И.Г. Вовк, Т.Ю. Бугакова

ОСНОВЫ СИСТЕМНО-ЦЕЛЕВОГО ПОДХОДА И ПРИНЯТИЕ РЕШЕНИЙ

Новосибирск
СГГА
2011

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

ГОУ ВПО «СИБИРСКАЯ ГОСУДАРСТВЕННАЯ ГЕОДЕЗИЧЕСКАЯ АКАДЕМИЯ»

И.Г. Вовк, Т.Ю. Бугакова

ОСНОВЫ СИСТЕМНО-ЦЕЛЕВОГО ПОДХОДА И ПРИНЯТИЕ РЕШЕНИЙ

Рекомендовано Сибирским региональным учебно-методическим центром высшего профессионального образования для межвузовского использования в качестве учебного пособия для студентов, обучающихся по специальностям 280101.65 «Безопасность жизнедеятельности в техносфере», 120101.65 «Прикладная геодезия», 120103.65 «Космическая геодезия» и по направлению подготовки бакалавров 120100.62 «Геодезия»

Новосибирск
СГГА
2011

УДК 519.87:519.71
В611

Рецензенты: доктор экономических наук, профессор, СибГУТИ
Г.Г. Шалмина
кандидат физико-математических наук, доцент НГИ
А.Ю. Губин
кандидат технических наук, доцент, СГГА *А.С. Суздалев*

Вовк, И.Г.
В611 Основы системно-целевого подхода и принятие решений [Текст]: учеб. пособие / И.Г. Вовк, Т.Ю. Бугакова. – Новосибирск: СГГА, 2011. – 152 с.

ISBN 978-5-87693-461-1

В предлагаемом учебном пособии рассматриваются методические основы системно-целевого подхода для изучения сложных систем и методы принятия решений в известных условиях и в условиях неопределенности и риска. Пособие будет полезно студентам, магистрантам, аспирантам и специалистам, которым необходимы начальные сведения по вопросам системного анализа, моделирования и принятия решений.

Печатается по решению редакционно-издательского совета СГГА

УДК 519.87:519.71

ISBN 978-5-87693-461-1

СОДЕРЖАНИЕ

ВВЕДЕНИЕ

Во второй половине XX века большие и сложные системы стали предметом изучения, управления и проектирования. Сложность любой системы проявляется в необъяснимости, непредсказуемости, неожиданности ее состояния или поведения.

Для выявления причин сложности систем потребовалось объединить усилия специалистов различных профилей, унифицировать и согласовать разнообразную конкретную информацию, планировать и реализовать изменения необходимые для устранения сложности, применяя экспериментальные, эмпирические и строгие математические методы, а также и неформальные эвристические процедуры, основанные на опыте и интуиции человека. Все перечисленные действия объединяются в одном определении – «системно-целевой подход (СЦП)». Он базируется на общей теории систем, теории управления, теории самоорганизации и эволюции и практически опирается на возможности современной математики и информационно-вычислительных технологий обработки информации.

Всякая деятельность людей целенаправленна и требует принятия решений для выполнения действий, направленных на достижение цели. Для этого необходимо, чтобы лица, принимающие решения, обладали знаниями и практическими навыками принятия решений в разнообразных условиях и ситуациях. Современная теория принятия решений – математическая дисциплина, основанная на достижениях современной математики. Наиболее понятное и простое изложение задач принятия решений осуществляется на математических моделях этих задач. Наиболее простые и типичные варианты моделей принятия решений рассмотрены в данной работе с позиций системно-целевого подхода. Эти модели хотя и не дают лучшего решения (его может просто не существовать), но на основе системного анализа задач принятия решений позволяют глубже проникнуть в сущность проблемы, исключить заведомо плохие решения и выявить информацию, необходимую для получения лучшего решения.

Системно-целевой подход и методы принятия решений все шире внедряются во все сферы человеческой деятельности. Это обстоятельство потребовало для ряда направлений высшего профессионального образования включить в программы подготовки специалистов курсы по системному анализу и моделированию. В настоящее время имеется немного учебников по этому направлению, изданы они малыми тиражами и труднодоступны для понимания не подготовленному читателю.

Предлагаемое учебное пособие условно состоит из двух частей. В первой части (1–4 раздел) рассматриваются методические основы системно-целевого подхода для изучения сложных систем, а во второй (5–8 раздел) – основные понятия и методы принятия решений в известных условиях и в условиях неопределенности и риска.

Разделы учебного пособия с 1 по 4 – результат совместной работы авторов. Разделы с 5 по 8 подготовил И.Г. Вовк.

Авторы надеются, что оно позволит получить начальное представление о методах системно-целевого подхода и процедурах принятия решений, их практическом применении. Пособие будет полезно студентам различных специальностей, магистрантам, аспирантам и специалистам, которым необходимы начальные сведения по вопросам системного анализа, моделирования и принятия решений.

1. ОСНОВНЫЕ ПОНЯТИЯ И ЗАДАЧИ СИСТЕМНО-ЦЕЛЕВОГО ПОДХОДА

Все окружающее нас физическое пространство представляет собой большую сложную систему, в которой природно-физические процессы, протекающие на Земле, процессы биосферы, техногенная, социальная деятельность человека находятся в сложной зависимости друг от друга. Не учитывать этой сложной зависимости при решении задач невозможно. В связи с этим возникает необходимость применения системно-целевого подхода (СЦП).

Системно-целевой подход ставит своей целью исследование функциональных связей и отношений объектов, процессов и явлений таким образом, чтобы изучаемые объекты стали бы более управляемыми, познаваемыми, а механизм взаимодействия этих объектов – применимым к другим объектам, процессам и явлениям. В самой природе науки лежит стремление к единству и синтезу знания. Изучение этого стремления, выявление особенностей этого процесса – одна из задач современных исследований в области теории научного знания.

В настоящее время системно-целевой подход стоит на стыке многих научных направлений и занимает междисциплинарную область. Так, например, 40 % информации специалисту необходимо черпать из смежных областей, подчас отдаленных. В технических науках речь идет о системотехнике, в кибернетике – о системах управления, в биологии – о биосистемах и их структурных уровнях, в социологии – о возможностях структурно-функционального подхода, в медицине – о системном лечении сложных болезней и т. д.

Системно-целевой подход объединяет две части познания и преобразования мира: системный анализ и системный синтез, которые базируются на двух взаимосвязанных друг с другом составляющих, формальной и понятийно-содержательной. Формальная составляющая использует формальный математический аппарат различного уровня строгости. Понятийно-содержательная составляющая концентрируется на основных понятиях, идеях, подходе, концепциях, возможностях, на основных методологических принципах, использует неформальное, основанное на интуиции и опыте человека, или «полуформальное» введение в суть рассматриваемых идей и понятий.

Системный анализ делает известными отдельные признаки сложного объекта как целостного образования и свойства его частей как самостоятельных предметов. **Системный анализ –** совокупность методов, используемых для обоснования решений по сложным проблемам. Эта совокупность включает не только формальные математические методы, но и *эвристические*, *экспертные* и *эмпирические* методы.

Синтез систематизирует представления, добытые в результате анализа. При этом именно системно-целевой подход выделяет и рассматривает те отличительные признаки и отношения между компонентами объекта, в силу которых они могут считаться частью какого-то целостного образования, и которые, следовательно, являются существенными для синтеза. **Синтез** – процесс (как правило, целенаправленный) соединения или объединения ранее разрозненных вещей или понятий в нечто качественно новое, целое.

1.1. Понятие системы

Основополагающим для системного анализа является понятие системы. Определений системы существует множество. Это объясняется тем, что определение – это языковая модель, и, следовательно, различия целей и требований к модели приводят к разным видам изложения одного и того же понятия.

Понятие «система» в переводе с греческого языка означает «целое, составленное из частей». В философских школах Древней Греции считалось, что часть всегда проще целого, что, изучив каждую из частей, можно узнать свойства целого. Однако в настоящее время выяснилось, что такое понимание системы недостаточно, целое обладает свойствами, которых нет ни у одной из частей.

Система – это совокупность частей (элементов и подсистем), обособленных от среды и объединенных общими ресурсами, связями, функциональной средой и целью существования и обладающая свойствами, отсутствующими у отдельных частей. Это определение позволяет все объекты и явления естественного или искусственного происхождения рассматривать как системы. Солнечная система, река, лес, самолет, автомобиль, завод, торговое предприятие – все это примеры естественных и искусственных систем.

Из данного определения следует, что всякая система предназначена для достижения некоторой цели. Что же такое «цель системы»? Для ответа на поставленный вопрос заметим, что все системы возникают вследствие потребности разрешения противоречий между необходимостью и возможностью их существования. Такое противоречие называют ***проблемой***, или ***проблемной ситуацией.***

Таким образом, ***цель системы*** – результат, который разрешает проблему (противоречие). Она может быть субъективной или объективной. Субъективные цели определяет человек, а объективные – природа [1]. Ни одна система не может существовать без цели. Нет цели – нет системы.

Объективная цель – будущее реальное состояние системы и окружающей среды, в которое они приходят в результате реализации объективных закономерностей. Например, падение материальных тел на Землю, износ технического оборудования в процессе эксплуатации, возникновение и развитие опасных процессов в техносфере.

Субъективная цель – образ желаемого будущего состояния системы и окружающей среды. Например, получение высшего образования, вертикальный взлет и посадка самолета и многие другие цели являются субъективными. Они могут быть достижимыми или недостижимыми. Недостижимыми целями являются те, которые противоречат законам природы. Например, создание вечного двигателя, материальной точки и т. д. – недостижимые субъективные цели.

Внутреннее строение объектов неоднородно, в них можно различать части, в которых, в свою очередь, могут быть выделены составные части и т. д. В связи с этим возникает необходимость структурирования системы.

Всякая система состоит из частей, которые могут быть простыми (элементарными) и сложными. ***Простыми, элементарными частями (элементами) системы*** считаются всякие условно неделимые и самостоятельно функционирующие части системы. Например, планеты – элементы Солнечной системы, деревья – элементы системы «лес», рабочие на заводе – элементы системы «завод».

Сложными частями (подсистемами) системы считают совокупность относительно однородных элементов, объединенных общими функциями и ресурсами. Малые планеты в Солнечной системе, бригады рабочих на заводе, устройства управления автомобилем образуют подсистемы в соответствующих системах.

Объединение элементов в подсистемы позволяет рассматривать не свойства отдельных элементов, мало отличающиеся между собой, а их общие свойства. Количественные оценки этих свойств называют ***интегральными характеристиками***. Они позволяют сократить число параметров, используемых при описании и оценке свойств подсистем и системы в целом. Например, при формулировке закона всемирного тяготения не рассматривается детальное строение космических тел, а принимается во внимание только их масса и расстояние между ними, при описании случайных явлений вместо множества значений случайных величин рассматривают их функции распределения или их числовые характеристики и т. д.

Понятие элементарности относительно, и то, что с одной точки зрения является элементом, с другой оказывается подсистемой, подлежащей дальнейшему разделению. Для различных целей одна и та же система может быть разбита на разные составные части.

В общем виде состав системы условно представлен на рис. 1.

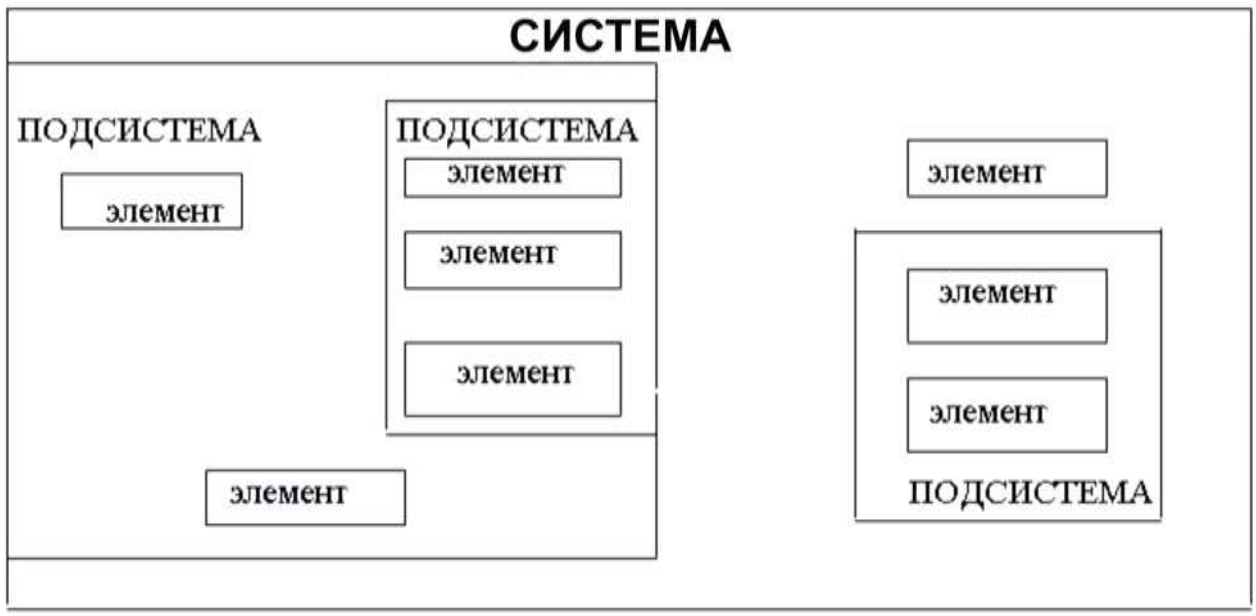


Рис. 1. Состав системы

Примером модели состава системы, может служить материнская плата компьютера (рис. 2), где четко прослеживается как совокупность отдельных элементов, так и выделение элементов в группы (подсистемы).

Рис. 2. Материнская (системная) плата

Способ определения элементарных объектов неоднозначен и зависит от уровня и глубины изученности объекта, назначения системы и других обстоятельств. Поэтому любой объект может быть представлен различными совокупностями элементарных объектов, каждый из которых обладает элементарными свойствами. Для любой системы характерно наличие свойства, присущего системе в целом, но отсутствующего у каждого элемента в отдельности. Это свойство называют ***свойством эмерджентности (системообразующим свойством)*** [1]. Этим свойством не обладает ни один из элементов системы в отдельности. Оно проявляется в результате взаимодействия системы с окружающей средой и по существу есть тот результат, для достижения которого предназначена система. Например, свойством эмерджентности естественной системы «Солнечная система» служит то, что Солнце, планеты и другие космические элементы, входящие в систему, рассматриваются относительно других космических систем как самостоятельный, неделимый объект. Свойством эмерджентности искусственной системы «завод», является способность завода производить определенный продукт или товар. Главная трудность в определении состава системы заключается в том, что разделение целостной системы на части является относительным, т. е. границы между подсистемами условны. Относительны и границы между системой и окружающей средой.

При удалении из системы какой-то ее части система теряет некоторые свойства (становится другой системой). Часть, изъятая из системы, тоже теряет свойства, которые проявлялись в результате ее взаимодействия с другими частями. Следовательно, система – не просто совокупность элементов, и, изучая каждый из них в отдельности, невозможно познать все свойства системы в целом.

Для описания или определения свойств объекта как целого необходима дополнительная информация. Эту информацию можно найти, если установить функции взаимосвязи (отношения) между элементарными объектами, удовлетворяющие определенным условиям. Множество необходимых и достаточных для достижения цели связей (отношений) между элементарными объектами называется **структурой объекта** [1].

Структура определяется, исходя из распределения функций между частями объекта, в ней отображаются отношения этих частей, как между собой, так и с внешней средой. Реальные объекты имеют между собой неисчерпаемое множество отношений, однако существенными для достижения цели являются лишь некоторые. Следовательно, и в структуре учитывается конечное число связей, важных по отношению к рассматриваемой цели. Построение и анализ структуры позволяет отвлечься от конкретной физической природы элементов и, проводя математические преобразования структуры, выявить некоторые общие закономерности, характеризующие свойства объекта.

Состав и структура системы дают наглядное представление о том, как устроен объект, из каких частей он состоит и как эти части связаны друг с другом. Принцип, по которому объединение элементов приводит к появлению новых свойств, отличных от свойств элементов, называют *принципом организации*. Различные принципы организации могут приводить к созданию систем, различающихся своими структурами, но тождественными по функциональному назначению. Сложность реальных объектов порождает разнообразие структур. Среди этого многообразия выделяют линейные, иерархические (древовидные), матричные и сетевые структуры, изображенные на рис. 3а, 3б, 3в, 3г [3, 4].

Линейная структура устанавливает связь (например, обмен информацией) между элементарными объектами в заданной последовательности, подчиненность последующего элемента предшествующему элементу.

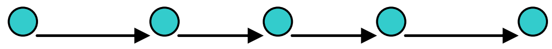


Рис. 3а. Основные типы структур. Линейная структура

В *иерархических* структурах элементарные объекты объединяются в группы, между которыми устанавливается соподчиненность, как в линейных структурах, но связи между элементарными объектами одной группы, как правило, запрещены, хотя допускаются и исключения, все элементы данной группы связаны только с одним элементом предшествующей группы.

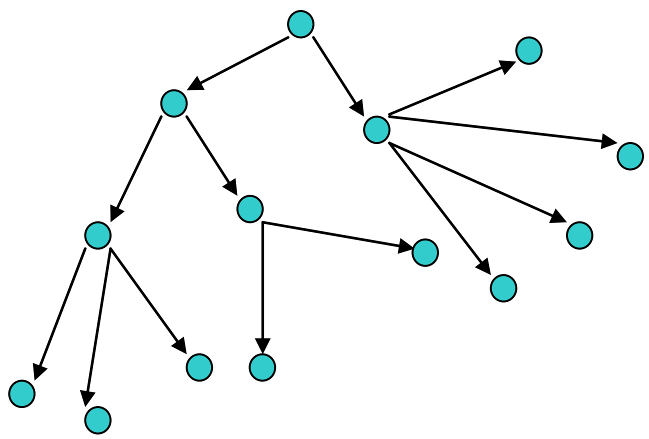


Рис. 3б. Основные типы структур. Иерархическая структура

Матричные структуры отображают отношения между элементарными объектами, которые могут быть представлены двумерными таблицами.

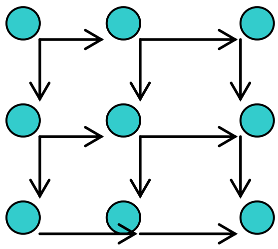


Рис. 3в. Основные типы структур. Матричная структура

Сетевые структуры являются обобщением иерархических структур. Они допускают произвольные связи между элементарными объектами, где каждый элементарный объект может подчиняться (зависеть) нескольким элементарным объектам.

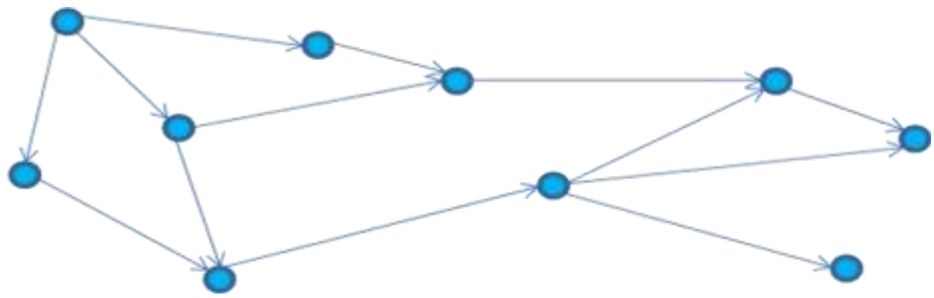


Рис. 3г. Основные типы структур. Сетевая структура

Примером видов структур может служить различный тип соединения компьютеров в сети.

Связь – одно из основных понятий в системно-целевом подходе. Система как единое целое существует именно благодаря наличию связей между ее элементами, т. е., иными словами, связи выражают законы функционирования системы. По характеру связи бывают прямыми и обратными, а по виду проявления (описания) детерминированными и вероятностными.

Прямые связи предназначены для заданной функциональной передачи вещества, энергии, информации или их комбинаций – от одного элемента к другому в направлении основного процесса.

Обратные связи, в основном, выполняют осведомляющие функции, отражая изменение состояния системы в результате управляющего воздействия на нее. Процессы управления, адаптации, саморегулирования, самоорганизации, развития невозможны без использования обратных связей. Схема обратной связи изображена на рис. 4.

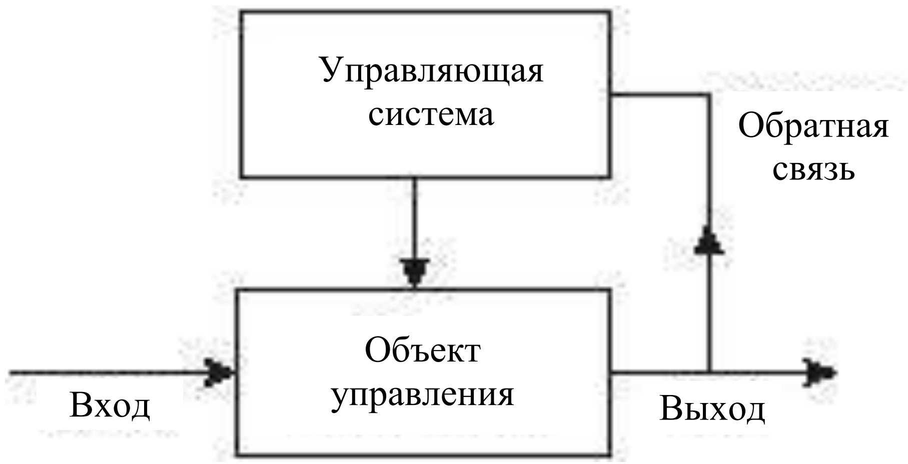


Рис. 4. Пример обратной связи

Обратная связь предназначена для выполнения следующих операций:

- сравнение данных на входе с результатами на выходе с выявлением их качественно-количественного различия;
- оценка содержания и смысла различия;
- выработка решения, вытекающего из различия;
- воздействие на вход.

С помощью обратной связи сигнал (информация) с выхода системы (объекта управления) передается в управляющую систему. Здесь этот сигнал, содержащий информацию о работе, выполненной объектом управления, сравнивается с сигналом, задающим критерии состояния объекта.

Основными функциями обратной связи являются:

1) противодействие тому, что делает сама система, когда она выходит за установленные пределы (например, реагирование на выход состояния объекта за пределы нормы);

2) компенсация возмущений и поддержание состояния устойчивого равновесия системы (например, воздействие на отдельные свойства объекта для удержания его состояния в допустимых пределах);

3) синтезирование внешних и внутренних возмущений, стремящихся вывести систему из состояния устойчивого равновесия, сведение этих возмущений к отклонениям одной или нескольких управляемых величин (например, выработка управляющих команд на изменение параметров системы в связи с изменением ее состояния).

Нарушение обратных связей в технических системах по различным причинам ведет к тяжелым последствиям.

Детерминированная связь, как правило, однозначно определяет причину и следствие, дает четко обусловленную формулу взаимодействия элементов. Недетерминированная связь определяет неявную, косвенную зависимость между элементами системы (рис. 5).

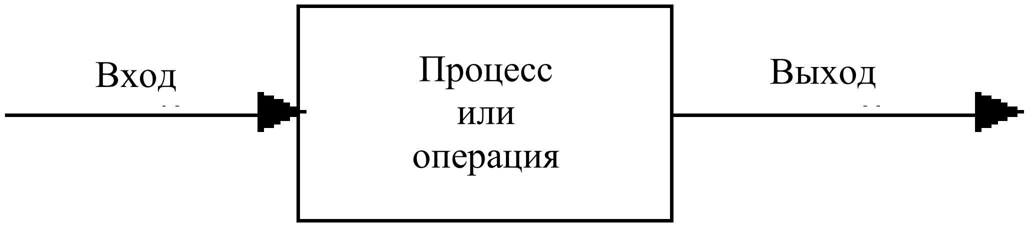


Рис. 5. Функционирование системы

Всякое целенаправленное действие состоит в преобразовании входных (известных) сигналов в значения выходных сигналов с учетом обратной связи.

Вход – каналы передачи ресурсов для протекания процесса.

Выход – результат процесса.

Процесс – преобразование входа в выход.

Система взаимодействует с внешней средой через входы и выходы. Из внешней среды на вход системы поступают ресурсы, необходимые для достижения цели. Формирующийся на выходе конечный результат поступает во внешнюю среду. Готовность к адекватному восприятию результата внешней средой называют **ингерентностью**.

Управление системой связано с понятием ограничения.

Ограничение обеспечивает соответствие между выходом системы и требованием к нему как к входу во внешнюю среду. Внешняя среда в данном случае является потребителем. Если заданное требование не выполняется, ограничение не пропускает его.

Процесс функционирования системы связан с понятием «проблемной ситуации», которая возникает, когда неудовлетворительность существующего положения осознана, но неясно, что следует делать для его изменения.

Исследование объекта как системы предполагает использование ряда системных представлений (категорий), среди которых основными являются следующие.

1. Структурное представление связано с выделением элементов системы и связей между ними.
2. Функциональное представление систем – выделение совокупности функций (целенаправленных действий) системы и ее компонентов, направленное на достижение определенной цели.
3. Макроскопическое представление – понимание системы как нерасчленимого целого, взаимодействующего с внешней средой.
4. Микроскопическое представление основано на рассмотрении системы как совокупности взаимосвязанных элементов. Оно предполагает раскрытие структуры системы.

Понятие элемент, подсистема, система взаимопреобразуемы, система может рассматриваться как элемент системы более высокого порядка (метасистема), а элемент при углубленном анализе – как система. То обстоятельство, что любая подсистема является одновременно и относительно самостоятельной системой, приводит к двум аспектам изучения систем: на макро- и микроуровнях.

При изучении на макроуровне основное внимание уделяется взаимодействию системы с внешней средой, причем системы более высокого уровня можно рассматривать как часть внешней среды. При таком подходе главными факторами являются целевая функция системы (цель) и мера достижения цели. При этом элементы системы изучаются с точки зрения организации их в единое целое, рассматривается их влияние на функции системы в целом.

На микроуровне основными становятся внутренние характеристики системы, характер взаимодействия элементов между собой, их свойства и значения для достижения цели.

Окружающий нас мир бесконечен в пространстве и во времени; человек существует конечное время, располагая при реализации цели конечными ресурсами (материальными, энергетическими, информационными, людскими, организационными, пространственными и временными).

Противоречия между неограниченностью желания человека познать мир и ограниченной (ресурсами, неопределенностью) возможностью сделать это, между бесконечностью природы и конечностью ресурсов человечества, имеют много важных последствий, в том числе и для самого процесса познания человеком окружающего мира. Одна из таких особенностей познания, которая позволяет постепенно, поэтапно разрешать эти противоречия – использование аналитического и синтетического образа мышления, т. е. разделение целого на части и представление сложного в виде совокупности более простых компонент, и наоборот, соединение простых и построение, таким образом, сложного. Это также относится и к индивидуальному мышлению, и к общественному сознанию, и ко всему знанию людей, и к самому процессу познания.

1.2. Классификация систем

Неисчерпаемое множество систем приводит к необходимости их разделения на подмножества, обладающие некоторыми общими свойствами. Эти подмножества называются классами, а сама процедура определения классов – классификацией.

Классификация – одна из проблем распознавания образов. Каждому классу приписывают некоторые характеристические признаки или свойства, которые называют критериями классификации, и по их наличию у объекта осуществляют классификацию. Как и всякая деятельность, классификация всегда имеет целевой характер и в то же время является условной.

В зависимости от критериев классификации существуют различные варианты классификации систем. Один из вариантов классификации приведен на рис. 6 [5].

На этом рисунке приведена классификация систем по пяти критериям: происхождение системы, состав системы, взаимодействие с окружающей средой, сложность и изменчивость системы. Дадим краткую характеристику указанных на рис. 6 типов систем.

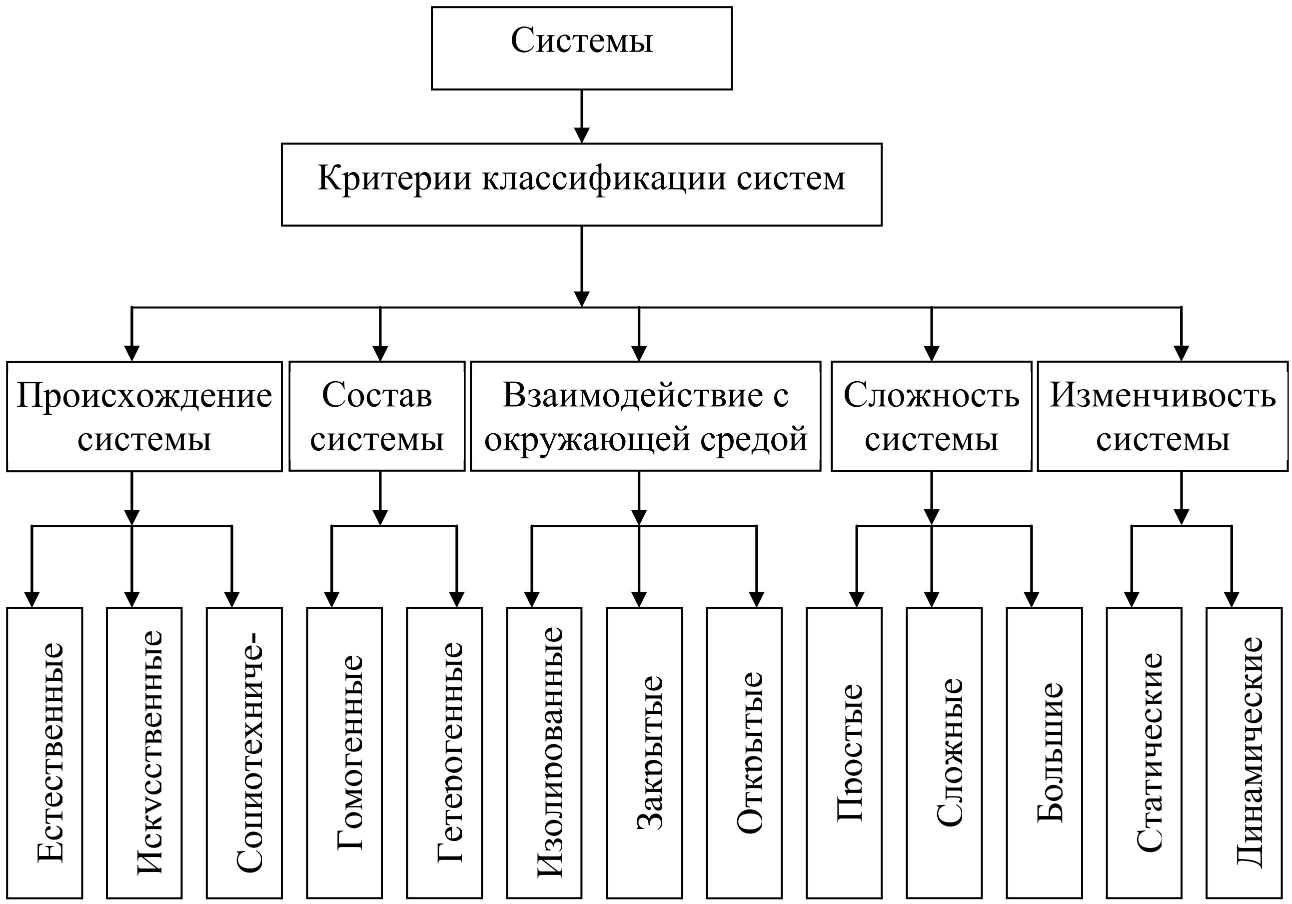


Рис. 6. Один из вариантов классификации систем

Рассмотрим системы, возникающие в природе в результате естественных процессов. Например, Солнечная система, планеты Солнечной системы, разнообразные растения и животные и многое другое. Можно ли их рассматривать просто как целое, составленное из частей? Конечно нет! Между частями Солнечной системы существуют гравитационные, электромагнитные, тепловые и другого рода взаимодействия, которые объединяют части Солнечной системы в целое. Части растений или животных объединяются в растение или животное вследствие взаимодействия своих частей для обеспечения возможности существования в природе.

В процессе развития цивилизации появились объекты, также составленные из частей, но искусственные, созданные человеком. Например, различного рода и назначения орудия труда и механизмы, производственные, общественные и политические организации, средства и способы преодоления времени и пространства. Такие искусственные объекты также не являются просто целым, составленным из частей. Их части неизбежно взаимосвязаны своим предназначением (целью). Например, пассажирский поезд состоит из локомотива и вагонов и, если не учитывать их взаимосвязи, то не возникает и целое. Аналогичная ситуация возникает, если не учитывать взаимосвязи между частями какого-либо производства.

Искусственные системы не обязательно должны состоять из материальных физических объектов. Например, алгоритм решения задачи или план преобразования некоторой системы. И алгоритм, и план состоят из частей, которые необходимо выполнить в определенной последовательности (в этом проявляется взаимосвязь частей), чтобы достичь желаемых целей, для реализации и алгоритма и плана требуются связи с внешней средой, откуда поступают ресурсы, необходимые для их реализации, и куда будут отправлены полученные результаты.

Система живая (биологическая) – система, в состав которой входят биологические элементы и подсистемы. Функционирование таких систем характеризуется целесообразностью, направленной на сохранение гомеостазиса (стабильности). Обеспечение стабильности достигается и осуществляется с помощью обратной связи [1].

Система «неживая» – система, в которой отсутствуют биологические элементы и подсистемы. Функционирование таких систем основывается на законах сохранения природы.

Система социотехническая (социальная, общественная) **–** система, возникающая в обществе. В состав таких систем входят люди или коллективы, интересы (цели) которых могут быть противоречивы. Это обстоятельство является причиной конфликтов в таких системах. Для разрешения конфликтов без разрушения системы необходимо найти способ сосуществования «противоборствующих» сторон.

Гомогенная система – система, состоящая из однородных и слабо связанных частей. Например, муравейник или системы памяти в компьютере, состоящие из однородных элементов.

Гетерогенная система – система, состоящая из отдельных, разнородных частей. Например, муравьи в муравейнике, человеко-машинные системы и др.

Изолированная система изолирована от среды. Поэтому такую систему можно только вообразить, но проверить ее реальность невозможно, так как невозможен опыт, в котором проявилось бы ее существование.

Примером такой системы может служить система материальных тел, взаимодействующих только между собой, но не с окружающей средой.

Закрытая система обменивается с внешней средой только информацией. Примером такой системы могут служить различного рода документы.

Система открытая – система, элементы (подсистемы) которой взаимодействуют с внешней средой, и в этом взаимодействии проявляется свойство *эмерджентности* системы. Открытая система всегда является частью некоторой другой, большей системы.

Простая система – гомогенная система. Например, бригада рабочих, выполняющих одинаковую работу.

Сложная система – система, состоящая из большого числа элементов и подсистем, в которой не хватает информации для эффективного управления, что приводит к неожиданным, непредсказуемым и нежелательным результатам. Например, производственное предприятие, система регулирования уличного движения и др.

Большая система – система, состоящая из большого числа взаимодействующих и многофункциональных элементов, имеющая многорежимный характер функционирования, стохастические связи с внешней средой и иерархическую структуру. Например, экономическая система, глобальная навигационная спутниковая система и т. д.

Статическая система – система, которая не изменяется с течением времени относительно своей цели. Система может быть статической на одном интервале времени и динамической на другом. Например, механическая система в состоянии равновесия является статической, пока на нее не окажут влияние силы, нарушающие равновесие, или живая система в состоянии гомеостазиса – статическая система, пока не возникнут условия нарушающие состояние гомеостазиса.

Динамическая система – система, у которой все или некоторые части изменяются во времени. Например, все живые системы.

Приведенный вариант классификации не является единственным, так как всякая классификация носит целенаправленный характер, в ней отражаются и некоторая условность целей классификации, и неполнота знаний о множестве классифицируемых объектов и другие субъективные и объективные факторы.

1.3. Понятие состояния системы, эволюции состояния

Мощность связей между элементами системы значительно превосходит силу связей с элементами, не входящими в систему, что позволяет выделить систему как целостный объект из внешней (окружающей) среды.

Внешняя среда (окружающая среда) **–** объекты, процессы и явления, не принадлежащие системе, но оказывающие влияние на возможность достижения системой цели. Они не принадлежат системе, но, взаимодействуя с ней, оказывают на нее влияние, вызывая ответную реакцию. Например, Космос – внешняя среда для Солнечной системы, рынок – внешняя среда для системы «завод». Взаимодействие системы с внешней средой осуществляется через входы и выходы системы [1].

Состояния системы – совокупность внутренних свойств, определяющих вместе с входами выходные реакции системы.

Обобщая все сказанное, всякую систему можно представить так, как это показано на рис. 7.

Рис. 7. Графическое представление системы

Состояние системы – это обобщенная характеристика системы. В каждый фиксированный момент времени система находится в определенном состоянии. Состояние системы характеризуется устойчивостью, стабильностью и живучестью.

Живучесть системы – способность системы противостоять действию возмущений, сохранять и восстанавливать (полностью или частично) свои свойства.

Стабильность системы – способность системы восстанавливать и поддерживать устойчивое состояние.

Устойчивость системы – способность системы, находящейся в некотором состоянии, возвращаться в это состояние после незначительных отклонений от него. Условия, при которых система находится в устойчивом состоянии, называют критериями устойчивости.

Для живых систем эти понятия объединяются в понятии «гомеостазис».

Гомеостазис – способность существования живых систем при изменении внешних условий.

При выходе системы из состояния устойчивости и стабильности возникают кризисы, катастрофы и катаклизмы.

Кризис системы – состояние системы, свидетельствующее о необходимости адаптации системы к заметно изменившимся условиям. Он характеризуется сохранением самых важных характеристик, незначительным ущербом элементам и свидетельствует о необходимости обновления системы.

Катастрофа системы – состояние системы, сопровождающееся значительным и резким изменением интегральных свойств системы

Катаклизм системы – разрушение системы и прекращение ее существования.

Состояние системы зависит от внешней среды, оказывающей влияние на состояние системы через входы системы, и от предшествующих ее состояний. Это значит, что система не может изменять свое состояние произвольно. Механизм изменения системой своего состояния связан с принципом Ле Шателье – Брауна, согласно которому любое внешнее воздействие на систему порождает ответную реакцию системы, направленную на ослабление влияния этого воздействия с целью сохранения своей устойчивости и стабильности. Каждому состоянию системы соответствует вполне определенное влияние внешних сил и предшествующих состояний системы [5].

В общем случае состояние системы есть функция времени. Существуют различные способы определения состояний [6, 7, 29, 30, 31].

1. Состояние системы определяется как совокупность состояний ее частей. Если каждая часть системы может находиться только в M состоя-

ниях, то система из N частей может иметь M^N состояний. Например, пусть система состоит из $N = 4$ элементов, каждый из которых может быть в $M = 2$ состояниях (например, быть исправным или неисправным). Такая система может находиться в одном из $2^4 = 16$ состояний.

2. Состояние системы определяется некоторым набором целых неотрицательных чисел (r = 0, 1, 2, 3…). Например, состояние системы контроля качества выпускаемой продукции определяется количеством контролируемых изделий, находящихся в системе, или состояние системы производства характеризуется количеством единиц выпускаемой продукции.

3. Состояние системы определяется набором действительных чисел. Например, состояние системы «положение спутника в пространстве» определяется тремя координатами или состояние экономической системы можно определить натуральными, денежными, социальными и другими числовыми показателями.

В общем случае считается, что состояние системы описывается набором характеристик $q_i (i = 1, 2, \ldots, k)$, каждая из которых принимает значения из некоторого множества Q_i , т. е. $q_i \in Q_i$. Следовательно, множество S возможных состояний системы определяется как прямое произведение множеств Q_i:

$$S = Q_1 \times Q_2 \times \ldots \times Q_k. \tag{1}$$

Это множество называют пространством состояний системы, или фазовым пространством. Каждая точка этого пространства называется *фазовой точкой*, она характеризует одно из состояний системы, ее координаты $(q_1, q_2, \ldots, q_k)$. Трехмерное эвклидово пространство – аналог фазового пространства в естественной физической интерпретации.

Пусть, например, состояние системы определяется двумя характеристиками:

$$x \in [a, b], \quad y \in [c, d]. \tag{2}$$

Пространство состояний системы с такими характеристиками (свойствами) есть прямоугольник, показанный на рис. 8.

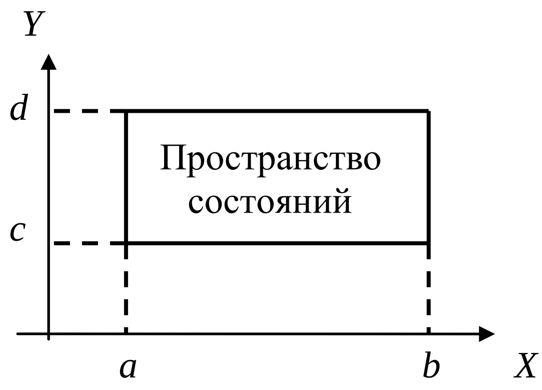


Рис. 8. Пространство состояний

Если свойства удовлетворяют неравенствам

$$x \le |a \cdot \cos t| \qquad y \le |a \cdot \sin t|, \quad t \in [0, 2\pi], \tag{3}$$

то, исключая параметр t из этих уравнений, найдем пространство состояний системы:

$$x^2 + y^2 \le a^2, \tag{4}$$

которое, очевидно, является кругом радиуса a. Предположим, что одно из свойств объекта изменилось. Пусть, например,

$$y \le |b \cdot \sin t|. \tag{5}$$

В результате этого изменяется и фазовое пространство:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} \le 1, \tag{6}$$

которое теперь есть область, ограниченная эллипсом. Полезно заметить, что в рассмотренном примере фазовое пространство – замкнутая область, в то время как уравнения характеристик состояния таковыми не являются.

Начальное состояние системы в пространстве состояний – это фазовая точка, координаты которой определяются начальными значениями характеристик (свойств) системы. Изменение состояния системы будет сопровождаться изменением хотя бы одной из фазовых координат и, вследствие этого, фазовая точка изменяет свое положение в фазовом пространстве, двигаясь по некоторой траектории. Эту траекторию называют фазовой траекторией. Например, пусть состояние системы определяется двумя характеристиками состояния:

$$x(a,t) := -a \cdot \cos(t); \qquad x1(a,t) := -a^{\frac{1}{2}} \cdot \sin(2 \cdot t);$$

$$y(b,t) := -b \cdot \cos(2 \cdot t); \qquad y1(b,t) := -b \cdot \cos(t) - 2.$$

Пространством состояний служит прямоугольник: $x \in [-a, a]$; $y \in [-b, b]$. Каждой паре свойств соответствует своя фазовая траектория, показанная на рис. 9, где a = 3, b = 3.

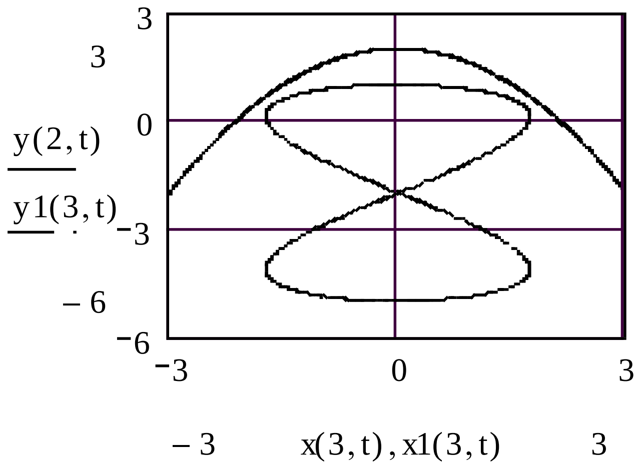


Рис. 9. Фазовые траектории системы в фазовом пространстве

В общем случае фазовых траекторий в фазовом пространстве может быть неограниченно много. Они позволяют получить представление о качественной картине изменения состояний системы, например, выявить непрерывные фазовые траектории, соответствующие плавному, эволюционному характеру изменения состояния системы, или найти области, где фазовая траектория разветвляется, свойства системы изменяются скачкообразно, что может приводить к непредсказуемым, опасным и нежелательным изменениям состояния системы.

1.4. Признаки существования системы

Не всякую совокупность элементов, взаимосвязанных друг с другом можно считать системой [1, 5]. Существуют критерии (признаки), по которым можно установить, что рассматриваемый объект в действительности является системой.

1. Всякая система обладает ***целостностью***, обособленностью от окружающей среды. Целостность системы не означает ее однородности и неделимости: в системе всегда различают составные части. Разделимость системы на части не означает, что эти части полностью изолированы друг от друга. Они образуют целое благодаря связям между ними.

2. Цельность системы обусловлена тем, что система как целое обладает такими свойствами, которых нет и не может быть у составляющих ее частей. Свойства системы не сводятся к свойствам ее частей, не являются простым объединением этих свойств. Свойство системы как целого проявляется в ее взаимодействии с окружающей средой (т. е. реализуется через внешние связи как функция системы), но само это свойство возникает и может существовать лишь благодаря взаимодействию частей (т. е. благодаря внутренним связям, структуре системы). Система – не простое объединение составляющих ее частей. Она обладает свойством эмерджентности, которого нет у составляющих ее частей.

3. Все части системы взаимосвязаны, и поэтому воздействие на какую-либо часть системы проявляется в изменении состояния других частей. Изъятие из системы какой-либо ее части влечет утрату и самой системой и изымаемой частью некоторых существенных свойств. В этом проявляется внутренняя целостность систем. Недопустимо без нарушения свойства целостности системы рассматривать части системы вне их взаимодействия с другими частями.

4. Система связана со средой, существует в ней, взаимодействует с ней, обменивается со средой веществом, энергией, информацией. Это значит, что все системы открыты, т. е. изолированных (замкнутых) от среды систем не существует. Замкнутые (изолированные) системы являются абстракцией, доказать их существование невозможно из-за невозможности взаимодействия с ними. Закрытые системы обмениваются со средой только информацией.

5. Открытость системы, ее связанность со средой означает, что она (система) является частью другой, бóльшей системы. В этом проявляется внешняя целостность систем. В результате мир представляется как система вложенных друг в друга, перекрывающихся частично или полностью, или разделенных, но взаимодействующих систем.

6. Внутренняя и внешняя целостность систем обобщаются, объединяются, синтезируются в понятии ***цели***, которая как бы диктует ***и структуру, и функцию*** системы. ***Функция системы*** интерпретируется, как проявление целеустремленности системы, ***структура системы*** выступает при этом как вариант реализации цели.

7. В результате внутренних и внешних взаимодействий системы изменяются. Многообразие процессов в системах велико. Многие явления в системах невозможно понять без учета их изменчивости (динамики).

Контрольные вопросы

1. Что такое системный анализ?
2. Что такое системный синтез?
3. Что такое проблемная ситуация?
4. Сформулировать понятие цели.
5. Сформулировать основные задачи структурно-целевого подхода.
6. Что такое система?
7. Что такое состав системы. Привести примеры.
8. Дать определение структуры системы. Привести примеры.
9. Дать определение целостности системы. В чем проявляется внешняя и внутренняя целостность системы?
10. Объяснить сущность понятия эмерджентности системы. Привести примеры.
11. Привести классификацию систем.
12. Перечислить признаки существования системы.
13. Что входит в понятие состояния системы? Привести примеры.
14. Что такое эволюция состояний?

2. ОБОБЩЕННАЯ СТРУКТУРА И СОДЕРЖАНИЕ СИСТЕМНО-ЦЕЛЕВОГО ПОДХОДА

Системно-целевой подход применим к системам любой природы и может рассматриваться как некоторая система, цель которой разработка рекомендаций для полной или частичной ликвидации проблемы. С этих позиций он имеет составные части, структуру, связи с внешней средой и состояния.

Практическое применение системно-целевого подхода основывается на следующих положениях [5, 32].

1. Каждый элемент системы рассматривается с учетом его роли и места в системе. Роль и место элемента в системе определяется его влиянием на достижение цели.

2. Всякую систему необходимо рассматривать во взаимосвязи с окружающей средой. Окружающую среду целесообразно разделить на ближнюю и дальнюю. Связи системы с ближней средой всегда более существенно влияют на достижение системой своих целей.

3. В каждой системе необходимо выявить, каким образом из свойств частей системы возникает свойство эмерджентности (внешняя целостность системы) и какими свойствами должны обладать части системы, чтобы реализовалось необходимое свойство целостности (внутренняя целостность системы).

4. В процессе исследования учитывать не все свойства системы, а только необходимые для достижения цели.

5. Для каждой системы необходимо исследовать условия устойчивости к внешним и внутренним возмущениям.

6. Основным аппаратом системно-целевого исследования служит метод моделирования, который позволяет представить себе функционирование будущей системы, сопоставить ее возможности с имеющимися ресурсами, сформулировать цепочку решений, в результате которых будут достигнуты цели.

Основные трудности при реализации системно-целевого подхода: необходимость учета большого числа факторов, влияющих на возможность достижения цели; недостаток и низкое качество исходной информации; ряд неопределенностей, обусловленных и поведением систем и отсутствием знаний об их поведении.

2.1. Обобщенная структура системно-целевого подхода

Обобщенная структура системно-целевого подхода соответствует известной структуре процесса познания и преобразования действительности: «наблюдение» – «абстрактное мышление» – «практика». На рис. 10 показан один из возможных вариантов представления состава и структуры системно-целевого подхода [8, 9].

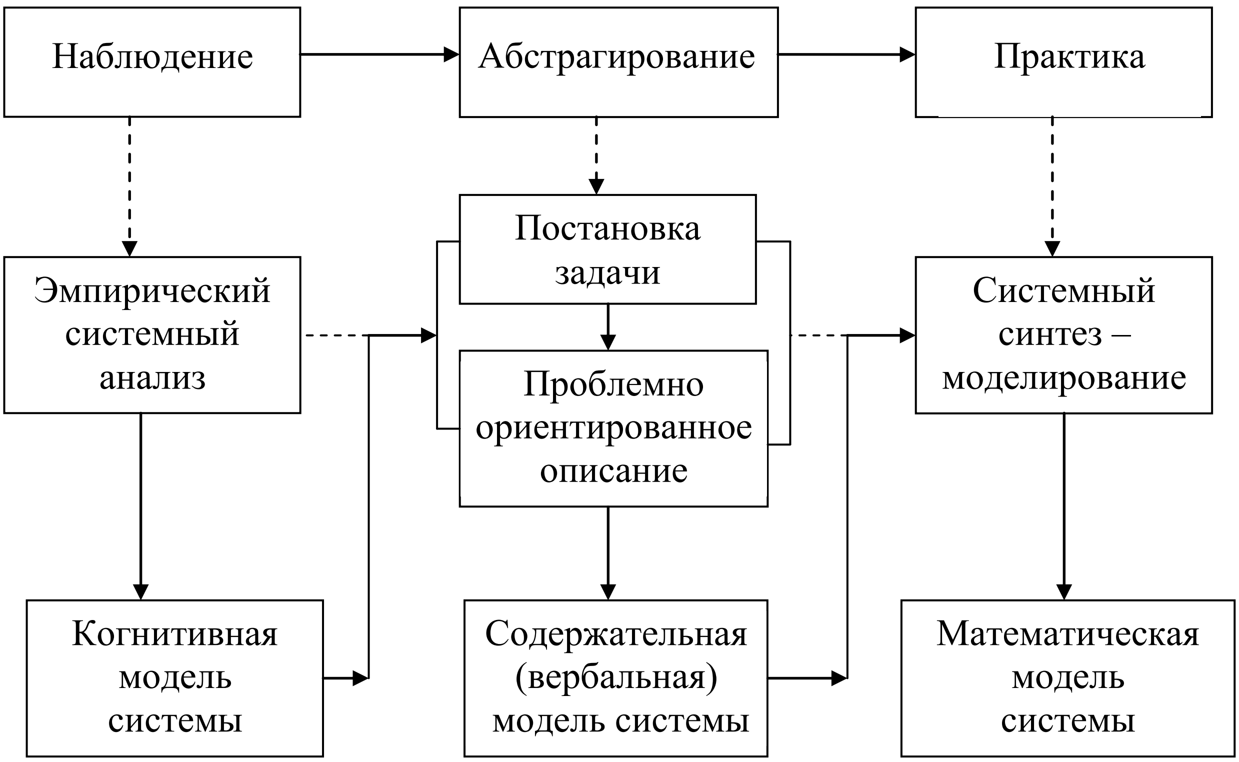


Рис. 10. Структура системно-целевого подхода

На этом рисунке три блока в верхней строке соответствуют трем этапам познания объективной реальности. Каждый блок этой строки пунктирной стрелкой соединен с соответствующим ему блоком системно-целевого подхода: «Эмпирический системный анализ» – «Постановка задачи и проблемно ориентированное описание исследуемой системы» – «Системный синтез – моделирование». Эти три блока и определяют состав системно-целевого подхода. Три блока в нижней строке обозначают результат работы каждого блока системно-целевого подхода. Стрелками на рис. 10 указаны связи между блоками.

Технология системно-целевого подхода имеет итеративный характер, т. е. при обнаружении в процессе исследования новой проблемы, противоречий или возникновении непредвиденных неопределенностей процедура системного исследования может возвращаться к ранее выполненным этапам.

Связи с внешней средой в процедуре системно-целевого подхода осуществляются через входы и выходы. Входы в процедуру системно-целевого подхода – это факты о состоянии системы в прошлом и настоящем, ее эволюции, их соответствии или не соответствии предъявляемым требованиям, а выходы – результат, характеризующий достижение или не достижение требуемой цели.

Рассмотрим содержание каждого этапа системно-целевого подхода.

2.2. Эмпирический системный анализ

Результаты эмпирического системного анализа должны обеспечить адекватность отображения реальности, необходимую для дальнейшего исследования. Плохо организованный и выполненный эмпирический системный анализ приводит к ошибочности результатов, полученных при реализации системно-целевого подхода.

В результате эмпирического системного анализа необходимо получить представление о сложности изучаемой системы, о возможных процессах во внешней среде, которые могут служить причинами несоответствия состояния системы, предъявляемым требованиям (имеющем место противоречии), о том, какие действия необходимо предпринять для ликвидации противоречия.

Поэтому эмпирический системный анализ направлен на решение следующих задач.

1. Выявление противоречия, которое необходимо ликвидировать и определение цели или целей исследования.

2. Описание факторов внешней среды, оказывающих наиболее существенное влияние на достижение цели.

3. Формулирование когнитивной модели.

Рассмотрим кратко содержание каждой задачи.

1. Выявление противоречия, которое необходимо ликвидировать и определение цели или целей исследования.

Всякое противоречие в системах возникает в результате внутренних процессов или взаимодействия с окружающей средой. В окружающей среде рано или поздно возникают неустойчивые неравновесные состояния (точки бифуркации). В техносфере возникновение неравновесного состояния чаще всего обусловлено прогрессом научно-технического и социально-общественного развития и связано с необходимостью изменений в структуре системы, условиях ее функционирования и с нарушениями технологии взаимодействия частей системы между собой и с внешней средой. Поэтому при выявлении противоречий необходимо иметь информацию об имеющей место ситуации. Такую информацию получают в результате сбора статистических сведений о прошлом и настоящем изучаемой системы, о системах, аналогичных данной. Полученная информация является основой для проведения эмпирического системного анализа и по существу эмпирический системный анализ – это целенаправленное изучение и преобразование прошлого. В результате проведения эмпирического системного анализа выявляют свойства, которые наиболее часто проявляются в анализируемых данных, изучают несоответствие имеющей место ситуации предъявляемым требованиям, определяют предпосылки возникновения противоречия, выявляют противоречия в наблюдаемой ситуации (несоответствие между необходимым и существующим состоянием) и формулируют проблему. В тех случаях, когда имеющихся данных недостаточно, для выявления противоречия могут быть использованы экспертные методы, основанные на интуиции и опыте человека.

Например, в процессе исторического развития человечество постоянно сталкивается с проблемой, которая проявляется в противоречии между фактически существующей производительностью труда и необходимостью ее увеличения для удовлетворения потребностей социально-технического развития. Результатом разрешения этой проблемы является все более высокая организация труда: от ручного труда к механизации, от механизации к автоматизации и от автоматизации к кибернетизации [1].

Уяснение и выявление существующей проблемы позволяет наметить цель исследования. Цель может быть не одна, их может быть несколько. Например, в производственных системах можно добиваться повышения производительности труда, снижения себестоимости продукции и повышения качества продукции. В сфере безопасности производства можно требовать снижения риска техногенных катастроф, уменьшения аварийности и травматизма на производстве, увеличения надежности технического оборудования и т. д. В общем случае цели определяются, исходя из потребностей практики, тенденций развития науки и техники и социально-экономической целесообразности. В настоящее время не существует формальных правил для выявления противоречия и выбора целей. Решение этих задач осуществляется на основании статистического материала о ситуациях, подобных исследуемой, понимания необходимости ликвидации проблемы, инженерного опыта и интуиции человека.

2. Описание факторов внешней среды, оказывающих наиболее существенное влияние на достижение цели.

Система как целое обладает свойствами, которых нет, и не может быть у составляющих ее частей. Эти свойства проявляются в ее взаимодействии с окружающей средой, т. е. реализуются через внешние связи как функция системы. Одной из причин возникновения проблемы является изменение состояния внешней среды и нарушение технологии взаимодействия системы с внешней средой. Множество факторов – предпосылок во внешней среде, оказывающих влияние на состояние системы, достаточно велико, но влияние их неодинаковое. Поэтому среди предпосылок необходимо выделить те, влияние которых наиболее существенно для достижения цели – ликвидации проблемы, и дать описание зависимости изменения ситуации от них. Для описания факторов внешней среды могут быть использованы статистические данные о прошлом и настоящем системы, результаты специально организованных экспериментов, опыт эксплуатации аналогичных систем и любые другие полезные сведения. Эти данные в процессе предварительного анализа могут обобщаться методами интерполяции, аппроксимации, статистического анализа и другими. Все ограничения на использование информации о связи системы с внешней средой необходимо четко указать, так как все полученные результаты будут справедливы только при соблюдении таких ограничений.

3. Формулирование когнитивной модели.

Информация, полученная на этапе эмпирического системного анализа, изучается и обобщается человеком. Из информации извлекаются знания. В результате человек формулирует представление о том, что необходимо предпринять для устранения проблемы. Такое представление называют когнитивной моделью[1]. Вследствие профессиональной ограниченности человека, его психологической инерции, темперамента, образования и культуры результаты такого обобщения субъективны, и поэтому могут в какой-то мере оказаться ошибочными. Однако эти ошибки удается выявить на последующих этапах системно-целевого подхода.

После создания когнитивной модели переходят ко второму этапу системного анализа – проблемно ориентированному описанию системы и постановке задачи исследования.

2.3. Постановка задачи системного анализа

Наиболее важный и ответственный этап системного анализа – постановка задачи (формулирование проблемной ситуации), т. е. выявление противоречия и всех возможных последствий его разрешения. Процесс постановки задачи идет практически непрерывно вплоть до получения решения, так как новая информация, получаемая в ходе решения, позволяет уточнить и изменить сформулированную когнитивную модель и первоначально поставленную задачу. Чтобы правильно и эффективно поставить задачу, необходимо представлять, из каких элементов состоит задача.

Возникновение проблемы в какой-либо области свидетельствует о существовании противоречия, которое необходимо устранить. Но предпринимать какие-либо действия по устранению противоречия возможно только с разрешения руководителя, ответственного за состояние дел в рассматриваемой области и имеющего желание достичь целей, разрешающих существующую проблему. Этого руководителя принято называть «лицо, принимающее решение» (ЛПР). Всякая система существует в некоторой

[1] Когнитология – наука об извлечении знаний из информации.

среде и обменивается с ней веществом, энергией и информацией. Такой обмен практически всегда сопряжен с различного рода условиями и ограничениями, которые необходимо уяснить и сформулировать при постановке задачи. И, наконец, должны существовать возможность выбора между различными стратегиями достижения целей и способ сравнения этих стратегий между собой.

Таким образом, имеем шесть основных элементов, составляющих задачу: лицо, принимающее решение о необходимости устранения противоречия; цели, достижение которых разрешает проблему, внешняя среда с ограниченными ресурсами; стратегии достижения целей; критерии и процедуры выбора стратегий из множества альтернатив. Все они обязаны быть принятыми во внимание при постановке задачи. Рассмотрим подробнее каждый из этих элементов и выясним, что нужно знать о каждом из них при постановке задачи.

Лицо, принимающее решение. Прежде всего, необходимо установить, кто обладает полномочиями для принятия, отмены или изменения необходимых решений. Если это коллектив, то необходимо определить его структуру, полномочия и интересы каждого члена коллектива. Необходимо также установить способ принятия решений: формально или нет, совместно или в определенной последовательности, большинством голосов или нет, обладает ли кто-нибудь правом «вето», кто и как оценивает предпринимаемые действия, кто несет ответственность за выполнение принятых решений, какие новые проблемы возможны после изменений, вызванных принятыми решениями, и т. д.

Цели, достижение которых разрешает проблему, формулируются, исходя из желаемых результатов принимаемых решений. Выявление целей превращает проблему в задачу выбора подходящих средств для их достижения. Правильно наметить цели удается не всегда и не сразу, они также могут изменяться со временем и при необходимости заменяться другими. Чтобы предвидеть возникновение новых проблем, в список целей включают не только желаемые (полезные), но и нежелательные (вредные) по своим последствиям цели, а также ранее достигнутые цели, которые необходимо сохранить в будущем. Например, цели, которые необходимо достигнуть – уменьшить производственные расходы, повысить надежность и безопасность производства, цели которые надо сохранить – сохранить ведущую роль в данной области производства, сохранить (не увеличить) уровень риска возникновения аварий на производстве. При определении целей системы необходимо принимать во внимание интересы всех заинтересованных участников, например, потребителей и конкурентов.

Внешняя среда – объекты, процессы и явления, не принадлежащие системе, но оказывающие влияние на возможность достижения системой цели. Они не принадлежат системе, но, взаимодействуя с ней через входы и выходы системы, оказывают на нее влияние, вызывая ответную реакцию. Входы и выходы системы – связь системы с окружающей средой, обеспечивающая передачу в систему и из системы вещества, энергии или информации. Основное ограничение, которое накладывается на связи среды и системы состоит в *невозможности действий, противоречащих законам природы, общества и целям системы.*

Ограничения могут быть субъективными и объективными. Субъективные ограничения можно изменять по нашему усмотрению, а объективные, связанные с законами природы, изменять нельзя. Например, объективным ограничением может служить невозможность вывоза со склада количества продуктов, превышающего их запас на складе, а субъективным ограничением – режим вывоза продуктов за указанное время (время определяется субъективно). Практически всегда необходимо рассматривать ограничения по ресурсам. В качестве ресурсов рассматривают вещество, энергию, информацию, пространство, время, кадры, фирмы и организации.

Стратегии достижения целей – возможные варианты получения результата. Например, пусть требуется составить спортивную команду из m человек при возможности выбора из n человек. Вариантов состава команды существует C_n^m, при условии, что все члены команды играют одинаковую роль. Спрашивается, какой вариант из этого множества следует выбрать? Для ответа на такой вопрос используют формальные методы и неформальные процедуры. К формальным методам определения стратегий относятся математические методы, патентный поиск, изучение специальной литературы, изучение опыта решения сходных проблем в других областях и др. Неформальные процедуры определения стратегий основаны на опыте и интуиции человека. К таким процедурам относятся метод экспертных оценок, интервьюирование заинтересованных лиц, метод мозгового штурма, метод ассоциативного мышления и т. д. Часто формальные и неформальные методы определения стратегий используются совместно. При рассмотрении конфликтных ситуаций необходимо определять стратегии для каждого участника конфликта. Стратегии противоборствующей стороны называют контрстратегиями. Внешняя среда (в частности природа) может рассматриваться как участник конфликта, и если не предусмотреть это, то выбранная стратегия может оказаться гибельной.

В некоторых случаях все существующие стратегии оказываются непригодными для достижения целей. В таких случаях приходится специально разрабатывать и оценивать новые стратегии. Список стратегий необходимо проанализировать, чтобы исключить из него заведомо невозможные, при реализации которых нарушаются поставленные ограничения по ресурсам, или которые не позволяют достигнуть некоторых из целей. Все исключенные стратегии необходимо зафиксировать с указанием причины исключения, чтобы сохранялась возможность в будущем устранить эти причины и может быть использовать эти стратегии.

Критерии и процедуры оценки стратегий достижения целей – это средство для сравнения различных стратегий и выбора наиболее полезной стратегии. Существует много различных способов сравнения и оценки стратегий. Наиболее распространенными являются: сравнение стратегий по значению некоторого критерия или системы критериев; язык бинарных отношений, основанный на сравнении всевозможных пар стратегий, и язык функций выбора [1].

Определение значения критериев, для какой-либо стратегии является оценкой ее пригодности для достижения цели. Поэтому только после определения критерия появляется возможность ставить задачи выбора и оптимизации. Однако при этом может не существовать единственной стратегии, удовлетворяющей всем критериям. В такой ситуации на систему критериев накладывают различного рода эвристические ограничения или все критерии связываются в один критерий, который и используется для оценки меры общей эффективности предпринимаемых действий и выбора лучшей стратегии.

При использовании языка бинарных отношений отдельная альтернатива не оценивается, а для каждой пары альтернатив устанавливается, какая из них предпочтительнее.

Язык функций выбора является наиболее общим методом оценки стратегий. Он позволяет получить решение в тех случаях, когда язык критериев и бинарных отношений применить не представляется возможным. Язык функций выбора позволяет рассматривать более сложные правила выбора, чем язык бинарных отношений или язык критериев [1].

Выбор стратегий из множества альтернатив – процедура принятия решения. Основанием для выбора служит множество стратегий достижения целей и методы сравнения и оценки стратегий достижения целей. Результат выбора существенно зависит от условий выполнения процедуры выбора. К условиям выбора обычно относят [1]:

– выбор в заданных условиях;
– выбор в условиях неопределенности и риска;
– выбор в условиях конфликта;
– режим выбора – однократный или повторяющийся;
– известны ли последствия выбора точно или нет;
– ответственность за выбор (индивидуальная или групповая);
– согласованность целей при многостороннем выборе;
– количество альтернатив, предъявленных для выбора;
– важность целей;
– функция распределения вероятности достижения цели для каждой стратегии;
– метод сравнения и оценки стратегий.

Рассмотрим пример.

Большинство естественных и искусственных систем включает людей и производственное оборудование, взаимодействующих в некоторой среде. Такие системы называют человеко-машинными системами (ЧМС) [5]. Они состоят из следующих элементов:

– человек (Ч);
– машина (М);
– производственная среда (С) – область пространства, где осуществляется производственный процесс.

Эти элементы взаимодействуют между собой соответственно целям системы по правилам, которые определены технологией (Т) производственного процесса. В результате этого взаимодействия проявляется целостность системы – свойство эмерджентности системы – способность производить некоторый продукт. Этим свойством не обладает ни один из элементов системы, взятый в отдельности, или даже их объединение без учета взаимосвязей. Данная система не изолирована от окружающей среды. Внешняя среда (ВС) воздействует на систему через ее входы $I(t)$, накладывая ограничения объективного и субъективного характера. В каждый момент времени система находится в некотором состоянии $S(t)$, которое изменяется в результате функционирования и развития системы и влияния внешней среды. Через свои выходы $E(t)$ система в свою очередь оказывает полезное или вредное влияние на окружающую среду. В ЧМС возникают разнообразные проблемы, связанные с безопасностью их функционирования, экономической эффективностью и др. Пусть требуется разрешить проблему снижения риска возникновения аварийных ситуаций.

Принимая во внимание изложенные обстоятельства, рассматриваемую систему представляют в виде схемы (рис. 11).

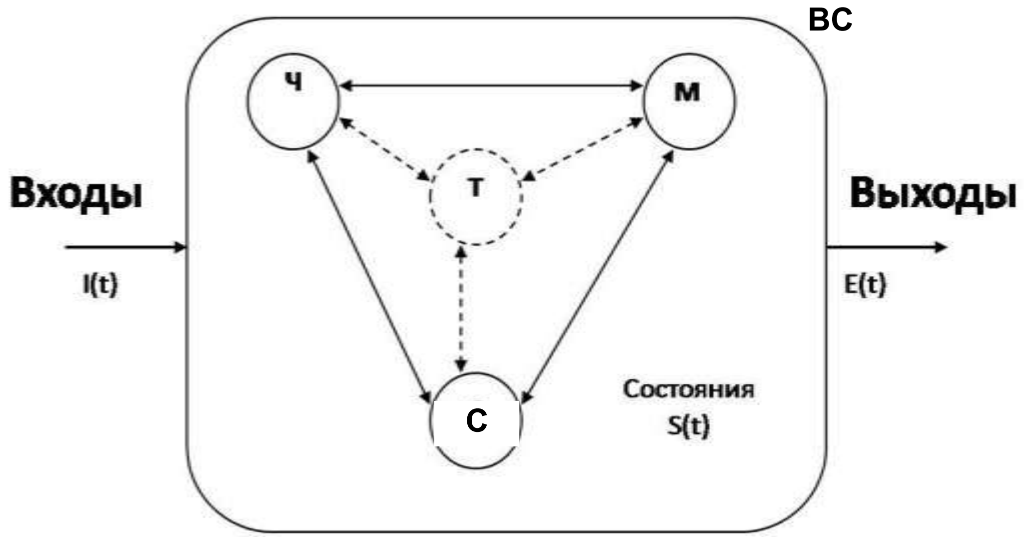


Рис. 11. Схема системы

Элементы этой системы при необходимости можно рассматривать как подсистемы данной системы.

Здесь ЛПР – персонал, управляющий системой; цели – повышение безопасности и снижение риска; внешняя среда – все, что не принадлежит системе, но влияет на достижение цели. Источником опасности служат или персонал (Ч), управляющий процессом, или машины (М), применяемые при выполнении процесса, или технологии (Т), используемые в производственном процессе, или процессы, возникающие в производственной среде (С), или процессы, возникающие во внешней среде (ВС).

Следовательно, стратегиями достижения цели служат мероприятия, выполнение которых снижает риск возникновения аварийных ситуаций по вышеперечисленным причинам, а также их корректные комбинации. Учитывая, что источников аварийности всего 5, встречаться они в принципе могут в любых сочетаниях, и каждый из них может или быть или не быть, общее число *M* стратегий можно вычислить по известной формуле комбинаторики:

$$M = \sum_{i=1}^{5} C_5^i = 2^5 - 1 = 31. \tag{7}$$

Необходимо проанализировать возможность реализации этих стратегий и исключить заведомо малоэффективные и те, возникновение которых невозможно по объективным причинам в принципе. После этого требуется сравнить стратегии между собой по эффективности достижения целей. В качестве критериев эффективности стратегий выберем вероятность возникновения аварийной ситуации и ущерб от аварии. Таким образом, задача сводится к оценке каждой из реализуемых стратегий (их не более 31) по этим двум критериям и выбору наиболее эффективной из них.

Процедуры выбора стратегий рассматриваются в разделе теории принятия решений.

После формулировки проблемы обычно приступают к определению одного или нескольких вариантов решения, которые заслуживают дальнейшего рассмотрения. Исследование эффективности различных вариантов выполняют на математической модели системы, в которой отображаются лишь свойства и отношения элементов системы между собой и с внешней средой, необходимые для достижения целей.

2.4. Проблемно ориентированное описание

Проблемно ориентированное описание предназначено для более четкой, чем на этапе эмпирического системного анализа, формулировки проблемы и определения целей исследования. Для этого необходимо объединить усилия специалистов различных профилей, унифицировать и согласовать разнообразную конкретную информацию, применяя экспериментальные, эмпирические и строгие математические методы, а также и неформальные эвристические процедуры, основанные на опыте и интуиции человека. Соответственно этому осуществляется описание системы с различных точек зрения. Перечислим основные направления реализации описания систем в системно-целевом подходе.

Морфологическое описание является средством формализации строения системы. Оно дает возможность четко отобразить структуру системы, выявить части системы, образующие единый комплекс, отобразить в наглядной форме связи между частями системы и их логическую последовательность, оценить степень важности отдельных частей и решить многие другие задачи. В нем отображаются состав системы и отношения элементов и подсистем между собой и с внешней средой. Оно не может быть исчерпывающим, так как зависит от глубины описания, уровня детализации, от учитываемых связей между главными подсистемами, между второстепенными подсистемами, между элементами, от типа связей (прямая или обратная связь), характера связей (позитивная, негативная). В значительной мере оно определяется назначением системы и целью исследования. Изучение морфологии начинается с элементного состава, т. е. с определения частей системы, рассматриваемых как неделимые. В состав системы включают элементы, необходимые для достижения цели. Морфологические свойства системы существенно зависят от внутренних связей между элементами системы и связей с внешней средой. *Обычно выделяют информационные, энергетические, вещественные связи.* В результате морфологического описания возникает описание *структуры* системы.

Наглядное представление морфологии системы дает граф, множество вершин которого – это части системы, а ребра или дуги графа – связи между частями системы. Схему ЧМС, представленную на рис. 11, можно интерпретировать и как морфологическое описание этой системы в виде графа. Формальное морфологическое описание осуществляется с помощью формальных методов представления графов: матриц смежности, матриц инцидентности, списков ребер, структур смежности и др.

Например, морфологическое описание ЧМС зададим матрицей смежности:

$$M = \begin{pmatrix} 0 & "Ч" & "М" & "С" & "Т" & "ВС" \\ "Ч" & 0 & 1 & 1 & 1 & 1 \\ "М" & 1 & 0 & 1 & 1 & 1 \\ "С" & 1 & 1 & 0 & 1 & 1 \\ "Т" & 1 & 1 & 1 & 0 & 1 \\ "ВС" & 1 & 1 & 1 & 1 & 0 \end{pmatrix}.$$

В первой колонке и первой строке этой матрицы вставлены имена вершин графа (имена частей ЧМС). Нулем отмечены части системы, формально не связанные между собой (части ЧМС не связаны сами с собой), а единицами – части, взаимодействующие между собой.

Задача морфологического описания системы состоит в том, чтобы получить заключение о структурных свойствах системы и ее подсистем. Для этого, в первую очередь, устанавливают наличие связи между элементами и направление связи. Исходя из наличия связей между элементами, определяется связность (целостность) системы, выявление циклов, определение длины пути между элементами. Если оказывается, что система не является связной, то ставится задача нахождения изолированных связных подсистем со списком входящих в них элементов. Данные о направлении связей между частями системы позволяют выполнять топологическую декомпозицию – выделение сильно связанных подсистем[2], выявление уровней в структуре[3] и определение их взаимосвязи и другие задачи.

Функциональное описание системы определяет роль и значение системы по отношению к другим системам и внешней среде. Функциональное описание системы основывается на следующих положениях [6, 29, 30, 31].

1. Любая система функционирует во времени, взаимодействует с внешней средой и находится в одном из возможных состояний. В начальный момент $t = t_0$ система находится в начальном состоянии $S_0 = S(t_0)$.

[2] Граф называется сильно связанным, если между его двумя любыми вершинами *i*, *j* существует путь – непрерывная последовательность дуг.

[3] Уровнем в графе называют непересекающиеся подмножества вершин, упорядоченные так, что если вершина входит в подмножество с номером *i*, то следующая за ней вершина – в подмножество с номером, большим *i*.

2. Состояние системы в фиксированный момент времени определяется предшествующими состояниями и входными (из внешней среды) сигналами, поступившими в данный момент и ранее.

3. Состояние системы и входные сигналы, относящиеся к данному и предшествующим моментам, определяют выходной сигнал системы.

Если состояние системы в момент времени $t > t_0$ не зависит от того, каким образом система пришла в состояние $S(t)$, то систему называют системой без последействия.

Для описания функционирования системы без последействия необходимо определить правила перехода системы из состояния в состояние и выходы системы в заданный момент времени t, если заданы состояния системы и входные сигналы в моменты времени, предшествующие моменту t.

Информационное описание дает представление о потоках информации в системе и характеризует организацию системы. Информация в системе представляется файлами. Функционирование системы генерирует одни файлы на основании других, т. е. происходит движение информации с целью получения результата. Все файлы подразделяют на входные, поступающие из внешней среды, выходные, представляющие результаты, и промежуточные – результаты переработки входной информации.

Входные и выходные файлы образуют информационный базис системы, который определяется функциями системы. Между файлами в системе определяют отношения включения и порядка. Отношение включения определяет, какие файлы необходимы для образования данного конкретного файла, а отношение порядка – последовательность образования файлов. Если файлам сопоставить вершины графа, а дугам отношения включения и порядка, то получим структуру, которая называется информационным графом, отражающим информационное взаимодействие элементов системы с внешней средой или между собой.

Анализ информационного графа системы позволяет получать дополнительную информацию о системе, извлекать новые знания о ней, решать информационно-логические задачи. Постановка и решение информационно-логических задач позволяет выявлять информационные и причинно-следственные связи в системе, проводить аналогии, выявлять дублирующие связи и избыточные элементы и т. д.

Объединение морфологического, функционального и информационного описания осуществляется в содержательной (вербальной) модели.

Кроме морфологического, функционального и информационного, в системно-целевом подходе существуют, например, системное, конструктивное, комплексное, проблемное, ситуационное, инновационное, нормативное, целевое, деятельностное, программно-целевое описание систем. Выбор описания зависит от того, в какой предметной области находится та или иная проблема, силами каких специалистов она будет решена.

2.5. Структура и этапы системного анализа и системного синтеза

Как методология решения сложных проблем и задач, системный анализ и системный синтез предполагает определенную последовательность операций: определение проблемы, поиск метода ее решения и само решение.

Структура системного анализа, приведенная на рис. 12, включает ряд этапов, последовательное решение которых приводит к достижению цели – решению проблемы [5].

Рис. 12. Структура системного анализа

1. На начальном этапе системного анализа проблемы или объекта нужно осознать саму проблему. Кроме того, необходимо определить всю цепочку целей: целей исследования, целей системы и т. д. Также следует выяснить целесообразность применения именно системного подхода к решению поставленной проблемы.

Общепризнано, что специалисты в конкретной области знания лучше всех из практического опыта знают, как изменить состояние и поведение конкретной системы: например, системные администраторы должны лучше всех знать поведение своего конкретного объекта – компьютерной сети. Однако без навыков системного анализа в сложных случаях трудно учесть все факторы, которые влияют на конечный результат. В результате задачи ставятся, как правило, в общих выражениях, либо, наоборот, формулируются очень узко. Полученные решения оказываются очень далекими от оптимальных, а зачастую и неправильными.

2. На следующем этапе системного анализа производят идентификацию соответствующих объектов или процессов и уточнение цели их системного исследования. Формулируется концептуальная система, осознается и обеспечивается ее целостность.

3. Производят сбор и обработку информации. Устанавливают границы системы. Иногда они очевидны, но чаще на практике приходится сталкиваться с условностями. Определяют внешнюю (окружающую) среду системы. В данном случае речь идет об установлении границ метасистемы, в которую изучаемая система входит как структурированная подсистема (элемент большой системы). Выявляют взаимодействие системы с окружающей средой и устанавливают ее связи с другими системами. Осуществляют декомпозицию цели на задачи и выбор способа решения каждой из этих задач. Выполняется декомпозиция системы для определения состава и структуры системы. Выявляются все основные отношения и связи между элементами и группами элементов. Описание интересующих аспектов системы дается в словесной форме и графически – различными схемами, диаграммами, рисунками и т. д. При этом необходимо сохранить осмысленность каждого выделенного элемента и системы в целом. Нужен разумный критерий декомпозиции, чтобы не пойти по пути полного перебора возможных вариантов.

Проводят предварительное исследование свойств системы и ее элементов. Осуществляется выбор переменных внутри системы, совокупность значений которых в каждый данный момент времени определяет состояние системы, в том числе и переменных на выходе системы. Выбор внутренних, входных и выходных переменных, естественно, определяется целью и характером решаемой задачи. Следует как можно точнее установить круг необходимых переменных для конкретной задачи, установить единицы их измерения, необходимую их точность.

Далее уточняется описание реакции системы (как внутри, так и выходе) на предполагаемые внешние воздействия, которые необходимо изучить и смоделировать. Изучается имеющаяся информация о системе: определяют ее достаточность для решения поставленной проблемы и необходимость дополнения ее экспериментальными данными.

4. На данном этапе выполняют разработку моделей выбранных объектов и процессов, проверку их соответствия оригиналам и цели исследования. Это этап математического моделирования системы с целью решения поставленной задачи.

5. Проведение модельных расчетов и экспериментов с целью исследования и совершенствования выбранных оригиналов. Проводится оценка различных сценариев развития системы при тех или иных внешних воздействиях, оценивают, насколько соответствует действительности построенная модель и ее работоспособность.

6. Последний этап – это теоретический системный синтез альтернативных рекомендаций и их внедрение в рассматриваемую сферу. Если стоит задача оптимального управления, то анализируются альтернативные варианты управления и выбирается наилучший. Здесь проявляется важное преимущество системного анализа – возможность воспроизведения с помощью математических моделей множества вариантов функционирования реальной системы при разных воздействиях.

Приведенная схема системного анализа – только одна из многих. К ней следует относиться творчески.

Контрольные вопросы

1. На чем основано практическое применение системно-целевого подхода?
2. Объяснить обобщенную структуру системно-целевого подхода.

3. Что входит в информационное описание системы?
4. Дать определение морфологического описания системы.
5. Объяснить, для чего необходимо функциональное описание системы.
6. Сформулировать задачу системного анализа.
7. Какова структура системного анализа?
8. Какие задачи решает эмпирический системный анализ?
9. Что такое проблемно ориентированное описание системы?

3. ОБЩИЕ ВОПРОСЫ МОДЕЛИРОВАНИЯ. МОДЕЛИРОВАНИЕ СИСТЕМ

Для выявления объективных закономерностей состояния и эволюции состояния систем необходимо экспериментирование. Но прямое физическое экспериментирование на реальных системах невозможно в принципе. Единственным средством получения информации о возможном изменении состояния систем служит моделирование. Модели систем строятся для теоретических целей анализа и для практических целей планирования, управления и прогноза. Они позволяют лучше осмыслить динамику изучаемых систем, выработать рекомендации по рационализации их структуры и методов прогнозирования и управления. Особое значение такие модели имеют при изучении регулирующих факторов: устойчивости, равновесия, роста, выявления обратных связей. В настоящее время все более широкое применение находят человеко-машинные системы для обучения рациональным формам и методам управления (деловые игры). Это способствует тому, что моделирование становится все более универсальным средством совершенствования экономических систем, проверки экономических положений и доктрин [7, 9].

Результаты моделирования служат основой для системного анализа проблем, возникающих в системах, и разработки мероприятий по их эффективному предупреждению или ликвидации.

3.1. Понятие объекта и его модели. Краткая характеристика моделей

Основными понятиями в моделировании служат понятия объекта и его модели. Все то, на что направлена человеческая деятельность, называют объектом. Объекты существуют вне нашего сознания и взаимодействуют между собой и с внешней средой. Для изучения объектов создают специальные наглядные и удобные для исследования схемы, основанные на гипотезах и аналогиях, которые называют моделями. Следовательно, моделирование – это замещение объекта его моделью с целью получения информации о свойствах объекта в результате изучения модели. При изучении сложных объектов невозможно и не нужно учитывать все его качества и свойства. Поэтому приходится отбирать (учитывать) только те, которые необходимы для достижения целей изучения или решения сформулированной задачи. В итоге получают модель объекта как некоторое вспомогательное средство, замещающее исходный объект. Моделями могут служить не только реальные объекты, но и абстрактные, идеальные построения. Если результаты моделирования подтверждаются практикой и могут быть использованы для прогнозирования изменений, происходящих с объектами, то говорят, что модель адекватна объекту, целям моделирования и критериям оценки модели. Любая модель связана некоторыми отношениями подобия с объектом-оригиналом [1, 7, 9].

Процесс моделирования предполагает наличие объекта моделирования, собственно модели, и человека – исследователя, который стремится достичь целей моделирования и выступает в роли экспериментатора, выполняющего эксперименты на модели. Эта ситуация может быть представлена схемой, приведенной на рис. 13.

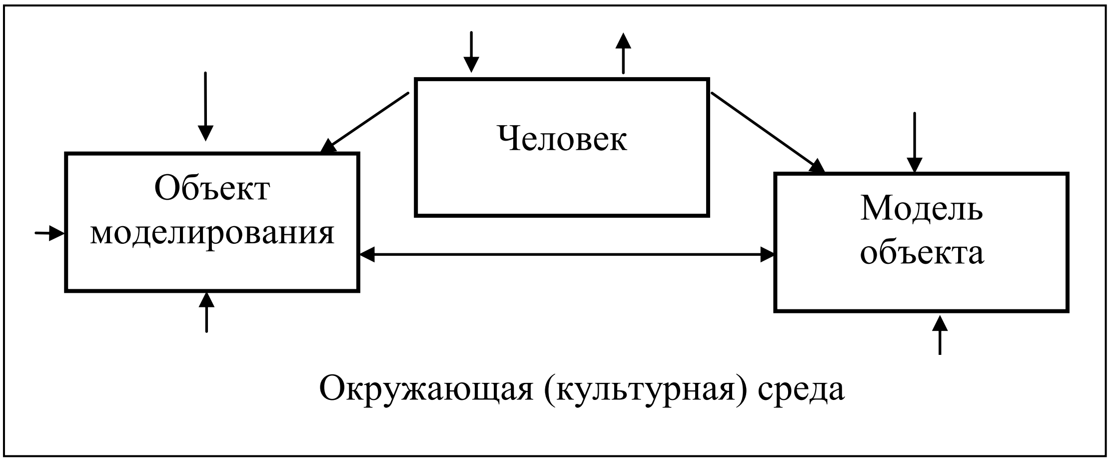


Рис. 13. Структура процесса моделирования

Часто построение модели начинают с «инвентаризации» всех имеющихся сведений об объекте моделирования, например, составления списка переменных и связывающих их ограничений, идентификации определяющих переменных, выбора вида модели и т. д. Любой объект может быть описан многими способами, исходные данные о нем известны приближенно. Поэтому модели объекта должны соответствовать уровню знаний о нем, возможностям использования и достижения целей моделирования.

Поскольку один и тот же объект разными людьми может восприниматься по-разному, то и создаваемая ими модель зависит от множества субъективных факторов – объема и качества знаний, эмоционального состояния и других причин. Поэтому модели одного и того же объекта, созданные разными людьми или даже одним и тем же человеком, но в разных условиях, отличаются между собой.

В зависимости от обстоятельств модель может сама стать объектом моделирования. Так, например, при моделировании закономерностей развития производства с целью прогнозирования этого развития используют статистические данные об изучаемом производстве, которые сами по себе являются моделью изучаемого производства.

Любой объект можно рассматривать и как единое целое и как объединенную совокупность элементов. Мощность связей между элементами изучаемого объекта значительно превосходит силу связей с элементами, не принадлежащими объекту, что позволяет выделить объект из окружающей среды как некоторую систему.

Система взаимодействует с окружающей средой через входы $\{X\}$ и выходы $\{Y\}$. В каждый момент времени t система имеет некоторое состояние $S(t)$, которое может изменяться под влиянием внешней среды или процессов внутри системы. Описанная ситуация изображена на рис. 7.

Однако, изучая каждый из элементов в отдельности, невозможно познать все свойства объекта в целом. У любого объекта как целого всегда имеется свойство, которого нет ни у одного из элементов или частей объекта. Оно возникает и существует за счет установления некоторых отношений (связей) между элементами и частями объекта. Сеть взаимосвязей между частями объекта, которая приводит к появлению свойств, отсутствующих у всех и каждой части объекта в отдельности называют структурой объекта. Так как модель объекта – это некоторое его отображение, то и структура объекта должна отображаться в структуре модели.

Структура модели должна подчиняться требованиям, предъявляемым к модели: она должна отображать структуру объекта соответственно целям моделирования. Различным целям моделирования могут соответствовать различные структуры одного и того же объекта. Например, участок физической поверхности Земли по-разному рассматривается и по-разному структурируется картографом, архитектором, строителем, крестьянином и т. д. Вместе с тем, модели одинаковой структуры могут использоваться для достижения различных целей (разные по содержанию документы могут иметь одинаковый шаблон).

Любая модель приближенно отображает объект. Эта приближенность обусловлена возможностью отображения в модели лишь конечного числа упрощенных свойств объектов. Поэтому одному объекту может соответствовать множество моделей (рис. 14). В каждой модели отображается только часть свойств объекта (на рис. 14 модели лишь частично пересекаются с объектом). В модели также имеются свойства, которых нет у объекта и свойства, соответствующие объекту лишь при определенных условиях. Поэтому всякая модель имеет истинное, условно истинное и ложное содержание. Разные объекты могут иметь одинаковые свойства, которые отображаются в одной модели. Такая модель служит отображением множества различных объектов.

Рис. 14. Отношения объекта и его моделей

Объект моделирования всегда существует в некоторой среде, взаимодействуя с ней и реагируя на ее воздействия. Желание предвидеть результаты взаимодействия объекта и окружающей его среды неизбежно приводит к моделированию состояний объекта и эволюции его состояний. Поэтому взаимосвязь объекта и среды необходимо учитывать при изучении объектов методом моделирования.

3.2. Классификация моделей

Существуют различные варианты классификации моделей в зависимости от критериев классификации. В настоящее время принято различать три класса моделей соответственно трем уровням организации материи: модели неживой материи, модели живой материи и модели общества – мыслящей, познающей себя материи [10].

На уровне неживой материи моделирование основывается на физических законах сохранения вещества, энергии, количества движения и т. д. Но эти законы не исчерпывают всех ограничений на выбор варианта модели и поэтому не обеспечивают единственность решения. Поэтому в процессе разработки модели соответственно целям моделирования и имеющимся ресурсам осуществляется сужение множества вариантов моделирования. Такие модели широко используются в физике и технике.

Принципы моделирования, справедливые в неживой материи, сохраняют свою силу и для моделирования живой материи, однако их уже недостаточно, чтобы объяснить процессы, происходящие в живых системах. Живые системы обладают целесообразным поведением, которое формируется с помощью обратной связи на основании информации о внешней среде так, чтобы сохранить свой гомеостазис, т. е. остаться внутри области параметров среды, где возможно существование системы. Изменение состояний живой системы, основанное на стремлении сохранить свой гомеостазис, не может быть выведено из принципов, определяющих состояние в неживых системах.

На общественном уровне организации материи при описании процессов в человеческом обществе информационные процессы и обратные связи оказываются намного сложнее, чем в живых, биологических системах. Люди имеют способность анализировать происходящие процессы, предвидеть последствия своей деятельности, формулировать гипотезы и прогнозировать поведение любой группы человеческого сообщества. Любые группы людей имеют свои цели (они могут быть взаимно противоречивые) и средства их достижения. В общественных системах циркулируют большие объемы информации, необходимой для принятия решений, направленных на достижение целей всей системы или ее частей. Обратные связи в общественных системах плохо формализуемы, а возникающие неопределенности обусловливают необходимость включения в математическую модель функционирующего «биологического» звена – человека (эксперта) [10]. Системы, с которыми ассоциирована группа субъектов, обладающих собственными целями, называют кибернетическими системами.

В физике и технике исследование математических моделей – один из основных методов исследования и проектирования. В биологических и общественных системах математические модели служат не для получения точных количественных оценок, а для определения допустимых границ и возможностей изменения состояния систем, тенденций их развития.

Другим критерием классификации моделей служит способ отображения объекта. По этому критерию все модели делятся на *изобразительные, аналоговые* и *символические* [7].

Изобразительная модель. Это простейший тип моделей. Они подобны оригиналу и лишь описывают объекты. Фотографии, картины, скульптуры являются изобразительными моделями. Вообще заметим, что часто изобразительная модель геометрически подобна объекту моделирования

В общем случае всякое изображение – это изобразительная модель некоторых свойств объекта в принятом масштабе. Например, на глобусе достаточно правильно отображены форма Земли, относительные размеры континентов, морей, островов и т. д. Однако, хотя изобразительная модель подобна оригиналу, она не отображает все его свойства. Модели этого класса хорошо приспособлены для отображения статических объектов или динамических объектов в фиксированный момент времени.

Аналоговая модель. Изобразительная модель не всегда позволяет отобразить все необходимые свойства объекта. На глобусе, например, нельзя наглядно представить геологическую структуру Земли. Тем не менее, эта задача решается использованием разноцветной окраски, т. е. подменой одного свойства другим в соответствии с заданным правилом. Такие модели широко используются в картографии, правила преобразования приводятся в легенде (системе условных обозначений). Модели, в которых совокупность одних свойств представляется с помощью совокупности других свойств, называют моделями-аналогами. Аналоговая модель – это вспомогательный объект, имеющий физическую природу, отличную от природы изучаемого объекта, и замещающий этот объект так, что его изучение дает полезные сведения об исходном объекте [11, 12]. Аналоговое моделирование основано на принципе подобия – сходства объектов по некоторым признакам. Простейшим примером модели-аналога служит обычный график, на котором различные физические свойства (время, вес, возраст, и т. д.) отображаются с помощью расстояния между точками. В общем случае результаты, полученные на моделях-аналогах, имеют вероятностный характер.

Символические модели. Это модели, в которых объекты описываются на формальном языке, состоящем из конечного набора символов, правил отношения между ними и правил интерпретации слов. С этих позиций алгоритмические языки, системы счисления, разнообразные шифры и коды представляют примеры символических моделей. Математические модели также относятся к классу символических моделей. Поэтому все законы природы, выраженные в математической или словесной форме, есть символические модели.

Другим вариантом классификации моделей может служить их деление на *познавательные* и *прагматические* [1, 7].

Познавательные модели являются формой организации и представления знаний, средством соединения новых и имеющихся знаний. В познавательной модели всегда отображается реальность. Вследствие этого для устранения расхождений между моделью и реальностью необходимо изменение модели. Примером этому может служить эволюция моделей мироздания: от библейских моделей к моделям Птолемея, Коперника – Кеплера – Ньютона и Эйнштейна.

Прагматические модели – средство или способ представления образцово правильных действий или их результата. Они играют роль стандарта и имеют нормативный характер. В прагматических моделях отображается, может быть не существующее, но желаемое и возможно осуществимое. Примером прагматических моделей служат уставы, кодексы законов, алгоритмы, технологические допуски и т. д. Целесообразность прагматических моделей заключается в определении целенаправленной деятельности для преобразования реальности с целью ее приближения к модели.

Еще одним способом классификации моделей служит их деление на *статические и динамические.*

Статические модели отображают какое-то одно, мгновенное состояние объекта, и этим они похожи на изобразительные модели. Примерами статических моделей служат любые изобразительные модели.

Динамические модели отображают процесс изменения состояний объекта, эволюцию состояний. Примером динамической модели может служить модель движений и деформаций природных или техногенных объектов: зданий, плотин ГЭС, земной поверхности, лунно-солнечных приливов и т. д.

Множественность вариантов классификации моделей свидетельствует о том, что одна и та же модель, в зависимости от критерия (правила) классификации, может быть отнесена к различным классам моделей, и что между классами моделей не всегда существует четкая граница, позволяющая выполнять однозначную классификацию моделей.

3.3. Обобщенная структура моделирования систем

Трудоемкость процесса моделирования систем, отличающихся разнообразием и сложностью, обусловливает необходимость рассмотрения общей структуры процесса моделирования. Вся процедура моделирования условно может быть разделена на пять этапов.

Этап 1. Формулируется задание на разработку модели. Для этого выполняется исследование объекта с целью выявления основных его свойств, осуществляется сбор, проверка и анализ доступных сведений об аналогичных системах, изучаются литературные источники о разработанных ранее моделях исследуемой системы, собранные сведения систематизируются и обобщаются в когнитивной модели – виртуальном образе системы.

В результате выполнения первого этапа составляется список вопросов, на которые должен быть получен ответ по результатам моделирования, и осуществляется содержательная постановка задачи моделирования, решение которой обеспечивает достижение целей в приемлемые сроки. По трудоемкости этот этап может занимать до 40 % всех ресурсов.

Этап 2. Осуществляется концептуальная постановка задачи, или семантическое (смысловое) моделирование исследуемой системы.

Концептуальная модель определяется перечнем вопросов, на которые нужно получить ответы, и гипотезами о поведении моделируемой системы [29, 30]. Концептуальная модель, представленная в виде диаграмм причинно-следственных отношений, является семантической моделью. Семантическая модель предоставляет возможность для предварительного контроля и анализа модели.

Этап 3. Цель данного этапа – проверка обоснованности концептуальной постановки задачи и корректности соответствующей семантической модели. Проверку проходят все принятые гипотезы и предположения о поведении системы, о существенности и несущественности свойств системы. Иногда уже на этом этапе могут быть достигнуты цели моделирования.

Этап 4. Выполняется математическое моделирование и выбирается метод реализации и исследования модели.

Этап 5. Качественный и количественный анализ модели. Качественный анализ модели необходим для выявления общих закономерностей ее функционирования и проверки адекватности полученных результатов объективной реальности и целям моделирования. Количественный анализ модели необходим для прогнозирования числовых характеристик и оценки эффективности различных стратегий ее преобразования.

3.4. Формальные модели систем

Основой всего системного анализа является построение моделей систем. Несмотря на многообразие реальных систем, различают всего несколько принципиально различных типов моделей систем: «черный ящик», модель состава, модель структуры, структурная схема системы и их корректные комбинации [1, 29, 30]. Эти модели являются формальными, они относятся к любым системам и, следовательно, ни к одной конкретно. Чтобы получить модель конкретной системы, необходимо наполнить формальную модель конкретным содержанием. Эта процедура носит неформальный характер и непосредственно зависит от исследователя, который должен определить элементный состав системы, существенные и несущественные свойства и отношения между частями системы. Рассмотрим эти формальные системы.

3.4.1. Модель «черного ящика»

«Черный ящик» – система, в которой внешнему наблюдателю доступны лишь ее связи с внешней средой, а внутреннее устройство и процессы, в ней протекающие, не известны. Система связана со средой и через эти связи взаимодействует со средой. Эти связи называют входами и выходами системы. В результате имеем модель системы, которая получила название модель ***«черного ящика»***. Ряд важных выводов о поведении системы можно сделать, наблюдая реакцию ***выходов*** на изменение ***входов***. Такой подход позволяет изучать системы, устройство которых неизвестно или слишком сложно.

Пусть на ***вход*** системы подаются входные воздействия $X_1, X_2, \ldots, X_n$, на выходе регистрируются выходные: $Y_1, Y_2, \ldots, Y_m$. Наблюдая достаточно долго за поведением такой системы, и, если потребуется, выполняя активные эксперименты над ней, т. е. изменяя по заданному сценарию входные воздействия, можно изучить систему настолько, что окажется возможным предсказать изменение ее ***выходов*** при любом изменении ***входов***. Однако, как бы тщательно не изучалось поведение ***«черного ящика»***, его внутреннее строение остается неизвестным, так как разные системы могут иметь одно и то же поведение. Они неотличимы для наблюдателя, которому доступны только их ***входы и выходы***.

Максимально формализованная модель ***«черного ящика»*** состоит из двух множеств **X** и **Y** входов и выходов системы. Никакие отношения между этими множествами задавать ***нельзя***. Построение модели ***«черного ящика»*** не является тривиальной задачей, так как непросто установить, сколько и какие ***входы и выходы*** следует включать в модель. Критерием выбора служит целевое назначение модели и существенность связи по отношению к этой цели. В некоторых случаях модель ***«черного ящика»*** оказывается единственно применимой для изучения систем.

3.4.2. Модель состава и структуры системы, структурная схема системы

Модель состава системы описывает, из каких подсистем и элементов она состоит. Состав системы определяется в зависимости от следующих параметров:

- уровень и глубина знаний о системе;
- назначение модели и ее цели;
- определение элемента системы;
- правила построения подсистем.

Модель состава системы, хотя и показывает, из каких частей состоит система, не всегда достаточна для решения необходимых задач. Это связано с тем, что система обладает свойствами, которых нет ни у одной из ее частей. Это обстоятельство обусловливает необходимость установления связей между частями системы, т. е. в определении структуры системы.

Объединяя модели ***«черного ящика»***, состава и структуры, имеем определение системы как совокупности взаимосвязанных элементов, обособленных от среды и взаимодействующих с ней как целое. Такую модель называют ***структурной схемой системы***. В ней перечисляются все элементы системы, все связи между элементами внутри системы и связи с окружающей средой (входы и выходы системы). Все рассмотренные модели отображают систему неизменной во времени. Их называют статическими моделями.

3.4.3. Динамические модели систем

Модели, отображающие какие-либо изменения, происходящие в системе с течением времени называют ***динамическими моделями*** систем. Для разных объектов и систем создано множество динамических моделей, в которых они отображаются с различной степенью детальности, подробности и точности. Наглядными примерами динамики системы служит ее функционирование и развитие [1, 29, 30]. ***Функционирование*** – это процессы в системе и окружающей ее среде, например, функционирует прибор, транспортная система, учебное заведение, а ***развитие*** – это необратимое, направленное, закономерное изменение системы, в результате которого возникает новое состояние системы, сопровождающееся изменением ее цели и структуры. Постепенное, медленное развитие системы называют эволюцией. В процессе эволюции системы приспосабливаются к изменениям состояния внешней среды, используя механизмы компенсации и сглаживания для сохранения своей живучести. Нарушения постепенности развития системы называют революцией. В процессе революционного развития происходят значительные качественные изменения системы, например, изменение ее структуры и целей.

При построении динамических моделей главная задача – отображение изменений, происходящих в системе. Например, в модели ***«черного ящи-***

ка» динамика отображается при определении входов и выходов как функций времени.

В динамических системах процессы могут протекать непрерывно, а могут быть дискретными, т. е. изменения в системе происходят в дискретные моменты времени. И в том и в другом случае предполагается возможность анализа поведения системы на некотором интервале времени. Иногда бывает известно, что система мгновенно преобразует вход в выход (выход $y(t)$ является функцией только $x(t)$ в тот же момент времени). Такие системы называют ***безынерционными***. В действительности безынерционные системы – это идеализация некоторой реальной системы, в которой выход определяется не только значением входа в данный момент времени, но и значениями, которые были на входе раньше. Кроме того, и сама система не остается неизменной вследствие действия на нее извне или в результате происходящих в ней внутренних процессов, что также необходимо отобразить в модели. В символическом виде эта модель имеет вид [1]:

$$X \xrightarrow{\varphi} S \xrightarrow{\eta} Y,$$

где X, Y, S – множества возможных значений входов, выходов и состояний системы соответственно; φ – оператор отображения входов в состояния системы; η – оператор отображения входов и состояний системы в ее выходы.

3.5. Роль и значение моделирования

Математическое моделирование – мощное средство решения разнообразных задач оптимизации технико-экономических и природных систем. Методы математического моделирования завоевывают все новые области приложения. Это обусловлено не только эффективностью методов математического моделирования, но и широчайшей общностью моделей, допускающей богатый спектр интерпретаций.

Процедура моделирования применяется человеком с незапамятных времен. По мере накопления знаний методы моделирования совершенствуются, становятся изощреннее, но, по-видимому, всегда сохранятся ограничения, не позволяющие получать абсолютно достоверные, однозначные результаты моделирования. Это обусловлено ограниченностью знаний об изучаемом объекте, множественностью целей моделирования, субъективности представления объекта человеком, неполнотой модели, связанной с ее целенаправленностью и неопределенностью реакции объекта на внешнее воздействие. Однако, несмотря на это, роль и значение моделирования постоянно возрастают за счет возможности замены «натурного» эксперимента экспериментом на модели [7]. Такая замена позволяет:

1) изучать недоступные для натурного эксперимента объекты, например, физические процессы на звездах или других планетах, технологические процессы, связанные с опасностью для экологии, явления, которые протекают либо очень быстро, либо очень медленно;

2) исследовать гипотетические объекты или реальные объекты в гипотетических условиях, например, новые, несуществующие конструкции или реальные объекты в экстремальных условиях;

3) выполнять имитацию на модели в широком диапазоне изменения параметров объекта и внешней среды, получая дополнительную информацию в условиях информационной неопределенности;

4) осуществлять обучение людей в различных областях науки и техники;

5) формировать справочные и экспертные системы;

6) создавать автоматизированные системы проектирования и управления производством;

7) объединять формальные и неформальные методы исследования, формируя основу для создания искусственного интеллекта, а также решать и многие другие задачи.

К недостаткам экспериментирования на модели можно отнести:

1) сложность модели и возможную высокую стоимость эксперимента;

2) трудности и неоднозначность интерпретации результатов модельного эксперимента;

3) трудности оценки достоверности результатов моделирования;

4) высокую специализацию модели и невозможность или трудности ее использования для других объектов.

Контрольные вопросы

1. Что такое модель? Привести примеры.
2. Что подразумевается под объектом моделирования?
3. Привести классификацию моделей.
4. Привести структуру процесса моделирования систем.
5. Что такое формальная модель системы?
6. Что такое модель «черного ящика»?
7. Какие процессы описывают при помощи динамических моделей систем? Привести примеры.
8. Что такое модель структуры системы?
9. Что входит в структурную схему модели системы?
10. Каково значение моделирования в современном мире?

4. ПРОЦЕДУРЫ ДЕКОМПОЗИЦИИ И АГРЕГИРОВАНИЯ ПРИ РЕАЛИЗАЦИИ СИСТЕМНО-ЦЕЛЕВОГО ПОДХОДА

Как отмечалось выше, системный анализ и синтез – основные процедуры системно-целевого подхода, в них содержатся операции разделения целого на части и объединения частей в целое. Эти операции называют декомпозицией и агрегированием соответственно.

Операция декомпозиции приводит к упрощению решения задачи о сложной системе путем разложения ее на какое-то число меньших (простых) частей. Чтобы обеспечить окончание процедуры, достаточно для простых объектов прекратить процесс декомпозиции и решить задачу непосредственно для них. Операция агрегирования из частей, полученных после декомпозиции, конструирует агрегат. Его можно представить в виде модели «черного ящика», имеющей конечное число входов и выходов и некоторое множество внутренних состояний. На входы из окружающей среды поступают сигналы. В зависимости от значений сигналов и от состояния, в котором находится агрегат, он переходит в новое состояние и выдает сигналы на выходы. При изменении входных сигналов изменяются состояния агрегата и выходные сигналы, агрегат функционирует во времени.

Операции декомпозиции и агрегирования – неформальные операции.

4.1. Декомпозиция системы

При возрастании сложности систем существенно затрудняется их исследование. Основой системно-целевого подхода является расчленение сложной проблемы на разрешимые задачи и рассмотрение их во взаимодействии [8, 13].

Процедура декомпозиции состоит в определении последовательности формальных и неформальных действий, позволяющих представить исходный объект в виде совокупности более простых частей, которые, взаимодействуя, обеспечивают достижение цели системы в целом.

Процесс декомпозиции имеет циклический характер – он может повторно применяться к ранее полученным результатам декомпозиции. Процесс декомпозиции прекращается, когда дальнейшая детализация или невозможна или не требуется. Основное требование, предъявляемое к результатам декомпозиции, состоит в представлении системы исчерпывающе подробно и с минимальным по возможности числом элементов и подсистем, обеспечивающим достижение цели. Все, что не требуется для достижения цели и не нарушает диалектического единства, может быть исключено из рассмотрения. В результате этого описание исходного объекта упрощается и достижение цели облегчается. Результаты декомпозиции удобно представлять в виде графа, подобно тому, как это изложено при структурном (морфологическом) описании систем.

Существуют различные варианты выполнения процедуры декомпозиции, основные из которых приведены на рис. 15.

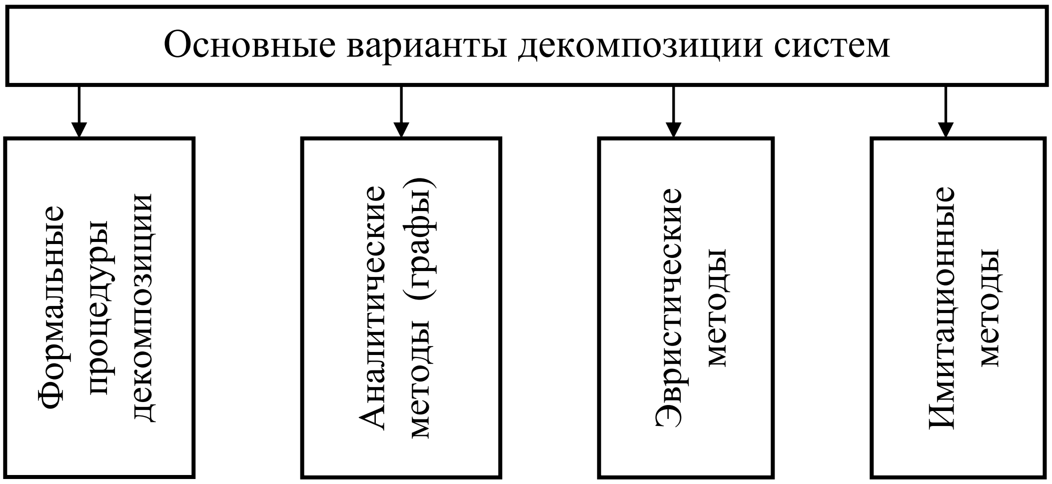


Рис. 15. Основные варианты декомпозиции

1. ***Формальная декомпозиция*** выполняется на основании принятой формальной модели системы. При этом рассматриваемый объект сравнивается с принятой формальной моделью и разделяется на столько частей, сколько их содержится в принятой модели. Например, если в качестве формальной модели выбрана модель типа «черный ящик», то результатом декомпозиции будет множество входов, множество выходов. Результаты декомпозиции в этом случае можно представить вектором входов и векто-

ром выходов. Вопрос о конкретном содержании частей системы при этом остается открытым до тех пор, пока содержательно не определена цель исследования.

В качестве примера формальной дкеомпозиции рассмотрим декомпозицию сложного события *S* на основании формальной модели состава и структуры. Возможный вариант декомпозиции показан на рис. 16.

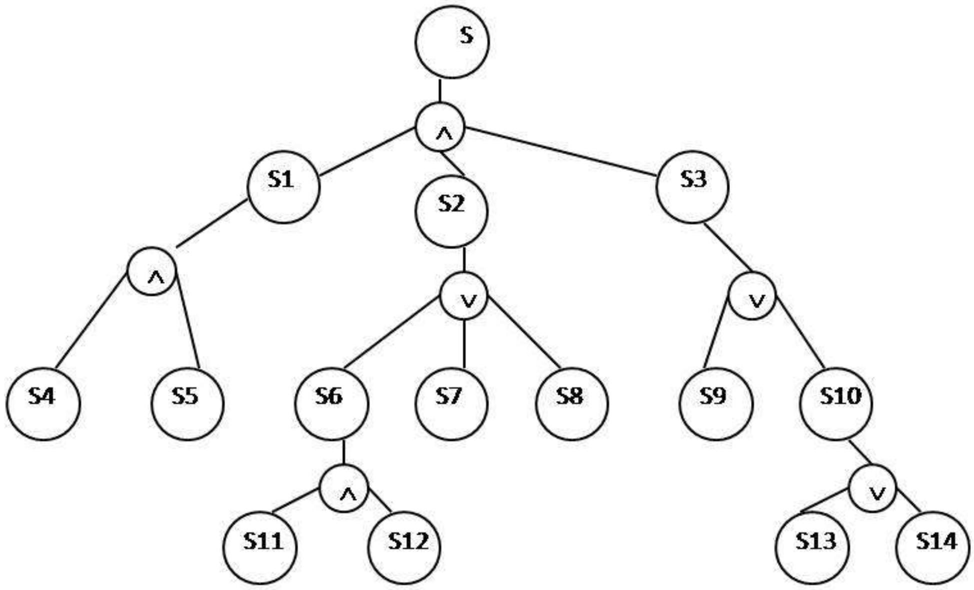


Рис. 16. Диаграмма декомпозиции

Первый цикл декомпозиции представляется выявлением трех предпосылок *S1*, *S2*, *S3* первого уровня, таких, что для появления события *S* необходимо появление каждой из них. Этот факт на рис. 16 показан в кружке логическим знаком конъюнкции.

На втором цикле декомпозиции каждая из предпосылок первого уровня, в свою очередь, рассматривается состоящей из предпосылок второго уровня. Предпосылка *S1* составляется из событий *S4* и *S5*, предпосылка *S2* – из событий *S6*, *S7*, *S8*, предпосылка *S3* – из событий *S9*, *S10*. Варианты взаимодействия предпосылок второго уровня показаны в кружках логическими операциями дизъюнкции и конъюнкции.

В третьем цикле декомпозиции события – предпосылки *S4*, *S5*, *S7*, *S8*, *S9* рассматриваются как неделимые, а события – предпосылки *S6*, *S10* оказалось возможным разделить на предпосылки третьего уровня *S11*, *S12, S13, S14*. Продолжать декомпозицию оказалось невозможным – все предпосылки нижнего уровня оказались элементарными, их деление нецелесообразно. В результате декомпозиции исходное событие *S* разделено на множество взаимосвязанных элементарных частей.

Чтобы придать результатам декомпозиции содержательный смысл, необходимо содержательно определить исходное событие *S*. Например, исходное событие *S* – безопасность производства в цехе № 5 завода металлоконструкций, или *S* – экономическая эффективность выпуска пылесосов на машиностроительном предприятии. Результаты декомпозиции будут различными в зависимости от конкретного содержания исходного события.

2. ***Аналитическая декомпозиция*** выполняется, когда имеется возможность конкретизировать понятие простоты, элементарности результата декомпозиции. Такая возможность часто возникает в математических задачах и при изучении технических систем. В математике аналитическая декомпозиция осуществляется разделением решения сложной задачи на совокупность решений простых задач и конструированием из них решения сложной задачи, а в технике – разделением системы на части, выполняющие простые элементарные операции, объединение которых в соответствии с принятой технологией обеспечивает достижение цели. Примером аналитической декомпозиции может служить представление функций рядами, в которых удерживается количество элементов – членов ряда, обеспечивающее достижение необходимой точности, или методы фильтрации ошибок экспериментальных исследований. В технике примером может служить конвейерная организация производства или блочно-иерархический подход к изучению и проектированию сложных систем.

Пример.

Изменение пространственно-временного состояния объекта определятся функцией (фазовой траекторией) по заданным дискретным временным рядам эмпирических данных. Для прогнозирования будущего состояния объекта обычно используются различные методы прогнозирования, основой которых является подбор такой непрерывной математической функции, заданной уравнением, которая бы с требуемой точностью описывала реальный процесс.

Одним из приемлемых методов прогнозирования является метод экспоненциального сглаживания (рис. 17а) [3, 4, 15]:

$$S_t = \alpha \cdot y_t + (1-\alpha) \cdot S_{t-1}. \qquad (8)$$

Здесь $\alpha \in [0,1]$ – постоянный коэффициент сглаживания; S_t – сглаженное значение y, отнесенное к моменту t; y_t – состояние объекта в момент t.

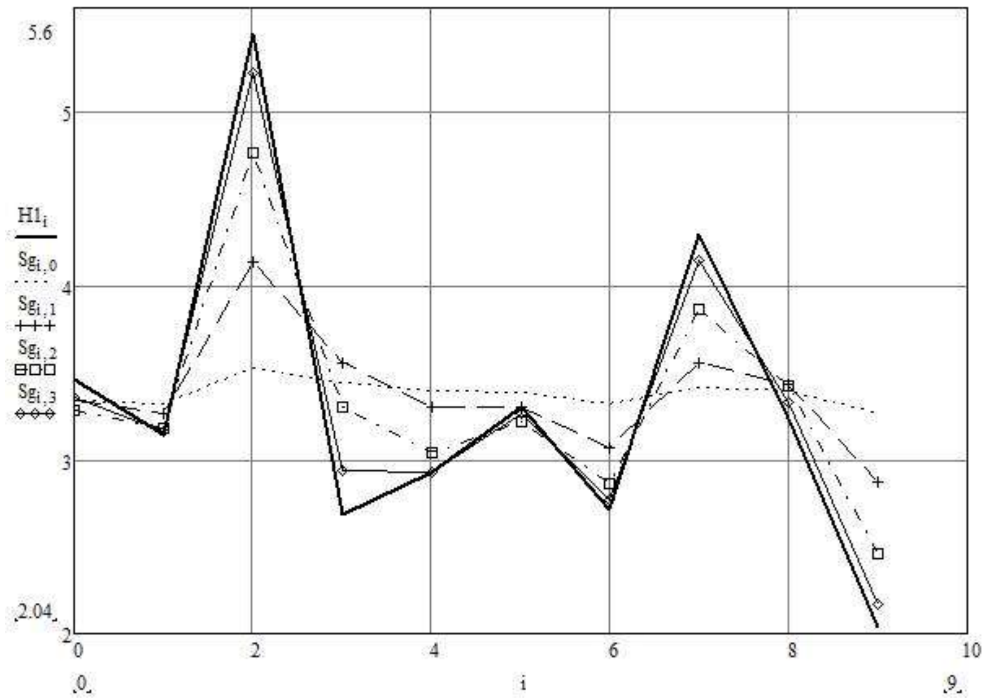
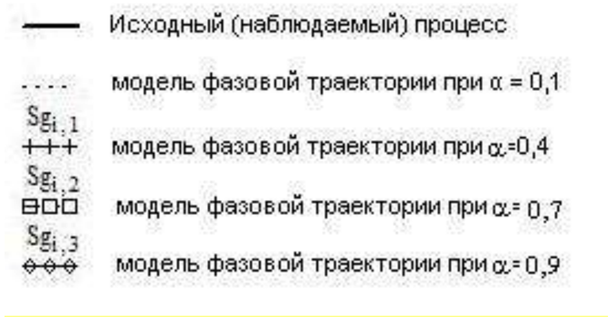


Рис. 17а. Прогнозирование состояния объекта методом экспоненциального сглаживания:



Начальное значение сглаживания S_0 при $t = 1$ может быть задано как среднее из эмпирических данных рассматриваемой функции. Выбор коэффициента сглаживания зависит от степени «доверия» эмпирическим данным. Отсутствие случайных ошибок данных прогнозируемого процесса предполагает выбор значения коэффициента α, близкого к единице. И, напротив, выбор значения α, близкого к нулю, осуществляется, когда прогнозируемая функция имеет стохастический характер.

Функция (8) позволяет получить прогнозируемое состояние объекта в дискретный момент времени. Описание наблюдаемого процесса для каждого значения коэффициента α определяет результат декомпозиции системы.

Для сравнения приведем результаты аппроксимации по методу наименьших квадратов многочленами третьей, пятой и седьмой степени (рис. 17б):

$$\sum(y - A3\cdot C3) = -5{,}556\times 10^{-13}; \qquad \sum(y - A3\cdot C3)^2 = 6{,}938;$$

$$\sum(y - A5\cdot C5) = 4{,}365\times 10^{-10}; \qquad \sum(y - A5\cdot C5)^2 = 4{,}674;$$

$$\sum(y - A7\cdot C7) = 1{,}106\times 10^{-7}; \qquad \sum(y - A7\cdot C7)^2 = 2{,}968.$$

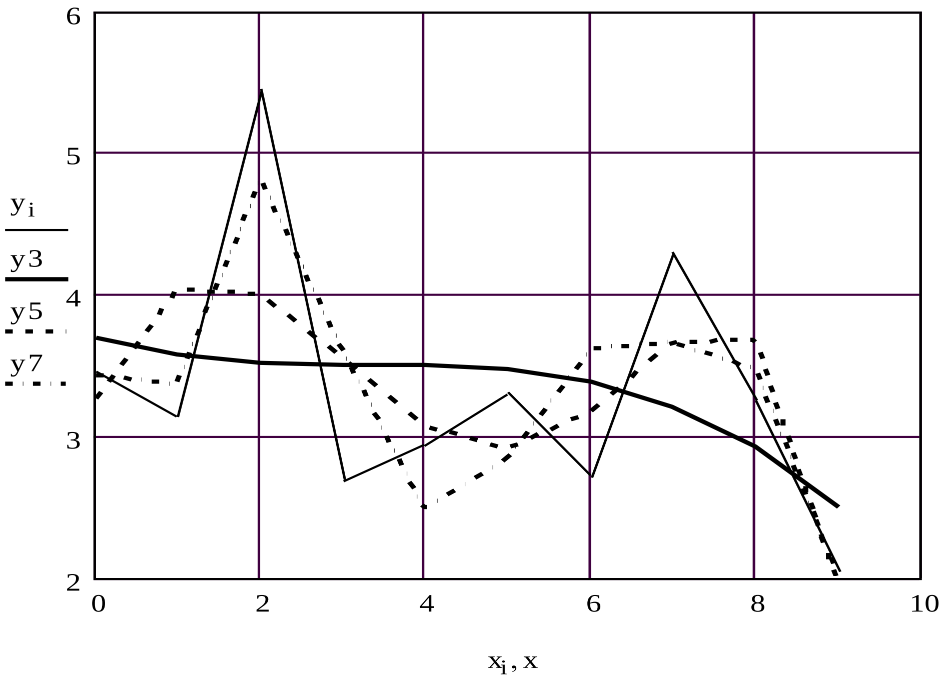


Рис. 17б. Прогнозирование состояния объекта методом наименьших квадратов:

y_i – исходный (наблюдаемый) процесс; $y3$ – модель фазовой траектории при α = 0,3; $y5$ – модель фазовой траектории при α = 0,5; $y7$ – модель фазовой траектории при α = 0,7

В отличие от метода экспоненциального сглаживания, сглаживание по методу наименьших квадратов осуществляется для всего множества экспериментальных данных в предположении, что в области аппроксимации характер аппроксимируемого процесса остается неизменным. При экспоненциальном сглаживании этого предположения не требуется.

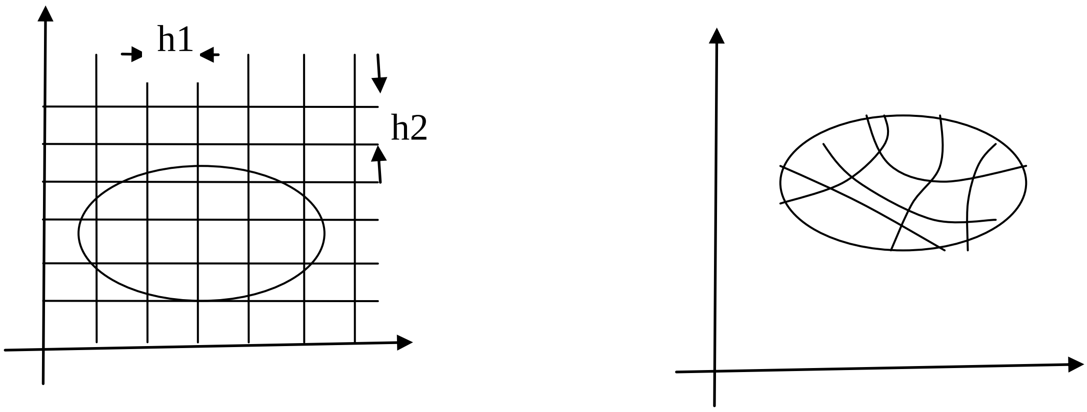
3. ***Эвристическая декомпозиция*** осуществляется экспертами – специалистами, обладающими большим субъективным запасом знаний, опыта и интуиции. Каждый эксперт самостоятельно генерирует варианты декомпозиции и определяет наиболее эффективные для достижения цели. Экспертные варианты могут обобщаться на основании коллективного обсуждения вариантов и выработки общего приемлемого варианта. При декомпозиции сложных систем может использоваться «метод мозгового штурма». Для этого создается специальная группа, в которую, как правило, включают 1–2 специалистов из других областей знания и малознакомых с рассматриваемой проблемой. Процедура «мозгового штурма» проходит в два этапа. На первом этапе генерируются и фиксируются любые идеи, критика которых на этом этапе запрещена. На втором этапе высказанные идеи изучаются и оцениваются высококвалифицированными специалистами, на основании мнения которых выбирается вариант декомпозиции.

Одним из способов декомпозиции является процедура дискретизации модели, которая состоит в преобразовании непрерывной информации в дискретную. При реализации этой процедуры обычно необходимо решить следующие задачи [7].

1. Разделить пространственные и временные области на конечное число элементарных участков.

2. Представить значения переменных конечным числом значений в избранных узловых точках, принадлежащих элементарным участкам, преобразовать непрерывную информацию (сигналы) в цифровую форму.

Сетка может быть прямоугольной или косоугольной, с постоянным или переменным шагом. Первая задача имеет два варианта решения. В первом варианте преобразуемая область покрывается сеткой, и область заменяется множеством точек – узлов сетки (рис. 18, *а*). Во втором варианте дискретизация области осуществляется ее разделением на непересекающиеся подобласти – конечные элементы. В пределах каждого конечного элемента выбирают конечное число узловых точек (рис. 18, *б*).



а *б*

Рис. 18. Дискретизация области:
а – с помощью сетки; *б* – с помощью конечных элементов

Решение второй задачи дискретизации модели выполняется заменой непрерывных переменных конечным числом их значений в узловых точках. В дальнейшем эти значения используются для аппроксимации соответствующих переменных в каждом конечном элементе и решении разнообразных задач, обеспечивающих достижение целей моделирования.

Дискретизация области широко используется, например, в цифровой картографии при моделировании рельефа.

4. ***Имитационная декомпозиция*** применяется для сложных систем, когда применение других методов невозможно. При имитационной декомпозиции система некоторым образом делится на взаимосвязанные части так, что сохраняется внутренняя и внешняя целостность системы, а декомпозиция частей выполняется каким-либо из вышеназванных методов декомпозиции. Например, вычисление интегралов методом Монте-Карло, когда процедуры аналитического или численного интегрирования заменяются (имитируются) случайным процессом.

По своей природе процедура декомпозиции часто имеет эвристический характер. Это значит, что результат декомпозиции зависит от опыта и компетентности специалиста, выполняющего декомпозицию системы. Разные специалисты по-разному выполнят декомпозицию одной и той же системы. Поэтому в качестве основания для декомпозиции используют формальные модели системы, сопоставляя объект декомпозиции с каждым элементом модели. Содержание каждого элемента декомпозиции зависит от содержания соответствующих элементов исходного объекта.

4.2. Агрегирование системы

Агрегирование – процедура, противоположная декомпозиции. В результате агрегирования устанавливают отношения на заданном множестве элементов, и образуется агрегат – такое объединение конструктивно и функционально унифицированных частей в целое, которое приводит к появлению нового качества за счет конкретных взаимосвязей между конкретными элементами агрегата. Другие связи приведут к возникновению других качеств.

Основная задача агрегирования состоит в определении структуры системы и значений параметров, наиболее эффективных относительно достижения поставленной цели. Эти задачи называют задачами структурного и параметрического синтеза.

Задача структурного синтеза относится к сложным задачам из-за большого числа факторов, влияющих на свойства и параметры агрегируемой структуры, и трудностей решения задач оптимизации большой размерности. Результатом структурного синтеза служит список необходимых для достижения цели частей системы и описание функциональных и информационных отношений между ними.

Задача параметрического синтеза заключается в определении наиболее эффективных значений параметров конструируемого агрегата с учетом всех ограничений, обусловленных взаимодействием с внешней средой и физическими законами его функционирования. По содержанию это задача нахождения вектора $\overline{\mathrm{X}}$ выходных параметров системы, таких, что целевая функция $\mathrm{F}(\overline{\mathrm{X}})$ при заданных ограничениях достигает наиболее эффективного значения. Как правило, эти задачи решаются методами математического программирования, среди которых широкое применение нашли методы комбинаторного анализа и, в частности, методы «ветвей и границ».

Агрегат, как и система, характеризуется множествами моментов времени $\boldsymbol{T}$, состояний $\boldsymbol{S}$ в каждый момент времени, множеством входных $\boldsymbol{X}$ и выходных $\boldsymbol{Y}$ сигналов. Поэтому агрегат рассматривается как унифицированная модель разнородных элементов системы, с помощью которой динамика системы раскрывается через динамику взаимодействующих между собой моделей элементов. Для создания агрегативной модели системы необходимо разработать агрегативные модели элементов и построить модель взаимодействия агрегатов [6, 14].

Агрегат можно представить как некоторое устройство («черный ящик»), имеющее конечное число входов и выходов и некоторое множество внутренних состояний. На входы из окружающей среды поступают сигналы. В зависимости от значений сигналов и от состояния, в котором находится агрегат, он переходит в новое состояние и выдает сигналы на выходы. При изменении входных сигналов изменяются состояния агрегата и выходные сигналы, агрегат функционирует во времени.

Всякая сложная система характеризуется многомерностью и разнообразием протекающих в ней процессов, имеющих смешанную природу происхождения. Однако не вся эта информация необходима для достижения поставленной цели. Поэтому в результате агрегирования необходимо получить описание системы, позволяющее с необходимой и достаточной для достижения цели полнотой описать систему. Такой агрегат называют ***конфигуратором*** [1].

В конфигураторе объединяются мнения специалистов различных областей знания о рассматриваемой проблеме. Полнота описания проблемы зависит от цели. Например, конфигуратором для определения положения точки ***n***-мерного пространства является совокупность ее координат. Если число координат окажется меньше размерности пространства, то положение точки не будет определено с необходимой полнотой. Разные системы координат – разные конфигураторы. Конфигуратором является описание экономической системы с финансовой, правовой, экологической, общественной и другой точки зрения.

Варианты агрегирования такие же, как и варианты декомпозиции: формальное, аналитическое, эвристическое и имитационное.

Формальное агрегирование. Как и при декомпозиции, агрегирование выполняется на основе какой-либо формальной модели, выбор которой зависит от условий и целей агрегирования. Разнообразие в выборе элементов и отношений между ними порождает разнообразие задач агрегирования. Если в качестве модели – основания для агрегирования принята модель «черный ящик», то результаты декомпозиции в этом случае можно представить графом с матрицей смежности, в которой число строк равно числу входов, а число столбцов – числу выходов. Элементам матрицы смежности, соответствующим связанным входам-выходам, приписывают значения 1, а не связанным – значения 0.

Аналитическое агрегирование. В результате аналитического агрегирования получают аналитическое описание системы, в котором отсутствуют свойства, не требующиеся для достижения цели. В общем случае аналитическое агрегирование требует на множестве элементов (результатов декомпозиции) установления некоторой системы отношений, обеспечивающих достижение цели. В результате этого образуется ***агрегат-оператор.*** Например, сложная функция может быть с заданной точностью определена частичной суммой ряда или аппроксимирована простой функцией. Агрегат-оператор в виде детерминированной числовой функции получают, когда агрегируемые признаки определены в числовых шкалах. С этих позиций любая функция многих переменных есть агрегат-оператор: на множестве элементов – независимых переменных – функция устанавливает правило их соответствия элементам другого множества. Выбор правила соответствия регламентируется целью агрегирования – результат агрегирования должен быть адекватным цели. ***Агрегат-структура*** чаще всего используется при синтезе систем. Они возникают при описании структуры уже существующей системы или определяются для проектируемой системы. При создании агрегата-структуры необходимо учесть все наиболее существенные отношения между элементами. Эти отношения определяются конфигуратором системы и, следовательно, структур должно быть столько, сколько языков включено в конфигуратор системы. Например, проект организационной системы включает структуры распределения власти, распределения ответственности и распределения информации. Эти структуры лишь с разных сторон описывают одну и ту же систему [1].

При статистическом анализе данных образуются агрегаты-статистики (функции выборочных значений). Они предназначены для извлечения из статистических данных полезной информации о некоторых параметрах.

В качестве примера выполним агрегирование события S, декомпозиция которого рассмотрена выше. Составим для этого события агрегат-оператор. Учитывая, что взаимодействие частей представляется логическими операциями дизъюнкции и конъюнкции, агрегат-оператор для события S – это логическая функция

$$F(S1, S2,..., S14) = S1 \wedge S2 \wedge S3 = (S4 \wedge S5) \wedge (S6 \vee S7 \vee S8) \wedge (S9 \vee S10) =$$
$$= (S4 \wedge S5) \wedge ((S11 \wedge S12) \vee S7 \vee S8) \wedge (S9 \vee (S13 \vee S14)),$$

где символами $\wedge$, $\vee$ обозначены операции логического умножения и логического сложения, соответственно. Эта логическая функция принимает значения «ложь» или «истина». Такие же значения получают и элементарные предпосылки: значение «истина» свидетельствует о реализации предпосылки, а значение «ложь» – о нереализации предпосылки. Сочетания предпосылок, при которых функция F принимает значение «истина», обеспечивают появление события S, а сочетания, при которых функция F принимает значение «ложь» – непоявление события S. Такой агрегат позволяет выявлять предпосылки, благоприятствующие появлению события S или исключающие возможность его появления.

Эвристическое агрегирование. Как и эвристическая декомпозиция, эвристическое агрегирование осуществляется экспертами-специалистами, обладающими большим субъективным запасом знаний, опыта и интуиции. Только теперь их задача состоит не в разделении системы на взаимосвязанные части, а в целенаправленном и эффективном объединении частей, полученных в результате декомпозиции [1, 2].

Имитационное агрегирование. В этом случае агрегирование рассматривают как выбор структуры частей системы и принципов их взаимодействия из некоторого небольшого числа перспективных вариантов, отбор которых осуществляется группой высококвалифицированных специалистов. Детальное исследование вариантов проводится для каждого из них в отдельности и осуществляется их сравнительный анализ и выбор окончательного варианта.

Конечным результатом системного исследования является принятие решения, направленного на ликвидацию проблемы. Как правило, существует множество вариантов решения, из которого необходимо сделать выбор оптимального или наиболее полезного варианта. Для этого необходимо оценить возможные последствия принимаемых решений, зависящие от условий (состояния внешней среды), в которых осуществляется выбор. Везде далее предполагается, что последствия принимаемых решений и соответствующие состояния среды известны или заданы.

Контрольные вопросы

1. Дать определение процедуры декомпозиции системы.
2. Дать определение процедуры агрегирования системы.
3. Объяснить, для чего необходимы процедуры декомпозиции и агрегирования систем.
4. Перечислить основные варианты декомпозиции систем.
5. Перечислить основные варианты агрегирования систем.
6. Дать определение формальной процедуры декомпозиции и агрегирования. Привести примеры.
7. Дать определение аналитического метода декомпозиции и агрегирования. Привести примеры.
8. Что представляет собой имитационный метод декомпозиции и агрегирования?
9. Что представляет собой эвристический метод декомпозиции и агрегирования? Привести примеры.

5. ОБЩИЕ ВОПРОСЫ ПРИНЯТИЯ РЕШЕНИЙ

Любые мероприятия или действия предпринимаются для достижения некоторой цели. При этом всегда имеется множество вариантов достижения цели, отличающихся между собой ресурсами, необходимыми для их реализации и условиями реализации. Из этого множества требуется выбрать вариант достижения цели. Рассмотрим некоторые примеры задач на принятие решения.

1. В любой человеко-машинной системе возникают аварийные ситуации, которые необходимо своевременно обнаружить и ликвидировать с минимальным ущербом. Объективно существует множество вариантов достижения этой цели. Требуется из этого множества выбрать наиболее полезный вариант.
2. Некоторое конструкторское бюро разрабатывает новый бытовой прибор (утюг, чайник, часы). Из множества вариантов организации производства такого прибора необходимо выбрать наиболее экономически эффективный вариант.
3. В городе проектируется телефонная станция. Требуется определить характеристики станции, при которых будут обслуживаться все абоненты со временем ожидания не более заданного.

Из приведенных примеров видно, что к задачам принятия решения можно приступать только после определения цели, ради достижения которой осуществляется выбор, и множества вариантов достижения цели. В дальнейшем будем считать, что эти сведения уже имеются.

Из-за неполноты или неточности необходимой информации выбранный вариант достижения цели может оказаться ошибочным и сопровождаться негативными последствиями: наносить ущерб материальным ресурсам, окружающей среде, жизни и здоровью людей. Поэтому выбор варианта действий для достижения цели не может быть произвольным, он должен быть обоснованным. Осуществление такого обоснованного выбора и составляет содержание теории принятия решения.

Задачи принятия решений разнообразны, но среди этого разнообразия выделяют три основных типа [16].

1. Выбор наиболее полезной альтернативы, например, альтернативы, обеспечивающей достижение поставленной цели в заданные сроки с минимальными затратами ресурсов.

2. Классификация альтернатив, т. е. объединение их в классы по заданной системе признаков.

3. Определение порядка на множестве альтернатив, т. е. ранжирование альтернатив по важности, затратам времени и ресурсов или другим признакам.

5.1. Основные проблемы принятия решений

Любое решение, несмотря на их разнообразие по содержанию, важности или сложности, направлено на достижение некоторой цели. Вариантов достижения цели, как правило, имеется несколько. Обозначим множество вариантов достижения цели X. Элементы этого множества называют альтернативами, стратегиями, планами и т. п.

Какой же из этих вариантов лучший? Как найти этот вариант? Для ответа на эти вопросы необходимо установить правило (принцип) выбора из множества альтернатив некоторого подмножества приемлемых вариантов, т. е. вариантов, более предпочтительных относительно достижения цели. Принцип такого выбора обозначим W. Этот принцип позволяет оценить эффективность каждой альтернативы – меру достижения цели при выборе той или иной альтернативы. Эффективность каждой альтернативы изменяется в зависимости от условий, в которых осуществляется выбор альтернативы. Эти условия называют состояниями внешней среды (иногда состояниями природы). Множество состояний внешней среды обозначим Y. Таким образом, в задачах принятия решений должны быть определены множество альтернатив X, принцип W выбора приемлемого варианта (вариантов) и множество состояний внешней среды Y.

Задачи принятия решений различаются в зависимости от информации о множестве X, принципе W и состоянии внешней среды Y. В результате принятия решения определяется некоторый компромисс, основанный на этой информации. Цель принятия решений – гарантировать логическую непротиворечивость выбранных концепций компромисса и основанных на них процедур выбора.

Когда число допустимых решений невелико и принцип W имеет «хорошие свойства», то решение можно найти методом прямого перебора вариантов или методами математического анализа. Однако при возрастании числа допустимых решений такая возможность практически может совсем исчезнуть. Для решения таких задач применяют специальный математический аппарат, называемый математическим программированием.

Часто решение ищется, когда не все условия известны заранее, некоторые из них содержат элемент неопределенности. Наиболее простым и благоприятным для расчетов является случай, когда неизвестные условия или факторы есть случайные величины или случайные функции, о которых имеются статистические данные, характеризующие их распределение. Наиболее трудным является случай, когда условия не обладают статистической устойчивостью, т. е. функции распределения неизвестных условий либо совсем неизвестны, либо вовсе не существуют. В этих условиях применяют эвристические методы.

В наиболее общем виде задача принятия решений формулируется в терминах и понятиях системно-целевого подхода. Процедура достижения цели рассматривается как некоторая система, функционирующая в некоторой среде. Варианты достижения цели интерпретируются как алгоритмы функционирования системы, не оказывающие влияния на состояние среды. Состояние системы в таких условиях определяется состоянием среды и алгоритмом функционирования.

В задачах принятия решений считается, что каждому фиксированному состоянию среды при фиксированном алгоритме функционирования системы соответствует единственный вариант достижения цели. Выбор конкретного алгоритма функционирования, при заданных ресурсах наилучшим образом соответствующего достижению цели системы, с учетом информации о состоянии среды называют принятием решения.

Математическая модель задачи принятия решения в общем случае формулируется следующим образом.

Пусть задано:

– X – множество допустимых алгоритмов функционирования (альтернатив, стратегий, решений, планов и т. д.), состоящее не менее чем из двух элементов;

– Y – множество возможных состояний среды;

– U – множество возможных исходов – состояний системы.

Состояние системы полностью определяется выбором алгоритма функционирования и состоянием среды, т. е. каждой паре (x, y), $x \in X$, $y \in Y$ соответствует определенный исход $u \in U$. Это значит, что существует отображение $F: X \times Y \to U$, или $u = F(x, y)$, которое определяет исход как функцию F.

Каждый вариант принимаемого решения, приводящий к некоторому исходу, необходимо оценить и выбрать наиболее полезный с точки зрения лица, принимающего решения (человек, фирма, система). Числовую оценку решения получают в результате анализа ***целевой функции*** $\varphi(u) = \varphi(F(x, y))$, такой, что если альтернатива u_k предпочтительнее альтернативы u_m, то $\varphi(u_k) > \varphi(u_m)$.

Существуют и другие методы оценки принимаемых решений, основанные на определении отношения предпочтения всевозможных пар решений, определении функции выбора или на сужении множества решений на основании доминирования по Парето.

Процедуру принятия решения обычно делят на три шага:

– построение математической модели задачи, т. е. определение отображения $F: X \times Y \to U$ и метода оценки решения;

– формулирование принципа выбора решения и нахождение решения;

– анализ результатов и выводы.

Многообразие ситуаций, в которых имеется необходимость принятия решений, порождает многообразие методов принятия решений. В настоящее время не существует универсального метода принятия решения. Различают следующие варианты задач принятия решения.

1. Детерминированные задачи принятия решений – принятие решения в условиях, когда состояние среды известно.

2. Принятие решения в условиях риска – данные о состоянии среды имеют стохастический характер.

3. Принятие решения в условиях неопределенности – известно только множество возможных состояний среды и функция реализации.

4. Принятие решения в игровых условиях – среда рассматривается как одна из активных подсистем.

5. Многокритериальные задачи принятия решений.
6. Многошаговые процедуры принятия решения.
7. Статистические процедуры принятия решений.
8. Коллективные решения.

В сложных ситуациях применяют любые корректные комбинации перечисленных вариантов.

Основным ресурсом, необходимым для принятия решения, является информация – совокупность сведений, данных, знаний и т. д. С течением времени информация стареет, рассеивается, искажается и подвергается другим негативным преобразованиям. Поэтому для эффективного и правильного выбора (принятия решения) необходимо, чтобы информация была актуальной (не устаревшей), правильной и точной (неискаженной), содержательной (содержать сведения, достаточные для принятия решения).

5.2. Классификация решений

Различным критериям классификации решений соответствуют различные варианты их классификации. Список вариантов задач принятия решений, приведенный в конце подразд. 5.1, можно рассматривать как вариант классификации решений. Другой вариант классификации приведен в табл. 1.

Таблица 1

Один из вариантов классификации решений

Критерий классификации	Результат классификации
По решаемым задачам	Информационные; организационные; оперативные
По принципам выработки решения	Алгоритмические; эвристические
По методам обоснования	Аналитические; статистические; математического программирования; игровые; синтетические и др.
По характеру исходной информации	Решения принимают в условиях: определенности (полной информации); неопределенности (неполной информации)

Приведенный список вариантов классификации решений не является исчерпывающим. Могут быть и другие варианты, основанные на критериях, не включенных в данный список.

5.3. Детерминированные задачи принятия решений

Детерминированные задачи принятия решений – задачи, в которых состояние среды известно и поэтому они сводятся к определению стратегии, доставляющей экстремум целевой функции, т. е. к задачам нахождения экстремума целевой функции. Такие задачи называют задачами оптимизации. Математические методы оптимизации начали развиваться в классической математике. Однако из-за отсутствия эффективных вычислительных средств на практике применялись только в самых простых случаях. Компьютеризация современного общества вызвала бурное развитие теории оптимизации и обеспечила ее широкое практическое использование.

5.3.1. Задачи оптимизации, общие сведения

Быстрое развитие и усложнение техники, увеличение масштабов и стоимости различных мероприятий, внедрение методов автоматизации и т. д. – все это приводит к необходимости научного анализа сложных целенаправленных процессов для выработки и принятия правильных и обоснованных решений. Под оптимизацией обычно понимают методы формализации и решения задач о выборе наилучшего в заранее предписанном смысле способа действия в конкретных условиях.

Классическая постановка задачи оптимизации состоит в следующем. В некотором пространстве P находят непустое множество M точек этого пространства, называемое допустимым множеством. Это множество составляют возможные варианты решений. В точках этого множества определяется вещественная функция $\varphi(x)$, называемая целевой функцией. Задача оптимизации состоит в нахождении точки $x_0 \in P$ (выборе варианта решения), где функция $\varphi(x)$ принимает экстремальное – максимальное или минимальное значение. Состояние среды считается точно известным, и поэтому множество Y содержит один элемент, соответствующий фактическому состоянию среды. Задача оптимизации может иметь одно решение или множество решений, или не иметь ни одного решения.

При решении задач оптимизации следует различать *абсолютный и локальный экстремумы*. В первом случае при достижении целевой функцией φ значения абсолютного экстремума, например, максимума Φ, неравенство $\Phi \geq \phi$ выполняется во всех точках области ее определения, а в случае локального максимума это неравенство выполняется лишь в некоторой окрестности точки, где достигается локальный экстремум.

Необходимым элементом решения задач оптимизации служит поиск экстремума функции одной переменной. Для этого обычно приходится решать уравнение вида $f'(x)=0$. На практике решение этого уравнения может оказаться сложным. Поэтому существенное значение приобретают методы оптимизации, не требующие вычисления производных. Когда о функции $f(x)$ известно лишь, что она непрерывна на отрезке $[a,b]$, то для определения экстремума необходимо исследовать поведение функции во всех точках отрезка. Так как практическое нахождение экстремума может быть приближенным, то поведение функции исследуют в конечном числе точек, способом выбора которых различаются методы оптимизации. В большинстве случаев задача сводится к построению последовательности вложенных отрезков $[a_n,b_n]$, стягивающихся к точке экстремума, или к построению последовательности точек $\{x_k\}$, где выполняются условия $f(x_{k+1}) \leq f(x_k)$ при поиске минимума и $f(x_{k+1}) \geq f(x_k)$ – при поиске максимума.

Для решения задач оптимизации разработан целый арсенал математических методов. На рис. 19 приведена схема классификации методов и задач оптимизации [9, 5].

При решении задач оптимизации используется понятие ***градиента***. Для любой дифференцируемой функции $f(x)$ ее градиентом в точке x называется вектор

$$\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \ldots, \frac{\partial f(x)}{\partial x_n} \right). \qquad (9)$$

Рис. 19. Классификация методов оптимизации

Вектор градиента $\nabla f(x)$ задает в точке x направление быстрейшего увеличения функции $f(x)$, а противоположный ему – направление $(-\nabla f(x))$, называемое антиградиентом, – направление скорейшего уменьшения этой функции. Точки, в которых градиент функции обращается в нуль, называются стационарными точками этой функции. Если экстремум функции $f(x)$ достигается внутри области D, то в точке оптимума u и ее градиент обращается в нуль, т. е.

$$\left.\frac{\partial f(x)}{\partial x_1}\right|_{x=u} = 0, \ldots, \left.\frac{\partial f(x)}{\partial x_n}\right|_{x=u} = 0. \tag{10}$$

Однако выполнение этих условий не является достаточным, так как они выполняются и в так называемых седловых точках.

Классические методы оптимизации применяют, если известно аналитическое выражение функции $f(x)$ и известно, что она, по крайней мере, дважды дифференцируема. Для определения экстремума функцию $f(x)$ в окрестности экстремальной точки x^* разлагают в ряд Тейлора:

$$f(x) - f(x^*) = \sum_{i=1}^{m} \frac{\partial f(x^*)}{\partial x_i}(x_i - x_i^*) + \frac{1}{2}\sum_{i=1}^{m}\sum_{k=1}^{m} \frac{\partial^2 f(x^*)}{\partial x_i \partial x_k}\left(x_i - x_i^*\right)\left(x_k - x_k^*\right). \quad (11)$$

Если x^* – точка максимума (минимума), то линейные члены в этом выражении равны нулю и, следовательно, сумма членов со вторыми частными производными должна быть отрицательной (положительной), т. е. $\overline{\Delta x}^T \cdot \Gamma \cdot \overline{\Delta x} < 0$ – в точке максимума и $\overline{\Delta x}^T \cdot \Gamma \cdot \overline{\Delta x} > 0$ – в точке минимума. В этих выражениях символом T обозначена операция транспонирования; Δx – вектор-столбец, составленный из разностей $x_i - x_i^*$; Γ – матрица Гессе, составленная из вторых производных функции $f(x)$ и имеющая вид:

$$\Gamma = \begin{Bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2}, \frac{\partial^2 f(x)}{\partial x_1 \partial x_2}, \ldots, \frac{\partial^2 f(x)}{\partial x_1 \partial x_m} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1}, \frac{\partial^2 f(x)}{\partial x_2^2}, \ldots, \frac{\partial^2 f(x)}{\partial x_2 \partial x_m} \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ \frac{\partial^2 f(x)}{\partial x_m \partial x_1}, \frac{\partial^2 f(x)}{\partial x_m \partial x_2}, \ldots, \frac{\partial^2 f(x)}{\partial x_m^2} \end{Bmatrix}. \quad (12)$$

Аналитические методы оптимизации требуют аналитического представления производной целевой функции, что может оказаться невозможным. Поэтому на практике часто используют численные методы оптимизации.

5.3.2. Принятие решений в задачах без неопределенностей

В задачах оптимизации без неопределенностей наиболее простая ситуация возникает, когда целевая функция есть функция одной переменной. В этом случае выбор осуществляется по одному критерию, и для принятия решения необходимо установить, существует ли оптимальное решение и как его найти.

Когда множество допустимых альтернатив конечно, решение существует (в конечном множестве всегда существует наибольший и наименьший элементы) и его можно найти методом прямого перебора элементов. Трудности возникают лишь из-за обширности множества альтернатив.

Когда множество альтернатив бесконечно, ответ зависит от свойств этого множества и свойств целевой функции. В приложениях задача нахождения оптимального решения сводится к задаче нахождения экстремума функции одной переменной в некоторой допустимой области.

Это классическая математическая задача, методы, решения которой достаточно хорошо известны. Для нахождения экстремума функции нужно продифференцировать ее по аргументам, приравнять производные нулю и решить полученную систему уравнений. В результате будут найдены критические точки функции, среди которых и находятся точки экстремума.

Таким образом, в рассматриваемом случае задача отыскания оптимального решения сводится к задаче поиска экстремума функции W, она может быть весьма сложной, но возникающие трудности не являются принципиальными, они всегда могут быть преодолены вычислительными средствами.

Поиск оптимальных решений для функций многих переменных в принципе аналогичен поиску оптимального решения для функции одной переменной, но реализация может быть существенно сложнее. В качестве примера рассмотрим задачу: необходимо найти экстремум целевой функции

$$f(\bar{x}) \rightarrow max \text{ в области, где } \overline{g(\bar{x})} < \bar{b}, \quad x_i \geq 0.$$

Максимальное значение $f(\bar{x})$, если оно существует, может быть в точках, принадлежащих множеству внутренних точек области допустимых решений, в которых все частные производные функции f равны нулю, множеству точек границы допустимой области и множеству точек допустимой области, где функция не дифференцируема. Для решения таких задач не существует единого эффективного алгоритма. Поэтому алгоритмы разрабатываются для отдельных типов задач. Методы принятия решений в задачах без неопределенностей достаточно подробно рассматриваются в курсе математического анализа

5.3.3. Принятие решений в задачах без неопределенностей при наличии ограничений

Рассмотрим теперь задачу нахождения оптимального решения при заданных ограничениях. Пусть требуется найти экстремум функции $f(x_1, x_2, ..., x_n)$ при условиях

$$\begin{cases} g_1(x_1, x_2, ..., x_n) = b_1, \\ g_2(x_1, x_2, ..., x_n) = b_1, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ g_m(x_1, x_2, ..., x_n) = b_m \end{cases}. \qquad (13)$$

Эти условия определяют область D допустимых решений.

Для решения таких задач применяют известный метод неопределенных множителей Лагранжа. Точка условного экстремума является стационарной точкой функции Лагранжа:

$$L(x_1, x_2, ..., x_n, \lambda_1, ..., \lambda_m) = f(x_1, x_2, ..., x_n) + \sum_{j=1}^{m} \lambda_j (g_j(x_1, x_2, ..., x_n), \qquad (14)$$

где λ_j – переменные, называемые множителями Лагранжа.

Обычно предполагается, что $n > m$ и разность $n - m$ называют числом степеней свободы данной задачи. В прикладных задачах $f(x_1, x_2, ..., x_n)$ часто интерпретируется как доход или стоимость, а правые части b_i, $i = 1, 2, ..., m$ – как затраты некоторых ресурсов [22, 21]. Тогда множители λ_i – отношение единицы стоимости к единице ресурса с номером i. Они показывают, как изменится максимальный доход или стоимость, если количество ресурса с номером i увеличится на единицу.

Таким образом, для нахождения условного экстремума необходимо решить систему уравнений, полученную приравниванием к нулю всех частных производных функции Лагранжа. Например, для функции $f(x, y)$ при одном условии (ограничении) $g(x, y) = 0$ функция Лагранжа имеет вид:

$$L(x, y, \lambda) = f(x, y) + \lambda g(x, y). \tag{15}$$

Следовательно, ищется экстремум функции $f(x, y)$ на линии, уравнение которой $g(x, y) = 0$.

Система уравнений, необходимая для нахождения условного экстремума, имеет вид:

$$\begin{cases} \dfrac{\partial f}{\partial x} + \lambda \dfrac{\partial g}{\partial x} = 0, \\ \dfrac{\partial f}{\partial y} + \lambda \dfrac{\partial g}{\partial y} = 0, \\ g(x, y) = 0. \end{cases} \tag{16}$$

Замечая, что $\overline{\nabla f} = \left(\dfrac{\partial f}{\partial x}, \dfrac{\partial f}{\partial y}\right)$, а $\overline{N} = \left(\dfrac{\partial g}{\partial x}, \dfrac{\partial g}{\partial y}\right)$ – вектор нормали N линии $g(x, y) = 0$, делаем вывод о том, что первые два уравнения эквивалентны векторному уравнению

$$\overline{\nabla f} = -\lambda \cdot \overline{N}. \tag{17}$$

Это означает, что точка условного экстремума находится в той точке линии g, где векторы $\overline{\nabla f}$ и $\overline{N}$ коллинеарные между собой (рис. 20).

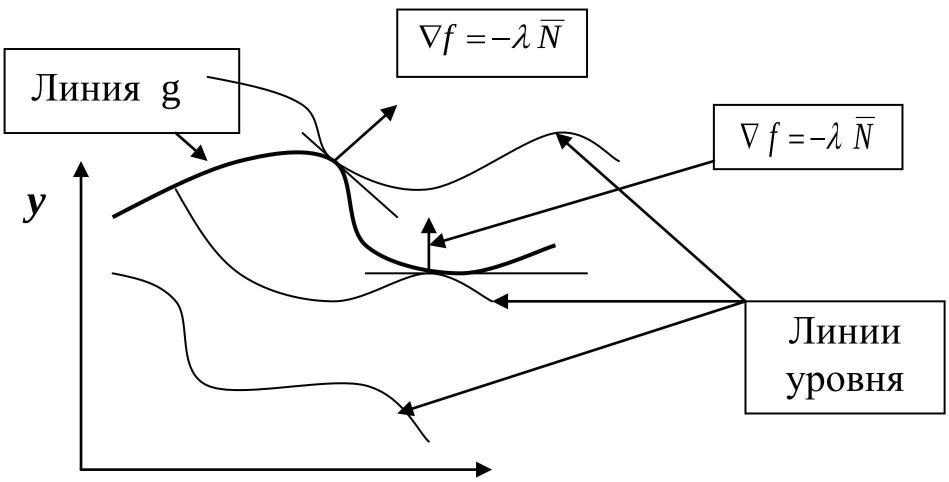


Рис. 20. Геометрическая интерпретация задачи на условный экстремум

В общем случае градиент целевой функции должен выражаться линейной комбинацией нормалей к гиперповерхностям $g_j(x_1, x_2, ..., x_m) = b_j$, $j = \overline{1, m}$.

Пример.

Пусть для ликвидации аварийной ситуации можно привлечь две фирмы. Стоимость ликвидации аварийной ситуации в этих фирмах обозначим $p_1(x)$ и $p_2(x)$, где x – количество ресурсов, необходимых для ликвидации аварийной ситуации. Всего имеется C единиц ресурсов. Как распределить ресурсы между фирмами, чтобы стоимость ликвидации аварийной ситуации была минимальной?

Пусть первая фирма получает x_1, а вторая – x_2 ресурсов соответственно.

Целевая функция задачи:

$$f(x_1, x_2) = p_1(x_1) + p_2(x_2) = min. \qquad (18)$$

Условие, которому должно удовлетворять решение:

$$g(x_1, x_2) = x_1 + x_2 = C. \qquad (19)$$

Функция Лагранжа

$$L(x_1, x_2) = f(x_1, x_2) + \lambda\, g(x_1, x_2) = p_1(x_1) + p_2(x_2) + \lambda(x_1 + x_2 - C). \qquad (20)$$

Находим частные производные функции Лагранжа по x_1, x_2, приравниваем их нулю и, добавляя уравнение $C - x_1 - x_2 = 0$, получаем систему уравнений

$$\left.\begin{aligned} &\frac{\partial L}{\partial x_1} = \frac{\partial f}{\partial x_1} - \lambda = \frac{\partial p_1}{\partial x_1} - \lambda = 0, \\ &\frac{\partial L}{\partial x_2} = \frac{\partial f}{\partial x_2} - \lambda = \frac{\partial p_2}{\partial x_2} - \lambda = 0, \\ &\frac{\partial L}{\partial \lambda} = -C + x_1 + x_2 = 0, \end{aligned}\right\} \qquad (21)$$

из решения которой найдем значения x_1, x_2 *и* λ. Для численного решения необходимо явно указать функции $p_1(x)$ и $p_2(x)$ и значение C.

Рассмотренные методы неприменимы, когда целевая функция и ограничения линейны, т. е. ее производные функции Лагранжа по всем аргументам постоянны и нигде в нуль не обращаются. Для решения таких задач применяют специальный математический аппарат линейного программирования.

5.3.4. Численные методы гладкой оптимизации

Численные методы оптимизации применяются, когда аналитическая оптимизация невозможна или очень сложна. Численные методы позволяют получать приближенные решения, но практически с любой заданной точностью.

Содержание этих методов состоит в следующем [22, 23]. Пусть непрерывно дифференцируемая целевая функция f задана во всех точках пространства. Для произвольной точки x, в которой $\nabla f \neq 0$, он задает направления градиента $\nabla f = g$ и антиградиента $(-\nabla f) = -g$. Сдвигаясь от точки x на малый шаг e в направлении единичного вектора d, можем записать неравенства:

$$f(x+eg) \geq f(x+ed) \geq f(x-eg). \tag{22}$$

Это означает, что движение в направлении градиента обеспечивает наибольший рост, а в направлении антиградиента – наибольшее уменьшение целевой функции. Эти направления называют направлениями *наискорейшего подъема* и *наискорейшего спуска* функции в заданной точке. Следовательно, отправляясь от заданной точки x^0, строим последовательность точек $\{x_i\}$ так, что перемещение от каждой точки x^{i-1} к следующей точке x^i производится в направлении наискорейшего подъема (при поиске максимума) или наискорейшего спуска (при поиске минимума). Трудность практического применения этой процедуры состоит в выборе величины шага e (при большой величине шага можно «проскочить» точку экстремума), и необходимости построения такой последовательности $\{ x^i \}$, которая сходится к точке экстремума u. Последнее требование означает, что по мере роста i расстояние $\left|x^i - u\right|$ от точки x^i до точки экстремума u должно стремиться к нулю. Это возможно лишь, если по мере приближения к точке экстремума величина шага e также стремится к нулю. Методы численной оптимизации отличаются способом определения шага *e*. Рассмотрим некоторые численные методы оптимизации.

Пропорционально-градиентный метод

Длина шага ε_i выбирается пропорциональной длине вектора градиента в точке x^{i-1}:

$$\varepsilon_i = \alpha \cdot \left|\nabla f\left(x^{i-1}\right)\right|, \ \alpha > 0, \ \alpha = const, \tag{23}$$

откуда следует, что

$$x^i = x^{i-1} \pm \alpha \cdot \nabla f\left(x^{i-1}\right), \tag{24}$$

знак (+) выбирается при поиске максимума, а знак (–) – при поиске минимума функции. При произвольном выборе начальной точки x^0 эта процедура необязательно приводит к экстремальной точке. Такой точкой может быть любая стационарная точка, в окрестности которой $\varepsilon_i \to 0$. Однако, если начальное приближение выбрано достаточно близко к точке экстремума, то сходимость может быть обеспечена подходящим выбором α.

Полношаговый градиентный метод

В этом методе каждый шаг подъема или спуска делается на максимально возможную длину, обеспечивающую требуемое направление изменения значения функции (ее увеличение или уменьшение). Поясним сказанное на примере.

Пусть требуется найти минимум функции $W = u^2 + q \cdot v^2$. Изобразим эту функцию на плоскости ее линиями уровня (рис. 21). Из начальной точки x^0 движение по направлению антиградиента происходит до точки x^1, в которой полупрямая, исходящая из точки x^0 в направлении антиградиента, касается некоторой линии уровня. Этот процесс повторяется в точке x^1 и т. д. В результате получаем ряд точек x^i, $i = 0, 1, 2, \ldots$, сходящийся к точке экстремума (минимума). В зависимости от способа определения точки касания получают различные версии рассматриваемого метода.

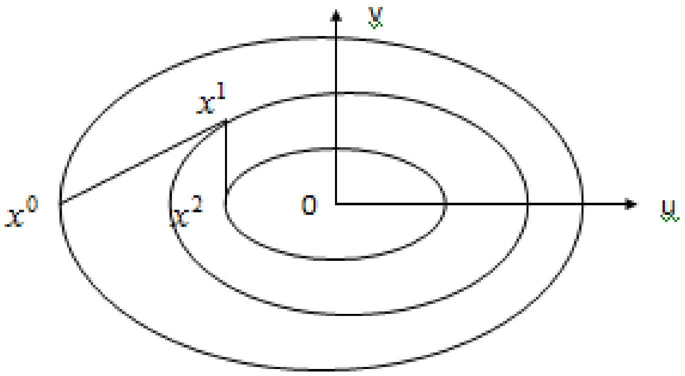


Рис. 21. Линии уровня функции W

Метод покоординатного спуска или подъема

При оптимизации недифференцируемых функций шаги выполняют по направлению одной из координатных осей. Для этого исследуют значение целевой функции на очередном шаге оптимизации в точке, одна из координат которой (с номером i) получила приращение $\pm\delta_i$, выбирая знак приращения так, чтобы значение функции изменялось в нужном направлении. Величину δ_i выбирают как в полношаговых методах, добиваясь максимально возможного изменения функции в нужном направлении. Сдвиги выполняют по всем координатам. Повторяя эту процедуру достаточное число раз, можно найти хорошее приближение экстремума целевой функции (если, конечно, такой экстремум существует).

5.3.5. Многокритериальные задачи принятия решений

В большинстве практических задач оценивание альтернатив по одному критерию оказывается упрощенным и малопригодным. Поэтому возникает необходимость их оценивания не по одному, а по нескольким качественно различным критериям, например, техническим, экономическим, социальным, экологическим и др. Гипотетически одна из альтернатив может оказаться лучшей по всем критериям. Очевидно, что она и будет наилучшей. Однако такие ситуации практически встречаются очень редко, и возникает вопрос, как же тогда принимать решение?

На рис. 22 приведена классификация многокритериальных задач принятия решения [1].

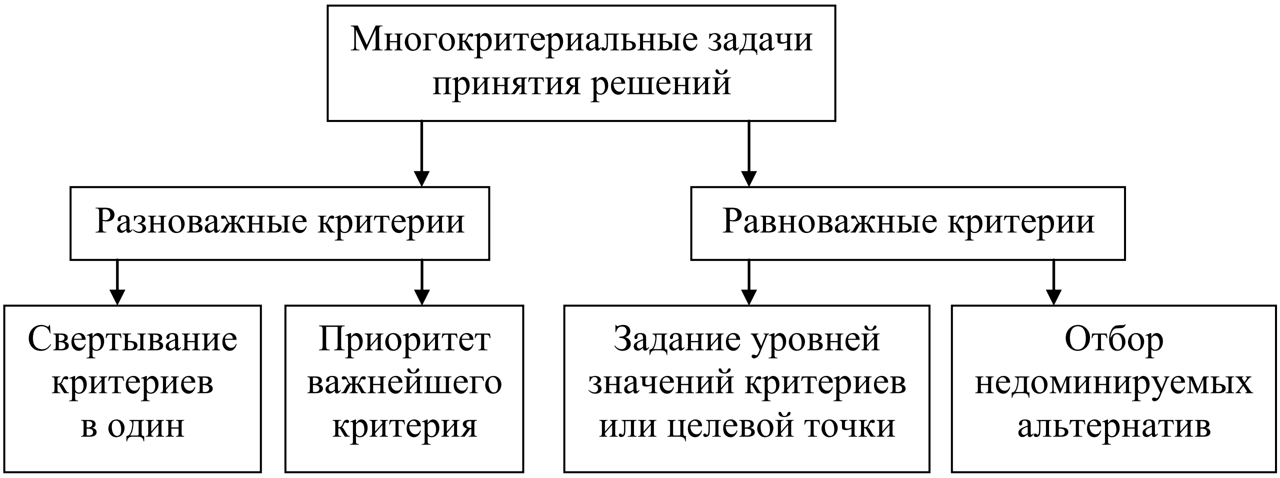


Рис. 22. Классификация многокритериальных задач принятия решений

Рассмотрим наиболее употребляемые методы решения многокритериальных задач.

Отбор недоминируемых альтернатив

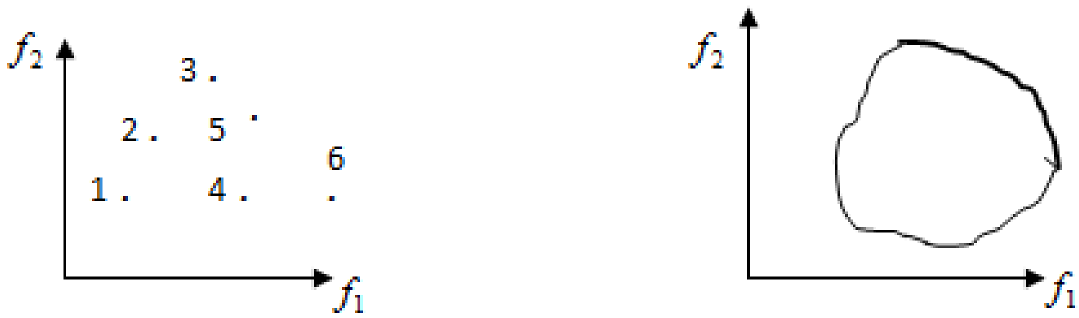
Критерий выбора альтернатив задается вектором оценки $\bar{f} = (f_1, f_2, ..., f_m)$, где f_j – оценка возможного решения по критерию с номером j. В результате этого сравнение любых двух решений заменяется сравнением их векторных оценок. Сравнение векторных оценок осуществляется на основании принципа доминирования по Парето [1, 16, 24].

Векторную оценку $\bar{f} = (f_1, f_2, ..., f_m)$ называют доминирующей по Парето векторную оценку $\bar{g} = (g_1, g_2, ..., g_m)$, если для всех $j = 1, 2, \ldots, m$ выполняется неравенство $f_j \geq g_j$, причем хотя бы для одного значения j последнее неравенство строгое. В результате сравнения решений все доминируемые (худшие по всем критериям) альтернативы отбрасываются, а все прочие (недоминируемые) оставляются. Они образуют множество недоминируемых альтернатив или множество Парето-оптимальных решений.

Пусть имеется некоторое множество Q векторных оценок. Векторная оценка из этого множества называется Парето-оптимальной, если в Q не существует никакой другой векторной оценки, доминирующей по Парето данную оценку.

Парето-оптимальность оценки означает, что она не может быть улучшена ни по одному из критериев, составляющих векторную оценку, без ухудшения, по какому-либо другому критерию, входящему в векторную оценку.

Для наглядности рассмотрим векторную оценку, составленную из двух критериев, каждый из которых желательно максимизировать. В этом случае векторная оценка каждого возможного решения изображается точкой на координатной плоскости. Когда множество допустимых решений дискретно, получим изображение, показанное на рис. 23, *а*.



а *б*

Рис. 23. Векторная оценка возможного решения:
а – множество допустимых решений дискретно;
б – множество допустимых решений непрерывно

Очевидно, что Парето-оптимальными являются оценки 3, 5 и 6 (рис. 23, *а*). Если множество допустимых решений непрерывно, то их векторные оценки заполняют некоторую область (рис. 23, *б*) и множество Парето-оптимальных решений изображается частью границы области, выделенной жирной линией. Таким образом, Парето-оптимальных оценок, как правило, получается множество, и выделить среди них единственную оптимальную оценку без дополнительной информации не представляется возможным, так как любые два Парето-оптимальных решения не сравнимы относительно доминирования по Парето. Это значит, что для любых двух Парето-оптимальных оценок всегда найдутся такие два критерия, по одному из которых предпочтительнее первый исход, а по другому – второй. В такой ситуации для принятия решения выбор оптимального решения из множества Парето-оптимальных осуществляет ЛПР или на основе дополнительной информации выполняется сужение множества Парето-оптимальных решений с помощью формальных и эвристических процедур. Рассмотрим некоторые из этих процедур.

Указание границ критериев

Дополнительная информация о некоторых оптимальных решениях задается указанием границ критериев. Например, в случае поиска решения, доставляющего максимум по всем критериям, такие неравенства имеют вид $f_j(a) \geq q_j, \ j = \overline{1,m}$, где q_j – нижняя граница критерия с номером j. При задании нижних границ критериев оптимальным считается только такое Парето-оптимальное решение, у которого по каждому из критериев с заданными нижними границами оценки не ниже этих оценок. В результате Парето-оптимальное множество сужается, однако, окончательный выбор из суженного множества производит лицо, принимающее решение, т. е. решение остается субъективным.

Субоптимизация

В этом случае из множества критериев выбирается один, наиболее важный, а по остальным назначаются границы. В качестве оптимального принимается решение, оптимальное по выбранному критерию при условии выполнения всех неравенств, определяющих границы остальных критериев. В результате применения этой процедуры задача многокритериальной оптимизации сводится к задаче однокритериальной оптимизации на суженном допустимом множестве. Выбор наиболее значимого критерия и назначение границ остальных критериев носит субъективный характер.

Лексикографическая оптимизация

Все критерии ранжируются и упорядочиваются по их относительной важности. Затем отбираются все решения, которые имеют оптимальную оценку по важнейшему критерию. Если такое решение одно, то его и считают оптимальным. Если таких решений несколько, то среди них выбирают те, которые имеют оптимальную оценку по второму важнейшему критерию и т. д., пока не останется единственное решение. Ранжирование критериев в этой процедуре играет важнейшую роль, так как первый, принятый за важнейший, критерий в значительной мере определяет выбор оптимального решения, а само ранжирование критериев зачастую имеет эвристический характер.

Сведение многокритериальной задачи к однокритериальной

Многокритериальные задачи принятия решений могут быть сведены к однокритериальным. Для этого множество заданных критериев необходимо преобразовать в один обобщенный критерий, выражающий, с точки зрения ЛПР, полезность заданной системы критериев. Основная трудность решения этой задачи состоит в необходимости соизмерения критериев, имеющих различную природу и определяемых в различных шкалах. Если эта трудность разрешена и все исходные критерии W_i $(i = 1,2,...,k)$ выражены в соизмеримых единицах, то для построения обобщенного критерия наиболее часто используются следующие варианты [18, 1].

1. Обобщенный критерий представляется дробью

$$U = \frac{W_1 \cdot W_2 \cdot ... \cdot W_m}{W_{m+1} \cdot W_{m+2} \cdot ... \cdot W_k}, \qquad (25)$$

в числителе дроби ставят те критерии, которые желательно увеличить, а в знаменателе – те, которые желательно уменьшить (например, отношение эффективность / стоимость).

2. Обобщенный критерий представляется в виде взвешенной суммы отдельных критериев эффективности

$$U = \sum_{i=1}^{k} a_i \cdot W_i, \qquad (26)$$

где a_i – весовые коэффициенты, имеющие знак «+» при тех критериях, которые желательно увеличить, и знак «–» – при тех, которые желательно уменьшить.

Общим недостатком составных критериев является то, что недостаток эффективности по одному из них всегда можно компенсировать за счет другого. Поэтому их использование может привести к неправильным рекомендациям.

3. Из множества показателей эффективности выделяют один, наиболее важный и стремятся получить оптимальное решение лишь по этому

единственному критерию. На остальные показатели эффективности накладывают ограничения вида:

$$W_i \geq w_i, (i = 2,..., m);\ W_j \leq w_j, (j = m+1,..., k) \quad (27)$$

и включают в число заданных условий.

4. Показатели эффективности располагают в порядке убывающей важности. Для простоты будем считать, что каждый из них нужно обратить в максимум (если это не так, достаточно изменить знак показателя). Сначала ищется решение, обращающее показатель W_1 в максимум. Затем назначается, исходя из практических соображений и точности исходных данных, некоторая уступка ΔW_1, ценой которой можно добиться максимума показателя W_2, т. е. на W_1 накладывается ограничение $W_1 \geq W_1^* - \Delta W_1$, где W_1^* – максимально возможное значение W_1, и при этом ограничении ищем решение, при котором достигается максимум показателя W_2. Этот процесс продолжается по мере необходимости. Этот метод поиска компромиссного решения хорош тем, что всегда известно, ценой какой уступки в одном показателе достигается выигрыш в другом. Значение уступки определяет ЛПР.

С формальной точки зрения построение обобщенного критерия представляет собой процедуру агрегирования частных критериев в один критерий. Главное требование, выполнение которого необходимо, состоит в сохранении для обобщенного критерия отношения доминирования. Это значит, что

$$(W_1, W_2,..., W_m) \overset{Par}{>} (U_1, U_2,..., U_m) \Rightarrow F(W_1, W_2,..., W_m) \overset{Par}{>} F(U_1, U_2,..., U_m),$$

где F – агрегат-оператор обобщенного критерия.

Рассматривая обобщенный критерий как функцию m переменных (частных критериев), определим для нее поверхности уровня:

$$F(W_1, W_2,..., W_m) = c. \quad (28)$$

При любых значениях частных критериев, удовлетворяющих данному уравнению, обобщенный критерий не изменяет своего значения, равного c.

Используя это обстоятельство, можно изменять значения частных критериев, сохраняя неизменным значение обобщенного критерия.

Когда мера эффективности критериев выражена в разнородных единицах, например, денежных и временных, то их необходимо привести к стандартной мере эффективности или преобразовать в безразмерные.

В первом случае методика решения этой задачи состоит в следующем. Выбирается наиболее важный критерий и мера эффективности этого критерия принимается в качестве стандартной. Чтобы выразить меру эффективности других критериев через стандартную меру, на плоскости строим график, на котором стандартная мера эффективности отображается по оси абсцисс, а мера эффективности другого критерия – по оси ординат. На основании или имеющегося опыта, или экспертных оценок определяют эквивалентные значения рассматриваемых критериев эффективности, и соответствующие точки используют для построения графика. Например, пусть стандартная мера эффективности (обозначим ее символом x) – затраты в тысячах рублей на мероприятия по охране труда в человеко-машинных системах, а мера эффективности второго критерия (обозначим его y) – вероятность возникновения аварийной ситуации. Пусть установлено, что эквивалентные значения этих критериев составляют:

$$x1 := 0{,}1; \qquad x2 := 1{,}1; \qquad x3 := 2{,}1;$$

$$y1 := 0{,}905; \qquad y2 := 0{,}333; \qquad y3 := 0{,}122.$$

По этим данным нетрудно построить график (рис. 24) зависимости мер эффективности двух критериев.

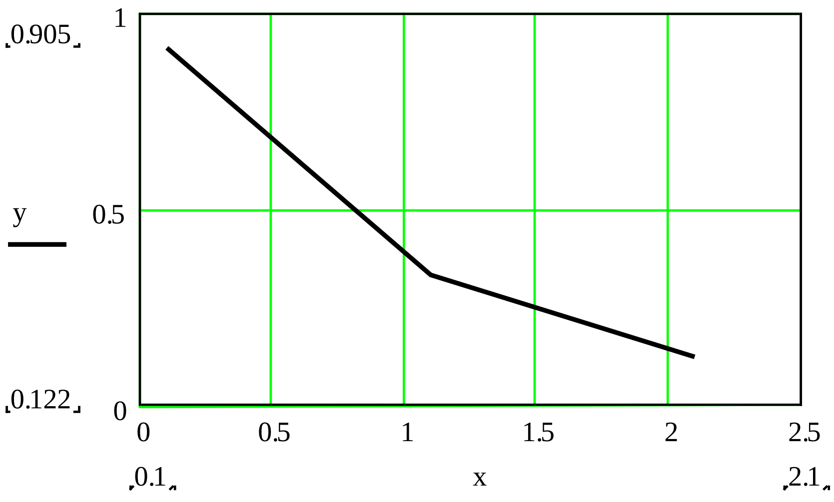


Рис. 24. График зависимости мер эффективности двух критериев

Аналогичные графики строят и для других критериев, обеспечивая, таким образом, выражение всех мер эффективности через меру эффективности основного критерия.

Во втором случае каждый критерий преобразуют к безразмерному виду, например, разделив значение критерия на величину диапазона его изменения.

Контрольные вопросы

1. Сформулировать задачу на принятие решения.
2. Перечислить основные варианты задач на принятие решения.
3. Привести примеры задач на принятие решений.
4. Дать определение среды в задачах принятия решений.
5. Дать формальное описание задачи принятия решений.
6. Дать определение функции реализации решения.
7. Дать определение целевой функции принятия решения.
8. Описать процедуру принятия решения.
9. Сформулировать задачу принятия решений в заданных условиях, когда целевая функция зависит от одной переменной.
10. Сформулировать задачу принятия решений в заданных условиях, когда целевая функция зависит от многих переменных.
11. Сформулировать задачу принятия решений при наличии ограничений.
12. Дать геометрическую интерпретацию задачи принятия решений при наличии ограничений.
13. Сформулировать задачу принятия решения при многих критериях.
14. Сформулировать принцип доминирования решения по Парето.
15. Дать определение Парето-оптимального множества решений. Привести геометрическую интерпретацию.
16. Сформулировать характеристическое свойство Парето-оптимального множества решений.
17. Сформулировать основные варианты преодоления неопределенности выбора решений из Парето-оптимального множества.
18. Объяснить смысл процедуры сужения Парето-оптимального множества, основанной на задании границ критериев выбора.

19. Объяснить смысл процедуры субоптимизации для сужения Парето-оптимального множества.

20. Объяснить смысл процедуры лексикографической оптимизации для сужения Парето-оптимального множества.

21. Дать определение обобщенного критерия оптимизации.

22. Сформулировать задачу алгебраического определения обобщенного критерия по частным критериям. Привести примеры.

23. Сформулировать задачу эвристического определения обобщенного критерия по частным критериям. Привести примеры.

24. Сформулировать необходимое требование при построении обобщенного критерия.

25. Рассказать, как строится обобщенный критерий, если меры эффективности используемых критериев различны.

6. МНОГОШАГОВЫЕ ПРОЦЕДУРЫ ПРИНЯТИЯ РЕШЕНИЙ

В различных областях теории и практики решение целесообразно искать не сразу, а последовательно, шаг за шагом. В этом случае поиск решения рассматривается как процесс, состоящий из нескольких этапов. Если на каждом шаге процедуры принятия решения результат полностью определен, то ее называют детерминированной, а, если не определен, но может быть предсказан с помощью некоторого распределения вероятностей, – стохастической. Многошаговые процедуры принятия решений относятся к задачам математического программирования. Целью математического программирования служит создание, если это возможно, аналитических методов решения экстремальных задач или разработка эффективных численных методов нахождения приближенного их решения, т. е. выбор программы действий. С этим обстоятельством связано и название «математическое программирование».

Под общей задачей математического программирования понимают задачу отыскания экстремума функции $f_0(x)$ при условиях $f_j(x) \le 0$, $j = 1,..., m$; $x \in Q$, где Q – некоторое множество в пространстве векторов x. Функция $f_0(x)$ называется целевой, а множество $X = \{ x \in Q;\ f_j(x) \le 0,\ j = 1,...,m \}$ – допустимым множеством. Отыскание характеристических свойств экстремума в этой задаче и является главным в математическом программировании. Эти свойства экстремума и численные методы решения определяются свойствами задач, которые зависят от свойств функций $f_j(x)$ и множества Q. В связи с этим в математическом программировании рассматривают задачи линейного, нелинейного, выпуклого, квадратичного, целочисленного и т. д. программирования.

6.1. Динамическое программирование, основные понятия и определения

Динамическое программирование – метод оптимизации многошаговых (многоходовых) операций. Одной из первых операций такого рода была задача о коммивояжере [25]. Суть ее состоит в следующем: имеется $n + 1$ городов $A_0, A_1, \ldots, A_n$, $(n > 1)$ с заданными между ними расстояниями $d_{i\,j}$ $(i, j = 0, 1, \ldots, n)$; требуется, отправляясь, например, из A_0, выбрать такой маршрут передвижения $A_0, A_{i1}, A_{i2}, \ldots, A_{in}, A_0$, при котором коммивояжер, побывав в каждом городе по одному разу, вернулся бы в исходный пункт A_0, проделав минимально возможный суммарный путь.

Эта задача может быть в принципе решена методом простого перебора всех возможных маршрутов. Однако число таких маршрутов равно $n!$ и при n, равном нескольким десяткам, практически исключает такую возможность. Проблема, следовательно, состоит в том, чтобы резко сократить перебор, отбрасывая заранее заведомо непригодные множества вариантов. Решение этой проблемы и составляет суть динамического программирования.

Рассмотрим основные принципы динамического программирования на примере задачи о коммивояжере [25]. Пусть имеется всего пять пунктов: $A_0, A_1, \ldots, A_4$, расстояния между которыми заданы в табл. 2.

Таблица 2

Значения расстояний между пунктами A_i, где $i = 0, 1 \ldots 4$

	A0	A1	A2	A3	A4
A0	0	300	250	200	400
A1	300	0	500	350	600
A2	250	500	0	250	200
A3	200	350	250	0	250
A4	400	600	200	250	0

Для определения кратчайшего пути коммивояжера будем развивать варианты его передвижения последовательно, пункт за пунктом. Начнем с вариантов, состоящих из трех участков $A_0, A_{i1}, A_{i2}, A_{i3}$, группируя их по последнему пункту A_{i3}. Некоторые из полученных вариантов можно отбросить, не развивая их дальше. Например, сравнивая варианты A_0, A_1, A_2, A_3 и A_0, A_2, A_1, A_3, имеем для первого суммарный путь: 300 + 500 + 250 = 1 050, а для второго: 250 + 500 + 350 = 1 100. Оба варианта относятся к одному и тому же множеству пунктов и оканчиваются в одном и том же пункте A_3 и в дальнейшем могут развиваться одинаково. Поэтому проигрыш 50 км, полученный для второго варианта, делает его заведомо хуже первого варианта при любом дальнейшем развитии. Следовательно, второй вариант может быть отброшен, а вместе с ним и все получающиеся из него варианты дальнейшего движения.

Результаты рассмотрения вариантов сведем в табл. 3.

Таблица 3

Результаты вариантов решения задачи определения кратчайшего пути

Варианты	Расстояние, км	Перспективен или нет	Варианты	Расстояние, км	Перспективен или нет
A0A2A3A1	850	да	A0A1A2A3	1 050	Да
A0A3A2A1	950	нет	A0A2A1A3	1 100	нет
A0A2A4A1	1 050	Да	A0A1A4A3	1 150	Да
A0A4A2A1	1 100	Нет	A0A4A1A3	1 350	Нет
A0A3A4A1	1 050	Нет	A0A2A4A3	700	Да
A0A4A3A1	1 000	Да	A0A4A2A3	850	Нет
A0A1A3A2	900	Да	A0A1A2A4	1 000	Да
A0A3A1A2	1 050	нет	A0A2A1A4	1 350	нет
A0A1A4A2	1 100	Да	A0A1A3A4	900	Да
A0A4A1A2	1 500	Нет	A0A3A1A4	1 150	Нет
A0A3A4A2	650	Да	A0A2A3A4	750	Нет
A0A4A3A2	900	нет	A0A3A2A4	650	да

Перспективных вариантов получилось 12. Развитие каждого из них однозначно, так как для каждого из них остается лишь один непосещенный пункт. Результаты сравнения перспективных 12 вариантов сведем в табл. 4.

Таблица 4

Результаты сравнения перспективных вариантов

Варианты	Расстояние, км	Перспективен или нет	Варианты	Расстояние, км	Перспективен или нет
A0A2A3A1A4	1 450	Нет	A0A1A2A3A4	1 300	Нет
A0A2A4A1A3	1 400	Нет	A0A1A4A3A2	1 400	Нет
A0A4A3A1A2	1 500	Нет	A0A2A4A3A1	1 050	Да
A0A1A3A2A4	1 100	Да	A0A1A2A4A3	1 250	Да
A0A1A4A2A3	1 350	Нет	A0A1A3A4A2	1 100	Да
A0A3A4A2A1	1 150	нет	A0A3A2A4A1	1 250	нет

В табл. 4 имеем 4 перспективных варианта. Дальнейшее их развитие заключается в возвращении в исходный пункт A_0. Для этих 4 окончательных маршрутов вычисляем суммарные расстояния и записываем в табл. 5.

Таблица 5

Суммарное расстояние маршрутов

Варианты	Расстояние, км	Варианты	Расстояние, км
A0A1A3A2A4A0	1 500	A0A1A2A4A3A0	1 450
A0A2A4A3A1A0	**1 350**	**A0A1A3A4A2A0**	**1 350**

Из этой таблицы видно, что существует два оптимальных маршрута следования коммивояжера, имеющих минимальную из всех вариантов длину, равную 1 350 км.

Рассмотренный пример демонстрирует основной прием динамического программирования – нахождение правил доминирования, которые позволяют производить сравнение вариантов развития последовательностей и заблаговременное отсеивание бесперспективных вариантов.

Методы динамического программирования применяются для решения задач в разнообразных областях науки и техники, но все эти задачи имеют отличительные черты, основные из которых следующие [18].

1. В любой момент времени состояние процесса описывается набором немногих параметров.

2. Операция выбора состоит в преобразовании этого набора параметров в такой же набор с другими численными значениями.

3. Прошлая история системы не имеет значения для определения будущих действий (Марковское свойство).

Содержание задачи динамического программирования состоит в последовательности выборов. Эту последовательность выборов называют ***стратегией***. Стратегии, наиболее желательные с точки зрения какого-либо заранее заданного критерия, называют ***оптимальными стратегиями***. Оптимальная стратегия обладает свойством: каково бы ни было начальное состояние и начальный выбор, остальные выборы должны составлять оптимальную стратегию относительно состояния, возникшего в результате первого выбора. Это свойство выражает ***принцип оптимальности*** динамического программирования.

Рассмотрим управляемую систему, которая под влиянием управления переходит из начального состояния S_0 в конечное состояние S_n. Разобьем процесс управления системой на n шагов, и пусть S_1, S_2,..., S_n – состояния системы после первого, второго,…, n-го шага. Состояние системы после шага с номером k характеризуется параметрами $p_{i,k}$ (i = 1, 2,..., q), которые называют ***фазовыми координатами***. Состояние S_k изображается точкой q-мерного пространства, называемого ***фазовым пространством***. Преобразования системы на шаге с номером k осуществляются в результате некоторых мероприятий u_k, переводящих систему из состояния S_{k-1} в состояние S_k. В задачах динамического программирования считается, что состояние системы после шага с номером k зависит только от предшествующего состояния системы S_{k-1} и управления u_k на данном шаге. Это свойство называют ***отсутствием последействия***. Представим эту зависимость в виде уравнения:

$$S_k = F_k (S_{k-1}, u_k). \tag{29}$$

Эти уравнения называют ***уравнениями состояния***. Функции $F_k(S_{k-1}, u_k)$ считаются известными.

Каждому варианту управления $U = (u_1, u_2, ..., u_k, ..., u_n)$, переводящему систему из состояния S_0 в состояние S_n, соответствует некоторая оценка эффективности:

$$W = W(S_0, U). \tag{30}$$

Эффективность шага с номером k обозначим $w_k(S_{k-1}, u_k)$. В задаче динамического программирования функция эффективности W должна быть аддитивной, т. е.

$$W = \sum_{k=1}^{n} w_k\left(S_{k-1}, u_k\right). \tag{31}$$

На каждом шаге управление может ограничиваться некоторыми условиями. Управления, удовлетворяющие этим ограничениям, называются допустимыми. Управление, при котором достигается экстремум значения W, называют оптимальным управлением.

Дадим геометрическую интерпретацию этой задачи (рис. 25). Для этого заметим, что состояние системы всегда можно описать некоторым числом параметров. Эти параметры принято называть фазовыми координатами системы. Тогда состояние системы можно изображать точкой S с этими координатами в условном фазовом пространстве (пространстве состояний). Размерность этого пространства зависит от числа фазовых координат. Для простоты будем считать, что состояние системы s зависит от двух параметров. Тогда фазовое пространство будет двумерным, а процесс будет изображаться перемещением точки S из $S_0 \in \tilde{S}_0$ в $S_\omega \in \tilde{S}_\omega$ по определенной траектории на фазовой плоскости. Эта траектория и будет изображать управление системой.

Характерным для динамического программирования является определенный методический прием. Процесс перемещения точки S из $\tilde{S}_0$ в $\tilde{S}_\omega$ разделяется на несколько шагов и затем осуществляется пошаговая оптимизация.

Модель динамического программирования называется дискретной или непрерывной, в зависимости от характера изменения управления u_k. В зависимости от числа параметров состояния $p_{i,k}$ $(i = 1, 2,..., q)$ и числа управляющих переменных на каждом шаге различают одномерные и многомерные модели динамического программирования. Число шагов в задаче может быть конечным или бесконечным.

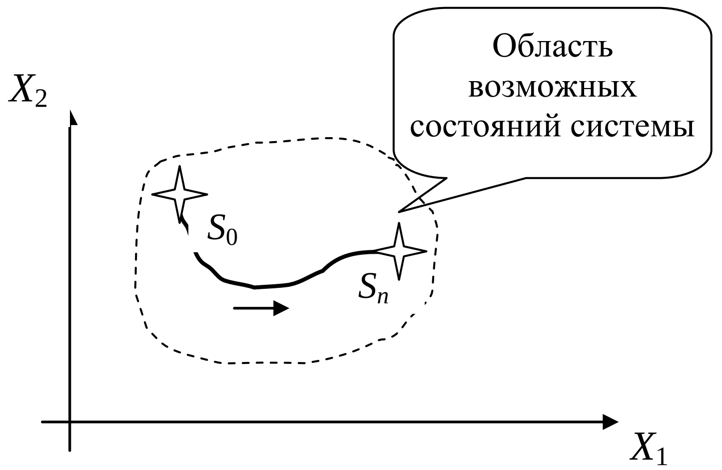


Рис. 25. Геометрическая интерпретация задачи динамического программирования

6.2. Принцип оптимальности. Уравнение Беллмана

Пусть имеется некоторая физическая система S, которая с течением времени меняет свое состояние, т. е. в системе происходит какой-то процесс. Мы можем управлять этим процессом, т. е. влиять на состояние системы. Такую систему называют *управляемой системой*, а способ воздействия на нее – *управлением* U. Предположим, что с процессом связана какая-то наша заинтересованность, выражаемая численно величиной W, которую называют *выигрышем*. Необходимо найти такое управление процессом, чтобы выигрыш был максимальным, т. е. найти

$$W_{\max} = \max_{U}\{W(U)\}. \tag{32}$$

Для решения этой задачи необходимо еще определить начальное S_0 и конечное S_ω состояния системы. Чаще всего эти состояния точно задать не удается, но всегда можно указать области $\tilde{S}_0, \tilde{S}_\omega$, которые их содержат. Таким образом, задача динамического программирования формулируется следующим образом [18].

Из множества возможных управлений U найти такое оптимальное управление u, которое переводит физическую систему S из начального состояния $S_0 \in \tilde{S}_0$ в конечное состояние $S_\omega \in \tilde{S}_\omega$ так, чтобы при этом выигрыш W *был максимальным.* Процедура построения оптимального управления делится на предварительную и окончательную стадии. На предварительной стадии для каждого шага определяется условное оптимальное управление $u(S)$ и условный оптимальный выигрыш $W(S)$ на всех оставшихся шагах, начиная с данного. На окончательной стадии для каждого шага определяется безусловное оптимальное управление.

Таким образом, основное функциональное уравнение динамического программирования на шаге с номером i имеет вид [18]:

$$W_i(S) = \max_{U_i}\{w_i(S,U_i) + W_{i+1}(\varphi_i(S,U_i))\}. \quad (33)$$

Управление $U_i = u_i(S)$, при котором этот максимум достигается, и есть условное оптимальное управление на i-м шаге, а $W_i(S)$ – условный оптимальный выигрыш на всех шагах, начиная с i-го. Из основного функционального уравнения следует, что значение $W_i(S)$ можно определить, если известна следующая за ней функция $W_{i+1}(S)$.

Учитывая это обстоятельство, заметим, что на последнем шаге значение $W_m(S) = \max\limits_{U_m}\{w_m(S,U_m)\}$. Здесь максимум необходимо искать не по всем возможным управлениям, а лишь по тем, которые приводят систему в заданную область конечных состояний $\tilde{S}_\omega$. То управление, при котором достигается максимум выигрыша $W_m(S)$, и есть условное оптимальное управление на последнем шаге. Теперь, двигаясь от последнего шага к первому и используя основное функциональное уравнение, можно определить цепочку условных оптимальных управлений и условных оптимальных выигрышей на всех шагах. На этом предварительная оптимизация закончена, и можно перейти ко второй стадии оптимизации – нахождению безусловного оптимального управления $u = \{u_1, u_2, ..., u_m\}$.

Пусть начальное состояние S_0 системы полностью известно. Тогда $W_{\max} = W_1(S)$ и $u_1 = u_1(S)$. Зная S_0 и управление u_1, найдем состояние системы после первого шага $S_1' = \varphi_1(S_0, u_1)$ и после этого – оптимальное управление на втором шаге и т. д. В результате определим все шаговые оптимальные управления, оптимальное управление операцией в целом и конечное состояние системы. Теперь можно записать основную последовательность действий при решении задачи динамического программирования.

1. Выбрать способ описания процесса, т. е. параметры, характеризующие состояние системы, фазовое пространство и способ деления операции на «шаги».

2. Записать выигрыш на шаге с номером i в зависимости от состояния S системы в начале этого шага и управления U_i: $w_i = w_i\{S, U_i\}$.

3. Записать для i-го шага функцию $w_i(S, U_i)$ перехода системы в новое состояние под действием управления U_i.

4. Записать основное функциональное уравнение, выражающее $W_i(S)$ через $W_{i+1}(S)$:

$$W_i(S) = \max_{U_i}\{w_i(S, U_i) + W_{i+1}(\varphi_i(S, U_i))\}. \quad (34)$$

5. Найти для последнего шага функции $W_m(S) = \max_{U_m}\{w_m(S, U_m)\}$ и соответствующее условное оптимальное управление $u_m(S)$.

6. Используя основное функциональное уравнение, найти последовательно $W_{m-1}(S), W_{m-2}(S), \ldots, W_1(S)$ и соответствующие им условные оптимальные управления $u_{m-1}(S), u_{m-2}(S), \ldots, u_1(S)$.

7. Если начальное состояние S_0 задано, найти оптимальный выигрыш $W_{\max} = W_1(S_0)$ и далее по цепочке:

$$S_0 \to u_1(S_0) \to S_1 \to u_2(S_1) \to \ldots\ S_{m-1} \to u_m(S_{m-1}) \to S_m.$$

8. Если задана лишь область, в которой находится начальное состояние S_0, то сначала нужно найти такое начальное состояние, при котором выигрыш $W_1(S)$ достигает максимума и далее, по цепочке, безусловные оптимальные управления.

Пример.

Пусть на протяжении периода в N лет необходимо в начале каждого года заказывать какое-то оборудование для выполнения определенных работ. Мы располагаем начальной суммой денег x, которая делится на две части: y и $x-y$, используемых: первая – для закупки оборудования типа A; вторая – для закупки оборудования типа B. Затрата суммы y на оборудование типа A даст $g(y)$ часов полезной работы, а затрата суммы $(x-y)$ на оборудование типа B даст $h(x-y)$ часов полезной работы. В конце года можно продать оборудование A в утиль и выручить сумму $ay\ (0<a<1)$, и продать в утиль оборудование типа B, выручив сумму $b(x-y),(0<b<1)$. Этот процесс покупки и продажи повторяется ежегодно в течение N лет. Требуется найти такое распределение денег по годам, которое даст наибольшее общее число часов полезной работы оборудования за N-летний период [18].

Задачу будем решать в соответствии с приведенным алгоритмом.

1. В качестве параметров, характеризующих состояние системы, естественно выбрать x и y. Фазовое пространство будет двумерным. Этапом или шагом служит один год. В процессе управления величины x и y изменяются из за перераспределения средств в начале каждого года и уменьшения средств за год, сказывающееся в конце каждого года. Управление U_i на i-м шаге состоит в определении средств x_i, y_i. Управление операцией $U=(U_1,U_2,...,U_N)$ составляется из всех шаговых управлений. Требуется найти такое управление (оптимальное распределение денег по годам) $u=(u_1,u_2,...,u_N)$, при котором общее число часов W полезной работы оборудования будет максимальным, т. е. $W=\sum_{i=1}^{N} w_i = W_{\max}$.

2. Состояние системы перед i-м шагом характеризуется одним параметром K– количеством средств, сохранившихся после предыдущих $i-1$ шагов. Управление U_i на i-м шаге состоит в направлении средств x_i на оборудование типа A, значение $y_i=K-x_i$. Выигрыш (доход) на шаге с номером i будет $w_i(K,x_i)=h(x_i)+g(K-x_i)$.

3. Под влиянием этого управления система перейдет из состояния K в состояние $K'=\varphi(x_i)+\psi(K-x_i)$.

4. Основное функциональное уравнение имеет вид:

$$W_i(K)=\max_{0\le x_i\le K}\{h(x_i)+g(K-x_i)+W_{i+1}(\varphi(x_i)+\psi(K-x_i))\},$$

где максимум берется по всем неотрицательным значениям x_i, не превосходящим фактический запас средств K.

5. Условный оптимальный выигрыш на последнем шаге $W_N(K)=\max\limits_{0\le x_N\le K}\{h(x_N)+g(K-x_N)\}$. Ему соответствует условное оптимальное управление $u_N(K)$, при котором этот максимум достигается. Это по существу количество средств $x_N(K)$, вкладываемых на N-м шаге в оборудование типа A.

6. Зная функцию $W_N(K)$, находим в соответствии с п. 4 условные оптимальные выигрыши на двух последних, на трех последних и т. д. шагах:

$$W_{N-1}(K)=\max_{0\le x_{N-1}\le K}\{h(x_{N-1})+g(K-x_{N-1})+W_N(\varphi(x_{N-1})+\psi(K-x_{N-1}))\};$$

$$W_{N-2}(K)=\max_{0\le x_{N-2}\le K}\{h(x_{N-2})+g(K-x_{N-2})+W_{N-1}(\varphi(x_{N-2})+\psi(K-x_{N-2}))\};$$

. . . .

$$W_1(K)=\max_{0\le x_1\le K}\{h(x_1)+g(K-x_1)+W_2(\varphi(x_1)+\psi(K-x_1))\}$$

и соответствующие им условные оптимальные управления

$$x_{N-1}(K),\ x_{N-2}(K),\ .\ .\ .\ ,\ x_1(K).$$

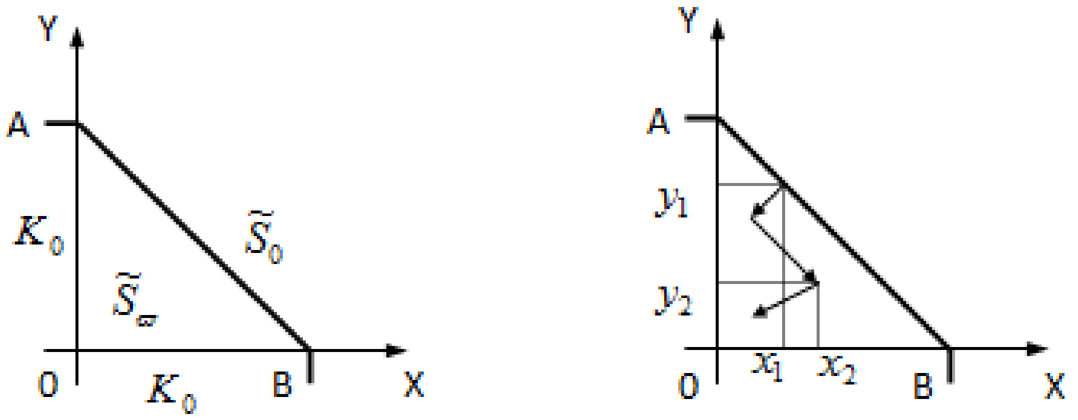
7. Начальное состояние K_0 (начальный запас средств) задано, поэтому максимальный доход (оптимальный выигрыш) будет $x_1=x_1(K_0)$. Состояние системы после первого шага $K_1=\varphi(x_1)+\psi(K_0-x_1)$. Оптимальное управление на втором шаге $x_2=x_2(K_1)$ и т. д. по цепочке

$$K_0\to x_1(K_0)\to K_1\to x_2(K_1)\to\ldots\to x_N(K_{N-1})\to K_N.$$

Совокупность средств, вложенных по годам в оборудование типа A: $x=(x_1,x_2,\ldots,x_N)$, представляет оптимальное управление процессом распределения ресурсов. Соответствующее количество средств, вложенных в оборудование типа B: $y=(K_0-x_1,K_1-x_2,\ldots,K_{N-1}-x_N)$.

Рассмотрим геометрическую интерпретацию полученного решения. Имеем двумерное фазовое пространство. По оси OX будем откладывать средства, вкладываемые в оборудование типа A, по оси OY – средства, вкладываемые в оборудование типа B. Их сумма не может превосходить количество начальных средств K_0, и поэтому фазовое пространство – часть плоскости XOY, заключенная внутри равнобедренного прямоугольного треугольника AOB с катетами K_0 (рис. 26, *а*).

Перед началом процесса сумма средств есть K_0, и поэтому область начальных состояний $\tilde{S}_0$ есть гипотенуза AB треугольника. На количество средств в конце периода N лет наложено единственное ограничение $0 \le X_N + Y_N < K_0$, и поэтому область конечных состояний $\tilde{S}_\omega$ системы – весь треугольник AOB, кроме гипотенузы. Изобразим фазовую траекторию состояния системы в фазовом пространстве (рис. 26, *б*).



а *б*

Рис. 26. Фазовая траектория состояния системы в фазовом пространстве:
а – начальное состояние; *б* – текущее состояние

Представим процесс изменения состояний системы состоящим из двух частей: распределения средств в начале каждого года и их использования в течение года. Тогда каждый шаг будет изображаться на фазовой траектории ломаной, состоящей из двух звеньев: на первом (когда идет перераспределение средств) точка S, характеризующая состояние системы, перемещается параллельно линии AB, а на втором, когда средства расходуются, – перемещается в сторону начала координат. Исключением служит первый шаг, когда сразу назначаются x_1, y_1 и начинается расходование ресурсов. Сумма абсциссы и ординаты последней точки траектории S_ω равно количеству средств, которое сохранится к концу периода при данном управлении.

Практические методы решения задачи динамического программирования рассматриваются в курсе математического программирования.

6.3. Задача линейного программирования

В задачах линейного программирования (ЛП) требуется найти экстремум линейной функции (целевой функции) при линейных ограничениях в виде равенств или неравенств. Общая постановка задачи и один из подходов к ее решению впервые были приведены в 1939 г. Л.В. Канторовичем. В виде задачи линейного программирования формулируются многочисленные задачи планирования в экономике, управления производственными и технологическими процессами, организации эффективной работы технических систем и др.

Наиболее распространенным примером задачи линейного программирования является задача планирования работы предприятия, выпускающего однородный продукт [18, 33]. Эта задача формулируется следующим образом: имеется n различных технологий и m видов ресурсов (сырье, оборудование, рабочая сила и т. д.) производства. Известны:

1) $\overline{C}^T = (c_1, c_2, ..., c_n)$ – вектор удельной производительности, где c_j – количество единиц продукта, которое можно получить в единицу времени при использовании технологии с номером j;

2) $A = \begin{Bmatrix} a_{11}, a_{12}, ..., a_{1n} \\ a_{21}, a_{22}, ... a_{2n} \\ .... \\ a_{m1}, a_{m2}, ..., a_{mn} \end{Bmatrix}$ – матрица удельного расхода ресурсов, где $a_{i,j}$ – расход ресурса с номером i при использовании технологии с номером j;

3) $\overline{B}^T = (b_1, b_2, ... b_m,)$ – вектор запаса ресурсов, где b_i – запас ресурса с номером i;

4) $\overline{X}^T = (x_1, x_2, ... x_n)$ – вектор, где x_j – время, в течение которого производство осуществляется по технологии с номером j.

Требуется найти такой план работы, при котором из имеющихся запасов было бы выпущено максимальное количество продукта.

Целевая функция сформулированной задачи имеет вид:

$$F = C^T \cdot X \to \max$$

при ограничениях:

$$A \cdot X \le B,$$
$$x_j \ge 0, \quad j = \overline{1, m}.$$

Каждой задаче ЛП соответствует двойственная задача, которая при тех же исходных данных формально может быть записана в следующем виде:

$$F = \overline{B}^T \cdot \overline{Y} \to min,$$

$$A^T \overline{Y} \ge C,$$

$$y_i \ge 0, \quad i = \overline{1, n},$$

где y_i – двойственные неизвестные, характеризующие скорость изменения целевой функции основной задачи при изменении правой части соответствующего ограничения. Двойственные неизвестные можно интерпретировать как удельные цены ресурсов. Тогда целевая функция двойственной задачи определит суммарные издержки производства, которые необходимо минимизировать.

Обе задачи обладают следующими свойствами.

1. В одной задаче ищут максимум целевой функции, а в другой – минимум.
2. Коэффициенты при переменных в целевой функции одной задачи являются свободными членами системы ограничений другой.
3. В задаче на *max* все ограничения неравенства имеют вид «≤», а в задаче на *min* – вид «≥».
4. Матрицы коэффициентов в системах ограничений обеих задач являются транспонированными относительно друг друга.
5. Число ограничений в одной задаче равно числу переменных в другой задаче.
6. Условия неотрицательности переменных имеются в обеих задачах.

Две задачи линейного программирования, имеющие перечисленные свойства, называются симметричными двойственными задачами.

Для перехода от одной задачи к другой необходимо:

1) если в исходной задаче ищется max целевой функции, то все неравенства ограничения приводят к виду «≤», а если min – к виду «≥»;

2) составить расширенную матрицу системы *A1*, в которую включить матрицу коэффициентов, столбец свободных членов и строку коэффициентов целевой функции;

3) найти матрицу $A1^T$, транспонированную относительно *A1*;

4) сформулировать двойственную задачу.

Например, пусть имеется задача линейного программирования.

Целевая функция:

$$50 \cdot x_1 + 40 \cdot x_2 \Rightarrow \max;$$

Ограничения на ресурсы:

$$\begin{array}{l} 20 \cdot x_1 + 10 \cdot x_2 \le 10 \cdot 40; \\ \left.\begin{array}{l} x_1 \le 15; \\ x_2 \le 30; \end{array}\right\} \end{array}$$

Условия неотрицательности:

$$x_1 \ge 0; \qquad x_2 \ge 0.$$

Составим двойственную задачу ЛП для рассмотренного примера.

1. Требования пункта 1 выполнены.
2. Составим расширенную матрицу системы:

$$A1 = \begin{pmatrix} 20 & 10 & 400 \\ 1 & 0 & 15 \\ 0 & 1 & 30 \\ 50 & 40 & \text{"F"} \end{pmatrix}.$$

3. Получим матрицу $A1^T$:

$$A1^T = \begin{pmatrix} 20 & 1 & 0 & 50 \\ 10 & 0 & 1 & 40 \\ 400 & 15 & 30 & \text{"Z"} \end{pmatrix}.$$

4. Сформулируем двойственную задачу.

Целевая функция

$$Z = 400y1 + 15y2 + 30y3 \Rightarrow \min$$

при ограничениях

$$\begin{aligned}
&20y1 + y2 \geq 50,\\
&10y1 + y3 \geq 40,\\
&y1 \geq 0, \quad y2 \geq 0, \quad y3 \geq 0.
\end{aligned}$$

Переменные y_k называются двойственными переменными. Оптимальные значения двойственных переменных характеризуют скорость изменения целевой функции прямой задачи при изменении правой части соответствующего ограничения. В примере двойственные оценки показывают, на какую величину возрастает прибыль, если объем соответствующего ресурса увеличится на 1. Пусть их значения $y_1 = 2{,}5$, $y_2 = 0$, $y_3 = 15$. Это означает, что дополнительная единица первого ресурса приносит 2,5 рубля прибыли, а увеличение на единицу третьего ресурса увеличивает прибыль на 15 рублей. Равенство нулю y_2 показывает, что его увеличение не влечет увеличения прибыли.

Величины двойственных оценок характеризуют предельный уровень цен, по которым имеет смысл покупать производственные ресурсы, чтобы выпуск продукции оставался выгодным. Например, двойственная оценка ограничения $x_1 \leq 15$ определяет верхнюю границу затрат на увеличение мощности на одну единицу. Если дополнительная единица производственной мощности будет обходиться дороже 2,5 рублей, то расширение мощностей невыгодно.

Связь между решениями двойственных задач устанавливается теоремами двойственности.

Задачи линейного программирования невозможно решить методами классического анализа, т. е. нельзя просто продифференцировать функцию F по аргументам x_1, $x_2, \ldots, x_n$ и решить полученную систему уравнений, так как функция F линейна, производные от нее по всем аргументам постоянны и нигде в нуль не обращаются. Экстремум функции F, если он существует, достигается всегда на границе области возможных значений $x_1, x_2, \ldots, x_n$, т. е. там, где начинают действовать ограничения.

Для решения задач рассмотренного вида применяют специальный математический аппарат линейного программирования, который позволяет эффективно обследовать границы области возможных решений и найти то решение, при котором линейная функция F обращается в максимум или в минимум.

Каноническая (основная) задача линейного программирования

В канонической задаче линейного программирования ограничения - неравенства преобразуют в ограничения – равенства. Эта задача формулируется следующим образом.

Имеется ряд переменных $x_1, x_2, \ldots, x_n$. Требуется найти такие неотрицательные значения этих переменных, которые удовлетворяют системе линейных уравнений $A \cdot X = B$ и, кроме того, обращают в максимум (минимум) линейную функцию

$$F = C^T \cdot X \to \max(\min). \qquad (35)$$

В этой задаче практически всегда число независимых уравнений m меньше числа неизвестных n. Тогда, если система совместна, у нее существует бесчисленное множество решений. При этом $n - m$ переменным мы можем придавать произвольные значения, а остальные m переменных выразятся через них (эти m переменных называют базисными). Если среди этих решений нет ни одного, для которого все $x_1, x_2, \ldots, x_n$ неотрицательны, то это значит, что задача линейного программирования не имеет допустимого решения. Если же такие решения существуют, то каждое из них допустимо. Как среди них найти такое, для которого линейная функция (35) обращается в максимум (минимум)?

Подробное изложение решения задачи линейного программирования имеется в литературе, посвященной курсу математического программирования.

Контрольные вопросы

1. Описать основные черты задачи динамического программирования.
2. Сформулировать задачу динамического программирования.

3. Дать содержательное определение задачи динамического программирования.

4. Привести пример задачи динамического программирования.

5. Дать геометрическую интерпретацию задачи динамического программирования.

6. Сформулировать принцип оптимальности задачи динамического программирования.

7. Дать определение уравнения состояния системы в задаче динамического программирования.

8. Дать определение фазовых координат и фазового пространства в задаче динамического программирования.

9. Объяснить свойство отсутствия последействия в задаче динамического программирования.

10. Объяснить содержание уравнения состояния в задаче динамического программирования.

11. Дать определение условного оптимального управления в задаче динамического программирования.

12. Дать определение условного оптимального выигрыша в задаче динамического программирования.

13. Сформулировать основную последовательность действий в задаче динамического программирования.

14. Сформулировать основное функциональное уравнение в задаче динамического программирования.

15. Определить функцию эффективности в задаче динамического программирования.

7. ПРИНЯТИЕ РЕШЕНИЯ В УСЛОВИЯХ НЕОПРЕДЕЛЕННОСТИ

Задачи принятия решений в условиях неопределенности возникают при необходимости действовать в не полностью определенной ситуации в самых разных областях человеческой деятельности: технике, экономике, биологии, экологии и т. д.

Формулируются они как поиск наиболее полезного решения на заранее заданном множестве допустимых решений. Основная сложность этой задачи в том, что последствия принимаемого решения зависят от неизвестной ситуации. Величину опасности (неприемлемости) последствий измеряют в условных единицах – потерях, которые может понести ЛПР. При этом основной исходной информацией для принятия решения является функция потерь – функция двух аргументов: выбираемой стратегии и состояния внешней среды (ситуации).

Основной проблемой при выборе решения в условиях неопределенности является преобразование функции потерь в другую функцию, зависящую только от одного аргумента – принимаемого решения. Правила такого преобразования называют критериями выбора решения в условиях неопределенности, оно полностью определяется ЛПР. В результате потери, сопутствующие каждой стратегии, как бы обобщаются относительно некоторого гипотетического состояния среды, и выбор наиболее полезной стратегии осуществляется на основании потерь, соответствующих этому гипотетическому состоянию среды.

Применимость различных критериев выбора зависит от типа неопределенности ситуации. В настоящее время наиболее изучены два типа неопределенностей: неопределенность состояния внешней среды (неопределенность природы) и неопределенность целенаправленного противодействия. Неопределенность состояния внешней среды (природы) имеет две версии реализации:

1) известно только множество Y состояний внешней среды, из которого оно может быть выбрано;

2) известно распределение вероятности на множестве возможных состояний внешней среды.

Первую из этих задач называют задачей принятия решения в условиях неопределенности, вторую – задачей принятия решений в условиях риска, а задачи принятия решения в условиях целенаправленного противодействия – играми.

7.1. Задача принятия решений в условиях неопределенности

Для преодоления возникающей проблемы необходимо сформулировать гипотезы о состоянии среды, позволяющие получить для каждой альтернативы числовую оценку полезности решения (потери) и по этой информации осуществить выбор. В простейшем случае множество X альтернатив и множество Y состояний среды – конечные множества, и целевую функцию можно задать таблицей (матрицей Q), строки которой соответствуют какой-либо альтернативе, а столбцы – состоянию среды. Тогда элемент $q_{i,j}$ матрицы Q – оценка эффективности альтернативы с номером i, соответствующая состоянию среды с номером j. Матрицу Q называют платежной матрицей, или матрицей потерь.

Для принятия решения необходимо для каждой стратегии ввести оценку, соответствующую какому-либо критерию выбора. Чаще всего используют один из следующих четырех критериев [18].

1. ***Критерий Лапласа*** основан на принципе равновозможности вариантов состояния среды и применяется, когда невозможно отдать предпочтение ни одному из них. Оценка альтернативы с номером i принимается равной среднему арифметическому элементов строки платежной матрицы с этим же номером:

$$L(i) = \frac{1}{m}\sum_{j=0}^{m-1} q_{i,j} \,. \tag{36}$$

Например, пусть дана матрица

$$\mathrm{U} = \begin{pmatrix} 0 & 2 & 5 & 3 & 7 & 1 \\ 6 & 3 & 1 & 1 & 9 & 1 \\ 0 & 5 & 5 & 1 & 4 & 0 \\ 7 & 4 & 8 & 8 & 5 & 4 \end{pmatrix}.$$

Критерий Лапласа для этой матрицы равен:

$$L = \begin{pmatrix} 3 \\ 3{,}5 \\ 2{,}5 \\ 6 \end{pmatrix}; \qquad \max(L) = 6.$$

Любые две альтернативы сравнимы между собой по критерию Лапласа: лучшая альтернатива имеет большую (меньшую) оценку, а оптимальная – максимальную (минимальную) оценку. Недостатки этого критерия связаны с эффектом сглаживания отдельных оценок при вычислении критерия.

2. ***Критерий Вальда*** основан на принципе минимакса (максимина). Минимакс – значение функции $f(x, y)$, которое она принимает, когда сначала выбирается максимум по y из множества Y, а затем – минимум по x из множества X.

Оценка альтернативы с номером i в соответствии с критерием Вальда выполняется по одной из формул:

$$W(i) = \begin{cases} \min\limits_i \max\limits_j q_{i,j}, \\ \max\limits_i \min\limits_j q_{i,j}. \end{cases} \tag{37}$$

Выбор одной из этих формул зависит от цели. Первая формула позволяет выбрать стратегию с наименьшими максимальными потерями, а вторая – с наибольшими минимальными потерями. Например, для матрицы U,

$$U = \begin{pmatrix} 4 & 8 & 6 & 2 & 7 & 4 \\ 0 & 5 & 5 & 7 & 6 & 7 \\ 1 & 3 & 2 & 1 & 7 & 5 \\ 2 & 0 & 7 & 2 & 5 & 1 \end{pmatrix},$$

найдем:

$$\max U_i = \begin{pmatrix} 8 \\ 7 \\ 7 \\ 7 \end{pmatrix}; \quad \min\max U = 7; \quad \min U_i = \begin{pmatrix} 2 \\ 0 \\ 1 \\ 0 \end{pmatrix}; \quad \max\min U = 2.$$

Содержательный смысл критерия Вальда состоит в том, что при его использовании выбранная стратегия обеспечивает или минимум максимально возможных потерь (min max), или максимум минимально возможных доходов (max min).

3. ***Критерий Гурвица*** основан на задании для любой альтернативы с номером строки i субъективной вероятности α наибольших (A) и $(1 - \alpha)$ наименьших (a) потерь, вычислении по этим данным значения

$$H(i) = \alpha A_i + (1 - \alpha)a_i \tag{38}$$

и выборе альтернативы с экстремальным значением H.

Например, при $\alpha = 0{,}5$

$$U = \begin{pmatrix} 7 & 7 & 9 & 5 & 2 & 7 \\ 3 & 6 & 0 & 2 & 5 & 7 \\ 4 & 6 & 4 & 6 & 5 & 6 \\ 5 & 1 & 4 & 4 & 6 & 1 \end{pmatrix}; \qquad H = \begin{pmatrix} 5{,}5 \\ 3{,}5 \\ 5 \\ 3{,}5 \end{pmatrix}; \qquad \max H = 5{,}5.$$

Критерий Гурвица учитывает только наилучший и наихудший исходы. Это обстоятельство относят к недостаткам данного критерия, как и субъективность определения значения α.

4. ***Критерий Сэвиджа*** основан на вычислении максимальной утерянной выгоды при различных вариантах действий, и выборе того варианта действий, который минимизирует максимальную утерянную выгоду. Матрица U преобразуется в матрицу R утерянной выгоды по правилу:

$$r_{i,j} = \max_j U_{i,j} - U_{i,j}. \tag{39}$$

Например,

$$U = \begin{pmatrix} 7 & 7 & 9 & 5 & 2 & 7 \\ 3 & 6 & 0 & 2 & 5 & 7 \\ 4 & 6 & 4 & 6 & 5 & 6 \\ 5 & 1 & 4 & 4 & 6 & 1 \end{pmatrix}; \; r = \begin{pmatrix} 2 & 2 & 0 & 4 & 7 & 2 \\ 4 & 1 & 7 & 5 & 2 & 0 \\ 2 & 0 & 2 & 0 & 1 & 0 \\ 1 & 5 & 2 & 2 & 0 & 5 \end{pmatrix}; \; \max r = \begin{pmatrix} 7 \\ 7 \\ 2 \\ 5 \end{pmatrix}; \; \min\max\ r = 2.$$

Оптимальной по критерию Сэвиджа считается альтернатива, минимизирующая утерянную выгоду:

$$S(i) = \min_i(\max_j q_{i,j} - q_{i,j}). \tag{40}$$

Рассмотрим пример.

Пусть требуется из четырех вариантов (A_1, A_2, A_3, A_4) выбрать проект технологии. Последствия, связанные с выбором, зависят от некоторого множества неопределенных факторов, которые определяют состояние среды. Пусть определено четыре варианта (B_1, B_2, B_3, B_4) состояний среды. Эффективность выбора какого-либо варианта при различных состояниях среды определяется платежной матрицей (табл. 6).

Таблица 6

Платежная матрица

	B_1	B_2	B_3	B_4
A_1	7	5	1	10
A_2	5	2	8	4
A_3	1	3	4	12
A_4	8	5	1	10

Какой вариант является оптимальным? Рассмотрим решение этой задачи при помощи рассмотренных критериев.

Таблица 7

Оптимальные варианты решения

	$L(i)$	$W(i)$	$H_\alpha(i)$, $\alpha = 1/2$	$S(i)$
X_1	23 / 4	1	11 / 2	**7**
X_2	19 /4	**2**	10 / 2	8
X_3	20 / 4	1	**13 / 2**	**7**
X_4	**24 / 4**	1	11 / 2	**7**

Оптимальные по каждому критерию варианты выделены в табл. 7 жирным шрифтом. Для разных критериев получаем разные оптимальные решения. Это обусловлено тем, что критерии основываются на различных гипотезах о состоянии среды. Использование той или иной гипотезы снимает неопределенность, но гипотеза – это лишь предположение, которое не обязательно совпадает с истинным состоянием среды.

Поэтому, если позволяют обстоятельства, рекомендуется рассмотреть весь диапазон возможных состояний среды и составить представление об эффективности решений в этом диапазоне. Решение, оптимальное для заданного диапазона состояний среды, называется локально оптимальным. Совокупность локально оптимальных решений для всего диапазона состояний среды и дает представление об эффективности решения и ее зависимости от состояния среды. Совершенно очевидно, что неопределенность при этом сохраняется. Поэтому неразумно предъявлять к точности решения слишком высокие требования, и лучше вместо строго оптимального решения выделить область приемлемых решений, которые оказываются несущественно хуже оптимального, и в пределах этой области произвести окончательный выбор.

7.2. Принятие решения в условиях риска

При втором типе неопределенности считают известными функции распределения вероятности возможных состояний среды. Вероятность состояния среды, представляется вектором $p = (p_1, p_2, \dots p_m)$, где p_i – вероятность наступления состояния среды с номером i. Такие задачи называют игрой с природой. В этом случае платежная матрица имеет вид (табл. 8).

Таблица 8

Платежная матрица с известными функциями распределения вероятности возможных состояний среды

	Состояние среды				
	1	…	j	…	m
Вероятность состояния	p_1	…	p_j	…	p_m
альтернатива					
1	$a_{1,1}$	…	$a_{1,j}$	…	$a_{1,m}$
…					
i	$a_{i,1}$	…	$a_{i,j}$	…	$a_{i,m}$
…					
n	$a_{n,1}$	…	$a_{n,j}$	…	$a_{n,m}$

Если выбрать альтернативу с номером i, то исход будет представлен случайной величиной

$$\xi_i = \begin{Bmatrix} a_{i,1},..., a_{i,m} \\ p_1,..., p_m \end{Bmatrix}, \tag{41}$$

и сравнение двух альтернатив эквивалентно сравнению соответствующих случайных величин. Числовые характеристики этой случайной величины характеризуют соответствующую альтернативу независимо от состояния среды. В этих условиях потери, связанные с ошибочным выбором альтернативы, характеризуются математическим ожиданием этой случайной величины, а рассеивание значений потерь относительно математического ожидания – средним квадратическим отклонением этой случайной величины.

Пусть, например, в данной «человеко-машинной системе» возможно использование двух технологий $a1$, $a2$, среда может принимать одно из трех состояний: $b1$, $b2$, $b3$. Известно, что состояние среды $b1$ реализуется с вероятностью 0,375, состояние $b2$ – с вероятностью 0,5, состояние $b3$ – с вероятностью 0,125. Вероятность реализации каждой технологии одинаковая и равна 0,5. Какую технологию необходимо выбрать, чтобы «стоимость» аварийной ситуации была минимальной? При возникновении аварии «стоимость» S аварийной ситуации оценивается платежной матрицей, приведенной в табл. 9.

Таблица 9

Платежная матрица

Вероятность состояния среды	0,375	0,5	0,125	Математическое ожидание ущерба $M\xi_{ai}$
	b_1	b_2	b_3	
a_1	16	18	14	17,5
a_2	8	15	40	15,5

В качестве критерия сравнения альтернатив примем математическое ожидание «ущерба» от аварийной ситуации, значение которой приведено в последней колонке платежной матрицы. Очевидно, что оптимальной альтернативой (ущерб наименьший) является альтернатива a_2.

Однако критерий математического ожидания ущерба предполагает, что имеется возможность многократной реализации рассматриваемой ситуации. В действительности эта возможность отсутствует, и решение принимается при однократной реализации. Поэтому возникает риск получить не лучшее решение. Величину риска каждого решения можно оценить по отклонениям реального ущерба от математического ожидания. Их значения приведены табл. 10.

Таблица 10

Отклонения реального ущерба от математического ожидания

	b_1	b_2	b_3	$M\xi_{ai}$
	0,375	0,5	0,125	
a_1	1,5	–0,5	3,5	17,5
a_2	8	1	–24	15,5

Приведенные данные свидетельствуют, что альтернативы a_1 и a_2, имея близкие значения ожидаемых выигрышей, по-разному ведут себя относительно возможных отклонений от них: для a_1 эти отклонения невелики, а для a_2 – значительны. Это значит, что, выбирая альтернативу a_2, мы рискуем больше, чем при выборе альтернативы a_1. Поэтому критерий выбора, основанный на величине математического ожидания (ожидаемого выигрыша), необходимо дополнить характеристикой риска – величиной возможных отклонений случайной величины от ее математического ожидания. В качестве характеристики риска принимают величину среднеквадратического отклонения σ.

Таким образом, в условиях риска выбор альтернативы характеризуется двумя показателями: ожидаемым выигрышем и показателем риска – его среднеквадратическим отклонением σ. Тогда фактически получается задача двухкритериальной оптимизации с частными критериями M и σ и ее решение основывается или на определении Парето-оптимального множества решений, или на построении обобщенного критерия.

7.3. Принятие решения в условиях конфликта

Задачи принятия решений в условиях конфликта изучаются в теории игр. Участвующие в конфликте стороны заинтересованы в том, чтобы скрыть от противника свои намерения. Основой теории игр является формализация понятий конфликта, принятия решения в нем и полезности (оптимальности) этого решения.

Содержательно конфликтом считают всякое явление, относительно которого можно говорить о его участниках, об их действиях, об исходах явления, к которым эти действия приводят, о сторонах, заинтересованных в этих исходах, и о сущности этой заинтересованности [11, 12].

Участников конфликта называют игроками, их возможные действия – стратегиями, исходы конфликта – ситуациями, заинтересованные стороны – участники конфликта, сущность заинтересованности – возможные предпочтения одной ситуации перед другой для каждого игрока. Конкретизация этих понятий приводит к разнообразным частным классам игр.

Принятием решения в теории игр считается выбор игроком своей стратегии. Вопрос о формализации понятия полезности (оптимальности) принимаемого решения является сложным. Единого представления о нем нет, и поэтому рассматриваются разные версии, но в основе каждой из них лежит интуитивное представление о полезности, как о чем-то устойчивом и справедливом. Формализация этих представлений определяется системой аксиом.

Ситуации, удовлетворяющие в игре требованиям полезности (оптимальности), называются решениями игры. Ограничимся рассмотрением игр двух участников с прямо противоположными (антагонистическими) интересами. Формально эта противоположность выражается в том, что при переходе от одной ситуации к другой увеличение выигрыша одного игрока влечет численно равное уменьшение выигрыша другого. Сумма выигрышей игроков в любой ситуации постоянна (ее можно считать равной нулю). Существует много явлений, для которых антагонистические игры являются хорошей моделью. Например, некоторые военные операции, спортивные игры, деловые решения в условиях конкуренции и т. д.

Антагонистическая игра задается множествами A и B стратегий игроков и вещественной функцией H, определенной на множестве $A \times B$

и являющейся функцией выигрыша первого игрока (функция выигрыша второго игрока равна $-H$). Процесс игры состоит в выборе игроками некоторых своих стратегий $a \in A$, $b \in B$, после чего первый игрок получает от второго сумму $H(a, b)$. Разумное (осторожное) поведение игроков в антагонистической игре осуществляется на основании принципа максимина [18].

Если

$$\max_{a \in A} \min_{b \in B} H(a,b) = \min_{b \in B} \max_{a \in A} H(a,b),$$

то у каждого игрока существует оптимальная осторожная стратегия. Эта стратегия называется «чистой» стратегией. Однако уже в самых простых ситуациях принцип максимина может не выполняться.

Если множества A и B конечны, то антагонистическая игра называется матричной игрой. Для нее всегда существуют оптимальные смешанные стратегии у обоих игроков. Если первый игрок имеет m стратегий, а второй – n стратегий, то матричная игра может быть задана $m \times n$ матрицей $A = \{a_{i,j}\}$. Матричные игры моделируют широкий круг конфликтных антагонистических ситуаций с двумя участниками и конечным множеством возможных действий у каждого из них. Иногда под одним из игроков понимается природа как совокупность всех обстоятельств, неизвестных принимающему решение другому игроку. Такие игры называют играми с природой. Они возникают при необходимости учета природных и других неконтролируемых факторов, не находящихся в распоряжении какого-либо лица. При этом природе приписывается роль сознательного противника – антагониста.

Для решения матричной игры необходимо найти оптимальные стратегии игроков и цену игры. Практическое правило для решения матричной игры следующее. К платежной матрице A добавим один столбец справа и одну строку снизу. В ячейки столбца запишем минимальное значение платежа соответствующей строки, а в ячейки строки – максимальное значение соответствующего столбца (табл. 11).

Первый игрок будет выбирать стратегию, которая гарантирует ему максимальный выигрыш независимо от выбора стратегии второго игрока. Поэтому он выберет стратегию, при которой он будет иметь максимальный выигрыш из возможных минимальных выигрышей (в столбце min это элемент, соответствующий стратегии R первого игрока), т. е. гарантированный выигрыш первого игрока:

$$v_1 = \max_i \min_j a_{i,j}\,. \tag{42}$$

Этот выигрыш v_1 называют нижней ценой игры, а соответствующую стратегию – максиминной стратегией первого игрока.

Таблица 11

Решение матричной игры

	Стратегии	Игрок 2		min
		S	T	
Игрок 1	P	–2	2	–2
	Q	–1	3	–1
	R	1	2	1
max		1	3	$v_1 = v_2 = 1$

Второй игрок, стараясь минимизировать проигрыш, выберет стратегию, гарантирующую наименьший из максимально возможных проигрышей, т. е. стратегию

$$v_2 = \min_j \max_i a_{i,j}\,, \tag{43}$$

которую называют минимаксной стратегией. Число v_2 называют верхней ценой игры. Если $v_1 = v_2 = v$, то число v называют ценой игры. Игры, для которых $v_1 = v_2 = v$, называют играми с седловой точкой. Седловых точек может быть несколько. Однако цена игры во всех седловых точках одна и та же. Если один из игроков применяет стратегию, соответствующую одной из седловых точек, то при этом ему обеспечен выигрыш не меньше цены игры. Если и второй игрок применяет стратегию, соответствующую какой-либо седловой точке, то и ему обеспечен проигрыш не более цены игры. Одностороннее отклонение от седловой точки невыгодно ни одному из игроков. Такое отклонение может в лучшем случае оставить выигрыш (проигрыш) неизменным, а в худшем случае – уменьшить (увеличить) его.

Таким образом, в играх с седловой точкой оптимальные стратегии обладают своеобразной устойчивостью: если один из игроков применяет свою оптимальную стратегию, то для другого всегда невыгодно отклоняться от своей оптимальной стратегии.

К классу игр, имеющих седловую точку, относятся игры, в которых каждый игрок знает результаты всех предыдущих ходов. Такие игры называются играми с полной информацией. Примерами игр с полной информацией служат шашки, шахматы, «крестики-нолики» и т. д. В теории игр доказано, что каждая игра с полной информацией имеет седловую точку, т. е. имеется пара оптимальных стратегий противников, дающая устойчивый выигрыш, равный чистой цене игры.

В монографии [10] показано, что если в игре участвует много игроков, то каждый игрок с номером i стремится сделать такой выбор, чтобы максимизировать свою целевую функцию f_i (свой выигрыш). Но значения целевой функции зависят и от выбора других игроков, т. е.

$$f_i = f_i(x_1, x_2, \ldots, x_n). \tag{44}$$

Если в точке $\hat{x} = (x_1, \hat{x}_2, \ldots, \hat{x}_n)$ выполняется условие

$$\max_{x_i} f_i(x_1, x_2, \ldots, x_n) = f_i(\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n), \tag{45}$$

то говорят, что в этой точке имеется ситуация равновесия и эту точку называют точкой равновесия в игре. Эти точки также называют точками устойчивости, поскольку, если игрок с номером i выберет стратегию, отличную от $\hat{x} = (\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n)$, то при условии, что остальные игроки сохранят свой выбор, проиграет, прежде всего, игрок с номером i, так как

$$f_i(\hat{x}_1, \hat{x}_2, \ldots x_i, \ldots, \hat{x}_n) \le f_i(\hat{x}_1, \hat{x}_2, \ldots \hat{x}_i, \hat{x}_n).$$

На этом основании формулируется принцип устойчивости Нэша: выбор рациональной стратегии должен производиться среди стратегий, соответствующих точкам равновесия. Это значит, что выигрыш каждого игрока в ситуации равновесия не меньше выигрыша, соответствующего другим возможным ситуациям, т. е. отступать от стратегии равновесия невыгодно ни одному из игроков.

Не всякая матричная игра приводит к решениям с единственным оптимальным выбором для обоих игроков. Это значит, что ни один из игроков не может выбрать чистую оптимальную стратегию. В этом случае каждый из игроков достигает уменьшения риска использованием смешанных стратегий. В теории игр смешанными стратегиями называют комбинированные стратегии, в которых используются чистые стратегии, чередующиеся случайным образом при заданном соотношении вероятностей (частот) использования. С этих позиций всякая чистая стратегия есть частный случай смешанной, в которой все стратегии, кроме одной, применяются с нулевыми частотами, а одна с частотой 1. Введение смешанных стратегий расширяет множество допустимых действий игроков, чтобы добиться существования решения игры.

Смешанной стратегией называют n-мерный вероятностный вектор **p**, координаты которого неотрицательны и сумма их равна 1. Значения координат вектора интерпретируются как вероятности выбора стратегии. Смешанные стратегии реализуются случайным выбором чистых стратегий с вероятностями, заданными в векторе **p**. Например, пусть для первого игрока смешанная стратегия задана вектором $(x_1, x_2, \ldots, x_n)$, а для второго – вектором $(y_1, y_2, \ldots, y_m)$. Так как какая-то из стратегий обязательно реализуется, то

$$\sum_i x_i = 1; \ \ \sum_i y_i = 1.$$

Это значит, что первый игрок будет выбирать стратегию с номером i с вероятностью x_i, а второй игрок – стратегию с номером j с вероятностью y_j. Полагая, что игроки выбирают стратегии независимо друг от друга, вероятность события (i, j) очевидно равна

$$p_{i,j} = x_i\, y_j.$$

В результате применения смешанных стратегий выигрыш (проигрыш) становится случайной величиной:

$$\xi = \begin{bmatrix} a_{i,j} \\ p_{i,j} \end{bmatrix}_{(i=\overline{1,m};\ \overline{j=1,m})}. \tag{46}$$

Математическое ожидание этой случайной величины и есть выигрыш (проигрыш) первого (второго) игрока в смешанных стратегиях (x, y):

$$M\xi = \sum_{i,j} p_{i,j} \cdot a_{i,j} = F_A(x, y). \tag{47}$$

Следовательно, исходной платежной матрице A соответствует новая платежная матрица $F_A(x, y)$ игры в смешанных стратегиях. В теории игр доказано, что игра в смешанных стратегиях всегда имеет седловую точку, т. е. такая игра всегда имеет цену, и каждый игрок имеет оптимальную смешанную стратегию.

Если игра размерности $m \times n$ не имеет в чистых стратегиях седловой точки, то при больших значениях m и n решить ее достаточно трудно. В некоторых случаях такую задачу удается упростить исключением дублирующих или заведомо невыгодных стратегий. Например, пусть имеем игру с матрицей (табл. 12).

Таблица 12

Матрица игры

$A \setminus B$	B_1	B_2	B_3	B_4
A_1	1	3	4	2
A_2	0	2	3	2
A_3	1	3	4	2
A_4	5	4	2	1

Очевидно, что стратегии A_1 и A_3 совпадают, и поэтому любую из них можно вычеркнуть. Сравнивая стратегии A_1 и A_2, убеждаемся, что первая из них доминирует вторую и, следовательно, стратегия A_2 заведомо невыгодная, и должна быть вычеркнута. После вычеркивания стратегий A_2 и A_3 получаем значения, приведенные в табл. 13.

Таблица 13

Матрица игры

$A \setminus B$	B_1	B_2	B_3	B_4
A_1	1	3	4	2
A_4	5	4	2	1

Для противника стратегии B_2 и B_3 – заведомо невыгодные, по сравнению со стратегией B_4, и тоже должны быть вычеркнуты. В итоге получаем матрицу игры (табл. 14).

Таблица 14

Матрица игры

$A \setminus B$	B_1	B_4
A_1	1	2
A_4	5	1

Наиболее простые конечные игры имеют размерность 2×2 и $2 \times m$. Рассмотрим игру 2×2 без седловой точки с матрицей (табл. 15).

Таблица 15

Матрица игры

$A \setminus B$	B_1	B_2
A_1	a_{11}	a_{12}
A_2	a_{21}	a_{22}

Требуется найти оптимальную смешанную стратегию игрока A:

$$S_A^* = \begin{pmatrix} A_1 & A_2 \\ p_1 & p_2 \end{pmatrix}.$$

Эта стратегия имеет свойство: при любых полезных стратегиях противника выигрыш будет равен цене игры υ. В игре 2×2 обе стратегии противника – полезные, так как иначе игра имела бы решение в чистых стратегиях (седловую точку). Следовательно, если игрок A придерживается своей оптимальной стратегии, то игрок B может пользоваться любой из своих чистых стратегий и при этом цена игры не изменится. Поэтому можем записать систему уравнений:

$$\begin{cases} a_{11} \cdot p_1 + a_{21} \cdot p_2 = \upsilon, \\ a_{12} \cdot p_1 + a_{22} \cdot p_2 = \upsilon, \\ p_1 + p_2 = 1, \end{cases} \qquad (48)$$

из решения которой найдем

$$p_1 = \frac{a_{22} - a_{21}}{a_{11} + a_{22} - a_{12} - a_{21}},$$
$$\nu = \frac{a_{11} \cdot a_{22} - a_{21} \cdot a_{12}}{a_{11} + a_{22} - a_{12} - a_{21}},$$
$$p_2 = 1 - p_1.$$

Для определения оптимальной смешанной стратегии игрока *B*, при известной цене игры, достаточно двух уравнений

$$\begin{aligned} & a_{11} \cdot q_1 + a_{12} \cdot q_2 = \nu, \\ & q_1 + q_2 = 1, \end{aligned} \tag{49}$$

из решения которых получим

$$q_1 = \frac{\nu - a_{12}}{a_{11} - a_{12}},$$
$$q_2 = 1 - q_1.$$

Решение игры 2 × 2 имеет простую геометрическую интерпретацию [18] (рис. 27).

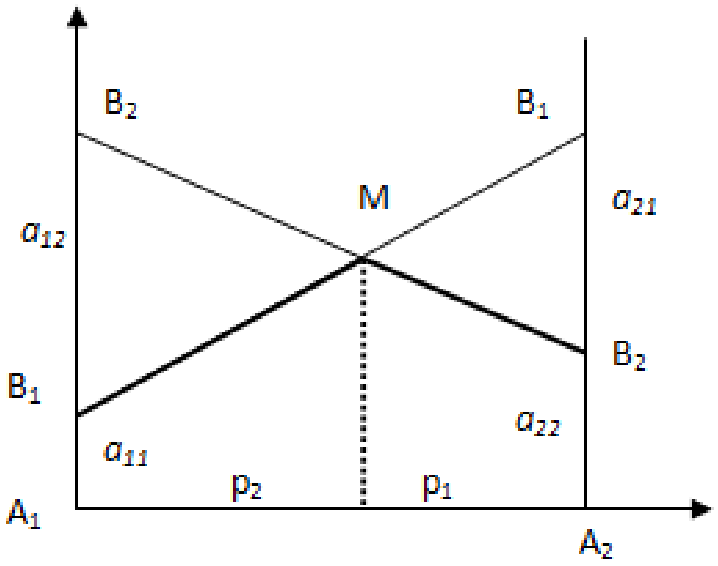


Рис. 27. Геометрическая интерпретация игры 2 × 2

Пусть имеется игра 2 × 2 с матрицей (табл. 16).

Таблица 16

Матрица игры

$A \backslash B$	B_1	B_2
A_1	a_{11}	a_{12}
A_2	a_{21}	a_{22}

На оси абсцисс возьмем отрезок длиной 1. Граничная точка слева будет соответствовать чистой стратегии A_1, а справа – чистой стратегии A_2. Через граничные точки проведем линии, перпендикулярные к оси абсцисс, и на них будем отмечать выигрыши при стратегии A_1 и A_2 соответственно.

Пусть противник применяет стратегию B_1. Тогда на оси A_1 и A_2 можем отметить точки соответствующих выигрышей a_{11} и a_{21}, через которые проведем линию B_1B_1. Аналогичным образом построим линию B_2B_2. Точки этих линий определяют выигрыш первого игрока в смешанных стратегиях при условии, что второй игрок применяет чистую стратегию B_1 или B_2 соответственно. Мы ищем смешанную стратегию для первого игрока, при которой его возможный минимальный выигрыш будет наибольшим, независимо от того, какую стратегию применит второй игрок. Для этого построим нижнюю границу выигрыша при стратегиях B_1 и B_2. На рис. 27 эта граница показана жирной линией. Эта линия выражает оптимальный выигрыш игрока A при любых его смешанных стратегиях. Найдем координаты точки M пересечения линий B_1B_1 и B_2B_2, в которой минимальный выигрыш игрока A достигает максимума. Ордината этой точки равна цене игры, а абсцисса равна p_2 – вероятности (частоте) применения стратегии A_2 в оптимальной смешанной стратегии S^*_A.

Таким образом, любая игра 2×2 может быть решена элементарными приемами. Совершенно аналогично может быть решена любая игра $2 \times n$. Матрица такой игры состоит из двух строк и n столбцов. Например, пусть $n = 4$. Как и в случае игры 2×2, строим нижнюю границу выигрыша (на рис. 28 жирная линия) и находим на ней точку N с максимальной ординатой. Эта точка дает решение игры $S_A^* = \begin{pmatrix} A_1 & A_2 \\ p_1 & p_2 \end{pmatrix}$.

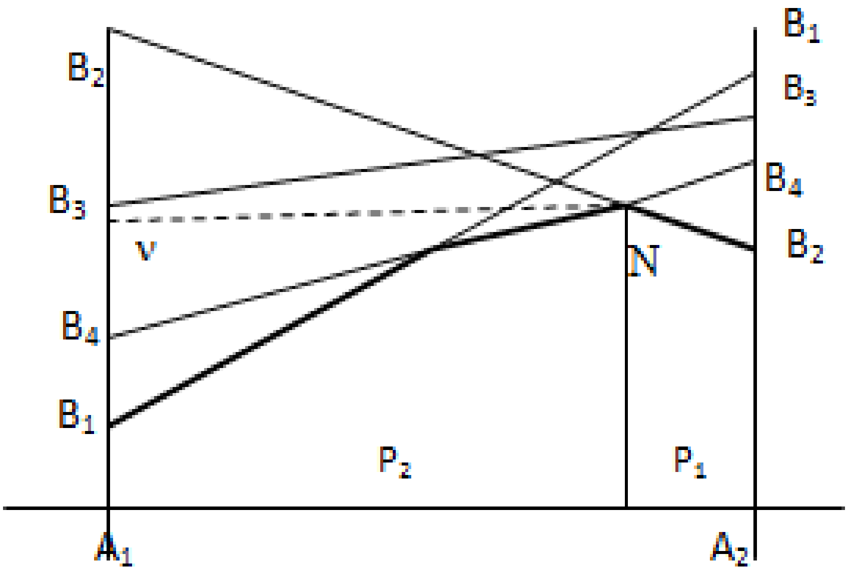


Рис. 28. Геометрическая интерпретация игры 2 × 4

Ордината точки N равна цене игры, а абсцисса равна вероятности (частоте) p_2 стратегии A_2. Анализируя рис. 28, замечаем, что оптимальная стратегия противника получается применением двух полезных стратегий B_2 и B_4, пересекающихся в точке N. Стратегия B_3 заведомо невыгодна, а стратегия B_1 невыгодна при оптимальной стратегии S_A^*.

В общем случае решение игры $m \times n$ довольно трудная задача. Доказано, что у любой конечной игры $m \times n$ имеется решение, число полезных стратегий в котором не превосходит наименьшего из чисел m или n. Трудности решения связаны главным образом с большим объемом вычислений. Поэтому при решении игр $m \times n$ применяют аналитические методы линейного программирования.

Рассмотрим сведение игры двух лиц с нулевой суммой к задаче линейного программирования.

Предположим, что цена игры положительна, т. е. $\nu > 0$. Если в платежной матрице имеются отрицательные элементы, то необходимо к платежной матрице прибавить положительное число d, такое, чтобы все ее элементы стали положительными. При этом оптимальные стратегии игроков не изменяются, и цена игры увеличится на d. Применяя оптимальную смешанную стратегию, первый игрок стремится при любой стратегии второго игрока получить выигрыш не менее цены игры ν, которая была бы максимальной. Такая ситуация может быть представлена моделью линейного программирования:

$$\nu \to \max, \ \Sigma p_i \cdot a_{i\,j} \geq \nu, \qquad i = 1, 2, \ldots, m, \qquad j = 1, 2, \ldots, n, \qquad p_i \geq 0. \qquad (50)$$

Обозначая

$$x_i = p_i / \nu, \qquad (51)$$

задачу переписывают в виде

$$x_1 + x_2 + \ldots + x_m \to \min; \qquad (52)$$

$$\Sigma a_{i\,j} \cdot x_i \geq 1, \qquad i = 1, 2, \ldots, m, \qquad j = 1, 2, \ldots, n, \qquad x_i \geq 0.$$

После решения этой задачи находят цену игры:

$$\nu = 1 / (x_1 + x_2 + \ldots + x_m)$$

и оптимальную смешанную стратегию для первого игрока.

Если исходная матрица увеличивалась на d, то для получения цены исходной игры значение ν нужно уменьшить на d.

Обозначим

$$y_j = q_j / \nu. \qquad (53)$$

Так как второй игрок старается минимизировать свой проигрыш, его поведение опишем двойственной задачей:

$$y_1 + y_2 + \ldots + y_n \to \max, \qquad (54)$$

$$\Sigma a_{j\,i} \cdot y_j \geq 1, \qquad i = 1, 2, \ldots, m, \qquad j = 1, 2, \ldots n, \qquad x_i \geq 0.$$

Таким образом, принятие решения в конфликтной ситуации может быть сведено к двойственной задаче линейного программирования.

Контрольные вопросы

1. Описать основные типы неопределенностей.
2. Дать определение понятию «риск».
3. Сформулировать задачу принятия решений в условиях неопределенности.

4. Описать процедуру принятия решения на основании:
   1) критерия Лапласа;
   2) критерия Вальда;
   3) критерия Гурвица;
   4) критерия Сэвиджа.
5. Почему оптимальные решения, соответствующие различным критериям не совпадают?
6. Описать, как можно уменьшить потери от решения, принимаемого в условиях неопределенности.
7. Сформулировать задачу принятия решений в условиях риска.
8. Что такое риск?
9. Что может служить показателем риска?
10. Описать процедуру принятия решения в условиях риска, когда альтернативы сравнимы по Парето.
11. Описать процедуру принятия решения в условиях риска, когда альтернативы несравнимы по Парето.
12. Сформулировать задачу принятия решения в конфликтных условиях.
13. В чем отличие задачи принятия решения в условиях конфликта от задачи принятия решения в условиях неопределенности и риска.
14. Описать матричную игру двух лиц.
15. Дать определение оптимальной стратегии и цены игры.
16. Дать определение игры с седловой точкой.
17. Дать определение оптимальной смешанной стратегии.
18. Привести геометрическую интерпретацию игры 2×2.
19. Привести геометрическую интерпретацию игры $2 \times n$.

8. СТАТИСТИЧЕСКИЕ ПРОЦЕДУРЫ ПРИНЯТИЯ РЕШЕНИЙ

Статистические процедуры принятия решений основаны на теории выбора наиболее полезного недетерминированного поведения в условиях неопределенности [20, 21]. Современная общая концепция статистических процедур принятия решений принадлежит А. Вальду [17, 26].

Полагая, что каждый исход эксперимента имеет стоимость, а за ошибочное решение необходимо платить штраф, А. Вальд предложил выбирать то решение, которое минимизирует риск, связанный с последствиями принимаемых решений. Так как величина риска зависит и от правила выбора решения и от функции распределения исходов, то риск оценивают математическим ожиданием всех потерь.

Выбор недетерминированного решения в условиях неопределенности осуществляется в два этапа [11]. На первом этапе интуитивно выбирается множество измеримых признаков $x = (x_1, x_2, \ldots, x_n)$ исхода эксперимента, а на втором осуществляется отображение множества наборов признаков x на множество Λ решений λ, принимаемых относительно исходов эксперимента, и выбор решения, минимизирующего риск.

В настоящее время наиболее разработаны три процедуры принятия недетерминированных решений в условиях неопределенности: метод последовательного анализа Вальда, Байесовские процедуры принятия решений и процедуры статистической проверки гипотез о функции распределения случайной величины и значениях числовых характеристик функции распределения.

8.1. Решения, основанные на проверке статистических гипотез

Статистическая проверка гипотез – это проверка предположений о законе и параметрах распределения генеральной совокупности по конечной выборке из этой совокупности. Правила или процедуры проверки статистических гипотез называются критериями согласия.

Решения, основанные на проверке статистических гипотез, предназначены для ответа на вопрос о согласованности эксперимента и некоторой теории. Предполагается существование выборки, состоящей из n наблюденных значений некоторой случайной величины, и необходимо решить, имеет ли распределение вероятностей этой величины некоторые заданные свойства. Пусть, например, это распределение $F(x)$ известно. Необходимо проверить статистическую гипотезу о том, что выборка получена из совокупности с этим распределением. Рассмотрим типичную простейшую процедуру проверки статистической гипотезы (критерий согласия) [27].

Предположим, что проверяемая гипотеза верна и, следовательно, при больших значениях n функция распределения выборки $F^*(x)$ должна мало отличаться от заданной функции $F(x)$. Меру различия между $F^*(x)$ и $F(x)$ обозначим D. Эта мера, как некоторая функция от выборочных значений случайной величины, сама является случайной величиной и имеет некоторое выборочное распределение. В математической статистике доказано, что при некоторых способах выбора D и больших значениях n эта функция распределения не зависит от $F(x)$. Допустим, что это выборочное распределение нам известно, и имеется возможность вычислить вероятность $P(D > D_0)$ того, что D превышает некоторое заданное значение D_0. Обозначим вероятность $P(D > D_0) = \alpha$ и выберем значение α настолько малым, чтобы можно было считать, что при единичном опыте событие $D > D_0$ практически невозможно. Вероятность α называют уровнем значимости критерия. Если для некоторой выборки из n значений окажется, что $D > D_0$ (в этом случае говорят, что отклонение значимо), то это означает, что произошло событие, имеющее вероятность α. Однако, по нашей гипотезе в единичном опыте это маловероятно и, следовательно, наша гипотеза противоречит опыту. Если же оказалось, что $D \leq D_0$, то считается, что это обусловлено случайными причинами и выдвинутая гипотеза не противоречит экспериментальным данным, по крайней мере, пока это подтверждается дальнейшими опытами. Геометрическая интерпретация изложенной методики проверки статистических гипотез приведена на рис. 29. На этом рисунке отмечены точки $-D_0$ и D_0, для которых выполнены условия $p(-D < -D_0) = 0{,}5\alpha$ и $p(D > D_0) = 0{,}5\alpha$, где α – некоторое малое число (на рисунке принято $\alpha = 0{,}05$). Статистическая гипотеза принимается, если значение $D \in [-D_0, D_0]$, и отклоняется, если $D \notin [-D_0, D_0]$. Область $[-D_0, D_0]$ называют областью допустимых значений (областью принятия гипотезы), а область $[-\infty, -D_0] \cup [D_0, \infty]$ – областью отклонения гипотезы.

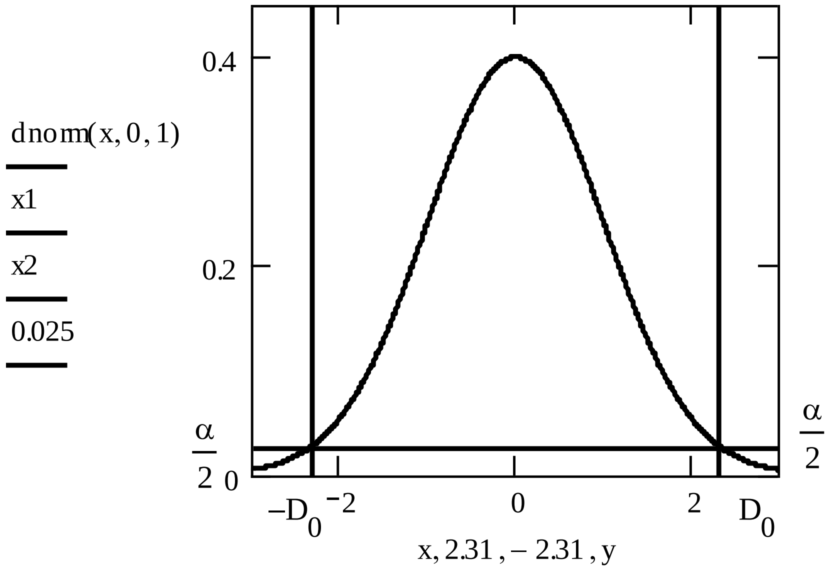


Рис. 29. Геометрическая интерпретация проверки статистических гипотез

В рассмотренном примере критерий проверки гипотезы двусторонний, так как значимы отклонения в обе стороны. Если отклонения значимы только в одну сторону, то строят односторонние критерии.

При проверке статистических гипотез возможны ошибки двух родов. Ошибка первого рода состоит в том, что гипотеза отклоняется, когда она верна. Вероятность этой ошибки равна α, она определяет заданную допустимую вероятность ошибки первого рода. Ошибка второго рода состоит в том, что гипотеза принимается, когда она не верна, вероятность ошибки второго рода обозначают β. Эту ситуацию можно представить в виде табл. 17.

Таблица 17

Ошибки при проверке гипотез

Гипотеза H_0	Верна	Неверна
Отвергается	Ошибка первого рода	Правильное решение
Принимается	Правильное решение	Ошибка второго рода

При статистической проверке гипотез желательно применять критерии, имеющие малые вероятности ошибок первого и второго рода. Если объем выборки ограничен, то вероятности ошибок первого и второго рода связаны – уменьшение одной из них влечет, как правило, увеличение другой. Поэтому на практике задают границу для вероятности ошибки первого рода и рассматривают критерии, минимизирующие вероятность ошибки второго рода, у которых вероятность ошибки первого рода не больше заданной границы.

Методы проверки статистических гипотез позволяют принимать решения о соответствии выборочной функции распределения некоторому теоретическому распределению, о соответствии числовых выборочных характеристик числовым характеристикам теоретических распределений и др.

8.2. Последовательные решения

Это статистический метод принятия решений, основанный на последовательном (пошаговом) принятии решений. Характерной чертой последовательного анализа является возможность выбора числа опытов непосредственно в процессе принятия решений в зависимости от значений поступающих данных. Теорию последовательных решений в начале 20 столетия разработал А. Вальд [17, 26].

Рассмотрим случайную величину ξ с дискретным распределением вероятностей $P(\xi, \theta)$. Задача состоит в том, чтобы по результатам наблюдений принять решение об истинном значении неизвестного параметра θ. В классических статистических методах число наблюдений для принятия решения о значении θ фиксируется заранее, а в последовательном анализе – является случайной величиной, не зависящей от будущего. В последовательном анализе наблюдения выполняются до тех пор, пока не станет возможным принять решение с заданной точностью, не зависящей от значений неизвестных параметров. По сравнению с классическими методами применение последовательного анализа для решения этой задачи дает экономию в числе наблюдений до 50 %.

Пусть, например, необходимо решить, приобретать или нет некоторую партию изделий. Предположим, что имеется возможность распознавать бракованное изделие, и что эта процедура стоит «*c*». Из партии, содержащей N изделий, извлекается случайная выборка объемом n изделий. Если в этой выборке бракованных изделий не более m, то приобретается вся партия. Стоимость этой процедуры $c \times n$.

Однако, если наблюдения производить последовательно, то число бракованных изделий может достигнуть значения m раньше, чем будут осмотрены все n изделий. Наблюдения прекращаются, как только число бракованных изделий достигнет m, или число годных окажется равным $n - m$. Одно из этих событий может наступить раньше, чем будут осмотрены все n изделий, и стоимость осмотра будет меньше $c \times n$.

Процедура последовательных решений состоит из двух шагов. Первый шаг – это выбор правила остановки, когда необходимо решить, следует ли принять решение, не делая наблюдений, или есть смысл провести хотя бы одно наблюдение. Если проведение наблюдений целесообразно, то для каждого набора наблюденных значений нужно сделать вывод: следует ли остановиться и принять решение, или же нужно продолжать наблюдения.

Второй шаг – выбор решающего правила. Если наблюдения проводить нет смысла, то принимается решение на основании априорных данных. Если же проводится хотя бы одно наблюдение, необходимо для каждого возможного набора наблюдений указать решение, минимизирующее риск – математическое ожидание всех убытков [17, 26].

Применение метода последовательных решений рассмотрим на примере последовательного критерия отношения правдоподобия для различения двух простых гипотез относительно значения параметра θ распределения случайной величины ξ с дискретным распределением вероятностей $P(\xi, \theta)$ [11, 12].

Предположим, что неизвестный параметр θ может принимать значения θ_0 и θ_1. Пусть H_0 – гипотеза о том, что $\theta = \theta_0$, а H_1 – о том, что $\theta = \theta_1$. Вектор последовательных независимых наблюдений случайной величины:

$$\xi = \left(\xi_1, \xi_2, \,..,\xi_{m-1} \;\; \xi_m\right). \tag{55}$$

Вероятность получения такой выборки для гипотез H_0 и H_1 равна соответственно

$$\begin{aligned} P_m^0 &= \mathrm{p}(\xi_1, \Theta_0) \cdot p(\,\xi_2, \Theta_0) \cdot \ldots \cdot p(\xi_m, \Theta_0); \\ P_m^1 &= p(\xi_1, \Theta_1) \cdot p(\xi_2, \Theta_1) \cdot \ldots \cdot p(\xi_m, \Theta_1). \end{aligned} \tag{56}$$

Отношение правдоподобия для выполненных наблюдений:

$$R_m = \frac{P_m^1}{P_m^0} = \frac{p(\xi_1, \Theta_1) \cdot p(\xi_2, \Theta_1) \cdot \ldots \cdot p(\xi_m, \Theta_1)}{p(\xi_1, \Theta_0) \cdot p(\xi_2, \Theta_0) \cdot \ldots \cdot p(\xi_m, \Theta_0)}. \qquad (57)$$

Если $R_m \to 0$, то следует принять гипотезу H_0, если $R_m \to \infty$, то следует принять гипотезу H_1. Если же $R_m \in [A, B]$, где $A > 0$, $B < \infty$, то числитель и знаменатель – одного порядка, и ситуация не позволяет сделать выбор.

Поэтому для определения последовательного критерия отношения правдоподобия для различения гипотез H_0 и H_1 выбирают постоянные A и B, удовлетворяющие условию:

$$0 < B < 1 < A < \infty.$$

На каждом шаге последовательной выборки вычисляется отношение R_m, сравнивается с числами A и B и выбирается одно из трех решений:

1) если $R_m \leq B$, то наблюдения прекращают, и принимается гипотеза H_0;
2) если $R_m \geq A$, то наблюдения прекращают, и принимается гипотеза H_1;
3) если $B < R_m < A$, то проводят следующее $m + 1$ наблюдение.

Постоянные A и B называют граничными точками последовательного критерия отношения правдоподобия. На практике удобнее вычислять $\log R_m$, так как в этом случае

$$\log R_m = \sum_{i=1}^{m} \log \frac{p(\xi_i, \Theta_1)}{p(\xi_i, \Theta_0)} = \sum_{i=1}^{m} z_i = S_m. \qquad (58)$$

и выбор решения осуществляют не по значению R_m, а по значению S_m и значениям $\log A$ и $\log B$. Значения чисел A и B выбирают не произвольно, а исходя из вероятностей ошибок первого и второго рола.

Пусть вероятность ошибки первого рода равна α, а вероятность ошибки второго рода равна β. Вектор (α, β) называют силой последовательного критерия. Точные значения $A = A(\alpha, \beta)$ и $B = (\alpha, \beta)$ определить затруднительно, и поэтому принимают приближенно:

$$A \approx \frac{1-\beta}{\alpha}, \ B = \frac{\beta}{1-\alpha}. \qquad (59)$$

Подробный и глубокий анализ задач принятия оптимальных статистических решений изложен в работе [17].

8.3. Байесовские процедуры принятия решений

Байесовский метод [11, 12, 28] – метод принятия решений о ненаблюдаемых характеристиках, основанный на знании априорного распределения вероятностей этих характеристик и на условном распределении результатов эксперимента при заданных значениях ненаблюдаемых характеристик.

В общем случае байесовский метод состоит в выборе решения с минимальным средним риском. Пусть $\{x\}$ – наблюдаемые реализации рассматриваемого случайного события или величины; di ($i = 1, 2, \ldots, n$) – возможные решения (гипотезы) о значениях искомых характеристик этого события. При использовании байесовского метода должны быть заданы априорные сведения: безусловные вероятности гипотез $p(di)$ и условные вероятности реализаций $p(x/di)$ при каждой из гипотез di. От этих априорных сведений легко перейти к условным вероятностям гипотез при наблюдаемой реализации $p(di/x)$, называемым апостериорными вероятностями. Наиболее вероятное решение в байесовском методе определяется максимальной апостериорной вероятностью. Если априорные вероятности гипотез неизвестны, то их можно оценить эмпирически по заданной выборке реализаций.

Байесовское решающее правило – статистическое решающее правило, обеспечивающее минимум среднего риска решения.

Байесовская теория принятия решений основана на предположении, что задача сформулирована в терминах теории вероятностей и известны все представляющие интерес вероятностные величины, условия задачи описаны с помощью небольшого числа параметров, значения которых хотя и не известны точно, но известны их распределения вероятностей [11, 12, 17].

Предположим, что ЛПР не представляется возможным предсказать какое из двух происшествий можно ожидать в ближайшем будущем, и он вынужден считать появление того или иного происшествия случайным событием. Используя терминологию теории решений, говорят, что вследствие этого среда пришла в одно из двух состояний, соответствующих возможным происшествиям. Обозначим ω_1 – состояние среды при появлении первого происшествия и ω_2 – при появлении второго. Пусть $P(\omega_1)$ вероятность того, что произойдет первое происшествие, и $P(\omega_2)$ вероятность, что произойдет второе происшествие. Эти вероятности отражают знание того, с какой степенью уверенности можно предсказать появление каждого из происшествий до их действительного появления. Эти вероятности называют априорными вероятностями.

Байесовский метод принятия решения в условиях неопределенности вытекает из теоремы Байеса, основывающейся на двух фактах из теории вероятностей. Если A и B – два любых события, то (рис. 30):

1) условная вероятность:

$$P(A/B) = P(A \text{ и } B) \,/\, P(B); \tag{60}$$

2) вероятность события A:

$$P(A) = P(A \text{ и } B) + P(A \text{ и (не } B)). \tag{61}$$

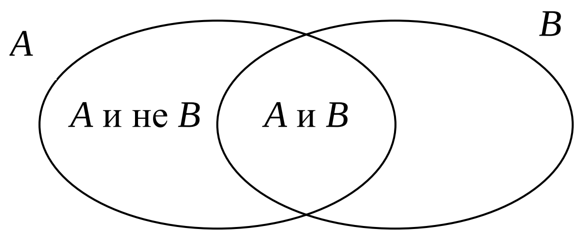


Рис. 30. Теорема Байеса

Пусть события B_1, B_2, …, B_n образуют полную группу несовместных событий. Вероятности событий B_j (j =1, 2, …, n) известны. Пусть событие A появляется вместе с одним из событий B_j. В этом случае события B_j выступают как единственно возможные и взаимно исключающие условия появления события A, или как гипотезы, в предположении которых (и только их) может произойти событие A. Тогда на основании формулы полной вероятности

$$P(A) = \sum_{j=1}^{n} P(B_j) \cdot P(A \mid B_j). \tag{62}$$

На основании факта 1: $P(A \text{ и } B_j) = P(A)P(B_j|A) = P(B_j)P(A|B_j)$ и, следовательно, при $P(A) \neq 0$,

$$P(B_j \mid A) = \frac{P(B_j)P(A \mid B_j)}{P(A)} = \frac{P(B_j)P(A \mid B_j)}{\sum_{j=1}^{n} P(B_j) \cdot P(A \mid B_j)}. \tag{63}$$

Формула (63) и выражает содержание теоремы Байеса. Условная вероятность $P(B_j/A)$ осуществления гипотезы B_j, вычисленная по формуле Байеса в предположении, что событие A произошло, называется *апостериорной* вероятностью.

В большинстве случаев при выборе решения привлекают дополнительную информацию – некоторые признаки x, которые рассматривают как случайные величины, распределение которых зависит от состояния среды. Пусть $p(x \mid \omega_j)$ – условная плотность распределения величины x в состоянии ω_j, т. е. функция плотности распределения случайной величины x при условии, что состояние среды есть ω_j.

Допустим, что известны и априорные вероятности $P(\omega_j)$, и условные плотности $p(x \mid \omega_j)$, которые в количественной мере оценивают правдоподобие состояния ω_j при данном x. Предположим, что имеется возможность измерить значение величины x, и спрашивается, в какой мере это измерение влияет на наше представление об истинном состоянии среды. Ответ на этот вопрос дает теорема Байеса.

Теорема Байеса показывает, как знание величины признака x позволяет из априорной вероятности $P(\omega_j)$ получить апостериорную вероятность $P(\omega_j|x)$. При наличии апостериорных вероятностей байесовское решающее правило состоит в выборе состояния среды ω_1, если $P(\omega_1|x) > P(\omega_2|x)$, и в выборе ω_2 в противоположном случае. Вероятность ошибки принимаемого решения, как и выше, равна меньшей из величин $P(\omega_1|x)$ или $P(\omega_2|x)$.

Если состояние среды есть ω_j, ($j = 1, 2, \ldots, s$) и реализуются некоторые мероприятия a_i ($i = 1, 2, \ldots, m$), то из-за ошибки в определении состояния среды возникают потери $\lambda(a_i/\omega_j)$. Так как $P(\omega_j/x)$ – вероятность того, что действительное состояние среды ω_j, то ожидаемые потери (условный риск), связанные с выполнением мероприятий a_i, равны [11, 12]:

$$R(a_i \mid x) = \sum_{j=1}^{s} \lambda(a_i \mid \omega_j) \cdot P(\omega_j \mid x). \qquad (64)$$

Формула (64) позволяет сформулировать Байесовское решающее правило: для минимизации общего риска требуется вычислить условный риск для всех значений a_i и выбрать то действие, при котором условный риск минимален. Минимальный общий риск называют байесовским риском, соответствующим наилучшему возможному образу действий.

Рассмотрим пример [16].

Пусть имеется две урны. В первой урне 70 зеленых и 30 белых шаров, а во второй – 70 белых и тридцать зеленых. Из урны, выбранной наугад, вытаскивают по одному с возвращением 12 шаров и регистрируют цвет. Оказалось, что было вытащено 8 зеленых и 4 белых шара. Спрашивается, какова вероятность того, что шары вытаскивали из первой урны?

До проведения опыта (вытаскивание шаров) априорная вероятность выбора первой или второй урны

$$P(U1) = P(U2) = 0{,}5.$$

После проведения опыта появилась новые данные о цвете 12 вытащенных шаров, которые позволяют рассчитать по формуле Байеса условную апостериорную вероятность выбора урны в зависимости от результатов опыта.

Обозначим A – событие: вытащено 8 зеленых и 4 белых шара. Вероятность этого события, при условии, что шары вытаскиваются из первой или второй урны, соответственно есть

$$P(A|U1) = (0{,}7)^8 \cdot (0{,}3)^4;$$

$$P(A|U2) = (0{,}7)^4 \cdot (0{,}3)^8.$$

Теперь по формуле Байеса найдем условную апостериорную вероятность того, что шары были вытащены из первой урны:

$$P(U1|\, A) = \frac{P(A|\, U1)\cdot P(U1)}{P(A|\, U1)\cdot P(U1) + P(A|\, U2)\cdot P(U2)} =$$

$$= \frac{0{,}7^8 \cdot 0{,}3^4 \cdot 0{,}5}{0{,}7^8 \cdot 0{,}3^4 \cdot 0{,}5 + 0{,}7^4 \cdot 0{,}3^8 \cdot 0{,}5} = 0{,}964.$$

В данном случае уточнение данных о состоянии среды за счет проведения опыта оказалось довольно значительным. Однако остается неясным, стоит ли новая информация, полученная по результатам эксперимента, затрат на ее получение?

Для ответа на этот вопрос сравним максимальный ожидаемый выигрыш до эксперимента с максимальным ожидаемым выигрышем после эксперимента [16].

Будем считать, что стоимость решения задана матрицей

$$A = (a_{ij}), \quad i = \overline{1,n}; \quad j = \overline{1,m}. \tag{65}$$

каждая строка i которой соответствует какому то варианту решения, а каждый столбец j – некоторому состоянию среды. Вероятности состояний среды заданы вектором

$$\mathbf{p} = (p_1, p_2, \ldots, p_m). \tag{66}$$

Максимальный ожидаемый выигрыш до эксперимента очевидно равен

$$M = \max_i \sum_{j=1}^{m} \mathbf{p}_j \cdot a_{ij}. \tag{67}$$

Предположим, что для уточнения состояния среды имеется возможность провести эксперимент, стоимость которого S. Пусть после эксперимента стало точно известно, что среда находится в состоянии j. После эксперимента максимальный выигрыш в состоянии j:

$$m_j = \max_i a_{ij}. \tag{68}$$

Максимальный ожидаемый выигрыш после эксперимента равен

$$\mu - S = \sum_{j=1}^{m} m_j \cdot p_j - S = \sum_{j=1}^{m} \max_i a_{ij} \cdot p_j - S \tag{69}$$

и, следовательно, эксперимент выгоден лишь тогда, когда $\mu - S > M$. Из этого неравенства следует, что эксперимент целесообразно проводить, если его стоимость не превосходит значения $\mu - M$.

Рассмотрим пример.

$$A = \begin{pmatrix} 1 & 3 & 2 & 5 \\ 2 & 4 & 7 & 4 \\ 8 & 2 & 1 & 1 \end{pmatrix}; \quad p = \begin{pmatrix} 0{,}2 \\ 0{,}3 \\ 0{,}4 \\ 0{,}1 \end{pmatrix}; \quad M = 4{,}8; \quad m = \begin{pmatrix} 8 \\ 4 \\ 7 \\ 5 \end{pmatrix}; \quad \mu = 6{,}1; \quad S = 1{,}3.$$

Из этого примера следует, что стоимость эксперимента не должна превышать 1,3 условных единиц.

Таким образом, байесовские процедуры принятия решений позволяют объединить имеющиеся априорные сведения о неизвестном параметре с данными эксперимента, что реализуется пересчетом априорного распределения в апостериорное распределение параметра. При этом риск принятия решения оказывается числом, а не функцией от неизвестного параметра.

Основная трудность применения байесовского метода состоит в необходимости знания априорного распределения параметра, которая преодолевается применением теории субъективной вероятности [17].

Контрольные вопросы

1. Что такое статистическая гипотеза?
2. Дать определение критериев согласия.
3. Для чего применяются процедуры статистической проверки гипотез?
4. Дать определение уровня значимости критерия согласия при проверке статистических гипотез.
5. Привести геометрическую интерпретацию процедуры проверки статистических гипотез.
6. Дать определение ошибок первого и второго рода.
7. Изложить процедуру проверки гипотезы о числовом значении математического ожидания при известной дисперсии.
8. Изложить процедуру проверки гипотезы о равенстве математических ожиданий двух нормальных генеральных совокупностей.
9. Дать определение условной вероятности.
10. Дать определение априорной вероятности.
11. Дать определение апостериорной вероятности.
12. Сформулировать теорему о полной вероятности.
13. Сформулировать теорему Байеса.
14. Сформулировать байесовское решающее правило основанное на знании апостериорных вероятностей.
15. Сформулировать байесовское решающее правило для минимизации общего риска.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Перегудов Ф.И., Тарасенко Ф.П. Основы системного анализа: учеб. – 2-е изд., доп. – Томск: Изд-во НТЛ, 1997. – 396 с.

2. Бугакова Т.Ю., Вовк И.Г. Системный анализ, моделирование и принятие решений: учеб. справочник. – Новосибирск: СГГА, 2010.

3. Бугакова Т.Ю. Оценка устойчивости состояний объектов по геодезическим данным методом фазового пространства: автореф. дис. канд. техн. наук. – Новосибирск: СГГА, 2005.

4. Бугакова Т.Ю. Оценка устойчивости состояний объектов по геодезическим данным методом фазового пространства: дис. канд. техн. наук. – Новосибирск: СГГА, 2005. – 163 с.

5. Белов П.Г. Системный анализ и моделирование опасных процессов в техносфере: учеб. пособие для студ. высш. учеб. заведений. – М.: Издательский центр «Академия», 2003. – 512 с.

6. Денисов А.А., Колесников Д.Н. Теория больших систем управления: учеб. пособие для вузов. – Л.: Энергоиздат, Ленингр. отд-ние, 1982. – 288 с.

7. Вовк И.Г. Введение в математическое моделирование: учеб. пособие. – Новосибирск: СГГА, 1997. – 45 с.

8. Бугакова Т.Ю., Вовк И.Г. Декомпозиция и агрегирование систем при реализации системно-целевого подхода // ГЕО-Сибирь-2010. Т. 1: Геодезия, геоинформатика, картография, маркшейдерия. Ч. 2: сб. матер. VI Междунар. науч. конгр. «ГЕО-Сибирь-2010», 19–29 апр. 2010 г., Новосибирск. – Новосибирск: СГГА, 2010. – С. 21–24.

9. Вовк И. Г. Математическое моделирование экономических систем (системный анализ и принятие решений): модульный курс. – Новосибирск: НГИ, 2008. – 60 с.

10. Моисеев Н.Н. Математические задачи системного анализа. – М.: Наука, 1981. – 488 с.

11. Энциклопедия кибернетики. – Киев: Укр. сов. энцикл., 1975. – Т. 2. – 622 с.

12. Энциклопедия кибернетики. – Киев: Укр. сов. энцикл., 1975. – Т. 1. – 607 с.

13. Бугакова Т.Ю., Вовк И.Г. Методические вопросы оценки техногенного геодинамического риска в прикладной информатике. Актуальные вопросы модернизации высшего образования // Сб. матер. регион. научн.-метод. конф., 11–12 февр. 2010 г., Новосибирск. – Новосибирск: СГГА, 2010. – 313 с.

14. Бусленко Н.П. и др. Лекции по теории сложных систем / Н.П. Бусленко, В.В. Калашников, Н.Н. Коваленко. – М.: Советское радио, 1973. – 438 с.

15. Чуев Ю.В. и др. Прогнозирование количественных характеристик процессов / Ю.В. Чуев, Ю.Б. Михайлов, В.И. Кузьмин. – М.: Сов. радио, 1975. – 400 с.

16. Райфа Г. Анализ решений (введение в проблему выбора в условиях неопределенности). – М.: Наука, 1977. – 408 с.

17. Де Гроот М. Оптимальные статистические решения. – М.: Мир, 1974. – 401 с.

18. Вентцель Е.С. Исследование операций. – М.: Сов. радио, 1972. – 522 с.

19. Подиновский В.В., Ногин В.Д. Парето-оптимальные решения многокритериальных задач. – М.: Наука, 1982. – 231 с.

20. Математическая энциклопедия / Гл. ред. И.М. Виноградов. – М.: Советская энциклопедия, 1984. – Т. 4.

21. Математическая энциклопедия / Гл. ред. И.М. Виноградов. – М.: Советская энциклопедия, 1982. – Т. 3.

22. Хемминг Р.В. Численные методы для научных работников и инженеров. – М.: Наука, 1968. – 400 с.

23. Демидович Б.П., Марон И.А. Основы вычислительной математики. – М.: Наука, 1963. – 659 с.

24. Кини Р., Райфа Х. Принятие решений при многих критериях: предпочтения и замещения. – М.: Радио и связь, 1981.

25. Глушков В.М. Основы безбумажной информатики. – Изд. 2-е, испр. – М.: Наука, 1987. – 552 с.

26. Вальд А. Последовательный анализ. – М.: Фитматгаз, 1960.

27. Крамер Г. Математические методы статистики. – М.: Мир, 1975. – 648 с.

28. Вовк И.Г. Байесовский подход при оценке состояния искусственных и естественных систем // ГЕО-Сибирь-2007. Т. 6. Исследования по общетехническим и гуманитарным проблемам: сб. матер. III Междунар. научн. конгр. «ГЕО-Сибирь-2007», 25–27 апреля 2007 г., Новосибирск. – Новосибирск: СГГА, 2007. – С. 133–135.

29. Моделирование систем: учеб. для вузов / С.И. Дворецкий, Ю.Л. Муромцев, В.А. Погонин, А.Г. Схиртладзе. – М.: Академия, 2009. – 320 с.

30. Советов Б.Я., Яковлев С.А. Моделирование систем: учеб. для вузов. – 4-е изд. – М.: Высш. шк., 2005. – 343 с.

31. Советов Б.Я., Яковлев С.А. Моделирование систем: практикум. – М.: Высш. шк., 2005. – 296 с.

32. Вовк И.Г. Методические основы системно-целевого подхода к изучению систем. Актуальные вопросы модернизации высшего образования: сб. матер. региональной научн.-метод. конф., 11–12 февраля 2010 г., Новосибирск. – Новосибирск: СГГА, 2010. – С. 140–142.

33. Исследование операций в экономике: учеб. пособие для вузов / Н.Ш. Кремер, Б.А. Путко, И.М. Тришин, М.Н. Фридман; под ред. проф. Н.Ш. Кремера. – М.: ЮНИТИ, 2006. – 407 с.

Учебное издание

Вовк Игорь Георгиевич

Бугакова Татьяна Юрьевна

ОСНОВЫ СИСТЕМНО-ЦЕЛЕВОГО ПОДХОДА И ПРИНЯТИЕ РЕШЕНИЙ

Редактор *Е.Н. Ученова*

Компьютерная верстка *Н.Ю. Леоновой*

Изд. лиц. ЛР № 020461 от 04.03.1997.
Подписано в печать. 17.05.2011. Формат 60 × 84 1/16
Печать цифровая.
Усл. печ. л. 8,84. Тираж 219 экз.
Заказ . Цена договорная.

Гигиеническое заключение
№ 54.НК.05.953.П.000147.12.02. от 10.12.2002.

Редакционно-издательский отдел СГГА
630108, Новосибирск, ул. Плахотного, 10.

Отпечатано в картопечатной лаборатории СГГА
630108, Новосибирск, ул. Плахотного, 8.