УДК 662.754:519.25:519.233.5

ВЫЯВЛЕНИЕ ЛИНЕЙНО-НЕЗАВИСИМЫХ ФАКТОРОВ УПРАВЛЕНИЯ СТАДИЕЙ ПОДГОТОВКИ СЫРЬЯ В ПРОЦЕССЕ ИЗОМЕРИЗАЦИИ ЛЕГКИХ БЕНЗИНОВЫХ ФРАКЦИЙ

Запропоновано процедуру встановлення лінійно-незалежних керуючих факторів шляхом аналізу власних значень кореляційної матриці даних та величин означень підматриць кореляційної матриці. Підхід проілюстровано на прикладі пускових операцій стадії підготування сировини технологічного процесу ізомеризації легких бензинових фракцій

Ключові слова: ізомеризація, статистичний аналіз, лінійно-незалежні фактори

Предложена процедура установления линейно-независимых управляющих факторов путем анализа собственных значений корреляционной матрицы данных и величин определителей подматриц корреляционной матрицы. Подход проиллюстрирован на примере данных пусковых операций стадии подготовки сырья технологического процесса изомеризации легких бензиновых фракций

Ключевые слова: изомеризация, статистический анализ, линейно-независимые факторы

It is offered the procedure of an establishment of linearly-independent operating factors by the analysis of own values of a correlation matrix of data and sizes of determinants of submatrixes of a correlation matrix. The approach is illustrated on an example of the data of starting operations of a stage of preparation of raw material of technological process of light naphtha fractions isomerization

Keywords: isomerization, the statistical analysis, linearly-independent factors

С.А. Кондратов
Доктор химических наук, профессор, заведующий кафедрой*
Контактный тел.: 8 (066) 948-51-40
E-mail: kondratov@rune.lg.ua

А.В. Сидоренко
Соискатель*
Контактный тел.: 8 (095) 398-04-64
E-mail: sidorenko78a@mail.ru

*Кафедра высшей математики и компьютерных технологий
Институт химических технологий Восточноукраинского национального университета им. Даля
ул. Ленина, 31, г. Рубежное, Украина, 93009

1. Введение

В химической и нефтехимической промышленности одной из наиболее ответственных операций является пуск многотоннажных установок, работающих в непрерывном режиме. Пуск всегда проводят в «полуручном» режиме, когда оператор выводит управляющие параметры на регламентные значения, и только после выхода технологического процесса на стационарный режим управление передается компьютерным системам.

В процессе управления пуском оператору приходится контролировать значительное количество технологических параметров. Избыточное количество контролируемых и регулируемых параметров сложной системы с одной стороны – необходимое условие управляемости. С другой стороны – это и её недостаток, так как одновременное управление всеми параметрами оказывается зачастую физически невозможным. В результате этого система в период пуска нередко подолгу находится в неустойчивом режиме, не выходя на проектные показатели, что нежелательно с точки зрения экономических показателей производства и его безопасности.

Возможный путь эффективного управления – снижение размерности задачи управления за счет нахождения минимального числа параметров, достаточных для эффективного управления процессом. Цель настоящей работы – разработка подходов к выявлению таких параметров на примере работы промышленной колонны подготовки сырья в процессе изомеризации легких бензиновых фракций.

2. Объекты и методы

Объектом исследований в настоящей работе является пусковой период технологической стадии подготовки сырья процесса изомеризации легких бензиновых фракций. Метод решения поставленной задачи базируется на статистической обработке результатов пусковых операций, полученных на одном из российских нефтеперерабатывающих заводах. В основу метода наших исследований положены идеи метода главных компонент (МГК) [1-4], основанного на предположении о том, что несколько измеряемых переменных сильно коррелируют друг с другом. В МГК полагают, что корреляция обусловлена не тем, что переменные взаимно определяют друг друга, а тем, что они связаны с некоторой величиной (латентной переменной), которую нельзя непосредственно измерить [1-4].

С математической точки зрения сущность метода главных компонент состоит в нахождении собственных векторов матрицы корреляции измеряемых переменных. Собственные векторы являются нагрузками латентных переменных на соответствующие измеряемые переменные, а сами латентные переменные являются линейными комбинациями наблюдаемых переменных. Вклад фактора в описание экспериментальных данных определяется отношением соответствующего ему собственного значения к сумме всех собственных значений корреляционной матрицы [1-4].

Подобный подход с успехом применяют для свертывания экспериментальных данных и для непосредственного выделения факторов, интерпретация которых в целом приводит к более глубокому пониманию изучаемых явлений [1-4]. Вместе с тем, с позиции выявления реальных значимых факторов и сокращения размерности задач управления, введение новых «латентных» переменных наоборот, затрудняет интерпретацию и не позволяет определить, какие технологические параметры надо контролировать, а какие являются их линейными комбинациями. Исходя из этих соображений, мы использовали другой подход. Опираясь на исследование собственных значений корреляционной матрицы аналогично МГК, определили количество линейно-независимых факторов (n), как количества, вносящее суммарный вклад в общую дисперсию на уровне 90-95 %. Как известно [1-3], нулевые или близкие к нулю значения собственных значений корреляционной матрицы свидетельствуют о наличии линейной зависимости между факторами. Кроме того, существует связь между собственными значениями квадратной матрицы и ее определителем. Поэтому в качестве меры линейной независимости факторов можно использовать величину определителя корреляционной матрицы, построенной на этих факторах. В настоящей работе мы рассчитывали величины всех определителей порядка n, построенных, как подмножества общей корреляционной матрицы размерности m>n (m – общее количество фиксируемых переменных). Число таких комбинаций составляет C_m^n. Затем путем ранжирования определителей по величине выбирали лучшие сочетания факторов. Все расчеты проводили в среде Excel под управлением соответствующих программ, составленных на языке VBA.3. Выявление количества значимых факторов

В настоящей работе использованы результаты двух пусковых операций установки изомеризации легких бензиновых фракций на одном из российских НПЗ по блоку фракционирования и подготовки сырья. В обеих группах данных одновременно контролировалось 18 технологических параметров (факторов). Их перечень и кодировка приведены в таблице 1. количество наблюдений составило 540 (выборка 1) и 2760 (выборка 2).

На первой стадии мы исследовали однородность полученных выборок. В настоящее время не разработаны критерии однородности многомерных выборок. Поэтому мы при сравнении использовали следующие соображения: если две многомерных выборки будут однородны, то должны быть однородными подвыборки по каждому из факторов, извлеченные из этих выборок. Исходя из этого, мы исследовали однородность по каждой из таких подвыборок. Однородность их проверяли по критерию Крамера-Уэлча:

$$T = \frac{\sqrt{mn}(\bar{x} - \bar{y})}{\sqrt{ns_x^2 + ms_y^2}}, \qquad (1)$$

где $\bar{x}, \bar{y}$ - выборочные средние, s_x^2, s_y^2 - выборочные дисперсии выборок объемом, соответственно, m и n. Критерий (1) при больших объемах выборок не зависит от вида функции распределения в выборках и имеет асимптотически нормальное распределение [5]. Для выборок большого объема условие принятия решения об однородности: T < 1,96 для уровня значимости 0,95 [5]. Результаты, приведенные в табл.1, свидетельствуют о том, что по большинству из переменных расчетные значения критерия (1) существенно превышают критические, что позволяет говорить о неоднородности полученных выборок.

Для каждой из выборок рассчитали корреляционные матрицы, их собственные значения и вычислили вклад каждого из факторов в общую дисперсию как отношение соответствующего собственного значения к сумме всех собственных значений. Было установлено, что для обоих групп данных из 18 значений только 6 оказались значимыми: их суммарный вклад в общую дисперсию составляет 90-95 % (рис. 1).

Учитывая, что данные наблюдений представляют собой случайные величины с некоторым неизвестным распределением, возникает вопрос об устойчивости результатов. Для решения этого вопроса использовали бутстреп-метод [6], который позволяет восстанавливать статистические характеристики неизвестных распределений путем многократного тиражирования исходной выборки со случайным выбором элементов.

Использовали следующий алгоритм единичного испытания: из исходной выборки, содержащей P строк из m значений, случайным образом извлекали P строк (используя процедуру выборки с возвращением), получили новую матрицу данных. Для нее рассчитывали корреляционную матрицу, определяли собственные значения, вычисляли и запоминали вклад в дисперсию наибольших 6 собственных значений. Каждое из испытаний проводили 1000 раз, получили набор данных, по которым строили бутстреп-распределение вклада 6 собственных значений (рис. 2).

Таблица 1

Технологические параметры, фиксируемые по блоку фракционирования и подготовки сырья и значения критерия (1) по этим факторам для двух выборок

Номер кода	Технологический параметр (фактор)	T
1	Расход сырья в колонну	4,05
2	Расход острого орошения	30,98
3	Расход кубового остатка	4,53
4	Расход пара на подогрев колонны	0,76
5	Расход дистиллята	7,74
6	Уровень в кубе колонны	4,61
7	Уровень в рефлюксной ёмкости	13,61
8	Давление верха колонны	6,92
9	Давление под первой тарелкой	3,34
10	Давление в рефлюксной ёмкости	15,78
11	Температура сырья из резервуара	42,92
12	Температура ввода сырья в колонну	8,75
13	Температура сырья в реактор (кубовый продукт)	8,65
14	Температура верха колонны	16,96
15	Температура ПГС из рибойлера	7,37
16	Температура куба колонны	13,25
17	Температура после воздушных конденсаторов	15,81
18	Температура в рефлюксной ёмкости	19,76

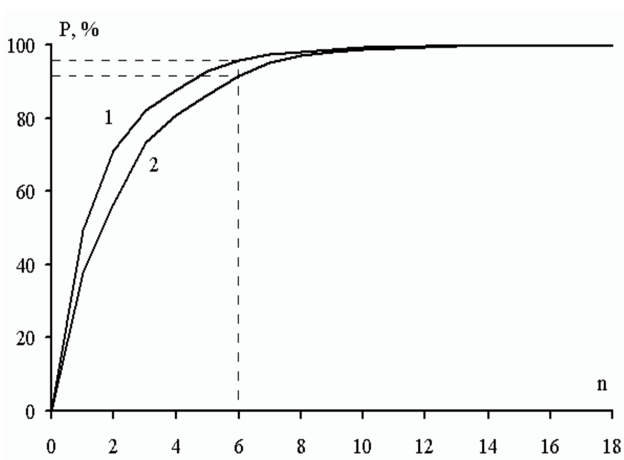

Рис. 1. Зависимость накопленного вклада в общую дисперсию от количества факторов для выборок 1 и 2

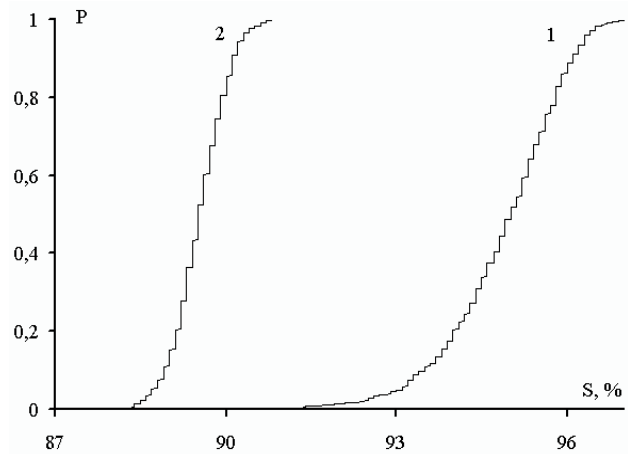

Рис. 2. Интегральная функция бутстреп-распределения вклада в общую дисперсию 6 наибольших собственных значений корреляционных матриц выборок 1 и 2

Как следует из рис. 2, для двух рассмотренных наборов данных вклад 6 собственных значений в общую дисперсию меняется в узких пределах. Это указывает, что, действительно, основной вклад в дисперсию вносят 6 факторов, и получаемые результаты статистически устойчивы. Таким образом, из 18 технологических факторов блока фракционирования и подготовки сырья процесса изомеризации легких бензиновых фракций только до 6 факторов являются линейно-независимыми, а остальные 12 следует рассматривать, как их линейные комбинации. Для управления процессом достаточно поддерживать на нужном уровне или изменять эти линейно-независимые факторы.

4. Выявление линейно-зависимых и линейно-независимых факторов

Для выявления линейно-зависимых и линейно-независимых факторов на основе исходной корреляционной матрицы порядка 18 построили все возможные подматрицы порядка от 2 до 6 и рассчитали их определители. В основу исследования положены следующие соображения. Корреляционная матрица положительно определена, поэтому ее собственные значения положительны [7]. На главной диагонали корреляционной матрицы располагаются единицы, и сумма собственных значений, равная следу матрицы, будет равна размерности матрицы. Следовательно, величины собственных значений ограничены снизу и сверху:

$$\min_{i\in[1;m]}(\lambda_i)\geq 0;\quad \max_{i\in[1;m]}(\lambda_i)\leq 1 \qquad (2)$$

Поскольку определитель матрицы равен произведению ее собственных значений,

$$\mathrm{Det}(A)=\prod_{i=1}^{m}\lambda_i\,, \qquad (3)$$

то определитель корреляционной матрицы также будет ограничен снизу и сверху. Для нахождения линейно-независимых факторов необходимо проанализировать значения определителей корреляционных матриц размерности n, построенных из элементов общей корреляционной матрицы размерности m>n. На основе исходной корреляционной матрицы порядка 18, мы построили все возможные корреляционные матрицы порядка 6 и рассчитали и упорядочили по убыванию их определители (общее количество C_{18}^{6} = 18564). По величине эти определители для n = 6 изменялись от ~ 10^{-6} до 0,84 (выборка 1) и до 0,37 (выборка 2). Мы условно разделили определители по величине на 2 группы: «хорошую» - от максимального значения до медианы, и «плохую» - от медианного значения до нуля.

Прежде всего, мы исследовали наиболее «плохие» комбинации, то есть, случаи достаточно сильной линейной зависимости между переменными. Для бинарных систем в группу самых плохих (наименьшие значения определителя) в обеих выборках попали следующие комбинации факторов:

Давление верха колонны (8) – давление под первой тарелкой (9),

Давление верха колонны (8) – температура верха колонны (14),

Давление верха колонны (8) – температура после воздушных конденсаторов (17),

Температура ввода сырья в колонну (12) – температура кубового продукта (13)

Температура верха колонны (14) – температура после воздушных конденсаторов (17)

Среди самих плохих комбинаций из 3 - 6 факторов представлены только комбинации, содержащие от 1 до 3 из отмеченных бинарных комбинаций. Эти результаты указывают, что линейно-зависимые факторы формируются преимущественно из бинарных факторов.

Далее мы проанализировали «хорошие» сочетания в комбинациях из 6 элементов.

Таблица 2

Частоты попадания факторов и их рейтинги

Ранг	Выборка 1		Выборка 2	
	Фактор	частота	Фактор	Частота
1	3(++)	180	3(++)	339
2	18(++)	89	18(++)	297
3	6(++)	51	11	268
4	10	50	6(++)	233
5	7(+)	46	13	185
6	5(+)	42	2	179
7	15	41	7(+)	179
8	1(+–)	38	12	177
9	4	35	5(+)	169
10	16	34	1(+–)	142
11	9	30	4	132
12	11	27	10	107
13	13	27	8	94
14	12	26	14	89
15	2	25	17	88
16	8	11	15	78
17	17	11	16	77
18	14	9	9	75

Объем «хорошей» группы для обеих выборок оказалось весьма небольшим, не более 2% от общего количества. Мы подсчитали частоту попадания каждого из факторов в «хорошую» группу и ранжировали эти частоты по убыванию. В таблице 2 показана первая десятка факторов (по рейтингу), наблюдаемая для обеих выборок. Как следует из табл.2, повторения высокого рейтинга наблюдаются для следующих пяти факторов:

- Расход кубового остатка (3)
- Температура рефлюксной емкости (18)
- Уровень в кубе колонны (6)
- Давление верха колонны (7)
- Расход дистиллята (5)

Из этих факторов первые 3 являются наиболее сильными, а факторы (6,7) проявляют влияние средней силы.

Шестой фактор устанавливается не столь надежно. Можно полагать, судя по данным табл. 2, что этот фактор - расход сырья в колонну (1). Он попадает в первую «десятку» в обеих выборках, но частота его появления невелика.

Отметим, что соотношение частот попадания факторов качественно хорошо корреспондируется с относительным вкладом компонентов в дисперсию (рис. 1).

Как известно [8], для установления корреляций двух факторов на фоне остальных используют коэффициенты частных корреляций, которые рассчитывают по формуле:

$$\rho_{i,j\bullet 1,2,..k,..} = \frac{\hat{R}_{ij}}{\sqrt{\hat{R}_{ii} \cdot \hat{R}_{jj}}} \qquad ,(4)$$

где $\hat{R}_{ij}$ - алгебраическое дополнение элемента (i,j) корреляционной матрицы (величина определителя, получаемого путем вычеркивания из корреляционной матрицы i-той строки и j-го столбца). Сравнивая выборочные обычный и частный коэффициент корреляции, можно делать выводы о том, насколько взаимозависимость между величинами Xi и Xj вызвана их собственной взаимосвязью [8].

Однако, вычисления частных коэффициентов корреляции показали, что в нашем случае этот метод оказался непригодным: все алгебраические дополнения оказались очень малыми по абсолютной величине (порядка 10-18÷10-16), что связано с наличием почти линейно-зависимых строк и столбцов.

Следствием этого является неустойчивость результатов и возможность появления вследствие этого ложных корреляций.

В отличие от этого, использованный нами последовательный анализ определителей корреляционной матрицы лишен этого недостатка и позволяет выявить наиболее значимые линейно-независимые факторы.

Таким образов, путем анализа определителей корреляционной матрицы данных удалось выявить 5÷6 наиболее значимых (линейно-независимых) факторов, изменяя величину которых, можно управлять остальными 12÷13 факторами процесса подготовки сырья, которые можно рассматривать, как их линейные комбинации.

Литература

1. С.А.Айвазян, В.М.Бухштабер, И.С.Енюков, Л.Д.Мешалкин; Под ред. С.А.Айвазяна. Прикладная статистика: Классификация и снижение размерности: Справ. изд. – М.: Финансы и статистика, 1989. – 607 с.
2. A. Gorban A., B. Kegl B., D. Wunsch D., A. Zinovyev A. Principal Manifolds for Data Visualisation and Dimension Reduction, LNCSE 58, Springer, Berlin – Heidelberg – New York, 2007 . – 332 p.
3. Эсбенсен К. Анализ многомерных данных. Избранные главы /пер. с англ. С.В.Кучерявского; Под ред. О.Е.Родионовой – Черноголовка: Изд-во ИПЧФ РАН, 2005. – 160 с.
4 Сычев С., Н. Сычев К.С., Ланин С.Н. Выявление характера межмолекулярных взаимодействий однозамещенных алифатических соединений методом главных компонент// Вестн. Моск. ун-та. Сер. 2. Химия, 2000. - Т. 41, № 1. – С. 28-31.
5 Орлов А.И. Прикладная статистика. Учебник. / А.И.Орлов.- М.: Издательство «Экзамен», 2004. – 656 с.
6 Эфрон Б. Нетрадиционные методы многомерного статистического анализа. – М.: Финансы и статистика, 1988 – 263 с.
7 Корн Т. Корн Г. Справочник по математике для научных работников и инженеров. М.: Наука, 1968. – 293 с.
8 М.Кендалл, Стьюарт А. Статистические выводы и связи. – М: Наука, 1973. – 900 с.