

УДК 519.242

А.А. Моисеев
**ИНТУИТИВНЫЙ МЕТОД РАНЖИРОВАНИЯ ФАКТОРОВ
ПО ЗНАЧИМОСТИ**

ГосНИИ химмотологии РФ

Эффективность применения дробных реплик полного факторного плана возрастает с ростом числа нефиктивных факторов. Она также зависит от удачного выбора системы смешивания линейных эффектов с эффектами взаимодействия факторов, а также рациональности схемы факторных экспериментов в случае значимости отдельных взаимодействий. При этом априорные сведения об указанных взаимодействиях зачастую оказываются не только желательными, но и необходимыми. В статье разработан несколько иной, интуитивный метод ранжирования факторов в линейно – регрессионной модели с использованием насыщенных планов и последующем отсеивании наименее нагруженных факторов. Его основным преимуществом, наряду с простотой реализации, является использование при анализе только нефиктивных факторов. Основным же недостатком, наряду с интуитивностью, является сравнительно низкая эффективность, обусловленная быстрым ростом числа необходимых экспериментов при возрастании числа факторов.

Ключевые слова: факторный эксперимент, линейно – регрессионная модель, насыщенный план, нагрузка, ранжирование, нефиктивные факторы, производящие функции.

Количество опытов в полном факторном эксперименте значительно превосходит число определяемых коэффициентов линейной модели, т.е. полный факторный эксперимент обладает большой избыточностью. В этих условиях возникает необходимость сократить число этих экспериментов за счет информации, которая малосущественна для построения указанной модели. При этом нужно стремиться к тому, чтобы матрица планирования не лишилась своих оптимальных свойств. Обычным подходом в решении этой задачи является использование дробных реплик полного факторного эксперимента. Именно с этой целью последние традиционно используются при переходе к линейно – регрессионной модели [1].

Использование для синтеза дробных реплик матриц полного факторного эксперимента позволяет использовать оптимальные свойства этих матриц – центрированность столбцов матрицы, соответствующих различным факторам, их ортонормированность и т.д. Наибольшей популярностью в двухуровневом планировании пользуются регулярные реплики полного факторного эксперимента, соответствующие насыщенным планам.

Причина этой популярности заключается в том, что использование ненасыщенных факторных планов приводит к снижению эффективности планирования, так как необходимое число факторных экспериментов

может значительно превышать число определяемых параметров. Поэтому так как правило используется подход, базирующийся на использовании насыщенных регулярных планов, включающих, в частности, фиктивные факторы. При подобном планировании часто используются дробные реплики полного факторного эксперимента. Соответствующие матрицы планирования при этом эквивалентны матрицам планирования полного факторного эксперимента, что позволяет использовать их оптимальные свойства – центрированность столбцов матрицы, соответствующих различным факторам, их ортонормированность, а также вытекающую из этого невырожденность матрицы планирования. А поскольку эта матрица является матрицей линейной системы относительно коэффициентов регрессии, отсюда вытекает однозначность расчета последних. Вследствие этого обычно исходят из сверхнасыщенных планов, число факторов в которых превышает число экспериментов, с последующим отбором значимых факторов с помощью, например, метода случайного баланса [2, 3].

Эффективность применения дробных реплик возрастает с ростом числа нефиктивных факторов. Она также зависит от удачного выбора системы смешивания линейных эффектов с эффектами взаимодействия факторов, а также рациональности схемы факторных экспериментов в случае значимости отдельных взаимодействий. При этом априорные сведения об указанных взаимодействиях зачастую оказываются не только желательными, но и необходимыми.

Здесь рассматривается несколько иной подход к решению задачи выделения существенных факторов, базирующийся на ранжировании последних по значимости с использованием насыщенных планов. Для этих планов число исследуемых факторов равно $k = 2^n - 1$, а число факторных экспериментов составляет $N = k + 1 = 2^n$. Насыщенному плану, в свою очередь, соответствует линейно – регрессионная модель вида $\hat{y} = b_0 + b_1 x_1 + \dots + b_k x_k$, коэффициенты которой определяются из системы уравнений:

$$\begin{cases} b_0 + b_1 x_{11} + \dots + b_k x_{k1} = y_1 \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ b_0 + b_1 x_{1k} + \dots + b_k x_{kk} = y_N \end{cases} \quad (1)$$

Матрица системы линейных уравнений относительно b_1, \dots, b_k является невырожденной матрицей насыщенного плана и вследствие этого указанная система разрешается однозначно. Значимость факторов

$x_1 \dots \dots \dots x_k$ в линейно – регрессионной модели будем определять относительными нагрузками $L_i = \frac{|b_i|}{\Delta}$, где $\Delta = y_{\max} - y_{\min}$ - диапазон измерений в серии, $i = 1 \dots k$.

Рассмотрим теперь ситуацию, когда число k исследуемых факторов удовлетворяет условию $2^n - 1 = l < k < 2^{n+1} - 1$, где n – число факторов полного факторного эксперимента, соответствующего регулярному насыщенному плану. Этой ситуации соответствует набор регулярных насыщенных планов для l факторов, выбранных из общего числа факторов k . Число этих планов и, следовательно, необходимое число серий экспериментов равно $P = C_k^l$. Воспользуемся для построения указанных планов методом производящих функций [4].

Этот метод применяется в комбинаторике и основывается на сходимости рядов. Он дает возможность просто описывать сложные комбинаторные последовательности находить для них явные формулы. Производящая функция представляет собой формальный степенной ряд, На практике, как правило, используются конечный ряд или ряд Тейлора для аналитической функции. В этом случае производящая функция может быть использована для построения и исследования последовательности коэффициентов ряда.

В случае комбинаторного анализа сочетаний используется тот факт, что при образовании сочетания осуществляется независимый отбор элементов. Правило умножения вероятностей позволяет использовать представление производящей функции в виде произведения линейных двучленов от независимой переменной. Коэффициенты разложения этого произведения соответствуют коэффициентам при различных сочетаниях.

Предположим в качестве примера, что исследуется влияние пяти факторов с помощью планов, охватывающих три фактора. Производящая функция в этом случае имеет вид $\Pi(t) = (1 + x_1 t) \dots (1 + x_5 t)$. После проведения перемножения находим для коэффициентов при степенях t :

$$\begin{aligned} &t^0 : 1 \\ &t^1 : x_1 + \dots + x_5 \\ &t^2 : x_1 x_2 + x_1 x_3 + x_2 x_3 + x_1 x_4 + x_2 x_4 + x_3 x_4 + x_1 x_5 + x_2 x_5 + x_3 x_5 + x_4 x_5 \\ &t^3 : x_1 x_2 x_3 + x_1 x_2 x_4 + x_1 x_3 x_4 + x_2 x_3 x_4 + x_1 x_2 x_5 + x_1 x_3 x_5 + x_1 x_4 x_5 + x_2 x_3 x_5 + x_2 x_4 x_5 + x_3 x_4 x_5 \\ &t^4 : x_1 x_2 x_3 x_4 + x_1 x_2 x_3 x_5 + x_1 x_2 x_4 x_5 + x_1 x_3 x_4 x_5 + x_2 x_3 x_4 x_5 \\ &t^5 : x_1 x_2 x_3 x_4 x_5 \end{aligned} \quad (2)$$

Эти коэффициенты интерпретируются как наборы насыщенных планов, соответствующих выборкам определенного числа факторов из пяти. В частности, выборкам, включающим три фактора из пяти, соответствует коэффициент при t^3 , т.е. набор насыщенных планов вида:

$x_1x_2x_3 + x_1x_2x_4 + x_1x_3x_4 + x_2x_3x_4 + x_1x_2x_5 + x_1x_3x_5 + x_1x_4x_5 + x_2x_3x_5 + x_2x_4x_5$
соответствующих репликам 2^{5-3} .

Для каждого из выбранных по такому принципу планов проводится серия троекратно повторяющихся экспериментов. Коэффициенты регрессии для каждого из этих планов рассчитываются однозначно в соответствии с (1), где $y_1 \dots y_N$ интерпретируются как медианы повторяющихся измерений. В свою очередь, по найденным коэффициентам рассчитываются нагрузки для факторов, охватываемых данной серией. Поскольку каждый фактор охватывается $P - 1$ серией экспериментов, для него получается то же число относительных нагрузок $L_{i1} \dots L_{iP-1}$, рассчитываемых с учетом изменения диапазона измерений от серии к серии. Характеристиками значимости факторов считаются максимальные нагрузки из полученных $L_{imax} = \max(L_{i1} \dots L_{iP})$, $i = 1 \dots k$, по которым и осуществляется ранжирование.

Относительные максимальные нагрузки $L_{1max} \dots L_{kmax}$ можно также использовать для рационального отсеивания незначимых коэффициентов. Предположим без ограничения общности, что ряд $L_{1max} \dots L_{kmax}$ представляет собой результат ранжирования нагрузок по убыванию, и образуем эквивалентный ряд $1 = \frac{L_{1max}}{L} \dots \frac{L_{kmax}}{L}$, где $L = L_{1max}$. Выставим порог по значимости, например, $h = 0.05$, и все факторы, удовлетворяющие условию $\frac{L_{imax}}{L} < h$, будем считать незначимыми. Этот метод отсеивания более прост в сравнении с традиционным статистическим отсеиванием [5], хотя и может считаться чисто интуитивным.

Резюмируем поведенное рассмотрение. В статье разработан интуитивный метод ранжирования факторов в линейно – регрессионной модели с использованием насыщенных планов и последующем отсеивании наименее нагруженных факторов. Его основным преимуществом, наряду с простотой реализации, является использование при анализе только нефиктивных факторов. Основным же недостатком, наряду с интуитивностью, является сравнительно низкая эффективность, обусловленная быстрым ростом числа необходимых экспериментов при возрастании числа факторов.

ЛИТЕРАТУРА

1. Адлер Ю.П., Маркова Е.В., Грановский Ю.В. Планирование эксперимента при поиске оптимальных условий, М., Наука, 1976, 279 с.
2. Хартман К., Лецкий Э., Шефер В. Планирование эксперимента в исследовании технологических процессов, М., Мир, 1977, 552 с.
3. Налимов В.В., Чернова Н.А. Статистические методы планирования экстремальных экспериментов, М., Наука, 1965, 340 с.
4. Риордан Дж. Введение в комбинаторный анализ, М., ИЛ, 1963, 288 с.
5. Ахназарова С.Л., Кафаров В.В. Методы оптимизации экспериментов в химической технологии, М., Высшая школа, 1985, 327 с.

A.A. Moiseev

INTUITIVE METHOD OF FACTORS SIGNIFICANCE RANKING

State Research Institute of Chimmotology

Effectiveness of fractional replicas application increases at non – fiction factors multiplication. It depends also on linear and non. – linear effects and factor experiments performance scheme in the presence of factors interactions. In this situation a priori information is desirable and even necessary. Developed here intuitive method of factors ranking in linear regression model based on regular plans and posterior rejection of least loaded factors. Its main advantage along with realization simplicity is only non fictive factors application. Base disadvantage along with intuitivism is swift effectiveness decrease at factors number rise.

Keywords: factor experiment, linear regression model, regular plan, relative load, ranking, non fictive factors, generating functions

REFERENCES

1. Adler Y. ea Planirovanie experimenta pri poiske optimalnykh uslovij (Experiment planning at optimal conditions search), М., “Science”, 1976, 279 p.
2. Hartmann K. ea Statistische Versuchplanung und –auswertung in der Stoffwirtschaft, Leipzig, Verlag fur Grundstoffindustrie, 1974, 552 p.
3. Nalimov V. ea Statisticheskije metody planirovaniya ekstremalnykh experimentov (Statistical methods of extreme experiments planning), М., “Science”, 1965, 3 40 p.
4. Riordan J., An introduction to combinatorial analysis, NY, Wiley, 1958, 288 p.
5. Achnazarova S. ea Metody optimizatsii experimentov v chimicheskoy technologii (Experiments optimization methods in chemical technology), М., “High school”, 1985, 327 p