УДК 519.71

МЕТОДЫ ОТЫСКАНИЯ ЗНАЧИМЫХ ФАКТОРОВ ПРИ МОДЕЛИРОВАНИИ ТЕХНОЛОГИЧЕСКИХ ПРОЦЕССОВ

© 2015 г. *О. В. Шестопал*

Южно-Российский государственный политехнический университет (Новочеркасский политехнический институт)

В статье с помощью модифицированного метода случайного баланса определяется комплекс показателей, включающий значимые входные параметры. Для этого определяются параметры оценки для построения модели. Полученная математическая модель проверяется на адекватность по критерию Пирсона. Такая модель позволяет привести массив исходных данных к виду, пригодному для построения математической модели другими более точными методами с целью прогнозирования, внедрения новых видов стали и создание основы для разработки системы автоматизированного управления качеством продукции.

Ключевые слова: *модифицированный метод случайного баланса; математическая модель; критерий Пирсона; моделирование технологического процесса.*

In the article, a set of indicators, which includes significant input parameters, is determined using a modified method of random balance. The parameters for the model evaluation were defined to complete this task. The resulting mathematical model is tested for adequacy by Pearson criterion. This model allows us to convert an array of raw data to a form suitable for the construction of a mathematical model by other more accurate methods to predict the introduction of new types of steel and provides a basis for the development of automated quality control.

Key words: *modified method of random balance; mathematical model; Pearson criterion; modeling process.*

Введение

Перед проведением работ по получению математической модели, согласно алгоритму из [7] во всех случаях рекомендуется сократить первоначальный список факторов до возможного минимума, так как с ростом числа факторов трудоемкость моделирования растет как степенная функция. Отсев факторов можно производить по двум критериям: факторы незначимые, то есть не влияющие на целевую функцию и внесенные в первоначальный список факторов ошибочно, и факторы коррелированные, то есть имеющие сильную внутреннюю связь.

1. Методы сокращения факторного пространства

Рассмотренные методы применялись при обработке первичных данных, отражающих технологический процесс выплавки стали. Первоначально список параметров сокращается за счет объединения однотипных факторов. Данную операцию выполняют эксперты-технологи, используя опыт, знания и собственную интуицию.

Вторым методом сокращения количества факторов была процедура выделения групп сильно коррелированных параметров, описанная в [8]. Естественно, что каждая группа таких факторов должна быть разбита, то есть один из факторов отброшен как не дающий дополнительной информации в будущей математической модели, а другой оставлен для дальнейшей работы. К сожалению, нет никаких формальных критериев, по которым можно судить, какой именно фактор должен быть отброшен, а какой оставлен — это в большей мере вопрос удобства дальнейшей работы, интуиции и опыта исследователя, возможностей измерительного оборудования и т. п. [5; 6].

С целью сокращения факторного пространства составляются корреляционные матрицы различных мер связи (коэффициентов корреляции, корреляционных отношений и др.) и по ним построены корреляционные плеяды. Их анализ [1] позволил сократить число факторов для проведения моделирования с 107 до 21. Критериями отбора были одновременное выполнение условий: максимальный коэффициент внутренней корреляционной связи и некоррелированность или слабая коррелированность с представителями других плеяд.

Эти параметры были включены как исходные данные для построения модели. Так как модель технологического процесса выплавки стали формируется по пассивным данным впервые, то в качестве метода моделирования был выбран модифицированный метод случайного баланса по пассивным данным (ММСБП), описанный в [8].

2. Модифицированный метод случайного баланса

Исследования показали, что процедура ММСБП дает хорошие результаты даже при значительном отклонении распределения выходной величины от нормального закона.

Любые факторные планы одним из предварительных этапов в планировании имеют переход от координат с абсолютными единицами измерения факторов к координатам с относительными единицами, где единичной мерой служит шаг ΔX_k, свой для каждого фактора. Тем самым достигается преобразование координат таким образом, что в факторном пространстве получаются концентрические гиперсферы, а не другие фигуры, а выбор вершин гиперкуба в качестве точек проведения активного эксперимента автоматически обеспечивает выбор только одной гиперсферы [3; 4].

В случае пассивного эксперимента каждый фактор X_k имеет в таблице экспериментальных данных целый диапазон значений от $X_k min$ до $X_k max$, которые могут быть рассмотрены как выборка с центром $\overline{X}_k$. С этой целью и для увеличения точности результатов будущих расчетов весь диапазон $X_k max - X_k min$ каждого фактора X_k следует разбить на три части таким образом, чтобы число попаданий в каждую из них было примерно одинаковым, при этом части следует кодировать символами –1, 0 и +1. Хотя вид закона распределения факторов не оговаривается, из практики известно, что в подавляющем большинстве случаев они унимодальные (одновершинные), и для них можно оговорить правило: все значения $X_k < \overline{X}_k - 0{,}5S_k$ будут относиться к области $x_k = -1$, все значения $X_k \geq \overline{X}_k - 0{,}5S_k$ — к области $x_k = +1$, а остальные значения X_k — к области $x_k = 0$ (здесь $\overline{X}_k$ — среднеарифметическое, S_k — среднеквадратическое отклонение числового фактора X_k, определенное по достаточно большому объему выборки, причем символом X_k обозначаются значения k-го фактора в абсолютных единицах, а x_k — в относительных. В результате исходная таблица с контрольно-измерительной информацией превращается в план квазиактивного эксперимента.

Гомокседаксичность в квазиактивном плане ММСБ нарушается, поэтому для расчетов оценок коэффициентов регрессии b_k и их дисперсии D_k следует использовать специальные выражения [2], учитывающие поправки на это нарушение гомоскедастичности (гетероскедаксичность) и являющиеся в этих условиях более эффективными, чем другие оценки:

$$b_k = \frac{\left(\frac{D_{2k}}{N_{2k}} + 2m_k^2\right)\mu_{1k} - \left(\frac{D_{1k}}{N_{1k}} + 2m_k^2\right)\mu_{2k}}{\frac{D_{1k}}{N_{1k}} + \frac{D_{2k}}{N_{2k}} + 4m_k^2}, \quad (1)$$

$$D_k = \frac{\frac{D_{1k}}{N_{1k}}\frac{D_{2k}}{N_{2k}} + \left(\frac{D_{1k}}{N_{1k}} + \frac{D_{2k}}{N_{2k}}\right)m_k^2}{\frac{D_{1k}}{N_{1k}} + \frac{D_{2k}}{N_{2k}} + 4m_k^2}. \quad (2)$$

Здесь:

$$\mu_{1k} = \frac{1}{N_{1k}}\sum_{j=1}^{N_{1k}} Y_j^{(1k)}; \quad (3.1)$$

$$\mu_{2k} = \frac{1}{N_{2k}}\sum_{j=1}^{N_{2k}} Y_j^{(2k)}, \quad (3.2)$$

где $\left\{Y_j^{(1k)}\right\}_{N_{1k}}$ и $\left\{Y_j^{(2k)}\right\}_{N_{2k}}$ — подмножества элементов выходной величины из общей выборки, для которых x_{kj} имеет соответственно положительный или отрицательный знак;

N_{1k}, N_{2k} — объем соответствующих подмножеств, причем $N_k = N_{1k} + N_{2k}$ — общий объем выборки для k-го фактора;

$m_k = \frac{1}{N_k}(\mu_{1k}N_{1k} + \mu_{2k}N_{2k})$ — оценка математического ожидания;

$$D_{1k} = \frac{1}{N_{1k}-1}\sum_{j=1}^{N_{1k}}(Y_j^{(1k)} - \mu_{1k})^2;$$

$$D_{2k} = \frac{1}{N_{2k}-1}\sum_{j=1}^{N_{2k}}(Y_j^{(2k)} - \mu_{2k})^2 —$$

дисперсии выходной величины соответственно при положительных и отрицательных значениях фактора x_k.

С помощью формул можно определить значимость каждой полученной оценки коэффициента регрессии по критерию Стьюдента. При выполнении условия

$$t_k = \frac{|b_k|}{\sqrt{D_k}} \ge t_{tab}(q;v_k), \qquad (4)$$

с уровнем значимости q и числом степеней свободы $v_k = N_k - 2$ оценки b_k признаются значимыми и должны быть включены в математическую модель. Результаты расчетов представлены в табл. 1.

Следует подчеркнуть, что в отличие от активного эксперимента величины m_k для каждого k-го фактора будут своими и, в силу этого, оценку b_0 следует искать как среднее арифметическое всех значений выходной величины Y, полученных экспериментальным путем.

3. Формирование модели из конечного числа факторов

Поскольку значимыми факторами следует признать x_3, x_4, x_8, x_{14}, x_{17}, x_{18}, а величина $b_0 = 64,07$, то искомая модель может быть представлена в виде:

$$\hat{Y} = 64,07 - 1,6x_3 - 3,04x_4 + \\ + 1,65x_8 + 1,59x_{14} - 1,51x_{17} + 2,18x_{18}. \qquad (5)$$

Окончательно вопрос о включении факторов в уравнение модели решается на стадии проверки адекватности модели, которая, в силу отклонения выходной величины Y от нормального закона, должно проводиться по критерию К. Пирсона:

$$\chi^2 = \sum_1^7 \frac{\left(n_j - \tilde{n}_j\right)^2}{\tilde{n}_j} = 9,55. \qquad (6)$$

Полученные значения χ^2 меньше табличного значения χ^2 (5%, 6) = 14,07. Таким образом, найденная модель правильно отражает

Таблица 1

Основные параметры оценки

Параметры	Факторы					
	x_3	x_4	x_8	x_{14}	x_{17}	x_{18}
N_{1k}	33	18	25	26	27	29
μ_{1k}	82,447	80,062	85,731	85,608	82,038	85,444
D_{1k}	24,057	6,333	21,518	15	13,183	15,927
N_{2k}	32	43	35	32	37	31
μ_{2k}	85,635	86,157	82,431	82,432	85,055	81,078
D_{2k}	21,775	49,581	26,518	27,409	30,79	29,255
m_k	84,015	84,359	83,806	83,855	83,782	83,188
b_k	–1,595	–3,045	1,649	1,588	–1,507	2,184
D_k	0,352	0,376	0,404	0,358	0,330	0,373
t_k	2,688	4,964	2,593	2,654	2,624	3,574
$t_{табл}$	1,97					

экспериментальные данные и может быть использована для анализа работы и для оптимизации исследуемого объекта. Любопытно отметить, что в данном случае критерий Фишера также подтвердил правильность нахождения модели:

$$F = \frac{S_{ad}^2}{S_p^2} = \frac{17,22}{20,94} < 1. \quad (7)$$

Заключение

Результаты исследования показали, что из 106 факторов наиболее значимыми с точки зрения управления процессом оказались шесть. Эти шесть параметров вошли в модель, полученную с помощью ММСБП, который основан на приведении таблицы числовых данных к квазиактивному плану-матрице факторов, оценки коэффициентов регрессии которых рассчитываются с учетом гетероскедастичности данных.

Очистка данных от грубых промахов и определение веса каждого значимого фактора в ММСБП делает массив исходных данных пригодным для моделирования другими, более точными методами.

Литература

1. *Boswijk H. P.* Asimptotic Theory for Integrated Processes. — Oxford University Press, 1999.

2. *Cameron A. C.* Regression Analysis of Count Data. / A. C. Cameron, P. K. Trivedi. — Cambridge University Press, 1998.

3. An-squared measure of Goodness of Fit for Some Common Nonlinear Regression Models. // Journal of Econometrics. — 1997. — №77. — Pp. 329–342.

4. *Dogerty K.* Introduction to Econometrics. — The 3th ed. — Oxford University Press, 2006.

5. *Phillips P. C.* Linear Regression Limit Theory for Nonstationary Panel Data. / P. C. Phillips, H. R. Moon. // Econometrica. — 1999. — Vol. 67, №5. — Pp. 1057–1111.

6. Stohastic check for control of electronic wares quality. // Trans. of 10th International Symposium on Applied stochastic Models and Data Analysis. Univ. de Techn. de Compiegne, France. — June 12–15 2001. — V. 1. — Pp. 387–390.

7. *Долгов Ю. А.* Схема математического моделирования технологического процесса плавки стали. / Ю. А. Долгов, Л. Я. Козак, О. В. Шестопал. // Радіоелектронні і комп’ютерні системи. — 2010. — №7 (48). — С. 157–160.

8. *Долгов Ю. А.* Статистическое моделирование: Учебник для вузов. — 2-е изд., доп. — Тирасполь: Изд-во Приднестр. ун-та, 2011. — 349 с.

Поступила в редакцию *2 марта 2015 г.*

Оксана Викторовна Шестопал — аспирант кафедры прикладной математики Южно-Российского государственного политехнического университета (Новочеркасского политехнического института).

Oksana Viktorovna Shestopal — graduate student at the Department of Applied Mathematics of the South-Russian State Polytechnical University (Novocherkassk Polytechnic Institute).

346428, г. Новочеркасск, ул. Просвещения, 132
132 Prosveshcheniya st., 346428, Novocherkassk, Rostov reg., Russia
Тел.: +7 8635 25 56 92; e-mail: npi_pm@mail.ru