

АЛГОРИТМ ПОДГОТОВКИ ТАБЛИЦЫ МНОГОМЕРНЫХ ПАССИВНЫХ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ ДЛЯ ПОЛУЧЕНИЯ АДЕКВАТНЫХ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ

Долгов Юрий Александрович

академик РАЕН,

доктор технических наук,

профессор Приднестровского государственного университета им. Т.Г. Шевченко,

г. Тирасполь

Балашова Юлия Владимировна

старший преподаватель

кафедры интегрированных компьютерных технологий и систем

Приднестровского государственного университета им. Т.Г. Шевченко,

г. Тирасполь

Тереценко Елена Владимировна

старший преподаватель

кафедры программного обеспечения вычислительной техники и автоматизированных систем

Приднестровского государственного университета им. Т.Г. Шевченко,

г. Тирасполь

Аннотация. Рассматривается алгоритм сокращения размерности факторного пространства экспериментальных данных без существенной потери полезной информации с целью получения адекватной многомерной математической модели и определения её информационной ёмкости.

Abstract. It is shown the algorithm of factor space dimension reducing of experimental data without any information loss to generation of adequate multivariate mathematical model and its information capacity.

Ключевые слова: размерность факторного пространства, адекватная математическая модель, информационная ёмкость.

Key words: factor space dimension, adequate mathematical model, information capacity.

Пассивный эксперимент представляет собой наблюдение и фиксацию числовых значений, параметров (факторов) и целевых функций объекта исследования при его естественном ходе без искусственного вмешательства экспериментатора [3]. Результатом является таблица, N строк представляет собой конкретные числовые значения целевой функции при конкретном наборе факторов, а M столбцов – числовые (бальные, дихотомические) значения каждого фактора в отдельности, то есть M выборок объёмом N каждая. Из этой таблицы предстоит извлечь скрытую информацию в виде многомерной регрессионной модели, что представляет достаточно сложную задачу.

Во-первых, вопреки классической теории планирования эксперимента [1], никакого искусственного изменения (управления) факторов нет, а имеет место беспорядочное, иногда довольно маленькое, варьирование, и чтобы отличить воздействие конкретного фактора на целевую функцию от шума эксперимента требуется достаточно длинная таблица.

Во-вторых, в первоначальном списке могут быть сильно коррелированные факторы, что, как известно [3,1], не способствует нахождению математической модели объекта исследования. Необходимо из каждой пары (или группы) сильно коррелированных факторов выбрать для дальнейших вычислений только один, что существенно сокращает размерность факторного пространства, а это, в свою очередь, существенно сокращает трудоёмкость нахождения модели. Наиболее простым способом осуществить эту операцию является метод корреляционных плеяд [4].

В-третьих, в первоначальном списке также могут оказаться факторы, не влияющие на целевую функцию, но, к сожалению, это можно выяснить либо непосредственно в ходе моделирования, либо дополнительными экспертными исследованиями.

Таким образом, алгоритм подготовки исходных пассивных экспериментальных данных к адекватному моделированию может быть представлен в виде нижеприведённого рисунка.

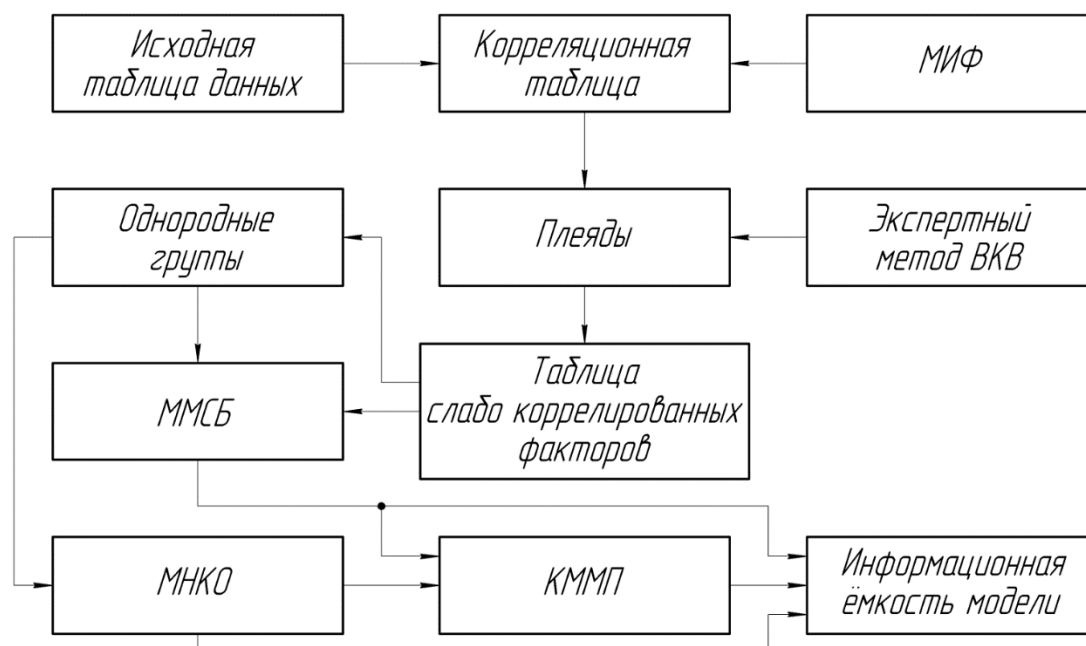


Рисунок 1. Алгоритм подготовки исходных данных

Из опыта нашей работы ясно, что корреляционную таблицу лучше всего вычислять не с помощью коэффициента линейной корреляции, пригодных только при числовых представлениях данных, а с помощью модифицированного индекса Фехнера (МИФ) [3,2], который практически совпадает по величине с коэффициентом корреляции в диапазоне значений $[0,2-1,0]$ но в отличие от него носит универсальный характер, то есть, пригоден для использования не только числовых данных, но и балльных, и дихотомических, чего коэффициент корреляции не может. Кроме того, МИФ нечувствителен к наличию в таблице данных грубых промахов как линейного, так и парного характера до 6-7% от объёма выборки, тем самым отпадает трудоёмкая процедура их поиска и замены.

Выделение из каждой плеяды главного фактора можно производить любым экспертным методом, однако, на наш взгляд, наиболее удобным для этого является метод весовых коэффициентов важности (ВКВ) [3], который не только вдвое точнее любого другого экспертного метода, но и обладает рядом только ему присущих преимуществ: удобство работы эксперта, определение коэффициента компетентности (непротиворечивости ответов) эксперта и, следовательно, возможности отсева заведомо ложных экспертиз, а также выявление отдельных факторов, по которым мнения экспертов разошлись, и доказательство непротиворечивости конечной ранжировки законам природы.

Что касается непосредственно адекватного моделирования, то укажем на три авторских метода: модифицированный метод случайного баланса (МНСБ) [3]; метод наименьших квадратов с предварительной ортогонализацией факторов (МНКО) [3]; комбинированный метод моделирования по пассивным данным (КММП) [3]. Кроме того, все эти три метода могут быть дополнительно оценены по качеству с помощью коэффициента информационной ёмкости [3], то есть, по количеству информации, которую модели выбрали из таблицы данных.

МНСБ – самый скоростной метод, менее трудоёмок, его коэффициенты одновременно являются весами факторов, однако он требует длинной таблицы данных, не даёт квадратичных членов, факторы представлены в относительном виде, что затрудняет использование модели, а информационная ёмкость составляет 15-30%.

МНКО – обладает повышенной трудоёмкостью, которая удваивается с каждым новым фактором; может использоваться при малом числе строк таблицы данных ($N \geq s$) при условии, что число столбцов хотя бы на единицу меньше; может работать при повышенной корреляции между факторами; может давать квадратичные, кубические и т.д. члены; модель представлена в естественных единицах измерения факторов; информационная ёмкость модели составляет 30-65%.

КММП является комбинацией МНСБ и МНКО; информационная ёмкость модели составляет 50-85%.

Выводы

Использование предложенного алгоритма гарантирует получения высококачественной математической модели, пригодной как для управления, например, технологическим процессом; так и для совершенствования объекта исследования.

Список литературы

1. Адлер Ю.П., Маркова Е.В., Грановский Ю.В. Планирование эксперимента при поиске оптимальных условий. -2-е изд., перераб., и доп.- М.: Наука, 1976 – 279 с.
2. Долгов Ю.А., Долгов А.Ю., Терещенко Е.В. Модифицированный индекс Фехнера как мера тесноты линейной связи (Вестник Приднестр. Ун-а. – 2015 – N3 – С. 92 – 98).
3. Долгов Ю.А. Статистическое моделирование: Учебник для вузов.-2 изд. доп.-Тирасполь: Полиграфист, 2011-352с.
4. Дружинин Г.В. Методы оценки и прогнозирования качества - М.: Радио и связь 1982-160с.