Федеральное агентство по образованию

Государственное образовательное учреждение высшего
профессионального образования

Воронежский государственный архитектурно-строительный университет

П.А. ГОЛОВИНСКИЙ, В.Я. МИЩЕНКО, Е.М. МИХАЙЛОВ

МАТЕМАТИЧЕСКИЕ МЕТОДЫ ПРИНЯТИЯ УПРАВЛЕНЧЕСКИХ РЕШЕНИЙ В СТРОИТЕЛЬСТВЕ

Учебное пособие

Рекомендовано редакционно-издательским советом Воронежского государственного архитектурно-строительного университета в качестве учебного пособия для студентов старших курсов и магистрантов, обучающихся по направлению 270100 «Строительство»

Воронеж 2008

УДК 69:51(075)
ББК 38в631.0я7
Г611

Головинский, П.А.
Математические методы принятия управленческих решений
Г611 **в строительстве** [Текст]**:** учеб. пособие / П.А. Головинский, В.Я. Мищенко, Е.М. Михайлов; Воронеж. гос. арх.- строит. ун-т. – Воронеж, 2008. – 92 с.

ISBN 978-5-89040-201-1

Излагаются разделы современной математики, которые используются при принятии решений в строительстве: описаны представление и обработка данных, методы оптимизации, теория очередей, сети и графы.

Пособие предназначено для студентов старших курсов и магистрантов, специализирующихся в области управления строительством.

Ил. 40. Табл. 8. Библиогр.: 10 назв.

УДК 69:51(075)
ББК 38в631.0я7

Рецензенты: кафедра «Экспертиза и управление недвижимостью» Курского государственного технического университета; П.Г. Грабовый, д. э н., проф., зав.кафедрой «Организация строительства» Московского государственного строительного университета

ISBN 978-5-89040-201-1

Введение

Настоящее пособие посвящено изложению только теоретических основ принятия решений в управлении строительством, но оно дает необходимый минимум знаний для практической работы.

Принятие решений при управлении в строительстве опирается на совокупность методов и технологий, среди которых важное место занимают математические модели. Их эффективное применение требует как определенных теоретических знаний, так и практических навыков решения соответствующих задач с использованием компьютерной техники. Учебное пособие рассчитано на обучение магистрантов в области строительства основам математических методов принятия решений в рамках короткого курса, читаемого в течение одного семестра. Объем курса, его характер и аудитория, на которую он ориентирован, определили объем, структуру и способ изложения материала. С точки зрения содержания, основное внимание уделено способам представления и первичной обработки данных, а также и базовым моделям принятия решений. Основной упор в изложении делался на пояснение идей и принципов, положенных в основу той или иной модели, а не подробным математическим доказательствам. Всюду, где это возможно, мы старались избегать излишне громоздких формул, отдавая предпочтение содержательному обсуждению моделей.

Опыт преподавания в течение ряда лет в Воронежском государственном архитектурно-строительном университете показал, что наилучшие результаты изучения курса достигаются при параллельном с изложением теории проведении лабораторно-практических занятий с применением компьютеров на основе использования прикладного математического пакета MatLab.

Глава 1. Методы обработки, оценки и представления данных

1.1. Моделирование

Процесс управления и принятия решений включает в себя большое число детерминированных и случайных параметров и связей. Полное воспроизведение этих связей в какой–либо модели невозможно. Всякая модель представляет собой отображение реального объекта на иную систему связей, которая сохраняет наиболее важные характеристики изучаемой системы.

В зависимости от решаемой задачи более важными становятся те или иные характеристики и взаимосвязи в системе. Соответственно данная система может описываться целым набором различных моделей. Иначе говоря, выбор модели не является единственным. В то же время модель, построенная для описания какой-то конкретной ситуации, может иметь более общий характер и применяться для описания иных ситуаций и систем. Фактически развивается ограниченное число типов моделей, с помощью которых исследователи пытаются перекрыть все многообразие реальных ситуаций.

Математическое моделирование заключается в построении абстрактно-математического образа системы и протекающих в ней процессов. Чаще всего математические модели управления строятся для следующих целей:

- имитации действия объекта или протекания процесса при различных параметрах с целью получения представления об изменении при этом тех или иных характеристик системы;

- нахождения при помощи модели оптимальных параметров и режимов процесса;

- прогнозирования развития системы во времени с учетом детерминированных и случайных параметров.

Все эти задачи математического моделирования являются общими независимо от природы моделируемых объектов. Моделирование технико-экономических систем обладает своими особенностями, что и вызвало построение целого набора специальных моделей.

Для построения математической модели необходимо представить параметры и связи между объектами в математической форме. Для работы с количественными показателями их следует, прежде всего, адекватно обрабатывать, то есть проводить статистический анализ данных. Статистический анализ данных с помощью современных компьютерных средств позволяет не только представить их в универсальной форме, но и выявить наличие или отсутствие взаимозависимости между ними.

1.2. Понятие о математической статистике

Методы математической статистики позволяют оценивать параметры системы и выделять в них случайную и закономерную составляющие. Хотя результаты наблюдений, зависящих от случайных факторов, нельзя достоверно предсказать, можно оценить шансы на их появление. Количественное описание шансов на тот или иной исход описывается на основе понятия *вероятности.*

Всё множество изучаемых элементов называется *генеральной совокупностью.* Если вся совокупность слишком велика, то приходится изучать часть этой совокупности. Такая группа элементов называется *выборкой* или *выборочной совокупностью.* Естественно, выборку следует делать так, чтобы она наилучшим образом представляла генеральную совокупность, то есть, как говорят, была *репрезентативной.* Наилучшей является случайная выборка, когда каждый объект имеет одинаковую вероятность быть выбранным. По выборке можно определить эмпирическую функцию распределения, оценить среднее значение и дисперсию выборки.

Описательная статистика хорошо представлена в таких пакетах, как STADIA STSTISTICA и STATGRAPHICS. Мы особенно рекомендуем прекрасный отечественный пакет STADIA, работающий полностью на русском языке, простой и удобный в использовании.

Важным средством анализа данных является графическое представление распределений – построение *гистограмм.* Стандартные графические пакеты позволяют также проводить проверку распределения на нормальность, то есть на соответствие гауссову нормальному распределению.

В теории вероятностей имеют дело со следующей ситуацией: если для некоторой системы выполняют ряд условий, которые называют *испытанием*, и в результате система переходит в некоторое определенное состояние, то говорят, что произошло *событие*. События можно разделить на следующие группы:

1. *Достоверное событие* – событие, которое обязательно происходит, если выполнена совокупность условий.

2. *Невозможное событие* – событие, которое никогда не наступит при данных условиях.

3. *Случайное событие* – наступление которого может произойти, а может не произойти.

Теория вероятностей не предсказывает появление того или иного случайного события, а определяет только *вероятность* его наступления. Случайные события, в свою очередь, можно разделить:

1. *На совместные,* когда два события могут наступить одновременно.

2. *На несовместные*, когда наступление одного события исключает появление другого.

Если в результате испытания должно появиться одно и только одно из несовместных событий, то говорят, что события образуют *полную группу событий.* Классическое определение вероятности строится на основе понятия равной вероятности (равновозможности) событий. Равновероятными считаются события, если нет оснований считать, что частота наступления одного события в результате серии испытаний больше, чем у другого.

1.3. Определение вероятности

Вероятность – это количественная характеристика шансов на исход исследуемого события. Вероятность события A обозначают как $P(A)$.

Пусть мы имеем дело с несовместными событиями, образующими полную группу. Назовем *элементарным исходом* каждый из возможных результатов испытания, а благоприятным исходом (события A) – исход, при котором наступает событие A. Тогда *вероятностью события A* называют отношение числа исходов, благоприятствующих событию A, к полному количеству всех равновозможных несовместных элементарных исходов, образующих полную группу.

Иначе говоря, если событие A подразделяется на m частных случаев, входящих в полное множество из n попарно несовместимых и равновозможных событий, то вероятность $P(A)$ события A равна

$$P(A) = \frac{m}{n}. \tag{1}$$

Вероятность $P(A)$ события A равняется отношению числа возможных результатов испытаний, благоприятствующих событию A, к числу всех возможных результатов испытаний. Вероятность удовлетворяет неравенству

$$0 \le P(A) \le 1. \tag{2}$$

Вероятность наступления хотя бы одного из независимых событий равна сумме вероятностей отдельных событий, то есть

$$P(A_1 + A_2) = P(A_1) + P(A_2). \tag{3}$$

1.4. Условная вероятность

Вероятность события A при наложении дополнительного условия, что произошло событие B, называется условной. Рассчитаем вероятность. Пусть событию A благоприятствуют n событий, B – k событий, а пересечению AB

– r событий (рис.1). Если произошло событие B, то это означает, что наступило одно из событий A, благоприятствующих B. При этом событию A благоприятствуют те и только те события, которые содержатся в AB. Таким образом,

$$P(A / B) = \frac{r}{k}. \tag{4}$$

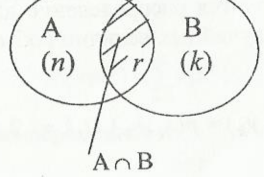

Рис. 1. Пересечение множеств событий

Поскольку B по условию является достоверным событием, то число таких возможных событий k образует весь набор исходных возможностей. Таким образом, условная вероятность

$$P(A / B) = \frac{r}{k} = \frac{r / n}{k / n}. \tag{5}$$

1.5. Случайные величины

Часто элементарному событию можно сопоставить число. Например, при случайном выборе человека из группы – возраст или вес человека, при случайном выборе частицы газа – ее скорость, при случайном выборе момента времени диффундирующей частицы – ее энергию.

Функция, определенная на пространстве элементарных событий, называется *случайной величиной.* В случае дискретного пространства элементарных событий любую случайную величину можно задать, занумеровав в каком-либо порядке все точки пространства и связав с каждой из них соответствующее значение X.

Пусть X – случайная величина, а $x_1, x_2, \ldots$ – ее значения. Совокупность всех элементарных событий, на которых X принимает определенное фиксированное значение x_i, образует событие $X = x_i$; его вероятность обозначается

$$P\{X = x_i\} = f(x_i),\ (i = 1,2,\ldots) \tag{6}$$

и называется распределением вероятностей случайной величины X. Ясно, что

$$f(x_i) \geq 0, \sum_i f(x_i) = 1. \tag{7}$$

Аналогичным образом вводится распределение для двух и более случайных величин. Так для двух случайных величин событием считается $X=x_i$; $Y=y_k$, а функция

$$P(X = x_i, Y = y_k) = p(x_i, y_k),\ (i,k = 1,2,\ldots) \tag{8}$$

называется *совместным распределением случайных величин* X и Y. Если случайная величина является непрерывной, то вместо функции распределения дискретной случайной величины вводят понятие *плотности вероятности*, а суммирование в (7) переходит в интегрирование.

Плотность распределения f непрерывной случайной величины x обладает свойствами:

$$f(x) \geq 0,\ \int f(x)dx = 1. \tag{9}$$

В природе и технике существуют самые разнообразные распределения, которые можно установить на основе проведения соответствующих полных или выборочных измерений.

Математическое ожидание $M(X)$ называется также ожидаемым значением, средним значением или первым моментом:

$$M(X) \equiv \overline{x} = \sum_n x_k P(x_k). \tag{10}$$

Для непрерывных величин

$$M(x) \equiv \overline{x} = \int x f(x) dx. \tag{11}$$

Математическое ожидание величины x^r называется r-м центральным моментом x или r-м моментом x, а

$$D(X) = M(x - M(x))^2 = M\langle (x - \overline{x})^2 \rangle = \sigma^2. \tag{12}$$

D – дисперсия или второй центральный момент x, σ называется среднеквадратичным отклонением случайной величины от среднего значения.

При решении задач теории вероятностей важную роль играют *производящие функции* для вероятностей P_n, определяемые их разложением в ряд Тейлора:

$$g(z) = \sum_{n=0}^{\infty} P_n z^n, \tag{13}$$

где z – вспомогательная переменная и

$$P_n = \frac{1}{n!} \frac{d^n g}{dz^n}\bigg|_{z=0}. \tag{14}$$

Производящую функцию (13) можно выразить через функцию случайной целочисленной переменной x в виде

$$g(z) = M(z^x). \tag{15}$$

Действительно,

$$M(z^x) = \sum_{n=0}^{\infty} P_n(x = n) z^n = g(z),$$

что и доказывает сформулированное утверждение.

1.6. Нормальное распределение

Одним из наиболее известных распределений является *гауссово нормальное распределение* (рис. 2):

$$f(x) = \frac{\exp(-x^2/\ 2\sigma^2)}{\sigma\sqrt{2\pi}}. \tag{16}$$

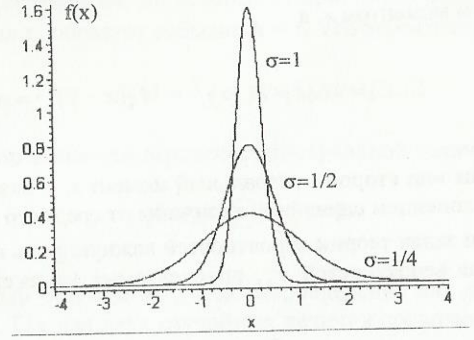

Рис. 2. Вид нормального гауссового распределения

Практическую ценность нормального распределения определяет *центральная предельная теорема Ляпунова:* если случайная величина X представляет собой сумму очень большого числа взаимно независимых случайных величин, влияние каждой из которых на всю сумму пренебрежимо мало, то X имеет распределение, близкое к нормальному.

В приложениях часто приходится иметь дело со случайными величинами, которые слагаются из большого числа независимых компонентов. При очень широких условиях такие величины имеют нормальное распределение вероятностей.

1.7. Распределение χ^2 (Хи-квадрат)

Если есть n нормально распределенных независимых случайных величин χ, причем математическое ожидание каждой из них равно 0, а среднеквадратичное отклонение – единице, то сумма квадратов этих величин

$$\chi^2 = \sum_{i=1}^{n} \chi_i^2 \qquad (17)$$

распределена по закону χ^2 (хи-квадрат) с $k=n$ степенями свободы. Если эти величины связаны одним линейным соотношением, например $\sum_i x_i = n\bar{x}$, то число степеней свободы $k=n-1$. Плотность этого распределения

$$f(x) = \begin{cases} 0, \text{ если } x \le 0, \\ \dfrac{1}{2^{k/2}\Gamma(k/2)} e^{-x/2} \cdot x^{(k/2)-1}, \text{ если } x > 0, \end{cases}$$

$\Gamma(x) = \int_0^x t^{x-1} e^{-t}\, dt$ – гамма функция, в частности $\Gamma(n+1) = n!$

Распределение χ^2 (рис. 3) определяется одним параметром – числом степеней свободы, с увеличением которого распределение плавно приближается к нормальному закону.

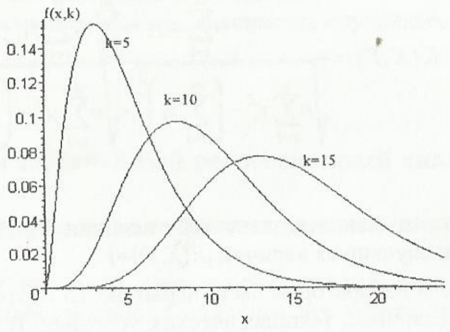

Рис. 3. Распределение χ^2

1.8. Корреляция

Две случайные величины могут быть независимыми или в той или иной мере зависеть друг от друга. Для характеристики меры зависимости двух случайных величин используется *коэффициент корреляции k(x,y)* случайных величин X и Y:

$$K(X,Y) = \frac{M((x-\bar{x})(y-\bar{y}))}{\sigma_x\sigma_y} = \frac{M(XY) - M(X)M(Y)}{\sigma_x\sigma_y} = \\ = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2 \sum_{i=1}^{n}(y_i-\bar{y})^2}}, \qquad (18)$$

где n – число пар точек $X=x_i$, $Y=y_i$.

Используя свойство дисперсии

$$D(X) = M(X^2) - \left[M(X)\right]^2, \tag{19}$$

формулу (18) можно представить в виде

$$K(X,Y) = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sqrt{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \cdot \sqrt{n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2}}. \tag{20}$$

Для статистических независимых величин $K(X,Y)=0$. При согласованном изменении случайных величин $|K(X,Y)|=1$.

Пример. В лаборатории было испытано 15 партий керамзита, изготовленного при различных технологических условиях. В результате испытаний был определен удельный вес керамзита в каждой партии (табл.1). Значение удельного веса является случайной величиной.

Таблица 1

Номер партии	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*
Уд. вес	*897*	*901*	*899*	*900*	*900*	*900*	*898*	*903*	*902*	*899*	*900*	*899*	*901*	*900*	*900*

Найти математическое ожидание m_x и дисперсию D.

Решение

1) $$m = \frac{1}{n}\sum_{i=1}^{n} x_i = (897 + 901 + 899 + 900 + 900 + 900 + 898 + 903 + 902 + 899 + 900 + 899 + 901 + 900 + 900)/15 = 899{,}93.$$

2) $$D = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - m)^2 = ((897 - 899{,}93)^2 + 2\cdot(901 - 899{,}93)^2 + 3\cdot(899 - 899{,}93)^2 + 6\cdot(900 - 899{,}93)^2 + (898 - 899{,}93)^2 + (903 - 899{,}93)^2 + (902 - 899{,}93)^2)/14 = 2{,}2.$$

Контрольные вопросы для самопроверки

1. В чем заключается математическое моделирование?
2. Для чего применяются методы математической статистики?
3. Что такое событие и какие виды событий Вы знаете?
4. Дайте определение вероятности.
5. Что называют случайной величиной?
6. Что называют математическим ожиданием?
7. Чем можно охарактеризовать зависимость случайных величин друг от друга?

Глава 2. Линейный регрессионный анализ

2.1. Приближение табличных значений функций

При анализе тех или иных зависимостей мы часто не знаем закона, управляющего этой зависимостью. Обозначим эту неизвестную функциональную зависимость величины y от числовых переменных $x^{(1)}, x^{(2)}, \ldots x^{(m)}$ как $y=f(x)$ и будем называть y откликом, а x – факторами, влияющими на отклик.

С геометрической точки зрения функция $f(x)$ задает поверхность в пространстве. Если $m=2$, то это поверхность в трехмерном пространстве $(x^{(1)}, x^{(2)}, y)$ (рис. 4).

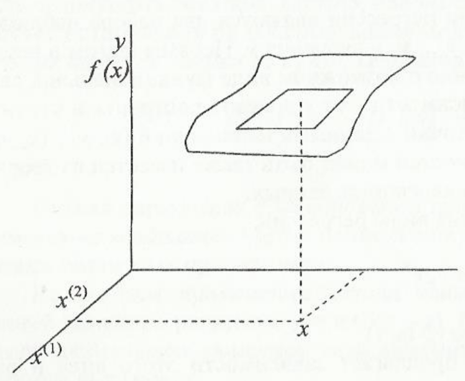

Рис. 4. Функциональная зависимость и ее линейное приближение

Простейшим подходом к заданию $f(x)$ является замена неизвестной поверхности участком плоскости. Такая замена может хорошо описывать поведение функции в некоторой окрестности точки x:

$$y \approx \sum_{j=1}^{m} c_j x^{(j)} + c_0 + \varepsilon , \quad (21)$$

где ε – погрешность приближения.

Задача определения коэффициентов $c=\{c_i\}$, обеспечивающих наилучшее представление функции $f(x)$, заданной набором наблюдений y_k, $x_k^{(i)}$ (k=1,2,3…n), называется *множественной линейной регрессией.*

2.2. Нелинейная регрессия

Если подбираемая функциональная зависимость

$$y = F(x,c) + \varepsilon \quad (22)$$

имеет более сложный вид, чем (21), то ее поиск путем определения коэффициентов c называется *нелинейной регрессией*. Задача нелинейной регрессии в общем случае является очень сложной. Самая простая регрессионная задача состоит в исследовании связи между одной независимой (одномерной) переменной x и одной зависимой переменной (откликом) y. Такая зависимость называется простой регрессией.

Исходными данными задачи регрессии являются два набора наблюдений $x_1, x_2, \ldots x_n$ – значения x, и $y_1, y_2, \ldots y_n$ – значения y. Первым шагом в решении задачи является предположение о возможном виде функциональной связи между x и y. Для подбора зависимости y от x полезно построить и изучить график, на котором изображены точки с данными наблюдений (x_1, y_1), (x_2, y_2), …(x_n, y_n). Примерный вид зависимостей может быть также известен из теории или предыдущих исследований аналогичных данных.

Часто используют следующие виды регрессий:

1) линейную - $y=a+bx$;
2) квадратичную - $y=a+bx+cx^2$;
3) степенную - $y=a_i x^i$;
4) логарифмическую - $y=a+b\ln(x)$;

Например, пакет STADIA предлагает зависимости этого вида и ряд других.

2.3. Оценка точности регрессии

После подбора регрессионной модели следует выяснить, насколько хорошо она описывает имеющиеся данные и найти наилучшее приближение. Единого правила для этого нет. Наиболее обоснованное решение можно принять путем сравнения значений y_i со значениями, полученными с помощью регрессионной функции. Оценку точности аппроксимации нелинейной зависимости можно оценить при помощи корреляционного соотношения

$$\eta = \sqrt{1 - \frac{\sum_{i=1}^{n}(y_i - f(x_i))^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}} . \quad (23)$$

Корреляционное отношение всегда удовлетворяет соотношению $0 \le \eta \le 1$. Если $\eta > K(X,Y)$, то кривая точнее аппроксимирует зависимость, чем прямая; для прямой $\eta = K(X,Y)$.

Разности ε_i между наблюдаемыми и предсказанными на основании регрессионной модели значениями называют *остатками.* Например, для линейной зависимости $y=a+bx$ значения остатков вычисляются как

$$r_i = y_i - f(y_i) = y_i - (a + bx_i),$$

где a и b – оценки коэффициентов a и b.

При выборе методов определения параметров регрессионной модели можно использовать различные критерии, которые обеспечивают минимальность совокупных остатков. Можно, например, оценивать точность по сумме модулей остатков или по максимальному модулю остатка. Одной из наиболее удобных является оценка по сумме квадратов:

$$\sum_{i=1}^{n}[y_i - f(x_i, c)]^2 \to \min . \quad (24)$$

Оценка параметров c на основании критерия (24) называется *методом наименьших квадратов*. Метод наименьших квадратов широко используется в самых различных приложениях.

Рассмотрим применение метода наименьших квадратов для случая простой линейной регрессии $y = a+b(x - x)$. Для нахождения оценок a и b по методу наименьших квадратов надо выяснить, при каких a и b достигается минимум выражения

$$S = \sum_{i=1}^{n}[y_i - a - bx_i + b\bar{x}]^2 . \quad (25)$$

Как известно, необходимым условием достижения минимума функции является равенство нулю ее частных производных:

$$\frac{\partial S}{\partial a}=0,\ \frac{\partial S}{\partial b}=0. \qquad (26)$$

При вычислении производных получим систему уравнений относительно a и b:

$$\sum_{i=1}^{n}[y_i-a-b(x_i+\bar{x})]=0,$$

$$\sum_{i=1}^{n}x_i[y_i-a-b(x_i+\bar{x})]=0. \qquad (27)$$

Решение системы уравнений (27) легко найти: $a=\bar{y}$, где

$$\bar{y}=\frac{1}{n}\sum_{i=1}^{n}y_i; \qquad (28)$$

$$b=\frac{\sum_{i=1}^{n}(y_i-\bar{y})(x_i-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}. \qquad (29)$$

В случае множественной линейной регрессии метод наименьших квадратов также приводит к системе линейных уравнений, которая легко решается на компьютере средствами линейной алгебры. В случае нелинейной регрессии получается система нелинейных уравнений, и поиск решения является сложной задачей. Не всегда ясно бывает и качество нелинейной регрессии.

Уверенность в том, что регрессия, даже линейная, правильно отражает опытные данные, никогда не бывает полной. Имеющийся опыт в этой области можно сформулировать в виде ряда эвристических требований. Первым и наиболее фундаментальным является предположение о том, что результаты отдельных измерений представляют собой независимые случайные величины. Проверка этого свойства невозможна методами статистики, и достоверность результата обеспечивается всей методикой опыта. Вторым важным предположением является равномерность распределения ошибки наблюдения как случайной величины. Это означает, что измерения отклика имеют равную точность при всех значениях фактора, если случайную составляющую отклика рассматривать как ошибку при его измерении. В противном

случае для произвольной изменчивости ошибок классическая регрессионная схема непригодна.

Очень важным является предположение о виде функциональной зависимости, используемой для регрессии. Важно выбрать зависимость $f(x,c)$ так, чтобы она отражала определенные закономерности, хотя бывают эффективные и чисто эмпирические подгоночные формулы. Выбор регрессионной зависимости является одной из основных задач в любом исследовании.

Сделать на основании регрессии определенные статистические выводы, например определить доверительный интервал для коэффициента регрессии, можно, выполнив предположения о гауссовой статистике, то есть о распределении случайных величин по нормальному закону. Современные пакеты типа STADIA позволяют легко проводить проверку на нормальность, например, распределения остатков r_i. Они также позволяют сделать заключение об адекватности регрессии.

Пример. На экспериментальном заводе по приготовлению строительных смесей было исследовано 7 вариантов приготовления некоторой смеси (табл. 2). Для ее приготовления используются два компонента (x_1, x_2) и отвердитель (y).

Определить функциональное соотношение между компонентами и отвердителем, необходимое для приготовления смеси.

Таблица 2

Опытные данные по соотношению между компонентами и отвердителем

x_1	87,28	88,42	87,82	87,22	88,58	87,46	87,91
x_2	12,89	12,37	12,98	13,07	11,70	12,57	13,02
y	0,233	0,228	0,234	0,235	0,222	0,230	0,234

Решение. Предположим, что соотношение компонентов и отвердителя связаны линейной связью, то есть $y=ax_1+bx_2+c$. Тогда, согласно методу наименьших квадратов, устремим к минимуму сумму квадратов отклонений фактических значений количества отвердителя от их предполагаемого значения и рассмотрим ее как некоторую функцию от переменных a, b, c:

$$S(a,b,c)=\sum_{i=1}^{7}\varepsilon^2=\sum_{i=1}^{7}(y_i-(ax_{1i}+bx_{2i}+c))^2\to\min.$$

Функция достигает своего экстремального значения, если ее частные производные обращаются в 0. Эти условия есть уравнения для нахождения коэффициентов *a*, *b*, *c*.

$$\frac{\partial S}{\partial a} = -2\sum x_1^i (y_i - (ax_1^i + bx_2^i + c)) = 0;$$

$$\frac{\partial S}{\partial b} = -2\sum x_2^i (y_i - (ax_1^i + bx_2^i + c)) = 0;$$

$$\frac{\partial S}{\partial c} = -2\sum y_i - (ax_1^i + bx_2^i + c) = 0.$$

Приводя подобные слагаемые, образуем и вычислим суммы

$$\sum_{i=1}^{7} y_i x_1^i = 141{,}89, \qquad \sum_{i=1}^{7} y_i x_2^i = 20{,}47,$$

$$\sum_{i=1}^{7} (x_1^i)^2 = 53979{,}41, \qquad \sum_{i=1}^{7} (x_2^i)^2 = 1122{,}89,$$

$$\sum_{i=1}^{7} x_1^i x_2^j = 7779{,}01, \qquad \sum_{i=1}^{7} x_2^j = 88{,}6,$$

$$\sum_{i=1}^{7} x_1^i = 614{,}69, \qquad \sum_{i=1}^{7} y_i = 1{,}616,$$

получаем систему уравнений

$$\begin{cases} 141{,}89 - 53979{,}41a - 7779{,}01b - 614{,}69c = 0, \\ 20{,}46 - 7779{,}01a - 1122{,}89b - 88{,}6c = 0, \\ 1{,}616 - 614{,}69a - 88{,}6y - 7z = 0, \end{cases}$$

решая которую, находим

$$a = -2.365 \cdot 10^{-4},\ b = 9.17 \cdot 10^{-3},\ c = 0.1356.$$

Контрольные вопросы для самопроверки

1. Что называют линейной регрессией?
2. Что называют нелинейной регрессией?
3. Что называют остатками?
4. В чем суть метода наименьших квадратов?

Глава 3. Временные ряды

3.1. Характеристики временных рядов

При описании данных о развитии того или иного процесса во времени используется понятие временного ряда. *Временным рядом* называется случайная функция времени $x(t)$, заданная в определенные моменты времени t_1, t_2, t_3 … своими значениями x_1, x_2, x_3 …

Временные ряды могут быть *многомерными*, когда одновременно регистрируется несколько характеристик одного процесса в одинаковые моменты времени. Временные ряды часто встречаются на практике. В экономике - это динамика цен, изменение объема продаж, изменение производительности труда, изменение урожайности и т. д. Временные ряды удобно представить графически. При этом они могут иметь самый разнообразный вид: от четко выраженной периодичности (график продаж в течение суток) до случайных колебаний на фоне монотонного изменения величины (рост урожайности).

Анализ временных рядов представляет собой обширную область статистики. Временные ряды, рассматриваемые в разных предметных областях, имеют различные свойства. Поэтому для их исследования используются различные методы анализа. Рассмотрим только ряд основных вопросов, относящихся к применению анализа временных рядов в реализации теории принятия решений в технико-экономических задачах.

Анализ временных рядов преследует весьма широкие цели. В первую очередь он позволяет дать сжатое описание структуры ряда на основе подбора статистической модели, описывающей ряд. Полученная модель позволяет предсказать будущие значения на основе предыдущих наблюдений, а также управлять процессом, порождающим временной ряд.

В реальности, однако, подобные цели далеко не всегда достижимы. Основным препятствием обычно является недостаточный объем статистических данных и изменение статистической структуры временного ряда с течением времени. Эти изменения лишают ценности прошлые наблюдения, поскольку они уже не помогают осуществлять прогноз динамики данных.

3.2. Анализ временных рядов

Анализ временных рядов проводится в несколько стадий:

1. Дается графическое представление временного ряда и проводится его визуальный анализ.
2. Из временного ряда удаляются закономерные составляющие зависимости ряда от времени – тренды и циклические (сезонные) составляющие.

3. Удаляются высокочастотные составляющие временного ряда (фильтрация).
4. Изучается случайная составляющая оставшегося временного ряда и для нее подбирается математическая модель.
5. Осуществляется прогнозирование будущего развития процесса, представленного временным рядом.
6. Проводится исследование взаимозависимости между различными временными рядами

Все эти функции выполняются в современных статистических компьютерных пакетах типа STADIA. Отметим также мощный компьютерный пакет ЭВРИСТА, специально ориентированный на анализ временных рядов.

При анализе временного ряда видимую его изменчивость стараются разделить на две составляющие – закономерную и случайную. Под закономерной (детерминированной) составляющей понимают последовательность значений d_i, вычисляемых по определенному правилу как функцию времени в момент t_i.

Наиболее часто используемые модели разложения временного ряда на детерминированную компоненту и случайную – это *аддитивная* и *мультипликативная модели*.
В аддитивной модели временной ряд

$$x_i = d_i + \varepsilon_i, \tag{30}$$

в мультипликативной модели

$$x_i = d_i \cdot \varepsilon_i, \tag{31}$$

где ε_i – случайная компонента. Эти модели допускают переход от одной к другой, поскольку логарифмирование мультипликативной модели сводит ее к аддитивной:

$$\ln x_i = \ln d_i + \ln \varepsilon_i. \tag{32}$$

Описание детерминированной компоненты временного ряда сильно зависит от области приложений.

В экономических задачах в *детерминированной компоненте* временного ряда d_i обычно выделяют три составляющие: тренд tr_i, сезонную компоненту s_i и циклическую компоненту c_i. В аддитивной модели

$$d_i = tr_i + s_i + c. \tag{33}$$

В ряде случаев к этим компонентам добавляют компоненту, называемую *интервенцией*. Под интервенцией понимают крупномасштабное и кратковременное воздействие на временной ряд.

Трендом временного ряда tr_i называют плавно изменяющуюся непериодическую компоненту ряда. *Сезонная компонента* отражает присущую тому или иному процессу повторяемость во времени. Она часто присутствует в экономических, метеорологических и других рядах и состоит из последовательности почти повторяющихся циклов.

Типичным примером сезонного эффекта является увеличение объема продаж накануне праздников, увеличение объема пассажирских перевозок городским транспортом в утренние и вечерние часы. Главный смысл выделения сезонных компонент заключается в сравнении всех значений временного ряда через определенный период.

Наиболее часто используемые модели тренда:

1. Линейная - $tr_i = a_0 + a_1 i$, (34)

2. Полиномиальная - $tr_i = a_0 + a_1 i + a_2 i^2 + \ldots + a_n i^n$, (35)

3. Логарифмическая - $tr_i = \exp(a_0 + a_1 i)$ (36)

(часто используется для описания процессов с постоянными темпами прироста),

4. Логистическая - $tr_i = a\,(1 + a_1 \exp(-a_2 i))^{-1}$, (37)

5. Гомперца - $\log(tr_i) = a_0 - a_1 a_2^{\,i}$, $(0 < a_2 < 1)$. (38)

Последние две модели задают кривые s-образной формы и соответствуют процессам с нарастающими темпами роста в начале и затухающими в конце процесса.

3.3. Анализ случайной компоненты ряда

Остановимся кратко на математических основаниях анализа случайной компоненты. Важным классом случайных процессов является *белый шум*, то есть временной ряд с нулевым средним и независимыми составляющими его случайными величинами x_i при всех i. Частным случаем такой величины является *гауссовский белый шум*, представляющий собой последовательность независимых нормально распределенных случайных величин с нулевым средним и общей дисперсией σ^2.

Последовательности независимых случайных величин далеко не всегда являются адекватными моделями реальных временных рядов. Часто применяются *процессы скользящего среднего (moving average)*. Процессом скользящего среднего со средним μ называют процесс

$$x_i = \varepsilon_i + \theta\varepsilon_{i-1} + \mu, \qquad (39)$$

где $\mu = \overline{x} = \langle x \rangle$; θ – числовой коэффициент; $i = 1,2,3,...n$; ε_i – независимые случайные одинаково распределенные величины с нулевым средним значением. У процесса скользящего среднего статистически зависимы только соседние величины. Значения процесса с разностью $\Delta i > 2$ статистически независимы, поскольку в их формировании участвуют только разные статистически независимые слагаемые ε_i.

Часто в прикладных задачах встречаются *процессы авторегрессии (autoregression)*. Процессом авторегрессии первого порядка (AR(1)) со средним значением μ называют процесс x_i, удовлетворяющий соотношению

$$x_i - \mu = \phi(x_i - \mu) + \varepsilon_i, \qquad (40)$$

где ϕ и μ – постоянные параметры. Если $|\phi| < 1$, то члены авторегрессии с ростом времени быстро становятся независимыми.

Если вероятностные свойства случайного процесса не меняются с течением времени, то он называется *стационарным случайным процессом*. Для оценки временных рядов широко используют математическое ожидание (среднее или первый момент):

$$\mu = \overline{x} = \langle x \rangle = \frac{1}{n}\sum_{i=1}^{n} x_i, \qquad (41)$$

и автокорреляционную функцию:

$$r(k) = \mathrm{corr}(x_i, x_{i+k}), \qquad (42)$$

где

$$r(k) = \frac{\sum_{i=1}^{n}(\overline{x}_0 - x_i)(\overline{x}_k - x_{i+k})}{\sqrt{\sum_{i=1}^{n}(\overline{x}_0 - x_i)^2 \sum_{i=1}^{n}(\overline{x}_k - x_{i+k})^2}}, \qquad (43)$$

$$\overline{x}_k = \frac{1}{n}\sum_{i=1}^{n} x_{i+k}.$$

Величину k часто называют *задержкой* или *лагом*.

Для того чтобы оценка (41) с ростом n приближалась к истинному значению (была *состоятельна*), достаточно, чтобы дисперсия

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(\overline{x} - x_i)^2 \to 0 \qquad (44)$$

при $n \to \infty$. О стационарности процесса позволяет судить анализ автокорреляционных функций процесса.

Сформулируем *теорему Слуцкого*: для стационарного процесса x_i оценка его среднего значения состоятельна тогда и только тогда, когда

$$\frac{1}{n}\sum_{k=1}^{m-1} r_i(k) \to 0 \qquad (45)$$

при $m \to \infty$, где $r(k)$ – автокорреляционная функция процесса с лагом k.

3.4. Практический анализ и построение прогноза

При анализе временного ряда следует начинать с его графического представления, а затем выделить из него тренд. Процесс удаления тренда и других нестационарных компонент может проходить в несколько этапов. На каждом этапе анализируется ряд остатков, полученный в результате вычитания тренда из исходного ряда. Признаком нестационарности ряда, кроме графического анализа ряда, служат не стремящиеся к нулю автокорреляционные функции и наличие ярко выраженных низкочастотных периодических пиков.

После максимального приближения ряда к стационарному виду подбирается статистическая модель объекта. Чаще всего используются параметрические модели авторегрессии – скользящего среднего (ARMA – модели). После подбора модели проводится оценка её адекватности и дисперсии остатков, которая далее используется для построения доверительных интервалов прогноза. Возможна ситуация, когда несколько моделей описывают временной ряд, так что результат неоднозначен. В ряде случаев модель подобрать не удается вообще.

Для выделения тренда чаще всего используется метод наименьших квадратов, который уже обсуждался в предыдущем параграфе. Часто нарушаются статистические предпосылки регрессионного анализа. При этом удается выделить плавные зависимости, но остатки обладают специфическими статистическими свойствами. Это не позволяет, например, оценить их дисперсию по остаточной сумме квадратов.

Наряду с методом наименьших квадратов для удаления тренда можно использовать и другие методы. Так, линейный тренд может быть удален путем перехода к ряду разностей соседних членов ряда. Этот метод предложен Дж. Боксом и Г. Дженкинсом в 1970 г.

Процедура перехода от ряда x_i при i=1,... n к ряду $y_i = x_i - x_{i-1} = \nabla x_i$ при $i = 2,...n$ называется взятием первых разностей. Оператор ∇ называется простым разностным оператором первого порядка. Длина ряда y_i первых разностей на единицу меньше исходного ряда x_i .

Аналогичным образом можно ввести разностный оператор второго и более высоких порядков. Разностные операторы более высоких порядков позволяют выделять из ряда полиномиальные тренды соответствующего порядка.

После удаления тренда следует оценить сезонную компоненту по ряду $x_i - tr_i$. Если период p последовательности известен, то сезонную компоненту простейшим образом можно оценить как среднее

$$S_i = \frac{1}{m+1}\sum_{l=1}^{m+1}(x_{i+lp} - tr_{i+lp}) \quad i = 1,2,..p. \qquad (46)$$

Усреднение производится по m+1 периоду, содержащему $n=(m+1)p$ значений ряда. Получив оценки сезонных компонент, их легко удалить из рассматриваемого ряда путем вычитания из начальных значений ряда.

При наличии в ряде циклической компоненты часто для сглаживания ряда используется *метод скользящих средних*. В нем производится замена исходного ряда средними значениями в окрестности i на интервале времени, длина которого выбрана заранее. Таким образом, с изменением i среднее скользит вдоль ряда.

В качестве среднего часто выбирается среднее арифметическое:

$$\overline{x}_i = \frac{1}{2m+1}(x_{i-m} + ... + x_i + x_{i+1} + ... + x_{i+m}). \qquad (47)$$

Скользящее среднее сглаживает исходный ряд и даёт представление об общей тенденции – тренде и циклической компоненте. Чем больше выбирается интервал усреднения, тем более гладкий вид имеет график скользящих средних. Поэтому необдуманное применение процедуры со слишком большим интервалом сглаживания приведет к потере сезонной компоненты. С другой стороны, соседние члены ряда скользящих средних сильно коррелируют. Причина этого заключена в том, что в формировании ряда участвуют одни и те же члены исходного ряда. Это может приводить к тому, что ряд

скользящих средних может порождать циклические компоненты, которые отсутствуют в исходном ряде (*эффект Слуцкого – Юла*).

Метод оценки сезонных компонент при использовании метода скользящих средних в целом аналогичен процедуре с использованием выделенного тренда. Если период последовательности известен, то для каждого сезона рассчитываются разности $x_i - \overline{x}_i$, а в качестве простейшей оценки берется простое среднее по $m+1$ периоду, то есть

$$S_i = \frac{1}{m+1}\sum_{l=1}^{m+1}(x_{i+lp} - \overline{x}_{i+lp}), \quad i = 1,...p. \qquad (48)$$

В аддитивной модели дальнейшее удаление сезонной компоненты сводится к ее вычитанию из исходного ряда. В мультипликативной модели $x_i = c_i \cdot s_i \cdot r_i$ переходят к логарифмам, превращая модель в аддитивную.

Бокс и Кокс в 1964 году использовали семейство преобразований, позволяющих изменить масштаб данных временного ряда. Преобразование Бокса – Кокса применимо только к положительным числовым рядам и определяется с помощью формулы

$$F(x,\lambda) = \begin{cases} \dfrac{(x^\lambda - 1)}{\lambda}, \lambda > 0, \\ \log x, \quad \lambda = 0. \end{cases} \qquad (49)$$

Преобразование Бокса – Кокса при $\lambda > 1$ растягивает расстояния между малыми значениями и сжимает его между большими по величине значениями. При $\lambda < 1$, наоборот, уменьшаются расстояния между меньшими значениями и увеличиваются между большими. Однако преобразование Бокса – Кокса существенно влияет на статистические свойства процесса и может значительно усложнить дальнейший подбор модели ряда.

Временные ряды можно анализировать с помощью пакетов STADIA, STATISTICA и STATGRAPHICS. Анализ временных рядов является специфической и обширной областью статистики, поэтому для их анализа используются также и специализированные пакеты. Одним из наилучших пакетов такого рода является пакет ЭВРИСТА.

Контрольные вопросы для самопроверки

1. Что называется временным рядом?
2. Какие виды временных рядов вы знаете?
3. Какие стадии можно выделить при анализе временных рядов?
4. Что называют «белый шум»?
5. Что называют «процесс скользящего среднего»?
6. Что называют «стационарный случайный процесс»?.

Глава 4. Многомерный статистический анализ

4.1. Многомерные данные

Многомерные методы представляют графические и вычислительные средства для классификации и объединения элементов в группы на основе сходства и близости данных, представленных в виде множества переменных, относящихся к этим элементам.

Важной задачей многомерного анализа является снижение размерности, то есть выделение в пространстве параметров явления наиболее значимых координат. Одним из самых мощных методов многомерного анализа является кластерный анализ. Кластерный анализ позволяет построить дерево классификации n объектов посредством иерархического объединения их в группы или кластеры. Классификация строится на основании анализа расстояний в пространстве m переменных, описывающих объекты. В результате исходное множество объектов разбивается на подмножества компактных кластеров. Кластерный анализ не даёт оценки адекватности получаемых классификаций.

4.2. Метрика

Исходные данные представляются в виде матрицы размером $n \times m$. Пусть индекс i нумерует объекты, а индекс k нумерует количественные признаки. Далее необходимо выбрать метод вычисления расстояния d_{ij} между объектами в многомерном пространстве. Вычисление расстояний между объектами с индексами i и j в зависимости от вида метрики производится по следующим формулам:

1. Для Евклидовой метрики

$$d_{ij} = \sqrt{\sum_{k=1}^{m} (x_{ik} - x_{jk})^2}\ ; \qquad (50)$$

2. Для суммы квадратов

$$d_{ij} = \frac{1}{n} \sum_{k=1}^{m} (x_{ik} - x_{jk})^2\ ; \qquad (51)$$

3. Для метрики Манхеттен

$$d_{ij} = \frac{1}{n} \sum_{k=1}^{m} \left| x_{ik} - x_{jk} \right|; \qquad (52)$$

4. Для метрики Канберра

$$d_{ij} = \sum_{k=1}^{m} \frac{\left| x_{ik} - x_{jk} \right|}{x_{ik} - x_{jk}}; \qquad (53)$$

5. Для метрики матриц Брея – Кортиса

$$d_{ijj} = \frac{\sum_{k=1}^{m} \left| x_{ik} - x_{jk} \right|}{\sum_{k=1}^{n} x_{ik} + \sum_{k=1}^{n} x_{jk}}. \qquad (54)$$

Расстояние между данным кластером с объектами i и всеми другими кластерами j может вычисляться в соответствии со следующими стратегиями: ближайшего соседа $\min(d_{ij})$, дальнего соседа $\max(d_{ij})$ и ряда других. Эффективные программы кластерного анализа содержатся в пакетах STADIA и STATISTICA.

4.3. Факторный анализ

Переменные, описывающие исследуемый объект, могут иметь достаточно условный характер, лишь качественно отражая взаимозависимости и внутреннюю структуру объекта. Такие параметры называются *факторами*. Среди факторов можно выделять те, которые оказывают на некоторый показатель наибольшее влияние. Выделение таких наиболее существенных факторов и составляет предмет *факторного анализа.* Этот метод успешно применяется в нечетких науках – таких, как психология, социология, экономика и др.

Факторный анализ включает несколько этапов. Вначале нужно выделить исходные факторы. Затем производится выделение *главных компонент.* Исходные данные, как и в кластерном анализе, представляются в виде значений m переменных для n объектов. Суть метода главных компонент состоит в следующем. Степень взаимозависимости переменных определяется уровнем коэффициента корреляции. В случае большого числа переменных нужно по имеющимся данным вычислить матрицу R_{ij} корреляций между исходными переменными. Для определения наиболее коррелированных комбинаций факторов находят собственные значения матрицы R_{ij}. Располагая собственные векторы в порядке убывания, получают новые координаты, в которых основных меньше, чем исходных. Это во многих случаях позволяет сократить необходимый набор переменных до двух – трех, что упрощает дальнейший анализ.

Исследование главных компонент наиболее плодотворно, когда все переменные имеют одинаковую природу и одинаковые единицы измерения, например характеризуют структуру экономической деятельности предприятия (в процентах). Подчеркнем, что новые переменные, полученные путем диагонализации матрицы корреляции, могут не иметь прямого смысла и простой интерпретации. Однако новые переменные указывают комбинации, при которых факторы максимально взаимодействуют.

4.4. Статистическое распознавание катастроф

Следуя В.И. Арнольду, будем называть *катастрофами* скачкообразные изменения, возникающие в виде внезапного ответа системы на плавное изменение внешних условий. Рассмотрим сборку Уитни, которая определяется решением уравнения

$$x^3 + ax + b = 0. \qquad (55)$$

Это означает, что при плавном изменении параметров a и b точка x будет находиться как корень уравнения (55). Графический образ уравнения показан на рис. 5.

Из рисунка видно, что в области сборки положение точки на верхнем или нижнем листе поверхности зависит от маршрута движений в координатах (a,b). При одинаковых параметрах (a,b) в конечном состоянии значения x значительно отличаются. С ростом b точка скачком переходит на нижний лист.

Поставим вопрос о статистической различимости сборки Уитни в области А, где катастрофа только зарождается. Он связан с надеждой распознать катастрофу ещё на ранних подступах к ней. В области А ещё нет двузначности функций $x(a,b)$. В дальнейшем параметр a мы свяжем со временем t: $a \equiv t$. В реальной ситуации имеются временные ряды x_i, b_i, значения которых соответствуют изменяющемуся времени t_i. Решение задачи заключается в попытке выяснить, не приближаемся ли мы к катастрофе, на основе анализа данных о временных рядах

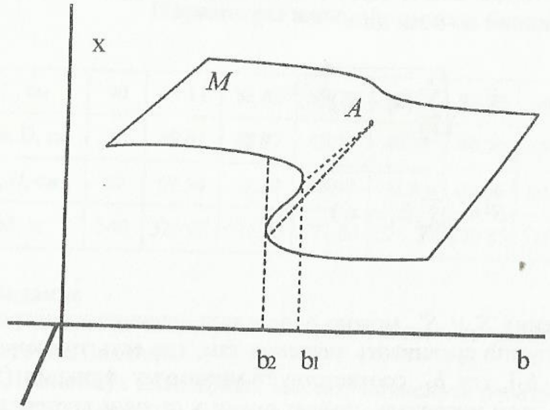

Рис. 5. Поверхность сборки

.Распознавание катастрофы состоит в проверке того, насколько хорошо точки лежат на поверхности М. Функция $x(t,b)$ является двузначной, в то время как функция $b(x, t) = -x^3 - ax$ - однозначной. Такая неадекватность представления $x = x(t,b)$ позволяет рассчитывать на распознавание развивающейся катастрофы. Можно строить кубические аппроксимации:

$$x = a_3 b^3 + a_2 b^2 + a_1 b + a_0 + d_1 t + d_2 t^2 \qquad (56)$$

и

$$b = -(c(x - x_0)^3 + t(x - x_0)) + b_0, \qquad (57)$$

а затем сравнивать их точность.

Если функция (57) значительно точнее аппроксимирует экспериментальные данные, чем (56), то далее ее можно исследовать на наличие максимумов и минимумов и определить наличие катастрофы. Обе зависимости (56) и (57) позволяют определить пары (x_i, b_i) и (x_i, b_i), которые несколько отличны в том смысле, что одна дает значения прямой функции, а другая – обратной. Поэтому их трудно сравнивать непосредственно. Для проведения такого сравнения определим $\tilde{x}_i$ по b_i на основании уравнения (56) и x_i' на основании решения уравнения (57), в котором выбирается корень s_i, наиболее близкий к x_i, то есть из условия $(|x_i - s_i|)$, где $i = 1$ или 1,2,3 в зависимости от числа корней уравнения (56), которых может быть 1 или 3. Затем оценить близость к экспериментальным данным по величине

$$\widetilde{S} = \sqrt{\sum_{i=1}^{n}(x_i - \widetilde{x}_i)^2}\ , \tag{58}$$

$$S' = \sqrt{\sum_{i=1}^{n}(x_i - x'_i)^2}\ . \tag{59}$$

Сравнивая значения S и S', можно определить наличие катастрофы в структуре данных. Разумно сравнивать значения там, где есть три корня, то есть на интервале $[b_1,b_2]$, где b_1 соответствует минимуму функции (56), а b_2 - её максимуму. На этом интервале следует оценить степень гистерезисности процесса. Для этого необходимо сконструировать безразмерный показатель $k= S'/S$. При k~1 нельзя сказать ничего определенного. При k>1 поверхность достаточно гладкая. При k<1 мы имеем дело с потенциальной катастрофой. Желательно проследить эволюцию k со временем. Два возможных сценария показаны на рис. 6. Первый - мы имеем дело с ложной катастрофой. Второй явственно указывает на наличие катастрофы.

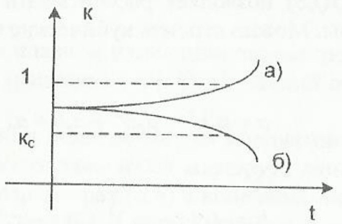

Рис. 6. Эволюция данных: а) – ложная катастрофа; б) – катастрофа

Локальный критерий катастрофы $k(t)$<1, который является необходимым условием. Допустимый уровень $k(t)$, при котором можно с уверенностью говорить о наличии катастрофы в структуре исходных данных, можно определить, видимо, только для конкретных задач.

Пример. По прибытии на строительную площадку партии железобетонных изделий, были произведены следующие измерения блоков ФБС 9-5-6: высота, ширина, длина и масса. Первая колонка табл.3 результатов содержит теоретические параметры, остальные содержат фактические результаты.

Таблица 3

Параметры железобетонных блоков

Длина, L, см	90	89.11	88.86	89.73	88.89	88.85	89.79	89.92	89.15	89.20
Ширина, D, см	50	49.61	48.97	49.24	49.75	49.70	48.73	49.57	49.47	48.88
Высота, H, см	60	59.50	59.39	58.97	58.83	60.00	59.02	59.50	59.92	58.79
Масса, M, кг	540	526.02	516.93	521.06	520.29	529.85	516.42	530.55	528.59	512.72

Задание

1. Используя кластерный анализ, вычислить расстояние между объектами методом Евклида.
2. Используя кластерный анализ, вычислить расстояние между данным кластером с объектами и всеми другими кластерами методом ближайшего соседа.

Найти, используя факторный анализ, степень взаимозависимости переменных.

Решение. Расстояния $d_{ij} = \sqrt{\sum_{k=1}^{9}(x_{ik} - x_{jk})^2}$.

Составим таблицу расстояний (табл. 4).

Таблица 4

Расстояния между объектами

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(2)	9,117							
(3)	5,04	4,25						
(4)	3,871	12,96	8,905					
(5)	9,676	1,149	4,669	13,53				
(6)	4,607	13,68	9,513	1,377	14,16			
(7)	2,608	11,69	7,615	1,319	12,24	2,161		
(8)	13,34	4,267	8,367	17,2	3,757	17,87	15,92	

Кластеры:

(5,2) → 1,149; (7,4) → 1,319: (7,6,4) → 1,377; (7,1,6,4) → 2,608; (8,5,2) → 3,757; (8,3,5,2) → 4,25; (8,7,1,6,4,3,5,2) → 5,04.

Проведем факторный анализ (табл. 5).

Таблица 5

Результаты факторного анализа

Переменная	Среднее	Ст. отклонение
Длина L, см	89,34	0,4256
Шир. D, см	49,28	0,3799
Высота H, см	59,39	0,4392
Масса M, кг	522,8	6,902

Вычислим корреляционную матрицу (табл. 6).

Таблица 6

Корреляционная матрица

Длина, L, см	Ширина, D, см	Высота, H, см	Масса, M, кг
Ширина, D, см	-0,181		
Высота, H, см	-0,443	0,784	
Масса, M, кг	0,019	0,934	0,834

Критическое значение = 0,6914. Число значимых коэффициентов = 3 (100 %).

Вычислим собственные значения и процент объясняемой дисперсии факторов (табл. 7).

Таблица 7

Собственные значения

Фактор	1	2	3
Собственные значения	2,77	1,059	0,1694
Дисперсия, %	69,26	26,47	4,234
Накопления, %	69,26	95,72	99,96

Контрольные вопросы для самопроверки

1. Что такое кластерный анализ?
2. Что такое факторный анализ?
3. Что называют катастрофой?

Глава 5. Методы исследования операций

5.1. Основные понятия исследования операций

Исследование операций – математическая дисциплина, занимающаяся построением, анализом и применением математических моделей принятия оптимальных решений. Основной задачей исследования операций является выбор в заданном множестве элементов, удовлетворяющих тем или иным критериям. При этом любой элемент множества называют допустимым решением, а выбранный элемент - оптимальным решением. Исследование операций подразумевает определенную последовательность действий в принятии решений и их реализации.

Типичные задачи исследования операций описаны ниже.

5.2. Задача о составлении рациона

Пусть имеется 4 вида продуктов, из которых необходимо составить паек, удовлетворяющий следующим требованиям:

1. В паек должны входить все виды продуктов.
2. Содержание белков, жиров, углеводов в пайке должно быть не менее установленных норм B.
3. Стоимость пайка не должна превосходить некоторой величины C.
4. Вес пайка не должен превышать величины P.
5. Паек должен быть минимального объема.
6. Паек должен иметь максимальную калорийность.

$x_к$ - единицы измерения, $k = 1,2,3,4$.

1. $x_к > 0$.
2. $\sum_{j=1}^{4} a_j \cdot x_j \geq B$.
3. $\sum_{k=1}^{4} c_k \cdot x_k \leq C$.
4. $\sum_{k=1}^{4} p_k \cdot x_k \leq P$.

Система линейных неравенств задает на плоскости многоугольник, а в пространстве - *выпуклый многогранник*. Выпуклой называется область, которая полностью лежит по одну сторону любой касательной.

5. $\sum_{k=1}^{4} V_k \cdot x_k \to \min V$.

6. $\sum_{k=1}^{4} q_k \cdot x_k \to \max Q$.

Условия 5 и 6, задающие в общем случае требование экстремума, называются целевыми функциями.

5.3. Задача о быстродействии

Рассмотрим вначале следующий **пример**. Корабль движется по реке со скоростью $\mathbf{V}(t)$, где t - текущий момент времени относительно течения, скорость которого постоянна как по величине, так и по направлению. Найти программу управления рулями корабля, при которой он достигает заданной конечной точки из заданной начальной точки за min время при условии $\mathbf{V}(t) = \mathbf{V}_0$. Это оптимальная задача быстродействия, т.е. задача о достижении заданного состояния системы за min время. Даная задача является типичной задачей оптимального управления динамической системой. Все задачи исследования операций делятся на 2 типа: детерминированные и стохастические задачи, где поведение системы описывается вероятностью или имеется неполная информация о состоянии системы.

В первом случае говорят о *параметрических задачах* исследования операций. Во втором в *стохастических задачах* находят решения, соответствующие наиболее желательным значениям параметра независимо от того, как конкретно реализуется неопределенность.

5.4. Задача о выборе наилучшей стратегии

Пусть имеется 2 игрока, первый из которых может избрать m стратегий, а другой n стратегий. В зависимости от выбранных стратегий выигрыш составляет величину A_{ij} ($1 \le i \le m$, $1 \le j \le n$). Если $A_{ij} > 0$, то выиграл игрок 1. Если $A_{ij} < 0$, то выиграл игрок 2. Критерием для каждого игрока является max его выигрыша.

5.5. Транспортная задача

Пусть в пунктах $a_1, a_2, \ldots, a_n$ находятся склады, в которых хранятся товары в количествах $X_1, X_2, \ldots, X_n$ соответственно. В пунктах $b_1, b_2, \ldots, b_m$ находятся потребители, которым нужно поставить эти товары в количествах не менее $Y_1, Y_2, \ldots, Y_m$ соответственно. Обозначим через d_{ij} стоимость перевозки единицы груза между пунктами a_i и b_j.

Исследуем операцию перевозки потребителям товаров в количествах, достаточных для того, чтобы удовлетворить потребности потребителей. Обозначим через x_{ij} количество товара, перевозимого из пункта a_i в пункт b_j. Для удовлетворения запросов потребителей необходимо выполнение систем неравенств

$$\sum_i x_{ij} \ge Y_j \qquad (60)$$

при разных значениях x_{ij}.

Со склада с номером i нельзя вывести больше имеющегося на нем запаса, то есть

$$\sum_j x_{ij} \le X_i \qquad (61)$$

Удовлетворить условиям (60) и (61), т.е. составить план, обеспечивающий запросы потребителей, можно бесчисленным множеством способов. Выбор одного из них допускается основывать на одной из возможных оценок решения. Одним из критериев может служить минимальность стоимости перевозок:

$$f(x) = \sum_{ij} d_{ij} x_{ij} \to \min . \qquad (62)$$

5.6. Задача об использовании ресурсов

Предприятие имеет в своем распоряжении определенное количество ресурсов: сырье, оборудование и т.п. в количестве $b_1, b_2, \ldots, b_n$. Пусть a_{ij} – число единиц ресурса i, необходимое для производства товара $j (j=1,2,\ldots m)$. Известно, что доход, получаемый от единицы товара j, есть c_j. Обозначим через x_j количество товара j. Тогда доход предприятия

$$S = \sum_{j=1}^{m} c_j x_j . \qquad (63)$$

Общее количество использованных ресурсов i равно

$$\sum_j a_{ij} x_j \le b_i , \qquad (64)$$

поскольку не превосходит запаса b_i. Математически задача о распределении ресурсов состоит в нахождении неизвестных x_j, удовлетворяющих условиям $x_j \geq 0$, условиям (6) и сообщающих максимальное значение функции S.

Первая и вторая задачи являются типичными задачам *линейного программирования*. Линейное программирование – это математическая дисциплина, изучающая методы нахождения наименьшего (или наибольшего) значения линейной функции нескольких переменных при условии, что последние удовлетворяют конечному числу линейных уравнений и неравенств.

5.7. Задача составления расписаний

Теория расписаний – это большое направление в дискретной математике и теории исследования операций. Она решает задачи выбора очередности и выделения определенного объема ресурса, который должен расходоваться на выполнение работ.

Одна из основных задач этого класса состоит в нахождении такого распределения ресурса и такого назначения очередности работ, при которых совокупность работ, составляющих проект, будет выполнена за минимальное время. Итак, пусть имеется перечень работ $P_1, P_2, P_3, \ldots, P_n$, необходимых для выполнения проекта, и требуемый ресурс для его выполнения. Ресурс может иметь различную природу. Это могут быть люди, оборудование, сырье, деньги. Таким образом, говоря о требуемом ресурсе, мы имеем в виду некоторый параметр, дающий перечень объемов ресурса различной природы. Кроме того, выполнение работ бывает обычно стеснено ограничениями. Их, как правило, удается разбить на две группы:

а) к первой группе относятся ограничения, описывающие взаимную зависимость работ. Это ограничения логического характера. Типичный пример – ограничения типа графа, когда выполнению работы номера i предшествует некоторая совокупность работ, без выполнения которых нельзя приступать к работе с номером i. Так, крыша здания не может быть построена раньше стен, а стены - раньше фундамента. Ограничения этого типа могут быть сформулированы на языке теории графов. Виды работ можно обозначить вершинами ориентированного графа. Тогда ребра графа будут показывать, какие работы и в какой последовательности следует выполнять;

б) второй тип условий связан с объемом ресурса, который может быть выделен на реализацию проекта. Обозначим $\mathbf{V}(t)$ – вектор ресурса, который может быть выделен в некоторый период t (t – номер месяца, квартала или года). Через $u^i(t)$ обозначим долю работы номера i, через $\mathbf{q}^i(u^i(t))$ – вектор

требуемого для выполнения работы ресурса. Тогда ограничения принимают форму

$$\sum_i \mathbf{q}^i(u^i(t)) \leq \mathbf{V}(t). \tag{65}$$

Если векторы $\mathbf{V}(t)$ заданы, то план реализации проекта сводится к следующему: для каждого интервала времени t должен быть указан перечень работ и доля $u^i(t)$ этих работ, которую необходимо выполнять так, чтобы суммарное время выполнения проекта было минимальным. Задача построения расписания – это задача *дискретного программирования*, в котором функции изменяются скачками, а аргумент (время) является дискретным.

5.8. Постановка задач оптимизации

Если целевая функция или ограничения заданы случайными функциями, то такие задачи относятся к *стохастическому программированию*. Если оптимальное решение ищется для ситуации, развивающейся во времени, то такая задача решается методами *динамического программирования*.

Таким образом, *исследование операций* - это класс методов, относящихся к выбору способа действия, варианта плана, то есть к принятию решений. В этом смысле термин «операция» означает любое целенаправленное действие.

Исследование операций является синтетической дисциплиной, в которой можно выделить три главные стадии.

1. *Построение математической модели.* На этом этапе строится математическое описание системы.

2. *Описание операции.* При этом цель действий формулируется в виде некоторой оптимизационной задачи

$$f(x) \to \max,$$

где x – элемент некоторого множества, определяемого природой модели. Поиск максимума осуществляется с учетом налагаемых природой объекта ограничений и дополнительных условий, а также возможных неопределенностей.

3. *Поиск решения оптимизационной задачи.* Строго говоря, только эта стадия исследования операции относится к самой математике.

Контрольные вопросы для самопроверки

1. Какие виды задач решаются методами исследования операций?
2. Что такое стохастическое программирование?
3. Что такое динамическое программирование?

Глава 6. Линейное программирование

6.1. Постановка задачи

При всем многообразии конкретного содержания задачи линейного программирования имеют общую математическую структуру. *Линейное программирование* изучает методы нахождения наименьшего (или наибольшего) значения линейной функции нескольких переменных при дополнительных условиях, имеющих вид линейных уравнений и неравенств.

Это означает, что дана система линейных уравнений

$$\left.\begin{aligned} & a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1, \\ & a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2, \\ & \ldots \\ & a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m \end{aligned}\right\} \tag{66}$$

и линейная функция

$$f = c_1x_1 + c_2x_2 + \cdots + c_nx_n. \tag{67}$$

Требуется найти такое неотрицательное решение

$$x_1 \geq 0,\ x_2 \geq 0,\ \ldots\ x_n \geq 0, \tag{68}$$

системы (66), при котором функция f принимает наименьшее значение.

В реальных задачах часто ограничения накладываются на линейную комбинацию переменных в виде

$$a_1x_1 + a_2x_2 + \ldots + a_nx_n + b_n \geq 0. \tag{69}$$

Вводя дополнительное неизвестное

$$x_{n+1} = a_1x_1 + a_2x_2 + \cdots a_nx_n + b_n \geq 0,$$

сводим неравенство (69) к совокупности линейного уравнения и условию $x_{n+1} \geq 0$, то есть приводим задачу к стандартной форме (66), (67), (68).

Часто в задаче линейного программирования требуется отыскать не минимум, а максимум линейной функции f. Так как

$$\max f = -\min(-f), \tag{70}$$

то одна задача сводится к другой заменой f на $-f$.

Любую задачу линейного программирования можно сформулировать так, что все ограничения будут иметь вид неравенств. Из равенств (66) можно выразить r неизвестных $x_1,\ x_2,\ \ldots,\ x_r$ в виде

$$\begin{aligned} & x_1 = d_{1,r+1}x_{r+1} + \ldots + d_{1,n}x_n + \beta_1, \\ & x_2 = d_{2,r+1}x_{r+1} + \ldots + d_{2,n}x_n + \beta_2, \\ & \ldots \\ & x_r = d_{r,r+1}x_{r+1} + \ldots + d_{r,n}x_n + \beta_r. \end{aligned} \tag{71}$$

Подставляя эти выражения для $x_1,\ x_2,\ \ldots,\ x_r$ в функцию f, мы выразим f также через $x_{r+1},\ \ldots,\ x_n$:

$$f = \gamma_0 + \gamma_{r+1}x_{r+1} + \cdots + \gamma_nx_n.$$

В силу неравенств (68) исходная задача теперь может быть сформулирована в следующем эквивалентном виде:

Дана система r неравенств

$$\begin{aligned} & x_1 = d_{1,r+1}x_{r+1} + \ldots + d_{1,n}x_n + \beta_1 \geq 0, \\ & x_2 = d_{2,r+1}x_{r+1} + \ldots + d_{2,n}x_n + \beta_2 \geq 0, \\ & \ldots \\ & x_r = d_{r,r+1}x_{r+1} + \ldots + d_{r,n}x_n + \beta_r \geq 0 \end{aligned} \tag{72}$$

с n-r неизвестными $x_{r+1},\ \ldots,\ x_n$, а также линейная функция

$$f = \gamma_0 + \gamma_{r+1}x_{r+1} + \cdots + \gamma_nx_n. \tag{73}$$

Требуется среди всех неотрицательных решений $x_{r+1} \geq 0,\ \ldots,\ x_n \geq 0$ этой системы найти минимизирующее функцию f.

Любое неотрицательное решение системы ограничений называется *допустимым*. Допустимое решение, дающее минимум функции f, называется *оптимальным*. Саму функцию f часто называют *линейной формой*.

Отметим, что задача о поиске экстремума линейной функции многих переменных (8) не может быть решена обычными методами математического анализа. Действительно, частные производные f равны коэффициентам при неизвестных, которые в нуль одновременно не обращаются. Это означает, что функция f не достигает экстремума во внутренних точках области, задаваемой системой неравенств (73). Таким образом, если минимум f существует, то он достигается на границе.

6.2. Геометрическая интерпретация

Задачи линейного программирования имеют геометрическое истолкование. Наиболее просто его понять в случае системы линейных неравенств с двумя переменными x и y. Тогда система ограничений имеет вид

$$ax + bx + c \geq 0 \,. \qquad (74)$$

Пусть для определенности $b>0$, тогда

$$y \geq -\frac{c}{b} - \frac{a}{b} x \,. \qquad (75)$$

Случаю равенства в (75) соответствует уравнение прямой, показанной на рис. 7.

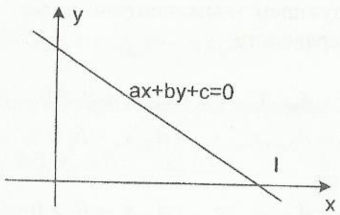

Рис. 7. Разбиение плоскости на две области неравенством (74)

Точки плоскости, удовлетворяющие неравенству (75), лежат выше прямой l.

Если имеется система неравенств

$$a_i x + b_i x + c_i \geq 0 \qquad i = 1,2,\ldots n \,, \qquad (76)$$

то она определяет пересечение системы полуплоскостей, представляющее собой некоторую многоугольную область M, как показано на рис. 8 и рис. 9.

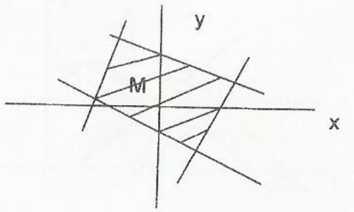

Рис.8. Ограниченная область решений системы (76)

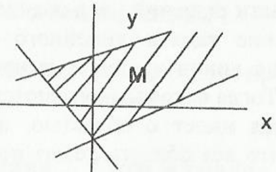

Рис.9. Неограниченная область решений системы (76)

Область решений M является *выпуклой*. Это означает, что она целиком лежит по одну сторону от любого отрезка, соединяющего две соседние вершины границы области M. Выпуклость многоугольника решений вытекает из того, что он получен путем пересечения нескольких полуплоскостей. В случае большего числа переменных область решений системы (72) образует выпуклый многогранник.

Выясним теперь геометрический смысл экстремума линейной формы f для двух переменных:

$$f = c_1 x + c_2 y \,. \qquad (77)$$

Зафиксируем временно некоторое значение f . Тогда этому значению соответствуют точки x и y, лежащие на прямой. Если значение f взять достаточно большим по модулю и отрицательным, то линия (77) будет лежать заведомо вне области M, если она замкнута, как показано на рис. 10.

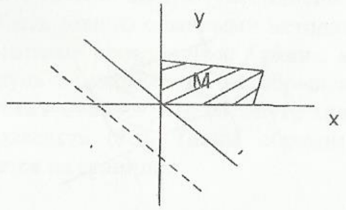

Рис. 10. Изменение взаимного расположения линии уровня и области решения при изменении f

Увеличивая значения f, мы будем смещать линии уровня параллельно друг другу, пока одна из них не коснется области решений. Точка касания линии уровня и области решений дает минимальное значение f в области M, то есть дает решение задачи линейного программирования. Возможен случай, когда встреча линии уровня с границей области M произойдет на отрезке ее границы. Тогда искомыми являются все точки этого отрезка.

Прямая, которая имеет с областью, по крайней мере, одну общую точку, притом так, что вся область лежит по одну сторону от этой прямой, называется *опорной* по отношению к этой области. С учетом этого исходная задача линейного программирования может быть сформулирована следующим образом: среди прямых уровня f найти опорную по отношению к области M, причем такую, чтобы вся область лежала со стороны больших значений f. Тогда любая из точек пересечения этой прямой с M дает решение задачи.

Геометрическая интерпретация задачи линейного программирования показывает, что ее можно решать последовательным переходом от одной вершины многогранника к другой с последовательным увеличением целевой функции f. Соответствующий эффективный алгоритм носит название *симплекс – метод* и реализован в виде компьютерных программ.

В ряде случаев переменные в задаче линейного программирования могут принимать только дискретные значения, как, например, построенные здания, количество машин, число бригад рабочих. В этом случае наиболее часто задача сначала заменяется непрерывной задачей линейного программирования, для решения которой существуют хорошо развитые методы. Дальнейший поиск целочисленного минимального решения проводится так называемым методом отсечения и заключается в нахождении целочисленной точки, которую первой пересекает линия уровня при изменении f, обеспечивающим ее движение от вершины многогранника в глубь области допустимых решений, как показано на рис. 11.

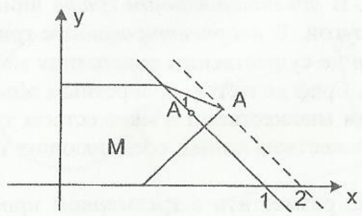

Рис. 11. Оптимальное решение для непрерывных переменных лежит в вершине A (пунктирная линия уровня), дискретное решение – точка A^1

Контрольные вопросы для самопроверки

1. Дайте характеристику методу линейного программирования.

Глава 7. Сети и графы

7.1. Задачи о сетях

Сети с точки зрения их геометрии представляют собой узлы, соединенные связями-дугами. Такие объекты в математике называются графами. Возникает вопрос, насколько такие структуры позволяют обеспечить представление и обработку данных, имеющих сложную иерархическую структуру, отображать различные связи и симметрии, способствовать представлению геометрических объектов разной степени сложности и реализовывать многовариантные стратегии поведения. Для ответа на эти и многие другие подобные вопросы необходимо иметь достаточно определенное представление об основных свойствах графов.

Отметим, что психолог Левин еще в 1936 г. высказал предположение о том, что «жизненное пространство» индивидуума можно представить в виде планарной (плоской) карты. На такой карте отдельные области представляют различные типы деятельности человека, например то, что он делает на работе, дома, или же его хобби (Lewin K. Principles of topological psychology. New York, 1939).

7.2. Общие свойства графов

Существует два основных вида графов: неориентированные и ориентированные. *Граф* представляет собой множество точек, часть из которых или все соединены линиями. В *ориентированном графе* линии имеют направление от одной точки к другой. В *неориентированном* графе линии не имеют направления. Для графов не существенны расстояния между точками и формы соединяющих линий. Граф является дискретным объектом и может быть задан двумя дискретными множествами – множеством точек, которые называются *вершинами*, и множеством линий, соединяющих некоторые вершины, - *ребрами*.

Любой граф можно разместить в трехмерном пространстве. Докажем это свойство. Разместим все вершины графа вдоль прямой линии. Выберем любую пару вершин графа. Их может соединять ребро. Если такое ребро есть, то проведем плоскость через прямую, на которой лежат вершины, и расположим ребро, соединяющее эти две вершины, в проведенной плоскости. Далее возьмем другую пару вершин, поместим их на прямую и повторим процедуру. В результате все выделенные ребра разместятся на веере плоскостей, пересекающихся по одной прямой.

Каждое ребро графа соединяет ровно две вершины. Графы используют во всех областях науки и техники, в частности при принятии решения и в задачах оптимизации.

Граф называется *простым*, если каждую пару вершин соединяет не более чем одно ребро. Граф называется *мультиграфом*, если хотя бы одну пару вершин соединяет более чем одно ребро. Ребра мультиграфа, соединяющие одну и ту же пару вершин, называются кратными. Примером мультиграфа может служить два населенных пункта, соединенных несколькими дорогами.

В простом графе ребро однозначно определяется парой вершин, которые оно соединяет, причем порядок вершин в паре не важен. В мультиграфе каждое ребро должно иметь свое собственное имя. Ребро, соединяющее вершину с самой собой, называется петлей. Граф, не имеющий ребер, называется пустым. Граф, в котором все вершины соединены между собой ребрами, называется полным.

7.3. Задание графа матрицами

Неориентированный граф задает два отношения между своими элементами: отношение смежности и отношение инцидентности. Две вершины называются смежными, если они соединены ребром. Таким образом, *смежность* - это бинарное отношение, так как относится к двум элементам: $V_j \in V$; тогда (V_iV_j) - ребро. Вводят *матрицу смежности* следующим образом: C_{ij}=1 – вершины имеют общее ребро; C_{ij}=0 – вершины не соединены. Матрица смежности полностью задает граф. Пример графа показан на рис. 12.

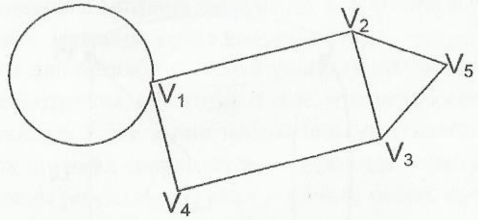

Рис. 12. Пример графа

Зададим его матрицу смежности:

$$C = \begin{pmatrix} 11010 \\ 10101 \\ 01011 \\ 10100 \\ 01100 \end{pmatrix}.$$

Коэффициенты на диагонали соответствуют петлям. Если по диагонали стоят нули, то, значит, петель нет. Для мультиграфа коэффициенты матрицы смежности равны числу ребер, соединяющих вершины.

Одному и тому же графу могут соответствовать разные матрицы смежности, так как порядок нумерации вершин графа не важен. Таким образом, между матрицами смежности и графами нет взаимно однозначного соответствия. Разным матрицам соответствуют разные графы, и в то же время одному графу могут соответствовать разные матрицы.

Инцидентность - это отношения между вершинами и ребрами. Ребро инцидентно каждой из вершин, которое оно соединяет. Инцидентность может быть задана прямоугольной бинарной матрицей D, в которой число строк равно числу вершин графа, а число столбцов - числу ребер: d_{ij}=1, если V_i инцидентно ребру l_i, d_{ij}=0, если V_i не инцидентно ребру l_i. В качестве примера рассмотрим граф, показанный на рис. 13. Соответствующая матрица инцидентности приведена в виде табл. 8.

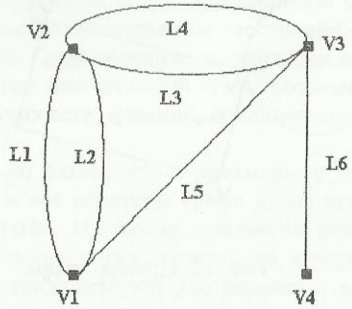

Рис. 13. Граф для построения матрицы инцидентности

Таблица 8

Матрица инцидентности имеет для данного графа вид

	l_1	l_2	l_3	l_4	l_5	l_6
V_1	1	1	0	0	1	0
V_2	1	1	1	1	0	0
V_3	0	0	1	1	1	1
V_4	0	0	0	0	0	1

Таким образом, можно однозначно задать граф при помощи *матрицы инцидентности*. Число ребер инцидентных вершине V_i называется степенью этой вершины. Вершина, степень которой равна единице, называется концевой или висячей. Граф называется подграфом некоторого исходного графа, если он содержит часть вершин исходного графа, соединённых так же, как это было в исходном графе. Матрица инцидентности подграфа легко получается вычёркиванием соответствующих строк и столбцов. Между матрицей инцидентности и матрицей смежности существует алгебраическая связь.

7.4. Ориентированные графы

Ориентированные графы (орграфы) задаются множеством вершин и множеством ориентированных рёбер. Ориентированное ребро графа соединяет вершину V_i с V_j и задаётся упорядоченной парой (V_i, V_j) с началом в V_i и концом в V_j. Ориентация ребра на графе указывается стрелкой. Примером орграфа может служить транспортная система с односторонним движением.

Для орграфа его бинарная матрица смежности C в общем случае не симметрична. Коэффициент матрицы c_{ij}=1, только если $V_i \rightarrow V_j$. Таким образом, матрица смежности для орграфа не обязательно является симметричной. Если же матрица симметрична, то это означает, что ребро соединяет не одну вершину, а их минимум - две и они направлены противоположно. Понятие инцидентности для орграфа сохраняется, но матрица инцидентности D различает начало и конец ребра: d_{ij}=-1, если j- начало ребра; d_{ij}=1, если j- конец ребра; d_{ij}=0, если вершины ребра соединены.

7.5. Пути и связность в графе

Путь в неориентированном графе – это последовательное соединение между собой первой вершины первого ребра (начала пути) и последней вершины последнего ребра (конец пути). Число ребер в пути называется его длиной (l). Длина может принимать только целые значения. Путь называется циклическим или *циклом*, если начало и конец пути совпадают. Путь называется *цепью*, если каждое ребро в нем встречается не более одного раза, и простой цепью, если любая вершина графа встречается в нём не более чем 1 раз, таким образом, простая цепь - это цепь, которая не пересекает сама себя.

Вершины V_i и V_j называются связными, если существует путь с началом в V_i и концом в V_j. В этом случае говорят также, что вершина V_j достижима из вершины V_i. Путь в ориентированном графе – это такая последовательность ориентированных ребер, при которой конец любого ребра совпадает с началом следующего ребра.

Имеет место следующая **теорема**: элемент $(i;j)=c_{ik}^{(l)}$ матрицы C^l, являющейся степенью матрицы смежности, равен числу путей длины l из $i \rightarrow j$. Доказательство производится методом индукции, при этом для l=1 утверждение теоремы очевидно, поскольку для матрицы смежности вершины i,j либо соединены напрямую и тогда l=1, либо не соединены и тогда l=0.

7.6. Деревья

Неориентированный граф без циклов называется неориентированным деревом. Если граф без циклов несвязный, он называется лесом. Пример дерева показан на рис. 14.

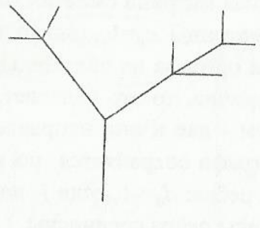

Рис. 14. Дерево

7.7. Планарный граф

Связный неориентированный граф называется планарным, если его можно уложить на плоскости. Граф укладывается на поверхности, если его можно изобразить на этой поверхности так, что никакие ребра не будут пересекаться. Плоским называется граф, который уже уложен на плоскости.

7.8. Стратегии поиска в пространстве состояний

Обычный поиск в пространстве состояний характеризует решение задачи как процесс нахождения пути решения. Под решением мы понимаем цепочку, ведущую к решению задач от исходного состояния к целевому. Поиск пространства состояний можно вести в двух направлениях:

- от исходных данных задач к целям;
- от целей к исходным данным.

При поиске на основе данных исследователь начинает процесс решения задачи, анализируя ее условие, а затем применяет допустимые ходы или правила изменения состояния. В процессе поиска правила применяются к фактам, которые, в свою очередь, используются для генерации новых фактов. Этот процесс продолжается, и в определенных случаях достигается цель.

Возможен альтернативный подход, когда движение в процессе решения ведется от цели, анализируются правила или допустимые ходы, ведущие к цели. Поиск от цели лежит в основе такого метода оптимизации, как *динамическое программирование*. Определив направление поиска (от данных или от цели), алгоритм поиска должен определить порядок исследования дерева или более общего графа, представляющего состояние и переходы в нижний уровень.

Альтернативными стратегиями являются поиск в глубину и поиск в ширину. Задача поиска в принципе может решаться простым перебором вариантов, однако практически это часто нереализуемо из-за большого числа возможных решений.

При поиске в глубину осуществляется последовательный переход от вершины к вершине при пошаговом удалении от начальной вершины графа. При этом в каждой вершине выбирается одна из ветвей с учетом имеющихся правил. Каждое следующее удаление на один шаг от вершины приводит в состояние, которое называется дочерним. Дочернее состояние генерируется правилами вывода, допустимыми ходами игры или другими операциями перехода состояния. Если после исследования всего графа цель не достигнута, то поиск потерпел неудачу.

При поиске в ширину узлы графа рассматриваются по уровням. При этом исследуются те состояния, пути к которым короче. Тем самым при поиске в ширину найденное решение будет оптимальным в смысле длины пути (количество ребер) к целевой вершине.

Найденное решение при поиске в глубину может соответствовать не самому короткому пути. Существуют и разнообразные комбинации этих методов.

7.9. Эвристический поиск

Эвристика определяется как область изучения методов и правил открытий и изобретений. Цель эвристических методов - заменить полный перебор задачи поиска рассмотрением наиболее перспективных состояний.

Эвристика чаще всего используется в двух ситуациях:

1. Проблема может не иметь точного решения из-за неопределенности в постановке задачи или в исходных данных. Примерами могут служить диагностика и системы технического зрения, когда визуальная сцена часто неоднозначна, как это бывает при оптическом обмане.

2. Проблема может иметь точное решение, но стоимость его поиска может быть непомерно высока.

Наиболее простой путь эвристического поиска – поиск ближайшего экстремума. Основой стратегии на поиске экстремума обычно оценивают не только текущее состояние поиска, но и его потомков. Для дальнейшего поиска выбирается наилучший потомок, при этом о его братьях и родителях просто забывают. Поиск прекращается, когда достигается состояние, которое лучше чем любой из его наследников. Такой поиск подобен тактике энергетического, но слепого альпиниста, поднимающегося по наиболее крутому склону до тех пор, пока он не сможет идти дальше. Так как в такой стратегии данные о предыдущих состояниях не сохраняются, то алгоритм не может быть продолжен путем частичного возврата из точки, которая привела к неудаче. Ясно, что основной недостаток такого алгоритма – тенденция, оставшаяся в локальном экстремуме, которая может не совпадать с глобальной.

Одним из самых распространенных алгоритмов является так называемый «жадный» эвристический алгоритм. "Жадный" алгоритм сохраняет списки пройденных состояний и возвращается после каждой пробы на шаг назад, чтобы в конечном итоге сделать шаг, наиболее приближающий к цели. Затем ситуация повторяется.

Контрольные вопросы для самопроверки

1. Что такое граф?
2. Какие основные элементы входят в состав графов?

Глава 8. Оптимизационные задачи на графах

8.1. Порождающие деревья

Если граф конечный и связный, то легко построить дерево (и, как правило, не одно), множество вершин которого совпадало бы с множеством всех вершин заданного графа, а все ребра дерева одновременно были бы ребрами этого графа. Это можно сделать, например, так: пометим произвольную вершину графа, выберем какое-нибудь исходящее из неё ребро графа и пометим вершину, в которое это ребро входит. Если в графе есть ещё вершины, выберем ребро, выходящее из этих двух вершин в какую-нибудь третью вершину графа и пометим её. Если есть еще непомеченные вершины, то повторим процедуру, то есть выберем ребро, соединяющее одну из помеченных вершин с одной из непомеченных. Так как число вершин графа конечно (равно *n*), то потребуется (*n*-1) шаг для исчерпания всех вершин (рис. 15). Полученное таким образом дерево порождает граф или называется *порождающим деревом* графа, или *остовом* графа, или его *стягивающим остовом*.

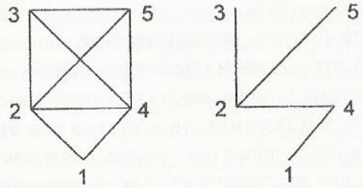

Рис. 15. Остов графа

Введем определение: множество всех ребер, выходящих из вершин графа, называется *звездой* этой вершины. Часто рассматривается задача, где каждому ребру графа приписано некоторое положительное число, которое называется его *весом*, тогда соответствующий граф называется *нагруженным графом* или *сетью*.

8.2. Задача о минимальном порождающем дереве

Если среди порождающих деревьев сети имеется хотя бы одно дерево, сумма длин всех дуг которого минимальна, то такое дерево называется *минимальным порождающим деревом* или *минимальным остовом.* Задача о нахождении минимального остова сводится к решению задачи о нахождении наиболее дешевой сети дорог, соединяющей некоторый набор населенных пунктов.

Пример. Пусть имеется некоторый набор населенных пунктов: A_1, A_2, A_3, …, A_7 , которые могут быть соединены в виде графа, как показано на рис. 16. Требуется построить сеть дорог минимальной суммарной длины, то есть минимальное порождающее дерево.

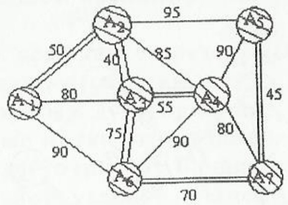

Рис. 16. Населенные пункты, соединенные минимальной по длине сетью дорог

Решение можно начать из любой вершины. Возьмем A_1; самый короткий путь (минимальный вес) - A_1A_2. Рассмотрим две вершины и найдем минимальный вес, то есть A_2A_3. Далее процесс повторяется.

8.3. Алгоритм построения минимального остова

ШАГ 1: Пометим произвольную вершину графа из ребер звезды, порожденной этой вершиной, выберем ребро минимальной длины, если имеются одинаковые минимальные, – то любое из них, и пометим вершину, в которую входит это выбранное ребро. В результате две вершины графа оказываются помеченными. Если других вершин в графе нет, то искомое порождающее дерево построено и задача решена, иначе требуется сделать следующий шаг.

ШАГ 2: Каждая из двух помеченных вершин графа порождает свою звезду. Рассмотрим все ребра этих звезд, кроме ребер, которые соединяют уже помеченные вершины. Из этих ребер выберем наименьшее и пометим вершину, в которую это ребро входит. Если вершины исчерпаны, то построение завершено и задача решена, если есть еще вершины, то второй шаг нужно повторить, рассматривая звезды, которые порождают помеченные вершины. После (n-1) шага все n вершин графа окажутся помеченными, то есть задача будет решена.

Замечание: Шаг 1 удобно начинать с ребра наименьшей длины в сети. Результат применения алгоритма показан на рис. 16 двойными дугами, отмечающими сеть минимальной длины.

8.4. Задача о кратчайшем маршруте между выбранными вершинами

Пусть дана сеть и требуется найти кратчайший маршрут, ведущий из данного узла к каждому из других узлов в сети.

Пример. Рассмотрим работу алгоритма на конкретном примере из 7-ми населенных пунктов (рис. 17).

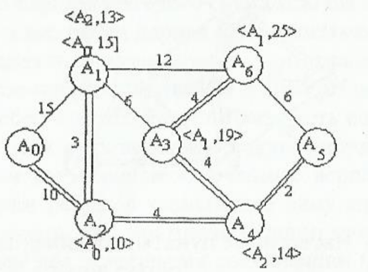

Рис. 17. Кратчайшая сеть для данного начального пункта

Требуется соединить A_0 со всеми пунктами кратчайшей сетью. Решение находится с помощью следующего алгоритма.

Шаг 1: дуга, связывающая A_0 и A_2, является самой короткой – это самый короткий путь в A_2. После окончания шага 1 узлы A_0 и A_2 имеют постоянные метки.

Шаг 2: отбирают все узлы, не имеющие постоянных меток, которые можно соединить с узлами A_0 и A_2 одной дугой. Это узлы A_1 и A_4. Сравним длины маршрутов из A_0 в эти точки. Длина маршрута A_0 , A_1 составляет 15, длина маршрута $A_0 A_2 A_1$ составляет 10+3=13. Так как 15>13, то временную метку $<A_0,15]$ заменяем на метку $<A_2,13]$, следовательно, узел A_4 получает метку $<A_2,14]$. Для узла A_1 маршрут $A_0 A_2 A_1$ является кратчайшим, следовательно, выделяем дугу $A_2 A_1$, а метку $<A_2,13]$ делаем постоянной, то есть $<A_2,13>$.

Шаг 3: отбираем все узлы, которые соединены с узлом A_1 одной дугой и не имеют меток. Таких узлов два - A_3 и A_6. Узел A_3 получает временную метку - $<A_1,19]$, а A_6 получает метку $<A_1, 25]$. Среди узлов с временными метками снова выбираем узел с наименьшим расстоянием от узла A_0. Это

узел A_4. Выделяем ребро. Далее процедуру повторяем. На 6-м шаге получаем данный рис. 17, где двойными линиями показан кратчайший маршрут.

Описание алгоритма. Сначала помечаем начальный узел и объявляем эту метку постоянной.

Шаг 1: рассмотрим все дуги, идущие из начального узла, и припишем всем узлам, с которыми соединены эти дуги, временные метки. Временная метка состоит из элементов, с которыми соединяется узел и которые уже имеют постоянную метку числа, равного сумме длин дуг, соединяющих текущий узел с начальным через узлы с постоянными метками. Далее среди всех узлов с временными метками выбираем узел, расстояние от которого от начального узла минимально. Если таких узлов несколько, выбираем любой. Объявляем временную метку выбранного узла постоянной.

Шаг 2: берем все узлы с постоянными метками и рассматриваем все дуги, идущие из этих узлов с постоянными метками. Вершины, в которые эти дуги приходят, мы пометим временными метками. Временные метки этих узлов содержат элемент «откуда» ребро пришло и расстояние от начального узла. В результате мы получим новый набор узлов с временными метками, образовавшимися на предыдущем шаге и на данном шаге. Все временные метки нужно сравнить и выбрать ту, которая имеет наименьшее расстояние до начального узла. Эту временную метку делаем постоянной и соединяем её постоянной дугой с узлом, указанным в метке. Дальше шаги повторяем до исчерпания сети. Данный алгоритм заканчивается за (n-1) шаг.

8.5. Задача о максимальном потоке

Сеть называется *ориентированной,* если ориентированы все ее дуги, то есть сеть – это нагруженный ориентированный граф.

Задача о потоках возникает при построении следующих моделей:

1. Перемещения товаров и грузов.
2. Перемещения денежных масс.
3. Передачи данных компьютерными сетями.
4. Водопроводных сетей.
5. Электрических сетей.
6. Транспортных потоков.

Узел сети, который является начальным для всех своих дуг (все эти дуги являются выходящими), называют *источником.* Узел, который является конечным для всех своих дуг (все дуги являются входящими), называется *стоком.* Все другие узлы такой сети называются *промежуточными.*

Припишем каждой дуге A_iA_k ориентированного графа (рис. 18) число $c_{ik} = c(A_iA_k)$, то есть пропускную способность этой дуги.

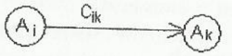

Рис. 18. Дуга ориентированного графа с указанием ее пропускной способности

Пропускной способностью называется максимально возможное количество продукта (например, жидкости или газа в трубопроводе), которое может быть доставлено из узла A_i в узел A_k за единицу времени. Такую сеть нередко называют *транспортной*.

В дальнейшем будем считать, что в транспортной сети есть ровно один источник и ровно один сток. Это ограничение не является существенным, так как все источники можно объединить в один виртуальный источник и также все стоки объединить в один виртуальный сток.

Рассмотрим транспортную сеть с одним источником A_0 и одним стоком A_n, а промежуточные узлы обозначим: $A_1, A_2, \ldots, A_{n-1}$. Будем говорить, что в сети задан поток величины ν, если каждой ориентированной дуге A_iA_k приписано неотрицательное число $\varphi_{ik} \geq 0$, которое называется потоком по дуге A_iA_k. При этом выполнены следующие условия:

1) $\varphi_{ik} \leq c_{ik}$, то есть поток по дуге не может быть больше ее пропускной способности;

2) сумма потоков по всем дугам, выходящим из источника, равна ν (равна входящим в сток)

$$\sum_{i\in\{A_0\}_-}\varphi_{0i} = \nu = \sum_{i\in\{A_n\}_+}\varphi_{in};$$

3) в стационарной сети продукты не накапливаются, поэтому для каждого узла k сумма входящих потоков равна сумме выходящих потоков:

$$\sum_{i\in\{A_k\}_+}\varphi_{ik} = \nu = \sum_{i\in\{A_k\}_-}\varphi_{kl},$$

как показано на рис. 19.

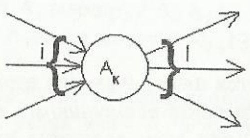

Рис.19. Потоки в узле сети

Третье условие называется *законом Кирхгофа*. Оно означает, что из каждого промежуточного узла выходит ровно столько продуктов, сколько в него поступило, то есть представляет собой математическую запись закона сохранения. Будем считать, что в сети нет равных дуг, то есть дуг, соединяющих одну и ту же упорядоченную пару узлов. В противном случае такие дуги объединяются в одну с общей пропускной способностью.

Рассмотрим произвольный путь, ведущий от источника A_0 в сток A_n.

Если ориентация всех дуг пути совпадает с направлением перемещения от узла A_0 к узлу A_n, то такой путь будем называть *ориентированным путем*. Если на пути есть противоположно ориентированные дуги, то такие дуги мы назовем обратными, в отличие от прямых дуг, направленных по пути.

Пример. Построим полный поток в сети, показанной на рис. 20. Рассмотрим путь: $A_0A_1A_4A_5$ – этот путь является ориентированным. Путь $A_0A_2A_1A_3A_5$ - не ориентирован.

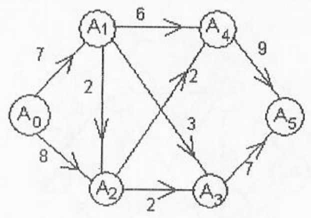

Рис. 20. Сеть с указанием пропускных способностей

Дуга сети называется *насыщенной*, если поток через нее равен пропускной способности. Поток в сети будем называть *полным*, если любой ориентированный путь из источника в сток содержит, по меньшей мере, одну насыщенную дугу. Если путь содержит ненасыщенные дуги, то поток по этому пути можно увеличить до пропускной способности (минимальной) дуги в этом пути.

Начальный поток примем равным нулю. Выберем ориентированный путь, ведущий из источника A_0 в сток A_5. В нем ни одна из дуг не является насыщенной. Увеличим поток до минимального значения $\varphi_{ik} = \min(c_{ik})$. Если поток при этом не будет полным, то выберем в сети следующий путь, содержащий ненасыщенные дуги и вновь увеличим поток до минимальной пропускной способности дуги в этом ориентированном пути. Поскольку сеть конечная, а пропускная способность дуг также конечна, то через конечной число ходов алгоритм завершится и построенный поток в сети окажется полным.

1. В рассматриваемом примере это достигается следующим образом. Выберем, например, путь $A_0A_1A_4A_5$, минимальная пропускная способность которого равна 6 (рис. 21).

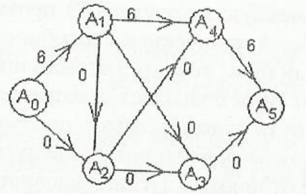

Рис. 21. Построенный поток в сети 1

Новая исходная сеть после вычета потока в выбранном пути для второго шага показана на рис. 22.

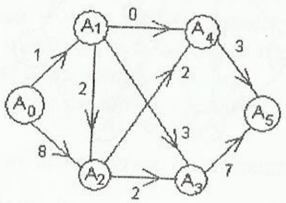

Рис. 22. Новая сеть 2

2. Выберем путь $A_0A_1A_2A_4A_5$, минимальная пропускная способность которого равна 1. Построенные потоки показаны на рис. 23.

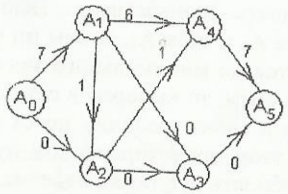

Рис. 23. Построенный поток в сети 3

Новая исходная сеть после вычета потока в выбранном пути для третьего шага показана на рис. 24.

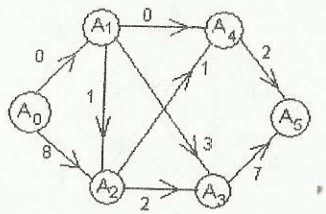

Рис. 24. Новая сеть 4

3. Выберем путь $A_0A_2A_3A_5$, min пропускная способность которого равна 2. Построенный на этом шаге поток в сети показан на рис. 25.

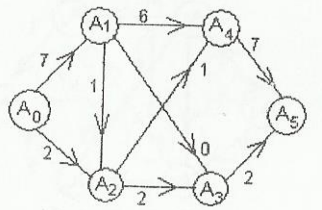

Рис. 25. Построенный поток в сети 5

Новая исходная сеть после вычета потока в выбранном пути для четвертого шага показана на рис. 26.

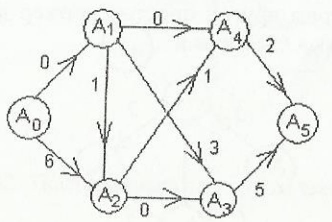

Рис. 26. Новая сеть 6

4. Выберем путь $A_0A_2A_4A_5$, минимальная пропускная способность которого равна 1. Построенный на этом шаге поток в сети показан на рис. 27.

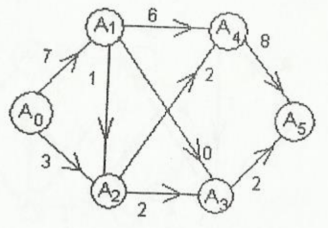

Рис. 27. Построенный поток 7

Новая исходная сеть после вычета потока в выбранном пути для пятого шага показана на рис. 28.

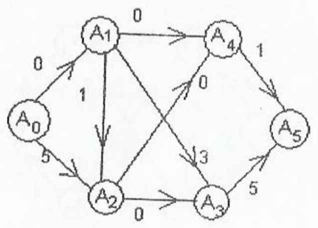

Рис. 28. Новая сеть 8

Таким образом, в сети на рис. 28 нет путей с ненулевой пропускной способностью и построенный поток полный: v=10.

Остается невыясненным вопрос о том, является ли этот поток максимальным из всех возможных.

Пример. Рассмотрим другой простой пример для сети, показанной на рис. 29, и решим его двумя способами.

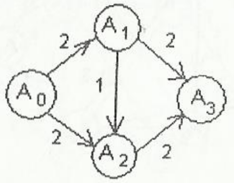

Рис. 29. Простая сеть

1-й способ: путь $A_0A_1A_2A_3$, показанный на рис. 30, имеет минимальную пропускную способность, равную 1.

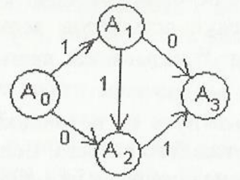

Рис. 30. Построенный на первом шаге поток

Для следующего шага исходная сеть имеет вид, показанный на рис. 31.

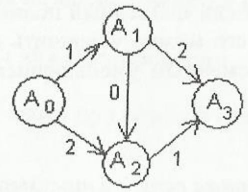

Рис. 31. Новая сеть

Выберем путь $A_0A_2A_3$, минимальная пропускная способность которого равна 1.

Полный поток в этом случае v = 2 (рис. 32).

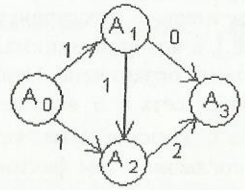

Рис. 32. Построенный на втором шаге поток

2-й способ: путь $A_0A_1A_2$ имеет min пропускную способность = 2. Путь $A_0A_2A_3$ имеет min пропускную способность = 2. Полный поток в этом случае v=4.

Как проверить, является ли построенный полный поток максимальным? Для этого имеются критерии, в основе которых лежит понятие *разреза* или *разделяющего сечения* сети. Обозначим через X некоторое множество узлов сети, соединенных с источником A_0, а $\overline{X}$ – остальные узлы сети, которые соединены с A_n. Совокупность весов всех дуг $\{X, \overline{X}\}$, начало которых принадлежит X, а конец $\overline{X}$, называется пропускной способностью разреза $c(X, \overline{X})$.

Заметим, что каждый путь из источника A_0 к стоку A_n содержит по меньшей мере одну дугу каждого разреза. Если из сети исключить все дуги какого-нибудь разреза, то не останется ни одного пути, связывающего источник A_0 со стоком A_n. Поэтому разрез сети нередко называют *разделяющим сечением*. Отсюда следует, что величина v произвольного потока φ не превышает пропускной способности $c(X, \overline{X})$ произвольного разреза. Более того, можно доказать теорему: для любой транспортной сети величина максимального потока из источника в сток совпадает с минимальной пропускной способностью разреза [4, 7]. Если найденный полный поток не совпадает с максимально возможным, то его можно увеличить для прямых дуг оставшегося непрямого пути с одновременным уменьшением на ту же величину потока для обратной дуги.

8.6. Реализация сетей в трехмерном пространстве

До сих пор мы ограничивались математическими графами и проблемы конкретной их реализации не рассматривались. Однако любая реальная физическая сеть состоит из элементов и соединяющих дуг, имеющих определенные размеры. Это накладывает определенные ограничения на возможности размещения сети в трехмерном пространстве. Соответствующая математическая задача была решена в работе Колмогорова А.Н., Бардзиня Я.М. «О реализации сетей в трехмерном пространстве» [Проблемы кибернетики. 1967. Вып.19. С.261-268.], в которой минимальное расстояние между дугами сети, а также между узлами ограничено. Ответ на поставленный вопрос дает следующая теорема: любая сеть с n вершинами может быть реализована в сфере радиуса $C\sqrt{n}$, где C константа, не зависящая от n. Эта теорема находится в удивительном согласии с тем фактом, что мозг устроен так, что основную его массу занимают волокна, а нейроны расположены лишь на его поверхности. Данная теорема подтверждает оптимальность в смысле объема такого строения нейронной сети. Ее, безусловно, необходимо учитывать и при попытках построения оптимальных по объему больших искусственных нейронных сетей.

8.7. Феномен «тесного мира»

Социолог Гарвардского университета С. Милграм обнаружил, что каждого человека на земном шаре можно связать с другим человеком цепочкой из шести знакомых [см. Евин И.А. Синергетика мозга. М.- Ижевск: НИЦ «Регулярная и хаотическая динамика», 2005. 198 с.]. Следовательно, несмотря на то, что на Земле живет более шести миллиардов людей, мир тесен (“small world”). Подобным свойством обладают и многие технические сети с высокой степенью кластеризации и малой средней длиной между узлами. Таким образом, к выводам, следующим из теоремы Колмогорова о наиболее плотной упаковке сетей, следует добавить, что если локально отказаться от требования оптимальности размещения в пространстве, а заменить это требованием быстрого установления связей между любыми узлами через малое число посредников, то наиболее эффективными становятся кластерные структуры. Итогом может служить вывод о максимальной эффективности кластерных структур, расположенных на вдоль замкнутой поверхности.

Контрольные вопросы для самопроверки

1. Какие задачи решаются при помощи графов.
2. Что такое сеть.

Глава 9. Принятие решений при неопределенности целей

9.1. Противоречивость целей

Назначение цели в задачах исследования операций и формализация цели, то есть выбор целевой функции, почти всегда является трудной проблемой. Часто бывает необходимо выбрать стратегию, которая делает максимальным доход при минимальных затратах. Таким образом, мы имеем одновременно две задачи:

$$\begin{aligned} &f(x) \to \max, \\ &-F(x) \to \max(F(x) \to \min), \end{aligned} \qquad (78)$$

где $f(x)$ и $F(x)$ – функции, характеризующие соответственно доход и затраты.

Такая задача, как правило, решений не имеет. В принципе для увеличения дохода необходимо увеличивать затраты. Следовательно, поставленные цели противоречат друг другу. Для того чтобы свести данную задачу исследования операций к стандартной задаче оптимизации, необходимо сформулировать дополнительные гипотезы, не вытекающие прямо из постановки задачи.

Остановимся на некоторых наиболее употребительных способах преодоления неопределенности целей, когда необходимо обеспечить максимальное значение функциям $f_1(x), f_2(x), \ldots, f_n(x)$ одновременно.

9.2. Линейная свертка

Вместо n отдельных критериев часто рассматривают один критерий вида

$$F(x) = \sum_{i=1}^{n} c_i f_i(x), \tag{79}$$

где c_i – некоторые положительные числа, обычно нормированные условием $\sum_{i=1}^{n} c_i = 1$.

Коэффициенты c_i являются результатами экспертизы. Они отражают представление оперирующей стороны о содержании компромисса, который она вынуждена принять. Таким образом, происходит ранжирование целей на основе назначения весовых коэффициентов.

9.3. Использование контрольных показателей

Часто в задачах планирования и проектирования задается некоторая система нормативов: $F_1, F_2, \ldots, F_n$. Это значит, например, что функции $f_i(x)$ должны достигать максимальных значений при условиях

$$f_i(x) \geq F_i, \quad i = 1, 2, \ldots n. \tag{80}$$

В таких случаях целевую функцию можно задать в виде

$$F(x) = \min \frac{f_i(x)}{F_i} \tag{81}$$

и искать вектор x, который обеспечивает максимальное значение $F(x)$. Смысл процедуры прост. Критерий (81) дает значение наихудшего показателя, а условие $F(x) \to \max$ делает его максимально хорошим.

Если на компоненты вектора x положены линейные ограничения вида

$$\sum_i a_{ij} x_j \leq b_i, \tag{82}$$

а функции $f_i(x)$ также линейные функции

$$f_i(x) = \sum_j d_{ij} x_j, \tag{83}$$

то задача сводится к линейному программированию.

9.4. Простейший способ преодоления неопределенности целей

Если система контрольных показателей имеет вид (80), то в ряде случаев выбирают основной критерий, например $f_1(x)$. Тогда снова получается однокритериальная задача:

$$f(x) \to \max \tag{84}$$

при условиях (80).

9.5. Метрика в пространстве целевых функций

Предположим, что решена система однокритериальных задач

$$f_i(x) \to \max, \quad i = 1, 2, \ldots n,$$

и нашли в i-й задаче вектор $x = x_i$, доставляющий максимальное значение критерию $f_i(x)$. Совокупность скалярных величин $f_i(x_i)$ определяют в пространстве критериев некоторую точку, которую назовем абсолютным максимумом.

Если векторы x_i различны, то такой абсолютный максимум недостижим. Однако можно поставить задачу о максимальном приближении к нему. В качестве меры близости введем расстояние в пространстве критериев, например, в виде евклидова расстояния:

$$h = \sqrt{\sum_i (f_i(x) - f_i(x_i))^2}. \tag{85}$$

Тогда h есть расстояние от точки $(f_1(x), f_2(x), \ldots, f_n(x))$ до точки $(f_1(x_1), f_2(x_2), \ldots, f_n(x_n))$ в пространстве критериев.

В качестве нового скалярного критерия можно теперь принять функцию (85). Ее минимизация позволяет ближе всего подойти к абсолютному максимуму.

9.6. Компромиссы Парето

Проанализировать решения многокритериальных задач можно и с несколько иных позиций, попытавшись исключить заведомо плохие решения. Один из путей решения подобных задач предложен итальянским экономистом В. Парето в 1904 г.

Пусть x_0 соответствует некоторому выбору и имеется лучший выбор x.

Тогда

$$f_i(x) \geq f_i(x_0),\ i = 1,2,\ldots,n. \tag{86}$$

Ясно, что выбор x предпочтительнее x_0. Множество всех таких x называют *множеством Парето*. Вектор x называют *наилучшим вектором результатов* или *вектором Парето*, если из $f_i(x) \geq f_i(x_0)$ для любого i следует $f_i(x) = f_i(x_0)$.

Предположим, что цели определяются двумя однозначными функциями:

$$f_1(x) \to \max,$$
$$f_2(x) \to \max. \tag{87}$$

Тогда каждому допустимому значению x отвечает одна точка на плоскости (f_1, f_2), (рис. 33).

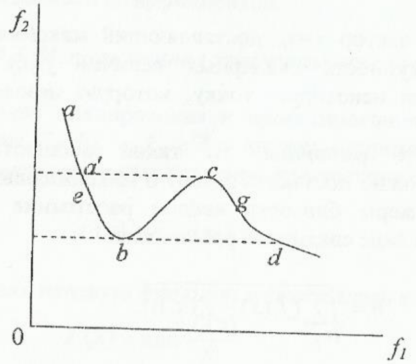

Рис. 33. Множество Парето для случая двух критериев

Равенства $f_1 = f(x)$, $f_2 = f(x)$ параметрически задают некоторую кривую $abcd$ в этой плоскости. Участок bc не принадлежит множеству Парето, поскольку на этом участке одновременно растут обе величины f_1 и f_2. На этом основании исключается и участок $a'b$, поскольку для любой точки e из участка $a'b$ найдется точка g участка cd, в которой обе функции больше, чем в точке e. Значит, множеству Парето могут принадлежать только участки aa' и cd.

В теории принятия решений существует *принцип Парето*, согласно которому выбирать в качестве решения можно только векторы x из множества Парето. Принцип Парето не определяет решение, но сужает множество возможных решений. Построение множеств Парето облегчает поиск компромисса в многокритериальных задачах.

Контрольные вопросы для самопроверки

1. Чем характеризуется ситуация неопределенности целей?
2. Какие методы принятия решений в условиях неопределенности целей Вы знаете?

Глава 10. Процессы массового обслуживания

10.1. Потоки в системах обслуживания

Существует большое число процессов, для которых характерно поступление объектов во *входящем потоке*. Эти объекты подвергаются в системе определенным операциям (*обслуживанию*) и затем покидают систему в *выходящем потоке*. Промежутки времени, через которые поступают объекты, и время обслуживания, как правило, носят случайный характер. При массовом обслуживании объектов в системе обслуживания могут возникать *очереди*.

Процессы массового обслуживания типичны для связи (телефон, телеграф, почта, компьютерные сети), транспорта (воздушные, наземные и морские перевозки), производственных процессов (ремонт и обслуживание оборудования) и т.п. Составными элементами процесса массового обслуживания являются:

1) входящий поток;

2) очередь;

3) система пунктов обслуживания;

4) выходящий поток.

Независимо от конкретной природы и характера объектов, поступающих в систему обслуживания, их называют *требованиями* или *заявками*.

Выходящий поток называют *простейшим*, если вероятность поступления того или иного числа требований в течение интервала t зависит только от протяженности этого интервала и не зависит от его расположения на оси времени. Считается, что требования поступают поодиночке и независимо друг от друга. Плотность распределения числа требований за время t для простейшего потока описывается распределением Пуассона:

$$P_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}, \tag{88}$$

где n число поступивших событий за время t.

Среднее число требований, поступающих за время t, равно

$$M(n) = \sum_{n=0}^{\infty} n \frac{(\lambda t)^n}{n!} e^{-\lambda t} = \lambda t \sum_{m=0}^{\infty} \frac{(\lambda t)^m}{m!} e^{-\lambda t} = \lambda t\,. \tag{89}$$

Из формулы (89) ясно, что λ представляет собой среднее число требований в единицу времени и его называют *интенсивностью* (*плотностью*) потока. Вероятность того, что в интервале времени t не поступит ни одного требования, равна $P_0(t) = e^{-\lambda t}$. Соответственно вероятность того, что интервал между двумя последовательными требованиями превысит некоторую величину δ, равна вероятности отсутствия требований в этом интервале, то есть также $e^{-\lambda t}$.

Пример. Пусть в бюро по обслуживанию поступает в среднем 12 заказов в час. Найдем вероятность того, что за 1 минуту в бюро не поступит ни одного заказа, а также вероятность поступления не более 3-х заказов за 10 минут.

Решение. Плотность потока заказов составит $\lambda = 12/60 = 0{,}2$ заказа в минуту. Тогда вероятность отсутствия заказов за минуту $P_0(1) = e^{-0{,}2} = 0{,}819$, а вероятность поступления не более трех независимых заказов за 10 минут составит

$$P = \sum_{k=0}^{3} P_k(10) = P_0(10) + P_1(10) + P_2(10) + P_3(10)\,.$$

Производительность системы обслуживания зависит от числа каналов и их быстродействия. Время обслуживания одного требования чаще всего считают случайной величиной, распределенной по экспоненциальному закону.

10.2. Марковские процессы обслуживания

Для процессов массового обслуживания с пуассоновским потоком требований характерно отсутствие последействия. Иначе говоря, будущее развитие зависит только от состояния в настоящий момент времени и не зависит от того, как происходило развитие в прошлом. Такие процессы называют *марковскими*.

Возможные состояния системы массового обслуживания будут следующие: S_0 – все каналы свободны; S_i - занято i каналов, где $i \leq m$ – полное число каналов (пунктов обслуживания); S_{m+r} – заняты все m каналов и r требований в очереди. Обозначим через $p_i(t)$ вероятность того, что в

момент t система находится в состоянии S_i (i=0,1,2,...,n). Для любого момента времени t сумма вероятностей нахождения системы во всех возможных состояниях равна единице, то есть

$$\sum_{i=0}^{n} p_i(t) = 1\,. \tag{90}$$

Марковский процесс описывается относительно вероятностей $p_0(t)$, $p_1(t)$, ..., $p_n(t)$ системой дифференциальных *уравнений Колмогорова.* При составлении этих уравнений удобно воспользоваться *графом состояний*, вершины которого соответствуют состояниям, а дуги - возможным переходам из состояния в состояние. Такой граф показан на рис. 34.

Рис. 34. Граф состояний системы массового обслуживания

Вероятность перехода системы из состояния S_i в состояние S_j обозначим λ_{ij}. Скорость изменения вероятности нахождения системы в k – ом состоянии в момент времени t можно записать в виде уравнения Колмогорова:

$$\frac{dp_k}{dt} = \lambda_{k-1} p_{k-1} - (\lambda_{k,k-1} + \lambda_{k,k+1}) p_k + \lambda_{k+1} p_{k+1}\,. \tag{91}$$

Смысл отдельных слагаемых в правой части уравнения (91) достаточно прозрачен. Первое и последнее слагаемые отвечают переходу системы из соседних состояний. Отрицательный член соответствует процессу обратного перехода. Система уравнений вида (91) может быть записана в матричной форме:

$$\frac{dp}{dt} = \Lambda p, \tag{92}$$

где $p=(p_0, p_1, p_2, ..., p_n)$ – вектор вероятности состояний и Λ - матрица скоростей переходов (плотности вероятности).

Составим уравнения Колмогорова для системы массового обслуживания с ожиданием. Для этого определим плотности вероятностей переходов и разметим граф этой системы (рис. 35), где m – число каналов обслуживания, r – длина очереди, μ - интенсивность обслуживания, ν - интенсивность ухода из очереди.

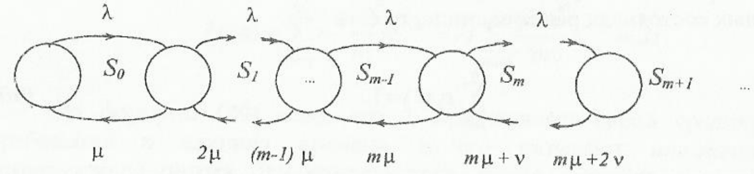

Рис. 35. Размеченный граф системы массового обслуживания

Пусть вероятности перехода $P_{i,i+1}$ из состояния i в состояние $i+1$ зависят исключительно от потока требований. При этом каждое новое требование либо поступает в канал обслуживания, либо получает отказ. Поэтому можно положить для простейшего входящего потока $\lambda_{i,\ i+1} = \lambda$. Переход из состояния $S_i = S_{m+r}$ в состояние $S_{i-1} = S_{m+r-1}$ $(r>1)$ определяется плотностью вероятности перехода, равной сумме вероятностей освобождения каналов и отказа от обслуживания. При отказе от обслуживания происходит уход из очереди очередного требования. Таким образом, плотность вероятности перехода в единицу времени из состояния S_{m+r} в S_{m+r-1} $(r>1)$ равна сумме интенсивностей освобождения каналов и отказа от обслуживания:

$$\lambda_{m+r,m+r-1} = \lambda_{i,i-1} = m\mu + r\nu\,. \qquad (93)$$

После разметки графа, как это показано на рис. 35, записываем систему дифференциальных уравнений:

$$\frac{dp_0}{dt} = -\lambda p_0 + \mu p_1;$$

$$\frac{dp_i}{dt} = \lambda p_{i-1} - (\lambda + i\mu) p_i + (i+1)\mu p_{i+1}$$

$$(1 \le i \le m-1); \qquad (94)$$

$$\frac{dp_{m+r}}{dt} = \lambda p_{m+r-1} - (\lambda + m\mu + r\nu) p_{m+r} + [m\mu + (r+1)\nu] p_{m+r+1} \quad (r \ge 0).$$

Если система массового обслуживания в начальный момент находилась в состоянии S_i, то начальными условиями являются соотношения $p_i=1$; $p_j=0$ $(i=1,2,...,m+r,...;\ j \ne i)$. Полученная система содержит неограниченное число уравнений. Она становится конечной, если накладываются ограничения на длину очереди, т.е. на величину r.

10.3. Стационарный режим обслуживания

При прошествии достаточно большого времени в системе часто устанавливается *стационарный режим,* когда $dp/dt=0$ и система дифференциальных уравнений (94) превращается в систему линейных алгебраических уравнений, то есть

$$-\lambda p_0 + \mu p_1 = 0;$$

$$\lambda p_{i-1} - (\lambda + i\mu) p_i + (i+1)\mu p_{i+1} = 0, \ (1 \le i \le m-1); \qquad (95)$$

$$\lambda p_{m+r-1} - (\lambda + m\mu + r\nu) p_{m+r} + [m\mu + (r+1)\nu] p_{m+r+1} = 0, (r \ge 0).$$

Система уравнений (95) дополняется нормировочным условием (90), где $p_i(t)=p_i=const$. Последовательное исключение неизвестных дает $p_1 = \frac{\lambda}{\mu} p_0 = \alpha p_0$, где $\alpha = \frac{\lambda}{\mu}$ – *приведенная плотность потока требований* (среднее число требований, поступающих за среднее время обслуживания одного требования),

$$p_i = \frac{\alpha^i}{i!} p_0 \ (1 \le i \le m). \qquad (96)$$

При $i>m$ тем же способом находим

$$p = \frac{\alpha^m}{m!} \times \frac{\alpha^r}{\prod_{i=1}^{r}(m + j\beta)} p_0 = \frac{\alpha^r}{\prod_{j=1}^{r}(m + j\beta)} p_m \ (r \ge 1), \qquad (97)$$

где $\beta = \frac{\nu}{\mu}$ – *приведенная плотность потока ухода из очереди* без обслуживания.

В соответствии с нормировочным условием имеем

$$\sum_{i=0}^{m} \frac{\alpha^i}{i!} p_0 + \frac{\alpha^m}{m!} \sum_{r=1}^{\infty} \frac{\alpha^r}{\prod_{i=1}^{r}(m + j\beta)} p_0 = 1. \qquad (98)$$

Отсюда легко определяется p_0. Средняя длина очереди $\bar{r}$ определяется как математическое ожидание числа находящихся в очереди требований:

$$\bar{r} = \sum_{r=1}^{\infty} r p_{m+r} = \frac{\alpha^m}{m!} \sum_{r=1}^{\infty} \frac{r\alpha^r}{\prod_{j=1}^{r}(m+j\beta)} p_0, \quad (r \geq 1). \tag{99}$$

Контрольные вопросы для самопроверки

1. Дайте определения основных понятий теории массового обслуживания.

Глава 11. Динамическое программирование

11.1. Принцип оптимальности

В целом ряде случаев операции разбиваются на ряд этапов. На каждом этапе имеется управление операцией, от результатов которой зависит как выполнение данной операции, так и операции в целом. *Динамическое программирование* представляет собой метод оптимизации процесса поэтапного управления.

Оптимизация многошагового процесса управления осуществляется на основе *принципа оптимальности*. Он состоит в том, что планирование проводится так, чтобы учитывался результат не только на данном этапе, но и общий результат в конце многошаговой операции.

Управление u всей операцией представляет собой совокупность шаговых управлений $u=(u_1,u_2, \ldots u_m)$, где m – число шагов. Эффективность операции (доход) F зависит от выбранной системы управлений, то есть является функцией от u:

$$F=F(u). \tag{100}$$

Задача состоит в таком выборе управления, чтобы F была максимальной. Управление u^*, при котором $F=F_{max}$, называется оптимальным. Идея решения такой задачи состоит в постепенной оптимизации. На каждом шаге оптимизация должна проводиться с учетом влияния всех последующих шагов на конечный результат. Исключение составляет последний шаг. Он может быть спланирован оптимально, независимо от предыдущих шагов. Затем можно оптимизировать предпоследний шаг и т.д.

Процесс *динамического программирования* выстраивается от конца к началу. На каждом шаге ищется оптимальное продолжение процесса относительно достигнутого в данный момент состояния системы, и такое управление называется условно оптимальным. При оптимизации управления многошаговый процесс проходится дважды:

1. От конца к началу, в результате чего находят условно-оптимальные управления на каждом шаге и условный оптимальный выигрыш на всех шагах.

2. От начала к концу, в результате чего определяют оптимальные шаговые управления.

11.2. Задача о распределении ресурсов

Пусть имеется начальное количество средств k_0, которое нужно распределять в течение m лет между производствами I и II. Если средства X вложить в первое производство, то за год будет получен доход $f(X)$, средства уменьшатся и к концу года от них останется некоторая часть $\varphi(X)<X$. Аналогично средства Y, вложенные в отрасль II, принесут за год прибыль $g(Y)$ и уменьшатся до величины $\psi(Y)<Y$. На следующий год вновь осуществляется вложение уже оставшихся средств без поступлений извне и без использования прибыли. Требуется найти способ управления ресурсами, при котором суммарная прибыль обоих производств за m лет будет максимальной.

Задача может быть решена методом динамического программирования. Естественным шагом процесса является в данном случае год. Управление u_i на i-м шаге состоит в выделении средств X_i и Y_i, вкладываемых в производства I и II. Управление операцией состоит из совокупности шаговых управлений за m лет.

Состояние системы перед i-м шагом характеризуется одним параметром k_{i-1} – количеством оставшихся к этому времени средств. Управление u_i на i-м шаге заключается в том, что первому производству выделяются средства X_i. Тогда второму производству остается величина $Y_i=k_{i-1}-X_i$. Прибыль на i-м шаге $W_i=f(X_i)+g(k_{i-1}-X_i)$.

Обозначим через $W_j(k)$ максимальную прибыль, которую можно получить начиная с j-го года по последний год включительно, при условии, что к началу j-го периода осталось k средств. Если в начале i-го периода выделено X_i средств в первое производство, то максимальная прибыль, которая может быть получена начиная с i-го года, будет равна сумме прибыли за i-й год и максимальной прибыли за оставшиеся годы, т.е.

$$W_i(k) = f(X_i)+g(k_{i-1}-X_i)+W_{i+1}(\varphi(X_i)+\psi(k_{i-1}-X_i)). \tag{101}$$

Максимальная прибыль за годы, начиная с i-го, будет получена, если выбрать X_i так, чтобы функция (101) принимала максимальное значение. Тем самым получаем связь между максимальной прибылью, после с $(i-1)$ –го года $W_i(k)$, и максимальной прибылью, которую можно получить после i-го года в виде функционального уравнения

$$W_i(k)=\mathbf{max}[f(X_i)+g(k-X_i)+W_{i+1}(\varphi(X_i)+\psi(k-X_i))], \quad 0 \le X_i \le k. \tag{102}$$

Это так называемое *основное функциональное уравнение динамического программирования*. Поскольку величина k_i заранее неизвестна, в уравнении она заменена k. Значение $W_i(k)$ нужно находить при всех k. Далее нужно найти значение $X_i(k)$ и соответствующее ему значение $Y_i(k)=k-X_i(k)$, при котором достигается оптимальное уравнение на i-м шаге (при условии $k_i=k$). Величина $W_i(k)$ является условно-оптимальной прибылью.

Условно-оптимальная прибыль на последнем шаге

$$W_m(k)=\mathbf{max}\{f(X_m)+g(k-X_m)\}, \quad 0 \le X_m \le k. \tag{103}$$

Ей соответствует условно-оптимальное управление $X_m(k)$, на котором достигается максимум. Для построения оптимального управления на предыдущих этапах воспользуемся уравнением (102), где W_{i+1} теперь имеет условно определенное значение. В итоге получаем условно-оптимальные управления $X_{m-1}(k)$, $X_{m-2}(k)$, $\dots X_1(k)$. Чаще всего функции $W_i(k)$ и $X_i(k)$ находятся в табличном виде.

Теперь можно по известному начальному запасу средств k_0 определить максимальную прибыль за весь период: $W_{max}=W_1(k_0)$. Затем можно найти оптимальное управление на первом шаге: $X_1^*=X_1(k)$ и $Y_1^*=k_0-X_1^*(k_0)$. По окончании первого этапа останутся средства в размере $k_1^*=\varphi(X_1^*)+\psi(Y_1^*)$. Находим затем оптимальное управление на втором шаге: $X_2^*=X_2(k_1^*)$, $Y_2^*=k_1^*-X_2(k_1^*)$, и остаток средств $k_2^*=\varphi(X_2^*)+\psi(Y_2^*)$ к концу второго года. Продолжая этот процесс, получим последовательность значений

$$k_0 \to X_1(k_0) \to k_1^* \to X_2(k_1^*) \to k_2^* \to \dots \to k_{m-1} \to x_m(k_{m-1}^*) \to k_m^*. \tag{104}$$

Величина k_m^* дает количество средств, оставшихся к концу планируемого периода.

В первое производство вкладывается по годам X_1^*, X_2^*, …, X_m^* средств, во второе производство - $Y_1^*=k_0-X_1^*$, $Y_2^*=k_1-X_2^*$,…, $Y_m^*=k_{m-1}^*-X_m^*$ средств.

Контрольные вопросы для самопроверки

1. Какие виды задач решаются методом динамического программирования?
2. Дайте характеристику принципа оптимальности.

Глава 12. Элементы теории игр

12.1. Конфликты как игры

В природе и обществе часто возникают конфликтные ситуации, в которых участвуют стороны с различными и даже с противоположными интересами. Конфликтные ситуации возникают в самых разных областях деятельности: при операциях купли-продажи, в большом и малом бизнесе, в спорте и т.д.

Математическая теория конфликтных ситуаций называется *теорией игр*. Ее задачей является выработка рекомендаций поведения, которые приводили бы к наибольшей выгоде той или иной стороны. Игры различаются по числу сторон на *парные* и *множественные*.

Мы рассмотрим только парные игры. Стороны назовем игроками A и B. Для описания игры необходимо сформулировать ее правила, объем информации каждой стороны о поведении другой и результат игры, к которому приводит каждая последовательность ходов. Парная игра называется *игрой с нулевой суммой*, при которой выигрыш одного игрока равен проигрышу другого. Мы будем рассматривать только такие игры.

Стратегией игрока называется совокупность правил, определяющих выбор его действий при каждом ходе в зависимости от сложившейся ситуации. Задача теории игр заключается в выборе стратегии, приводящей к наибольшему выигрышу в предположении, что второй игрок также придерживается наилучшей стратегии.

Пусть игрок А имеет возможность применять m стратегией A_1, A_2 … , A_m, а игрок B – n стратегий B_1, B_2,…,B_n. Обозначим через a_{ij} выигрыш игрока A, если он выберет стратегию A_i, а игрок B – стратегию B_j. В этом случае игра полностью описывается матрицей a_{ij}, которая называется *платежной матрицей* или *матрицей игры*. Выигрыш игрока A означает проигрыш игрока B, для которого платежная матрица $b_{ij}=-a_{ij}$. Предположим, что игрок A выбрал стратегию A_i. Естественно, что игрок B выберет стратегию, при которой выигрыш α_i игрока A будет минимальным, то есть $\alpha_i = \min\limits_i a_{ij}$. В свою очередь игрок A предпочтет стратегию A_k, при которой выигрыш будет наибольшим и $\alpha_k = \max\limits_j a_{ij} = \max\limits_i \min\limits_j a_{ij}$. Стратегия A_k игрока A называется *максиминной.* Она гарантирует игроку A выигрыш α_k. Как бы ни играл игрок B, выигрыш игрока A не будет меньше этого числа. Величина α_k называется *нижней ценой игры*.

Рассмотрим теперь поведение игрока B. Если он выберет стратегию B_j, то A примет стратегию, приводящую к наибольшему выигрышу $\beta_j = \max\limits_i a_{ij}$. Тогда для B следует выбрать такую стратегию, при которой выигрыш A будет наименьшим, то есть $\beta_l = \min\limits_j \max\limits_i a_{ij}$ Стратегию B_l называют *минимаксной* , и она гарантирует игроку B, что его проигрыш будет не более величины β_l, как бы ни играл игрок A. Величина β_l называется *верхней ценой игры.*

12.2. Основное неравенство и игра с седловой точкой

По определению $\alpha_k = \min\limits_j a_{kj}$, то есть α_k не больше любого элемента из k-й строки. Поэтому $\alpha_k \le \alpha_{kl} \le \max a_{il} = \beta_l$. Отсюда следует $\alpha_k \le \beta_l$ – основное неравенство. Из основного неравенства вытекает ряд важных следствий.

1. Нижняя цена игры не превосходит верхней цены игры.
2. Если игрок A придерживается максиминной стратегии, а игрок B - минимаксной стратегии, то выигрыш a_{kl} заключен между нижней и верхней ценами игры $\alpha_k \le a_{kl} \le \beta_l$.
3. Если нижняя и верхняя цена игры совпадают ($\alpha_k = \beta_l$), то основное неравенство превращается в равенство и элемент α_{kl} является минимальным в строке и максимальным в столбце. При этом по аналогии с непрерывными функциями говорят, что игра имеет *седловую точку*. Справедливо и *обратное утверждение*: если игра имеет седловую точку, то нижняя и верхняя цены совпадают.

Если игра имеет седловую точку и игрок A придерживается максиминной стратегии, то игроку B невыгодно отклоняться от минимаксной стратегии и обратно, если B придерживается минимаксной стратегии, то A невыгодно отклоняться от максиминной стратегии. Однако имеются игры без седловых точек, для которых сформулированные утверждения неверны.

12.3. Игры с вероятностным выбором стратегии

Часто игра может происходить многократно, и игроки в процессе игры могут менять стратегии. Предположим, что игрок A применяет стратегии A_i с заданными вероятностями p_i. Поскольку стратегии A_i представляют собой полную систему событий, то

$$\sum_{i=1}^{m} p_i = 1. \tag{105}$$

В свою очередь игрок B применяет стратегии B_j с частотами (вероятностями) q_j и

$$\sum_{j=1}^{n} q_j = 1. \tag{106}$$

Смешанные стратегии представляют собой математическую модель изменчивой и гибкой тактики игрока, при которой противостоящий ему игрок не может узнать заранее то положение, с которым ему придется встретиться в игре, ибо перед каждой партией производится случайный выбор одной из чистых стратегий с помощью некоторого механизма, производящего этот выбор с определенными и заранее заданными вероятностями.

Наборы чисел $(p_1,\dots,p_m)$ и $(q_1,\dots,q_n)$ называются *смешанными стратегиями.* Если некоторое p_i=1, а все остальные вероятности нулевые, то это означает, что игрок A постоянно придерживается одной стратегии A_i. Такая стратегия называется *чистой*. Аналогично определяется чистая стратегия B игрока B_j, когда q_j=1.

Решением игры в смешанных стратегиях называется такая пара стратегий игроков A и B, называемых оптимальными, что если один из них придерживается своей оптимальной стратегии, то другому также невыгодно отклоняться от этой стратегии.

Выигрышем игрока A естественно назвать среднее значение выигрыша на множестве реализуемых стратегий. Если игроки A и B играют независимо, то вероятность того, что игрок A примет стратегию A_i, а игрок B – стратегию B_j есть произведение вероятностей этих независимых событий p_i, q_j. Поэтому среднее значение v выигрыша игрока A равно математическому ожиданию

$$v = \sum_{i,j} a_{ij} p_i q_j . \tag{107}$$

Если используются оптимальные смешанные стратегии, то соответствующее значение v называется *ценой игры.*

Существует связь между теорией игр и линейным программированием, методы которого можно использовать для анализа игр в смешанных стратегиях. Пусть все числа a_{ij} положительные. Если это не так, то к элементам матрицы a_{ij} можно прибавить такое положительное число c, что все элементы матрицы станут положительными. Зафиксируем некоторую смешанную стратегию $(p_1, p_2, ..., p_m)$ игрока A. Число ζ назовем *гарантированным выигрышем игрока А* при выбранной стратегии, если выигрыш A не будет меньше ζ при любой стратегии игрока B. Пусть B выбирает чистую стратегию (1,0,…,0). Тогда по формуле выигрыш

$$v = a_{11}p_1 + a_{21}p_2 + ... + a_{m1}p_m .$$

По свойствам величины ζ

$$a_{11}p_1 + a_{21}p_2 + ... + a_{m1}p_m \geq \zeta .$$

Проверив все чистые стратегии игрока B, приходим к системе неравенств

$$a_{1j}p_1 + a_{2j}p_2 + ... + a_{mj}p_m \geq \zeta , \tag{108}$$

Если неравенства (108) выполнены, то при любой смешанной стратегии $(q_1, q_2, ..., q_n)$ игрока B выигрыш

$$v = \sum_{i,j} a_{ij} p_i q_j = \sum_j \left(\sum_i a_{ij} p_j \right) q_j \geq \sum_j \zeta q_j = \zeta \sum_j q_j = \zeta . \tag{109}$$

То есть $v > \zeta$ и величина ζ является гарантированным выигрышем. Выполнение неравенств (108) является необходимым и достаточным условием того, чтобы величина ζ была гарантированным выигрышем.

Естественно в качестве оптимальной стратегии A принять ту, которая дает максимальный гарантированный выигрыш ζ. Введем обозначения $x_i = p_i / \zeta$, тогда неравенства (108) примут вид

$$\sum a_{ij} x_j \geq 1 . \tag{110}$$

Равенство (105) переходит в

$$\sum_i x_i = 1/\zeta = F . \tag{111}$$

Таким образом, задача заключается в минимизации линейной формы (111) при ограничениях (110). Тем самым она приняла стандартный вид для линейного программирования. Подобным же образом может быть поставлена и решена задача о наименьшем неизбежном проигрыше η игрока B. Для

оптимальных стратегий $\zeta = \eta$. Тем самым мы приходим к основной *теореме Неймана теории игр*: всякая конечная игра имеет решение.

12.4. Выбор стратегии

Бывают конфликтные ситуации, в которых одна из сторон действует неопределенно. Игру такого типа называют игрой с природой. Пусть игрок B теперь является природой. Если игрок А выбирает чистую стратегию A_0, то математическое ожидание выигрыша равно $\sum a_{ij} q_j$. Поэтому наиболее выгодной будет та стратегия, при которой достигается

$$\max \sum_j a_{ij} q_j . \tag{112}$$

Если информация о состояниях природы мала, то разумно считать, что все состояния природы равновероятны. Тогда игрок А выбирает ту стратегию, при которой достигается

$$\max \frac{1}{n} \sum_{j=1}^{n} a_{ij} . \tag{113}$$

Критерий (113) выбирает стратегию, для которой среднее арифметическое элементов соответствующей строки максимальное.

Существуют и другие виды критериев для выбора наилучшей стратегии. По *критерию Вальда* выбирается стратегия, обеспечивающая $\max_i \left(\min_j a_{ij} \right)$. Этот критерий предполагает, что природа будет действовать наихудшим для A способом.

Критерий Гурвица позволяет смягчить требования критерия Вальда и записать их в виде

$$\max_i \left[\lambda \min_j a_{ij} + (1 - x) \max_j a_{ij} \right], \tag{114}$$

где $0 \leq \lambda \leq 1$.

Критерий Сэвиджа основан на минимальности *рисков* $r_{ij} = \max_i a_{ij} - a_{ij}$ (разницей между наибольшим выигрышем игрока A и самим выигрышем). Критерий запишем в виде

$$r = \min(\max_i a_{ij} - a_{ij}) . \tag{115}$$

Пример. Пусть отношения заказчика строительства объекта и субподрядчика, выполняющего строительные работы на данном объекте, описываются матрицей игры

$$\begin{array}{c} \\ A_1 \\ A_2 \\ A_3 \end{array} \begin{array}{c} \begin{array}{ccc} B_1 & B_2 & B_3 \end{array} \\ \begin{bmatrix} 3 & 1 & 4 \\ 5 & 7 & 2 \\ 7 & 9 & 3 \end{bmatrix} \end{array}$$

Заданы вероятности стратегий игрока $A(p_1, \ p_2, \ p_3)$=(0;0,6;0,4) и $B(q_1,q_2,q_3)$=(0,3;0,2;0,5).

Определить:

1. Среднее значение выигрышей игрока A.
2. Величину гарантированного выигрыша игрока A.

Решение. Выигрыши игрока A при выборе разных фиксированных стратегий игроком B составит в среднем

$$\sum_i a_{il} p_i = \begin{cases} 3\cdot 0 + 5\cdot 0{,}6 + 7\cdot 0{,}4 = 5{,}8; \ l = 1, \\ 1\cdot 0 + 7\cdot 0{,}6 + 2\cdot 0{,}4 = 5; \ l = 2, \\ 4\cdot 0 + 2\cdot 0{,}6 + 3\cdot 0{,}4 = 2{,}44; \ l = 3. \end{cases}$$

Гарантированный выигрыш игрока A $\varsigma = \min\limits_l \sum\limits_i a_{il} p_i = 2{,}4$.

Среднее значение выигрыша игрока A $v = \sum\limits_{i,j} a_{ij} p_i q_j$ =3,94.

Контрольные вопросы для самопроверки

1. Что называют конфликтной ситуацией?
2. Что называют теорией игр?
3. Что называют стратегией (чистой стратегией) игрока?
4. Какие виды игр Вы знаете?

Глава 13. Сетевое планирование

13.1.Планирование работ

При выполнении крупномасштабных технических проектов, строительстве предприятий, управлении производственными объединениями

постоянно приходится решать задачи планирования сложного комплекса работ. Например, строительство промышленного объекта включает в себя создание подъездных путей, подготовку строительной техники, подвоз материалов и т.д. В целом задачи планирования такого рода работ имеют следующие общие черты:

1. Работа представляет собой комплекс более элементарных работ.
2. Работы друг друга обусловливают, то есть не могут выполняться в произвольном порядке и для начала одних работ требуется предварительное выполнение некоторых других.

Типичными параметрами, которыми описываются такие ситуации, являются:

1. Время, необходимое для выполнения всего комплекса работ и отдельных звеньев.
2. Стоимость всего комплекса работ и отдельных звеньев.
3. Сырьевые, энергетические и прочие ресурсы, затрачиваемые на выполнение работ.

Задача о рациональном планировании работ включает в себя следующие вопросы:

1. Когда начинать и когда заканчивать отдельные работы?
2. Как распределить имеющиеся средства между отдельными звеньями?
3. Что может послужить препятствиями для выполнения работы в срок и как их устранить?

Точное решение некоторых задач указанного выше вида осуществляется с помощью сетевого планирования. Для построения математической задачи сетевого планирования необходимо задать исходную информацию о работах, которые требуется выполнить, в *структурную таблицу комплекса работ*. Эта таблица показывает, что работа a_i может быть выполнена только после окончания работ a_{i1}, a_{i2},.... Некоторые виды работ не опираются ни на какие другие. В противном случае данный комплекс работ невозможно начать до окончания определенных работ.

13.2. Сетевой график

Структурную таблицу комплекса работ всегда можно упорядочить так, чтобы следующие по номерам работы опирались только на предыдущие. Для этого сначала работы разбивают на типы, так чтобы работы (k+1) типа опирались только на работы типа не выше k, а затем работам приписывают новые номера, начиная с 1-го типа, затем переходя ко 2-му и т.д. В дальнейшем предполагаем, что задана уже упорядоченная структурная таблица. Структурную таблицу можно изобразить графически в виде *сетевого графика*. Основными элементами графика являются *узлы*, которые изображают кружками, и *работы*, которые изображают стрелками. Узлы будем обозначать заглавными буквами с индексами: A_0, A_1, ... Узел A_0 соответствует началу работ. Из этого узла проводим по одной стрелке на каждую работу первого типа a_1, a_2, ..., a_k. В конце стрелки рисуем узлы $A_1, A_2, \ldots, A_k$ (узел A_i соответствует состоянию, когда выполнена работа a_i), как показано на рис. 36.

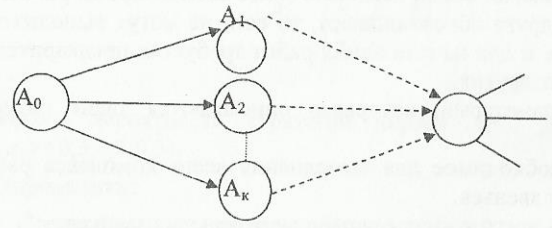

Рис. 36. Схема стрелок и узлов на начальном этапе работ

Затем переходим к работам второго рода и так далее. Для установления минимального времени, необходимого для выполнения данного комплекса работ, рассматривают *временные сетевые графики* комплекса работ. Для построения временного сетевого графика выбирается ось времени, которую примем горизонтальной. Тогда проекция вектора, изображающего некоторую работу a_i, на ось времени должна равняться длительности этой работы. Вспомогательные узлы вида $A_{i1,i2,\ldots ik}$, которые в обычных сетевых графиках используются для обозначения окончаний работ $i_1 i_2 \ldots i_k$ k-го вида, теперь не используются, а все стрелки сводят в тот из узлов $A_{i1,i2,\ldots ik}$, проекция которого на ось времени имеет наибольшее значение (рис. 37)

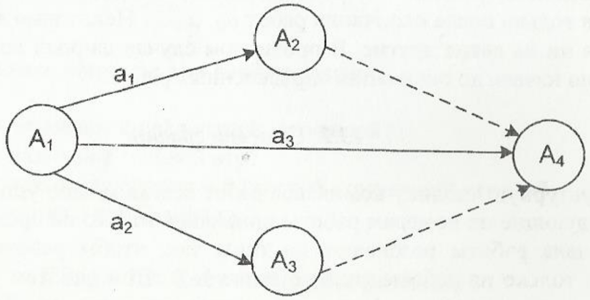

Рис. 37. Временной сетевой график с пунктирным указанием фиктивных работ

Узлы соединяют стрелками даже в том случае, когда реальную работу не производят, и тогда они обозначают фиктивные работы, отражая только последовательность этапов работ.

Начальный узел A_0 удобно размещать в начале координат. Тогда временной сетевой график сразу показывает, какое минимальное время требуется для выполнения всего комплекса работ, это время равно T – координате конечного узла A.

13.3. Критические пути и работы

Введем понятие *критического пути*, которым называется цепочка из последовательно выполняемых работ и узлов, для которых суммарная длительность работ составляет T. В примере, изображенном на рис. 38, критическим является путь A_0, A_1, A_3, A_5, A_7, A_9, A.

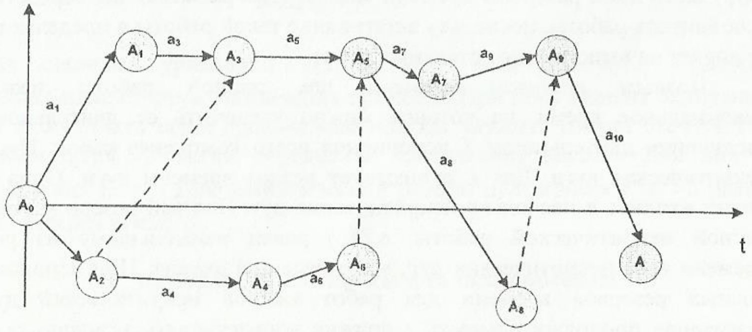

Рис. 38. Временной сетевой график с начальным и конечным узлами

Критических путей может быть несколько, хотя в примере, показанном на рис. 38, такой путь один.

Вводят понятия *критическая работа* и *критический узел* как работу и узел, входящие в некоторый критический путь. Очевидно, что увеличение длительности любой из критических работ приводит к увеличению длительности всего комплекса работ. Для некритических работ имеются некоторые резервы времени. *Некритической дугой* называется цепочка некритических узлов и работ (включая фиктивные), начинающаяся и заканчивающая критическими узлами. В приведенном на рис. 38 примере, имеются три некритические дуги: $A_0 A_2 A_3$; $A_0 A_2 A_4 A_6 A_5$; $A_5 A_8 A_9$.

13.4. Резервы времени

Поскольку при выполнении критических работ резервов времени нет, то возможные резервы времени относятся только к некритическим работам. *Частичным резервом времени* при выполнении некритической работы называется максимальное время, на которое можно увеличить ее длительность без изменения моментов начала остальных работ. Частичные резервы времени находят по временному сетевому графику.

В примере, показанном на рис. 38, частичный резерв времени работы a_6 составляет одну единицу, если следующий узел a_5, опирающийся на работу a_6, удален от узла A_6 на одну единицу (по оси t). Таким образом, в общем случае у работы a_i имеется частичный резерв времени, если из узла A_i выходят только фиктивные работы. При этом частичный резерв равен минимальной продолжительности этих фиктивных работ. Частичный резерв времени может быть равен нулю, как, например, для некритических работ a_2 и a_4. Частичным резервом времени может распоряжаться непосредственный исполнитель работы, поскольку затягивание такой работы в пределах резерва не влияет на выполнение остальных работ.

Полным резервом времени для данной работы называется максимальное время, на которое можно увеличить ее длительность без увеличения длительности T исполнения всего комплекса работ. Пусть L – некритическая дуга. Для L существует резерв времени дуги. Одна работа может входить в несколько некритических дуг. Полный резерв времени для данной некритической работы $\Delta T(L)$ равен наименьшему из резервов времени всех некритических дуг, в которые она входит. При использовании полных резервов времени для работ каждой некритической дуги их суммарная продолжительность t должна удовлетворять условию $t \leq \Delta T(L)$. При использовании полных резервов следует согласовывать друг с другом разные работы. Поэтому полным резервом времени должен распоряжаться руководитель всего комплекса работ.

При большом количестве работ построить сетевой график сложно. Полезно иметь чисто алгебраические правила вычисления возможного начала t_i и окончания T_i работы a_i, а также момента наиболее позднего возможного без увеличения длительности комплекса работ начала t'_i и окончания T'_i. Обозначим τ_i длительность работы a_i. Тогда $T_i = t_i + \tau_i$, $T'_i = t'_i + \tau_i$. Пусть работа a_i опирается на работы a_{i1}, a_{i2}, a_{i3}…. Тогда

$$t_i = \max\{T_{i1}, T_{i2}, \ldots\}. \qquad (116)$$

Это равенство означает, что a_i можно начинать, как только закончатся все работы a_{i1}, a_{i2}, a_{i3}...

Равенство

$$T_i = t_i + \tau_i, \qquad (117)$$

и (116) позволяет последовательно определить время начала t_i и время окончания T_i для всех работ от a_1 до a_n. Начальными условиями служат равенства t_i=0 (все работы начинаются в нулевой момент времени). Тогда для работ 1-го типа $T_i = \tau_i$ и далее в соответствии с описанной схемой находят все t_i, T_i.

Подобным же образом можно построить алгоритм нахождения времени позднего начала t'_i и окончания T'_i работ, используя уравнения

$$T'_i = t'_i + \tau \qquad (118)$$

и

$$T'_i = \min\{t'_{i1}, t'_{i2} t'_{i3}, \ldots\}. \qquad (119)$$

Из системы уравнений (118), (119) величины t'_i, T'_i находят последовательно, стартуя с окончания процесса. При этом момент окончания работ T может быть задан произвольно и будет служить точкой отсчета. По найденным t_i, T_i, t'_i, T'_i легко определить критические работы. Для них и только для них $t_i = t'_i$. Для некритической работы a_i величина $t'_i - t_i > 0$ равна полному резерву времени.

Контрольные вопросы для самопроверки

1. Назовите основные элементы сетевого графика.
2. Назовите временные параметры сетевого графика.

Глава 14. Генетические алгоритмы и эволюционное программирование

14.1. Генетические понятия

Генетические алгоритмы используются для оптимизации сложных систем. Идея генетических алгоритмов основана на эволюционной теории, согласно которой эволюция живых организмов определяется следующими факторами:

1. Случайной изменчивостью,
2. Наследственностью,

3. Естественным отбором.

Определим ряд основных понятий, используемых в теории генетических алгоритмов.

- *Ген* - единица наследственной информации, не делимая в функциональном отношении, которая передается от родителей к потомкам.
- *Аллель* - фиксированная форма гена.
- *Геном* – совокупность всех генов данного организма.
- *Мутация* – случайное наследственное изменение отдельного гена.
- *Естественный отбор* – процесс, направленный на повышение вероятности оставления потомства одной формы организма, по сравнению с другими.
- *Изменчивость* – разнообразие признаков и свойств у особей групп особей любой степени родства.
- *Популяция* - совокупность особей определенного вида, внутри которого осуществляется случайное скрещивание.
- *Селекция* – форма искусственного отбора, при котором эволюция направляется факторами внешней среды.
- *Эволюция* – процесс постепенного и непрерывного изменения форм организмов от одного состояния к другому.

14.2. Генетический алгоритм

Представляет собой адаптивный поисковый метод, основанный на селекции лучших элементов в популяции подобно эволюционной теории Ч. Дарвина. Основой генетических алгоритмов служит модель биологической эволюции и методы случайного поиска. Эволюционный поиск с точки зрения приобретения информации – это последовательное преобразование одного конечного множества промежуточных решений в другое. Цель разработки генетических алгоритмов состоит в том, чтобы понять механизмы развития и адаптации естественных биологических и интеллектуальных систем, а также использовать эволюционные модели для решения научно-технических задач оптимизации. Задача оптимизации понимается как поиск абсолютного max(min) некоторой целевой функции.

Все генетические алгоритмы работают на основе начальной информации, в качестве которой выступает множество исходных альтернативных решений P, которые называются *исходной популяцией*. Популяция $P'=\{P_1, P_2,\ldots, P_n\}$ - множество элементов P_i, соответствующих некоторой генерации генетического алгоритма t =0,1,2…, а N – размер популяции. Каждый элемент популяции P_i представляет собой одну или несколько хромосом индивидуальной особи, то есть одно или несколько альтернативных решений.

Хромосомы состоят из генов $P_i=\{g_1, g_2, \ldots, g_v\}$ которые составляют части закодированного решения. Позиция определенного гена в хромосоме называется *локусом*. Функциональное назначение генов определяет аллель.

Гены могут иметь числовые или функциональные значения. Генетический материал хромосом обычно кодируется на основе двоичного кода $\{0,1\}$:

$$P_i = \{001\ 001\ 101\}.$$

Элементы популяции генетических алгоритмов часто называют *родителями*. Родителей выбирают из популяции на основе заданных правил, а затем смешивают (скрещивают) для производства потомства. Дети и родители в результате генерации, то есть одного цикла эволюции, создают новую популяцию. *Генерация* – это процесс реализации одной итерации алгоритма, которая называется поколением.

14.3. Эволюция в популяции

Это процесс чередования поколений, в ходе которого хромосомы изменяют свои значения так, что в результате каждое новое поколение наилучшим образом приспосабливается к внешней среде. Каждый элемент популяции имеет определенный уровень качества, который характеризуется значением целевой функции (функции полезности). Эта функция используется в генетических алгоритмах для сравнения между собой альтернативных решений и выбора лучших.

Основная задача генетических алгоритмов состоит в оптимизации целевой функции. Каждая популяция обладает наследственной изменчивостью, что означает возможность случайных отклонений в генах и хромосомах в каждом поколении. При этом наследственные признаки закрепляются, если они имеют приспособительный характер, то есть обеспечивают большее значение целевой функции. Отбраковка менее приспособленных потомков составляет суть селекции. Генетический алгоритм обеспечивает также адаптацию к изменяющейся окружающей среде, то есть к изменяющейся целевой функции.

При использовании традиционных методов оптимизации всякую новую изменившуюся задачу обычно приходится решать заново. При эволюционном подходе оптимизация часто может быть продолжена с помощью использования механизмов дополнения и видоизменения популяции. Генетический алгоритм обладает также тем достоинством, что он сходится к результатам значительно быстрее, чем простые алгоритмы случайного поиска типа Монте-Карло. Напомним, что алгоритм Монте-Карло заключается в выборе примеров случайным образом и сравнение качества альтернатив между собой.

14.4. Канонический генетический алгоритм

Канонический генетический алгоритм состоит в выполнении следующих шагов:

1. Задается функция $f(P_i)$, определяющая эффективность каждого найденного решения при значениях параметров решения P_i. P_i кодируется как вектор, который называется хромосомой. В хромосоме каждый элемент (элемент вектора) представляет собой ген. Ген кодируется в двоичном представлении.

2. В соответствии с ограничениями, налагаемыми на параметры условиями задачи, инициализируется исходная популяция P^0 потенциальных решений, состоящая из некоторого количества хромосом N, число которых задается в начале работы алгоритма и в процессе эволюции обычно не меняется.

3. Каждой хромосоме в популяции, на основе вычисления целевой функции $f(P_i)$, присваивается вероятность воспроизведения p_i. Одним из простейших способов определения p_i является пропорциональный отбор, при котором

$$p_i = \frac{f(p_i)}{\sum_{j=1}^{N} f(p_j)}. \qquad (120)$$

4. В соответствии с вероятностями воспроизведения p_i создается новая популяция хромосом, причем с большей вероятностью воспроизводятся наиболее эффективные элементы. Хромосомы производят потомков, используя операцию рекомбинации. Операция *рекомбинации* состоит из двух операторов – кроссинговера, при котором хромосомы скрещиваются, обмениваясь частями строк, и *оператора мутации*, который осуществляет вероятностные изменения генов.

5. Генетический алгоритм останавливается, если получено удовлетворительное решение, т. е. найдено решение с заданной точностью или закончилось время, отведенное на эволюцию (параметр t вышел за допустимую границу).

14.5.Оператор кроссинговера

Производит скрещивание хромосом и обмен генетическим материалом между родителями для получения потомков. Этот оператор служит для исследования новых областей, пространства параметров и улучшения существующих параметров, обеспечивая эволюционное приспособление.

Простейший одноточечный кроссинговер производит обмен частями, на которые хромосома разбивается точкой, выбираемой случайно (рис. 39).

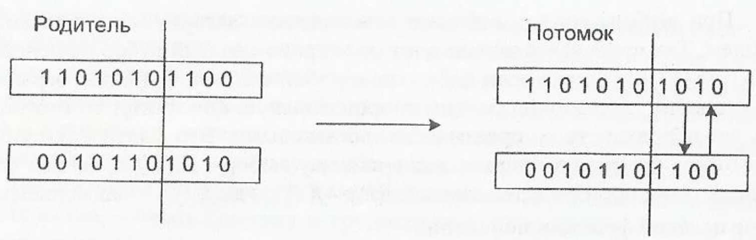

Рис. 39. Пример кроссинговера

Двухточечный кроссинговер обменивает куски строк, попавшие между двумя точками.

14.6. Оператор мутации

Применяется к каждому биту информации хромосомы с малой вероятностью $p = 10^{-3}$ и меньше, в результате чего бит изменяет значение на противоположное: $0 \to 1$, $1 \to 0$. Мутации нужны для расширения пространства поиска, то есть для предотвращения полной потери разнообразия в результате получения чистых аллелей - популяции, в которой один или несколько генов одинаковы у всех элементов популяции.

Наиболее часто в генетических алгоритмах используется бинарная кодировка для представления элемента популяции. Такой выбор связан с тем, что двоичный код наиболее прост в обработке компьютером. Существуют альтернативные двоичному коду схемы кодировки, в том числе использование представления чисел с помощью плавающей запятой при фиксированной разрядности. В зависимости от свойств конкретной решаемой задачи предпочтительными могут быть те или иные способы кодирования. Отметим также, что довольно часто применяются коды Грея.

Коды Грея построены таким образом, что каждый символ кодируется при помощи четырех бинарных символов, как показано на рис. 40.

0	0	0	1
0	0	1	0
1	0	0	1

Рис. 40. Примеры элементов кода Грея

При выборе метода отбора в генетических алгоритмах возникает ряд проблем. Так чаще всего используют пропорциональный отбор на основании формулы (120). Однако если добавить к любой целевой функции, ограниченной вариации, достаточно большую произвольную константу, то можно сделать все вероятности p_i практически одинаковыми. Это фактически уничтожает отбор, приводя эволюцию к случайному выбору. Для устранения данного недостатка производится замена: $f(P_i) \rightarrow f(P_i')$, где $f(P_i')$ – наименьшее значение целевой функции популяции.

Еще одна из математических проблем, связанная с пропорциональным отбором, состоит в том, что такая процедура не может гарантировать асимптотическую сходимость в пределе большого числа поколений к глобальному оптимуму. Наилучшая хромосома в популяции может быть потеряна в любом поколении. Поэтому результаты эволюции, достигнутые в ряде поколений, могут быть утрачены. Одним из способов преодоления этого явления является использование элитного отбора, который всегда сохраняет наилучшую хромосому в популяции.

Контрольные вопросы для самопроверки

1. Какими факторами определяется идея генетического алгоритма? Что является ее основанием?
2. Перечислите основные понятия и определения, применяемые в теории генетических алгоритмов.
3. Что является целью применения генетического алгоритма? Каковы принципы его работы?
4. Какими достоинствами обладает генетический алгоритм в отличие от использования традиционных методов оптимизации?
5. В чем заключается принцип работы канонического генетического алгоритма?
6. Какие операторы позволяют осуществлять кодировку элементов?

Заключение

Настоящее пособие посвящено основным математическим моделям и методам, применяемым в управлении строительным производством и при его планировании. Эффективное решение задач планирования и управления в строительстве в настоящее время связано с необходимостью быстро обрабатывать огромный объем информации, что практически невозможно без применения средств вычислительной техники и программного обеспечения, в основе которого лежат математические методы и модели. Задача предлагаемого пособия - дать основные навыки в применении математических методов и моделей, которые используются в управлении строительным производством, показать принципы моделирования и обосновать применение математических методов. В пособии доступно изложены основные принципы построения математических моделей, помогающие понять принцип работы методов и алгоритмов, лежащих в основе математического обеспечения пакетов прикладных программ, используемых в управлении строительством.

Построение математических моделей, адекватно описывающих процесс строительства, - очень сложная и трудоемкая задача, окончательного решения которой в настоящее время пока нет. Математические модели и методы, применяемые в строительстве, постоянно совершенствуются, но без знания принципов работы и построения базовых моделей, методов и алгоритмов проводить исследования в этой области невозможно.

Надеемся, что наше пособие поможет освоить основные принципы работы математических моделей и методов, применяемых в управлении строительным производством и при его планировании, и создаст необходимую базу для продолжения исследований в этой области.

Библиографический список

1. Гмурман, В.Е. Теория вероятностей и математическая статистика / В.Е. Гмурман. - М: Высшая школа, 1977. – 479 с.
2. Арнольдь, В.И. Теория катастроф / В.И. Арнольдь. - М: Наука, 1990. – 128 с.
3. Сигорский, В.П. Математический аппарат инженера / В.П. Сигорский. - Киев: Техника, 1975. – 768 с.
4. Волков, И.К. Исследование операций / И.К. Волков, Е.А. Загоруйко. - М.: Изд-во МГТУ им. Н.Э.Баумана, 2004. – 440 с.
5. Оуэн, Г. Теория игр / Г. Оуэн. - М.: Едиториал УРСС, 2004. – 216 с.
6. Бурков, В.Н. Теория графов в управлении организационными системами / В.Н. Бурков, А.Ю. Заложнев, Д.А. Новиков. - М.: Синтег, 2001. – 124 с.
7. Шикин, Е.В. Исследование операций / Е.В. Шикин. - М.: Изд-во Проспект, 2006. – 280 с.
8. Вентцель, Е.С. Исследование операций / Е.С. Вентцель. - М.: Высшая школа, 2001. – 208 с.
9. Харари, Ф. Теория графов / Ф. Харари. - М.: Едиториал УРСС, 2003. – 296 с.
10. Гладков, Л.А. Генетические алгоритмы / Л.А. Гладков, В.В. Курейчик, В.М. Курейчик. - М.: Физматлит, 2006. – 320 с.

ОГЛАВЛЕНИЕ

Учебное издание

Головинский Павел Абрамович
Мищенко Валерий Яковлевич
Михайлов Евгений Михайлович

МАТЕМАТИЧЕСКИЕ МЕТОДЫ ПРИНЯТИЯ УПРАВЛЕНЧЕСКИХ РЕШЕНИЙ В СТРОИТЕЛЬСТВЕ

Учебное пособие

для студентов старших курсов и магистрантов, обучающихся по направлению «Строительство»

Редактор Аграновская Н.Н.

Подписано в печать 22.10.2008 Формат 60х84 1/16. Уч.-изд. л. 5,6.
Усл.-печ. л. 5,7. Бумага писчая. Тираж 240 экз. Заказ № 575.

Отпечатано: отдел оперативной полиграфии Воронежского государственного архитектурно-строительного университета
394006 Воронеж, ул.20-летия Октября, 84