

Интеллектуальный анализ многофакторных зависимостей на основе геометрических образов

Аннотация Использование геометрических образов процессов в системах поддержки принятия управленческих решений для снижения рисков и повышения эффективности предоставляет широкие возможности. Наличие эффектов взаимодействия в моделях, представляемых многофакторными зависимостями, позволяет описывать сложные технико-экономические системы. Для исследования процессов во временной зависимости использованы методы динамики сплошных сред, в которых поведение среды описывается с помощью уравнений Навье-Стокса, имеющих чувствительность при возникновении турбулентного режима - сценария перехода к хаосу. Исследование динамики процесса с использованием геометрических образов и имитационного моделирования позволяет определить уровень устойчивости процесса и принять своевременное управляющее решение.

Ключевые слова: модель, динамика, устойчивость, геометрический образ, спектр.

Введение

Считается, что отличие информации от данных состоит в том, что она появляется в результате обработки данных при решении конкретных задач. Суть обработки, в основном, заключается в отсеке данных, не имеющих отношения к предмету рассмотрения, в исключении аномалий, классификации, группировки по сходным признакам, выявлении закономерностей и попытке получения на их основе новой информации, позволяющей принимать решения по изменению, корректировке существующих процессов. Процесс поддержки принятия решений как раз и включает этапы оперативной аналитической обработки и интеллектуального анализа данных. Проблема интеллектуального анализа многомерных данных, одним из измерений которых является, например, время, использующего мощные методы, например, байесовские сети, метод k-ближайшего соседа, в частности, неприменимы для задач многомерного интеллектуального анализа, так как основаны на определении сходства детализированных примеров и не способны работать с агрегированными данными [1].

Эффективным методом анализа и исследования процессов с помощью моделей является использование геометрических образов на основе имеющихся данных [2]. Использование когнитивной машинной графики в системах поддержки принятия управленческих решений для снижения рисков и повышения эффективности управления предоставляет широкие возможности. Суть визуализации, в частности, может состоять в проецировании многомерных данных на двумерную плоскость и дальнейшем исследовании процесса во временной зависимости. При этом исходно, взаимосвязь между количественными данными может выявляться с помощью простых статистических моделей. Как правило, выделяется определяемый, оказывающий влияние на эффективное функционирование процесса или системы, показатель и факторы от которых этот показатель зависит.

Для случая однофакторной или двухфакторной зависимости достаточно просто изобразить геометрический образ модели, позволяющий исследовать взаимосвязи и выявлять закономерности, например, линия как образ в линейной модели и парабола в полиномиальной модели второго порядка. Двухфакторная модель

может быть геометрически представлена в виде поверхности, в которой факторы являются осями на плоскости, а определяемый показатель - амплитудой третьей координаты. Проекция поверхности показателя на плоскость факторов изображается в виде линий равного выхода (изолиний), во всех точках которых показатель имеет постоянное значение независимо от координат. Одновременное изучение корреляции нескольких факторов проводится на основе использования методов множественной корреляции. Линейное уравнение множественной зависимости может быть записано, например, так

$$Y = b_0 + b_1x_1 + b_2x_2 + \dots + b_mx_m,$$

или так

$$Y = a_1x_1 + a_2x_2 + \dots + a_mx_m.$$

По параметрам полученного уравнения можно оценить долю каждого из факторов в изменении уровня результативного показателя, либо путем прямой оценки по величине коэффициентов регрессии при каждом из факторов, либо по коэффициентам эластичности. Коэффициенты уравнения множественной зависимости показывают абсолютный уровень влияния факторов на величину определяемого показателя и характеризуют степень влияния каждого фактора на него при фиксированном уровне других факторов, входящих в модель. Исследовать в целом процесс, описываемый подобным уравнением, используемым зачастую в качестве производственной функции, невозможно. Именно наличие эффектов взаимодействия в подобных моделях позволяет описывать реальные сложные технико-экономические системы и принимать решения, учитывающие диалектические противоречия в таких системах. Это обычно недоступно исследователю, анализирующему поведение систем с помощью малофакторных зависимостей [3].

1. Имитационное моделирование для анализа многофакторных зависимостей

В сложных технико-экономических системах обычно учитывается множество факторов и один из способов преодолеть ограничения для построения геометрического образа рассматриваемого процесса - это использование методов

имитационного моделирования. Данные методы имеют ряд преимуществ: в имитационной модели фактически допустимы любые изменения, каждый фактор может варьироваться по усмотрению исследователя; ошибки, допущенные в модели или в исходных данных, не имеют тех катастрофических последствий, какие иногда могут возникнуть в физическом опыте [4]. Предполагая, что процесс, описываемый какой-либо функцией, имеющей много факторов, оказывающих влияние на основной показатель и непосредственно связанных друг с другом, развивается во времени, можно воспользоваться имитацией для его исследования. Одна из разновидностей обработки данных связана с преобразованием их в частотное представление или спектр. Спектр может получиться в результате разложения исходной функции, зависящей от времени, в базис, например, спектра Фурье, получаемый на основе синуса. Каждая составляющая спектра с определенной частотой и амплитудой, полученная в результате подобного разложения участвует в формировании результата. При использовании уравнения множественной корреляции для разложения по спектру, коэффициенты при составляющих спектра - это коэффициенты регрессии, показывающие степень, с которой соответствующие функции коррелируют с данными. Применение составляющих спектра в двух координатах позволит получить изображение в виде двумерной плоскости с определяемым показателем, отображаемым амплитудой третьей координаты.

2. Пример отображения геометрического образа многофакторной зависимости

Например, данные процесса, полученные в ходе экспериментов (Табл. 1), имеют разный по величине диапазон изменения и разные единицы измерения. Для устранения размерности и приведения данных к виду, который в дальнейшем продемонстрирует степень влияния каждого из факторов на показатель, выбранный в качестве результирующего, воспользуемся простым методом обезразмеривания по максимальному значению каждого из показателей. В результате корреляционно-регрессионного анализа получено уравнение регрессии без константы для обезразмеренных данных, в кото-

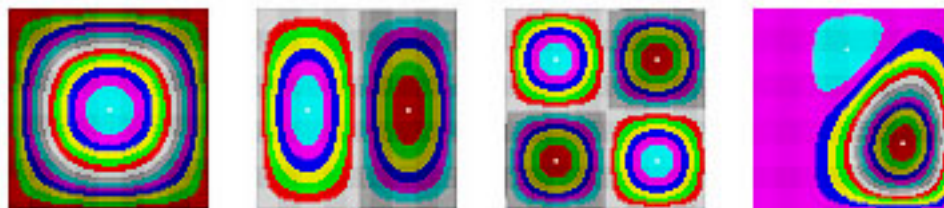


Рис. 1. Моды спектра базиса i, j : 1,1, 1,2, 2,2 и суммарный геометрический образ топологии исследуемого процесса

рых последний столбец выбран за определяемый показатель:

$$Y = 0,013 \cdot x_1 - 1,45 \cdot x_2 - 0,65 \cdot x_3 + 0,95 \cdot x_4 + 1,38 \cdot x_5 + 0,55 \cdot x_6$$

Табл. 1. Исходные данные исследуемого процесса

x_1	x_2	x_3	x_4	x_5	x_6	Y
43	3,93	19852	13	3320	241	10
70	1,06	27809	16	1041	179	15
32	5,81	20246	24	5060	226	20
66	4,19	11504	18	4195	126	40
94	4,86	13997	12	4489	201	12
77	3,62	13489	17	3349	249	24
88	3,64	17153	24	3046	159	14
54	1,99	15815	15	1650	187	26
54	4,5	21829	25	3528	125	22
44	2,18	12508	14	2018	310	34
56	1,77	16851	24	1776	210	35
45	3,27	18640	13	2837	117	16
80	3,24	31103	31	2850	197	18
30	0,63	18650	15	1430	123	13

Достоверность полученной зависимости достаточно велика, коэффициент детерминации порядка 90%. Графики зависимости показателя Y и значений, полученных из уравнения регрессии для данных, прошедших процедуру обезразмеривания, демонстрировали, что уравнения достаточно точно описывают исследуемый процесс. Не углубляясь в детали о степени доверия в соответствии с критериями Фишера или Стьюдента к полученным коэффициентам влияния факторов, используем полученные данные для дальнейшего анализа. Проецируем уравнение регрессии с обезразмеренными данными на ортогональный тригонометрический базис, удовлетворяющий условию нулевых значений в пределах ограниченной единичными размерами плоскости:

$$\psi_{ij} = \cos\left\{\frac{i\pi x}{L} - \left[l + (-1)^i\right]\frac{\pi}{4}\right\} \cdot \cos\left\{\frac{j\pi y}{C} - \left[l + (-1)^j\right]\frac{\pi}{4}\right\},$$

где i и j - моды базиса, $C=1$ и $L=1$ - относительные размеры плоскости: ширина и длина.

Обезразмеренные коэффициенты уравнения регрессии со своими знаками закрепляются за модами базиса. Фактор с коэффициентом наибольшим по величине, что говорит о максимальном его влиянии на процесс, закрепляется за основной модой спектра затем остальные - пропорционально убыванию степени влияния. В результате получен геометрический образ топологии процесса (Рис. 1), который подвергается последующему анализу.

3. Исследование многофакторной зависимости в динамике

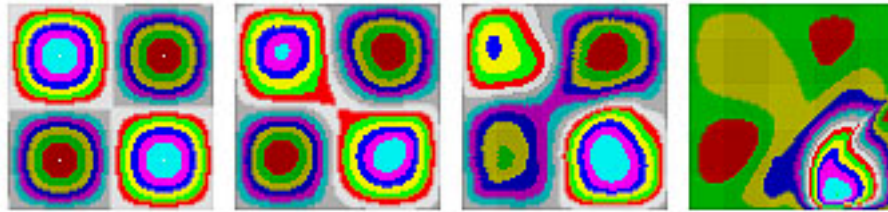
Для исследования процесса, отображенного в виде геометрического образа, во временной зависимости использованы методы динамики сплошных сред, в которых поведение среды описывалось с помощью уравнений Навье-Стокса, имеющих чувствительность при возникновении турбулентного режима - сценария перехода к хаосу. Уравнения двумерной динамики записывались [5] в переменных функциях тока, завихренность - в форме обобщенного уравнения Гельмгольца с добавлением ротора поля вынуждающих или мотивирующих динамику процесса сил:

$$\frac{\partial \omega}{\partial t} + \frac{\partial \psi}{\partial y} \frac{\partial \omega}{\partial x} - \frac{\partial \psi}{\partial x} \frac{\partial \omega}{\partial y} = k \left(\frac{\partial^2 \omega}{\partial y^2} + \frac{\partial^2 \omega}{\partial x^2} \right) + \text{rot} F_{xy}$$

$$\left(\frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial x^2} \right) = -\omega,$$



Рис.2. Геометрические образы исследуемого процесса в динамике

Рис.3. Автоколебательный режим процесса
(исходный образ, автоколебания, хаос)

где ψ - функция тока, ω - завихренность, t - относительное время, F_{xy} - поле вынуждающих динамику процесса сил, rot - векторная операция ротора, x, y - линейные безразмерные координаты среды, $k \sim 1/Re$ - коэффициент пропорциональный обратному значению числа Рейнольдса, определяемому по характерному размеру среды и скорости.

Для численного решения использовалась явная двухслойная схема с центральными разностями по пространственным переменным. Решение рассматривалось на равномерной сетке в пределах плоскости. Анализировались фазовые портреты рассматриваемого в динамике процесса, включающие геометрические образы и энергетический параметр - средняя по слою скорость взаимодействия. Как показывают изображения (Рис.2), процесс претерпевает со временем изменение, но в итоге устанавливается в новое устойчивое стационарное состояние (относительная скорость: 5,54, 4,1, 1,0, 0,06, 0,01), следовательно, данный процесс стабилен и не подвержен кризисам.

Численные исследования динамики процессов с разной структурой геометрических образов показали, что при наличии структуры взаимодействия составляющих процесса в виде четырех контуров, наблюдаются автоколебания в виде попарного замыкания противоположных из них (Рис.3), при увеличении интенсивности и при определенных параметрах, приводящие

либо к устойчивому стационарному состоянию либо к хаосу.

Два контура взаимодействий составляющих процесса также приводили его к режиму автоколебаний. Наличие фактора, превышающего остальные по величине, что отображается в виде одиночного контура, способствует повышению стабильности. Многоконтурная структура образа также повышает стабильность процесса по причине хорошей диссипации энергии в системе. Данные численных исследований подтверждаются результатами физических экспериментов [6, 7].

Заключение

Сложные технико-экономические системы обычно характерны множеством факторов, оказывающих влияние на процессы. Состояние процесса может быть стабильным или нестабильным. Кризис, как показывает опыт, не является редким событием, любая система, любой процесс по мере развития приходит к состоянию насыщения, что неизбежно выливается в какое-либо преобразование. Параметры системы могут быть изменены при своевременном выявлении предкризисной ситуации посредством управляющих, корректирующих воздействий. Исследование динамики процесса с использованием геометрических образов и имитационного моделирования позволяет оп-

ределить уровень устойчивости процесса и принять своевременное управляющее решение.

Литература

1. Parsaye K. OLAP and Data Mining: Bridging the Gap // Database Programming and Design. - 1997. - № 2. - P. 30-37.
2. Горохов В.Л., Лукьянец А.А., Чернов А.Г. Современные методы когнитивной визуализации многомерных данных. - Томск: Некоммерческий фонд развития региональной энергетики, 2007. - 216 с.
3. Вознесенский В.А. Статистические методы планирования эксперимента в технико-экономических исследованиях. - 2-е изд., перераб. и доп.—М.: Финансы и статистика, 1981. - 263 с.
4. Яковлев Е.И. Машинная имитация. М.: Наука, 1975-156 с.
5. Шкадов В.Я., Запрянов З.Д. Течения вязкой жидкости. М.:Изд. Моск. университета., 1984. - 200с.
6. Альмухаметов В.Ф. Моделирование и исследование двумерных автоколебательных процессов. Цветные металлы. 2011, №6.- С.75-77.
7. Альмухаметов В.Ф., Кирко И.М., Хрипченко С.Ю. Физическое моделирование неустойчивого состояния границы раздела электролит-металл в алюминиевых электролизерах. Доклады АН СССР. 1988, т.302, №4, с. 845-847.

Альмухаметов Вакиль Федорович. Профессор кафедры информационных технологий и автоматизированных систем Пермского национального исследовательского политехнического университета. Автор более 60 печатных работ и трех монографий. Область научных интересов: моделирование процессов, интеллектуальный анализ, теория хаоса.
Email: almval51@mail.ru.