

УДК 519.71

СРЕДСТВА РЕАЛИЗАЦИИ ПРОЦЕДУРЫ ОТСЕВА ГРУБЫХ ПРОМАХОВ

Л. Я. КОЗАК

Ludmilayaroslavovna@gmail.com

ГОУ ВО «Приднестровский государственный университет им. Т. Г. Шевченко» (ПГУ)

Поступила в редакцию 01.09.2016

Аннотация. Описана последовательность приведения исходных показателей технологического процесса выплавки стали к виду, пригодному для построения математической модели с целью прогнозирования, внедрения новых видов стали и создания основы для разработки системы автоматизированного управления качеством продукции.

Ключевые слова: пассивный эксперимент, двумерное распределение, математическая модель, грубые промахи, моделирование технологического процесса.

ВВЕДЕНИЕ

На практике увеличение числа повторных измерений часто приводит к появлению в результатах ошибочных значений, резко отличающихся от остальных измерений. Главным образом это вызвано антропогенным фактором, устранение которого по тем или иным причинам не возможно. Такие ошибки должны и могут быть устранены на этапе первичной обработки данных, для этого применяются различные методы отсева грубых промахов.

Данная проблема возникла при применении статистических методов обработки пассивных данных технологического процесса выплавки стали на Молдавском металлургическом заводе (г. Рыбница, Молдова).

ПРИМЕНЕНИЕ МЕТОДА ДВУМЕРНОГО РАСПРЕДЕЛЕНИЯ

Размер таблицы исходных данных составляет около 500 000 значений. Такая длинная таблица с множеством чисел может содержать и ошибочные данные, поэтому перед дальнейшей работой все значения должны быть проверены на грубые промахи любым из известных способов, а выявленные промахи удалены, иначе статистический анализ может дать неверные выводы.

Таблица двумерного распределения (табл. 1) дает дополнительную возможность избавиться от грубых промахов, которые невозможно вы-

явить в одномерных выборках. Ее удобно применять для исследования распределения парной выборки (X_i, Y_i) двух случайных величин. Для этого обе выборки упорядочиваются и разбиваются на k интервалов. После этого строится таблица с числом внутренних клеток $k \times k$, строки которой соответствуют интервалам выборки X , а столбцы – интервалам выборки Y . Далее из двух оригинальных (неупорядоченных) выборок последовательно выбираются пары (X_i, Y_i) и определяются номера интервалов I_x, I_y , в которые попадают значения X_i и Y_i . На пересечении строки I_x и столбца I_y ставится отметка о попадании (увеличивается счетчик). Исчерпав обе выборки, для каждой ячейки подсчитывается количество попаданий. Таким образом, полученная таблица будет содержать дискретные частоты появления пар (X, Y) [1].

Для удобства данную таблицу расширяют, добавляя столбец слева, в который записывают среднее значение \bar{X}_j интервала I_j выборки X ; строку сверху, содержащую среднее значение \bar{Y}_l интервала I_l выборки Y ; два столбца справа и еще одну строку снизу. В предпоследнем и последнем столбцах для каждой строки подсчитываются сумма частот по строкам и среднее значение \bar{Y}_j , последняя же строка содержит суммы частот по столбцам. Таким образом, предпоследний столбец последней строки будет содержать общее число элементов двумерной выборки [2–4].

Таблица 1

Таблица двумерного распределения

\bar{X}_j	\bar{Y}_l									$\sum_{l=1}^k n_{jl}$	\bar{Y}_j
	52,7	56,2	59,7	63,2	66,7	70,2	73,7	77,2	80,7		
140		5	1							6	58,8
164		6	4	2	2					14	59,7
188	4	5	7	5	2					23	59,1
212	2	8	11	10	5	1				37	60,7
236	3	4	9	4	4	2				26	60,8
260	1	2	3	5	7	4	3	1		26	65,6
286				4	5	3	4	1	1	21	69,5
308						6	5	5		13	74,9
332								4	6	10	79,3
$\sum_{j=1}^k n_{jl}$	10	30	35	30	25	16	12	11	7	176	–

Для двумерной таблицы грубым промахом считаются пары, попавшие в клетку, отстоящую от основного массива данных одновременно не менее чем на одну пустую клетку по горизонтали и по вертикали. Другими словами, если в таблице двумерного распределения есть одиноко стоящая заполненная клетка, вокруг которой есть хотя бы по одной пустой клетке – то это двумерное выделяющееся значение [5].

Стандартные статистические пакеты не позволяют реализовать данный метод. Применение других методов отсева промахов требует соблюдения нормального закона распределения, что невозможно гарантировать в исследуемых данных.

Применение пакета MathModel (разработчики НИЛ «Математическое моделирование», г. Тирасполь, Молдова) позволяет провести всю цепочку работ над экспериментальными данными, начиная от очистки данных от грубых промахов, отбора информативных параметров, до получения различных статистических математических моделей и оценки качества путем сравнения их информационных емкостей.

Данные измерения могут быть записаны в файле на диске в любом из трех форматов: двоичный (*.prq), текстовый (*.txt) или файл баз данных (*.dbf). В двоичном файле первые два числа целого типа и должны содержать количество строк и столбцов данных в таблице измерений [6].

Программа реализует достаточное количество методов статистического моделирования

по пассивным данным, но есть ряд недостатков:

1. При создании нового проекта нет возможности вручную ввести значения для вычислений.
2. Ограничено количество параметров для расчетов, которое не должно превышать 50.
3. Программа не позволяет сохранять промежуточные данные, так как не использованы базы данных.
4. Не предусмотрена сетевая версия программного продукта.

СРЕДСТВА РЕАЛИЗАЦИИ МЕТОДА

Разработанный нами программный комплекс позволяет устранить эти недочеты.

Программа состоит из серверной и клиентской частей (рис. 1).

Серверная часть работает под управлением СУБД Oracle 10g (данная СУБД установлена на заводе) и представляет собой набор таблиц и хранимых подпрограмм, выполняющих хранение и обработку исходных данных соответственно. При проектировании комплекса было решено, что большая часть вычислений будет выполняться средствами СУБД Oracle, так как данный подход позволяет сократить расходы системных ресурсов, необходимых для передачи данных клиенту и их возврату в измененном виде.

Клиентская часть представлена набором функционально разделенных модулей.



Рис. 1. Структурная схема программного комплекса

В качестве среды разработки был использован Borland C++ Builder версии 6.0. Результатом сборки проекта является исполняемое приложение, предоставляющее пользовательский интерфейс для выполнения вычислительных задач, возложенных на программный комплекс [7].

Взаимодействие приложения и СУБД Oracle осуществляется при помощи программного интерфейса доступа к данным ActiveX Data Objects (сокращенно ADO), разработанного компанией Microsoft. Данный интерфейс является универсальным и позволяет взаимодействовать с различными СУБД. Необходимым условием его функционирования является наличие подходящего для данной СУБД драйвера поставщика данных. На данный момент существует два подобных драйвера, реализующих интерфейс ADO для СУБД Oracle. Первый из них – поставщик данных разработанный компанией Microsoft. Второй – драйвер, содержащийся в установочном пакете СУБД Oracle. В результате проведенных тестирований выяснилось, что драйвер, предоставленный компанией Microsoft, не может выполнить поставленную задачу, так как не способен установить

соединение с сервером БД и приводит к появлению сбоев в работе программы. В свою очередь драйвер компании Oracle обеспечивает стабильную работу.

Серверная часть представлена в виде набора служебных и пользовательских таблиц, хранимых подпрограмм, ограничений и последовательностей.

Данные объекты образуют целостную структуру, позволяющую, в полной мере, выполнять поставленные задачи. На рис. 2 отображена структурная схема базы данных, из которой видно, что работа комплекса построена на трех служебных таблицах: FILTER_METHODS, FILTER_METHOD_ARGUMENT и EXPERT_FORMS.

Таблица FILTER_METHODS используется для хранения информации о хранимых функциях реализующих алгоритмы отсева грубых промахов. К такой информации относятся: имя хранимой функции (поле OBJECT_NAME), краткое описание или название метода отсева (METHOD_SHORT), развернутое описание метода отсева (METHOD_DESC).

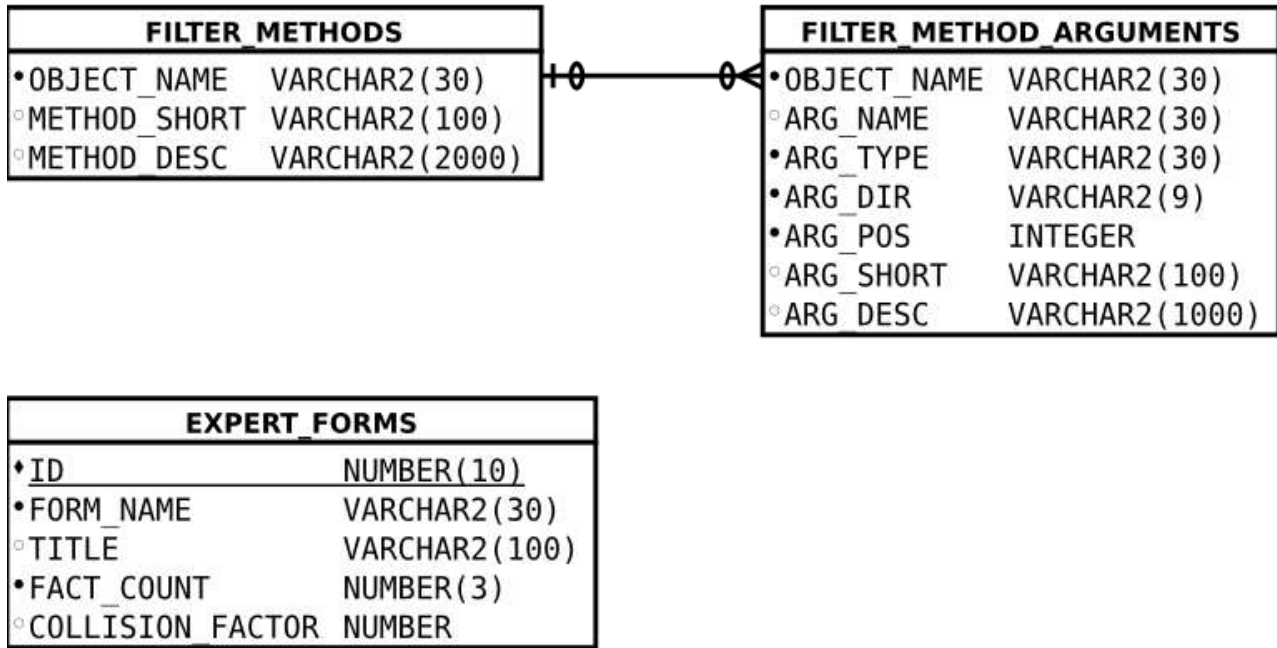


Рис. 2. Структурная схема БД

Поля описаний содержат информацию о функциях отсева на естественном для пользователя языке. Содержимое данной таблицы используется в клиентской программе для автоматического формирования списка доступных методов отсева грубых промахов [4, 7].

Для автоматического вызова подпрограммы необходимо хранить список и типы ее параметров. Для этого предназначена таблица FILTER_METHOD_ARGUMENTS, каждая запись которой хранит информацию об одном параметре подпрограммы. Связь параметров с хранимой подпрограммой устанавливается через поле OBJECT_NAME. В целом структура таблицы описана полями, представленными в табл. 2.

Чтобы избежать нарушения целостности, на поля ARG_DIR и ARG_POS таблицы FILTER_METHOD_ARGUMENTS налагаются соответствующие ограничения INOUT_CHECK_CONSTRAINT и ARGPOS_CHECK_CONSTRAINT соответственно. Возможные значения поля ARG_DIR ограничены набором ('IN', 'OUT', 'INOUT'), а значения поля ARG_POS ограничены только положительными числами.

Клиентская часть комплекса использует информацию из таблицы FILTER_METHOD_ARGUMENTS для автоматического создания полей формы пользовательского интерфейса, что позволяет пользователю несколькими простыми действиями выполнить вызов хранимой подпрограммы, не беспокоясь о порядке и типах ее параметров [8].

Таблица 2

Список полей таблицы FILTER_METHOD_ARGUMENTS

Имя столбца	Описание
OBJECT_NAME	имя хранимой подпрограммы
ARG_NAME	имя параметра подпрограммы
ARG_TYPE	тип параметра
ARG_DIR	направление передачи аргумента;
ARG_POS	позиция параметра в списке
ARG_SHORT	название параметра на естественном языке
ARG_DESC	описание параметра на естественном языке

ЗАКЛЮЧЕНИЕ

Данный программный продукт является средством реализации процедуры отсева грубых промахов. Это позволило исключить из исходной таблицы данных плавки, несоответствующие стандартам. В дальнейшем на основании этих данных будут построены математические модели с целью прогнозирования, внедрения новых видов стали, и создания основы для разработки системы автоматизированного управления качеством продукции. Полученные результаты позволяют оптимизировать процесс построения модели многофакторного процесса.

СПИСОК ЛИТЕРАТУРЫ

1. **Boswijk H. P.** Asimptotic Theory for Integrated Processes. Oxford: Oxford University Press, 1999. [H. P. Boswijk. *Asimptotic Theory for Integrated Processes*, (in English). Oxford: Oxford University Press, 1999.]
2. **Cameron A. C., Trivedi P. K.** Regression Analysis of Count Data. Cambridge: Cambridge University Press, 1998. [A.C. Cameron, P.K. Trivedi. *Regression Analysis of Count Data*, (in English). Cambridge: Cambridge University Press, 1998.]
3. **Dogerty K.** Introduction to Econometrics. The 3-th Ed. Oxford: Oxford University Press, 2006. [K. Dogerty *Introduction to Econometrics. The 3-th Ed*, (in English). Oxford: Oxford University Press, 2006.]
4. **Долгов Ю. А.** Статистическое моделирование: Учебник для вузов. 2-е изд., доп. Тирасполь: Изд-во Приднестр.ун-та, 2011. 349 с. [Y.A. Dolgov. *Statistical Modeling: Textbook for universities. The 2-nd Ed*, (in Russian). Tiraspol: Publishing house Pridnestr. University Press, 2011.]
5. **Cameron A. C., Windmeijer A. D.** An-squared measure of Goodness of Fit for Some Common Nonlinear Regression Models // *Journal of Econometrics*. 1997. № 77. P. 329–342. [A.C. Cameron, A.D. Windmeijer. An-squared measure of Goodness of Fit for Some Common Nonlinear Regression Models, (in English), in *Journal of Econometrics.*, no. 77, pp. 329-342, 1997.]
6. **Phillips P. C. B., Moon H. R.** Linear Regression Limit Theory for Nonstationary Panel Data // *Econometrica*. 1999. № 67. P. 1057–1111. [P.C.B. Phillips, H.R. Moon. Linear Regression Limit Theory for Nonstationary Panel Data, (in English), in *Journal of Econometrics.*, no. 67, pp. 1057-1111, 1999.]
7. **Долгов Ю. А., Козак Л. Я., Шестопад О. В.** Схема математического моделирования технологического процесса плавки стали // *Радіоелектронні і комп'ютерні системи*. 2010. № 7. С. 157–160. [Y.A. Dolgov, L.Y. Kozak, O.V. Shestopal, Driving mathematical modeling of technological process of melting steel, (in Russian), in *Electronic and computer systems*, no. 7, pp. 157-160.]

8. **Stochastic** check for control of electronic wares quality // *Trans. of 10-th International Symposium on Applied stochastic Models and Data Analysis*. June 12-15 2001. Univ. de Techn. de Compiègne, France. vol. 1. P. 387–390. [Stochastic check for control of electronic wares quality (in English), in *Trans. of 10-th International Symposium on Applied stochastic Models and Data Analysis*, June 12-15 2001, vol. 1, pp.387-390.]

ОБ АВТОРЕ

КОЗАК Людмила Ярославовна, доц. каф. информатики и программной инженерии. Дипл. информатик-экономист (ПГУ, 2001). Канд. техн. наук по моделированию (БГТУ, 2013). Иссл. в обл. мат. моделирования технологических процессов.

METADATA

Title: The selection procedure of significant factors in modeling process

Authors: L. Y. Kozak

Affiliation:

Pridnestrovian State University (PGU), Russia.

Email: Ludmilayaroslavovna@gmail.com.

Language: Russian.

Source: Vestnik UGATU (scientific journal of Ufa State Aviation Technical University), vol. 21, no. 1 (75), pp. 33-37, 2017. ISSN 2225-2789 (Online), ISSN 1992-6502 (Print).

Abstract: In this 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's 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's criterion, modeling process.

About author:

KOZAK, Ludmila Yaroslavovna, Ph.D. (Technology), associate professor of the department of computer science and software engineering, the Pridnestrovian State University. Dipl. Master of Technics & Technology (PGU, 2001). Cand. of Tech. Sci. (BGU, 2013).