=p+2}^t (\theta_{i,i}(u) - \tau_{i,i}(u))^2.$$

За меру качества модели в целом принимается $\chi(f) = \max_{i \in V} \chi_i(f)$. Модель f есть ϵ -модель, если $\chi(f) \leq \epsilon$. Выделим в F_p подкласс ϵ -моделей $F_{\epsilon,p}$. Введем понятие сложности модели f . В отличие от переменной x_k назовем x_k^{1-q} аргументом. Аргумент x_k^{1-q} в уравнении i модели f считается действующим, если стоящий перед ним в i -м уравнении модели f коэффициент отличен от нуля. Обозначим через $Q_i(f)$ множество действующих аргументов i -го уравнения. Сложность модели f назовем число $S(f) = \|\cup_{i \in V} Q_i(f)\|$. Таким образом, $S(f)$ есть общее число действующих аргументов модели f .

ЗАДАЧА 3. Даны $\epsilon > 0$ и набор таблиц T_1, \dots, T_t . Найти ϵ -модель минимальной сложности.

Алгоритмы решения задач 1, 2 и 3 приводятся в [3].

Подчеркнем, что авторегрессионная модель, определенная выше, решает задачу одновременного описания нескольких временных рядов.

Расчет оптимального управления. Пусть для набора таблиц T_1, \dots, T_t построено несколько авторегрессионных моделей вида (2). Среди d_1, \dots, d_m выбирается некоторый объект d_u и фиксируется одна из построенных моделей. Задается область допустимых управлений в виде конечных вещественных интервалов для каждой переменной управления. Границы интервалов управления могут быть постоянными или изменяющимися с течением времени. В

качестве начальных значений переменных x_0^1, \dots, x_n^1 при $l=1, \dots, t_0$ берутся числа $\tau_{1,0}(u), \dots, \tau_{1,n}(u)$ из таблиц T_1, \dots, T_t .

ЗАДАЧА 4 [11, с.19-22]. Зная начальные значения переменных x_0^1, \dots, x_n^1 при $l=1, \dots, t_0$ ($t_0 > p+1$), выбрать такие значения переменных управления x_{s+1}^1, \dots, x_n^1 при $l=t_0+1, \dots, t_1$ из заданной области управления, которые максимизируют (или минимизируют) значение целевой переменной в момент времени t_1 .

С п е к т р а л ь н ы й и а м п л и т у д н ы й а н а л и з м о д е л е й. Задана авторегрессионная модель

$$\begin{aligned} \tilde{x}(l+1) = & A_0 \tilde{x}(l) + A_1 \tilde{x}(l-1) + \dots + A_p \tilde{x}(l-p) + \\ & + B_{-1} \tilde{u}(l+1) + B_0 \tilde{u}(l) + \dots + B_p \tilde{u}(l-p) + \tilde{d}, \end{aligned} \quad (3)$$

$$\tilde{u}(l+1) = G \tilde{u}(l) + \tilde{\delta}. \quad (4)$$

Здесь \tilde{x} - вектор фазовых переменных, $\tilde{x} = (x_0, x_1, \dots, x_n)$, \tilde{u} - вектор переменных управления, $\tilde{u} = (x_{s+1}, \dots, x_n)$, уравнение (3) есть запись системы (2) в векторном виде. Решение системы (3)-(4) можно получить, используя начальные значения векторов \tilde{x} и \tilde{u} и рекуррентные соотношения (3), (4). Но наша задача заключается в получении аналитической зависимости векторов \tilde{x} и \tilde{u} от параметра l . Если ввести новые переменные, то система (3)-(4) может быть переписана в виде равенства

$$\tilde{x}'(l+1) = D \tilde{x}'(l), \quad l = p+1, p+2, \dots, \quad (5)$$

в котором \tilde{x}' есть вектор размерности $r = p(s+1) + 2(n+1)$, а размерность матрицы D есть $r \times r$. Компоненты \tilde{x}' с номерами от 0 до s есть фазовые переменные, а с номерами от $(s+1)(p+1)$ до $(s+1)(p+1) + (n-s)-1$ - переменные управления. Для скалярного уравнения эта редукция описана в [6, §5.3]. Решение системы (5) записывается в виде $\tilde{x}'(l) = D^{l-p} \tilde{x}'(p)$. Используя технику, развитую в теории функций от матриц [12, гл.У], можно записать элементы матрицы D^{l-p} в виде выражений от l . В простейшем случае, когда все собственные значения λ_j матрицы D различны, получаем равенство

$$\tilde{x}'(l) = \sum_{j=1}^r \tilde{\epsilon}_j \lambda_j^l, \quad (6)$$

в котором $\tilde{\epsilon}_j$ есть r -мерные векторы с комплексными компонентами.

В [10, гл.У, §4] приводится вывод формулы (6) для случая скалярного конечно-разностного уравнения. Равенство (6) может быть преобразовано к виду

$$\tilde{x}'(1) = \sum_{j=1}^q \{ \tilde{c}_j^{(1)} \rho_j^1 \cos(\omega_j 1) + \tilde{c}_j^{(2)} \rho_j^1 \sin(\omega_j 1) \} + \sum_{j=q+1}^{\infty} \tilde{c}_j^{(0)} \rho_j^1, \quad (7)$$

где ρ_j есть модуль собственного числа λ_j , ω_j - его аргумент, и $\tilde{c}_j^{(k)}$ есть r -мерные вещественные векторы. Формула (7) дает аналитическую зависимость вектора \tilde{x}' от параметра 1 . В векторе \tilde{x}' нас интересуют не все компоненты, а только те, которые описывают поведение \tilde{x} и \tilde{u} , т.е. компоненты с номерами от 0 до v и от $(s+1)x$ до $(s+1)(p+1)+(n-s) - 1$. Назовем их совокупность существенным подвектором уравнения (5).

ЗАДАЧА 5. Дан набор матриц $A_0, A_1, \dots, A_p, B_{-1}, \dots, B_p, G$ и вектора \tilde{x} и \tilde{c} . Требуется найти собственные значения матрицы D .

ЗАДАЧА 6. Зная собственные значения матрицы D и начальный отрезок траектории, порожденной моделью (3)-(4), найти представление существенного подвектора системы (5) в виде (7).

Экзамен моделей, расчет вариантов. Регрессионная модель связывает целевую переменную x_0 с независимыми переменными x_1, \dots, x_n . Зафиксируем значения независимых переменных, т.е. положим $x_i = \tau_i$. Сопоставим набору τ_1, \dots, τ_n точку $\tilde{x} = (\tau_1, \dots, \tau_n) \in R^n$. Под расчетом варианта понимается вычисление значения регрессии в точке \tilde{x} . Для расчета траектории, порожденной авторегрессионной моделью, задаются начальные значения фазовых переменных $\tilde{x}(1) = (x_0^1, x_1^1, \dots, x_n^1)$ при $l=1, \dots, p+1$ и фиксируются значения переменных управления $\tilde{u}(1) = (x_{n+1}^1, \dots, x_n^1)$ при $l=1, \dots, t$ ($t > p+1$). Из соотношений (2) вычисляем $\tilde{x}(p+2), \dots, \tilde{x}(t)$. Экзамен модели заключается в сравнении вычисленных значений целевой переменной с теми ее значениями, которые содержатся в контрольных таблицах "объект-переменная".

§3. Состав программного обеспечения

Пакет "СИНТЕЗ" включает в себя библиотеку фортрановских модулей, состоящую из 31 программы, и набор JCL-процедур на языке управления заданиями [13]. Размеры программ от 10 до 1800 операторов. Часть программ играют вспомогательную роль и служат для записи информации в стандартные наборы данных, выделения необхо-

димых подмножеств данных, обработки результатов и вывода их на АЦПУ и т.п. Имеется так называемая "программа привязки", кото-рая содержит задания на создание необходимых наборов данных на диске, а также задания на трансляцию и редактирование фортранов-ских модулей. Скомпилированные и отредактированные программы по-мещаются в библиотеку загрузочных модулей. Исходя из решаемой за-дачи, пользователь системы выбирает ту или иную JCL-процедуру, ко-торая осуществляет вызов необходимых модулей и запуск их в работу в нужной последовательности. Пользователь может сам составлять уп-равляющие программы для решения своих задач. Обмен информацией между последовательно работающими программами осуществляется че-рез наборы данных на диске.

В основных программах используется динамическое распределе-ние памяти. Это позволяет решать задачи больших размерностей. Об-рабатываемые таблицы экспериментальных данных могут содержать до 50000 чисел в вещественном формате.

Пакет предназначен для эксплуатации на ЭВМ ЕС с использова-нием накопителя на магнитных дисках НМД-ЕС-5061 и накопителя на магнитных лентах НМД-ЕС-5010 и работает под управлением опера-ционной системы ОС ЕС версии 6.1.

Пакет выполнен в лаборатории технической кибернетики Новоси-бирского государственного университета. Помимо авторов настоящей статьи, над пакетом работали программисты Е.В.Агафонов, В.В.Залю-бовский, М.Г.Зюзин, Ю.Г.Жилин.

Л и т е р а т у р а

1. ДЕМИДЕНКО Е.З. Линейная и нелинейная регрессии. -М.: Фи-нансы и статистика, 1981.- 302 с.
2. БЕРЕЗИН И.С., ЖИДКОВ Н.П. Методы вычислений. -М.: Наука, 1966, т.1. - 632 с.
3. УСТОЖАНИНОВ В.Г. Модели прогноза минимальной сложности. -В кн.: Анализ разнотипных данных (Вычислительные системы, вып. 99). Новосибирск, 1983, с.88-101.
4. УСТОЖАНИНОВ В.Г. О минимизации описания обучающей выбор-ки. -В кн.: Тезисы XXI областной научно-технической конференции НТО Р и ЭС им.Попова. Новосибирск, 1978, с. 44-45.
5. УСТОЖАНИНОВ В.Г. "КОНЪЮНКЦИЯ" - программа построения ло-гических отделителей. -В кн.: Анализ разнотипных данных (Вычисли-тельные системы, вып.99). Новосибирск, 1983, с.117-119.
6. АНДЕРСОН Т. Статистический анализ временных рядов. -М.: Мир, 1976. - 755 с.

7. ЗАГОРУЙКО Н.Г., ЕЛКИНА В.Н., ТИМЕРКАЕВ В.С. Алгоритм заполнения пропусков в эмпирических таблицах (алгоритм "ЗЕТ"). - В кн.: Вычислительные системы. Вып. 61. Эмпирическое предсказание и распознавание образов. Новосибирск, 1975, с.3-27.

8. ИВАХНЕНКО А.Г. Индуктивный метод самоорганизации моделей сложных систем. - Киев: Наукова думка, 1982. - 296 с.

9. ЗАГОРУЙКО Н.Г., УСТЮЖАНИНОВ В.Г., ЖИЛИН Ю.Г. Идентификация моделей прогноза в условиях малой обучающей выборки. - В кн.: Тезисы докладов 2-й Всесоюзной школы-семинара "Программно-алгоритмическое обеспечение прикладного статистического анализа". М., 1983, с.140-142.

10. ГЕЛЬФОНД А.О. Исчисление конечных разностей. - М.: Наука, 1967. - 375 с.

11. БОЛТЯНСКИЙ В.Г. Оптимальное управление дискретными системами. - М.: Наука, 1973. - 446 с.

12. ГАНТМАХЕР Ф.Р. Теория матриц. - М.: Наука, 1967. - 575 с.

13. Единая система электронных вычислительных машин. Операционная система. Язык управления заданиями. Описание языка. П 51.804.006Д2, 1980 г.

Поступила в ред.-изд.отд.
19 апреля 1985 года