Федеральное агентство по образованию
Тверской государственный технический университет

Ю.Н. Матвеев

Основы теории систем и системного анализа

Учебное пособие

Часть 1

Издание первое

Допущено Учебно-методическим объединением по образованию в области прикладной информатики в качестве учебно-методического пособия для студентов высших учебных заведений, обучающихся по специальности «Прикладная информатика (по областям)» и другим специальностям.

Тверь 2007

УДК 519.8(075.08)
ББК 22.18.я7

Матвеев, Ю.Н. Основы теории систем и системного анализа: учебное пособие / Ю.Н. Матвеев. Ч. 1. 1-е изд. Тверь: ТГТУ, 2007. 100 с.

В первой части изложены основные положения системологии и системного анализа. Рассмотрен ряд методов и алгоритмов решения задач классификации. Приведены сведения о задачах исследования операций. Подробно рассмотрен вопрос о решении задач линейного программирования. Теоретический материал проиллюстрирован достаточным количеством примеров.

Во второй части предполагается изложение материалов, посвященных решению задач целочисленного и динамического программирования, задач исследования систем массового обслуживания и теории игр.

Учебное пособие предназначено для студентов вузов, слушателей курсов повышения квалификации, а также для широкого круга читателей, которые самостоятельно хотят приобрести знания в области системологии и системного анализа.

Рецензенты: заведующий кафедрой прикладной информатики Уральского социально-экономического института Академии труда и социальных отношений г. Челябинска, доктор технических наук, доцент В.И. Забейворота; профессор кафедры информационных технологий Тверского государственного университета, доктор технических наук В.А. Масюков.

Юрий Николаевич Матвеев

Основы теории систем и системного анализа

Учебное пособие
Часть 1
Издание первое

Редактор В.А. Крылова
Корректор И.В. Шункова
Технический редактор Г.В. Комарова

Подписано в печать 19.06.07		
Формат 60x84/16		Бумага писчая
Физ. печ. л. 6,25	Усл. печ. л. 5,81	Уч.-изд. л.5,43
Тираж 150 экз.	Заказ № 49	С – 48

Редакционно-издательский центр
Тверского государственного технического университета
170026, г. Тверь, наб. А. Никитина, 22

ISBN 5-7995-0374-0

Введение

В различных областях науки и техники широко используется понятие «система». Под системой понимается нечто целое, составленное из частей или множества элементов, образующее некое единство.

Однако с 30-40 гг. XX в. системы являются предметом исследования логико-математической дисциплины, которая занимается изоморфизмом системных понятий, законов и моделей в различных областях целенаправленной деятельности людей. Эта дисциплина имеет различные названия: системология, общая теория систем, системотехника и т.п.

Системология активно развивается, так как она занимается актуальными проблемами разработки математических методов решения междисциплинарных задач.

В системологии очень приблизительно можно выделить два аспекта [1].

В границах первого аспекта системологию можно рассматривать как расширение и обобщение теории управления.

В рамках второго аспекта системология реализует кибернетический или структуралистский подход в исследовании систем.

При таком подходе особое внимание обращается на структурные характеристики, а не на характеристики функций системы. Предлагается формализация семантики и логики общесистемных понятий, позволяющая определить их иерархическую классификацию. На основе этой классификации параллельно с ней разрабатывается соответствующая классификация системных задач и методов решения.

В данном учебном пособии изложены материалы, позволяющие студенту получить представление о системологии, классификационном анализе и об одном из разделов исследования операций, а именно о решении задач линейного программирования.

Учебное пособие разработано в соответствии со стандартом ГОС ЕН.Ф.05 и учебным планом для специализации 080801.65 Прикладная информатика (в экономике).

Основой учебного пособия является курс лекций по дисциплине «Теория систем и системный анализ», читаемый автором в течение ряда лет в Тверском государственном техническом университете.

1. Основные понятия и исходные модели системного подхода

Под системным подходом понимается практическая процедура диалектико-материалистического метода мышления и организации деятельности специалистов при решении сложных проблем, в том числе связанных с управлением.

Главное назначение системного подхода состоит в сокращении затрат времени и числа ошибок при поиске вариантов решения сложных проблем. Эти ошибки имеют место при ориентации только на «здравый смысл». В то же время выводы, полученные на основе системного подхода, не должны противоречить «здравому смыслу».

Именно поэтому для успешного применения проектировщиками системного подхода необходимо наличие формализованных (алгоритмизированных) процедур анализа и синтеза и знание объекта исследования.

Анализ системных исследований показывает, что основная их трудность связана с нахождением адекватных понятийных средств представления исследуемых объектов как систем. Если такие понятийные средства удается найти, то последующая разработка технических средств исследования (соответствующих формальных аппаратов) оказывается хотя и сложной, но вполне разрешимой задачей. Конечными продуктами применения системного подхода являются системные описания изучаемых или проекты создаваемых объектов. Эти системные описания должны быть получены путем анализа и (или) синтеза в заданное время и способствовать большей эффективности и надежности управления.

1.1. Понятия о системах, больших системах, системном анализе

Объекты исследования целесообразно рассматривать с некоторых общих позиций как в концептуальном, так и в терминологическом смысле. В основе этой общности лежит понятие «система», которое не отличается формальностью, необходимой строгостью и универсальностью. Система (греч.) означает целое, составленное из частей соединение [2]. В более поздние исторические периоды понятие системы изменялось и наполнялось различным содержанием:

- это комплекс взаимодействующих элементов или совокупность элементов, находящихся в определенных отношениях друг с другом и со средой;
- средство или способ решения проблемы;
- множество вещей, свойств и отношений.

Условимся под системой понимать совокупность объектов, взаимодействующих друг с другом и с внешней средой, объединенных для достижения некоторой цели. С системой связан ряд часто используемых фундаментальных понятий:

- цель – результат, достигаемый использованием системы;

- элемент – простейшая часть системы, имеющая определенное функциональное назначение. Деление системы на элементы весьма условно;
- подсистема – совокупность элементов, способная выполнить относительно независимые функции в составе системы и содействующая достижению общей стоящей перед системой цели;
- связь – понятие, отражающее характер взаимодействия элементов друг с другом и с внешней средой;
- структура – отражение наиболее существенных взаимоотношений между элементами и подсистемами;
- состояние в момент t_0 – минимальный набор сведений о системе, позволяющий вместе с некоторым возможным входным воздействием, заданным на временном отрезке $[t_0, t_1]$, однозначно определить процесс на выходе системы для $t \in [t_0, t_1]$ при $t_1 \ge t_0$;
- поведение – закономерность перехода из одного состояния в другое с учетом входов и выходов системы;
- равновесие – способность системы сохранять свое состояние в отсутствие внешних воздействий;
- устойчивость – способность системы возвращаться в состояние равновесия после того, как она была выведена внешними воздействиями из этого состояния, а затем эти воздействия исчезли.

Понятие системы широко используется в самых различных отраслях знаний. Однако в связи со сложными научными и техническими проблемами, возникшими несколько десятилетий назад в энергетике, космонавтике, экологии, экономике, военном деле, возникли задачи исследования систем, функционирующих в ситуациях, когда число факторов, влияющих на поведение системы достаточно велико, а цели, которые должна достигнуть система, противоречивы и конфликтны.

Подобные сложные задачи привели к разработке самостоятельного аппарата исследований таких систем, а сами системы были выделены в отдельный класс и названы *большими (сложными)*. *Большая система* определяется как иерархически организованная и целенаправленно функционирующая совокупность большого числа информационно связанных и взаимодействующих элементов. Большие системы отличаются наличием характерных признаков[3]:

1. Большое количество $(>10^3)$ взаимодействующих элементов. Связи между элементами системы более существенны, чем с элементами, не принадлежащими системе. Это позволяет выделить большую систему как нечто самостоятельное из окружающей среды.
2. Невозможность полного формального описания системы как из-за сложности подобной операции, так и ее нецелесообразности в связи с громоздкостью и практической неприменимостью полученных результатов в случае успеха. Сложность описания определяется и тем, что

взаимодействие элементов друг с другом и внешней средой, как правило, подчинено стохастическим закономерностям.

3. Непостоянство структуры системы и ее функционирования. Система «развивается», и соответствующие изменения нельзя заранее предусмотреть.

4. Многокритериальность, нечеткое задание самих критериев эффективности. К тому же критерии не всегда могут быть четко сформулированы.

5. Эрготический характер системы, заключающийся в том, что часть функций выполняется автоматически, а другая – человеком, в частности лицом, принимающим решение (ЛПР).

6. Наличие интегративного качества, которое заключается в том, что система в целом обладает свойствами, не присущими ни одному ее элементу в отдельности. Это означает, что по результатам изучения свойств отдельных подсистем невозможно сделать заключение о свойствах системы в целом.

7. Иерархичность структуры, проявляющаяся в выделении в системе отдельных уровней и упорядочении взаимодействий между уровнями в порядке от высшего к низшему. На нижних уровнях используется более детальная и конкретная информация об отдельных сторонах работы системы, на высоких – обобщенная информация о работе системы в целом, необходимая для принятия решения по работе системы.

Особо следует подчеркнуть, что при исследовании конкретной системы бессмысленна попытка установить все из перечисленных признаков и на основании этого считать ее большой.

Основным методом исследования больших систем является системный анализ. Основные принципы системного анализа[3]:

- выявить и четко формализовать основную цель, достигаемую в результате принятия решения;
- рассмотреть проблему как целое, как единую систему и выявить все возможные последствия и взаимосвязи каждого частного решения;
- выявить и проанализировать возможные альтернативные пути достижения цели;
- цели, стоящие перед отдельными частями системы (подсистемы), не должны вступать в конфликт с глобальной целью большой системы.

Теоретическая дисциплина, которая занимается большими системами, с точки зрения системного анализа (подхода) носит название «исследование операций».

Большие системы являются наиболее сложными объектами математического моделирования и вычислительного эксперимента. Для решения проблем повышения эффективности их функционирования возникают различные задачи, среди которых целесообразно выделить:

- анализ системы, предназначенный для определения поведения или эффективности системы по ее известной математической модели;
- синтез систем, заключающийся в определении ее структуры и параметров (или только параметров при заданной структуре) при условии достижения заданного уровня эффективности. Чаще всего это задачи определения экстремумов показателей эффективности работы системы;
- цифровое моделирование и идентификация процессов с определенными детерминированными или вероятностными свойствами.

Последние из вышеприведенных задач исследования системы могут быть успешно решены с применением методов теории планирования эксперимента.

1.2. Основные понятия системного подхода

При изучении и преобразовании окружающей действительности необходимо получить первоначальное представление о ней в наиболее простом виде, а затем усложнять это представление только по мере необходимости, т.к. степень детальности (глубина) описания определяется практическими потребностями и может быть совершенно различной – от содержательного представления о назначении объекта до его формализованной математической модели [4].

Следовательно, должны быть найдены эффективные способы многоуровневого описания объекта.

Несмотря на то, что реальный окружающий нас мир есть бесконечное множество объектов и отношений (связей) между ними, очевидно, простейшим и наиболее удобным будет представление о реальной действительности в виде взаимодействия только двух элементов: в одном из них сосредоточено все то, что нас интересует в настоящее время; в другом – все остальное.

Бесконечное множество связей между этими элементами можно свести к одной входной и одной выходной связи, которые являются самыми существенными в определенном отношении.

Интересующий нас элемент назовем системой, другой элемент – внешней средой или просто средой. На рис. 1.1 представлена простейшая схема взаимодействия системы и внешней среды.

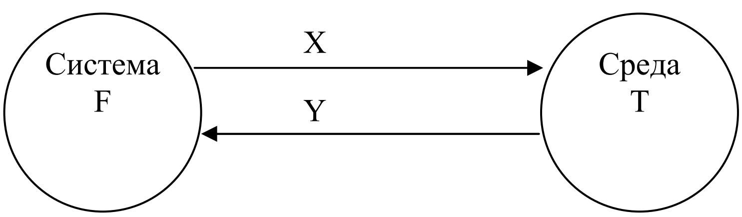


Рис. 1.1. Схема взаимодействия «система-среда»

По мере необходимости будем расчленять систему и среду на составляющие их части, соблюдая при этом обязательное условие:

внутренние элементы и связи между составными частями должны выделяться по степени своей существенности и значимости относительно главной (внешней) связи, ради которой собственно и была выделена система. На рис. 1.2 представлена детализация моделей системы и среды.

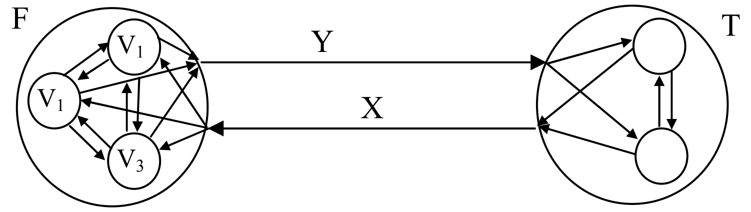


Рис. 1.2. Детализация моделей системы и среды

По входной и выходной связям между системой и средой происходит обмен путем взаимной передачи материальных, энергетических или информационных элементов. Элементы, передаваемые системой во внешнюю среду, называются конечными продуктами системы, а из среды в систему – ресурсами.

С точки зрения внешней среды система нужна этой среде как источник удовлетворения своих потребностей в конечных продуктах. Отсюда вывод: новую систему порождает наличие неудовлетворенной потребности – проблемной ситуации.

Проблемная ситуация определяется как возникшее или назревающее неудовлетворительное состояние элементов внешней среды, которое собственными средствами (внутренними системами во внешней среде) на заданном пространственно-временном интервале не может быть ликвидировано, то есть осознание проблемной ситуации можно считать исходным систематизирующим фактором, т.к. система есть средство решения возникшей проблемы.

Следующим шагом в создании системы будет определение ее цели. Под целью понимается информационный образ конечного продукта (целевого состояния) системы.

Цели объектов управления в автоматизированных системах имеют сложный характер. Глобальной целью системы называется информационный образ средства ликвидации проблемной ситуации, выраженный в терминах реализации неудовлетворенной потребности внешней среды (увеличение выпуска продукции, улучшение обслуживания населения, повышение качества подготовки специалистов и т.д.).

Во взаимоотношениях между системой и средой можно выделить два основных характерных режима:

1. Система полностью удовлетворяет имеющиеся потребности среды. Имеет место функционирование системы.

2. Система не удовлетворяет внешние потребности среды. Возникает потребность в развитии системы путем замены ее глобальной цели.

Под функционированием системы понимается процесс перехода системы или ее элементов из состояния в состояние при условии постоянства заданной цели.

Развитие системы связано с необходимостью изменения поведения системы при изменении цели.

Понятия функционирования и развития являются относительными: элементы системы могут находиться в состоянии развития, в то время как сама система функционирует.

Первый и второй способ деятельности системы будем рассматривать как ее внешнюю функцию.

Функция определяется как способ достижения системой поставленной цели. Задание способа достижения цели – функции системы предопределяют внутреннее устройство, конструкцию системы. Главной характеристикой внутреннего устройства системы является структура системы – совокупность внутренних связей между элементами системы.

Потребности системы в восстановлении элементов, вышедших из строя, в новых элементах, обеспечивающих перестройку структуры в случае развития системы, удовлетворяются внешней средой. Источник восполнения и дополнения элементов структуры системы через внешнюю среду называется внешними условиями.

Таким образом выявляется логическая последовательность действий при анализе и синтезе:

проблемная ситуация – цель – функция – структура – внешние условия.

Рассмотрим детально структуру системы, представленную совокупностью связей между ее элементами.

Отметим тот факт, что на практике существуют системы, имеющие одну и ту же цель, но различающиеся по структуре.

С другой стороны, можно предположить, что если несколько систем имеют одинаковые цели, то эти системы должны обязательно иметь нечто общее в своей структуре.

Это позволяет определить такие системы как системы одного назначения.

Назовем это общее формальной структурой системы (инвариантным аспектом системы). Под формальной структурой будем понимать совокупность существенных отношений, необходимых и достаточных для достижения системой заданной цели.

Например, формальной структурой любого производства являются предметы труда, орудия труда, рабочая сила и отношения, связывающие их.

Если рассматриваемые системы имеют одну и ту же цель, то можно сделать вывод о том, что цель системы однозначно определяет формальную структуру системы.

Подставив в функциональные места вместо формальных элементов соответствующие материальные элементы, получим материальную структуру или конструкцию системы.

Например, на предприятии в качестве орудий труда с заданной функцией могут выступить прессы или станки различного типа; рабочее место слесаря одного разряда могут занимать люди с различными нормами поведения, уровнями образования и культуры. Из вышесказанного делаем вывод, что одной формальной структуре может соответствовать множество материальных структур.

1.3. Модели систем

Под моделью системы будем понимать любую другую систему, обладающую той же формальной структурой при условиях:

1) между системными характеристиками (функцией, структурой, элементами) модели и оригиналом существует соответствие;

2) модель более доступна для оперирования имеющимися средствами, чем оригинал.

Следует оговориться, что модель обладает только выделенными свойствами, а оригинал и многими другими, т.к. их материальные структуры различны. Для описания систем в наибольшей степени должны применяться формальные, в том числе и языковые модели. Для построения формальных моделей необходимы формальные языки описания. Каждый формальный язык выполняет функции описания и классификации. При описании объекта необходимо найти в нем признаки, содержащиеся в понятиях данного языка.

Для классификации объекта необходимо подвести под одно из понятий языка специфический набор признаков, обнаруженных в объекте при его описании.

Способ классификации объектов называется классификатором. Если для описания системы используется несколько языков, то для каждого из них образуют свой классификатор.

Очевидно, что одним из основных направлений применения средств вычислительной техники сегодня является использование ЭВМ в системах классификации и распознавания образов. Это вызвано целым рядом причин. Во-первых, необходимостью обработки и упорядочивания больших объемов разнородной информации, с которой приходится иметь дело в процессе классификации и распознавания. Ведь именно от упорядоченности информации зависит скорость ее обработки, следовательно, и время принятия решений, которое должно стремиться к минимуму, особенно в условиях риска. Во-вторых, необходимостью автоматизации отдельных видов человеческой деятельности. Это может быть вызвано как повышенной опасностью для жизни человека или невозможностью использования человеческих ресурсов, так и ускорением работы определенного звена сложной технологической системы. В-третьих, для ускорения и оценивания принятия решений в различных ситуациях.

Ответ на вопрос: зачем вообще нужна классификация – еще более очевиден. Классификация и распознавание образов (объектов, сигналов, ситуаций, явлений или процессов) – едва ли не самые распространенные задачи, которые человеку приходится решать практически ежесекундно от первого до последнего дня своего существования. Распознавание текстов и анализ изображений, устройство автопилота и банкомат, научные исследования и прогнозирование – и это далеко не полный перечень того, где применяются системы классификации.

Классификация необходима в автоматизированных системах, предназначенных для управления сложными технологическими процессами, для использования в криминалистике, медицине, военном деле и т.д.

Для решения этих задач человек использует огромные ресурсы своего мозга, включая одновременно около 7-8 млрд. нейронов. Именно это дает возможность людям мгновенно узнавать друг друга, с большой скоростью читать печатные и рукописные тексты, безошибочно водить автомобили в сложном потоке уличного движения, разгадывать коды…

Для решения данных задач с помощью автоматизированных систем используется огромное множество разнообразных математических методов и алгоритмов, обладающих высокой степенью сложности и требующих ресурсов высокопроизводительных ЭВМ.

Разнообразие областей, в которых необходима классификация и отсутствие универсальных методов решения поставленной задачи, требует индивидуального подхода к каждой проблемной ситуации. Поэтому число методов и алгоритмов классификации постоянно растет, учитывая качество и скорость систем распознавания.

1.4. Структура системы распознавания

В настоящее время большие системы распознавания – это технические средства, предназначенные для выявления признаков объектов и измерения описывающих их параметров; совокупность алгоритмов распознавания, преобразующих входную информацию об объектах в определенные выводы; вычислительная техника, привлекаемая для реализации этих алгоритмов; коллективы специалистов, осуществляющие первичную формализацию исходной априорной информации, а также как полученных апостериорных данных, так и формальных решений задачи распознавания на всех уровнях системы.

В общем смысле распознавание представляет собой задачу преобразования входной информации, в качестве которой уместно рассматривать некоторые параметры, признаки распознаваемых образов, в выходную, представляющую собой заключение о том, к какому классу относится распознаваемый образ (объект). Упрощенная структура системы распознавания представлена на рис. 1.3.

Рис. 1.3. Упрощенная структура системы распознавания

В качестве входной информации используются признаки распознаваемых объектов, причем стараются сделать так, чтобы число признаков было минимальным, а информации, заложенной в них, достаточно для получения результата с высокой достоверностью.

Результатом работы классификатора является решение о принадлежности объекта одному из классов. При этом очень важно, чтобы точность классификации была как можно больше, а время достижения результата как можно меньше.

1.4.1. Классификация систем распознавания

Любая классификация основывается на определенных принципах. С точки зрения общности классификации систем распознавания рационально рассматривать в качестве классификационного принципа свойства информации, используемой в процессе распознавания.

В [5] приводится классификация систем распознавания:

1. По степени однородности информации об объектах:
   - простые;
   - сложные.
2. По количеству первоначальной априорной информации об объектах:
   - системы без обучения;
   - системы с обучением;
   - самообучающиеся системы.
3. По характеру информации о признаках объектов:
   - детерминированные;
   - вероятностные;
   - логические;
   - структурные.
4. По алгоритмам распознавания:
   - детерминированные;
   - вероятностные;
   - логические;
   - структурные;
   - комбинированные.

Обобщенная схема классификации систем распознавания представлена на рис. 1.4.

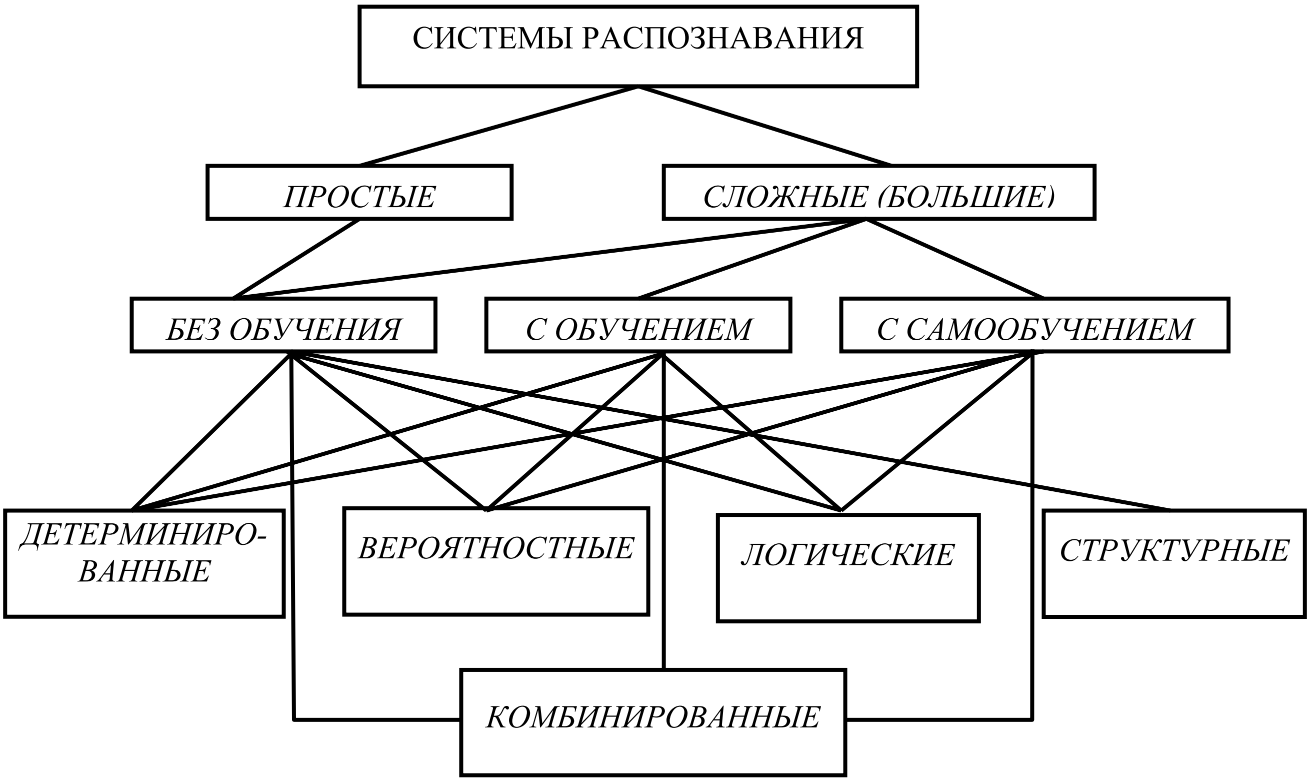


Рис. 1.4. Классификация систем распознавания

1.4.2. Классификация по степени сложности

Системы распознавания можно подразделить на простые и сложные в зависимости от того, какая информация используется для описания объектов. Если признаки, на языке которых произведено описание объектов и алфавита классов, имеют единую физическую природу, то такие системы, как правило, называют простыми. Если же для описания распознаваемых объектов используется физически разнородная информация, то такие системы являются более сложными.

Разделение систем распознавания на простые и сложные достаточно условно. Например, многоуровневая система распознавания, использующая физически однородную информацию для описания объектов и классов, может оказаться на практике весьма сложной.

Простые системы

К простым системам распознавания и классификации относят, например, автоматические распознающие устройства, в которых признаки рабочего словаря представляют собой лишь те или иные линейные размеры распознаваемых объектов; автоматы для размена монет, где в качестве признака, используемого для распознавания монет, берется их масса; автоматические устройства, предназначенные для определения бракованных деталей, где в качестве признаков, применяемых для описания классов бракованных и небракованных деталей, используются либо некоторые линейные размеры, либо масса и т.д.

Сложные системы

К ним относят, например, системы медицинской диагностики, в которых в качестве признаков (симптомов) могут использоваться данные анализа крови и кардиограмма, температура и динамика кровяного давления и т.п.; системы, предназначенные для распознавания образцов военной техники вероятного противника; системы распознавания состояний динамических объектов и т.д.

Сложные системы распознавания бывают одноуровневыми и многоуровневыми. В одноуровневых системах апостериорная информация о признаках классифицируемых объектов $x_1, x_2, ..., x_n$ определяется прямыми измерениями непосредственно на основе обработки результатов экспериментов.

В многоуровневых системах апостериорная информация о признаках может определяться на основе косвенных измерений. Здесь признаки каждого нового уровня могут быть определены с помощью признаков предыдущего уровня.

1.4.3. Классификация по типам признаков

Если в качестве принципа классификации использовать характер информации о признаках распознаваемых объектов, которые подразделены на логические, вероятностные, детерминированные и структурные, то в зависимости от того, на языке каких признаков производится описание этих объектов или же какой алгоритм распознавания реализован, системы распознавания могут быть подразделены на детерминированные, вероятностные, логические, структурные и комбинированные. Рассмотрим кратко основные из них.

Детерминированные признаки и системы

Детерминированные признаки – это признаки, принимающие конкретные числовые значения, которые могут быть рассмотрены как координаты точки, соответствующей данному объекту, в n-мерном пространстве признаков. При использовании в системах классификации таких признаков ошибки измерений должны быть учтены до подачи признаков на вход классификатора.

Детерминированные системы – это такие системы, которые в качестве признаков объектов и классов используют детерминированные признаки. Кроме того, в таких системах для построения алгоритмов распознавания используются геометрические меры близости, основанные на измерении расстояний между распознаваемым объектом и эталонами классов. В общем случае применение детерминированных методов распознавания предусматривает наличие координат эталонов классов в признаковом пространстве либо координат объектов, принадлежащих соответствующим классам.

Вероятностные признаки и системы

Вероятностные признаки – это признаки, случайные значения которых распределены по всем классам объектов, при этом решение о принадлежности распознаваемого объекта к тому или другому классу может приниматься только на основании конкретных значений признаков данного объекта, определенных в результате проведения соответствующих опытов. Признаки распознаваемых объектов следует рассматривать как вероятностные и в случае, если измерение их числовых значений производится с такими ошибками, что по результатам измерений невозможно с полной определенностью сказать, какое числовое значение данная величина приняла.

Вероятностные системы – это такие системы, которые в качестве признаков объектов и классов используют вероятностные признаки. Кроме того, в данных системах для построения алгоритмов распознавания используются вероятностные методы распознавания, основанные на теории статистических решений. В общем случае применение вероятностных методов распознавания предусматривает наличие вероятностных зависимостей между признаками распознаваемых объектов и классами, к которым эти объекты относятся.

Логические признаки и системы

Логические признаки распознаваемых объектов можно рассматривать как элементарные высказывания, принимающие два значения истинности (истина – ложь) с полной определенностью. К логическим признакам относятся прежде всего признаки, не имеющие количественного выражения. Эти признаки представляют собой суждения качественного характера типа наличия или отсутствия некоторых свойств или некоторых элементов у распознаваемых объектов или явлений. В качестве логических признаков можно рассматривать, например, такие симптомы в медицинской диагностике, как боль в горле, кашель и т.д. К логическим можно отнести также признаки, у которых важна не величина признака у распознаваемого объекта, а лишь факт попадания или непопадания ее в заданный интервал. В пределах этих интервалов появление различных значений признаков у распознаваемых объектов предполагается равновероятным. На практике логические признаки подобного рода имеют место в таких ситуациях, когда либо ошибками измерений можно пренебречь, либо интервалы значений признаков выбраны таким образом, что ошибки измерений практически не оказывают влияния на достоверность принимаемых решений относительно попадания измеряемой величины в заданный интервал. Например, в области технической диагностики решение о выходе из строя технических устройств принимается лишь тогда, когда фактические значения некоторых параметров (признаков) превышают заданные интервалы. Отклонение же значений параметров от номинала, не сопровождающееся выходом за пределы соответствующих интервалов, является информацией о том, что устройство функционирует нормально.

Логические системы – это такие системы, которые в качестве признаков объектов и классов используют логические признаки. Кроме того, в этих системах для построения алгоритмов распознавания используются логические методы распознавания, основанные на дискретном анализе и базирующемся на нем исчислении высказываний. В общем случае применение логических методов распознавания предусматривает наличие логических связей, выраженных через систему булевых уравнений, в которой переменные – логические признаки распознаваемых объектов, а неизвестные величины – классы, к которым эти объекты относятся.

1.4.4. Классификация по наличию априорной информации об объектах и классах

В работах [5, 6], посвященных распознаванию образов, по наличию априорной информации об объектах и классах системы распознавания делятся на системы без обучения, с обучением и самообучающиеся.

Системы без обучения

В таких системах, как правило, априорной информации достаточно для того, чтобы определить априорный алфавит классов, построить априорный словарь признаков и на основе непосредственной обработки исходных данных произвести описание каждого класса на языке этих признаков.

В [5] сказано, что построение систем распознавания без обучения возможно при наличии полной априорной информации, которая включает:

1) сведения о природе распознаваемых объектов или явлений;
2) решения, которые могут приниматься на основе результатов распознавания;
3) данные для построения априорного словаря признаков;
4) ограничения на создание измерительной аппаратуры;
5) зависимости между классами и признаками априорного словаря.

Пункты 1 и 2 являются исходными для определения принципа классификации и собственно проведения классификации.

Системы с обучением

Построение систем с обучением необходимо в том случае, когда отсутствует полная первоначальная априорная информация. Ее объем позволяет подразделить объекты на классы и определить априорный словарь признаков. Однако объем априорной информации недостаточен для того, чтобы в признаковом пространстве путем непосредственной обработки исходных данных построить описания классов объектов $K_1, K_2, ..., K_m$ на языке априорного словаря признаков. Такими описаниями могут быть, например, разделяющие функции $F_i(x_1, x_2, ..., x_n)$, $i = \overline{1, m}$, априорные вероятности появления объектов различных классов $P(K_i)$, условные плотности распределений $f_i(x_1, x_2, ..., x_n)$, $i = \overline{1, m}$, эталоны и т.д.

Изначально все множество объектов подразделено на классы $K_1, K_2, ..., K_m$ и определен вектор $\overline{X} = \{x_1, x_2, ..., x_n\}$, компоненты которого и составляют априорный словарь признаков.

Требуется на основании предъявления системе распознавания объектов обучающей выборки с указанием классов, которым они принадлежат, построить в многомерном признаковом пространстве гиперповерхность, разделяющую это пространство на области D_i, соответствующие классам $K_i, i = \overline{1, m}$. При этом разделение должно осуществляться в соответствии с заранее выбранным критерием оптимизации.

Если обучающая выборка достаточно представительна, то, по мнению авторов [5], в пределе подобная процедура приводит к достаточно точному описанию классов и, следовательно, возможности определения таких границ классов, придерживаясь которых можно достичь потенциально достижимой точности работы системы распознавания.

1.5. Этапы разработки классификатора

Процесс классификации можно разделить на два этапа:

1. Разработка классификатора.
2. Определение состояния объекта с помощью данного классификатора, т.е. этап функционирования.

Этап разработки классификатора схематично представлен в виде двух блоков на рис. 1.5.

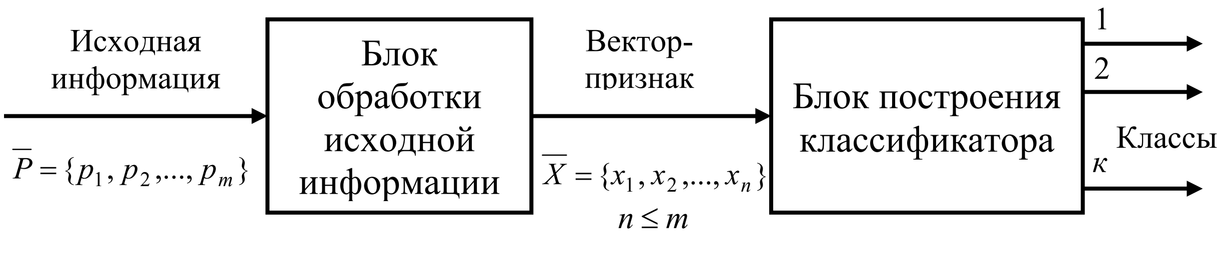


Рис. 1.5. Этап разработки классификатора

На вход блока обработки исходной информации поступает множество признаков объекта $\overline{P} = \{p_1, p_2, ..., p_m\}, |P| = m$, которыми могут быть, например, показания датчиков, тренды, результаты анализов и т.д. На выходе имеем вектор-признак $\overline{X} = \{x_1, x_2, ... x_n\}$, $|X| = n$, причем $\overline{X} \subseteq \overline{P}, n \le m$.

Кроме того, число признаков объекта может быть довольно велико, следовательно, их обработка будет требовать больших затрат машинного времени, что непременно скажется на эффективности классификатора. Поэтому программное средство, реализующее функции данного блока, должно также корректно понижать размерность пространства измерений от нескольких тысяч до сотен или даже десятков признаков.

Следовательно, на данный блок возлагаются функции, во-первых, определения информационной значимости признаков, во-вторых, выявления взаимозависимости признаков, в-третьих, отбрасывания незначимых признаков. Сделать это корректно не так просто в связи с тем, что все признаки взаимосвязаны и ценность одних может меняться весьма значительно при отбрасывании других, даже несущественных признаков.

Основные задачи блока обработки исходной информации:

1. Определение информационной значимости признаков.
2. Выявление взаимозависимости признаков.
3. Корректное понижение размерности признакового пространства.

На вход блока построения классификатора поступает так называемый вектор-признак объекта $\overline{X} = \{x_1, x_2, ...x_n\}$, $|X| = n$, который содержит только те признаки, без которых организация процесса распознавания с требуемой точностью невозможна. На выходе данного блока должно быть решение об отнесении распознаваемого объекта к определенному классу, или же, иными словами, класс распознаваемого объекта K_j, $j = \overline{1, l}$.

Таким образом, основная задача блока построения классификатора состоит в разработке структуры, позволяющей определять состояние объекта по его признакам $\overline{X} = \{x_1, x_2, ...x_n\}$. Данная структура может представлять собой совокупность возможных состояний объекта в виде классов, разделяющих функций, критериев и т.д. в зависимости от наличия априорной информации, типов используемых признаков и методики разработки. Важно отметить, что все детерминированные признаки, подаваемые на вход данного блока, должны быть формализованы[1].

Основные функции первого этапа:

1. Ввод исходной информации.
2. Обработка исходной информации.
3. Формализация признаков.
4. Разработка алгоритма классификации.
5. Определение возможных состояний объекта (классов, критериев).
6. Анализ качества работы классификатора.

Этап определения состояния объекта с помощью классификатора схематично представлен в виде двух блоков на рис. 1.6.

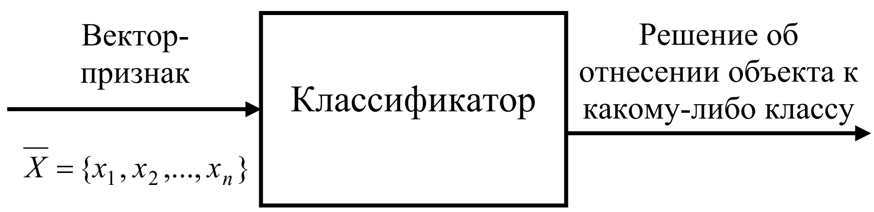


Рис. 1.6. Этап определения состояния объекта

[1] Формализация данных необязательна в случае употребления определенных мер расстояния.

На втором этапе вектор-признак $\overline{X} = \{x_1, x_2, ..x_n\}$ подается на вход классификатора, который на основании этих данных по алгоритму определяет класс, к которому принадлежит распознаваемый объект. Необходимо отметить, что алгоритм определения объекта с помощью классификатора отличается от алгоритма разработки классификатора. Иными словами, первый определяет структуру классификатора, а второй – поиск альтернативных вариантов структуры.

Основные функции второго этапа:

1. Обработка исходной информации в реальном времени.
2. Определение состояния классифицируемого объекта на основании его признаков с помощью специального алгоритма.

Первый этап может занимать значительное количество как машинного, так и реального времени и потребовать знаний экспертов в различных областях. На данном этапе первостепенными являются вопросы эффективности работы классификатора.

На втором этапе используется готовый классификатор, который уже в реальном времени должен определять класс динамического объекта. Здесь важным является не только эффективность определения класса объекта, но и быстродействие классификатора.

1.6. Основные задачи классификации

Классификация сложных объектов или явлений требует создания специальных систем распознавания – сложных динамических систем, состоящих в общем случае:

- из коллектива подготовленных специалистов, иначе их еще называют экспертами;
- совокупности технических средств получения и обработки информации;
- алгоритмов задач распознавания;
- априорной информации об объектах.

В процессе проектирования и построения классификаторов возникают задачи:

1. Определение признаков объектов.
2. Составление априорного алфавита классов.
3. Разработка априорного словаря признаков.
4. Описание классов объектов на языке признаков.
5. Разбиение априорного пространства признаков на области, соответствующие классам априорного алфавита классов.
6. Построение рабочего алфавита признаков и классов.
7. Оценка эффективности работы системы.

Задача 1. Заключается в том, чтобы определить полный набор признаков, характеризующих объекты или явления, для распознавания которых разрабатывается данная система. Названная совокупность признаков должна быть сформирована без относительно каких-либо ограничений, связанных как с получением априорной информации, необходимой для исходного описания классов объектов, так и с получением апостериорной информации о конкретных объектах, подлежащих распознаванию. Наоборот, первоначально необходимо определить все признаки, хотя бы в малейшей мере характеризующие объекты или явления.

Задача 2. Заключается в проведении первоначальной классификации распознаваемых объектов или явлений и в составлении априорного алфавита классов. Главным в этой задаче является выбор правильного принципа классификации, удовлетворяющего требованиям, предъявляемым к системе классификации, которые в свою очередь зависят от того, какие решения могут приниматься системой управления по результатам распознавания неизвестных объектов или явлений. При решении последующих задач из априорного алфавита классов формируется рабочий алфавит классов системы.

Задача 3. Заключается в разработке априорного словаря признаков, в который войдут только те признаки, относительно которых может быть получена априорная информация, необходимая для описания классов на языке этих признаков.

Задача 4. Состоит в описании всех классов априорного алфавита классов на языке признаков, включенных в априорный словарь признаков. Данная задача может быть решена с помощью методов обучения, самообучения или непосредственной обработки исходных данных.

В зависимости от вида признаков распознаваемого объекта строятся и описания классов. Если признаки объекта вероятностные, то для описания классов потребуется определить характеристики распределений: априорные вероятности $P(K_i)$ принадлежности объекта классу K_i, функции плотности вероятности $f_i(x_1, x_2, ..., x_N)$ значений параметров $x_1, x_2, ..., x_N$ при условии, что объект принадлежит классу K_i.

Если признаки объекта детерминированные, то классы могут быть описаны с помощью эталонов. В качестве эталона используют точку n-мерного пространства, равноудаленную от всех точек, описывающих объекты, принадлежащие данному классу.

Если признаки распознаваемых объектов логические и имеют количественные выражения, то для описания классов объектов на языке признаков необходимо определить диапазоны значений признаков Δx_j^i, $j = \overline{1, N}$, соответствующие классам $K_i, i = \overline{1, m}$. При этом каждый из отрезков может рассматриваться как элементарное логическое высказывание A,B,C,… Если признаки распознаваемых объектов есть суждения качественного характера, то каждый из них рассматривается как элементарное логическое высказывание A’,B’,C’... Для описания классов на языке этих признаков необходимо выяснить, какими из них характеризуется каждый класс, после этого установить зависимости в форме булевых соотношений между признаками A,B,C,…; A’,B’,C’... и классами $K_1, K_2, ..., K_m$.

Задача 5. Задача разбиения априорного пространства признаков на области, соответствующие классам априорного алфавита классов, должна быть решена оптимальным, насколько это возможно, способом, например, чтобы обеспечить минимальное значение ошибок, неизбежно сопровождающих распознавание поступающих на вход системы распознавания объектов или явлений. Если разбиение объектов на классы $K_1, K_2, ..., K_m$ произведено, требуется выделить в пространстве признаков области $D_{i,} i = \overline{1, m}$, эквивалентные классам. Это означает, что если объект, имеющий признаки $x_1, x_2, ..., x_N$, относится к классу K_i, то представляющая его точка в пространстве признаков принадлежит области D_i.

Задача 6. В [5] представляет собой общую постановку проблемы распознавания. Суть ее заключается в разработке такого алфавита классов и такого словаря признаков (их авторы называют оптимальными), которые в условиях ограничений на построение системы распознавания обеспечивают максимальное значение показателя эффективности системы распознавания, принимающей в зависимости от результатов распознавания неизвестных объектов соответствующие решения.

Задача 7. Состоит в выборе показателей эффективности системы распознавания и оценке их значений. В качестве показателей эффективности системы могут быть выбраны, например, среднее время распознавания, вероятность правильной работы, величина расходов, связанных с получением апостериорной информации и т.д.

Эффективность работы системы, как правило, оценивается на основании экспериментальных данных, полученных в ходе исследований, либо реальной или смоделированной системы, либо с помощью ее физической или математической модели.

1.7. Методы классификации

В процессе развития теории распознавания появляется достаточно большое количество методов, позволяющих решать поставленные задачи, так что классификация этих методов является условной и неоднозначной.

Авторы [5, 6] приводят различную типологию методов классификации. Одни различают параметрические, непараметрические и эвристические методы, другие выделяют группы методов, исходя из исторически сложившихся школ и направлений в данной области.

Исторически сложилось так, что теория распознавания образов развивалась по двум направлениям: детерминистскому и статистическому, хотя чаще всего строго различить их не удается.

Каждый из существующих методов классификации не является универсальным, то есть может быть применим только для решения определенного класса задач. Кроме того, каждому методу присущи свои достоинства и недостатки. Стремление создать универсальный метод или преодолеть недостатки ранее разработанных – все это объясняет существование большого количества методов классификации.

Большой интерес сегодня по-прежнему представляют методы Байеса и кластерного анализа.

Выбор того или иного метода определяется исходя из количества и типа признаков, природы распознаваемых объектов, целей классификации, вида желаемого результата, требований к точности и т.д.

1.7.1. Байесовская классификация

Байесовский классификатор основан на критерии, который представляет собой правило, в соответствии с которым классификация производится таким образом, чтобы обеспечить минимум среднего риска. Применение критерия Байеса целесообразно в случае, когда система распознавания многократно осуществляет распознавание неизвестных объектов или явлений в условиях неизменного признакового пространства, при стабильном описании классов и неизменной платежной матрице.

Минимум риска, усредненного по множеству решений задачи распознавания неизвестных объектов, обеспечивается тогда, когда решения о принадлежности объекта классу K_1 или K_2 принимаются в соответствии с правилом: если измеренное значение признака у данного объекта расположено в области R_1, то объект относится к классу K_1, если в области R_2, – к классу K_2.

Стратегию, основанную на данном правиле, и минимальный средний риск называют байесовскими.

Байесовская стратегия может быть описана следующим образом. Пусть в результате опыта установлено, что значение признака у распознаваемого объекта составляет величину $x = x^0$. Тогда условная вероятность принадлежности объекта классу K_1 (условная вероятность первой гипотезы в соответствии с теоремой гипотез или формулой Байеса)

$$P(K_1 / x^0) = \frac{P(K_1) f_1(x^0)}{f(x^0)}. \tag{1.1}$$

Условная вероятность принадлежности объекта классу K_2 (условная вероятность второй гипотезы)

$$P(K_2 / x^0) = \frac{P(K_2) f_2(x^0)}{f(x^0)}, \tag{1.2}$$

где $f(x^0) = P(K_1)f_1(x^0) + P(K_2)f_2(x^0)$ – совместная плотность распределения вероятностей значений признака x по классам, в свою очередь величины $P(K_1)$ и $P(K_2)$ – апостериорные вероятности принадлежности распознаваемого объекта классам K_1 и K_2 соответственно.

Таким образом, байесовский подход к решению задачи распознавания состоит в вычислении условных апостериорных вероятностей и принятии решения на основании сравнения их значений. Именно такой подход обеспечивает минимум среднего риска и минимум ошибочных решений.

Если число классов больше двух и равно m, то апостериорная вероятность отнесения объекта к K-му классу

$$P(K_i / x) = \frac{P(K_i)f_i(x^0)}{\sum_{i=1}^{m} P(K_i)f_i(x^0)}. \qquad (1.3)$$

Когда объект Q характеризуется n признаками $x_j, j = \overline{1,n}$, и признаки распознаваемого объекта приняли значения $x_1 = x_1^0$; $x_2 = x_2^0$; …; $x_n = x_n^0$, вероятность того, что объект относится к i-му классу

$$P(K_i / Q) = \frac{P(K_i)f_i(x_1^0, x_2^0, ..., x_n^0)}{\sum_{i=1}^{m} P(K_i)f_i(x_1^0, x_2^0, ..., x_n^0)}. \qquad (1.4)$$

Достоинства:

- относительная простота реализации для большого числа классов;
- пространство признаков для известных законов распределения.

Если аналитические выражения для функции плотности вероятностей признаков известны, то задача сводится к вычислению наиболее вероятного класса принадлежности, что не представляет большого труда.

Недостатки:

- ограниченность применения;
- сложность реализации для неизвестных распределений.

Ограниченность применения метода Байеса заключается в том, что он может быть применим только в случае наличия априорной информации об объектах и классах, то есть когда известны условные плотности вероятности. При решении практических задач точное численное значение этих вероятностей получить очень сложно.

Существенным ограничением методов теории статистических решений является сложность их практической реализации. Если плотности вероятностей не удается представить аналитически, то необходимо хранить в памяти их значения для каждой точки n-мерного пространства. Объем требуемой памяти при этом так велик, что построение распознающих машин, основанных на вычислении коэффициента подобия, практически невозможно без дополнительных упрощений.

1.7.2. Кластерный анализ

Кластерный анализ – ряд математических методов многомерного анализа, позволяющих на основе множества показателей, характеризующих ряд объектов, сгруппировать их в классы (кластеры) таким образом, чтобы объекты, входящие в один класс, были более однородными, сходными с объектами, входящими в другие классы.

Кластерный анализ позволяет выявить внутреннюю структуру распределения многомерных объектов по совокупности показателей. Он отвечает на следующие вопросы: относятся ли все наблюдения к одной совокупности (однородная выборка), или же выборка неоднородная, т.е. данные взяты из разных совокупностей (кластеров); сколько различных кластеров и какие наблюдения относятся к какому кластеру? Тем самым появляется возможность составлять однородные выборки объектов для дальнейшего статистического анализа.

Применяя методы кластерного анализа, можно решать такие задачи классификации:

- первоначальная классификация, которая проводится с целью составления априорного алфавита классов и анализа структуры данных;
- выделение подклассов или уровней классификации путем группировки или классификации объектов;
- разделение пространства признаков на области, соответствующие классам, с целью выявления пересекающихся и непересекающихся областей, которые определяются с помощью гиперсфер и межкластерных расстояний;
- классификация объектов в случае отсутствия какой бы то ни было информации о количественном и качественном составе классов.

Кроме того, методы кластерного анализа достаточно удобны для построения иерархической системы классификации многомерных объектов. Иными словами, если требуется провести классификацию по нескольким признакам, то производят ранжирование по степени важности и проводят классификацию по первому признаку, затем полученные подклассы разбивают на подклассы по второму и т.д. Подобным образом строится большинство комбинационных статистических группировок.

К *достоинствам* методов кластерного анализа следует отнести широту их применения в различных областях, работоспособность при полном отсутствии априорной информации об объектах и классах, возможность классификации многомерных объектов, которые очень трудно разделить в пространстве с помощью линейных и нелинейных функций и т.д.

Однако существенным *недостатком* методов кластерного анализа является необходимость вычисления матрицы расстояний, что всегда сопряжено с временными затратами и большими ресурсами памяти. Данный недостаток несущественен при малом числе объектов и классов и становится значительным при их увеличении.

1.7.3. Дискриминантный анализ

Основная идея дискриминантного анализа – составить функцию исходных показателей, которая обеспечивает оптимальное в некотором смысле разделение объектов, относящихся к разным классам. Не исключено, что среди исходных переменных уже имеется такой определяющий признак, и по его значениям можно безошибочно относить объекты к тому или иному классу. В общем же случае дискриминантную функцию составляют в виде некой линейной комбинации исходных показателей, коэффициенты которой подбирают из условия наибольших различий между известными классами.

Использование дискриминантного анализа для решения задач классификации объектов по признакам предполагает наличие по крайней мере двух этапов:

1. Построение дискриминантных функций путем анализа объектов и обучения системы.
2. Классификация неизвестных объектов.

Первый этап является самым трудоемким и определяющим весь дальнейший процесс классификации. Если на первом этапе дискриминантные функции найдены верно, то есть они обеспечивают требуемое качество классификации, то на втором этапе, используя входные данные, остается лишь вычислять их значения и по ним определять класс принадлежности объекта. Скорость работы такого классификатора на втором этапе определяется характером вычислений и количеством построенных функций (классов).

Формирование дискриминантной функции, разделяющей два или несколько классов объектов, основывается обычно на одном из методов:

- построение решающих правил;
- построение линейных разделяющих функций;
- потенциальные функции и другие.

Для реализации каждого из этих методов необходима обучающая выборка. Обучающая выборка – это множество объектов, заданных значениями признаков, и принадлежность которых к тому или иному классу достоверно известна учителю и сообщается учителем «обучаемой» системе. По обучающей выборке система строит решающие правила, функции, рассчитывает критерии и т.д. Качество таких систем оценивается по контрольной (экзаменационной) выборке, в которую входят объекты, заданные значениями признаков, и принадлежность которых тому или иному образу известна только учителю. Предъявляя обучаемой системе для контрольного распознавания объекты экзаменационной выборки, учитель в состоянии дать оценку вероятностей ошибок распознавания, то есть оценить качество обучения. К обучающей и контрольной выборкам предъявляются определённые требования. Важно, чтобы объекты экзаменационной выборки не входили в обучающую выборку (иногда, правда, это требование нарушается, если общий объём выборок мал и увеличить его либо невозможно, либо чрезвычайно сложно). Обучающая и экзаменационная выборки должны достаточно полно представлять генеральную совокупность (гипотетическое множество всех возможных объектов каждого образа). Например, при обучении системы медицинской диагностики в обучающей и контрольной выборках должны быть представлены пациенты различных половозрастных групп, с различными анатомическими и физиологическими особенностями, сопутствующими заболеваниями и т.д.

Перечисленные выше методы всегда применяются при наличии априорной информации об объектах и классах или после первоначальной классификации, когда получена предварительная структура системы. Они используют максимум дополнительной информации и обеспечивают высокую надежность и относительную простоту классификации в случае небольшого количества линейно разделяемых классов. Однако при увеличении числа классов и признаков объектов становятся труднореализуемыми и малопригодными. Кроме того, в случае линейно неразделяемых классов сложность реализации данных методов заставляет искать новые пути решения поставленной проблемы.

Определенные ограничения на процесс построения дискриминантных функций накладывает размерность обучающей выборки.

Что касается вычислительных ресурсов, требуемых для реализации методов дискриминантного анализа, то они вполне приемлемы и определяются в основном размерностью обучающей выборки и количеством классовых областей, которые в свою очередь определяют число разделяющих функций или решающих правил.

1.8. Математический аппарат для разработки алгоритма классификации

1.8.1. Основные критерии кластеризации

Разбиение объектов на кластеры всегда осуществляется в соответствии с некоторым, заранее выбранным критерием качества.

Основными критериями качества, применяемыми в кластерном анализе, являются:

- внутригрупповая сумма квадратов отклонений;
- дисперсия объектов внутри класса;
- расстояние между объектами класса;
- мера сходства объектов кластера и другие.

Критерий качества кластеризации в той или иной мере отражает неформальные требования:

- внутри групп объекты должны быть тесно связаны между собой;
- объекты разных групп должны быть далеки друг от друга;
- при прочих равных условиях распределения объектов по группам должны быть равномерными.

Первые два требования выражают стандартную концепцию компактности классов разбиения; последнее требование состоит в том, чтобы критерий не навязывал объединения отдельных групп объектов.

1.8.2. Общая внутриклассовая дисперсия

Минимизация внутриклассовой дисперсии – это один из самых распространенных критериев в кластерном анализе, хотя используется и в задачах дискриминантного и дисперсионного анализа. Естественная сфера его применения – выделение кластеров в евклидовом пространстве, имеющих шарообразную форму.

$$D = \sum_{i=1}^{m} \sum_{j=1}^{n} \sigma_{ji}^2 , \tag{1.5}$$

где σ_{ji}^2 – дисперсия j-го признака в i-м классе;

n – число признаков;

m – число классов.

При этом дисперсия каждого признака может быть определена так:

$$\sigma^2 = \frac{\sum_{j=1}^{k} x_j^2}{k} - \frac{\left(\sum_{j=1}^{k} x_j \right)^2}{k^2} , \tag{1.6}$$

где x_j – признак объекта;

k – число объектов в классе.

1.8.3. Внутригрупповая сумма квадратов отклонений

Внутригрупповая сумма квадратов отклонений (ВСКО) в общем виде вычисляется следующим образом:

$$\begin{aligned} W &= \sum_{i=1}^{m} (X_i - \overline{X})^2 = \sum_{i=1}^{m} (X_i^2 - 2X_i \overline{X} + \overline{X}^2) = \sum_{i=1}^{m} X_i^2 - \sum_{i=1}^{m} 2X_i \overline{X} + \sum_{i=1}^{m} \overline{X}^2 = \\ &= \sum_{i=1}^{m} X_i^2 - 2\sum_{i=1}^{m} X_i \sum_{i=1}^{m} \frac{X_i}{m} + \sum_{i=1}^{m} \left(\sum_{i=1}^{m} \frac{X_i}{m} \right)^2 = \\ &= \sum_{i=1}^{m} X_i^2 - \frac{2}{m} \sum_{i=1}^{m} X_i \sum_{i=1}^{m} X_i + \frac{1}{m^2} \sum_{i=1}^{m} \left(\sum_{i=1}^{m} X_i \right)^2 = \\ &= \sum_{i=1}^{m} X_i^2 - \frac{2}{m} \left(\sum_{i=1}^{m} X_i \right)^2 + \frac{m}{m^2} \left(\sum_{i=1}^{m} X_i \right)^2 = \sum_{i=1}^{m} X_i^2 - \frac{1}{m} (\sum_{i=1}^{m} X_i)^2 , \end{aligned} \tag{1.7}$$

где X_i представляет собой измерение i-го объекта;

m – число объектов в классе.

Очевидно, что чем меньше ВСКО кластера, тем плотнее объекты располагаются в кластере и тем больше объекты «схожи» друг с другом.

С учетом того что каждый объект обладает несколькими признаками, выражение примет вид

$$W = \frac{1}{n}\sum_{j=1}^{n}\left[\sum_{i=1}^{z} x_{ji}^{2} - \frac{1}{z}(\sum_{i=1}^{z} x_{ji})^{2}\right], \qquad (1.8)$$

где x_{ji} – j-й признак i-го объекта;

n – число признаков объекта;

z – количество объектов в классе.

1.8.4. Квадраты внутриклассовых и межклассовых расстояний

Квадраты внутриклассовых расстояний, как и внутриклассовая дисперсия, – один из самых распространенных критериев кластерного анализа, хотя может использоваться и в задачах дискриминантного анализа:

$$F = \sum_{i,j \in K} d_{ij}^{2}\ , \qquad (1.9)$$

где d_{ij} – расстояние между объектами i и j класса K .

Данный критерий используют для определения размеров классов, а также сходства объектов внутри класса. При этом чем больше величина F, тем больше размер класса и тем меньше сходство объектов в нем.

Часто вместо квадратов внутриклассовых расстояний с целью выявления компактных удаленных групп объектов $F = F_{in} - F_{out}$ используют средние внутриклассовые и межклассовые расстояния [7]:

$$F_{in} = \frac{\sum\limits_{i,j \in K_l} d_{ij}}{\sum\limits_{l=1}^{k} n_l^2}\ , \qquad (1.10)$$

где F_{in} – среднее внутриклассовое расстояние;

n_l – число объектов l класса;

k – число классов;

d_{ij} – расстояние между объектами i и j класса S_l.

$$F_{out} = \frac{\sum\limits_{i \in S_q, j \in K_l} d_{ij}}{\sum\limits_{l<q} n_l n_q}\ , \qquad (1.11)$$

где F_{out} – среднее межклассовое расстояние;

n_l, n_q – число объектов l и q классов соответственно;

d_{ij} – расстояние между объектами i и j классов S_l и S_q.

В качестве показателя эффективности, отражающего основные требования классификации с точки зрения кластерного анализа, используют отношение суммы средних внутриклассовых расстояний к сумме средних межклассовых.

$$F = \frac{\sum\limits_{k=1}^{m} D_{in}}{\sum\limits_{i=1}^{m} \sum\limits_{j=i+1}^{m} D_{out}}, \qquad (1.12)$$

где $D_{in} = \sum\limits_{i,j \in K} d_{ij}$ – среднее внутриклассовое расстояние;

m – количество классов;

d_{ij} – расстояние между объектами i и j класса k;

n – число объектов в классе;

$D_{out} = \sum\limits_{j=1}^{n2} \sum\limits_{i=1}^{n1} d_{ij} / n_1 n_2$ – среднее расстояние между классами K_1 и K_2;

d_{ij} – расстояние между объектами i и j, $i \in K_1, j \in K_2$;

где n_1 – число объектов в классе K_1;

n_2 – число объектов в классе K_2.

1.9. Формализация данных

Мера расстояния вычисляется на основании исходных показателей признаков $x_i, i = \overline{1, n}$, имеющих разную физическую природу и различную размерность. Для того чтобы их можно было комбинировать, необходимо представлять данные в таком формате, который бы не влиял на определения расстояний между объектами. Иными словами необходима формализация данных, переводящая их в безразмерные величины.

Однако в результате формализации картина классификации может полностью измениться, как показано на рис. 1.7.

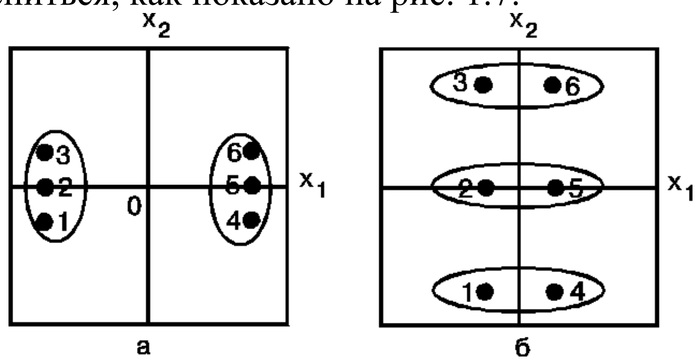


Рис. 1.7. Расслоение объектов в пространстве двух показателей до (а) и после (б) нормировки

Здесь при одних масштабах визуально выделяется два кластера (1,2,3) + (4,5,6), а при других – три совершенно иных кластера (1,4) + (2,5)+ + (3,6).

Такие курьезные ситуации описаны в книге [7] и могут иметь место для всех нормировок, связанных с конкретной выборкой.

1.9.1. Стандартизация всех переменных

Наиболее популярной является стандартизация всех переменных. Именно этот тип нормировки будет предложен исследователю в первую очередь. При стандартизации в качестве нормы принимается выборочное среднеквадратическое отклонение σ:

$$x_j^{'} = \frac{x_j - \overline{x_j}}{\sigma}. \tag{1.13}$$

Стандартизация – привычная операция. Согласно закону больших чисел, стандартизованные значения, как правило, не превосходят по модулю 3, поэтому после стандартизации сразу видны все сомнительные и выпадающие данные (их стандартизованные значения больше 3). Это относится к *достоинствам* данной нормировки.

Однако есть и *недостатки*:

при стандартизации все привязывается к данной выборке наблюдений, а мы вовсе не уверены, что наша выборка репрезентативная (представительная);

после стандартизации все показатели становятся равноправными с дисперсиями, равными единице. Но до стандартизации показатели имели разную изменчивость даже в относительных единицах (по коэффициенту вариации). Не будет ли ошибкой такое уравнивание их по изменчивости?

1.9.2. Альтернативная нормировка

Альтернативная нормировка учитывает различия в вариации отдельных показателей:

$$x_j^{'} = \frac{x_j}{\overline{x_j}}. \tag{1.14}$$

Данная нормировка лишена некоторых недостатков предыдущей, но имеет свои.

Проблемы могут возникнуть, если среднее какого-либо показателя близко к нулю $\overline{x_j} \approx 0$.

1.9.3. Модификация альтернативной нормировки

Модификация альтернативной нормировки выглядит следующим образом:

$$x_j^{'} = \frac{x_j}{x_{j\max} - x_{j\min}}. \tag{1.15}$$

Теперь знаменатель никогда не будет равен нулю, но вместо улучшения полученная нормировка оказалась хуже стандартизации, т.к. среднее квадратичное отклонение Sx_j – более устойчивая оценка изменчивости, чем размах $x_{j\max} - x_{j\min}$.

1.9.4. Перспективная нормировка

Перспективная нормировка представляет собой отношение

$$x_j' = \frac{x_j}{a_j}, \tag{1.16}$$

где a_j – некоторое эталонное значение, например оптимальное, которое должен назначить специалист. Но специалисту большей частью неизвестны эталонные значения, а проблема определения оптимальных параметров чаще всего является конечной целью исследования.

1.9.5. Нормировка по размаху

Данный способ позволяет перейти в безразмерное признаковое пространство, учитывая возможные границы (размах) изменения признака. В этом случае новая формализованная величина определяется согласно выражению

$$x_j' = \frac{x_j - x_{j\min}}{x_{j\max} - x_{j\min}}. \tag{1.17}$$

В этом случае значение каждого признака будет лежать в пределах [0;1]. Расстояние между объектами по признакам изменяется, но остается пропорциональным исходным данным.

1.9.6. Функции расстояния

Неотрицательная вещественнозначная функция $d(X_i, Y_j)$ называется функцией расстояния (метрикой), если:

1) $d(X_i, X_j) \geq 0$ для всех X_i и X_j;

2) $d(X_i, X_j) = 0$ тогда и только тогда, когда $X_i = X_j$;

3) $d(X_i, X_j) = d(X_j, X_i)$;

4) $d(X_i, X_j) \leq d(X_i, X_k) + d(X_k, X_j)$.

Значение $d(X_i, X_j)$ для заданных X_i и X_j называется расстоянием между X_i и X_j, где X_i и X_j могут быть и векторами в n-мерном пространстве.

Расстояния между парами векторов $d(X_i, X_j)$ могут быть представлены в виде симметричной матрицы расстояний:

$$D = \begin{Bmatrix} 0 & d_{12} & \cdots & d_{1n} \\ d_{21} & 0 & \cdots & d_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ d_{n1} & d_{n2} & \cdots & 0 \end{Bmatrix}, \tag{1.18}$$

причем диагональные элементы матрицы $d_{ii} = 0$ для $i = \overline{1, n}$.

Рассмотрим основные функции расстояний, применяемых при определении близости объектов.

1.9.7. Евклидово расстояние

Евклидово расстояние – наиболее общий и часто употребляемый тип расстояния. Оно попросту является геометрическим расстоянием в многомерном пространстве и вычисляется следующим образом:

$$d(X,Y)=\sqrt{\sum_{i=1}^{m}(X_i-Y_i)^2}\ , \qquad (1.19)$$

где $X=(X_1,\ldots,X_m)$ и $Y=(Y_1,\ldots,Y_m)$ – объекты, между которыми определяется расстояние;
или в векторном виде

$$d(X,Y)=\sqrt{(X-Y)^T(X-Y)}\ , \qquad (1.20)$$

где X,Y – векторы-признаки объектов.

Заметим, что евклидово расстояние (и его квадрат) вычисляется по исходным, а не по стандартизованным данным. Это обычный способ его вычисления, который имеет определенные преимущества (например, расстояние между двумя объектами не изменяется при введении в анализ нового объекта, который может оказаться выбросом). Тем не менее на расстояния могут сильно влиять различия между осями (единицами измерения параметров), по координатам которых вычисляются эти расстояния [6].

1.9.8. Квадрат евклидова расстояния

Иногда может возникнуть желание возвести в квадрат стандартное евклидово расстояние, чтобы придать большие веса более отдаленным друг от друга объектам. Это расстояние вычисляется следующим образом:

$$d(X,Y)=\sum_{i=1}^{m}(X_i-Y_i)^2\ , \qquad (1.21)$$

где $X=(X_1,\ldots,X_m)$ и $Y=(Y_1,\ldots,Y_m)$ – объекты, между которыми определяется расстояние;
или в векторном виде

$$d(X,Y)^2=(X-Y)^T(X-Y)\ , \qquad (1.22)$$

где X,Y – векторы-признаки объектов.

Эвклидово расстояние и его квадрат целесообразно использовать для анализа количественных данных [7].

1.9.9. Манхэттенское расстояние

Это расстояние является средним разностей по координатам. В большинстве случаев эта мера расстояния приводит к таким же результатам, как и для обычного расстояния Евклида. Однако отметим, что для этой меры влияние отдельных больших разностей (выбросов) уменьшается (так как они не возводятся в квадрат). Манхэттенское расстояние, иначе еще называемое как l_1 – норма, вычисляется по формуле

$$d(X,Y)=\sum_{i=1}^{m}\left|X_i-Y_i\right|, \qquad (1.23)$$

где $X=(X_1,\dots,X_m)$ и $Y=(Y_1,\dots,Y_m)$ – объекты, между которыми определяется расстояние.

1.9.10. Расстояние Чебышева

Это расстояние может оказаться полезным, когда желают определить два объекта как «различные», если они различаются по какой-либо одной координате (каким-либо одним измерением). Расстояние Чебышева вычисляется по формуле

$$d(X,Y)=\max\left|X_i-Y_i\right|. \qquad (1.24)$$

1.10. Меры сходства

Понятием, противоположным расстоянию между объектами X_i и X_j, является сходство.

Неотрицательная вещественнозначная функция $s(X_i,Y_j)=s_{ij}$ называется мерой сходства, если:

1) $0\le s(X_i,X_j)\le 1$ для всех $X_i\ne X_j$;
2) $s(X_i,X_i)=1$;
3) $s(X_i,X_j)=s(X_j,X_i)$.

Сходства между парами векторов $s(X_i,X_j)$ могут быть представлены в виде симметричной матрицы сходства:

$$S=\left\{\begin{matrix} 1 & s_{12} & \cdots & s_{1n} \\ s_{21} & 1 & \cdots & s_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ s_{n1} & s_{n2} & \cdots & 1 \end{matrix}\right\}, \qquad (1.25)$$

причем диагональные элементы матрицы s_{ii}=1 для $i=\overline{1,n}$.

1.11. Типы расстояний

При проведении кластерного анализа обычно определяют расстояние на множестве объектов, а алгоритмы кластерного анализа формулируют в терминах этих расстояний.

Выделяют типы расстояний:

- между объектами;
- классами;
- объектом и классом.

1.11.1. Расстояние между объектами

Каждый из представленных выше типов расстояний имеет свою специфику и соответственно способы определения этих расстояний. Так, например, расстояние между объектами можно интерпретировать как расстояние между парой точек в n-мерном евклидовом пространстве. Данный тип расстояния наиболее прост и является частным случаем расстояния между классами, содержащими лишь по одному объекту, и между объектом и классом, также содержащим единственный объект.

1.11.2. Расстояние между классами

Для определения расстояния между парой кластеров могут быть сформулированы различные методы, определяемые на основе расстояний между объектами:

- расстояние между ближайшими соседями – ближайшими объектами кластеров;
- расстояние между самыми далекими соседями;
- среднее расстояние между кластерами;
- расстояние между центрами кластеров или центроидный метод (центр объединенного кластера вычисляется как среднее центров объединяемых кластеров, без учета их объема);
- среднее расстояние между всеми объектами пары кластеров с учетом расстояний внутри кластеров;
- метод медиан – тот же центроидный метод, но центр объединенного кластера вычисляется как среднее всех наблюдений;
- метод Варда (в качестве расстояния между кластерами берется прирост суммы квадратов расстояний объектов до центров кластеров, получаемый в результате их объединения).

Рассмотрим наиболее распространенные меры близости и расстояния, характеризующие взаимное расположение отдельных групп объектов, изображенных на рис. 1.8.

Меры близости отличаются от расстояний тем, что они тем больше, чем более похожи объекты.

Пусть имеется два класса объектов $X = \{X_1, X_2, ..., X_i, ..., X_{n1}\}$ и $Y = \{Y_1, Y_2, ..., Y_j, ..., Y_{n2}\}$.

Обозначим через $D = \{d(X_i, Y_j), i = \overline{1, n_1}, j = \overline{1, n_2}\}$ множество всех расстояний.

Величина $D_1(X, Y) = \min d(X_i, Y_j)$, $i = \overline{1, n_1}, j = \overline{1, n_2}$ называется *минимальным локальным расстоянием* между кластерами X и Y, или расстоянием между ближайшими соседями (ближайшими объектами кластеров).

Величина $D_2(X, Y) = \max d(X_i, Y_j)$, $i = \overline{1, n_1}, j = \overline{1, n_2}$ называется *максимальным локальным расстоянием* между кластерами X и Y, или расстоянием между самыми далекими соседями (отдаленными объектами кластеров).

Величина $D_3(X, Y) = d(\overline{X}, \overline{Y})$, где $\overline{X}$ и $\overline{Y}$ – центры классов X и Y соответственно, называется *центроидным расстоянием* между кластерами X и Y, или расстоянием между центрами кластеров.

Величина $D_4(X,Y)=\sum_{j=1}^{n2}\sum_{i=1}^{n1}d(X_i,Y_j)/n_1n_2$ называется *средним расстоянием между кластерами* X *и* Y.

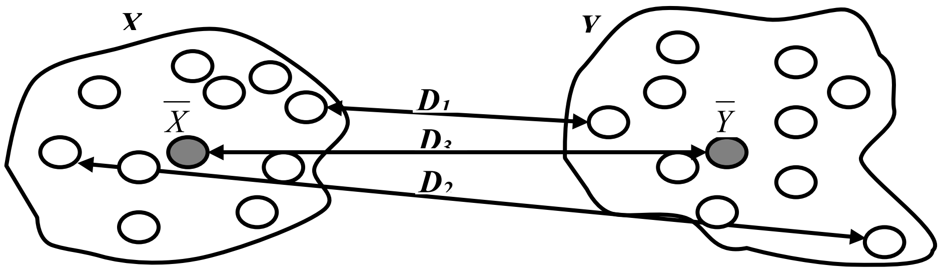


Рис. 1.8. Различные способы определения расстояния между кластерами X и Y:
D₁ – по ближайшим объектам; *D₂* – по самым далеким объектам;
D₃ – по центрам тяжести

Выбор той или иной меры расстояния между кластерами влияет, главным образом, на вид выделяемых алгоритмами кластерного анализа геометрических группировок объектов в пространстве признаков. Так, алгоритмы, основанные на расстоянии ближайшего соседа, хорошо работают в случае группировок, имеющих сложную, в частности цепочечную структуру. Расстояние дальнего соседа применяется, когда искомые группировки образуют в пространстве признаков шаровидные облака. И промежуточное место занимают алгоритмы, использующие расстояния центров тяжести и средней связи, которые лучше всего работают в случае группировок эллипсоидной формы.

1.11.3. Расстояние между объектом и классом

Для определения расстояния между объектом и кластером могут быть сформулированы различные методы.

Особый интерес представляет двухгрупповой метод, предложенный Сокалом и Миченером [8]. В данном методе связь между объектом S и классом K выражается в виде среднего коэффициента сходства между объектом Q и всеми объектами, входящими в класс K.

Обозначим объекты, входящие к класс K через $X_1, X_2, ..., X_{n_K}$, а через $\overline{X}$ – центр класса K. Тогда, если средний коэффициент сходства выразить через евклидово расстояние, среднее расстояние D_{SK} между объектом $S \notin K$ и всеми объектами из K

$$D_{SK}=\frac{1}{n_K}\sum_{j=1}^{n_K}(X_j-S)^T(X_j-S). \qquad (1.26)$$

Далее

$$\begin{aligned} D_{SK} &= \frac{1}{n_K}\sum_{j=1}^{n_K}(X_j-\overline{X}+\overline{X}-S)^T(X_j-\overline{X}+\overline{X}-S)= \\ &=\frac{1}{n_K}\sum_{j=1}^{n_K}(X_j-\overline{X})^T(X_j-\overline{X})+(\overline{X}-S)^T(\overline{X}-S). \end{aligned} \qquad (1.27)$$

Первое слагаемое правой части – внутригрупповая дисперсия объектов из K, второе слагаемое представляет собой квадрат расстояния между объектом S и центром класса K. Процедура последовательной кластеризации заключается в том, что объект $S \notin K$, для которого D_{SK} минимально, присоединяется к классу K. Из (1.24) легко видеть, что среднее расстояние D_{QK} минимизирует расстояние между объектом S и центром класса K, если два класса имеют сравнимые дисперсии. Для классов с различными дисперсиями объединение происходит в первую очередь с кластером меньшей дисперсии.

Данный способ определения расстояния позволяет учесть не только «отдаленность» объекта от класса, но и близость объектов внутри класса, что позволяет разрешить проблему «равенства расстояний» и минимизирует вероятность «сближения» классов после классификации данного объекта.

Если учесть, что каждый объект обладает, как правило, набором различных характеристик (признаков), то выражение (1.26) можно записать в виде

$$D_{SK} = \frac{1}{n_K} \sum_{j=1}^{n_K} \sum_{i=1}^{p_K} (x_{ij} - y_i)^2 , \tag{1.28}$$

где p_K – количество признаков объекта;

x_{ij} – i-й признак j-го объекта, принадлежащего классу K;

y_i – i-й признак классифицируемого объекта.

Учитывая (1.26), выражение (1.27) будет иметь вид

$$\begin{aligned} D_{SK} &= \frac{1}{n_K} \sum_{j=1}^{n_K} \sum_{i=1}^{p_K} (x_{ij} - \overline{X} + \overline{X} - y_i)^2 = \\ &= \frac{1}{n_K} \sum_{j=1}^{n_K} \sum_{i=1}^{p_K} (x_{ij} - \overline{X})^2 + (x_{ij} - y_i)^2. \end{aligned} \tag{1.29}$$

1.12. Виды алгоритмов кластерного анализа

Алгоритмы кластерного анализа отличаются большим разнообразием.

Это могут быть, например, алгоритмы, реализующие полный перебор сочетаний объектов или осуществляющие случайные разбиения множества объектов.

В то же время большинство таких алгоритмов состоит из двух этапов. На первом этапе задается начальное (возможно, искусственное или даже произвольное) разбиение множества объектов на классы и определяется некоторый математический критерий качества автоматической классификации. Затем, на втором этапе, объекты переносятся из класса в класс до тех пор, пока значение критерия не перестанет улучшаться.

Многообразие алгоритмов кластерного анализа обусловлено также множеством различных критериев, выражающих те или иные аспекты качества автоматического группирования. Функционалы качества и конкретные алгоритмы автоматической классификации достаточно полно и подробно рассмотрены в специальной литературе. Эти функционалы и алгоритмы характеризуются различной трудоемкостью и подчас требуют ресурсов высокопроизводительных компьютеров. Разнообразные процедуры кластерного анализа входят в состав практически всех современных пакетов прикладных программ для статистической обработки многомерных данных.

Поскольку методы кластерного анализа формализованы и дают легко воспроизводимые способы создания классификаций, они пользуются широкой популярностью.

Целью изучения кластерного анализа является выявление и диагностирование состояния объектов (процессов) на основе измерения нескольких непрерывных переменных.

Все методы классификации состояния процесса по сходству и различию соответствующих параметров имеют две главные особенности:

- каждый метод требует количественного определения, или меры, относительного сходства между состояниями процесса;
- при заданных количественных показателях сходства требуется алгоритм для вычисления коэффициентов сходства, чтобы обнаружить однородные группы или классы.

Для любого метода определения сходства при наличии многих переменных существует ряд различных алгоритмов классификации.

1.12.1. Метод построения эталонов

Для каждого класса $K_i, i=\overline{1,m}$ по обучающей выборке строится эталон E_i, имеющий значения признаков. Эталон – это усреднённый по обучающей выборке абстрактный объект. Абстрактным он называется потому, что может не совпадать не только ни с одним объектом обучающей выборки, но и ни с одним объектом генеральной совокупности.

Распознавание осуществляется следующим образом. На вход системы поступает объект S, принадлежность которого к тому или иному классу системе неизвестна. От этого объекта измеряются расстояния до эталонов всех классов $d(E_i, S), i=\overline{1,m}$, и система относит объект к тому классу, расстояние до эталона которого минимально $S \in K_i, i = \min\limits_{i} d(E_i, S)$. Расстояние измеряется в той метрике, которая введена для решения определённой задачи классификации.

1.12.2. Метод дробящихся эталонов

Этот метод отражает стремление к безошибочному распознаванию обучающей выборки, с использованием покрытий обучающей выборки каждого образа простыми фигурами, усложняющимися по мере необходимости [6].

Один из вариантов этого метода предусматривает использование в качестве покрывающих фигур набора гиперсфер. Для каждого из m классов строится сфера минимального радиуса, покрывающая все его обучающие реализации.

Сделать это можно так. Строится эталон каждого класса. Вычисляется расстояние от эталона до всех объектов данного класса, входящих в обучающую выборку. Выбирается максимальное из этих расстояний. Строится гиперсфера с центром в эталоне и радиусом. Она охватывает все объекты данного класса. Такая процедура проводится для всех классов (образов).

Значения радиусов этих сфер и расстояний между их центрами позволяют определить классы, сферы которых не пересекаются со сферами других классов. Такие сферы считаются эталонными (см. образ 1 на рис. 1.9), а их центры и радиусы запоминаются в качестве «эталонов первого поколения».

Если два образа пересекаются, но в области пересечения не оказалось ни одной реализации обучающей выборки, то такое пересечение считается фиктивным, центры и радиусы этих сфер также вносятся в список эталонов первого уровня. При этом область пересечения считается принадлежащей сфере с меньшим радиусом (сфера 2 на рис. 1.9).

Если в зоне пересечения оказались точки только одного образа, то эта зона считается принадлежащей этому образу (сфера 3 на рис. 1.9). Точка будет считаться относящейся к образу 4, если она попадает в сферу 4 и не попадает в сферу 3.

Если же область пересечения содержит точки разных образов, то для этих точек строятся «эталоны второго уровня» (сферы 4, 5 и 6 на рис. 1.9). Если и они пересекаются, то для точек из зоны пересечения строятся «эталоны третьего поколения». Процедура дробления эталонов продолжается до получения заданной надежности распознавания обучающей последовательности. Опыт показал, что даже в очень сложных случаях для хорошего распознавания обучающей выборки бывает достаточно в среднем не более трех уровней эталонов.

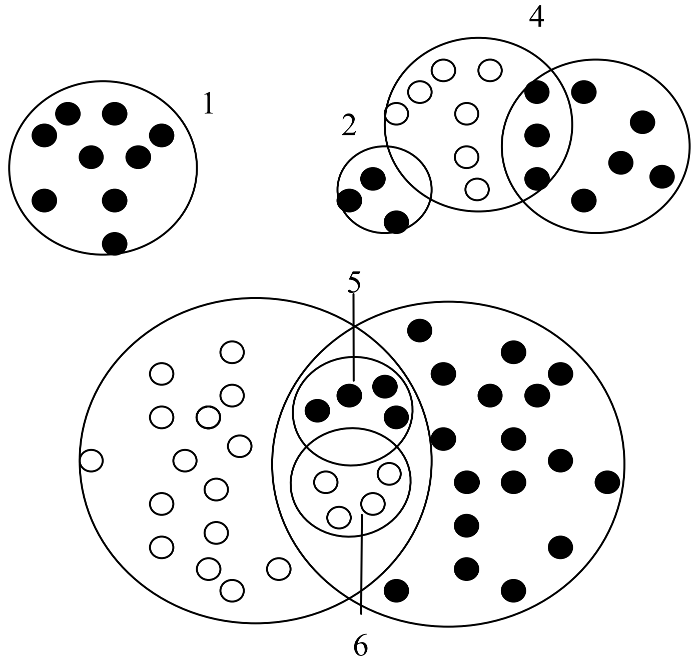


Рис. 1.9. Метод дробящихся эталонов

Распознавание осуществляется следующим образом. Определяется местонахождение объекта относительно гиперсфер первого уровня. При попадании объекта в гиперсферу, соответствующую одному и только одному образу, процедура распознавания прекращается. Если же объект оказался в области перекрытия гиперсфер, которая при обучении содержала объекты более чем одного образа, то переходим к гиперсферам второго уровня и проводим действия такие же, как для гиперсфер первого уровня. Этот процесс продолжается до тех пор, пока принадлежность неизвестного объекта тому или иному образу не определится однозначно. Правда, это событие может и не наступить. В частности, неизвестный объект может не попасть ни в одну из гиперсфер какого-либо уровня. В этих случаях учитель должен включить в решающие правила соответствующие действия. Например, система может либо отказаться от решения об однозначном отнесении объекта к какому-либо образу, либо использовать критерий минимума расстояния до эталонов данного или предшествующего уровня и т.п. Какой из этих приёмов эффективнее, сказать трудно, т.к. метод дробящихся эталонов носит в основном эмпирический характер.

1.12.3. Кластеризация полным перебором

Кластеризация полным перебором заключается в полном переборе всех возможных разбиений на кластеры и отыскании такого разбиения, которое ведет к оптимальному значению целевой функции. На практике такие алгоритмы применяются крайне редко или вообще не применяются за исключением тех случаев, когда число объектов n и число кластеров m невелико.

Число способов разбиения n объектов на m классы определяется как

$$S(n,m)=\frac{1}{m!}\sum_{j=0}^{m}C_m^j(-1)^j(m-j)^n\ , \qquad (1.30)$$

где C_m^j – число сочетаний j объектов из m.

При полном переборе целевую функцию W придется вычислять S раз, а затем искать такое разбиение, которое приводит к минимуму W.

Несомненным достоинством данного алгоритма является то, что в конце работы он дает оптимальный результат. Однако при большом количестве классов и объектов затраты вычислительных и временных ресурсов настолько велики, что алгоритм становится непригодным.

1.12.4. Классический последовательный алгоритм

Классический последовательный алгоритм относится к иерархическим алгоритмам группирования, основанным на понятиях «сходство» и «расстояние».

Основная идея всех иерархических алгоритмов заключается в следующем. На первом шаге каждый из n объектов считается отдельным кластером. Затем два наиболее близких объекта (кластера) объединяются в один новый кластер, и число кластеров становится равным n-1. Эта процедура повторяется, пока все объекты не объединятся в один кластер.

Результаты работы иерархических алгоритмов обычно оформляются в виде дендрограммы (рис. 1.10), в которой приведены номера объединяемых объектов и значения меры сходства, при которых эти объекты были объединены.

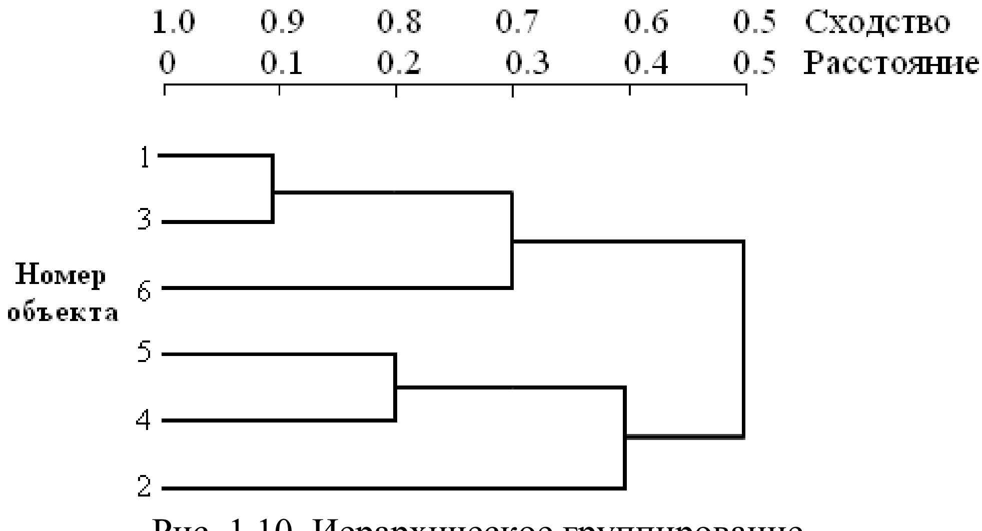


Рис. 1.10. Иерархическое группирование

По такой дендрограмме можно определить:

- составы кластеров при заданном уровне сходства;
- составы кластеров при их заданном количестве.

1.13. Модель этапов системного подхода

При использовании моделей для исследования систем возникают два характерных случая:

1. Модель достаточно точно (адекватно) отображает поведение системы. Система является простой по отношению к модели.

2. Модель неадекватна относительно своего прототипа. В этом случае считается, что система является сложной по отношению к модели.

Используя принцип «от простого к сложному», введем понятие модели уровней описания системы.

Самым простым представлением модели является уровень «черного ящика». На уровне «черного ящика» изучаются свойства входов, выходов, а также процесса, связывающего вход и выход системы.

Следующий уровень модели – это перечисление состава входов, выходов и элементов систем.

На третьем уровне осуществляется дальнейшая детализация модели путем рассмотрения структуры системы.

Уровни описания – «черный ящик», состав, структура – применимы как к системе в целом, так и к ее элементам. Эффективность использования моделей для целей системного проектирования в значительной мере зависит от развитости языка их описания.

Количественно все компоненты систем могут быть охарактеризованы как монокомпоненты (одно свойство, одно отношение, один элемент) и поликомпоненты (много свойств, отношений, элементов).

По составу компоненты систем оцениваются как статические (находящиеся в состоянии относительного покоя) и динамические (изменяющиеся).

Динамические компоненты подразделяются на функционирующие, т.е. изменение компонентов системы не приводит к смене качества соответствующего компонента, и развивающиеся (изменение приводит к смене качества компонента).

Структурно, т.е. по характеру отношений с другими явлениями, компоненты систем оценивают как детерминистские и случайные; простые и сложные.

Система является детерминистской, если ее поведение необходимо обусловлено конечным множеством входящих в нее элементов и отношений между ними, т.е. поведение детерминистских систем полностью объяснимо и предсказуемо на основе информации об указанном конечном множестве.

Система является случайной, если в обусловленности ее поведения участвуют объекты, не входящие в конечное множество составляющих данной системы. Понятия простой и сложной системы введены выше.

С учетом вышеизложенного могут быть предложены количественные и качественные основания классификации свойств элементов и отношений (табл. 1).

Таблица 1. Классификатор свойств, элементов и отношений системы

Основание классификации	Категориальные характеристики и их шифр
Количество	1а – моно 1б – поли
Поведение и устойчивость во времени	2а – статические 2б – динамические 2б1 – функционирующие 2б2 – развивающиеся
Обусловленность отношений с другими явлениями	3а – детерминистские 3б – случайные
Предсказуемость поведения (отношение к модели)	4а – простые 4б – сложные

В соответствии с таблицей система, классифицированная как (P_0): 1а, 2б1, 3а, 4а – это многофункционирующая, детерминистская, простая система.

Использование моделей глубины (уровней) описания позволяет осуществить дальнейшую их формализацию, т.е. перейти от содержательного описания системы к математическому. Для этих целей следует использовать характеристики системы, т.е. набор выявленных в результате описания признаков системы.

Число, выражающее отношение между данной характеристикой и некоторым избранным эталоном, называется параметром системы.

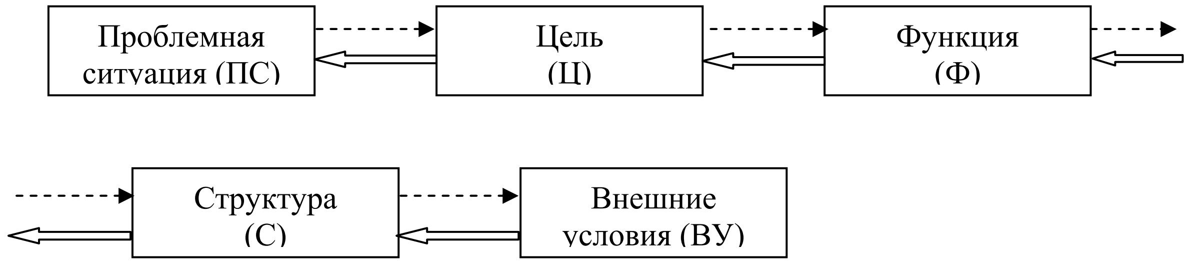
В этом случае качественные характеристики можно оценивать параметрами. Параметры можно установить, либо пользуясь двоичной системой (0,1), либо, например, ранговыми шкалами.

Совокупность параметров позволяет изобразить состояние системы точкой некоторого фазового пространства.

Изменение состояния системы отображается некоторой фазовой траекторией в этом фазовом пространстве. В свою очередь и цель системы может быть представлена некоторой областью фазового пространства, движение к которой может происходить по различным траекториям.

Способ определения предпочтительности двух состояний системы относительно целевого принято называть критерием качества состояния системы.

Логическая последовательность проведения анализа и синтеза системы представляет собой модель этапов системного подхода (рис. 1.11).



– реализация систем

– исследование или проектирование

Рис. 1.11. Модель этапов системного подхода

Однако для эффективного использования этой модели (этапов системного подхода) необходимо каждый элемент, представленный на уровне «черного ящика» классифицировать по составу, создать язык описания.

В качестве первого этапа системной деятельности должен быть получен ответ на вопрос: зачем нужны имеющаяся и создаваемая вновь системы? Для этого требуется выявить проблемную ситуацию. Применение модели глубины описания (см. табл. 1) позволяет образовать классификацию проблемных ситуаций, которая целиком совпадает с понятиями базовой классификации. Но некоторые понятия требуют дальнейшей детализации.

Во-первых, в количественном отношении проблемные ситуации редко встречаются каждая в отдельности, а это, как правило, взаимосвязанное множество или комплексная проблемная ситуация.

Во-вторых, относительно свойств развития и функционирования хорошо известно, что проблемная ситуация в своем развитии проходит три стадии: скрытое, явное развитие и превращение в постоянно действующий отрицательный фактор.

Своевременное выявление назревающей проблемы является одной из актуальнейших задач.

В-третьих, относительно свойств сложности проблемную ситуацию необходимо вычленить и сформулировать, а это далеко не простая задача, т.к. она не всегда повторяется, в ней может быть неизвестен состав существенных факторов, а также связи между факторами, многие характеристики не имеют количественного измерения и т.д. Справедливо считается, что если правильно сформулировать проблему, значит наполовину ее решить. В этом смысле следует ввести в классификатор понятие стандартной и нестандартной, а также структуризованной (с градациями «слабо-хорошо») и неструктуризованной проблемной ситуации.

Хорошо структуризованные ситуации имеют явно выраженные количественные связи между существенными параметрами. Для разрешения таких ситуаций широко используются математические методы обоснования решений, развиваемых в исследовании операций.

Слабоструктуризованные ситуации характеризуются наличием качественных и количественных элементов, причем качественные элементы доминируют.

Для разрешения неструктуризованных проблемных ситуаций применяется методология системного подхода.

Неструктуризованные ситуации содержат лишь описание важнейших признаков и характеристик, взаимосвязи между которыми еще надлежит установить, поэтому для разрешения таких ситуаций широко применяются эвристические методы.

Классификатор свойств проблемных ситуаций с учетом их специфики приведен в табл. 2.

Таблица 2. Классификатор свойств проблемных ситуаций

Шифр	Наименование
1	Количество
1а 1б 1б1	Одиночная (моно) Множественная (поли) Комплексная
2	Поведение и устойчивость во времени
2а 2б 2б1 2б2 2б2а	Статическая Динамическая Сложившаяся (функционирующая) Развивающаяся Назревающая
3	Обусловленность (по составу)
3а 3а1 3б 3б1	Детерминистская Стандартная Случайная Нестандартная
4	Предсказуемость поведения
4а 4а1 4б 4б1 4б2	Простая Структуризованная Сложная Неструктуризованная Слабоструктуризованная

Рассматривая проблемные ситуации по составу, необходимо выделить возможности, предоставляемые внешней средой системе в виде ресурсов и ограничений как в настоящее время, так и в прогнозируемом будущем, а также потребности внешней среды, которые в настоящее время оцениваются нормативами, планами и фактически сложившимся уровнем потребления, а в будущем – перспективными (прогнозными) нормативами и планами.

Различие между потребностями и возможностями в настоящем и будущем отражают структуру проблемной ситуации.

С учетом всего вышеизложенного проблемная ситуация может быть определена, например, как (P_1): 1б1; 2б2а; 3а1; 4а1 – комплексная, назревающая, стандартная, структуризованная.

Второй этап системного подхода – выявление цели.

Этот этап состоит в формировании ответа на вопрос: что должна выполнять система для ликвидации проблемной ситуации, т.е. какова цель системы?

Цель, представленная в виде совокупности параметров, задается на информационный вход системы. Система преобразует исходные ресурсы таким образом, чтобы полученный конечный продукт соответствовал заданной цели.

Возможная классификация целей системы приведена в табл. 3.

Таблица 3. Классификация целей системы

Шифр	Наименование
1	Количество
1а	Моно (одноцелевого назначения)
1а1	Субстанционального состава
1а1а	Овеществленная
1а1а1	Материальная
1а1а2	Энергетическая
1а1а3	Информационная
1а1б	Неовеществленная
1а1б1	Услуга *
1а2	Функционального состава
1а2а	Элемент
1а2б	Отношение
1б	Поли
1б1	Комплексная
2	Поведение и устойчивость во времени
2а	Статическая
2б	Динамическая
2б1	Функционирующая
2б2	Развивающаяся
3	Обусловленность отношений
3а	Детерминистская
3а1	Стандартная
3б	Случайная
4	Предсказуемость поведения
4а	Простая
4а1	Структуризованная
4б	Сложная
4б1	Неструктуризованная
4б2	Слабоструктуризованная

* Под услугой понимается конечный продукт деятельности, полученный без изменения исходных свойств предметов труда (перевозки, ремонт, техническое обслуживание и т.п.).

Пример

Цель классифицируется как (P_2): 1а1а3;2а;3а1;4б1 – информационная, статическая, стандартная, неструктуризованная.

На третьем этапе рассматриваются функции системы, т.е. проявление избранного свойства в динамике, приводящее к достижению цели. Иначе говоря, необходимо дать ответ на вопрос: как должна работать система, чтобы достигнуть цели? В общем случае для реализации конечного продукта во внешней среде система должна последовательно реализовать полный жизненный цикл, в состав которого входят:

выявление потребности в конечном продукте;

производство конечного продукта;

обеспечение потребления конечного продукта.

Классификация (один из вариантов) свойств функций приведена в табл. 4.

Таблица 4. Классификатор свойств функции

Шифр	Наименование
1	Количество
1а 1а1 1а2 1а3 1б 1б1	Моно Выявление потребности в конечном продукте Производство конечного продукта Потребление конечного продукта Поли Комплексная
2	Поведение и устойчивость во времени
2а 2б 2б1	---------- Динамическая Постоянная функция
3	Обусловленность отношений
3а 3а1 3б 3б1	Детерминистская Стандартная Случайная Нестандартная
4	Предсказуемость поведения
4а 4а1 4б 4б1 4б2	Простая Структуризованная Сложная Неструктуризованная Слабоструктуризованная

Пример

Внешняя функция системы классифицируется как (P_3): 1а2;2б2;3б1;4б – развитие производства конечного продукта, нестандартная, сложная.

На четвертом этапе предполагается проникновение внутрь «черного ящика», образующего систему и выявление таких элементов и отношений между ними, которые обеспечивают целенаправленное функционирование системы.

Элементы любого содержания (физические тела, идеи, люди и пр.), необходимые для обеспечения функции, называются частями системы. Совокупность частей образует элементный состав системы. Упорядоченное множество отношений между частями, существенных по отношению к цели, которое необходимо для обеспечения функции, образует структуру. Структура системы имеет два аспекта:

статический – это отношения между элементами (пространственное расположение, различие в каких-либо качествах, иерархия и т.д.);

динамический – взаимодействие частей друг с другом во времени.

Классификатор свойств структур представлен в табл. 5.

Конструкция системы в самом общем смысле представляет собой структурно упорядоченный состав, который необходимо и достаточно обеспечивает функцию системы.

Таблица 5. Классификатор свойств структур

Шифр	Наименование
1	Количество
1а 1б	Моно Поли
2	Поведение и устойчивость во времени
2а 2б 2б1 2б2	Статическая Динамическая Функционирующая Развивающаяся
3	Обусловленность отношений
3а 3а1 3а2 3б 3б1	Детерминистская Постоянная Временная Случайная Нестандартная
4	Предсказуемость поведения
4а 4б	Простая Сложная

Пример

Структура системы классифицируется как (P_4): 1б;2а;3а2 – полиструктура статическая, временная.

На пятом этапе учитываются внешние условия (ВУ), действующие на входы функционирующей системы. Система является открытой для вещества, энергии и информации, которые воздействуют на ее вход и выступают либо как ограничения, либо как ресурсы.

Классификатор свойств внешних условий приведен в табл. 6 и практически совпадает с классификатором для свойств конечных продуктов, представленных в табл. 3.

Таблица 6. Классификатор свойств внешних условий

Шифр	Наименование
1	Количество
1а 1а1 1а1а 1а1а1 1а1а2 1а1а3 1а1б 1а1б1 1а2 1а2а 1а2б 1б	Моно (одноцелевого назначения) Субстанционального состава Овеществленные Материальные ресурсы Энергетические ресурсы Информационные ресурсы Неовеществленные Услуга Функционального состава Элемент Отношение Поли
2	Поведение и устойчивость во времени
2а 2б 2б1 2б2	Статическая Динамическая Функционирующая Развивающаяся
3	Обусловленность отношений
3а 3а1 3б	Детерминистская Стандартная Случайная
4	Предсказуемость поведения
4а 4а1 4б 4б1 4б2	Простая Структуризованная Сложная Неструктуризованная Слабоструктуризованная

Пример

Внешние условия классифицируются как (P_5): 1а1а2;2б1;4б – энергетические ресурсы функционирующие, сложные.

Если внешняя среда не обеспечивает возможности выбора удовлетворяющих средств (необходимых ресурсов), то создание системы невозможно. Системный подход в этом случае обосновывает отказ от нереальной поставленной цели.

1.14. Этапы системного подхода

Этапы системного подхода содержат основные характеристики системных объектов, позволяют сформулировать одно из определений понятия системы. Система – это целенаправленная функционирующая конструкция, способная к разрешению проблемной ситуации при определенных внешних условиях.

На основе рассмотренной модели можно предложить укрупненный алгоритм, позволяющий осуществить анализ действующих либо синтез вновь создаваемых систем.

Этот алгоритм условно называется алгоритмом системного подхода. На рис. 1.12 представлена блок-схема алгоритма этапов системного подхода.

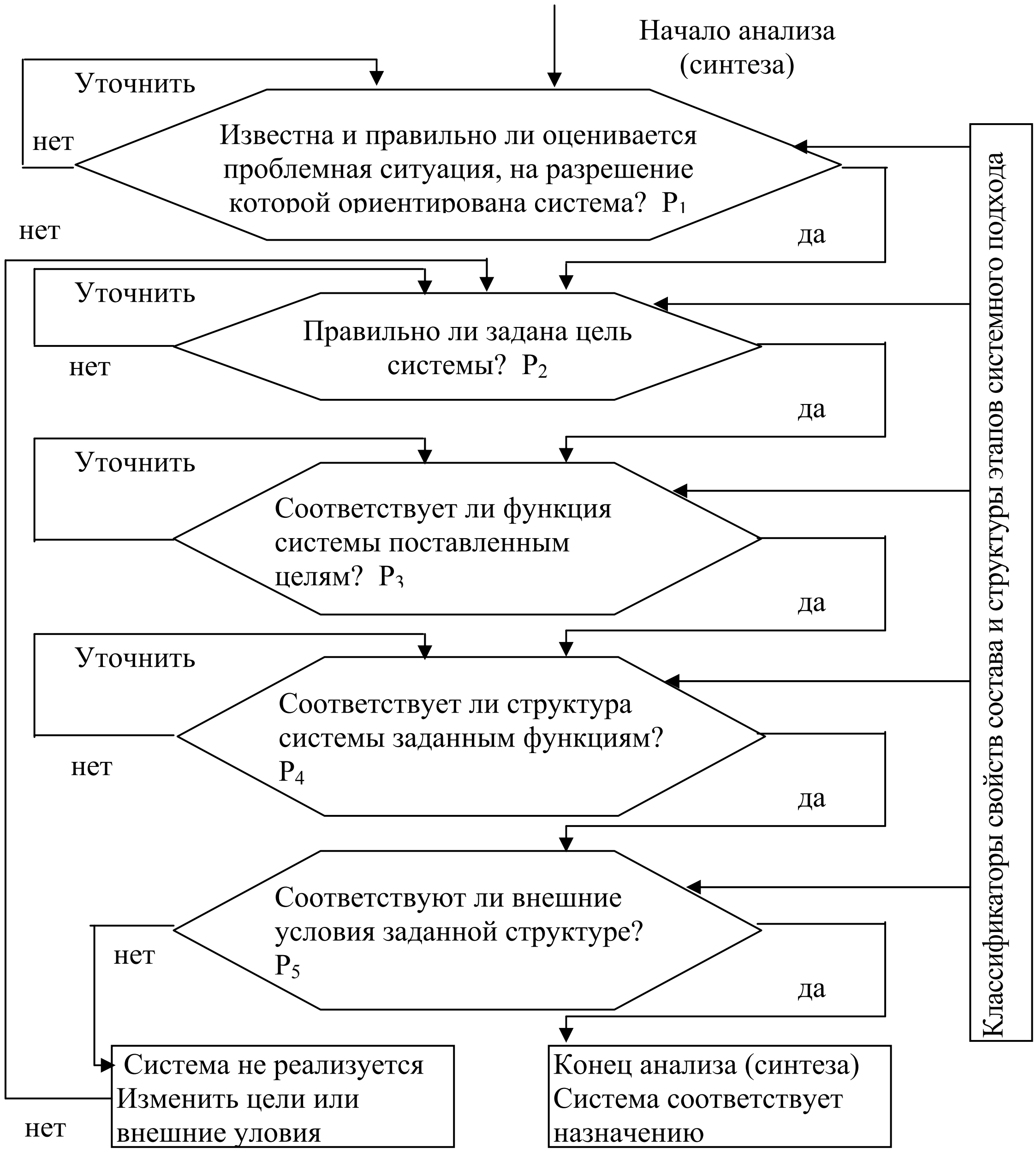


Рис. 1.12. Блок-схема алгоритма этапов системного подхода

Для изображения алгоритма используются логические схемы алгоритмов (ЛСА). При применении ЛСА вершинам графа, изображающего последовательность действий, ставится в соответствие некоторый оператор, а дугам – последовательность действий над операторами.

Операторы делятся:

на функциональные, обозначаемые A_i, $i=1,n$ и однозначно обеспечивающие переход к следующему оператору;

логические, обозначаемые P_j, $j=1,m$ и проверяющие некоторые условия.

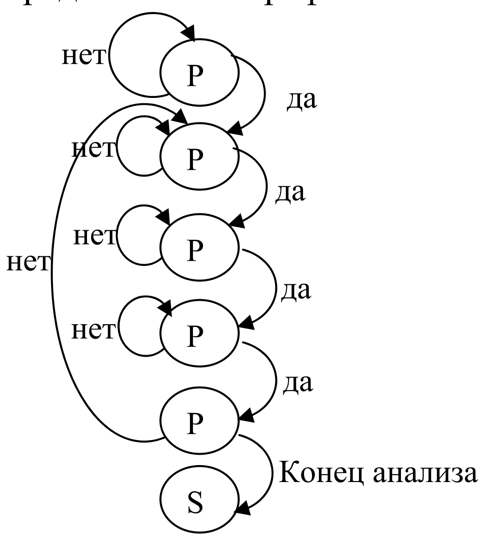
Если условие выполнено (да), то следующим выполняемым оператором будет оператор, записанный справа от P_j $j=1,m$.

Если же условие не выполняется (нет), то необходимо выполнить оператор, к которому идет стрелка от оператора P_j $j=1,m$. При этом справа от оператора P_j $j=1,m$ ставится стрелка вверх, над которой пишется номер этого оператора P_j, а перед оператором $P_к$, к которому должна идти стрелка, ставится стрелка концом вниз, отмеченная тем же номером.

Для большей компактности записи вводится оператор безусловного перехода, который обозначается Ω_ρ, $\rho=1,e$. После выполнений этого оператора всегда происходит переход по стрелке, находящейся справа от него. Конечный оператор обозначается буквой S. Блок-схема алгоритма этапов системного подхода с использованием операторов ЛСА (логических схем алгоритмов) будет иметь вид

$$\underset{1}{\downarrow} P_1 \overset{1}{\uparrow} \underset{2,5}{\downarrow} P_2 \overset{2}{\uparrow} \underset{3}{\downarrow} P_3 \overset{3}{\uparrow} \underset{4}{\downarrow} P_4 \overset{4}{\uparrow} P_5 \overset{5}{\uparrow} S . \qquad (1.31)$$

В традиционном представлении граф-схема имеет вид



Рассмотрим условный пример, где каждому из операторов P_1, P_2, P_3, P_4, P_5 присвоены соответствующие свойства из таблиц 2 – 6.

Таблица 7. Сводная таблица классификаторов

ПС	P_1	1б1	2б1	3б1	4а1
Ц	P_2	1а1а3	2а	3а	4б1
Ф	P_3	1а2	2б2	-	4б
С	P_4	1б	2а	3а1	-
ВУ	P_5	1а1а2	2б1	-	4б

Анализ элементов табл. 7 позволяет выявить противоречия и несоответствия в изучаемой системе. Так, в примере имеются расхождения между нестандартной проблемной ситуацией (ПС-3б1) и строго заданной целью (Ц-3а), характером развивающихся функций (Ф-2б2 (разв.)) и статической структурой (С-2а), отсутствуют некоторые характеристики (Ф-3, С-4, ВУ-3). Для ликвидации выявленных противоречий необходимо применить циклические процедуры алгоритма этапов системного подхода. Уровень абстракций примера можно снизить путем введения значений параметров свойств, а также за счет дальнейшей конкретизации классификаторов свойств и перехода к описанию состава элементов. Для этого необходима конкретизация класса изучаемых систем.

1.15. Основные информационные модели организационных систем

Применение системного подхода для практических целей, конечно, не может быть ограничено предложенными выше моделями.

В зависимости от конкретной сферы приложения модели системного подхода развиваются и дополняются.

Системы, в которых важнейшим элементом являются люди, объединенные в коллективы и преследующие определенные цели, называются системами организационного типа.

Специфика систем организационного типа состоит в том, что в них наряду с наличием внешней цели системы ее составные части имеют собственные цели, которые могут частично или даже полностью не совпадать и даже противоречить основной цели системы.

С учетом специфической особенности система организационного типа может быть определена как система, назначением которой является согласование действий целеустремленных (социальных групп и личностей) и нецелеустремленных элементов (средств и предметов деятельности) с глобальной целью получения основного конечного продукта. Эта специфическая функция системы организационного типа предопределяет основные свойства структуры такой системы: неформальность, саморазвитие, самоуправляемость.

В отличие от структурно детерминированной системы, в которой при заданном составе элементов ее функционирование полностью определяется структурой (т.к. элементы могут находиться только в двух состояниях: работает, не работает), в системе организационного типа функционирование зависит не только от структуры, но и от поведения ее элементов. Система становится неформальной.

Особенности организационных систем состоят:

в постоянной способности к развитию;

необходимости прогнозирования возможных направлений развития частей системы;

обеспечении оптимального соотношения между развитием частей системы и системы в целом;

определении оптимального соотношения между развитием и функционированием системы.

Возникающая постоянная потребность в согласовании интересов элементов системы обеспечивается самоуправлением в организационной системе.

Для этого необходимо создать (выделить) в системе управляющую и управляемую подсистему. Иначе говоря, надо выделить объект управления (управляемая часть) и субъект управления (управляющая часть).

Под управлением понимается целенаправленное воздействие субъекта управления на объект посредством решений, обеспечивающих получение конечного продукта (достижение цели) системы.

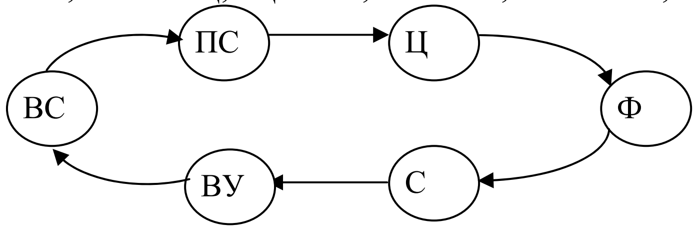
Управление должно обеспечить либо удержание системы у достигнутой цели (функционирование), либо перевод системы в новое состояние, к новой цели (развитие).

1.16. Синтез универсального алгоритма системной деятельности

Выполнение каждого этапа, устанавливающего формальный порядок действий при анализе и синтезе систем, требует осуществления процедуры выбора и принятия промежуточного решения. В этой процедуре каждый предыдущий элемент модели этапов системного подхода логично рассматривать как цель анализа или синтеза, а каждый последующий – как средство ее реализации.

Такой порядок действий можно представить в виде последовательности цепочек:

ВС → ПС; ПС → Ц; Ц → Ф; Ф → С; С → ВУ; ВУ → ВС.



Здесь ВС – внешняя среда;
ПС – проблемная ситуация;
Ф – функция;
С – структура;
ВУ – внешние условия;
Ц – цель.

Если применить при выборе варианта для каждого элемента общую схему деятельности, то эта схема, с учетом специфики решаемых задач, должна иметь:

на входе – цели исследования, альтернативы (способы) их реализации;

на выходе – выбранный вариант;

в составе структуры – группу исследователей, исходную информацию, модели и критерии для оценки последствий выбранного варианта и вероятности его достижения.

Схема выбора вариантов представлена на рис. 1.13.

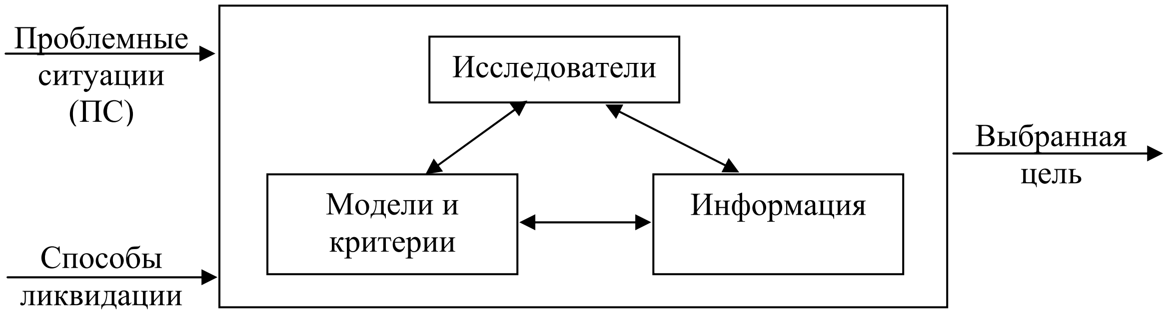


Рис. 1.13. Схема выбора вариантов

Схема соответствует модели процесса анализа, представленной Э. Квейдом, которая включает в себя [3]:

1) цель (или цели) лица, принимающего решение о выборе;
2) альтернативы (способы) достижения цели;
3) затраты (ресурсы), необходимые для реализации выбранного способа достижения цели;
4) модель (или модели), необходимую для того, чтобы оценить затраты для каждого варианта или степень достижения желаемого результата;
5) критерий (правило), в соответствии с которым альтернативы располагаются в порядке их предпочтительности.

2. Основы исследования операций

Потребности практики управления сложными системами вызвали к жизни специальные научные методы их анализа, которые удобно объединять под названием «исследование операций» (ИСО), – это применение математических, количественных методов для обоснования решений по управлению во всех областях целенаправленной человеческой деятельности.

2.1. Основные понятия и принципы исследования операций

Операцией называется всякое мероприятие (система воздействий), объединённое единым замыслом и направленное к достижению какой-то цели. Операция – всегда управляемое мероприятие, т.е. параметры (факторы), влияющие на исход операции должны быть контролируемыми и управляемыми.

Всякий определённый набор численных значений факторов операции называется решением, которое может быть удачным и неудачным. Оптимальными называют решения, по тем или иным причинам предпочтительные перед другими.

Цель исследования операций [9] – количественное обоснование принимаемых решений по управлению системой. Следует особо отметить, что само принятие решения выходит за рамки исследования операций и относится к компетенции ответственного лица, которому предоставлено право окончательного выбора и на которого возложена ответственность за выбор того или иного решения.

Параметры $x_i, i=\overline{1,n}$, совокупность которых образует решение, называются элементами решения, т.е. $\bar{x}^T = (x_1, x_2, ..., x_n)$.

Элементы решения принадлежат некоторому множеству X – множеству возможных решений, т.е. $\bar{x}^T \in X$.

С точки зрения математики, речь идет о том, как в множестве решений X выделить те решения $\bar{x}^{0T} = \left(x_1^0, x_2^0, ..., x_n^0\right)$, которые эффективнее других.

Но чтобы сравнивать между собой по эффективности разные решения, необходимо иметь какой-то количественный критерий, который называется показателем эффективности, критерием оптимальности, целевой функцией (ЦФ) $W(\bar{x})$. Показатель эффективности должен отражать цель операции, и в зависимости от цели необходимо либо максимизировать, либо минимизировать этот показатель.

Исследование операций немыслимо без математических моделей. Разработка модели операции – это искусство, доступное далеко не каждому исследователю.

2.1.1. Классификация моделей и методов исследования операций (ИСО)

Как прикладное направление кибернетики ИСО предназначено для количественного обоснования решений в любой области целенаправленной деятельности человека, и потому арсеналом его средств могут являться модели и количественные методы из любой области знаний, связанные единым целевым назначением, обусловленным решаемой проблемой. Поэтому ИСО не располагает каким-то специфическим собственным математическим аппаратом, несмотря на то, что можно называть целый ряд моделей и методов, появившихся впервые из практических потребностей, связанных именно с обоснованием решений в той или иной области деятельности человека (линейное программирование, теория игр и т.д.). Принятие решения – это, как правило, сложный акт, требующий мобилизации всего морально-психологического и интеллектуального потенциала ЛПР, при этом нервное напряжение тем выше, чем больше важность принимаемого решения и чем меньше время на его подготовку. За это время ЛПР с помощью системы поддержки решения (СПР), которой он располагает, должен сформулировать, оценить и сравнить ряд наиболее вероятных вариантов решения. Естественно, что роль моделей и методов ИСО будет зависеть от тех удобств и возможностей, которые они могут обеспечить ЛПР по обоснованию и выбору варианта решения. В связи с этим представляется целесообразным в качестве основного признака классификации моделей и методов ИСО выбрать степень свободы ЛПР по влиянию на параметры, определяющие ход операции достижения главных ее целей. По этому признаку все модели (задачи) ИСО можно разделить на три группы: оценочные, оптимизационные и игровые. Для записи общей формальной постановки задач каждой из этих групп введем некоторые обозначения.

Свое влияние на операцию ЛПР (исследователь) может осуществлять через некоторую группу управляемых параметров (вектор, матрица A*). На ход операции также будет оказывать влияние ряд факторов, обусловленных условиями ее протекания и внешними воздействиями. В моделях эти факторы могут быть представлены тремя группами параметров: ~a ~, ~b ~,~Y~ (векторы, матрицы).

Постоянные параметры ~a~ отражают фиксированные условия операций (характеристики средств, используемых в ней, и другие мало меняющиеся факторы).

Параметры ~b~ связаны со случайными факторами(воздействие внешней среды и др.).

Параметры Y – группа переменных, не контролируемых ЛПР. Область изменения параметров Y обычно известна.

Параметры X и Y можно назвать активными, параметры ~a~ и ~b~ – пассивными. Полагая, что группа параметров Y контролируется (управляется) некой другой стороной B (X-стороной A), указанные три группы моделей и задач ИСО можно записать в наиболее общих формальных постановках:

оценочные	оптимизационные	игровые
$F(X,Y,a,b)$	$F(X,a,b)\to\max_x$	$F_A(X,Y,a,b)\to\max_x$
$X,Y = const$	$X \in D_x$	$F_B(X,Y,a,b)\to\max_y$
a, b-параметры	a, b-параметры	$X \in D_x$, $Y \in D_y$
		a, b-параметры,

где X, Y – решения (альтернативы, управляемые параметры) соответственно сторон A и B; D_X, D_y – области допустимых значений соответственно для X и Y.

Поскольку в оценочных моделях векторы решений X, Y полагаются фиксированными, то они могут быть включены в группу пассивных (постоянных или случайных) параметров. По этой причине в записи ЦФ для оценочных моделей векторы X, Y могут быть опущены.

Примечание. В оптимизационных моделях записано, что A всегда стремится максимизировать критериальную функцию. Это не снижает общности моделей, т.к. при необходимости минимизировать ЦФ достаточно изменить её знак на противоположный, после чего задача снова решается на максимум.

В оценочных моделях ЦФ может быть векторной:

$$F(a,b)=(F_r\{a,b)),\ r=\overline{1;R},$$

т.е. для данной модели и ситуации, определяемой параметрами a и b, вычисляется множество частных критериев R, которые используются ЛПР при сравнительной оценке «просчитанных» вариантов решений. В ходе сравнительного анализа ЛПР выбирает рациональный (по его представлениям и с учетом всей дополнительной информации) вариант, который и закладывает в основу своего решения. Такой метод принятия решения называют вариантным.

С помощью оценочных методов ИСО получают только количественные оценки по ряду критериев (F_r) для сравниваемых вариантов, однако сам выбор множества альтернативных вариантов для их оценок и сравнения по многим критериям являются далеко не тривиальной задачей, требующей не только системного подхода, но также опыта и искусства. Обычно к оценочным методам ИСО прибегают в тех случаях, когда оцениваемая ситуация при ее формализации приводит к очень сложным моделям или вообще не поддается аналитическому описанию, так что некоторые ее элементы приходится просто имитировать с помощью специальных вспомогательных моделей с привлечением ЭВМ.

Если оценочные модели предоставляют информацию ЛПР на вопрос о том, какого исхода операции следует ожидать в данных условиях (a, b, X, Y), то оптимизационные модели позволяют ответить на вопрос: какое решение X в данных условиях (a, b, Y) следует принять, чтобы степень достижения цели была максимальной? Такой ответ более приемлем для ЛПР, т.к. избавляет его от необходимости сравнения целого множества вариантов. Это становится возможным благодаря тому, что в оптимизационных задачах используется только один критерий, записанный в виде целевой функции (ЦФ).

Игровые модели ИСО основаны на предположении, что каждая из сторон A и B имеет свой единый критерий, отражающий ее цель. Однако каждая из ЦФ зависит не только от своих управляемых параметров, но и от аналогичных параметров другой стороны.

Итак, в зависимости от степени свободы управляемых параметров X и Y каждая из трех групп моделей может быть преобразована в одну из предыдущих. Так, при X, Y = const получаем оценочную модель. Оптимизационную модель получаем в двух случаях: или $F_A = F_B$, или одна из групп управляемых параметров имеет постоянное значение. Первый случай означает, что цели обеих сторон (лиц, игроков) совпадают и, значит, имеем их коалицию как одно лицо. Второй случай соответствует выбору решения одной из сторон, исходя из своих интересов, при фиксированном решении другой.

Представляет интерес частный случай игровых моделей, наиболее разработанный в теоретическом плане. Если выигрыш стороны А равен проигрышу стороны В:

$$F_A\{X, Y, a, b) = -F_B(X, Y, a, b),$$

то отпадает необходимость в задании двух платежных функций.

Все три группы моделей обладают определенной общностью, поэтому дальнейшая классификация по другим признакам будет применима для всех групп. Все модели могут зависеть или не зависеть от случайных факторов (параметров b), в соответствии с чем их называют вероятностными (стохастическими) или детерминированными. Детерминированная модель обуславливает однозначную (детерминированную) связь между входными и выходными величинами. В вероятностных моделях эта однозначность может соблюдаться только между характеристиками этих величин (математическое ожидание, дисперсия и др.) или вообще быть неоднозначной. В последнем случае каждый вариант расчетов (при одной и той же исходной информации), как правило, будет иметь другие результаты. Модель, основанную на случайных результатах, называют статистической. Для обеспечения достоверности результата, получаемого статистическим методом, производят многократные расчеты (при одной и той же исходной информации) и полученные результаты усредняют. Действие фактора случайности при усреднении результатов уменьшается, при этом достижима любая точность. На указанном принципе основаны статистические модели и методы (статистическое моделирование).

2.2. Основы линейного программирования (ЛП)

2.2.1. Задачи линейного программирования

Самыми простыми задачами исследования операций являются такие, где выбор показателя эффективности W достаточно явно диктуется целью операции, а условия проведения операции известны заранее (детерминированный случай). В этом случае показатель эффективности зависит только от двух групп параметров: заданных условий α и элементов решения $\bar{x}$, т.е. $W = W(\alpha, \bar{x})$.

Если обозначить все возможные решения как $\bar{x}^T = (x_1, x_2, ..., x_n)$, то задача состоит в том, чтобы найти значения $x_1^0, x_2^0, ..., x_n^0$, которые бы обращали величину W в минимум или максимум.

Такие задачи отыскания значений параметров, обеспечивающих экстремум функции при наличии ограничений, наложенных на аргументы, носят общее название задач математического программирования [9].

Трудности, которые возникают при решении задач математического программирования, зависят:

от вида функциональной зависимости $W(\bar{x})$;

размерности n вектора решения $\bar{x}^T = (x_1, x_2, ..., x_n)$;

вида и количества ограничений, наложенных на элементы решения.

Среди задач математического программирования самыми простыми и более изученными являются задачи линейного программирования (ЛП), для которых характерны [9]:

показатель эффективности W, линейно зависимый от элементов решения $x_1, x_2, ..., x_n$;

ограничения, накладываемые на элементы решения и имеющие вид линейных равенств или неравенств относительно $x_1, x_2, ..., x_n$.

2.2.2. Примеры задач линейного программирования.

Задача о пищевом рационе

Имеется четыре вида продуктов $\Pi_1, \Pi_2, \Pi_3, \Pi_4$. Стоимость единицы каждого продукта равна соответственно $\tilde{N}_1, \tilde{N}_2, \tilde{N}_3, \tilde{N}_4$. Из этих продуктов требуется составить пищевой рацион, который должен содержать: белков не менее b_1 единиц; углеводов не менее b_2 единиц; жиров не менее b_3 единиц. Для продуктов $\Pi_1, \Pi_2, \Pi_3, \Pi_4$ содержание белков, углеводов, жиров (в единицах на единицу продукта) известно: $a_{ij}, i = \overline{1,4}, j = \overline{1,3}$. Первый индекс – номер продукта, второй – питательного вещества (белки, углеводы, жиры).

Требуется рассчитать пищевой рацион, т.е. определить, какое количество продуктов $\Pi_1, \Pi_2, \Pi_3, \Pi_4$ может обеспечить хотя бы необходимые потребности в белках, жирах, углеводах, но чтобы суммарная стоимость этого рациона была минимальной.

Составим математическую модель операции. Обозначим через x_1, x_2, x_3, x_4 искомые количества продуктов $\Pi_1, \Pi_2, \Pi_3, \Pi_4$ в рационе. Показатель эффективности составления рациона – стоимость рациона, которую необходимо минимизировать. Стоимость рациона линейно зависит от количества продуктов в рационе:

$$W = c_1 x_1 + c_2 x_2 + c_3 x_3 + c_4 x_4, \text{ или } W = \sum_{i=1}^{4} c_i x_i . \qquad (2.1)$$

Ограничения по белкам, углеводам и жирам запишутся соответственно:

$$\begin{aligned} &a_{11}x_1 + a_{21}x_2 + a_{31}x_3 + a_{41}x_4 \ge b_1, \\ &a_{12}x_1 + a_{22}x_2 + a_{32}x_3 + a_{42}x_4 \ge b_2, \\ &a_{13}x_1 + a_{23}x_2 + a_{33}x_3 + a_{43}x_4 \ge b_3. \end{aligned} \qquad (2.2)$$

Естественно, что искомые количества продуктов $x_i \ge 0, i = \overline{1,n}$.

Таким образом, поставленная задача расчета рациона сводится к следующему: определить такие неотрицательные значения переменных $x_1^0, x_2^0, x_3^0, x_4^0$, чтобы они удовлетворяли ограничениям-неравенствам и одновременно приводили в минимум линейную функцию этих переменных W. Это типичная задача ЛП.

Задача о планировании производства

Предприятие производит изделия трех видов: $И_1, И_2, И_3$. Запланирован выпуск продукции в количестве b_1 единиц изделий вида $И_1$, b_2 единиц изделий вида $И_2$, b_3 единиц изделий вида $И_3$. План выпуска изделий может быть перевыполнен, но с учетом предполагаемого спроса количества изделий не может быть более $\beta_1, \beta_2, \beta_3$ единиц.

На производство изделий используется 4 вида сырья: S_1, S_2, S_3, S_4 в количествах B_1, B_2, B_3, B_4 соответственно. На единицу *к-го* изделия расходуется a_{ij} единиц *j-го* сырья: $i = \overline{1,3}$, $j = \overline{1,4}$.

Прибыль, получаемая при реализации изделий вида $И_1$, равна C_1, вида $И_2$ – C_2, $И_3$ – C_3. Требуется определить, какое количество изделий $И_1, И_2, И_3$ необходимо произвести, чтобы обеспечить выполнение или перевыполнение плана (при отсутствии «затоваривания»), а суммарная прибыль при этом была максимальна.

Обозначим $x_1, x_2, x_3 \ge 0$ количества единиц изделий $И_1, И_2, И_3$ соответственно, которые необходимо выпустить предприятию.

Обязательность выполнения плана запишется в виде ограничений-неравенств

$$x_1 \ge b_1, x_2 \ge b_2, x_3 \ge b_3. \tag{2.3}$$

Предельно допустимое количество продукции: $x_1 \le \beta_1, x_2 \le \beta_2, x_3 \le \beta_3$, или

$$\begin{aligned} b_1 \le x_1 \le \beta_1, \\ b_2 \le x_2 \le \beta_2, \\ b_3 \le x_3 \le \beta_3. \end{aligned} \tag{2.4}$$

Ограничения по видам и количествам сырья будут иметь вид

$$\begin{aligned} a_{11}x_1 + a_{21}x_2 + a_{31}x_3 \le B_1, \\ a_{12}x_1 + a_{22}x_2 + a_{32}x_3 \le B_2, \\ a_{13}x_1 + a_{23}x_2 + a_{33}x_3 \le B_3, \\ a_{14}x_1 + a_{24}x_2 + a_{34}x_3 \le B_4. \end{aligned} \tag{2.5}$$

Суммарная прибыль предприятия

$$W = C_1x_1 + C_2x_2 + C_3x_3 \to \max. \tag{2.6}$$

Таким образом, имеем задачу линейного программирования: определить такие неотрицательные значения переменных x_1, x_2, x_3, чтобы они удовлетворяли ограничениям (2.3 – 2.5) и обращали бы в максимум линейную функцию (2.6) этих переменных.

Задача о загрузке оборудования

На фабрике имеется два типа станков в количестве N_1 и N_2, на которых изготавливают три вида тканей: T_1, T_2, T_3. Производительность станков обоих видов по разным тканям равна $a_{ij}, i = \overline{1,2}, j = \overline{1,3}$ (i – тип станка, j – вид ткани). Каждый метр ткани вида T_1 приносит при реализации доход C_1; вида T_2 – доход C_2; T_3 – C_3.

Принято решение, согласно которому фабрика должна ежемесячно производить не менее b_1 метров ткани вида T_1, не менее b_2 метров ткани T_2, не менее b_3 метров ткани T_3. «Перевыполнение» плана не должно превышать $\beta_1, \beta_2, \beta_3$ метров тканей T_1, T_2, T_3 соответственно. Необходимо распределить загрузку станков для изготовления тканей T_1, T_2, T_3, чтобы суммарный месячный доход фабрики был максимальным.

Какие параметры операции выбрать в качестве решения? Количество тканей x_1, x_2, x_3? Но в этом случае никак не учитывается производительность станков! Поэтому обозначим x_{11}– количество станков типа 1, на которых будут изготавливать ткань T_1; x_{12} – количество станков типа 1 для изготовления ткани T_2 и т.д. Всего элементов решения шесть:

$$\begin{cases} x_{11} x_{12} x_{13}, \\ x_{21} x_{22} x_{23}. \end{cases} \tag{2.7}$$

Условия выполнения плана запишутся как

$$\begin{cases} a_{11}x_{11} + a_{21}x_{21} \ge b_1, \\ a_{12}x_{12} + a_{22}x_{22} \ge b_2, \\ a_{13}x_{13} + a_{23}x_{23} \ge b_3; \end{cases} \tag{2.8}$$

ограничения на перевыполнение плана –

$$\begin{cases} a_{11}x_{11} + a_{21}x_{21} \le \beta_1, \\ a_{12}x_{12} + a_{22}x_{22} \le \beta_2, \\ a_{13}x_{13} + a_{23}x_{23} \le \beta_3; \end{cases} \tag{2.9}$$

ограничения на количество станков и их полную загрузку –

$$\begin{cases} x_{11} + x_{12} + x_{13} = N_1, \\ x_{21} + x_{22} + x_{23} = N_2; \end{cases} \tag{2.10}$$

выражение для суммарного дохода –

$$W = C_1(a_{11}x_{11} + a_{21}x_{21}) + C_2(a_{12}x_{12} + a_{22}x_{22}) + C_3(a_{13}x_{13} + a_{23}x_{23}). \tag{2.11}$$

Теперь математическую модель операции можно записать как

$$W = C_1(a_{11}x_{11} + a_{21}x_{21}) + C_2(a_{12}x_{12} + a_{22}x_{22}) + C_3(a_{13}x_{13} + a_{23}x_{23}) \to \max \quad (2.12)$$

при ограничениях

$$\begin{aligned} & b_1 \le a_{11}x_{11} + a_{21}x_{21} \le \beta_1, \\ & b_2 \le a_{12}x_{12} + a_{22}x_{22} \le \beta_2, \\ & b_3 \le a_{13}x_{13} + a_{23}x_{23} \le \beta_3, \\ & x_{11} + x_{12} + x_{13} = N_1, \\ & x_{21} + x_{22} + x_{23} = N_2, \end{aligned} \quad (2.13)$$

где

$$x_{ij} \ge 0, i = \overline{1,2}, j = \overline{1,3}. \quad (2.14)$$

Итак, мы рассмотрели несколько задач исследования операций. Можно выделить характерные черты, которые объединяют эти задачи:

- элементы решения $x_1, x_2, ..., x_n$ должны быть неотрицательными;
- требуется определить такие значения $x_1^0, x_2^0, ..., x_n^0 \ge 0$, чтобы выполнялись некоторые ограничения, имеющие вид линейных неравенств или равенств относительно переменных $x_j, j = \overline{1,n}$, и при этом некоторая линейная функция W тех же переменных $x_j, j = \overline{1,n}$ обращалась бы в минимум или максимум.

Для решения подобных задач разработан специальный математический аппарат, который носит название линейного программирования (планирования). А нужен ли специальный математический аппарат? Может быть, необходимо, как это принято при поиске экстремумов функций многих переменных в математике, взять частные производные по всем переменным $x_j, j = \overline{1,n}$ функции W, приравнять их к нулю и найти точку экстремума? По знаку вторых производных или методом приращений определить вид экстремума. Но в виду линейности функции W сделать этого нельзя. Производные по аргументам $x_j, j = \overline{1,n}$ функции W будут константами, равными $C_j, j = \overline{1,n}$, т.е. ни одна частная производная не будет равна нулю ни при каких значениях $x_j, j = \overline{1,n}$. Максимум или минимум функции W, если он существует, достигается всегда где-то на границах возможных значений $x_j, j = \overline{1,n}$, т.е. там, где начинают действовать ограничения.

Математический аппарат линейного программирования позволяет последовательно и довольно быстро обследовать границы области возможных решений и найти на этих границах решение задачи или доказать, что оно не существует.

2.3. Основная задача линейного программирования

Задача линейного программирования (ЗЛП) с ограничениями-равенствами называется основной задачей линейного программирования (ОЗЛП) [10].

ОЗЛП ставится следующим образом: имеется ряд переменных $x_1, x_2, ..., x_n$. Требуется определить такие неотрицательные значения этих переменных, которые удовлетворяли бы системе линейных уравнений

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n = b_1, \\ a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n = b_2, \\ .... \\ a_{m1}x_1 + a_{m2}x_2 + ... + a_{mn}x_n = b_m \end{cases} \tag{2.15}$$

и, кроме того, обращали бы в минимум линейную функцию

$$W = C_1x_1 + C_2x_2 + ... + C_nx_n. \tag{2.16}$$

Случай, когда необходимо определить максимум линейной функции, легко сводится к предыдущему. Для этого необходимо только изменить знак функции W и рассматривать функцию

$$W' = -W = -C_1x_1 - C_2x_2 - ... - C_nx_n, \tag{2.17}$$

т.е. $\max W = \min(-W)$ и $\max W = -\min(W)$.

Будем называть допустимым решением ОЗЛП любую совокупность переменных $x_1 \geq 0, x_2 \geq 0, ..., x_n \geq 0$, которая удовлетворяет уравнениям-ограничениям (2.15).

Оптимальным решением будем называть то из допустимых решений $x_1^0, x_2^0, ..., x_n^0$, при котором линейная функция (2.17) обращается в минимум.

Рассмотрим прежде всего вопрос о существовании допустимых решений ОЗЛП.

Определение

Рангом матрицы называется наибольший порядок отличного от нуля определителя, который можно получить, вычёркивая из матрицы какие-то строки и столбцы.

Матрицей системы уравнений

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n = b_1, \\ a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n = b_2, \\ .... \\ a_{m1}x_1 + a_{m2}x_2 + ... + a_{mn}x_n = b_m \end{cases} \tag{2.18}$$

называется матрица, составленная из коэффициентов $a_{ij}, i = \overline{1,m}, j = \overline{1,n}$ при $x_j, j = \overline{1,n}$:

$$A=\left\|a_{ij}\right\|, i=\overline{1,m}, j=\overline{1,n}=\begin{vmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ . & . & . & . \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{vmatrix}. \tag{2.19}$$

Расширенной матрицей системы линейных уравнений называется матрица A, дополненная столбцом свободных членов $b_i, i=\overline{1,m}$:

$$A_p=\begin{vmatrix} a_{11} & a_{12} & \ldots & a_{1n} & b_1 \\ a_{21} & a_{22} & \ldots & a_{2n} & b_2 \\ . & . & . & . & . \\ a_{m1} & a_{m2} & \ldots & a_{mn} & b_m \end{vmatrix}. \tag{2.20}$$

В линейной алгебре доказывается, что для совместности системы линейных уравнений (2.18) необходимо и достаточно, чтобы ранг матрицы системы A был равен рангу её расширенной матрицы A_p. Этот общий ранг r называется рангом системы и численно равен количеству линейно независимых уравнений-ограничений ОЗЛП.

Очевидно, что ранг системы не может быть больше числа уравнений m:

$$r \le m,$$

и ранг системы не может быть больше общего числа переменных n:

$$r \le n.$$

Структура задачи ЛП существенно зависит от ранга системы ограничений (2.15).

Рассмотрим случай, когда $r=n$, т.е. $m=n$. В этом случае число линейно независимых уравнений, входящих в систему (2.15), равно числу переменных. Система уравнений-ограничений ОЗЛП в этом случае имеет вид

$$\begin{cases} a_{11}x_1+a_{12}x_2+\ldots+a_{1n}x_n=b_1, \\ a_{21}x_1+a_{22}x_2+\ldots+a_{2n}x_n=b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{n1}x_1+a_{n2}x_2+\ldots+a_{nn}x_n=b_n. \end{cases} \tag{2.21}$$

Так как $r=n$, то определитель

$$\det A=\begin{vmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ . & . & . & . \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{vmatrix} \tag{2.22}$$

не равен 0.

Известно, что в этом случае система (2.21) имеет единственное решение. Но если в этом решении хотя бы одна из переменных $x_j, j=\overline{1,n}$

отрицательна, тогда полученное решение ОЗЛП недопустимо и, соответственно, ОЗЛП не имеет решения.

Если все $x_j, j=\overline{1,n}$ неотрицательны, то найденное решение является допустимым и оптимальным для ОЗЛП, потому что других решений нет. Этот тривиальный случай не интересует исследователей операций.

Мы будем исследовать случаи, когда $r<n$, т.е. когда число независимых уравнений, которым удовлетворяют переменные $x_j, j=\overline{1,n}$, меньше числа самих переменных. В случае, если система совместна, существует бесчисленное множество решений. При этом $k=n-r=n-m$ переменным можно давать произвольные значения, естественно, из области их определения ($x_j \geq 0, j=\overline{1,n}$). Эти k переменных называются свободными. Остальные $r=m=n-k$ переменных выражаются через свободные переменные и называются базисными.

2.4. Геометрическая интерпретация ОЗЛП

Пусть число свободных переменных n на два больше, чем независимых уравнений m, которым они (свободные переменные) должны удовлетворять, т.е. $n-m=2$.

Тогда $k=n-m=2$ переменных выбираем в качестве свободных, а m переменных сделаем базисными и выразим их через свободные. Получим $m=n-2$ уравнений вида

$$\begin{cases} x_3 = a_{31}x_1 + a_{32}x_2 + \beta_3, \\ x_4 = a_{41}x_1 + a_{42}x_2 + \beta_4, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots \\ x_n = a_{n1}x_1 + a_{n2}x_2 + \beta_n. \end{cases} \qquad (2.23)$$

Дадим ЗЛП геометрическую интерпретацию (рис. 2.1). По осям $0x_1, 0x_2$ будем откладывать значения свободных переменных. Так как по условию $x_1 \geq 0, x_2 \geq 0$, то допустимые значения свободных переменных располагаются в первом квадранте. Отметим это штриховкой, обозначающей допустимую сторону каждой координатной оси.

Рис. 2.1. Геометрическая интерпретация ЗЛП

Базисные переменные $x_3, x_4, ..., x_n$ также должны быть неотрицательными, т.е.

$$\begin{cases} x_3 = a_{31}x_1 + a_{32}x_2 + \beta_3 \geq 0, \\ x_4 = a_{41}x_1 + a_{42}x_2 + \beta_4 \geq 0, \\ .... \\ x_n = a_{n1}x_1 + a_{n2}x_2 + \beta_n \geq 0. \end{cases} \quad (2.24)$$

Изобразим условия (2.21) геометрически. Рассмотрим одно из этих условий, например $x_3 = a_{31}x_1 + a_{32}x_2 + \beta_3 \geq 0$. Положим, $x_3 = 0$. Тогда $a_{31}x_1 + a_{32}x_2 + \beta_3 = 0$. Это уравнение прямой линии в координатах $x_1 0 x_2$. На этой прямой $x_3 = 0$ (см. рис. 2.1). По одну сторону от прямой $x_3>0$, по другую $x_3<0$ в зависимости от коэффициентов a_{31}, a_{32} и β_3. Отметим штриховкой ту сторону от прямой $x_3 = 0$, где $x_3 > 0$.

Пример 1

$$x_3 = 2 - x_1 - x_2,$$
$$x_3 = 0 \to 2 - x_1 - x_2 = 0 \to x_2 = 2 - x_1,$$
$$x_1 = 0 \to x_2 = 2,$$
$$x_2 = 0 \to x_1 = 2.$$

Пусть $x_1 = x_2 = 0$, $x_3 = 2 - 0 - 0 = 2 > 0$. Если $x_1 = 2; x_2 = 2$, то $x_3 = 2 - 2 - 2 = -2 < 0$.

Штриховка направлена вниз у линии $x_3 = 0$. Аналогично построим прямые $x_4 = 0, x_5 = 0, ..., x_n = 0$ и отметим штриховкой допустимую сторону, где соответствующая базисная переменная $x_i, i = \overline{1,m}$ больше нуля. Таким образом, получено n прямых: две оси координат $x_1 = 0, x_2 = 0$ и n–2 прямых ($x_3 = 0, x_4 = 0, ..., x_n = 0$). Каждая из этих n прямых определяет допустимую полуплоскость, лежащую по одну ее сторону. Часть плоскости $x_1 0 x_2$, принадлежащая одновременно всем этим полуплоскостям, образует область допустимых решений (ОДР).

Область допустимых решений ОЗЛП всегда представляет собой выпуклый многоугольник (для n - m = 2). Выпуклой называется фигура (многоугольник), которая обладает следующим свойством: если две точки А и В отрезка АВ принадлежат этой фигуре, то и весь отрезок АВ принадлежит ей.

На рис. 2.1 приведен пример, когда ОДР ОЗЛП существует, т.е. система уравнений-ограничений ОЗЛП имеет неотрицательные решения.

Но могут быть случаи, когда неотрицательных решений системы не существует. Это означает, что не существует и ОДР (рис. 2.2).

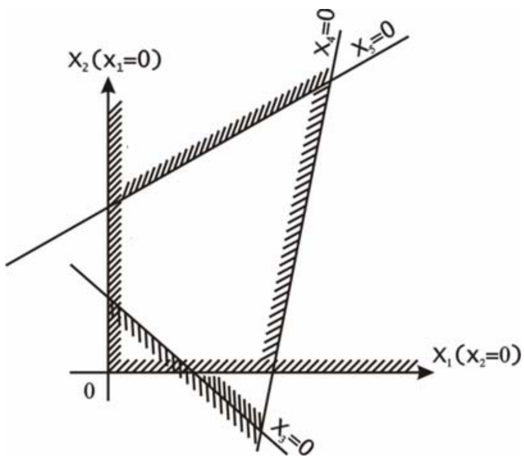


Рис. 2.2. Геометрическая интерпретация ЗЛП, не имеющей решения

Пример 2

В задаче ЛП семь переменных $x_j \geq 0, j = \overline{1,7}$. Имеется $m = 5$ ограничений:

$$\begin{cases} x_1 - x_2 + x_3 = 4, \\ 2x_1 - x_2 - x_3 - x_4 = -5, \\ x_1 + x_2 - x_5 = -4, \\ x_2 + x_6 = 5, \\ 2x_1 - 2x_2 - x_6 + 2x_7 = 7. \end{cases}$$

Требуется построить ОДР, если она существует.

Решение

Выберем в качестве свободных переменных $x_1 \geq 0, x_2 \geq 0$ и выразим через них базисные переменные x_3, x_4, x_5, x_6, x_7. Из первого уравнения $x_3 = -x_1 + x_2 + 4$, третьего – $x_5 = x_1 + x_2 + 4$, четвертого – $x_6 = -x_2 + 5$.

Подставим во второе уравнение $x_3 = -x_1 + x_2 + 4$. Имеем

$$2x_1 - x_2 - (-x_1 + x_2 + 4) - x_4 = -5,$$

$$2x_1 - x_2 + x_1 - x_2 - 4 - x_4 = -5,$$

$$x_4 = 3x_1 - 2x_2 + 1.$$

Подставим в пятое уравнение $x_6 = -x_2 + 5$. Имеем

$$2x_1 - 2x_2 - (-x_2 + 5) + 2x_7 = 7,$$

$$2x_1 - 2x_2 + x_2 - 5 + 2x_7 = 7,$$

$$x_7 = -x_1 + 0{,}5x_2 + 6.$$

$$\begin{cases} x_3 = 0;\; x_2 = x_1 - 4, \\ x_4 = 0;\; x_2 = 1{,}5x_1 + 0{,}5, \\ x_5 = 0;\; x_2 = -x_1 - 4, \\ x_6 = 0;\; x_2 = 5, \\ x_7 = 0;\; x_2 = 2x_1 - 12. \end{cases}$$

Для случая n – m = 2 оказалось возможным построить ОДР (см. рис. 2.3).

Теперь рассмотрим вопрос о нахождении из числа допустимых (предполагаем, что ОДР существует) оптимального решения, т.е. решения, которое приводит в минимум линейную функцию

$$W = c_1x_1 + c_2x_2 + ... + c_nx_n. \tag{2.25}$$

Рассмотрим случай, когда n-m=2, и дадим геометрическую интерпретацию поиску оптимального решения ОЗЛП. Положим, что $x_1 \ge 0, x_2 \ge 0$ – свободные переменные, а $x_3 \ge 0, x_4 \ge 0, ..., x_n \ge 0$ – базисные переменные. Подставим выражение для $x_3, x_4, ..., x_n$ (2.24) в выражение для W (2.25), приведем подобные члены и выразим как линейную функцию только свободных переменных x_1 и x_2.

Получим

$$W = \gamma_0 + \gamma_1x_1 + \gamma_2x_2\ , \tag{2.26}$$

где γ_0 – свободный член, которого в первоначальном виде у функции W не было.

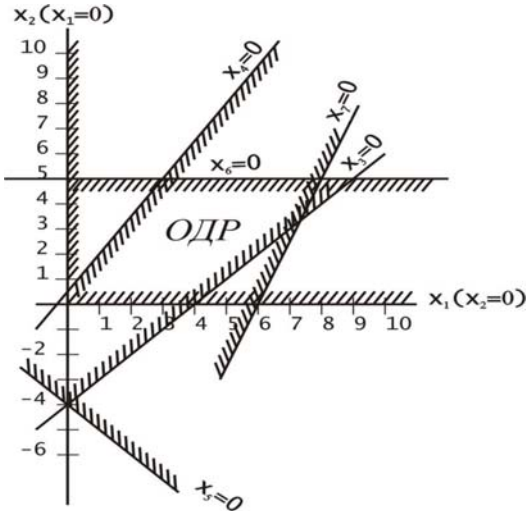


Рис. 2.3. Геометрическая интерпретация решения ЗЛП для примера 2

При переходе к выражению через свободные переменные x_1 и x_2 этот свободный член мог появиться. Очевидно, что линейная функция (2.26) достигает минимума при тех же значениях x_1 и x_2, что и функция $W' = \gamma_1x_1 + \gamma_2x_2$, т.к. $W' = W - \gamma_0$, где γ_0 не зависит от x_1 и x_2. Минимумы функций W и W' достигаются при одних и тех же значениях x_1 и x_2 и отличаются друг от друга на величину γ_0.

Приравняем W' к некоторой постоянной величине C:

$$W' = \gamma_1 x_1 + \gamma_2 x_2 = C, \qquad или \qquad x_2 = -\frac{\gamma_1}{\gamma_2} x_1 + \frac{C}{\gamma_2}.$$

Это уравнение прямой линии в координатах $\boldsymbol{x}_1\boldsymbol{x}_2$. Угловой коэффициент этой прямой равен $-\frac{\gamma_1}{\gamma_2}$. Если изменить значение константы C на C_1, то угловой коэффициент прямой $W' = C_1$ будет равен $-\frac{\gamma_1}{\gamma_2}$:

$$W' = \gamma_1 x_1 + \gamma_2 x_2 = C_1,$$

$$x_2 = -\frac{\gamma_1}{\gamma_2} + \frac{C_1}{\gamma_2}.$$

При изменении C на C_1 прямая W' перемещается параллельно самой себе. Если положить, что $W' = 0$ (основная прямая), то при перемещении этой прямой параллельно самой себе линейная функция W' будет убывать в одну сторону перемещения и, наоборот, возрастать в другую.

Построим основную прямую $W' = 0$ на плоскости $x_1 0 x_2$. Угловой коэффициент прямой равен $-\frac{\gamma_1}{\gamma_2}$. Выясним, при перемещении в какую сторону параллельно самой себе W' убывает, т.е. движется в сторону минимума.

Рассмотрим различные варианты.

1. Пусть $W' = 2x_1 + 2x_2$. Приравняем $W' = 0$. Имеем $x_2 = -x_1$. Направление убывания W' показано стрелками (рис. 2.4) для случая $\gamma_1>0$, $\gamma_2>0$.

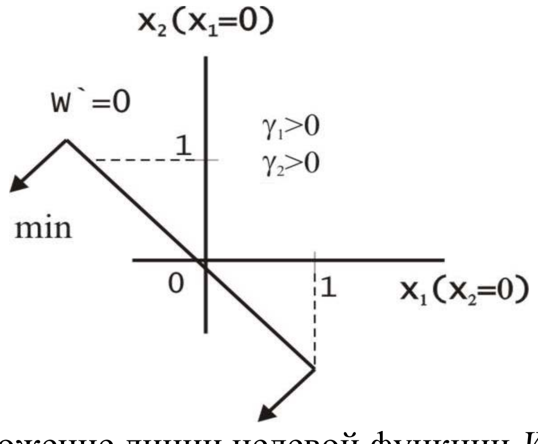


Рис. 2.4. Расположение линии целевой функции W при $\gamma_1>0$, $\gamma_2>0$

2. Пусть $W' = -2x_1 + 2x_2$ для $W' = 0 \rightarrow x_2 = x_1$ (рис. 2.5).

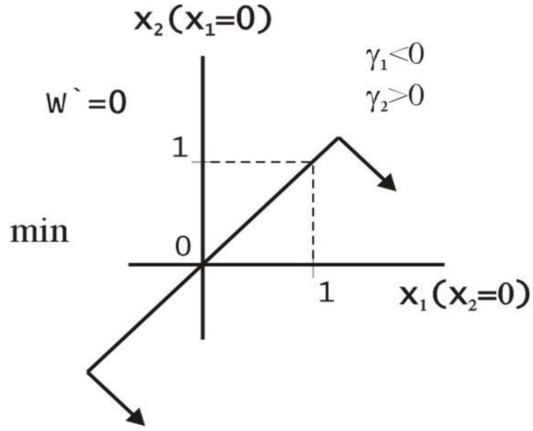


Рис. 2.5. Расположение линии целевой функции W при γ_1<0, γ_2>0

3. Пусть $W' = 2x_1 - 2x_2$ для $W' = 0 \rightarrow x_2 = x_1$ (рис. 2.6).

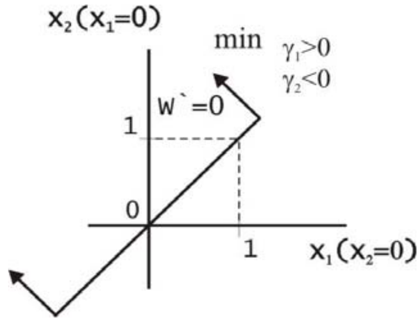


Рис. 2.6. Расположение линии целевой функции W при γ_1>0, γ_2<0

4. Пусть $W' = -2x_1 - 2x_2$ для $W' = 0 \rightarrow x_2 = -x_1$

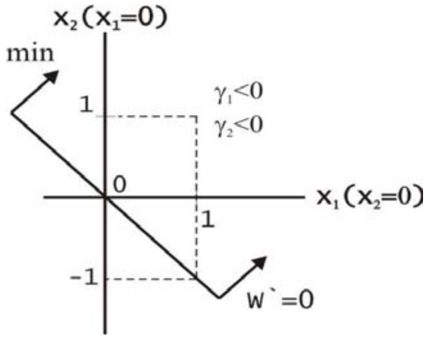


Рис. 2.7. Расположение линии целевой функции W при γ_1<0, γ_2<0

Положение на плоскости $x_1 0 x_2$ основной прямой $W' = 0$ и направление убывания линейной формы W' определяются величинами и знаками коэффициентов γ_1 и γ_2 при свободных переменных x_1 и x_2 в выражении W'.

Дадим геометрическую интерпретацию нахождения оптимального решения ОЗЛП среди допустимых. Пусть имеется ОДР и основная прямая $W' = 0$ (рис. 2.8). Известно направление убывания линейной формы W'.

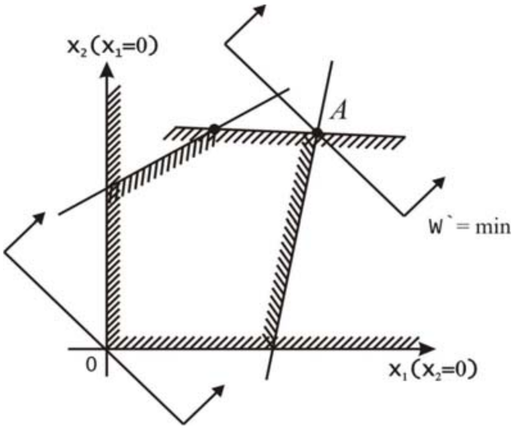


Рис. 2.8. Геометрическая интерпретация нахождения оптимального решения ОЗЛП

При перемещении основной прямой в направлении, указанном стрелками, линейная форма W' будет убывать. Очевидно, что наименьшего своего значения W' достигнет, когда прямая $W' = C^0$ будет проходит через наиболее удаленную от начала координат точку ОДР. Координаты (в данном случае т. А) x_1^0 и x_2^0 определяют оптимальное решение ОЗЛП. Зная оптимальные значения свободных переменных x_1^0 и x_2^0 и подставив их в уравнения, можно определить оптимальные значения базисных переменных:

$$\begin{cases} x_3^0 = a_{31}x_1^0 + a_{32}x_2^0 + \beta_3, \\ x_4^0 = a_{41}x_1^0 + a_{42}x_2^0 + \beta_4, \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ x_n^0 = a_{n1}x_1^0 + a_{n2}x_2^0 + \beta_n, \end{cases},$$

а также оптимальное (в смысле минимума) значение линейной функции W': $W'_{\min} = \gamma_0 + \gamma_1 x_1^0 + \gamma_2 x_2^0$.

Пример 3

Найти оптимальное решение ОЗЛП для ограничений примера 2, которое обращает в минимум линейную целевую функцию

$$W = x_1 - x_2 + 2x_3 - x_4 - 3x_5 + x_6 - 2x_7. \qquad (2.27)$$

Решение

В примере 2 уравнения-ограничения были разрешены относительно базисных переменных x_3, x_4, x_5, x_6, x_7, которые были выражены через свободные переменные x_1 и x_2:

$$\begin{cases} x_3 = -x_1 + x_2 + 4, \\ x_4 = 3x_1 - 2x_2 + 1, \\ x_5 = x_1 + x_2 + 4, \\ x_6 = -x_2 + 5, \\ x_7 = -x_1 + 0{,}5x_2 + 6. \end{cases}$$

Подставим эти выражения в целевую функцию и, приводя подобные члены, получим

$$W = x_1 - x_2 + 2(-x_1 + x_2 + 4) - (3x_1 - 2x_2 + 1) - 3(x_1 + x_2 + 4) + (-x_2 + 5) -$$
$$-2(-x_1 + 0{,}5x_2 + 6) = (x_1 - 3x_1 - 3x_1) + (-x_2 + 2x_2 - 3x_2) + (2 + 8 - 3 - 12 + 5 - 12) =$$
$$= -5x_1 - 2x_2 - 12; \quad W = -5x_1 - 2x_2 - 12.$$

Вернемся к ОДР, построенной для примера 2 (см. рис. 2.3). Основная прямая для $W \to W' = -5x_1 - 2x_2$. Построим основную прямую $W' = 0$, $W' = 0 \to x_2 = -\frac{5}{2}x_1$ (рис. 2.9).

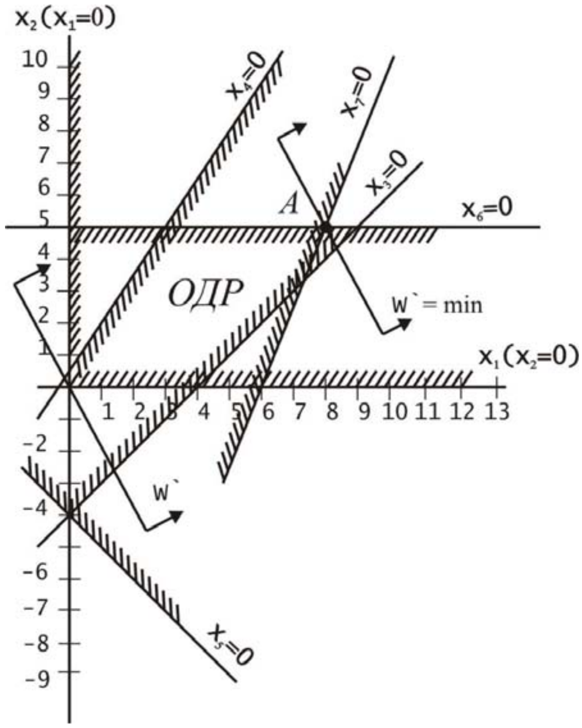


Рис. 2.9. Геометрическая интерпретация нахождения оптимального решения ОЗЛП для примера 3

Прямая $W' = 0$ параллельна самой себе, и в т. А имеем оптимальное решение. В т. А $x_6^0 = 0$ и $x_7^0 = 0$, тогда $x_6^0 = 0 \to \; + x_2 = 5 \to x_2^0 = 5$.

Отсюда $x_7^0 = 0 \to -x_1 + \frac{1}{2}x_2 + 6 = 0 \to -x_1 + \frac{5}{2} + 6 = 0 \to x_1^0 = 8{,}5$.

Таким образом,

$$x_1^0 = 8{,}5; \quad x_2^0 = 5,$$
$$x_3^0 = -8{,}5 + 5 + 4 = 0{,}5,$$
$$x_4^0 = 3 \cdot 8{,}5 - 2 \cdot 5 + 1 = 16{,}5,$$
$$x_5^0 = 8{,}5 + 5 + 4 = 17{,}5.$$

Оптимальное решение:

$$x_1^0 = 8{,}5; \; x = 5; \; x_3^0 = 0{,}5; \; x_4^0 = 16{,}5; \; x_5^0 = 17{,}5.$$

Определим минимум W:

$$W^0_{\min} = -5 \cdot 8{,}5 - 2 \cdot 5 - 12 = -64{,}5.$$

В итоге получим

$$W^0_{\min} = x^0_1 - x^0_2 + 2x^0_3 - x^0_4 - 3x^0_5 + x^0_6 - 2x^0_7 = 8{,}5 - 5 + 2 \cdot 0{,}5 - 16{,}5 - 3 \cdot 17{,}5 + $$
$$+ 0 + 2 \cdot 0 = 4{,}5 - 69 = -64{,}5.$$

Анализ этого частного случая ОЗЛП (k = n-m = 2) позволяет сделать выводы:

1. Решение ОЗЛП, если оно существует, не может быть расположено внутри ОДР. Оно может находиться только на границе ОДР.

2. Решение ОЗЛП может быть неединственным (оптимальным). Если основная прямая параллельна той стороне многоугольника допустимых решений, где достигается минимум W', то этот минимум достигается в любой точке этой стороны, т.е. имеет место бесчисленное множество оптимальных решений.

3. ОЗЛП может не иметь оптимального решения, даже если ОДР существует, но эта ОДР открытая или неограниченная. Геометрическая интерпретация этого варианта представлена на рис. 2.10.

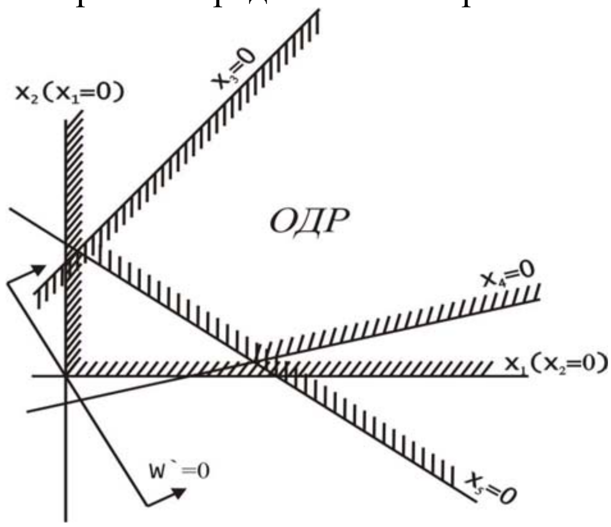


Рис. 2.10. Геометрическая интерпретация решения ОЗЛП (оптимальное решение отсутствует)

4. Оптимальное решение ОЗЛП достигается в одной из вершин многоугольника допустимых решений. Случай бесчисленного множества оптимальных решений также удовлетворяет этому выводу. Решение ОЗЛП, находящееся в одной из вершин многоугольника ОДР, называется опорным решением, а сама эта вершина – опорной точкой.

5. Если число свободных переменных в ОЗЛП равно k = 2, а число базисных переменных равно m и решение (оптимальное) ОЗЛП существует, то оно всегда расположено в вершине ОДР, где по крайней мере две из переменных $x_j, j = \overline{1,n}$ равны нулю. Но бывают случаи, когда в опорной точке пересекаются более двух прямых-ограничений. Тогда в

оптимальном решении равны нулю не две, а больше переменных. Это случай ОЗЛП называется вырожденным.

6. Интуитивно понятно, что для поиска оптимального решения ОЗЛП необходимо проанализировать опорные решения (опорные точки ОДР) и выбрать из них то, где линейная функция W достигнет минимума.

Подобные выводы можно сделать для любых количеств переменных ЗЛП и уравнений-ограничений при условии $m<n$.

Оптимальное решение, если оно существует, находится не внутри, а на границе ОДР, в одной из опорных точек. В каждой опорной точке не менее k переменных ($k = n - m$) ЗЛП равны нулю.

Для ускорения процедуры отыскания оптимального решения необходимо отыскать сначала опорные точки (опорные решения) и, целенаправленно анализируя эти опорные решения, определить минимум целевой функции W.

2.5. Задача линейного программирования с ограничениями-неравенствами. Переход к ОЗЛП и обратно [10]

В большинстве практических задач ЛП ограничения заданы не уравнениями, а неравенствами. Переход от ЗЛП общего вида к ОЗЛП осуществляется следующим образом.

Пусть имеется ЗЛП с x_j, $j = \overline{1,n}$ переменными. Ограничения задачи ЛП имеют вид линейных неравенств. Знаки неравенств могут быть как ≥, так и ≤.

Зададим все ограничения неравенства в стандартной форме записи:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n \ge b_1, \\ a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n \ge b_2, \\ .... \\ a_{m1}x_1 + a_{m2}x_2 + ... + a_{mn}x_n \ge b_m. \end{cases} \tag{2.28}$$

Считаем, что все неравенства линейно независимы, т.е. никакое из них нельзя представить в виде линейной комбинации других. Требуется найти такую неотрицательную совокупность значений $x_1, x_2, \ldots, x_n$, которая удовлетворяла бы неравенствам системы (13) и, кроме того, обращала бы в минимум линейную функцию

$$W = c_1x_1 + c_2x_2 + ... + c_nx_n.$$

От поставленной таким образом ЗЛП легко перейти к ОЗЛП. Перепишем систему (2.28) в виде

$$\begin{cases} y_1 = a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n + b_1, \\ y_2 = a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n + b_2, \\ .... \\ y_m = a_{m1}x_1 + a_{m2}x_2 + ... + a_{mn}x_n + b_m, \end{cases} \tag{2.29}$$

где y_1, y_2, …, y_m – новые переменные, которые называются дополнительными.

Из системы (2.29) следует, что дополнительные y_i, $i=\overline{1,m}$, так же как и основные x_j, $j=\overline{1,n}$, должны быть неотрицательными.

Появилась новая постановка задачи ЛП: определить такие неотрицательные значения $n+m$ переменных x_j, $j=\overline{1,n}$ и y_i, $i=\overline{1,m}$, которые удовлетворяли бы системе уравнений (2.29) и обращали бы в минимум линейную функцию

$$W = c_1x_1 + c_2x_2 + ... + c_nx_n + 0 \cdot y_1 + 0 \cdot y_2 + ... + 0 \cdot y_m.$$

Это ОЗЛП. Уравнения в системе заданы в виде, когда базисные переменные y_i $i=\overline{1,m}$ уже выражены через свободные x_j, $j=\overline{1,n}$. Общее количество переменных увеличилось на m, но задача сведена к ОЗЛП.

Пример 4

Определить минимум $W = x_1 - 2x_2 - 3x_3$ при ограничениях

$$\begin{cases} 2x_1 - x_2 + 3x_3 \le 6, \\ x_3 - 3x_2 \le -1, \\ x_5 - 2x_4 + x_1 \ge -1, \\ x_5 - x_1 \le 0, \\ x_j \ge 0,\ j=\overline{1,n}. \end{cases} \tag{2.30}$$

Привести эту задачу к ОЗЛП.

Решение

Приведем неравенства (2.30) к стандартной форме:

$$\begin{cases} -2x_1 + x_2 - 3x_3 + 6 \ge 0, \\ -x_3 + 3x_2 - 1 \ge 0, \\ x_5 - 2x_4 + x_1 + 1 \ge 0, \\ -x_5 + x_1 \ge 0. \end{cases}$$

Введем дополнительные переменные:

$$\begin{cases} y_1 = -2x_1 + x_2 - 3x_3 + 6, \\ y_2 = 3x_2 - x_3 - 1, \\ y_3 = x_1 - 2x_4 + x_5 + 1, \\ y_4 = x_1 - x_5. \end{cases} \tag{2.31}$$

Задача ЛП сводится к поиску $x_j \ge 0, j=\overline{1,5}$, $y_i \ge 0$, $i=\overline{1,4}$, которые бы удовлетворяли (2.31) и приводили бы к минимуму $W = x_1 - 2x_2 - 3x_3$.

Переход от уравнений-ограничений к неравенствам может быть произведен следующим образом:

имеем

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n = b_1, \\ a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n = b_2, \\ .... \\ a_{m1}x_1 + a_{m2}x_2 + ... + a_{mn}x_n = b_m; \end{cases}$$

перепишем

$$\begin{cases} b_1 \geq a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n \geq b_1, \\ b_2 \geq a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n \geq b_2, \\ .... \\ b_m \geq a_{m1}x_1 + a_{m2}x_2 + ... + a_{mn}x_n \geq b_m. \end{cases}$$

Пример 5

Эта система после введения $2m$ дополнительных переменных будет выглядеть так:

$$\begin{cases} x_1 + x_2 = 1, \\ x_2 - 2x_3 = -3, \\ x_3 - x_4 + x_5 = 1; \end{cases}$$

$$\begin{cases} x_1 + x_2 - 1 \geq 0, \\ -x_1 - x_2 + 1 \geq 0, \\ x_2 - 2x_3 + 3 \geq 0, \\ -x_2 + 2x_3 - 3 \geq 0, \\ x_3 - x_4 + x_5 - 1 \geq 0, \\ -x_3 + x_4 - x_5 + 1 \geq 0. \end{cases}$$

Вводим дополнительные переменные $y_i \geq 0,\ i = \overline{1,6}$:

$$\begin{cases} y_1 = x_1 + x_2 - 1, \\ y_2 = -x_1 - x_2 + 1, \\ y_3 = x_2 - 2x_3 + 3, \\ y_4 = -x_2 + 2x_3 - 3, \\ y_5 = x_3 - x_4 + x_5 - 1, \\ y_6 = -x_3 + x_4 - x_5 + 1. \end{cases}$$

Физический смысл дополнительных переменных, особенно при таких преобразованиях, не всегда понятен и требует тщательного осмысливания.

2.6. Симплекс-метод решения задачи линейного программирования

Идея симплекс-метода относительно проста. Пусть в ОЗЛП имеется n переменных и m независимых линейных ограничений, заданных уравнениями. Известно, что оптимальное решение ОЗЛП, если оно существует, достигается в одной из вершин ОДР (опорной точке), где не менее $k = n\text{-}m$ переменных задачи равны нулю [10].

Выберем в ОЗЛП какие-то k переменных в качестве свободных и выразим m базисных переменных через свободные.

Пусть свободные переменные имеют вид x_1, x_2, …, x_k , тогда базисные переменные запишем так:

$$\begin{cases} x_{k+1} = a_{k+1,1}x_1 + a_{k+1,2}x_2 + ... + a_{k+1,k}x_k + \beta_{k+1}, \\ x_{k+2} = a_{k+2,1}x_1 + a_{k+2,2}x_2 + ... + a_{k+2,k}x_k + \beta_{k+2}, \\ x_n = a_{n,1}x_1 + a_{n,2}x_2 + ... + a_{n,k}x_k + \beta_n. \end{cases} \quad (2.32)$$

Положим, $x_1 = 0, x_2 = 0, ..., x_k = 0$. Получим

$$x_{k+1} = \beta_{k+1}, x_{k+2} = \beta_{k+2}, ...x_n = \beta_n.$$

Это решение ОЗЛП. Оно допустимо, если все свободные члены неотрицательные, т.е.

$$\beta_{k+1} \geq 0, \beta_{k+2} \geq 0, ..., \beta_n \geq 0.$$

Предположим, что это так. Тогда полученное решение опорное. Выясним, оптимально ли полученное решение?

Выразим линейную функцию W через свободные переменные x_1, x_2, …, x_k:

$$W = \gamma_0 + \gamma_1 x_1 + \gamma_2 x_2 + ... + \gamma_k x_k \to \min. \quad (2.33)$$

Для полученного опорного решения $x_1 = 0$, $x_2 = 0$, …, $x_k = 0$; $W = \gamma_0$.

Посмотрим, нельзя ли улучшить (оптимизировать) решение, т.е. уменьшить функцию W, увеличивая в положительную сторону какие-то переменные x_1, x_2, …, x_k. (Уменьшить их нельзя, т.к. они станут отрицательными, что недопустимо по условиям ЛП).

Если все $\gamma_j, j = \overline{1,k}$ в (2.33) неотрицательны (≥0), то при увеличении каких-либо переменных x_1, x_2, …, x_k в положительную сторону невозможно уменьшить W (значение W будет возрастать). Следовательно, найденное опорное решение

$$x_1 = 0, x_2 = 0, x_k = 0, x_{k+1} = \beta_{k+1}, x_{k+2} = \beta_{k+2}, ..., x_n = \beta_n$$

является оптимальным и $W^0_{\min} = \gamma_0$.

Вывод

Если в (2.33) все $\gamma_j, j = \overline{1,k}$ неотрицательны, то полученное ранее опорное решение ОЗЛП является оптимальным. Наоборот, если среди $\gamma_j, j = \overline{1,k}$ есть отрицательные, то увеличение при них каких-то переменных из x_1, x_2, …, x_k приведет к уменьшению W, т.е. улучшит решение. Например, пусть в (2.33) $\gamma_1 \leq 0$. Значит, если увеличивать x_1, т.е. переходить от полученного опорного решения к другому, где $x_1 \neq 0$, а какая-то переменная $x_{k+p}=0$, можно добиться уменьшения значения W. Однако увеличивать x_1 следует осторожно, чтобы переменные x_{k+1}, x_{k+2}, …, x_n (в них входит x_1 как свободная) не стали отрицательными.

Проанализируем случаи, когда увеличение в положительную сторону x_1 может сделать какие-то из переменных x_{k+1}, x_{k+2}, …, x_n отрицательными. Это может произойти, если в каком-то (каких-то) уравнении системы (2.32) коэффициент $a_{p,1}$ при x_1 отрицателен, если же коэффициенты при x_1 строго положительны, то x_1 можно увеличивать беспредельно, а значит линейная функция W не ограничена снизу, и оптимального решения ОЗЛП не существует.

Допустим, что среди уравнений системы (2.32) имеются такие, в которых коэффициент при x_1 отрицателен.

Выберем одно из уравнений, где базисной переменной является $x_e=a_{e_1}x_i+a_{e_2}x_2+...+a_{e_k}x_k+\beta_e$. Как отмечено выше, $\beta_e>0$, а $a_{e_1}<0$. Оставляем $x_2=x_3=\ldots=x_k=0$.

Имеем

$$x_e=a_{e_1}x_1+\beta_e. \tag{2.34}$$

Из (2.34) видно, что увеличивать x_1 можно только до величины

$$x_e=0,\ x_1=-\frac{\beta_e}{a_{e_1}}.$$

Если x_1 будет больше $-\dfrac{\beta_e}{a_{e_1}}$, то x_e станет отрицательной, что недопустимо по условиям ОЗЛП. Определим ту из переменных x_{k+1}, x_{k+2}, …, x_n, в которой коэффициент при x_1 в уравнениях отрицателен.

Тогда очевидно, что раньше всех в нуль обратится та базисная переменная, для которой отношение $-\dfrac{\beta_e}{a_{e_1}}\to\min$.

Пусть этой переменной будет x_r. Тогда имеет смысл для улучшения решения ОЗЛП «переразрешить» систему (2.32), выведя x_1 из числа свободных переменных и сделав ее базисной, и, наоборот, базисную переменную x_r сделать свободной в новом решении, где x_r будет равна 0. Необходимо перейти от опорного решения $x_1=0, x_2=0, x_k=0$, $x_{k+1}=\beta_{k+1}, x_{k+2}=\beta_{k+2},...,x_n=\beta_n$ к другому опорному решению $x_2=x_3=...=x_k=x_r=0,\ x_{k+1}=\beta_{k+1}, x_{k+2}=\beta_{k+2}, x_{r-1}=\beta_{r-1},\ x_{r+1}=\beta_{r+1},...,$ $x_n=\beta_n$.

Получим новую систему уравнений типа (2.32). Тогда можно выразить через новые свободные переменные и целевую функцию

$$W=\gamma_0'+\gamma_2'x_2+...+\gamma_k'x_k+\gamma_r'x_r. \tag{2.35}$$

Проанализируем коэффициенты $\gamma_2', \gamma_3', ..., \gamma_k', \gamma_r'$ в (2.35). Если они неотрицательны, то полученное опорное решение оптимально в смысле минимума *W*. Задавая новые свободные переменные $x_2 = x_3 = ... = x_k = x_2 = 0$, получим минимум $W = \gamma_0'$. Если же среди коэффициентов $\gamma_j, j = \overline{2, k+1}$ есть отрицательные, система (2.32), полученная из (2.35), вновь «переразрешается» относительно других базисных переменных и так далее, пока не будет найдено оптимальное в смысле минимума значение *W*. Особо следует отметить, что все вышесказанное справедливо для опорных решений, все $\beta_p > 0, p = \overline{k+1, n}$.

Пример 6

Определить

$$\min W = 5x_1 - 2x_3 \tag{2.36}$$

при ограничениях

$$\begin{cases} -5x_1 - x_2 + 2x_3 \le 2, \\ -x_1 + x_3 + x_4 \le 5, \\ -3x_1 + 5x_4 \le 7, \\ x_1, x_2, x_3, x_4 \ge 0. \end{cases} \tag{2.37}$$

Решение

Приведем неравенства к стандартному виду умножением левых и правых частей неравенств системы (2.27) на –1:

$$\begin{cases} 5x_1 + x_2 - 2x_3 \ge -2, \\ -x_1 - x_3 - x_4 \ge -5, \\ 3x_1 - 5x_4 \ge -7. \end{cases}$$

Вводя дополнительные переменные $y_1 \ge 0$, $y_2 \ge 0$, $y_3 \ge 0$ и назначив их базисными, получим

$$\begin{cases} y_1 = 5x_1 + x_2 - 2x_3 + 2, \\ y_2 = -x_1 - x_3 - x_4 + 5, \\ y_3 = 3x_1 - 5x_4 + 7. \end{cases} \tag{2.38}$$

Общее число переменных $n = 7$, число уравнений $m = 3$. Поэтому число свободных переменных $k = n - m = 7 - 3 = 4$.

Пусть свободные переменные x_1, x_2, x_3, x_4. Положим, $x_1 = x_2 = x_3 = x_4 = 0$. Из (2.38) получаем $y_1 = 2, y_2 = 5, y_3 = 7$.

Имеем опорное решение (все переменные ≥ 0) $x_1 = x_2 = x_3 = x_4 = 0$; $y_1 = 2, y_2 = 5, y_3 = 7$. Для такого опорного решения $W = 0$.

Полученное решение неоптимальное в смысле минимума *W*, потому что в (2.36) коэффициент при x_3 отрицателен, значит, увеличивая x_3 в положительную сторону, можно уменьшить *W*.

Определим, насколько (до какого численного значения) можно увеличивать x_3, чтобы полученное решение было допустимым и опорным.

Рассмотрим систему (2.38). В уравнении для y_1 и y_2 коэффициенты $a_{13}=-2$ и $a_{23}=-1$. При увеличении x_3 переменные y_1 и y_2 могут стать отрицательными, что недопустимо.

Какая из переменных, y_1 или y_2, быстрее станет отрицательной при увеличении x_3? При каком минимальном значении x_3 переменная y_1 или y_2 станет равной нулю?

Положим, в (2.38) $y_1=0 \to 5x_1+x_2-2x_3+2=0$. Если $x_1=x_2=0$ (опорное решение), то x_3 = 1 и y_1 = 0, $y_2=0 \to x_1-x_3-x_4+5=0$. Если $x_1=x_4=0$, то x_3 = 5 и y_2 = 0. Значит, наиболее угрожаемая, чувствительная базисная переменная – y_1 (она становится равной 0 при x_3=1). Вводим y_1 в число свободных переменных, а x_3 – базисных. «Переразрешаем» систему (2.38) относительно базисных переменных x_3, y_2, y_3. При этом x_1, x_2, y_1, y_4 станут свободными переменными. Из первого уравнения (2.38) имеем

$$x_3=\frac{5}{2}x_1+\frac{1}{2}x_2-\frac{1}{2}y_1+1. \tag{2.39}$$

Подставим выражение (2.39) для x_3 во второе уравнение (2.38):

$$y_2=x_1-\frac{5}{2}x_1-\frac{1}{2}x_2+\frac{1}{2}y_1-1-x_4+5=-\frac{3}{2}x_1-\frac{1}{2}x_2+\frac{1}{2}y_1-x_4+4.$$

Уравнение y_3 в (2.38) не содержит x_3 и не изменится. Получаем систему

$$\begin{cases} x_3=\frac{5}{2}x_1+\frac{1}{2}x_2-\frac{1}{2}y_1+1, \\ y_2=-\frac{3}{2}x_1-\frac{1}{2}x_2+\frac{1}{2}y_1-x_4+4, \\ y_3=3x_1-5x_4+7. \end{cases} \tag{2.40}$$

Выразим линейную функцию (2.36) через новые свободные переменные x_1, x_2, x_3, x_4:

$$W=5x_1-2x_3=5x_1-2(\frac{5}{2}x_1+\frac{1}{2}x_2-\frac{1}{2}y_1+1)=-x_2+y_1-2,$$

$$W=-x_2+y_1-2. \tag{2.41}$$

Для данного опорного решения $x_1=x_2=y_1=x_4=0$; $x_3=1$, $y_2=4$, $y_3=7$, $W=0+0-2=-2$.

Новое значение W = –2, прежнее W = 0. Является ли полученное значение W оптимальным? Нет, потому что коэффициент при x_2 в (2.41) отрицателен, т.е. $\gamma_2'=-1$. Значит нужно ввести x_2 в состав базисных переменных, а одну из базисных переменных системы (2.40) сделать свободной. Этой переменной будет y_2, т.к. в уравнении системы (2.40) коэффициент при x_2 отрицателен. В уравнении для x_3 коэффициент при x_2 положителен, а в уравнении для y_3 переменной x_2 нет.

Делаем x_2 базисной переменной, а y_2 свободной:

$$y_2 = -\frac{3}{2}x_1 - \frac{1}{2}x_2 + \frac{1}{2}y_1 + 4,$$
$$x_2 = -3x_1 - 2y_2 + y_1 + 8.$$

Подставим полученное для x_2 выражение в уравнение для x_3 в (2.40) и получим

$$x_3 = \frac{5}{2}x_1 + \frac{1}{2}(-3x_1 - 2y_2 + y_1 - x_4 + 8) - \frac{1}{2}y_1 + 1 = \frac{5}{2}x_1 + \frac{3}{2}x_1 - y_2 +$$
$$+ 4 - x_4 + 1 = x_1 - y_2 - x_4 + 5.$$

В уравнение (2.40) для y_3 переменная x_2 не входит. Получим новую систему

$$\begin{cases} x_3 = x_1 - y_2 - x_4 + 5, \\ x_2 = -3x_1 - 2y_2 + y_1 + 8, \\ y_3 = 3x_1 - 5x_4 + 7. \end{cases} \tag{2.42}$$

Выразим W из (2.41) через свободные переменные системы (2.42):

$$\begin{aligned} W &= -x_2 + y_1 - 2 = -(-3x_1 - 2y_2 + y_1 + 8) + y_1 - 2 = \\ &= 3x_1 + 2y_2 + 8 - 2 = 3x_1 + 2y_2 - 10. \end{aligned} \tag{2.43}$$

Подставив значение переменной x_3 из (2.42) в (2.36), получим

$$\begin{aligned} W &= 5x_1 - 2x_3 = 5x_1 - 2(x_1 - y_2 - x_4 + 5) = 5x_1 - 2x_1 + \\ &+ 2y_2 + 2x_4 - 10 = 3x_1 + 2y_2 + 2x_4 - 10. \end{aligned} \tag{2.44}$$

Задавая свободные переменные $x_1 = x_4 = y_1 = y_2 = 0$, получим в обоих случаях $W = -10$.

Это решение оптимально, т.к. в выражениях (2.43) и (2.44) для W все коэффициенты при свободных переменных положительны.

Оптимальное решение

$$x_1^0 = 0, x_2^0 = 8, x_3^0 = 5, x_4^0 = 0, y_1^0 = 0, y_2^0 = 0, y_3^0 = 7, W_{\min}^0 = -10.$$

2.7. Табличный алгоритм замены базисных переменных на свободные и наоборот

Процедура «переразрешения» системы уравнений-ограничений ОЗЛП относительно новых базисных переменных может быть существенно упрощена. Этого можно достигнуть, если производить вычисления по алгоритму в стандартных таблицах. Рассмотрим алгоритм на конкретном примере. Вопросы работоспособности алгоритма в общем случае не обсуждаются, а считается, что алгоритм справедлив и для общего случая.

Рассмотрим систему трех уравнений-ограничений какой-то ОЗЛП, записанных в стандартной форме:

$$\begin{cases} y_1 = a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + a_{14}x_4 + b_1, \\ y_2 = a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + a_{24}x_4 + b_2, \\ y_3 = a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + a_{34}x_4 + b_3, \end{cases} \quad (2.45)$$

где переменные x_1, x_2, x_3, x_4 являются свободными, а y_1, y_2, y_3 – базисными.

Пусть требуется превратить переменную x_2 в базисную, а базисную переменную y_3 сделать свободной. Такую замену переменных символически обозначим как $x_2 \leftrightarrow y_3$.

Ранее была рассмотрена процедура замены переменных способом подстановки. Такая процедура достаточно громоздкая, требующая внимания, вероятность совершения ошибок в расчетах велика. Кроме того, каждый раз нужно проделывать одни и те же операции, выполняемые по определенным правилам. Эти правила реализованы в виде табличного алгоритма.

Для упрощения расчетов по этому алгоритму необходимо преобразовать систему (2.45) к виду

$$\begin{cases} y_1 = b_1 - (-a_{11}x_1 - a_{12}x_2 - a_{13}x_3 - a_{14}x_4), \\ y_2 = b_2 - (-a_{21}x_1 - a_{22}x_2 - a_{23}x_3 - a_{24}x_4), \\ y_3 = b_3 - (-a_{31}x_1 - a_{32}x_2 - a_{33}x_3 - a_{34}x_4). \end{cases} \quad (2.46)$$

Обозначим $\alpha_{11} = -a_{11}; \alpha_{12} = -a_{12}; ....; \alpha_{31} = -a_{31}; ......; \alpha_{34} = -a_{34}$. Тогда система (2.46) примет вид

$$\begin{cases} y_1 = b_1 - (\alpha_{11}x_1 + \alpha_{12}x_2 + \alpha_{13}x_3 + \alpha_{14}x_4), \\ y_2 = b_2 - (\alpha_{21}x_1 + \alpha_{22}x_2 + \alpha_{23}x_3 + \alpha_{24}x_4), \\ y_3 = b_3 - (\alpha_{31}x_1 + \alpha_{32}x_2 + \alpha_{33}x_3 + \alpha_{34}x_4). \end{cases} \quad (2.47)$$

Запишем уравнения системы (2.47) в виде стандартной таблицы:

	Своб. член	x_1	x_2	x_3	x_4
y_1	b_1	α_{11}	α_{12}	α_{13}	α_{14}
y_2	b_2	α_{21}	α_{22}	α_{23}	α_{24}
y_3	b_3	α_{31}	α_{32}	α_{33}	α_{34}

Производим замену $x_2 \leftrightarrow y_3$. Выделим жирными линиями столбцы, в которых расположен x_2, и строку, в которой находится y_3. Столбец x_2 называется разрешающим, а строка y_3 – разрешающей, коэффициент α_{32} – разрешающим элементом.

	Своб. член	x_1	x_2	x_3	x_4
y_1	b_1	α_{11}	α_{12}	α_{13}	α_{14}
y_2	b_2	α_{21}	α_{22}	α_{23}	α_{24}
y_3	b_3	α_{31}	α_{32}	α_{33}	α_{34}

После выполнения замены $x_2 \leftrightarrow y_3$ в строке, которая является разрешающей, должен появиться x_2, а в разрешающем столбце – y_3. Преобразуем разрешающую строку. Разрешим третье уравнение системы (2.47) относительно x_2:

$$y_3 = b_3 - (\alpha_{31}x_1 + \alpha_{32}x_2 + \alpha_{33}x_3 + \alpha_{34}x_4),$$

$$\alpha_{32}x_2 = b_3 - (\alpha_{31}x_1 + y_3 + \alpha_{33}x_3 + \alpha_{34}x_4),$$

$$x_2 = \frac{b_3}{\alpha_{32}} - (\frac{\alpha_{31}}{\alpha_{32}}x_1 + \frac{1}{\alpha_{32}}y_3x_2 + \frac{\alpha_{33}}{\alpha_{32}}x_3 + \frac{\alpha_{34}}{\alpha_{32}}x_4). \qquad (2.48)$$

Для преобразования остальных строк подставим в первое уравнение системы (2.47) вместо x_2 выражение (2.48):

$$y_1 = b_1 - (\alpha_{11}x_1 + \alpha_{12}x_2 + \alpha_{13}x_3 + \alpha_{14}x_4) \to y_1 = b_1 - \left\{\alpha_{11}x_{11} + \alpha_{12}\left[\frac{b_3}{\alpha_{32}} - \right.\right.$$

$$\left.\left. - \left(\frac{\alpha_{31}}{\alpha_{32}}x_1 + \frac{1}{\alpha_{32}}y_3 + + \frac{\alpha_{33}}{\alpha_{32}}x_3 + \frac{\alpha_{34}}{\alpha_{32}}x_4\right)\right] + \alpha_{13}x_3 + \alpha_{14}x_4\right\} = b_1 - \left\{\alpha_{11}x_1 + \frac{\alpha_{12} \cdot b_3}{\alpha_{32}} - \right.$$

$$\left. - \frac{\alpha_{12} \cdot \alpha_{31}}{\alpha_{32}}x_1 - \frac{\alpha_{12}}{\alpha_{32}}y_3 - \frac{\alpha_{12} \cdot \alpha_{33}}{\alpha_{32}}x_3 - \frac{\alpha_{12} \cdot \alpha_{34}}{\alpha_{32}}x_4 + \alpha_{13}x_3 + \alpha_{14}x_4\right\} =$$

$$= \left(b_1 - \frac{\alpha_{12} \cdot b_3}{\alpha_{32}}\right) - \left[\left(\alpha_{11} - \frac{\alpha_{12} \cdot \alpha_{31}}{\alpha_{32}}\right)x_1 - \left(\frac{\alpha_{12}}{\alpha_{32}}\right)y_3 + \left(\alpha_{13} - \frac{\alpha_{12} \cdot \alpha_{33}}{\alpha_{32}}\right)x_3 + \right.$$

$$\left. + \left(\alpha_{14} - \frac{\alpha_{12} \cdot \alpha_{34}}{\alpha_{32}}\right)x_4\right].$$

Аналогичным образом преобразуются и остальные строки. Полученные результаты занесем в таблицу, в которой замена $x_2 \leftrightarrow y_3$ произведена:

	Своб. член	x_1	y_3	x_3	x_4
y_1	$b_1 - \frac{\alpha_{12} \cdot b_3}{\alpha_{32}}$	$\alpha_{11} - \frac{\alpha_{12} \cdot \alpha_{31}}{\alpha_{32}}$	$-\frac{\alpha_{12}}{\alpha_{32}}$	$\alpha_{13} - \frac{\alpha_{12} \cdot \alpha_{33}}{\alpha_{32}}$	$\alpha_{14} - \frac{\alpha_{12} \cdot \alpha_{34}}{\alpha_{32}}$
y_2	$b_2 - \frac{\alpha_{22} \cdot b_3}{\alpha_{32}}$	$\alpha_{21} - \frac{\alpha_{22} \cdot \alpha_{31}}{\alpha_{32}}$	$-\frac{\alpha_{22}}{\alpha_{32}}$	$\alpha_{23} - \frac{\alpha_{22} \cdot \alpha_{33}}{\alpha_{32}}$	$\alpha_{24} - \frac{\alpha_{22} \cdot \alpha_{34}}{\alpha_{32}}$
x_2	$\frac{b_3}{\alpha_{32}}$	$\frac{\alpha_{31}}{\alpha_{32}}$	$\frac{1}{\alpha_{32}}$	$\frac{\alpha_{33}}{\alpha_{32}}$	$\frac{\alpha_{34}}{\alpha_{32}}$

Теперь можно сформулировать алгоритм преобразования коэффициентов стандартной таблицы:

1) в новой таблице разрешающий элемент заменяется на обратную ему величину;

2) вместо коэффициентов разрешающей строки старой таблицы проставляются коэффициенты, численно равные частному от деления прежних элементов на разрешающий элемент;

3) вместо коэффициентов разрешающего столбца старой таблицы заносятся коэффициенты численно равные частному от деления прежних элементов, взятых с противоположным знаком, на разрешающий элемент;

4) в остальные ячейки заносятся величины, численно равные алгебраической сумме коэффициента старой таблицы, и произведения элемента, расположенного в новой таблице в той же строке, что и первое слагаемое суммы, но из столбца, соответствующего разрешающему в старой таблице, на коэффициент разрешающей строки старой таблицы, расположенный в том же столбце, что и первое слагаемое суммы. Словесной изящностью последний пункт не отличается, хотя на самом деле преобразование довольно простое.

Пример 7

В системе уравнений-ограничений ОЗЛП осуществить замену $x_1 \leftrightarrow y_2$:

$$\begin{cases} y_1 = x_1 - x_2 + 2x_3 - 5, \\ y_2 = 2x_1 - x_2 + 1, \\ y_3 = 2x_2 - x_3 - 1, \\ y_4 = -x_1 - x_3 + 2. \end{cases} \qquad (2.49)$$

Решение

Запишем систему (2.49) в стандартной форме:

$$\begin{cases} y_1 = -5 - (-x_1 + x_2 - 2x_3), \\ y_2 = 1 - (-2x_1 + x_2), \\ y_3 = -1 - (-2x_2 + x_3), \\ y_4 = 2 - (x_1 + x_3). \end{cases} \qquad (2.50)$$

Внесем коэффициенты системы (2.50) в таблицу:

$y_2 \updownarrow$

$x_1 \leftrightarrow$ (строка y_2)

	Своб. член	$\boldsymbol{x_1}$	$\boldsymbol{x_2}$	$\boldsymbol{x_3}$
$\boldsymbol{y_1}$	–5	–1	1	–2
$\boldsymbol{y_2}$	1	$\underline{\boldsymbol{2}}$	1	0
$\boldsymbol{y_3}$	–1	0	–2	1
$\boldsymbol{y_4}$	2	1	0	1

Разрешающий коэффициент для замены $x_1 \leftrightarrow y_2$ расположен на пересечении строки y_2 и столбца x_1: $\alpha_{21} = -2$. X_1 – разрешающий столбец, y_2 – разрешающая строка. Строим таблицу:

	Своб. член	$\boldsymbol{y_1}$	$\boldsymbol{x_2}$	$\boldsymbol{x_3}$
$\boldsymbol{y_1}$	$-4/2$	$-1/2$	$-1/2$	–2
$\boldsymbol{x_1}$	$-1/2$	$-1/2$	$-1/2$	0
$\boldsymbol{y_3}$	–1	0	–2	1
$\boldsymbol{y_4}$	$5/2$	$1/2$	$1/2$	1

«Переразрешение» системы (2.50) $x_1 \leftrightarrow y_2$ завершено. Аналогичным способом можно сделать любую замену $x_j \leftrightarrow y_i$. В задаче ЛП кроме ограничений есть и целевая функция $W = c_0 + c_1x_1 + c_2x_2 + ... + c_nx_n$, которую в ОЗЛП необходимо минимизировать. Если производится замена $x_j \leftrightarrow y_i$, то это означает, что одна из свободных переменных x_r становится базисной, а одна из базисных переменных y_p – свободной. Если в целевой функции W $x_j, j = \overline{1,n}$ свободные переменные, то после замены $x_r \leftrightarrow y_p$ функция W примет вид

$$W = c_0^{'} + c_1^{'}x_1 + c_2^{'}x_2 + ... + c_p^{'}y_p + ... + c_n^{'}x_n.$$

Для такой замены переменных W полностью применим тот же алгоритм, что и для ограничений ОЗЛП. Приводим функцию W к стандартному виду:

$$W = c_0 - (\gamma_1x_1 + \gamma_2x_2 + ... + \gamma_nx_n),$$

где $\gamma_1 = -c, \gamma_2 = -c_2, ..., \gamma_n = -c_n$.

Вводим в стандартной таблице коэффициентов дополнительную строку W:

	Своб. член	x_1	x_2	…	x_n
W	c_0	γ_1	γ_2	…	γ_n
y_1	b_1	α_{11}	α_{12}	…	α_{1n}
…	…	…	…	…	…
y_m	b_m	α_{m1}	α_{m2}	…	α_{mn}

Преобразование коэффициентов производится точно так же, как и для других строк стандартной таблицы. Однако строка W имеет характерную особенность по сравнению с другими строками таблицы, а именно: в строке W разрешающий элемент выбирать нельзя!

Пример 8

Произвести замену переменных $x_1 \leftrightarrow y_2$ в ОЗЛП:

$$W = -x_1 + 2x_2 - x_3 + 1 \to \min,$$
$$y_1 = x_1 - x_2 + x_3 - 1,$$
$$y_2 = \tfrac{1}{2}x_1 - x_3 - 3,$$
$$y_3 = 3x_2 - 2x_3,$$
$$x_{1,2,3} \ge 0.$$

Преобразуем условия ОЗЛП в стандартную форму записи:

$$W = 1 - (x_1 - 2x_2 + x_3),$$
$$y_1 = -1 - (-x_1 + x_2 - x_3),$$
$$y_2 = -3 - (-\tfrac{1}{2}x_1 + x_3),$$
$$y_3 = 0 - (-3x_2 + 2x_3).$$

Заполним стандартную таблицу:

$y_2 \updownarrow$

		Своб. член	x_1	x_2	x_3
	W	1	1	–2	1
	y_1	–1	–1	1	–1
$x_1 \leftrightarrow$	y_2	–3	***–1/2***	0	1
	y_3	0	0	–3	2

Разрешающий элемент –1/2, разрешающая строка y_2, разрешающий столбец x_1. Строим новую таблицу:

	Своб. член	y_2	x_2	x_3
W	–5	2	–2	3
y_1	5	–2	1	–3
x_1	6	–2	0	–2
y_3	0	0	–3	2

Заполняем ячейки, используя алгоритм «переразрешения». Замена $x_1 \leftrightarrow y_2$ осуществлена.

Используя табличный алгоритм «переразрешения», можно решить любую задачу ЛП. Определение решения ОЗЛП состоит из двух этапов:

1. Отыскание опорного решения.
2. Поиск оптимального решения.

В процессе выполнения первого этапа выясняется, имеет ли данная ОЗЛП допустимые решения (имеет ли она вообще решение). Если задача имеет допустимые решения, то определяется опорное решение. На втором этапе путем перебора опорных решений выясняется, ограничена ли снизу целевая функция. Если целевая функция не ограничена, оптимального решения не существует. Если целевая функция ограничена, то после ряда замен $x_j \leftrightarrow y_i$ находится оптимальное решение.

2.8. Определение опорного решения ОЗЛП

Пусть ОЗЛП приведена к стандартной форме записи:

$$W = c_0 - (\gamma_1 x_1 + \gamma_2 x_2 + ... + \gamma_n x_n);$$

$$\begin{cases} y_1 = b_1 - (\alpha_{11} x_1 + \alpha_{12} x_2 + ... + \alpha_{1n} x_n), \\ y_2 = b_2 - (\alpha_{21} x_1 + \alpha_{22} x_2 + ... + \alpha_{2n} x_n), \\ .... \\ y_m = b_m - (\alpha_{m1} x_1 + \alpha_{m2} x_2 + ... + \alpha_{mn} x_n), \end{cases} \tag{2.51}$$

где $x_j, j = \overline{1,n}$ – свободные переменные; $y_i, i = \overline{1,m}$ – базисные. Ранее было отмечено, что в каждой опорной точке (вершине) ОДР по крайней мере n переменных из $n+m$ должны быть равны нулю.

Положим, $x_1 = x_2 = ... = x_n = 0$, тогда

$$y_1 = b_1; \; y_2 = b_2; \ldots; \; y_m = b_m. \tag{2.52}$$

Если в (2.52) все $b_i \geq 0, i = \overline{1,m}$, то опорное решение получено. Если некоторые $b_r < 0, r = \overline{1,p}$, то решение не только не является опорным, но и вообще недопустимо, т.к. $y_r < 0, r = \overline{1,p}$. Осуществляя замены $x_j \leftrightarrow y_i$, шаг за шагом мы или получим опорное решение, или докажем, что задача ЛП не имеет решения. Задача ЛП не имеет решения, когда система уравнений (2.51) не совместима с неравенствами $x_j \geq 0, j = \overline{1,n}$ и

$y_i \geq 0, i = \overline{1,m}$. Внешним признаком отсутствия допустимого решения является отсутствие отрицательных коэффициентов в какой-то строке y_r таблицы, где $b_r < 0$. Тогда

$$y_r = b_r - (\alpha_{r1}x_1 + \alpha_{r2}x_2 + ... + \alpha_{rn}x_n),$$

где $b_r < 0$ и $\alpha_{r1} \geq 0$, $\alpha_{r2} \geq 0, ...,$ $\alpha_{rn} \geq 0$, $x_j \geq 0$, $j = \overline{1,n}$. В этом случае y_r не может стать неотрицательной величиной.

Пример 9

Пусть $y_4 = -10 - (2x_1 + 2x_2 + 3x_3)$. При $x_1 = x_2 = x_3 = 0 \rightarrow y_4 = -10$.

Допустим, x_1 = 1, x_2 = 2, x_3 = 3. Тогда $y_4 = -10 - (2 + 4 + 9) = -25$, следовательно, опорного решения нет.

В случае, когда опорного решения нет, необходимо производить замены $x_j \leftrightarrow y_i$ так, чтобы после каждого «переразрешения» происходило приближение к границе ОДР, т.е. с каждым шагом число отрицательных базисных переменных должно убывать или, по крайней мере, убывали бы по абсолютной величине отрицательные свободные члены. Для замены $x_j \leftrightarrow y_i$ необходимо выбрать разрешающий элемент. Существует ряд способов выбора разрешающего элемента для наиболее быстрого движения к опорному решению. Суть одного из них заключается в следующем.

Пусть в одном из уравнений системы (2.51) свободный член $b_r < 0$. Рассматриваем в этой строке коэффициенты α_{rj} при свободных переменных $x_j, j = \overline{1,n}$. Напомним, что если все $\alpha_{rj} > 0, j = \overline{1,n}$, то задача ЛП не имеет решения.

Пусть $\alpha_{rp} < 0$. Выбираем столбец x_p в качестве разрешающего, где рассматриваем все коэффициенты $\alpha_{ip}, i = \overline{1,m}$, которые имеют одинаковые знаки (плюс или минус) с соответствующими свободными членами $b_i, i = \overline{1,m}$. В качестве разрешающего выбирается тот коэффициент, для которого отношение к нему соответствующего свободного члена минимально:

$$\alpha_{rp} \rightarrow \min \frac{\alpha_{ip}}{b_i}, i = \overline{1,m}.$$

Выбор разрешающего элемента однозначно определяет разрешающую строку. Далее применяется процедура «переразрешения», т.е. замена $y_r \leftrightarrow x_p$.

Пример 10

Определить (если оно существует) опорное решение ОЗЛП. Линейная функция W для простоты не приводится.

$$\begin{cases} y_1 = 1 - (-x_1 - 2x_2 + x_3), \\ y_2 = -5 - (-2x_1 + x_2 - x_3), \\ y_3 = 2 - (x_1 + x_2), \\ y_4 = 1 - (-x_2 + x_3). \end{cases}$$

Решение

Запишем условия задачи в виде стандартной таблицы, где $b_2 = -5 < 0$:

$y_3 \updownarrow$

	Своб. член	x_1	x_2	x_3
y_1	1	–1	–2	1
y_2	–5	–2	1	–1
$x_1 \leftrightarrow$ y_3	2	1	1	0
y_4	1	0	–1	1

В строке y_2 имеются $\alpha_{21} < 0, \alpha_{23} < 0$. Это означает, что ОЗЛП имеет решение (пока). Выбираем столбец x_1 ($\alpha_{21} = -2$) в качестве разрешающего и выбираем разрешающий элемент по правилу $\alpha_{rp} \to \min \dfrac{b_r}{\alpha_{rp}}, r = \overline{1,m}$:

$$\frac{-5}{-2} = 2{,}5 \to \frac{b_2}{\alpha_{21}},$$

$$\frac{2}{1} = 2 \to \frac{b_3}{\alpha_{31}},$$

$$2 < 2{,}5.$$

Разрешающий элемент $\alpha_{31} = 1$. Производим замену $x_1 \leftrightarrow y_3$ и получаем таблицу, где $b_2 < 0$ ($b_2 = -1$).

$\updownarrow y_2$

	Своб. член	y_3	x_2	x_3
y_1	3	1	-1	1
$x_3 \leftrightarrow$ y_2	-1	2	3	-1
x_1	2	1	1	0
y_4	1	0	-1	1

По абсолютной величине b_2 меньше, чем был первоначально abs (– 1) < abs (–5).

В строке y_2 с $b_2 = -1 < 0$ есть отрицательный коэффициент $\alpha_{23} = -1 < 0$. Столбец x_3 разрешающий.

Выбираем разрешающий элемент:

$$\frac{b_1}{\alpha_{13}} = \frac{3}{1} = 3; \frac{b_2}{\alpha_{23}} = \frac{-1}{-1} = 1; \frac{b_4}{\alpha_{43}} = \frac{1}{1} = 1.$$

Имеем два одинаковых отношения. Разрешающим выбираем $\alpha_{23} = -1$. Строка y_2 разрешающая. Производим замену $y_2 \leftrightarrow x_3$ и получаем таблицу:

	Своб. член	$\boldsymbol{y_3}$	$\boldsymbol{x_2}$	$\boldsymbol{y_2}$
$\boldsymbol{y_1}$	2	3	2	1
$\boldsymbol{x_3}$	1	–2	–3	–1
$\boldsymbol{x_1}$	2	1	1	0
$\boldsymbol{y_4}$	0	2	2	1

Опорное решение получено: $y_3 = x_2 = y_2 = 0, y_1 = 2, y_4 = 0, x_3 = 1, x_1 = 2.$

Пример 11

Найти опорное решение ОЗЛП (без строки W):

$$\begin{cases} y_1 = -4 - (-x_1 + 2x_2), \\ y_2 = -3 - (x_1 - x_2 + x_3), \\ y_3 = -10 - (2x_1 + x_2 + x_3), \\ y_4 = -2 - (-x_1 + x_2). \end{cases}$$

Решение

Запишем условия в стандартную таблицу:

$y_4 \updownarrow$

		Своб. член	$\boldsymbol{x_1}$	$\boldsymbol{x_2}$	$\boldsymbol{x_3}$
	$\boldsymbol{y_1}$	–4	–1	2	0
	$\boldsymbol{y_2}$	–3	1	–1	1
	$\boldsymbol{y_3}$	–10	2	–1	1
$x_1 \leftrightarrow$	$\boldsymbol{y_4}$	–2	–1	1	0

В строке y_1 $b_1 = -4 < 0$, опорного решения нет (пока). Коэффициент $\alpha_{11} = -1 < 0$. Значит столбец x_1 разрешающий. Определим разрешающий элемент:

$$\frac{b_1}{\alpha_{11}} = \frac{-4}{-1} = 4, \ \frac{b_4}{\alpha_{41}} = \frac{-2}{-1} = 2, \ 2 < 4,$$

Следовательно, $\alpha_{41} = -1$ – разрешающий элемент, строка y_4 разрешающая. Производим замену $x_1 \leftrightarrow y_4$ и получаем таблицу:

	Своб. член	y_4	x_2	x_3
y_1	–2	–1	1	0
y_2	–5	1	0	1
y_3	–14	2	1	1
x_1	2	–1	1	0

В строке y_3 $b_3 = -14 < 0$. Однако все $\alpha_{31}, \alpha_{32}, \alpha_{33} > 0$. Запишем $y_3 = -14 - (2x_1 + x_2 + x_3)$. Очевидно, что ни при каких x_1, x_2, x_3 > 0 y_3 не может стать нулевой или положительной величиной.

Вывод

Система ограничений не совместна, следовательно, ОЗЛП не имеет решений (допустимых решений не существует).

2.9. Поиск оптимального решения ОЗЛП

Займёмся отысканием такого опорного решения (первое получено), которое обращает в минимум линейную функцию

$$W = c_0 - (\gamma_1 x_1 + \gamma_2 x_2 + ... + \gamma_n x_n).$$

Покажем на примерах, как это делается.

Пример 12

Определить решение ОЗЛП. Пусть целевая функция $W = 0 - (-x_1 + x_2 + x_3)$ при ограничениях

$$\begin{cases} y_1 = 2 - (x_1 + x_2 - 2x_3), \\ y_2 = 1 - (x_1 - x_2 + x_3), \\ y_3 = 5 - (x_2 + x_3), \\ y_4 = 2 - (2x_1 - x_2), \\ x_{1,2,3,4} \geq 0. \end{cases}$$

Решение

Запишем условия ОЗЛП в виде стандартной таблицы:

$y_1 \updownarrow$

		Своб. член	x_1	x_2	x_3
	W	0	–1	1	–2
$x_2 \leftrightarrow$	y_1	2	1		–2
	y_2	1	1	–1	1
	y_3	5	0	1	1
	y_4	2	2	–1	0

Опорное решение достигнуто: $x_1 = x_2 = x_3 = 0,\, y_1 = 2,\, y_2 = 1,\, y_3 = 5,$ $y_4 = 2$, значение W = 0. Но оно не является оптимальным, т.к. увеличение x_2 в W приводит к уменьшению W против нуля в опорном решении.

Правила нахождения оптимального решения ОЗЛП:

1. Если все $b_i \geq 0$ в симплекс-таблице, а в строке W (не считая свободного члена) нет ни одного положительного элемента, то оптимальное решение достигнуто.

2. Если в строке W есть положительный коэффициент $\gamma_p > 0$, а в столбце P нет ни одного коэффициента $\alpha_{pi} > 0, i = \overline{1,m}$, то линейная функция W не ограничена снизу и оптимального решения ОЗЛП не существует.

3. Если в столбце $\gamma_p > 0$ имеются $\alpha_{pi} > 0, i = \overline{1,m}$, то необходимо произвести замену свободной переменной P на базисную Z, используя тот же принцип выбора разрешающего элемента, как при поиске опорного решения.

Из таблицы примера 12 видно, что столбец x_2 является разрешающим ($\gamma_2 = 1 > 0$).

Найдем разрешающий элемент:

$$\frac{b_1}{\alpha_{12}} = \frac{2}{1} = 2,$$

$$\frac{b_3}{\alpha_{32}} = \frac{5}{1} = 5,$$

$$2 < 5.$$

Разрешающим элементом является $\alpha_{12} = 1$, следовательно, разрешающая строка – y_1. Производим замену $y_1 \leftrightarrow x_2$, получаем таблицу:

	Своб. член	$\boldsymbol{x_1}$	$\boldsymbol{y_1}$	$\boldsymbol{x_3}$
$\boldsymbol{W}$	–2	–2	–1	0
$\boldsymbol{x_2}$	2	1	1	–2
$\boldsymbol{y_2}$	3	2	1	–1
$\boldsymbol{y_3}$	3	–1	–1	3
$\boldsymbol{y_4}$	4	3	1	–2

Оптимальное решение получено: $x_1^0 = x_3^0 = y_1^0 = 0, x_2^0 = 2, y_2^0 = 3, y_3^0 = 3,$ $y_4^0 = 4, W_{\min}^0 = -2.$

Вернёмся к п. 2 правил нахождения оптимального решения ОЗЛП. Если коэффициент $\gamma_p > 0$ в строке W, а в столбце P нет ни одного коэффициента $\alpha_{pi} > 0, i = \overline{1,m}$, то увеличение свободной переменной $x_p \geq 0$ приводит к неограниченному уменьшению линейной функции W.

Одновременно ни при каких $x_p \geq 0$ ни одна из базисных переменных не сможет принять нулевое численное значение. Например, $y_1 = 10 - (-x_1 - x_2)$. Любое увеличение x_1 или x_2 приводит к увеличению y_1.

2.10. «Вырожденный» случай ОЗЛП

«Вырожденным» случаем называют ситуацию, когда один или более свободных членов $b_i, i = \overline{1,m}$ в уравнениях-ограничениях при получении опорного решения равны нулю. Это означает, что в данном опорном решении равны нулю не только свободные, но и некоторые базисные переменные.

Пример 13

Решить ЗЛП:

$$W = 2x_1 - x_2 \to \min$$

при ограничениях

$$\begin{aligned} y_1 &= x_1 - x_2, \\ y_2 &= -x_2 + x_3 + 2, \\ y_3 &= x_3 + x_4 + 1, \\ x_{1,2,3,4} &\geq 0. \end{aligned}$$

Решение

Записываем условия примера в стандартной форме:

$$\begin{aligned} W &= 0 - (-2x_1 + x_2) \to \min, \\ y_1 &= 0 - (-x_1 + x_2), \\ y_2 &= 2 - (x_2 - x_3), \\ y_3 &= 1 - (-x_3 - x_4). \end{aligned}$$

Вводим коэффициенты в таблицу:

			$y_1 \updownarrow$		
	Своб. член	$\boldsymbol{x_1}$	$\boldsymbol{x_2}$	$\boldsymbol{x_3}$	$\boldsymbol{x_4}$
$\boldsymbol{W}$	0	–2	1	0	0
$x_2 \leftrightarrow$ $\boldsymbol{y_1}$	0	–1	1	0	0
$\boldsymbol{y_2}$	2	0	1	–1	0
$\boldsymbol{y_3}$	1	0	0	–1	–1

Столбец x_2 разрешающий, т.к. $\gamma_2 > 0$. Разрешающий элемент

$$\min\left\{\frac{b_1}{\alpha_{12}}; \frac{b_2}{\alpha_{22}}; \frac{b_3}{\alpha_{32}}\right\} = \min\left\{\frac{0}{1}; \frac{2}{1}; \frac{1}{0}\right\} = \frac{0}{1} \to \alpha_{12} = 1.$$

Разрешающая строка y_1. Делаем «переразрешение» $x_2 \leftrightarrow y_1$:

	Своб. член	x_1	y_1	x_3	x_4
W	0	–1	–1	0	0
x_2	0	–1	1	0	0
y_2	2	1	–1	–1	0
y_3	1	0	0	–1	–1

При замене $x_2 \leftrightarrow y_1$ уменьшения линейной функции не произошло. Переменную $x_2 = 0$ заменяем на переменную $y_1 = 0$ и получаем оптимальное решение:

$$x_1^0 = y_1^0 = x_3^0 = x_4^0 = 0, x_2^0 = 0, y_2^0 = 2, y_3^0 = 1, \ W_{\min}^0 = 0.$$

При наличии «вырождения» может оказаться, что замена $x_p \leftrightarrow y_r$ приводит только к перестановке переменных. Линейная функция W не уменьшается. В некоторых случаях последовательное применение правила выбора разрешающего элемента приводит к тому, что после нескольких итераций можно вернуться к исходной таблице. Это называется «зацикливанием», эффективным способом борьбы с которым является смена разрешающего элемента.

2.11. Двойственная задача линейного программирования

2.11.1. Понятие двойственности

С каждой задачей ЛП тесно связана другая задача, называемая двойственной. Первоначальная задача называется исходной (основной, прямой). Решение двойственной задачи может быть получено из решения исходной и наоборот.

Связь исходной и двойственной задач состоит в том, что коэффициенты $C_j, j = \overline{1,n}$ целевой функции исходной задачи являются свободными членами в системе ограничений двойственной задачи, свободные члены $b_i, i = \overline{1,m}$ системы ограничений исходной задачи служат коэффициентами функции цели двойственной задачи. Матрица коэффициентов системы ограничений двойственной задачи представлена как транспонированная матрица коэффициентов системы ограничений исходной задачи.

Рассмотрим постановку исходной и двойственной задач на примере задачи использования ресурсов.

Предприятие имеет m видов ресурсов в количестве $b_i, i = \overline{1,m}$ единиц. Из этих m видов ресурсов производится n видов продукции. Для производства единицы j-го вида продукции используется a_{ij} единиц i-го вида ресурса. Стоимость единицы j-го вида продукции – $C_j, \ j = \overline{1,n}$.

Следует рассчитать программу выпуска продукции, которая обеспечила бы её максимальный выпуск в стоимостном выражении.

Обозначим как $x_j \geq 0, j = \overline{1,n}$ количество единиц j-го вида продукции. Тогда решение задачи заключается в нахождении вектора $\overline{x}^T = (x_1, x_2, ..., x_n)$, который удовлетворяет ограничениям

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n \leq b_1, \\ a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n \leq b_2, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ a_{m1}x_1 + a_{m2}x_2 + ... + a_{mn}x_n \leq b_m, \\ x_j \geq 0, \ j = \overline{1,n} \end{cases} \tag{2.53}$$

и приводит к максимуму линейную функцию

$$W = c_1x_1 + c_2x_2 + \ldots + c_nx_n. \tag{2.54}$$

Оценим ресурсы, необходимые для изготовления продукции. Обозначим как $z_i \geq 0, i = \overline{1,m}$ стоимость единицы i-го вида ресурса. Тогда стоимость всех видов ресурсов, затраченных на изготовление единицы продукции j-го вида, равна $\sum_{i=1}^{m} a_{ij}z_i$. Стоимость затраченных ресурсов не может быть меньше стоимости единицы готовой продукции (для нормального производства), поэтому должно выполняться неравенство $\sum_{i=1}^{m} a_{ij}z_i \geq C_j, j = \overline{1,n}$. Стоимость всех имеющихся ресурсов есть величина $\sum_{i=1}^{m} b_i z_i$. Тогда решение двойственной задачи заключается в нахождении вектора $\overline{Z}^T = (z_1, z_2, ..., z_m)$, который удовлетворяет ограничениям

$$\begin{cases} a_{11}z_1 + a_{21}z_2 + ... + a_{m1}z_m \geq C_1, \\ a_{12}z_1 + a_{22}z_2 + ... + a_{m2}z_m \geq C_2, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ a_{1n}z_1 + a_{2n}z_2 + ... + a_{mn}z_m \geq C_n, \\ z_i \geq 0, i = \overline{1,m} \end{cases} \tag{2.55}$$

и приводит к минимуму линейную функцию

$$F = b_1z_1 + b_2z_2 + \ldots + b_mz_m. \tag{2.56}$$

Рассмотренные исходная и двойственная задачи могут быть интерпретированы с экономической точки зрения следующим образом.

Исходная задача

Сколько единиц и какого вида продукции $x_j \geq 0, j = \overline{1,n}$ необходимо произвести предприятию, чтобы при заданных стоимостях $C_j, j = \overline{1,n}$ единицы продукции и размерах имеющихся ресурсов, равных продукции

$b_i, i=\overline{1,m}$ единиц, максимизировать выпуск продукции в стоимостном выражении.

Двойственная задача

Какова должна быть цена единицы каждого из m ресурсов, чтобы при заданных количествах ресурсов $b_i, i=\overline{1,m}$ и величинах стоимости единицы продукции j-го вида $C_j, j=\overline{1,n}$ минимизировать общую стоимость затрат?

Переменные $z_i \geq 0, i=\overline{1,m}$ двойственной задачи называются оценками, или учётными, неявными, теневыми ценами.

Многие задачи ЛП первоначально ставятся в виде исходных или двойственных задач, поэтому говорят о паре двойственных задач ЛП.

2.11.2. Виды математических моделей двойственных задач ЛП

Математические модели пары двойственных задач могут иметь один из видов:

Тип задачи	Исходная задача	Двойственная задача
Несимметричные задачи	$W_{\min} = \overline{C}^T \overline{x}$ $A\overline{x} = \overline{B}$ $\overline{x} \geq 0$	$F_{\max} = \overline{B}^T \overline{z}$ $A^T \overline{z} \leq \overline{C}$ $\overline{z}$ не ограничен по знаку
	$W_{\max} = \overline{C}^T \overline{x}$ $A\overline{x} = \overline{B}$ $\overline{x} \geq 0$	$F_{\min} = \overline{B}^T \overline{z}$ $A^T \overline{z} \geq \overline{C}$ $\overline{z}$ не ограничен по знаку
Симметричные задачи	$W_{\min} = \overline{C}^T \overline{x}$ $A\overline{x} \geq \overline{B}$ $\overline{x} \geq 0$	$F_{\max} = \overline{B}^T \overline{z}$ $A^T \overline{z} \leq \overline{C}$ $\overline{z} \geq 0$
	$W_{\max} = \overline{C}^T \overline{x}$ $A\overline{x} \leq \overline{B}$ $\overline{x} \geq 0$	$F_{\min} = \overline{B}^T \overline{z}$ $A^T \overline{z} \geq \overline{C}$ $\overline{z} \geq 0$

Приведем без доказательства теоремы двойственности.

Теорема 1

Если исходная и двойственная задачи ЛП имеют опорные решения, то:

1) существует оптимальное решение $x_j^0, j=\overline{1,n}$ исходной задачи;
2) оптимальное решение $z_i^0, i=\overline{1,m}$ двойственной задачи;
3) имеет место соотношение

$$\sum_{j=1}^{n} C_j x_j^0 = \sum_{i=1}^{m} b_i z_i^0 , \tag{2.57}$$

т.е. max W = min F.

Теорема 2

Если линейная функция цели любой из пары задач не ограничена (ЗЛП не имеет оптимального решения), то другая задача не имеет решения.

Теорема 3 (о дополнительной нежёсткости)

Если $x_j, j=\overline{1,n}$ – решение исходной задачи ЛП, а $z_i, i=\overline{1,m}$ – решение двойственной задачи, то оба решения являются оптимальными тогда и только тогда, когда выполняются соотношения

$$z_i^0(\sum_{j=1}^{n} a_{ij}x_j^0 - b_i) = 0, i = \overline{1,m}, \tag{2.58}$$

$$x_j^0(\sum_{i=1}^{m} a_{ij}z_i^0 - C_i) = 0,\ j = \overline{1,n}. \tag{2.59}$$

Если при подстановке компонент оптимального решения x_j в систему ограничений исходной задачи i-е ограничение обращается в неравенство, то $z_i^0 = 0$, а если $z_i^0 > 0$, то i-е ограничение исходной задачи удовлетворяется оптимальным решением x_j^0 как строгое равенство.

2.12. Поиск оптимального решения двойственной задачи

Решение двойственной задачи (оптимальное, если оно существует) содержится в последней симплекс-таблице, полученной при решении исходной задачи ЛП.

Пример 14

Найти min $W = -x_1 + 3x_2 + 2x_3$ при ограничениях

$$\begin{cases} x_1 + x_2 + 2x_3 \ge -5, \\ 2x_1 - 3x_2 + x_3 \le 3, \\ 2x_1 - 5x_2 + 6x_3 \le 5, \\ x_{1,2,3} \ge 0. \end{cases}$$

Перепишем условия исходной задачи в виде

$$W = -x_1 + 3x_2 + 2x_3 \to \min$$

при

$$\begin{cases} x_1 + x_2 + 2x_3 \ge -5 - z_1, \\ -2x_1 + 3x_2 - x_3 \le -3 - z_2, \\ -2x_1 + 5x_2 - 6x_3 \le -5 - z_3, \\ x_{1,2,3} \ge 0. \end{cases}$$

Тогда двойственная задача будет иметь вид

$$F = -5z_1 - 3z_2 - 5z_3 \to \max\ ;$$

$$\begin{cases} z_1 - 2z_2 - 2z_3 \le -1 \quad v_1 \ge 0, \\ z_1 + 3z_2 + 5z_3 \le 3 \quad v_2 \ge 0, \\ 2z_1 - z_2 - 6z_3 \le 2 \quad v_3 \ge 0, \\ z_{1,2,3} \ge 0. \end{cases}$$

Решим исходную задачу:

$$W = 0 - (x_1 - 3x_2 - 2x_3) \to \min,$$

$$\begin{cases} y_1 = 5 - (-x_1 - x_2 - 2x_3), \\ y_2 = 3 - (2x_1 - 3x_2 + x_3), \\ y_3 = 5 - (2x_1 - 5x_2 + 6x_3). \end{cases}$$

Заносим коэффициенты в таблицу:

$y_2 \updownarrow$

		Своб. член	x_1	x_2	x_3
	W	0	1	–3	–2
	y_1	5	–1	–1	–2
$x_1\leftrightarrow$	y_2	3	2	–3	1
	y_3	5	2	–5	6

Опорное решение достигнуто: $x_1 = x_2 = x_3 = 0$, $y_1 = 5$, $y_2 = 3$, $y_3 = 5$, а оптимальное нет, т.к. $\gamma_1 = 1 > 0$.

Столбец x_1 разрешающий, строка y_2 разрешающая, т.к. $\min\left\{\frac{3}{2};\frac{5}{2}\right\}=\frac{3}{2}$, $\alpha_{21} = 2$ – разрешающий элемент. Делаем замену $x_1 \leftrightarrow y_2$, результаты заносим в таблицу:

	Своб. член	y_2	x_2	x_3
W	–3/2	–1/2	–3/2	–5/2
y_1	13/2	1/2	–5/2	–3/2
x_1	3/2	1/2	–3/2	1/2
y_3	2	–1	–2	5

Оптимальное решение исходной задачи получено: $W^0_{\min} = -3/2$ при $x_1^0 = 3/2, x_2^0 = 0, x_3^0 = 0, y_1^0 = 13/2, y_2^0 = 0, y_3^0 = 2.$

Решим симплекс-методом двойственную задачу:

$$F = 0 - (5z_1 + 3z_2 + 5z_3) \to \max$$

при

$$\begin{cases} v_1 = -1 - (z_1 - 2z_2 - 2z_3), \\ v_2 = 3 - (z_1 + 3z_2 + 5z_3), \\ v_3 = 2 - (2z_1 - z_2 - 6z_3). \end{cases}$$

Заносим результаты в таблицу:

$v_1 \updownarrow$

		Своб. член	z_1	z_2 -2	z_3
	F	0	5	3	5
$z_2 \leftrightarrow$	v_1	–1	1		–2
	v_2	3	1	3	5
	v_3	2	2	–1	–6

Опорного решения нет, столбец z_2 разрешающий, строка v_1 разрешающая, $\alpha_{12} = -2$ – разрешающий элемент.

Делаем замену $z_2 \leftrightarrow v_1$ и получаем таблицу:

	Своб. член	z_1	v_1	z_3
F	–3/2	13/2	3/2	2
z_2	1/2	–1/2	–1/2	1
v_2	3/2	5/2	3/2	2
v_3	5/2	3/2	–1/2	–5

Оптимальное в смысле max *F* решение получено: $F_{\max}^0 = -3/2$ при $z_1^0 = 0, z_2^0 = 1/2, z_3^0 = 0, v_1^0 = 0, v_2^0 = 3/2, v_3^0 = 5/2.$

Проанализировав решения исходной и двойственной задач, имеем:

1) min *W* = max *F* = –3/2;

2) оптимальные значения переменных двойственной задачи численно равны взятым по абсолютной величине коэффициентам в оптимальном решении при тех дополнительных переменных (базисных), которые перешли в разряд свободных.

Если соответствующая z_i дополнительная переменная y_i осталась базисной в оптимальном решении, это значит, что данная $z_i^0 = 0$.

Двойственная задача ЛП необходима при анализе решения основной (исходной) задачи на чувствительность. Но эта проблема выходит за рамки данного пособия. Некоторое представление об анализе на чувствительность оптимального решения ЗЛП может быть получено из [9].

Библиографический список

1. Клир, Дж. Системология. Автоматизация решения системных задач: пер. с англ / Дж. Клир. М.: Радио и связь, 1990. 544 с.

2. Герасимов, И.Г. Структура научного исследования: (философ. анализ познавательной деятельности в науке) / И.Г. Герасимов. М.: Мысль, 1985. 215 с.

3. Месарович, М. Теория иерархических многоуровневых систем / М. Месарович, Д. Мако, И. Такахара. М.: Мир, 1973. 344 с.

4. Перегузов, Ф.И. Введение в системный анализ / Ф.И. Перегузов, Ф.П. Тарасенко. М.: Высшая школа, 1989. 367 с.

5. Горелик, Л.Л. Методы распознавания: учебное пособие / Л.Л. Горелик. В.А. Скрипкин. 2-е изд., перераб. и доп. М.: Высшая школа, 1984. 208с.

6. Дуда, Р. Распознавание образов и анализ сцен: пер. с англ. / Р. Дуда, П. Харт. М.: Мир, 1976. 512 с.

7. Мандель, И.Д. Кластерный анализ / И.Д. Мандель. М.: Финансы и статистика, 1988. 176 с.

8. Дюран, Б. Кластерный анализ: пер. с англ / Б.Дюран, П. Одел. М.: Статистика, 1977. 128 с.

9. Вагнер, Г. Основы исследования операций. В 3-х т. / Г. Вагнер. М.: Мир, 1972. Т. 1. 336 с.

10. Вентцель, Е.С. Исследование операций / Е.С. Вентцель. М.: Советское радио, 1972. 552 с.

Оглавление