

Рассмотрим графики месячного распределения интенсивности нагрузок в Эрлангах (Эрл.), по up-link и down-link на примере двухдиапазонных трехсекторных базовых станций. Результаты за период с 1 июня по 30 июня 2008 г. представлены на рис. 1–12.

Базовая станция № 2 находится в центральном районе города. В данном случае она обеспечивает покрытие в основном коммерческим предприятиям. Базовая станция № 1 расположена ближе к спальному району г. Красноярска, интенсивность нагрузки заметно слабее, чем на БС № 2. Интенсивность входящей и исходящей нагрузки в различные дни месяца неодинакова, причем кроме случайных колебаний имеют место и регулярные колебания. Из графиков видно, что максимальные значения нагрузки приходятся на рабочие дни недели, на пятницу – спад нагрузки, в субботу и воскресенье – минимальные значения. Вполне оправдан спад нагрузки в период с 12 по 15 июня, так как отмечались празднования Дня Независимости Российской Федерации. Стоит обратить внимание на то, что 12 июня, БС № 2 была загружена. Это обус-

ловлено миграцией населения с окраин в центр города на празднования.

Диапазон 900 МГц недогружен (рис. 1–6). Максимальное значение ЧНН для сектора 1 – 900 МГц на БС № 1 – 94,5 Эрл. (20 июня, пятница). Максимальное значение ЧНН для сектора 1 – 900 МГц на БС № 2 – 634,5 Эрл. (12 июня, четверг). Минимальное значение ЧНН для сектора 2 – 900 МГц на БС № 1 – 10,9 Эрл. (1 июня, воскресенье). Минимальное значение ЧНН для сектора 2 – 900 МГц на БС № 2 – 41,2 Эрл.

Таким образом, можно сделать следующие выводы: анализ нагрузки на базовых станциях сотовой сети показал, что базовая станция вторая загружена в шесть раз больше, чем первая; при высокой загрузке базовой станции большие потери вызовов абонентов; анализ нагрузки необходим для корректировки места расположения базовых станций, с целью создания оптимальных режимов работы абонентов и снижения потерь вызовов.

M. N. Petrov, D. G. Mikhailov

ANALYSIS OF LOAD ON THE CELLULAR NETWORKS OF STANDARD GSM IN KRASNOYARSK

The systems of this kind came not long ago and some companies such as «ETK», «MTS», «Megafon», «Beeline» conduct service. The customer number has sizably exceeded the stationary number. Cellular systems have some features in contrast to stationary. That is why the teletraffic results can be used for analysis, design and calculation of cellular systems. It is necessary to carry out statistics analysis for showing up the regularities of modification capacity taking into account the peculiarities.

Keywords: the cellular, network, loading, losses, service.

УДК 519.876

И. М. Митасов, А. Н. Завьялкин

МЕТОД ПОЛНОГО ПЕРЕБОРА В ЗАДАЧЕ МНОГОФАКТОРНОГО РЕГРЕССИОННОГО АНАЛИЗА

Рассматривается задача последовательного усложнения многофакторной регрессионной модели с полным перебором всех возможных сочетаний факторов при заданном их количестве. Предлагается программно реализуемый алгоритм полного перебора, для построения наилучшей модели при фиксированном количестве факторов, что позволяет выбрать по известным критериям значимости оптимальную модель из множества наилучших моделей для различного количества факторов.

Ключевые слова: алгоритм, факторы, модель, выборка, оптимальная.

Важной проблемой многофакторного регрессионного анализа является выбор наиболее значимых факторов и построение модели, определяющей значения моделируемого показателя в зависимости от значений этих факторов.

Проблема состоит в выборе количества факторов и состава факторов при заданном количестве.

Обычно рассматривается полный перебор для двух

факторов, а затем дополнение полученных выборок факторов с помощью различных алгоритмов сокращенного перебора.

В работе предлагается метод решения указанной проблемы на основе программно реализуемого алгоритма полного перебора выборок факторов произвольной длины.

1. Постановка задачи. Линейная многофакторная модель регрессионного анализа имеет следующий вид:

$$y = \langle A, X \rangle + c, \quad (1)$$

где y – моделируемый скалярный показатель; A – вектор параметров модели; X – вектор факторов; c – скалярный параметр модели.

Построение модели (1) заключается в определении значений параметров A, c на основе выборки V следующего вида:

$$V = \{y^i, X^i; i = \overline{1, m}\}, \quad (2)$$

где m – количество наблюдений значений характеристик y, X .

Пусть вектор X представляет собой совокупность n факторов – x_1, x_2, \dots, x_n . Выбирая различные по составу выборки из n факторов по k ($k \leq n$) и формируя на основе (2) соответствующие выборки V^k , получаем методом наименьших квадратов (НК) k -факторные модели показателя y .

Пусть k^* – заданное экспертами максимальное количество факторов, которое целесообразно включать в модель (1), $k^* \leq n$. Если k^* не задано, то $k^* = n$.

Требуется рассмотреть значения $k = 1, k^*$, для каждого фиксированного k построить все возможные модели и выбрать наилучшую модель $(A_{\text{опт}}^k, c_{\text{опт}}^k)$ по критерию следующего вида:

$$S(A_{\text{опт}}^k, c_{\text{опт}}^k) = \min_{l=1, C_n^k} S(A_l^k, c_l^k),$$

где (A_l^k, c_l^k) определяются методом НК на основе выборки $V_l^k \in V$ следующего вида:

$$V_l^k = \{y^i, x_j^i; i = \overline{1, m}; j \in (j_1^l, \dots, j_k^l)\}.$$

Таким образом, будут получены k^* наилучших моделей. Оптимальная модель среди k^* наилучших моделей определяется по некоторому критерию значимости увеличения количества факторов.

2. Построение k -факторной оптимальной модели. Количество различных моделей при фиксированном k определяется числом сочетаний C_n^k .

Если существует алгоритм последовательного выбора k факторов из n факторов, который приводит к полному перебору всех возможных выборок, то оптимальная k -факторная модель может быть найдена известной процедурой поиска минимального элемента среди C_n^k чисел, представляющих собой значение критерия НК для соответствующей модели.

Очевидно, что алгоритм выбора факторов сводится к алгоритму выбора номеров этих факторов, т. е. построение k -факторной оптимальной модели сводится к построению программно-реализуемого алгоритма генерирования выборок длины k , различных по составу из совокупности чисел $(1, 2, \dots, n)$.

Рассмотрим следующий алгоритм генерирования выборок длины k , различных по составу из совокупности чисел $(1, 2, \dots, n)$:

1. В качестве исходной, для выборок из совокупности $\{1, 2, \dots, n\}$ по k чисел, рассматривается выборка $\{1, 2, \dots, k\}$.

2. Последовательное увеличение на единицу k -го элемента исходной выборки позволяет получить исходную матрицу D выборок

$$D = \begin{pmatrix} 1 & 2 & \dots & k \\ 1 & 2 & \dots & k+1 \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ 1 & 2 & \dots & n \end{pmatrix}.$$

Матрица D содержит $(n - k + 1)$ строк-выборок. Все строки представляют собой выборки, различные по составу.

3. Выбирается первая строка матрицы D и из нее формируется матрица D^1 по следующему алгоритму:

3.1. На первом шаге последовательно увеличиваются на единицу элементы $(k - 1, k)$ первой строки матрицы D пока на k -м месте не появится число n . Например, первой строкой матрицы D^1 будет выборка $(1, 2, \dots, k, k + 1)$.

3.2. На втором шаге последовательно увеличиваются на единицу элементы $(k - 2, k - 1, k)$ первой строки матрицы D и строк матрицы D^1 , для которых эта операция не приводит к получению на k -м месте числа, большего n , пока на k -м месте не появится число n .

3.3. Процесс формирования различных по составу выборок из первой строки матрицы D завершается на $(k - 1)$ -м шаге последовательным увеличением на единицу элементов $(1, 2, \dots, k)$ первой строки матрицы D и строк матрицы D^1 , для которых эта операция не приводит к получению на k -м месте числа большего n , пока на k -м месте не появится число n .

4. Выбирается вторая строка матрицы D и из нее формируется матрица D^2 по алгоритму, аналогичному 3.1, 3.2, 3.3.

5. Процесс выбора строк матрицы D и формирования из них выборок по алгоритму, аналогичному 3.1, 3.2, 3.3, продолжается до строки $(n - k)$.

Для обоснования алгоритма необходимо доказать следующее.

Утверждение 1. Все выборки, получаемые в силу пп. 1–5 алгоритма различны по составу.

Утверждение 2. Алгоритм позволяет получить произвольную выборку (j_1, j_2, \dots, j_k) из множества выборок, различных по составу.

Доказательство. Получение новых выборок на основе первой строки матрицы D согласно алгоритму сводится к сложению этой строки со следующими строками:

$$\begin{aligned} &(0, 0, \dots, 1, 1); (0, 0, \dots, 2, 2); \dots; (0, 0, \dots, n - k, n - k); \\ &(0, 0, \dots, 1, 1, 1); \dots; (0, 0, \dots, n - k, n - k, n - k); \\ &(0, 0, \dots, 1, 1, 1, 1); \dots (1, 1, \dots, 1) \dots (n - k, n - k, \dots, n - k). \end{aligned}$$

Так как все вышеперечисленные строки различны, то их сложение с одной и той же строкой приведет к получению различных строк. Таким образом, матрица D^1 состоит из различных строк.

Формирование матриц D^2, D^3, \dots, D^{n-k} представляет собой добавление к исходным строкам этих матриц выше перечисленных строк, которые использовались для формирования строк матрицы D^1 путем их сложения с исходной строкой матрицы D^1 . Так как исходные строки матриц D^2, D^3, \dots, D^{n-k} отличаются только k -м элементом, то устранить это различие добавление к ним строк, изменяющих не менее двух элементов исходной выборки, невозможно.

Следовательно, строки матриц $D^1, D^2, D^3, \dots, D^{n-k}$ представляют собой различные по составу выборки из n по k чисел.

Для доказательства утверждения 2 рассмотрим произвольную выборку (j_1, j_2, \dots, j_k) из множества выборок, различных по составу.

Не ограничивая общности, можно предполагать, что $j_1 < j_2 < \dots < j_k$. Тогда

$$j_i \geq j_{i-1} + 1, \quad i = \overline{2, k},$$

следовательно

$$j_1 + (k-1) \leq j_k \leq n. \quad (3)$$

На первом шаге рассмотрим исходную выборку $(1, 2, \dots, k)$ и прибавим ко всем ее элементам число (j_1-1) , тогда получим выборку $(j_1, j_1+1, \dots, j_1+(k-1))$. Такое преобразование исходной выборки соответствует алгоритму и допустимо в силу (3).

На втором шаге прибавим ко всем элементам полученной выборки, начиная со второго, число $((j_2-j_1)-1)$, тогда получим выборку $(j_1, j_2, j_2+1, \dots, j_2+(k-2))$. Такое преобразование исходной выборки соответствует алгоритму и допустимо, так как $j_2+(k-2) \leq j_k$.

Продолжая этот процесс, на $(k-1)$ -м шаге имеем выборку $(j_1, j_2, \dots, j_{k-1}, j_{k-1}+1)$. Прибавим к последнему элементу этой выборки число $(j_k-j_{k-1}-1)$ и получим произвольную выборку путем преобразования исходной выборки по пп. 1–5 алгоритма.

Таким образом, доказано, что все выборки, получаемые по пп. 1–5 алгоритма различны по составу и что любая выборка из C_k^n различных по составу выборок может быть получена на основе пп. 1–5 алгоритма.

Алгоритм последовательного формирования выборок номеров k факторов, включаемых в модель (1) позволяет найти k -факторную оптимальную модель следующим образом:

1. Рассматривается первая выборка номеров факторов и на основе исходной выборки V формируются статистические данные V_1^k . На основе этих данных методом НК строится первая модель вида (1) и вычисляется ее качество: $S_1 = S(A_1^k, c_1^k)$. Эта модель считается оптимальной и ее параметры запоминаются.

2. Рассматривается вторая выборка номеров факторов и на основе исходной выборки V формируются статистические данные V_2^k . На основе этих данных методом НК строится вторая модель вида (1) и вычисляется ее качество: $S_2 = S(A_2^k, c_2^k)$. Если $S_2 < S_1$, то вторая модель оптимальная, ее параметры запоминаются, а параметры

предыдущей модели стираются. Если $S_1 \leq S_2$, то предыдущая модель остается оптимальной и рассматривается очередная модель.

3. Процесс рассмотрения новых выборок номеров факторов и сравнения качества новой модели и наилучшей на предыдущем шаге продолжается пока не будут рассмотрены все возможные выборки из n по k .

3. Выбор оптимальной модели из множества наилучших моделей с различным количеством факторов. Пусть на основе алгоритмов, изложенных в п. 2, получены k^* наилучших моделей, содержащих $1, 2, \dots, k^*$ факторов.

Пусть $S_{\text{опт}}^k$ – значение $S(A_{\text{опт}}^k, c_{\text{опт}}^k)$ для оптимальной k -факторной модели.

Рассмотрим, например, следующий критерий значимости:

– модель, имеющая k параметров и значение критерия качества $S_{\text{опт}}^k$ лучше моделей, имеющих i параметров ($i < k$), и значение критерия качества $\{S_{\text{опт}}^i; i = \overline{1, k-1}\}$, если выполнена следующая система условий:

$$S_{\text{опт}}^k \leq S_{\text{опт}}^i \quad (i/k); \quad i = \overline{1, k-1}. \quad (4)$$

Условия (4) являются достаточно жесткими, но их можно смягчить, рассматривая вместо (i/k) коэффициент $(i+\alpha)/k, 0 < \alpha < 1$.

4. Результаты программной реализации алгоритмов. Разработанные алгоритмы [1] реализованы в виде программы на ЯОН C++Builder 6.0.

Рассматривалась следующая модельная задача.

Пусть $m = 4, k^* = 2, n = 5$. Необходимо построить все однофакторные и двухфакторные модели. Найти наилучшую однофакторную и двухфакторную модель. Затем, среди наилучших, найти оптимальную по критерию значимости (4).

Числовые данные

$$X^1 = \begin{pmatrix} 1 \\ 3 \\ 5 \\ -2 \end{pmatrix}, \quad X^2 = \begin{pmatrix} 2 \\ 7 \\ 4 \\ 5 \end{pmatrix},$$

$$X^3 = \begin{pmatrix} 5 \\ 6 \\ 8 \\ -3 \end{pmatrix}, \quad X^4 = \begin{pmatrix} 6 \\ 7 \\ 2 \\ 3 \end{pmatrix}, \quad X^5 = \begin{pmatrix} 4 \\ 9 \\ 5 \\ -1 \end{pmatrix},$$

$$y^1 = 4, \quad y^2 = 6, \quad y^3 = 7, \quad y^4 = 11.$$

Результаты расчетов представлены в табл. 1–2.

Таблица 1

Однофакторные модели

Номер модели	Номера факторов	Параметры модели	Ошибка модели	Наилучшая модель
1	5	$c = 9,093\ 596$ $a_5 = -0,492\ 610$	13,684 729	
2	1	$c = 7,915\ 888$ $a_1 = -0,523\ 364$	18,672 897	
3	2	$c = 4,576\ 923$ $a_2 = 0,538\ 462$	22,230 769	
4	3	$c = 8,885\ 714$ $a_3 = -0,471\ 429$	10,442 857	3
5	4	$c = 10,441\ 176$ $a_4 = -0,471\ 429$	16,058 824	

Сравнивая наилучшие модели по критерию (4), получаем, что оптимальной по составу и количеству факторов является следующая модель:

$$y = 1,947\,781 \cdot X_1 - 1,612\,272 \cdot X_3 + 10,040\,470.$$

Таким образом, в работе получены следующие результаты: предложен и обоснован алгоритм полного перебора факторов в задаче выбора оптимального состава и количества факторов для многофакторной регрессионной модели; рассмотрены результаты применения программной реализации алгоритма на модельной задаче.

Библиографический список

1. Митасов, И. М. Автоматизированная система построения линейной многофакторной регрессионной модели оптимальной по количеству и составу факторов / И. М. Митасов, А. Н. Завьялкин. М. : ОФАП. Свидетельство № 11399 от 12.09.2008.

Таблица 2

Двухфакторные модели

Номер модели	Номера факторов	Параметры модели	Ошибка модели	Наилучшая модель
1	1, 5	$c = 9,088\,919$ $a_1 = -0,007\,596$ $a_5 = -0,488\,382$	13,684 093	
2	1, 2	$c = 5,262\,663$ $a_1 = -0,557\,164$ $a_2 = 0,602\,750$	13,980 463	
3	2, 3	$c = 6,858\,491$ $a_2 = 0,433\,962$ $a_3 = -0,452\,830$	8,018 868	
4	3, 4	$c = 11,325\,592$ $a_3 = -0,411\,919$ $a_4 = -0,595\,092$	4,670 465	
5	4, 5	$c = 10\,305\,481$ $a_4 = -0,399\,371$ $a_5 = -0,354\,897$	11,935 759	
6	1, 3	$c = 10,040\,470$ $a_1 = 1,947\,781$ $a_3 = -1,612\,272$	0,0639 687	6
7	2, 4	$c = 7,679\,245$ $a_2 = 0,745\,283$ $a_4 = -0,896\,226$	9,132 075	
8	3, 5	$c = 9,053\,191$ $a_3 = -0,390\,578$ $a_5 = -0,115\,502$	10,223 404	
9	1, 4	$c = 11,453\,795$ $a_1 = -0,537\,954$ $a_4 = -0,780\,528$	8,321 782	
10	2, 5	$c = 5,237\,928$ $a_2 = 1,040\,825$ $a_5 = -0,687\,445$	1,528 095	

I. M. Mitasov, A. N. Zavyalkin

METHOD OF FULL SORTING OUT IN THE PROBLEM OF REGRESSIVE POLIFACTOR-ANALYSIS

The paper is considers the problem of the sequential complication of the regressive polyfactor-model with full sorting out of all the possible factors combinations at their given number. There is offered the program-realized algorithm of full sorting out for the construction of the best model with the fixed number of factors, that permits to select an optimal model from the set of the best models for the different number of factors according to the well-known criteria of significans.

Keywords: algorithm, factors, model, sorting out, optimum.