МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ

Учреждение образования
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИНФОРМАТИКИ И РАДИОЭЛЕКТРОНИКИ

Кафедра информатики

Е.С. Лукин

Прикладная теория информации

Учебное пособие
для студентов специальности «Информатика»

Минск 2002

УДК 621.391.1(075)
ББК 32.811 я 73
Л 84

Рецензент
проректор Высшего государственного колледжа связи,
канд. физ.-мат. наук, доц. В.Н. Теслюк

Лукин Е.С.

Л 84 Прикладная теория информации: Учеб. пособие для студентов специальности «Информатика». — Мн.: БГУИР. 2002. — 42 с.: ил.

ISBN 985-444-428-7

Учебное пособие содержит математический аппарат, применяемый для описания детерминированных и случайных сигналов; принципы дискретизации и квантования сигналов. Описаны методы количественной оценки информации. Даны примеры применения принципов теории информации к созданию и использованию методов и средств кодирования информации.

УДК 621.391.1(075)
ББК 32.811 я 73

ISBN 985-444-428-7

Содержание

Введение

Теория информации как наука существует с середины XX века, с момента появления основополагающей работы — К. Шеннон «Математическая теория связи» (1948) — прошло около 55 лет. У Шеннона были предшественники, например, Р. Хартли, впервые предложивший в 1928 году количественную меру информации, или В.А. Котельников, сформулировавший в 1933 году важнейшую теорему о возможности представления непрерывной функции совокупностью ее значений в отдельных точках отсчета. Были и современники, и последователи, например А.Н. Колмогоров, внесший огромный вклад в статистическую теорию колебаний, являющуюся математической основой теории информации. Работы по развитию теории информации продолжаются и в настоящее время.

Теория информации быстро разделилась на фундаментальную и прикладную.

Фундаментальная теория информации, это:

— анализ сигналов как средства передачи сообщений и оценка переносимого «*количества информации*»;

— анализ информационных характеристик источников сообщений и каналов связи и обоснование принципиальной возможности кодирования и декодирования сообщений, обеспечивающих предельно допустимую скорость передачи сообщений по каналу связи как при отсутствии, так и при наличии помех.

Прикладные результаты приводятся здесь только для пояснения основ теории.

Прикладная теория информации (ПТИ) основывается на практических результатах, полученных при рассмотрении фундаментальных законов. Можно определить ПТИ двояко. Первое определение (узкое) — разработка конкретных методов и средств кодирования сообщений. Второе, более широкое: предметом теории информации является изучение любых процессов, связанных с получением, передачей, хранением, обработкой и использованием информации.

Второе определение затрагивает проблемы буквально всех наук (от математики до педагогики). Идеи теории информации широко используются в различных научных дисциплинах потому, что в основе своей эта теория — математическая. Основные ее понятия (энтропия, количество информации, пропускная способность) определяются только через вероятности событий, которым может быть приписано самое различное физическое содержание.

Целью настоящего учебного курса являются в основном задачи первого типа, т.е. вопросы кодирования.

Существует множество определений понятия ***информация***, от наиболее общего философского (информация есть отражение реального мира), до узкого практического (информация есть все сведения, являющиеся объектом хранения, передачи и преобразования). Имеется множество точек зрения на суть информации, одна из которых рассматривает информацию как некоторую философскую категорию, такую же общую, как понятия материи или энергии (Н. Винер).

Есть вопрос, на который трудно дать однозначный ответ**:** ***информация — это свойство некоторого объекта*** (системы) (В.М. Глушков, А.Н. Колмогоров, У. Эшби) ***или информация появляется только тогда, когда объект изучается неким разумным существом***. Существует ли информация независимо от того, воспринимается ли она, зависит ли ее восприятие от индивидуальных способностей воспринимающего?

Противоречие частично снимается, если рассматривать информацию как некое потенциальное свойство объекта (системы). Так, беря в руки книгу, можно извлечь из нее какую-то информацию, но книгу можно использовать разными способами, в том числе и исключающими получение информации.

Отдельные философы понимают под словом информация только то, что воспринято и осмыслено, то, что служит для управления объектами или процессами.

Информация существует только в форме материально-энергетических сигналов. Информацию, представленную в формализованном виде, позволяющем осуществить ее обработку с помощью технических средств, называют ***данными***.

Этапы обращения информации. Роль информации может ограничиваться неопределенным эмоциональным воздействием на человека, но в чисто технических (автоматических) и человеко-машинных (автоматизированных) системах она чаще всего используется для выработки управляющих воздействий. При обращении информации в системах можно выделить отдельные этапы.

Этап восприятия информации: осуществляется извлечение и анализ информации об объекте (процессе) и формирование образа объекта, проводится его опознание и оценка. Полезный сигнал отделяется от ***шума***, т.е. мешающей информации. Происходит ***выявление*** или ***измерение*** полезного сигнала.

Этап подготовки информации: проводятся операции нормализации, аналого-цифрового преобразования, шифрование. В результате восприятия и подготовки получается сигнал в форме, удобной для передачи или обработки.

Этап передачи и хранения: информация пересылается либо из одного места в другое, либо от одного момента времени до другого. Поскольку задачи, возникающие на этих этапах, близки друг другу, хранение информации часто в самостоятельный этап не выделяется. Для передачи на расстояние используются каналы различной физической природы. Для хранения используются магнитные и другие носители. Извлечение сигнала на выходе канала, подверженного действию шумов, носит характер вторичного восприятия.

Этап обработки информации: выявляются ее взаимозависимости, представляющие интерес для системы. Формализуемый процесс обработки может выполняться техническими средствами без участия человека. В системах управления целью обработки является решение задачи выбора управляющих воздействий (этап принятия решения).

Этап отображения информации: должен предшествовать этапам, связанным с участием человека. Цель — предоставить человеку информацию в форме, доступной для его органов чувств.

Этап воздействия: информация используется для осуществления необходимых изменений в системе.

Информационные системы. Совокупность средств информационной техники и людей, объединенных для достижения определенных целей, называют информационной системой. Системы бывают автоматическими (например, телефонные АТС) или автоматизированными (человеко-машинными).

Автоматизированная информационная система (АИС) становится автоматизированной системой управления (АСУ), если входящая в систему информация извлекается из какого-либо объекта, а выходящая — используется для изменения состояния того же объекта.

Большинство АИС и АСУ являются локальными, т.е. системами ограниченных размеров – от устройства (например, видеокамера) до размеров предприятия. Однако сегодня такие системы интегрируются и взаимодействуют на региональном и глобальном уровнях.

Системы становятся территориально рассредоточенными, иерархичными по функциям. Обеспечение взаимодействия рассредоточенных систем требует протяженных высокоскоростных и надежных каналов связи, а большой объем информации — ЭВМ высокой производительности. Развитие таких систем уменьшает роль телефона, телеграфа, почты и т.д.

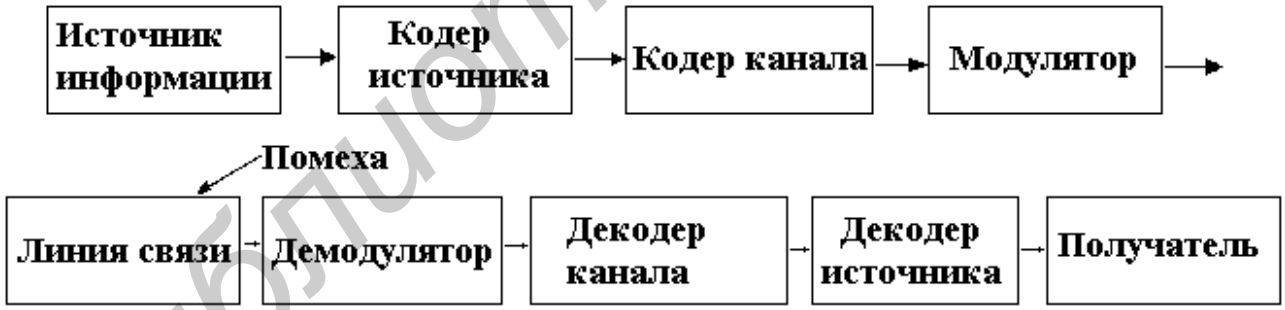

Схема системы передачи информации

Информация поступает в АИС в форме сообщений. *Сообщением* называют совокупность знаков или первичных сигналов, содержащих информацию. Источник сообщений образуют источник информации (ИИ) (исследуемый или наблюдаемый объект) и первичный преобразователь (ПП) (датчик, человек-оператор и т.п.), воспринимающий информацию о его состоянии или протекающем процессе. Различают дискретные и непрерывные сообщения.

Дискретные сообщения формируются в результате последовательной выдачи источником отдельных элементов — знаков. Множество различных знаков

называют *алфавитом источника сообщений*, а число знаков — *объемом алфавита*. В частности, знаками могут быть буквы языка.

Непрерывные сообщения не разделены на элементы. Они описываются непрерывными функциями времени (речь, TV и т.д.).

Преобразование сообщения в сигнал, удобный для передачи по данному каналу связи, называют *кодированием в широком смысле слова*. Операцию восстановления сообщения по принятому сигналу называют *декодированием*.

Под *линией связи* понимают любую физическую среду (воздух, металл, магнитную ленту и т.п.), обеспечивающую поступление сигналов от передающего устройства к приемному. Сигналы на выходе линии связи могут отличаться от переданных вследствие затухания, искажения и воздействия помех. *Помехами* называют любые мешающие возмущения, как внешние (атмосферные, промышленные помехи), так и внутренние (источник — аппаратура связи), вызывающие отклонения принятых сигналов от переданных. Эффект воздействия помех на различные блоки системы стараются учесть изменением характеристик линии связи. Поэтому источник помех условно относят к линии связи.

Из смеси сигнала и помехи *приемное устройство* выделяет сигнал и посредством декодера восстанавливает сообщение, которое в общем случае может отличаться от посланного. Меру соответствия принятого сообщения посланному называют *верностью передачи*. Обеспечение заданной верности передачи сообщений — важнейшая цель системы связи.

Принятое сообщение с выхода системы связи поступает к абоненту-получателю, которому была адресована исходная информация.

Совокупность средств, предназначенных для передачи сообщений, называют *каналом связи*. Для передачи информации от группы источников, сосредоточенных в одном месте, к группе получателей, расположенном в другом, часто используют только одну линию связи, организовав на ней требуемое количество каналов. Такие системы называют *многоканальными*.

Уровни проблем передачи информации. Обмен информацией предполагает использование некоторой системы знаков, например, естественного или искусственного (формального) языка. Информация о непрерывных процессах также может быть выражена посредством знаков. Изучение знаковых систем наукой о знаках и языках (семиотикой) проводится по крайней мере на трех уровнях.

На ***синтаксическом уровне*** рассматривают внутренние свойства текстов, т.е. отношения между знаками, отражающие структуру данной знаковой системы. Внешние свойства текстов изучают на семантическом и прагматическом уровнях.

На ***семантическом уровне*** анализируют отношения между знаками и обозначаемыми ими предметами, действиями, качествами, т.е. смысловое содержание текста, его отношение к источнику информации.

На ***прагматическом уровне*** рассматривают отношения между текстом и теми, кто его использует, т.е. потребительское содержание текста.

Решение проблем синтаксического уровня помогает создать теоретические основы построения систем связи с показателями работы, близкими к предельно возможным. Это чисто технические проблемы совершенствования методов передачи сообщений и их материального воплощения — сигналов, проблемы доставки получателю сообщений как совокупности знаков, при этом полностью абстрагируется их смысловое и прагматическое содержание.

Прикладная теория информации решает проблемы именно этого, синтаксического, уровня. Она опирается на понятие «количество информации», являющееся мерой частоты употребления знаков, которая никак не отражает ни смысла, ни важности передаваемых сообщений.

На семантическом уровне формализуют смысл передаваемой информации, судят о близости информации к истине, оценивают ее качество. Эти проблемы чрезвычайно сложны, т.к. смысловое содержание информации больше зависит от получателя, чем от семантики сообщения. Одно и то же сообщение может быть понято по-разному, поскольку зависит от личности и уровня знаний лица, эту информацию получившего. Мы еще не умеем измерять семантическую информацию.

На прагматическом уровне интересуют последствия от получения информации абонентом. Потребительская ценность полученной информации различна для разных получателей. Имеет значение скорость доставки, наличие реального масштаба времени для обмена информацией и др. Ряд проблем этого уровня решается. Предложены меры оценки потребительской информации, ведутся исследования в области старения информации, т.е. потери ее ценности в процессе доставки, и т.д.

1. Сигналы

Для передачи сообщения по каналу связи ему необходимо поставить в соответствие определенный сигнал. В информационных системах под *сигналом* понимают физический процесс, отображающий (несущий) сообщение. Для передачи дискретных сообщений иногда применяют принцип: одно сообщение — один сигнал, специальный для данного сообщения. Однако в такой системе невозможно передать произвольное сообщение. Применение *знаков* дает возможность передавать бесконечно большой объем информации конечным числом образцовых сигналов, соответствующих алфавиту источника.

Для надежности распознавания образцовых сигналов их число сокращают до минимума. Поэтому представляют исходные знаки в другом алфавите с меньшим числом знаков, называемых *символами*. При обозначении этой операции используется тот же термин «*кодирование*», рассматриваемый в узком смысле. Устройство, выполняющее такую операцию, называют кодирующим, или *кодером*. Так как алфавит символов меньше алфавита знаков, то каждому знаку соответствует некоторая последовательность символов, которую назовем

кодовой комбинацией. Число символов в кодовой комбинации называют ее *значностью*, число ненулевых символов — *весом*.

Для операции сопоставления символов со знаками исходного алфавита используется термин «*декодирование*». Техническая реализация осуществляется декодирующим устройством, или *декодером*. В простейшем случае кодирующее и декодирующее устройства могут отсутствовать.

Передающее устройство преобразовывает непрерывные сообщения или знаки в сигналы, удобные для прохождения по линии связи (хранения). При этом один или несколько параметров выбранного носителя изменяют в соответствии с передаваемой информацией. Такой процесс называют *модуляцией*. Он осуществляется *модулятором*. Обратное преобразование сигналов в символы производится *демодулятором*.

Виды модуляции

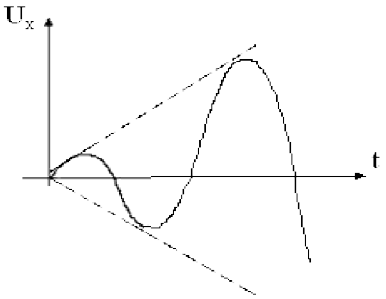

Рис. 1.1. Амплитудная модуляция

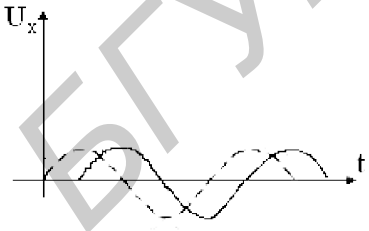

Рис. 1.2. Фазовая модуляция

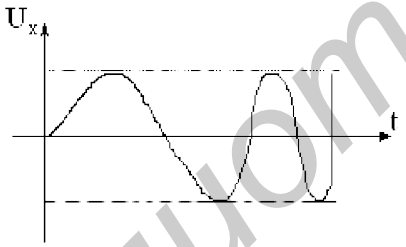

Рис. 1.3. Частотная модуляция

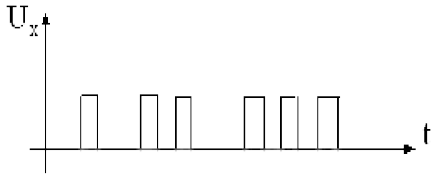

Рис. 1.4. Импульсная модуляция

Существуют и другие виды модуляции: амплитудно-импульсная, частотно-сдвиговая, сдвига фазы, спектральная и т.д.

Как носители информации используются колебания различной природы, чаще всего гармонические, включая частный случай — постоянное состояние (частота $\omega = 0$). В технических информационных системах получили распространение носители в виде электрического напряжения или тока. Говоря о модели сигнала, будем иметь в виду, в первую очередь, электрический сигнал.

При использовании гармонических электрических колебаний информативными могут стать такие параметры, как амплитуда, частота, фаза. Колебания принято подразделять на детерминированные и случайные.

Детерминированными называют колебания, которые точно определены в любые моменты времени.

Для *случайных* колебаний невозможно предсказать их параметры. Такие колебания могут быть как сигналами, так и помехами.

При изучении свойств каналов связи, сигналов и помех мы отвлекаемся от их конкретной физической природы, содержания и назначения, заменяя моделями. *Модель* — это выбранный способ описания объекта, отражающий существенные с точки зрения решаемой задачи факторы. Мы будем изучать математические модели, т.к. этот класс моделей позволяет получать количественные характеристики изучаемых процессов.

Детерминированным называется сигнал, который нет смысла передавать, т.е. сигнал, у которого можно точно предсказать изменения значения информативного параметра. (Несущий информацию сигнал обязан быть случайным). Результаты анализа детерминированных сигналов являются основой для изучения более сложных случайных сигналов. Иногда детерминированные сигналы имеют и самостоятельное значение. Они создаются для измерения, отладки систем информационной техники, выполняя роль эталонов.

Формы представления детерминированных сигналов

Сигналы подразделяют на *дискретные*, *непрерывные* и *дискретно-непрерывные*.

Сигнал считают *дискретным* по данному параметру, если число значений, которое может принимать этот параметр, конечно (или счетно). Если множество возможных значений параметра образует континуум, то сигнал считают *непрерывным*. Сигнал, дискретный по одному параметру и непрерывный по другому, называют *дискретно-непрерывным*.

Как математическая модель используются:

— непрерывная функция непрерывного аргумента (например, времени);

— непрерывная функция дискретного аргумента, например функция, значения которой отсчитывают только в определенные моменты времени;

— дискретная функция непрерывного аргумента, например функция времени, квантованная по уровню;

— дискретная функция дискретного аргумента, принимающая одно из конечного множества возможных значений в дискретные моменты времени.

Часто модели сигналов представляются как совокупности элементарных (базисных) функций по времени. Это вызвано тем, что на модели исследуется прохождение реальных сигналов через интересующие исследователей системы. Чаще всего исследуются инвариантные по времени линейные системы.

При анализе прохождения сложного сигнала $U(t)$ через такие системы его представляют в виде взвешенной суммы базисных функций $\varphi_k(t)$ (или соответствующего ей интеграла):

$$U(t) = \sum C_k \varphi_k(t); \qquad t \subseteq [t_1, t_2]. \tag{1.1}$$

При выбранном наборе базисных функций сигнал $U(t)$ полностью определяется совокупностью безразмерных коэффициентов C_k. Такие совокупности чисел называют *дискретными спектрами сигналов*. На интервале $[t_1, t_2]$ (1.1) справедливо как для сигналов, неограниченных по времени, так и для сигналов конечной длительности. За пределами $[t_1, t_2]$ сигнал конечной длительности может периодически повторяться, поэтому и там он не равен нулю в общем случае. Для любого момента времени ограниченный по времени сигнал может быть представлен:

$$U(t) = \int_{-\infty}^{\infty} S(\alpha)\varphi(\alpha, t)d\alpha, \qquad (1.2)$$

где $\varphi(\alpha, t)$ — базисная функция с непрерывно изменяющимся параметром α. В этом случае имеется непрерывный (сплошной) спектр сигнала, который представляется спектральной плотностью $S(\alpha)$. Размерность ее обратна размерности α. Аналогом безразмерного коэффициента C_k здесь является величина $S(\alpha)d\alpha$.

Совокупность методов представления сигналов в виде (1.1) и (1.2) называют *обобщенной спектральной теорией сигналов*. В рамках линейной теории спектры являются удобной аналитической формой представления сигналов.

Для теоретического анализа базисные функции $\varphi_k(t)$ нужно выбирать так, чтобы они имели простое аналитическое выражение, обеспечивали быструю сходимость ряда (1.1) для любых сигналов $U(t)$ и позволяли легко вычислять значения коэффициентов C_k. Базисные функции не обязательно должны быть действительными, их число может быть неограниченным. Обычно сигнал представляется суммой ограниченного числа $(0 \le k \le n)$ действительных линейно независимых базисных функций.

Ортогональное представление сигналов

Вычисление спектральных составляющих сигнала существенно облегчается при выборе в качестве базиса системы ортогональных функций.

Систему функций $\psi_0(t)$, $\psi_1(t)$, …, $\psi_k(t)$, …, $\psi_n(t)$ называют *ортогональной* на отрезке $[t_a, t_b]$, если для всех $k = \overline{0, n}$; $j = \overline{0, n}$, за исключением случая $k = j$, удовлетворяется условие

$$\int_{t_a}^{t_b} \psi_k(t)\psi_j(t)dt = 0. \qquad (1.3)$$

Эта система будет ортонормированной, если для всех $j = \overline{0, n}$ справедливо:

$$\int_{t_a}^{t_b} \psi_j^2(t)dt = 0. \qquad (1.4)$$

Определим коэффициенты C_k при представлении сигнала $U(t)$ совокупностью ортонормированных функций в виде $U(t) = \sum_k C_k \psi_k(t)$, $t \subseteq [t_1, t_2]$, отсюда:

$$C_k = \int_{t_1}^{t_2} U(t)\psi_i(t)dt\,. \qquad (1.5)$$

В теоретических исследованиях обычно используют полные системы ортогональных функций, обеспечивающие сколь угодно малую разность непрерывной функции $U(t)$ и представляющего ее ряда при неограниченном увеличении числа его членов. Разность оценивают по критерию

$$\delta = \int_{-\infty}^{\infty} [U(t) - \sum_k C_k \psi_k(t)]^2 dt\,. \qquad (1.6)$$

При этом говорят о сходимости ряда $\sum_k C_k \psi_k(t)$ к функции $U(t)$.

Широко известной ортонормированной системой является совокупность тригонометрических функций аргументов, кратных $\frac{1}{\sqrt{2\pi}}$:

$$\frac{1}{\sqrt{\pi}}\cos\left(k\frac{2\pi}{T}t\right); \quad \frac{1}{\sqrt{\pi}}\sin\left(k\frac{2\pi}{T}t\right); \quad k = 1, 2, 3\ldots$$

Это разложение было первым, носит название ряда Фурье, поэтому соотношение (1.5) часто называют обобщенным рядом Фурье, а значения C_k — обобщенными коэффициентами Фурье.

Временная форма представления сигнала

Временным представлением сигнала называют такое разложение сигнала $U(t)$, при котором в качестве базисных функций используются единичные импульсные функции — дельта-функции. Математическое описание:

$$\delta(t - \xi_1) = \begin{cases} \infty, & t = \xi_1; \\ 0, & t \neq \xi_1; \end{cases} \quad \int_{-\infty}^{\infty} \delta(t - \xi_1)dt = 1, \qquad (1.7)$$

где $\delta(t)$ — дельта функция, отличная от нуля в момент времени $t = \xi_1$.

Такая математическая модель соответствует абстрактному импульсу бесконечно малой длительности и безграничной величины. Ортогональность совокупности таких импульсов очевидна, т.к. они не перекрываются по времени.

Частотная форма представления сигнала

Экспоненциальные базисные функции в преобразовании Фурье комплексно-сопряженными парами позволяют представить сложный детерминированный сигнал в виде суммы гармонических составляющих:

$$e^{j\omega t} + e^{-j\omega t} = 2\cos\,\omega t \qquad \text{(формула Эйлера).} \tag{1.8}$$

Поскольку ω имеет смысл круговой частоты, результат такого преобразования называют частотной формой представления сигнала.

Из-за этого преимущества разложение сигналов по системе гармонических базисных функций легло в основу классической спектральной теории сигналов.

Спектры периодических сигналов

Периодических сигналов, естественно, ***не существует***, т.к. любой реальный сигнал имеет начало и конец. Однако при анализе сигналов в установившемся режиме можно исходить из предположения, что они существуют бесконечно долго, и принять в качестве модели таких сигналов периодическую функцию времени.

Пусть функция $u(t)$ задана в интервале времени $t_1 \le t \le t_2$ и удовлетворяет условиям Дирихле (на любом конечном интервале функция должна быть непрерывной или иметь конечное число точек разрыва первого рода, а также конечное число экстремальных точек). В точках разрыва t_0 функцию $u(t)$ следует считать равной $u(t_0) = S[u(t_0 + 0) + u(t_0 - 0)]$, период повторения $T = \dfrac{2\pi}{\omega_1} = t_2 - t_1$; $[-\infty \le t \le +\infty]$.

Если в качестве базисных выбраны экспоненциальные функции, то выражение (1.5) запишем в виде

$$U(t) = \frac{1}{2}\sum_{k=-\infty}^{\infty} A(jk\omega_1)e^{jk\omega t}\,; \quad A(jk\omega_1) = \frac{2}{T}\int_{t_1}^{t_2} U(t)e^{-jk\omega t}dt. \tag{1.9}$$

Соотношение (1.9) представляет собой ряд Фурье в комплексной форме, содержащей экспоненциальные функции как с положительным, так и с отрицательным параметром (двустороннее частотное представление). Составляющие с отрицательными частотами являются следствием комплексной формы записи вещественной функции.

Функцию $A(jk\omega_1)$ принято называть *комплексным спектром* периодического сигнала $U(t)$. Этот спектр дискретный, т.к. функция $A(jk\omega_1)$ определена на числовой оси только для целых значений k. Значения функции $A(jk\omega_1)$ при конкретном k называют комплексной амплитудой.

Модуль комплексного спектра $A(jk\omega_1)$ называют *спектром амплитуд*, а функцию $\varphi(k\omega_1)$ — *спектром фаз*.

Если известны спектр амплитуд и спектр фаз сигнала, то в соответствии с (1.9) он восстанавливается однозначно. В практических приложениях более значимым является спектр амплитуд, а информация о фазах составляющих часто несущественна (кроме случаев с применением фазовой модуляции сигнала).

Поскольку $A(jk\omega_1)$ и $\varphi(k\omega_1)$ отличны от нуля только при целых k, спектры амплитуд и фаз периодического сигнала являются дискретными.

Воспользовавшись формулой Эйлера: $e^{-jk\omega t} = \cos k\omega t - j\sin k\omega t$, выразим комплексный спектр $A(jk\omega_1)$ в виде действительной и мнимой частей:

$$A(jk\omega_1) = \frac{2}{T}\left[\int_{t_1}^{t_2} U(t)\cos k\omega_1 t dt - j\int_{t_1}^{t_2} U(t)\sin k\omega_1 t dt\right] = A_k - jB_k, \qquad (1.10)$$

где $A_k = \frac{2}{T}\int_{t_1}^{t_2} U(t)\cos k\omega_1 t dt$; $B_k = \frac{2}{T}\int_{t_1}^{t_2} U(t)\sin k\omega_1 t dt$.

Спектр амплитуд $A(k\omega_1) = \sqrt{A_k^2 + B_k^2}$ является четной функцией k, т.е.

$$A(k\omega_1) = A(-k\omega_1). \qquad (1.11)$$

Поскольку четность A_k и B_k противоположна, спектр фаз $\varphi(k\omega_1) = \operatorname{arctg}\frac{B_k}{A_k}$ — функция нечетная, т.е. $\varphi(k\omega_1) = -\varphi(-k\omega_1)$.

При $k = 0$ получаем постоянную составляющую: $\frac{A_0}{2} = \frac{1}{T}\int_{t_1}^{t_2} U(t)dt$.

От двустороннего спектрального представления перейдем к одностороннему (не имеющему отрицательных частот), объединяя комплексно-сопряженные составляющие. В этом случае получим ряд Фурье в тригонометрической форме:

$$U(t) = \frac{A_0}{2} + \sum_{k=1}^{\infty} A(k\omega_1)\cos(k\omega_1 t - \varphi_k). \qquad (1.12)$$

Отдельные составляющие в (1.12) называют *гармониками*. Спектр амплитуд и спектр фаз гармонического сигнала удобно представлять наглядно спектральными диаграммами. На диаграмме спектра амплитуд каждой гармонике ставится в соответствие вертикальный отрезок, длина которого пропорциональна амплитуде, а расположение — частоте этой составляющей. Спектр периодического сигнала характеризует совокупность гармоник, кратных основной частоте ω_1. Аналогично на диаграмме спектра фаз обозначают значения фаз гармоник. Эти спектры отображаются совокупностями линий и носят название *линейчатых*.

Спектры непериодических сигналов

Любой физически реализуемый сигнал ограничен во времени и обладает конечной энергией. Функции, отображающие реальные сигналы, удовлетворяют условиям Дирихле и абсолютно интегрируемы, т.е.

$$\int_{-\infty}^{\infty} |U(t)| dt \le M, \qquad (1.13)$$

где M — конечная величина.

Получим конкретный вид спектрального преобразования для непериодического сигнала, проследив изменения, происходящие в спектре периодической последовательности импульсов $U_1(t)$ при увеличении периода их повторения.

Решение (1.12) показывает, что абсолютные значения амплитуд спектральных составляющих при увеличении периода уменьшаются. Так как частоты состав-

ляющих спектра кратны основной частоте, то при ее уменьшении линии на спектральной диаграмме сближаются. Спектральное представление для одиночного импульса $U_1(t)$ получим как следствие увеличения периода сигнала до бесконечности.

Запишем преобразования Фурье для периодической функции $U_1(t)$ в форме (1.9): при $T \to \infty$, $U_1(t)$ переходит в $U(t)$, частота ω_1 уменьшится до $d\omega$, а $k\omega_1$ превращается в текущую частоту ω. Заменяя суммирование интегрированием, найдем:

$$U(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} U(t) e^{-j\omega t} dt \right] e^{j\omega t} d\omega .$$

Обозначив интеграл в квадратных скобках $S(j\omega)$**,** получим формулы для прямого и обратного интегрального преобразования Фурье:

$$S(j\omega) = \int_{-\infty}^{\infty} U(t) e^{-j\omega t} dt ; \qquad U(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(j\omega) e^{j\omega t} d\omega . \tag{1.14}$$

Величину $S(j\omega)$ называют комплексной спектральной плотностью или спектральной характеристикой. Она имеет размерность [амплитуда/частота]. На каждой конкретной частоте амплитуда соответствующей составляющей равна нулю. Сравнивая (1.10) и (1.14), находим, что бесконечно малому интервалу частоты $d\omega$ соответствует составляющая с бесконечно малой комплексной амплитудой $dA(j\omega)$:

$$dA(j\omega) = \frac{1}{\pi} S(j\omega) d\omega . \tag{1.15}$$

Комплексная спектральная характеристика может быть записана

$$S(j\omega) = S(\omega) e^{-j\varphi(\omega)}, \tag{1.16}$$

где $S(\omega) = |S(j\omega)|$ называется *спектральной плотностью амплитуд,* или *спектром непериодического сигнала*.

Так как составляющие расположены на всех частотах, то спектр непериодического сигнала является непрерывным. Спектральная характеристика может быть представлена следующим образом:

$$S(j\omega) = \int_{-\infty}^{\infty} U(t) \cos\omega t dt - j \int_{-\infty}^{\infty} U(t) \sin\omega t dt = A(\omega) - jB(\omega), \tag{1.17}$$

где

$$A(\omega) = \int_{-\infty}^{\infty} U(t) \cos\omega t dt ; \qquad B(\omega) = \int_{-\infty}^{\infty} U(t) \sin\omega t dt . \tag{1.18}$$

Модуль спектральной характеристики $S(\omega)$ определяется как

$$S(\omega) = \sqrt{|A(\omega)|^2 + |B(\omega)|^2} \tag{1.19}$$

и представляет собой четную функцию частоты.

Для фазы спектральной характеристики $S(j\omega)$ получаем

$$\varphi(\omega) = \operatorname{arctg} \frac{B(\omega)}{A(\omega)} . \tag{1.20}$$

Из (1.14) следует, что $A(\omega)$ — четная функция частоты, а $B(\omega)$ — нечетная, поэтому функция $\varphi(\omega)$ относительно частоты нечетна.

Комплексная форма интегрального преобразования Фурье легко приводится к тригонометрической:

$$U(t)= \frac{1}{\pi}\int_0^{\infty} S(\omega)\cos(\omega t - \varphi(\omega))d\omega . \qquad (1.21)$$

Преимущество тригонометрической формы записи Фурье-преобразования заключается в возможности физического истолкования.

Соотношения между длительностью импульсов и шириной их спектров

Спектр одиночного прямоугольного импульса при увеличении его длительности τ от 0 до ∞ сокращается от безграничного у дельта-функции до одной спектральной линии в начале координат (постоянное значение сигнала). Это свойство сокращения ширины спектра сигнала (при увеличении его длительности и наоборот) справедливо для сигналов любой формы. Оно вытекает непосредственно из особенностей прямого и обратного интегральных преобразований Фурье, у которых показатель степени экспоненты в подынтегральных выражениях имеет переменные t и ω в виде произведения.

Рассмотрим функцию $U(t)$ определенной продолжительности и функцию $U(\lambda t)$, длительность которой при $\lambda > 1$ будет в λ раз меньше. Считая, что $U(t)$ имеет спектральную характеристику $S(j\omega)$, найдем характеристику $S_\lambda(j\omega)$ для $U(\lambda t)$:

$$S_\lambda(-j\omega) = \int_{-\infty}^{\infty} U(\lambda t)e^{-j\omega t}dt = \frac{1}{\lambda}\int_{-\infty}^{\infty} U(t')e^{-\left(\frac{j\omega t'}{\lambda}\right)}dt' = \frac{1}{\lambda}S\left(\frac{j\omega}{\lambda}\right), \qquad (1.22)$$

где $t' = \lambda t$.

Следовательно, спектр укороченного в λ раз сигнала ровно в λ раз шире. Коэффициент $\frac{1}{\lambda}$ перед $S\left(\frac{j\omega}{\lambda}\right)$ изменяет только амплитуду гармонических составляющих и на ширину спектра не влияет.

Другой важный вывод: *длительность сигнала и ширина его спектра не могут быть одновременно ограничены конечными интервалами*; если длительность сигнала ограничена, то спектр его неограничен, и наоборот, сигнал с ограниченным спектром длится бесконечно долго. Справедливо соотношение

$$\Delta t \Delta f = C, \qquad (1.23)$$

где Δt — длительность импульса; Δf — ширина спектра импульса; C — постоянная величина, зависящая от формы импульса (при ориентировочных оценках $C = 1$).

Реальные сигналы ограничены во времени, генерируются и передаются устройствами, содержащими инерционные элементы (емкости и индуктивности), поэтому не могут содержать гармоники сколь угодно высоких частот.

Случайный процесс как модель сигнала

Мы рассматривали модели известных функций времени. Случайные составляющие, всегда существующие в реальном входном сигнале, считались пренебрежимо малыми и не принимались во внимание.

Однако однозначная функция времени только тогда будет нести информацию, когда она выбрана из множества возможных. Поэтому при моделировании сигналов необходимо использовать статистические методы. Это необходимо и потому, что часто недопустимо пренебрегать воздействием помех. Воздействие помех на полезный сигнал проявляется в непредсказуемых искажениях его формы. Математическая модель помехи представляется также в виде случайного процесса, определяемого параметрами, выявленными на эксперименте. Вероятностные свойства помехи, как правило, отличаются от свойств полезного сигнала, что и лежит в основе методов их разделения.

Под *случайным (стохастическим)* процессом подразумевается случайная функция времени $U(t)$, значения которой в каждый момент времени случайны. Точно предсказать, какой конкретно будет реализация этого процесса в очередном опыте принципиально невозможно. Могут быть определены лишь статистические данные, характеризующие все множество возможных реализаций, называемое ансамблем. Появляется возможность судить о поведении информационной системы по отношению ко всему ансамблю возможных реализаций.

Случайный процесс, у которого множество состояний составляет континуум, а изменения состояний возможны в любые моменты времени, называют *непрерывным случайным процессом*. Случайный процесс с конечным множеством состояний, которые могут изменяться в произвольные моменты времени, называют дискретным. Среди случайных процессов с дискретным множеством состояний выделим такие, у которых статистические зависимости распространяются на ограниченное число k следующих друг за другом значений. Они называются *обобщенными марковскими процессами k-го порядка*.

Вероятностные характеристики случайного процесса. Случайный процесс $U(t)$ может быть описан системой n обычно зависимых случайных величин U_i (i = 1, 2, …, n), взятых в различные моменты времени t_1, …, t_n. При неограниченном увеличении числа n такая система эквивалентна случайному процессу $U(t)$.

Исчерпывающей характеристикой такой системы является n-мерная плотность вероятности реализации такого многофакторного события. На практике, чтобы снизить трудоемкость, чаще рассматривают одно или двухмерную вероятностную систему.

Одномерная плотность вероятности $p_1(U_1; t_1)$ случайного процесса $U(t)$ характеризует распределение одной случайной величины U_1, взятой в произвольный момент времени t_1.

Двухмерная плотность вероятности $p_2(U_1; U_2; t_1; t_2)$ позволяет определить вероятность совместной реализации любых 2 значений случайных величин U_1 и

U_2 в произвольные моменты времени t_1 и t_2 и, следовательно, оценить динамику развития процесса.

Описывают случайный процесс через *математическое ожидание* и *дисперсию.*

Математическим ожиданием $m_u(t_1)$ в момент времени t_1 называют среднее значение случайной величины $U(t_1)$ по всему множеству возможных реализаций:

$$m_u(t_1) = M\{U(t_1)\} = \int_{-\infty}^{\infty} U_1 p_1(U_1; t_1) dU_1 . \qquad (1.24)$$

Степень разброса случайных значений процесса от среднего значения характеризуется дисперсией $D_u(t)$ и среднеквадратичным отклонением $\sigma_u(t)$:

$$D_u(t) = M\{[U(t) - m(t)]^2\} = M\{[U(t)]^2\};\ D_u(t) = \sigma^2{}_u(t), \qquad (1.25)$$

где $U(t) = U(t) - m(t)$ — центрированная случайная величина.

Случайные процессы могут иметь одинаковые математические ожидания и дисперсии, однако резко различаться по быстроте изменений своих значений во времени. Для оценки степени статистической зависимости мгновенных значений процесса в произвольные моменты времени и используется *автокорреляционная* (или просто *корреляционная*) функция $R_u(t_1, t_2)$. При конкретных аргументах t_1 и t_2 она равна корреляционному моменту значений процесса $U(t_1)$ и $U(t_2)$:

$$R_u(t_1, t_2) = M[U(t_1)\ U(t_2)]. \qquad (1.26)$$

В силу симметричности этого выражения относительно аргументов справедливо равенство $R_u(t_1, t_2) = R_u(t_2, t_1)$.

Для сравнения случайных процессов вместо корреляционной функции удобно пользоваться нормированной функцией автокорреляции, которая равна 1:

$$\rho_u = \frac{R_u(t_1, t_2)}{\sigma_u(t_1)\sigma_u(t_2)} = 1. \qquad (1.27)$$

Дисперсию случайного процесса можно рассматривать как частное значение автокорреляционной функции.

Стационарные и эргодические случайные процессы. Стационарными называются процессы, плотность вероятности в которых не зависят от времени (таких процессов в природе не существует).

Процесс $U(t)$ принято называть стационарным, если выполняется условие постоянства математического ожидания и дисперсии, а корреляционная функция не зависит от начала отсчета времени и является функцией только $\tau = t_2 — t_1$.

Случайные процессы, наблюдаемые в устойчиво работающих реальных системах, имеют конечное время корреляции. Если для случайного процесса перечисленные требования не выдерживаются, но на интересующем нас интервале времени изменением этих параметров можно пренебречь, его называют квазистационарным.

Свойство эргодичности состоит в том, что каждая реализация случайного процесса несет практически полную информацию о свойствах всего ансамбля

реализаций, что позволяет упростить процедуру определения статистических характеристик, заменяя усреднение значений по ансамблю реализаций усреднением значений одной реализации за длительный промежуток времени:

$$m_u = \lim_{T\to\infty}\frac{1}{T}\int_0^T u(t)dt = u_0; \qquad D_u = \lim_{T\to\infty}\frac{1}{T}\int_0^T [u(t)-u_0]^2\,dt = u_0; \tag{1.28}$$

$$R_{\text{u}}(\tau) = \lim_{T\to\infty}\frac{1}{T}\int_0^T [u(t)-u_0][u(t+\tau)-u_0]dt\,, \tag{1.29}$$

где $u(t)$ — конкретная реализация случайного процесса $U(t)$.

Исследование случайных процессов в их временном представлении лежит в основе корреляционной теории сигналов.

Преобразование непрерывных сигналов в дискретные

В любую систему информация поступает в виде сигналов. Различные данные о физических процессах преобразуются в электрические сигналы с помощью датчиков. Как правило, это непрерывно изменяющийся ток или напряжение, хотя возможны и импульсные сигналы (радиолокация, компьютеры).

Информацию можно хранить, передавать и обрабатывать как в виде непрерывных сигналов, так и дискретных. Сегодня предпочтение отдается дискретным сигналам, поэтому непрерывные сигналы дискретизируют.

Под *дискретизацией* понимают преобразование функции непрерывного времени в функцию дискретного времени. Роль координат при этом часто выполняют мгновенные значения функции, отсчитанные в определенные моменты времени.

Под квантованием подразумевают преобразование некоторой величины с непрерывной шкалой значений в величину, имеющую дискретную шкалу значений. Оно сводится к замене любого значения одним из конечного множества разрешенных значений, называемых уровнями квантования. При проведении операций дискретизации и квантования вид сигнала изменяется принципиально, поскольку можно вместо импульсов передавать числа, представляющие собой номер квантового уровня, номеру их дискретизации

Причины перехода к дискретному и цифровому выражению информации. Для конкретных задач управления или исследования обычно требуется значительно меньше информации, чем ее поступает с непрерывно выдающих сигналы датчиков. Имея априорную информацию о законах управления системой, можно определить промежуток времени, когда следует проводить отсчет.

Все сигналы с любых датчиков снимаются в условиях флуктуаций значений параметров и наличия погрешностей измерения. Это определяет уровни квантования сигналов. Зачастую можно и нужно еще и огрубить эти сигналы, поскольку специфика решаемых в системе задач не требует их точных значений. Рациональное выполнение операций дискретизации и квантования приводит к снижению затрат на передачу, обработку и хранение информации.

При передаче и обработке информации в цифровой форме существует принципиальная возможность снижения вероятности получения ошибочного результата до сколь угодно малых значений. Это возможно потому, что:

— применимы такие методы кодирования, которые обеспечивают обнаружение и исправление ошибок;

— можно избежать свойственного аналоговым сигналам эффекта накопления искажений в процессе их передачи и обработки, поскольку квантованный сигнал легко восстановить до первоначального значения всякий раз, когда величина накопленного искажения становится значимой.

Практическая реализация наиболее эффективна для двух уровней сигнала.

Цифровая форма информации помогает унифицированной ее обработке и хранению, уменьшает стоимость благодаря массовости изготовления и повышает надежность работы аппаратуры.

Общая постановка задачи дискретизации

В самом общем случае представление непрерывного сигнала $u(t)$ на интервале T совокупностью координат $(c_1, c_2, \ldots, c_N)$ может быть записано в виде $(c_1, c_2, \ldots, c_N) = A[u(t)]$, где A — оператор дискретного представления сигнала, реализуемый устройством, называемым дискретизатором.

Операция восстановления непрерывной функции $u^*(t)$ (воспроизводящей функции) по совокупности координат $(c_1, c_2, \ldots, c_N)$, отображающей исходный сигнал с некоторой погрешностью $\delta(t) = u(t) - u^*(t)$:

$$u^*(t) = B[c_1, c_2, \ldots, c_N],$$

где B — оператор восстановления, реализуемый устройством восстановления сигнала.

Задача дискретизации в математическом плане сводится к совместному выбору пары операторов A и B, обеспечивающих заданную точность восстановления сигнала. Практическое применение нашли линейные операторы:

$$c_j = \int_T \xi_j(t)u(t)dt = Au(t), \qquad (1.30)$$

где $\{\xi_j(t)\}_{j=1}^N$ — система весовых функций.

Воспроизводящая функция представляется аппроксимирующим полиномом:

$$u^*(t) = \sum_{j=1}^{N} c_j \varphi_j(t) = B(c_1, c_2, \ldots, c_N), \qquad (1.31)$$

где $\{\varphi_j(t)\}_{j=1}^N$ — система базисных функций.

При одном и том же операторе представления A для восстановления могут использоваться различные операторы B. Из (1.30) следует, что произведения $[\xi_j(t)\varphi_j(t)]$ должны иметь размерность, обратную времени.

Применяются различные методы дискретизации. В части алгоритмов используются характеристики случайного процесса как модели сигнала. Методы дискретизации рассматривают и с позиций полезности для решения теоретических вопросов передачи и преобразования сигналов, и с позиции возможности их технической реализации. В теоретическом плане важны методы, обеспечивающие минимальное количество координат при заданной погрешности воспроизведения (методы оптимальной или предельной дискретизации). Эти методы применяются для случаев высоких уровней импульсных помех.

Чаще используется замена сигнала $u(t)$ на совокупность его мгновенных значений $u(t_j)$, взятых в определенные моменты времени t_j (j = 1, 2, …, N). Роль весовых функций $\xi_j(t)$ играют дельта-функции Дирака. Поскольку дельта-функция технически нереализуема, отсчеты берут в некотором интервале времени, значительно меньшем шага дискретизации.

Если шаг дискретизации остается постоянным во всем диапазоне преобразования, дискретизация считается равномерной.

Теорема Котельникова

Правило выбора предельного шага при равномерной дискретизации с использованием модели сигнала с ограниченным спектром носит название теоремы В.А. Котельникова (другие названия: теорема отсчетов или теорема Найквиста).

Теорема устанавливает принципиальную возможность полного восстановления детерминированной функции с ограниченным спектром по ее отсчетам и указывает предельное значение интервала времени между отсчетами, при которой такое восстановление возможно. Формулировка:

Функция u(t), допускающая преобразование Фурье и имеющая непрерывный спектр, ограниченный полосой частот от 0 до $F_c = \omega_c/2\pi$, полностью определяется дискретным рядом своих мгновенных значений, отсчитанных через интервалы времени:

$$\Delta t = \frac{1}{2F_c}. \qquad (1.32)$$

Физическая основа теоремы выявляется при рассмотрении связи между формой функции и шириной ее спектра. Если спектр безграничен, ее значения могут изменяться произвольно. Сокращение высокочастотной части спектра до частоты ω_1 равнозначно устранению из временной функции деталей, которые могли быть созданы этими высокочастотными составляющими. Поскольку значения функций в пределах малого Δt не могут изменяться существенно, можно ограничиться значениями функции, взятыми через этот интервал времени.

2. Количественная оценка информации

Энтропия как мера неопределенности выбора

Факт получения информации всегда связан с уменьшением разнообразия или неопределенности. Установим количественные меры неопределенности для информации и выясним ее свойства.

Дискретный источник информации может в каждый момент времени случайным образом принять одно из конечного множества возможных состояний. Различные состояния u_i «*реализуются вследствие выбора их источником*». Ансамбль состояний U характеризуется суммой вероятностей их появления:

$$\sum_i p_i = 1. \qquad (2.1)$$

Введем меру неопределенности выбора состояния источника. Ее можно рассматривать и как меру количества информации. За такую меру можно было бы взять число состояний источника (при их равновероятности). Тогда эта мера отвечала бы условию монотонного возрастания при увеличении числа возможных состояний источника.

Однако такая мера не отвечает требованию аддитивности: *Если два независимых источника с числом равновероятных состояний N и M рассматривать как один источник, одновременно реализующий пары состояний $n_i m_j$, то неопределенность объединенного источника должна равняться сумме неопределенностей исходных источников:*

$$f(NM) = f(N) + f(M). \qquad (2.2)$$

Соотношение (2.2) выполняется, если в качестве меры неопределенности источника с равновероятными состояниями принять логарифм числа состояний:

$$H(U) = \log N. \qquad (2.3)$$

Тогда при $N = 1$ и $H(U) = 0$ требование аддитивности выполняется (Р. Хартли, 1928 г.). Основание логарифма не имеет принципиального значения и определяет только масштаб или единицу неопределенности. Технические соображения подсказывают выбор основания логарифма — 2. При этом единица неопределенности называется битом (от англ. *binary digit*). Иногда используется дит (от *decimal*).

Пример: Определить минимальное число взвешиваний для выявления одной фальшивой монеты среди 27: $H(U) = \log_2 27$. Одно взвешивание: три возможных исхода, неопределенность: $H(U') = \log_2 3$. Поэтому $H(U) = 3\log_2 3 = H(U')$, т.е. требуется 3 взвешивания.

Предложенная мера удачна, но широко не применяется, т.к. использует слишком грубую модель источника информации (равновероятную).

К. Шеннон предложил более широко используемую меру:

$$H(U) = -C\sum_i p_i \log p_i , \qquad (2.4)$$

где C — произвольное положительное число.

Ее называют энтропией дискретного источника информации, или энтропией конечного ансамбля. Это единственный функционал (утверждение К.

Шеннона, строго доказанное Л.Я. Хинчиным), удовлетворяющий всем требованиям к мере неопределенности (мере информации).

Для двоичной системы измерения, приняв $C = 1$, получим

$$H(U) = -\sum_i p_i \log_2 p_i , \qquad (2.5)$$

Формальная структура (2.4) совпадает с энтропией физической системы (Больцман). Согласно второму закону термодинамики энтропия замкнутого пространства определяется как

$$H = -\frac{1}{M_П} \sum_{i=1}^{N} m_i \ln \frac{m_i}{M_П}, \qquad (2.6)$$

где $M_П$ — число молекул в данном пространстве; m_i — число молекул, обладающих скоростью от v до $v + \Delta v$.

Так как $m_i / M_П$ есть вероятность того, что молекула имеет скорость от v до $v + \Delta v$, то (2.6) можем переписать: $H = -\sum_{i=1}^{N} p_i \ln p_i$.

Совпадение имеет глубокий физический смысл, поскольку в обоих случаях величина H характеризует степень разнообразия состояний системы.

Мера Шеннона является естественным обобщением меры Хартли на случай ансамбля с неравновероятными состояниями. Она позволяет учесть статистические свойства источника информации.

Некоторые свойства энтропии:

1. Энтропия является вещественной и неотрицательной величиной, т.к. для любого p_i ($1 \le i \le N$) изменяется в интервале от 0 до 1, $\log p_i$ отрицателен и, следовательно, $-p_i \log p_i$ положительно.

2. Энтропия — величина ограниченная. Для слагаемых $-p_i \log p_i$ в диапазоне $0 < p_i < 1$ ограниченность очевидна. Предел для $-p_i \log p_i$ при $p_i \to 0$, по правилу Лопиталя, равен 0.

3. Энтропия обращается в ноль, если вероятность одного из состояний равна 1.

4. Энтропия максимальна, когда все состояния источника равновероятны, что доказывается использованием метода неопределенных множителей Лагранжа.

5. Энтропия источника u с двумя состояниями u_1 и u_2 изменяется от 0 до 1, достигая максимума при равенстве их вероятностей:

$$p(u_1) = p = p(u_2) = 1 - p = 0{,}5.$$

6. Энтропия объединения нескольких статистически независимых источников информации равна сумме энтропий исходных источников.

7. Энтропия характеризует среднюю неопределенность выбора одного состояния из ансамбля и ничто больше (при оценке неопределенности воздействия лекарств безразлично, выздоровеет 90%, а 10% умрет, или наоборот).

Энтропия может характеризовать не только дискретный, но и непрерывный источник информации. Энтропию для такого источника называют дифференциальной энтропией:

$$H(U) = \int_{-\infty}^{\infty} p(u)\log p(u)du - \lim_{\Delta u \to 0} \log \Delta u . \qquad (2.7)$$

Эта величина при $\Delta u \to 0$ стремится к бесконечности (неопределенность выбора из бесконечного числа возможных состояний (значений) бесконечно велика).

Первый член в правой части (2.7) имеет конечное значение, зависящее только от закона распределения U, и не зависит от шага квантования Δu. Он имеет точно такую же структуру, как и энтропия дискретного источника.

Второй член зависит лишь от шага квантования Δu. Он ответственен за то, что $H(U)$ обращается в бесконечность.

К трактовке (2.7) известны два подхода.

Первый состоит в том, что в качестве меры неопределенности непрерывного источника принимают первый член. Эта величина получила название дифференциальной энтропии непрерывного источника. Ее можно трактовать как среднюю неопределенность выбора случайной величины U с произвольным законом распределения по сравнению со средней неопределенностью выбора случайной величины U', изменяющейся в диапазоне, равном 1, и имеющей равномерное распределение.

Условная энтропия непрерывного источника может быть выражена как

$$H_V(U) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(u,v)\log\left[\frac{p(u,v)}{p(v)}\right]dudv - \lim_{\Delta u \to 0}\Delta u . \qquad (2.8)$$

При втором подходе для количественного определения информационных свойств непрерывного источника предлагается принять во внимание практическую невозможность обеспечения бесконечно большой точности различения значений непрерывной величины U. Поэтому все бесконечное число значений U в пределах заданной точности измерений следует рассматривать как одно значение.

Из средней неопределенности выбора источником некоторого значения в этом случае необходимо вычесть среднюю неопределенность того же источника, полученную при условии, что мы знаем результаты определения с некоторой точностью ε. Тогда информационные свойства непрерывного источника будут оцениваться разностью безусловной (2.7) и условной (2.8) энтропий. Такая разность является мерой снятой неопределенности, называемой количеством информации.

Количество информации как мера снятой неопределенности

Передача информации диктуется желанием устранить неопределенность относительно последовательности состояний, реализуемых источником. Передача информации инициируется либо самим источником, либо осуществляется по запросу. Информация проявляется всегда в форме сигналов. Сигналы, поступающие с выхода первичного преобразователя источника информации на вход канала связи, принято называть сообщениями, в отличие от сигналов, формирующихся на входе линии связи.

Отдельные первичные сигналы с выхода источника сообщений называют *элементами сообщений*. Каждому элементу сообщения соответствует определенное состояние источника информации. Если источник информации реализует множество состояний параллельно (лист бумаги с текстом), первичный преобразователь обеспечивает их последовательное отображение элементами сообщений (произнесение звуков человеком).

Основное понятие теории информации — количество информации — рассматривается здесь применительно к передаче отдельных статистически несвязанных элементов сообщения. Дискретный источник информации сообщений при этом полностью характеризуется ансамблем

$$Z = \begin{pmatrix} z_1 & .... & z_i & ... & z_N \\ p(z_1) & ... & p(z_i) & ... & p(z_N) \end{pmatrix}.$$

Непрерывный источник информации характеризуется одномерной плотностью распределения случайной величины: z — $p(z)$.

Передача информации от дискретного источника

Как меняется неопределенность относительно состояния источника сообщения при получении адресатом элемента сообщения с выхода канала связи? Алфавиты передаваемых и принимаемых элементов сообщения будем считать идентичными.

Вследствие воздействия помех полученный элемент сообщения в общем случае отличается от переданного. Обозначим принимаемые элементы сообщения буквами $w_1, \ldots, w_i, \ldots, w_n$.

Априорная неопределенность (неопределенность до получения элемента сообщения) относительно состояния источника не является полной. Предполагается, что адресату известен алфавит элементов сообщения, а из прошлого опыта он знает вероятности их появления. Считая, что состояния источника реализуются независимо, априорная частная неопределенность появления элемента сообщения z_i: $H(z_i) = -\log p(z_i)$; где $p(z_i)$ — априорная вероятность появления элемента сообщения z_i.

Обычно считают, что между элементами сообщения и помехой статистические связи отсутствуют, искажения отдельных элементов сообщения являются событиями независимыми и адресату известна совокупность условных вероят-

ностей $p\left(\frac{z_i}{w_i}\right)$ ($1 \le i \le N$, $1 \le j$) того, что вместо элемента сообщения z_i будет принят элемент сообщения w_j.

При получении конкретного элемента сообщения w_j адресату становится известным значение условной вероятности $p\left(\frac{z_i}{w_i}\right)$, называемой апостериорной (послеопытной) вероятностью реализации источником элемента сообщения z_i. Это позволяет найти апостериорную частную неопределенность, остающуюся у адресата и относящуюся к выдаче источником элемента сообщения z_i после получения конкретного элемента сообщения w_j:

$$H\left(\frac{z_i}{w_i}\right) = -\log p\left(\frac{z_i}{w_i}\right). \qquad (2.9)$$

Определим частное количество информации $I(z_i)$**,** получаемое при приеме элемента сообщения w_j относительно некоторого реализованного источником элемента сообщения z_i, как разность частных неопределенностей, имевшихся у адресата до и после получения элемента сообщения:

$$I(z_i) = H(z_i) - H\left(\frac{z_i}{w_j}\right) = -\log\frac{p\left(\frac{z_i}{w_j}\right)}{p(z_i)}. \qquad (2.10)$$

Анализ (2.10) позволяет сделать следующие заключения.

1. Частное количество информации растет с уменьшением априорной и увеличением апостериорной вероятностей реализации элемента сообщения источником, что находится в соответствии с интуитивным представлением.

2. Частное количество информации об элементе сообщения может быть не только положительным, но и отрицательным, а также нулем, что зависит от соотношения априорной $p(z_i)$ и апостериорной $p\left(\frac{z_i}{w_j}\right)$ вероятностей. Если вероятность того, что источником был реализован элемент сообщения z_i, увеличилась после приема элемента сообщения w_j, т.е. $p\left(\frac{z_i}{w_j}\right) > p(z_i)$, то полученное частное количество информации положительно. Если эта вероятность не изменилась, т.е. $p\left(\frac{z_i}{w_j}\right) = p(z_i)$, то имевшая место неопределенность тоже не изменилась и частное количество информации равно 0. Случай $p\left(\frac{z_i}{w_j}\right) < p(z_i)$ соответствует увеличению неопределенности относительно реализации z_i после по-

лучения элемента сообщения w_j, и частное количество информации отрицательно.

3. В случае отсутствия помехи апостериорная вероятность $p\left(\frac{z_i}{w_j}\right) = 1$. Частное количество информации численно совпадает с априорной неопределенностью реализации данного элемента сообщения z_i: $I(z_i) = H(z_i) = -\log p(z_i)$. Это максимальное частное количество информации, которое можно получить об элементе сообщения z_i.

4. Частное количество информации относительно реализации источником элемента сообщения z_i, содержащееся в принятом источником элементе сообщения w_j, равно частному количеству информации относительно w_j, содержащемуся в элементе сообщения z_i:

$$I(z_i, w_j) = \log \frac{p(z_i / w_j)}{p(z_i)} = \log \frac{p(z_i, w_j)}{p(z_i)\,p(w_j)} = I(w_j, z_i). \qquad (2.11)$$

Для большинства применений важны усредненные характеристики функционирования информационных систем.

Среднее количество информации, содержащееся в любом принятом элементе сообщения относительно переданного (реализованного) источником через вероятности: $I(z, w) = \sum_{i=1}^{N}\sum_{j=1}^{M} p(z_i w_j)\log\frac{p(z_i w_j)}{p(z_i)\,p(w_j)}$.

Если частный характер количества информации специально не оговаривается, всегда идет речь о количестве информации, приходящемся *в среднем на один элемент сообщения*.

Передача информации от непрерывного источника

Количество информации, получаемой от непрерывного источника по каналу с помехами, определяется точно так же, как выше, но с использованием понятия дифференциальной энтропии (2.7).

Среднее количество информации, содержащееся в каждом принятом значении случайной величины w источника, имеющего непрерывное множество состояний относительно переданного значения случайной величины z, можно получить как разность априорной и апостериорной дифференциальных энтропий:

$$I(z, w) = H(z) - H_w(z). \qquad (2.12)$$

Основные свойства количества информации:

1. Несмотря на то, что частное количество информации может быть величиной отрицательной, количество информации неотрицательно.

Действительно, если $H_w(z) \le H(z)$, тогда $I(z, w) = H(z) - H_w(z) \ge 0$.

2. При отсутствии статистической связи между случайными величинами z и w: $H_w(z) = H(z)$, следовательно, в этом случае $I(z, w) = 0$ (принятые элементы сообщения не несут никакой информации относительно переданных).

3. Количество информации в w относительно z равно количеству информации в z относительно w.

4. При взаимно однозначном соответствии между множествами передаваемых и принимаемых элементов сообщений (в отсутствие помехи) апостериорная энтропия равна нулю и количество информации численно совпадает с энтропией источника: $I(z, w) = H(z)$.

Это максимальное количество информации о состоянии дискретного источника. Для непрерывного источника оно бесконечно.

Информационные характеристики источника сообщений и канала связи

Основные понятия и определения. Источники сообщений и каналы связи в системах передачи имеют разнообразную структуру и физическую природу. Для выяснения общих закономерностей необходимо абстрагироваться от их конкретного физического воплощения и оперировать формализованными понятиями.

Источник дискретных сообщений формирует дискретные последовательности из ограниченного числа элементарных сообщений. На выходе источника непрерывных сообщений образуются непрерывные сообщения. Источник полностью определяется статистическими данными о формируемых сообщениях.

Под каналом связи подразумевают совокупность устройств и физических сред, обеспечивающих передачу сообщений из одного места в другое (или во времени). Каналы также подразделяются на дискретные и непрерывные.

Так как в процессе передачи дискретных сообщений модулятором в соответствии с поступающей последовательностью символов осуществляется изменение информативного параметра непрерывного сигнала, то часть дискретного канала является непрерывным каналом связи. Именно модулятор (демодулятор), а также кодирующие и декодирующие устройства отличают дискретный канал от непрерывного.

Если вредным влиянием помех в канале можно пренебречь, то его можно назвать *каналом без помех* и проводить анализ *идеализированного канала*. В таком (идеальном) канале каждому сообщению на входе однозначно соответствует определенное сообщение на выходе и наоборот (не надо вводить сообщений w_j). Более сложная модель — канал с помехами.

Канал считается заданным, если известны статистические данные о сообщениях на его входе и выходе и ограничения, накладываемые на входные сообщения физическими характеристиками канала. Канал прямой передачи, дополненный обратным каналом, называют каналом с обратной связью.

Информационные характеристики источника дискретных сообщений

Модели источника дискретных сообщений. Потребителя чаще всего интересует не одно конкретное состояние источника, а последовательности со-

стояний (телеграммы, ТВ-сюжеты и пр.). Для построения моделей непрерывных и дискретных случайных процессов необходимо знать объем l алфавита знаков (z_1, z_2, …, z_l), из которых источником формируются сообщения, и вероятности создания им отдельных знаков с учетом возможной взаимосвязи между ними.

При доказательстве основных положений теории информации Шенноном использовалась модель, называемая эргодическим источником сообщений. Предполагается, что создаваемые им сообщения математически можно представить в виде эргодической случайной последовательности.

Такая последовательность удовлетворяет условиям стационарности и эргодичности, т.е. вероятность отдельных знаков и их сочетаний не зависит от расположения их по длине сообщения, и статистические закономерности, полученные при исследовании одного, достаточно длинного сообщения, с вероятностью, близкой к единице, справедливы для всех сообщений, создаваемых источником. Из статистических характеристик в данном случае интересна средняя неопределенность в расчете на один знак последовательности.

Стационарный источник сообщений, выбирающий каждый знак формируемой последовательности независимо от других знаков, всегда является эргодическим. Его также называют источником без памяти.

На практике чаще встречаются источники, у которых вероятность выбора зависит от того, какие знаки были выбраны источником раньше (источники с памятью). Для описания используют *цепи Маркова*.

Цепь Маркова порядка n характеризует последовательность событий, вероятности которых зависят от того, какие n событий предшествовали данному. При объеме алфавита l число R различных состояний источника не превышает l^n. Обозначим состояния через S_1, …, S_q, …, S_R, а вероятности выбора в состоянии S_q знака z_i через $p_q(z_i)$.

Когда корреляционные связи наблюдаются только между двумя знаками (простая цепь Маркова), максимальное число различных состояний источника равно объему алфавита: $R = l$ и $p_q(z_i) = p\left(\frac{z_i}{z_q}\right)$, где $1 \le q \le l$. Энтропия источника сообщений:

$$H(Z) = -\sum_{q=1}^{l} p(z_q) \sum_{i=1}^{l} p\left(\frac{z_i}{z_q}\right) \log p\left(\frac{z_i}{z_q}\right). \qquad (2.13)$$

При наличии корреляционной связи между тремя знаками, состояния источника определяются двумя предшествующими знаками. Дадим обозначение для произвольного состояния источника S_{kh}. Тогда $p(S_{kh}) = p(z_k, z_h)$; и $p_q(z_i) = p\left(\frac{z_i}{z_k z_h}\right)$.

Величина энтропии источника для этого случая

$$H(Z) = -\sum_{k=1}^{l}\sum_{h=1}^{l} p(z_k z_h)\sum_{i=1}^{l} p\left(\frac{z_i}{z_k z_h}\right)\log p\left(\frac{z_i}{z_k z_h}\right). \qquad (2.14)$$

Аналогично могут быть получены выражения для энтропии источника и при более протяженной корреляционной связи между знаками.

Свойства эргодических последовательностей знаков

Любую последовательность знаков всегда можно разбить на две группы: *типичные и нетипичные*. К типичным относятся такие последовательности знаков, которые при достаточно большом N отличаются тем, что вероятности их появления практически одинаковы, причем вероятность p любой такой последовательности удовлетворяет неравенству:

$$\left| \log \frac{(1/p)}{N} - H(z) \right| < \mu , \qquad (2.15)$$

где $H(z)$ — энтропия источника сообщений.

(2.15) называют *свойством асимптотической равномерности* длинных последовательностей. При $N \to \infty$ источник сообщений с вероятностью, сколь угодно близкой к единице, выдает только типичные последовательности, принимаемое во внимание число последовательностей равно $1/p$. Неопределенность создания каждой такой последовательности равна $\log(1/p)$. С учетом неопределенности, приходящейся на один знак, это — энтропия источника: $\log\frac{(1/p)}{N} = H(z)$.

Для общего случая это утверждение доказывается с привлечением цепей Маркова. Важно также и то, что за исключением случая равновероятного и независимого выбора букв источником, когда нетипичные последовательности просто отсутствуют, типичные последовательности при достаточно большом N составляют незначительную долю от общего числа возможных последовательностей.

К. Шеннон показал, что рассмотренные свойства длинных последовательностей помогают осуществлять эффективное кодирование информации.

Избыточность. Из-за разной вероятности использования источником знаков алфавита существует также недоиспользование их как переносчиков информации. Известная априорная информация о вероятностях выбора отдельных знаков и их сочетаний приводит к уменьшению средней неопределенности выбора источником знака и, как следствие, — переносимого им количества информации. При равновероятном и некоррелированном выборе ту же информационную нагрузку на знак можно обеспечить, используя алфавит меньшего объема.

Избыточность алфавита источника и ее мера D показывают, насколько хорошо используются знаки данного источника:

$$D = \frac{H_{\max}(z) - H(z)}{H_{\max}(z)}, \qquad (2.16)$$

где $H_{\max}(z)$ — максимально возможная энтропия, равная log l; $H(z)$ — энтропия источника.

Если избыточность равна 0, то формируемые сообщения оптимальны в смысле наибольшего количества переносимой информации. Для передачи информации в количестве I при отсутствии помех в этом случае необходимо $k_1 = \frac{I}{H_{\max}(z)}$ знаков.

Избыточность нельзя рассматривать как признак несовершенства источника сообщений. Обычно она является следствием его физических свойств. Например, особенности артикуляции не позволяют формировать слова, состоящие из произвольных сочетаний букв.

Последствия избыточности неоднозначны. С одной стороны, избыточные сообщения требуют дополнительных затрат на передачу, например, увеличения времени или расширения ширины спектра канала связи, что нежелательно. С другой стороны, при использовании сообщений, подчиняющихся априорно известным ограничениям, появляется возможность обнаружения и исправления ошибок, которые приводят к нарушению этих ограничений, и наличие избыточности способствует повышению помехоустойчивости сообщений. Высокая избыточность большинства естественных языков обеспечивает, например, надежное общение людей даже при наличии у них акцентов и дефектов речи.

Для автоматических систем избыточность в большинстве случаев необходимо устранять, поскольку алгоритмы обнаружения и исправления ошибок, базирующихся на статистических закономерностях работы источников, оказываются неоправданно сложными для реализации. Для повышения помехоустойчивости вводится «рациональная» избыточность, позволяющая обнаруживать и исправлять ошибки наиболее простыми средствами.

Производительность источника дискретных сообщений

Под производительностью подразумевают количество информации, вырабатываемое в единицу времени. Она же — скорость создания сообщений или поток входной информации.

Обозначим длительность выдачи знака z_i, формируемого источником в состоянии S_q, через τ_{qz}. Тогда средняя длительность выдачи источником одного знака

$$\tau_н = \sum_{q=1}^{R} p(S_q) \sum_{i=1}^{l} p_q(z_i) \tau_{qz} . \qquad (2.17)$$

Производительность источника $\bar{I}(z)$ можно выразить как $\bar{I}(z) = \frac{H(z)}{\tau_н}$.

Повышение производительности источника возможно не только за счет увеличения энтропии, но и за счет снижения средней длительности формирования знака. Длительность знака желательно выбирать обратно пропорционально вероятности его появления.

Информационные характеристики дискретных каналов связи

Пропускная способность дискретного канала без помех. Для теории и практики важно, до какого предела и как можно повысить скорость передачи информации по каналу.

Пропускная способность канала $C_д$ равна той максимальной скорости передачи, которую можно достигнуть при самых совершенных способах передачи и приема: $C_д = \max \bar{I}(V, U) = \max V_\tau I(V, U)$; $V_\tau = \dfrac{1}{\tau_{ср}}$ — техническая скорость передачи.

При заданном алфавите символов и фиксированных основных характеристиках канала (полосе частот, средней и пиковой мощности передатчика) остальные характеристики должны быть выбраны так, чтобы обеспечить наибольшую скорость передачи по нему сигналов. Пропускная способность канала, как и скорость передачи, измеряется числом двоичных единиц информации в секунду.

В отсутствие помех есть взаимно однозначное соответствие между множеством символов на выходе канала и на его входе ($I(V, U) = H(U)$). Максимум возможного количества информации на символ $\log m$, где m — объем алфавита символов, откуда пропускная способность дискретного канала без помех:

$$C_д = V_\tau \log m. \qquad (2.18)$$

Для увеличения скорости передачи и приближения ее к пропускной способности канала последовательность букв сообщения должна подвергнуться такому преобразованию в кодере, при котором различные символы в его выходной последовательности появлялись бы по возможности равновероятно, а статистические связи между ними отсутствовали бы. Доказано, что это выполнимо для любой эргодической последовательности букв, если кодирование осуществлять блоками такой длины, при которой справедлива теорема об их асимптотической равновероятности.

Расширение объема алфавита знаков приводит к повышению пропускной способности канала.

Пропускная способность дискретного канала с помехами. При наличии помех соответствие между множествами символов на входе и выходе канала связи перестает быть однозначным.

Скорость передачи информации по каналу с помехами, если объем алфавита входных символов U равен m_1, а выходных символов V — m_2:

$$\bar{I}(V,U) = V_\tau \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} p(V_j U_i) = \log \frac{p(V_j U_i)}{p(V_j)\, p(U_i)}. \qquad (2.19)$$

Предельное значение $C_д$ скорости передачи информации по каналу называют *пропускной способностью* дискретного канала связи с помехами:

$$C_д = \max_{p(u)} V_\tau \bar{I}(V,U), \qquad (2.20)$$

где $p(u)$ — множество распределений вероятностей выходных сигналов.

Важно, что при наличии помех пропускная способность канала определяет наибольшее количество информации в единицу времени, которое может быть передано со сколь угодно малой вероятностью ошибки. Произвольно малая вероятность ошибки достижима только в пределе, когда длина блоков становится бесконечной.

Предельные возможности канала никогда не используются. Степень загрузки характеризуется *коэффициентом использования канала*

$$\lambda = \frac{\bar{I}(z)}{C_д}, \qquad (2.21)$$

где $\bar{I}(z)$ — производительность источника сообщений; $C_д$ — пропускная способность канала связи.

Согласование физических характеристик и статистических свойств источника сообщений и канала связи

Конкретный канал связи обладает определенными физическими параметрами: временем, в течение которого он предоставляется для передачи сигнала T_k, шириной полосы пропускания сигнала F_k и допустимым превышением сигнала над помехой в канале H_k.

Превышение характеризуется разностью максимально допустимого сигнала в канале и уровня помех (в логарифмическом масштабе). Для проводных каналов превышение в основном определяется пробивным напряжением и уровнем перекрестных помех, для радиоканалов — возможностями выявления сигнала на соответствующем расстоянии.

Произведение этих параметров принято называть объемом (емкостью) канала и обозначать V_k:

$$V_k = T_k \cdot F_k \cdot H_k. \qquad (2.22)$$

Аналогично объему канала вводится понятие объема (емкости) передаваемого сигнала:

$$V_c = T_c \cdot F_c \cdot H_c. \qquad (2.23)$$

Необходимо: $V_c \le V_k$.

Когда канал имеет полосу пропускания меньше, чем практическая ширина спектра, подлежащего передаче, ее можно уменьшить за счет увеличения длительности сигнала. Объем сигнала при этом останется постоянным.

Пример: Запись сигнала на магнитную ленту и проигрывание при меньшей скорости, когда ширина спектра канала равна полосе пропускания. **Обратный пример:** Экономия времени передачи сигнала на широкополосном канале — проигрывание на большей скорости. При низком уровне превышения сигнала — многократное его повторение.

Согласование статистических свойств источника сообщений и канала связи проводится с целью улучшения качества системы передачи. Оценка качества происходит по достоверности и средней скорости передачи.

Достоверность дискретного канала оценивается значением вероятности ошибочного приема одного символа. Достоверность характеризует помехоустойчивость информационной системы.

Под скоростью передачи подразумевают среднее количество информации, передаваемое по каналу в единицу времени. Именно эта (а не техническая) скорость формирования символов подлежит согласованию с пропускной способностью канала. Скорость передачи характеризует эффективность системы.

В простейшем случае преобразование сообщений в сигналы происходит для обеспечения простоты, надежности и эффективности аппаратуры, а также для защиты их от несанкционированного доступа (*шифрование*). Шифрование может происходить как на уровне знаков, так и на уровне символов.

В более сложном случае за счет введения в канал связи кодирующего (и декодирующего) устройства возможно увеличение эффективности и помехоустойчивости системы передачи информации (Шеннон).

3. Кодирование информации

Кодирование информации при передаче по дискретному каналу без помех

Кодирование как процесс. Любому дискретному сообщению или знаку можно приписать какой-либо порядковый номер. Измерение аналоговой величины, выражающееся в сравнении ее с образцовыми мерами, также приводит к числовому представлению информации. Передача или хранение сообщений при этом сводится к передаче или хранению чисел.

Общепризнанным в настоящее время является позиционный принцип образования системы счисления. Значение каждого символа (цифры) зависит от его положения — позиции в ряду символов, представляющих число. Число выразится как:

$$Q = \sum_{i=1}^{l} a_i m^{i-1} = a_i m^{l-1} + a_{i-1} m^{l-1} + ... + a_2 m^1 + a_1 m^0 , \qquad (3.1)$$

где m — основание системы счисления; i — номер разряда данного числа; l — количество разрядов; a_i — количество единиц i-го разряда в числе.

Чем больше основание системы счисления, тем меньшее число разрядов требуется для представления данного числа, значит — меньшее время для его передачи. Однако с ростом основания аппаратура должна иметь большее число устойчивых состояний и выше стоимость ее создания.

Целесообразно выбрать систему, обеспечивающую минимум произведения количества разных сигналов m на количество разрядов l для выражения любого числа. Для $Q \approx 60\,000_{10}$ приведем таблицу:

m	l	$m \times l$
1	60 000	60 000
2	16	32
3	10	30
4	8	32
16	4	64
40	3	120
60 000	1	60 000

Отчетливо видно, что наиболее эффективной системой является ***троичная!*** Незначительно уступают двоичная и четверичная. Существенно хуже 10-ричная и более. Технически удобнее (надежнее) создавать двоичные устройства.

Технические средства представления информации в цифровой форме

Аналого-кодовые преобразователи. Только для IBM PC выпускается более 2 000 плат, позволяющих преобразовать сигналы от различных датчиков в цифровую форму (замена КАМАК на ISA и другие шины компьютеров), и проводить их дальнейшую обработку.

Пример: Кодовые датчики геометрических координат (линейные и угла поворота, абсолютных координат и относительного перемещения).

Подавляющее большинство сигналов от первичных датчиков получают в непрерывной аналоговой форме (изменения постоянного напряжения или тока). Для их представления в цифровой форме используют преобразователи напряжение — код, и именно такой преобразователь подразумевают под термином «***аналого-цифровой преобразователь***» (АЦП) по умолчанию. Виды АЦП:

— преобразователи последовательного счета, где ток предварительно преобразуется в импульсы, число которых соответствует измеряемому значению;

— преобразователи поразрядного уравновешивания, где последовательно изменяется на единицу младшего разряда уровень компенсирующего напряжения и сравнивается с измеряемым сигналом;

— преобразователи считывания, где входное напряжение u_x подается на схемы сравнения опорного делителя напряжения (дешифратора). Измерение происходит при сравнении измеряемого напряжения с опорным.

Кодирование как средство криптографического закрытия информации

Серьезная защита невозможна без системы охраны территории, регулирования доступа в помещения, устройств идентификации пользователей и т.п. Методы защиты непосредственно передаваемой информации представляют собой такое преобразование сообщений, при котором их исходное содержание становится доступным лишь при наличии ключа и выполнении обратного преобразования.

Рассмотрим методы криптографического закрытия информации.

Шифр *прямой подстановки* — наиболее простой, не обеспечивает высокой степени защиты. Шифр *Вижинера* состоит в том, что ключом к нему служит некоторое слово, подписываемое буква за буквой под шифруемым текстом. Цифровые коды букв текста и ключа складываются и передаются. Для длинных ключей шифр достаточно надежен. Шифрование *гаммированием* — цифровые знаки текста сообщения складываются с псевдослучайной последовательностью чисел, именуемой гаммой. Наиболее часто этот метод применяется для шифрования двоичных сообщений. Надежность определяется в основном длиной n неповторяющейся части гаммы. Если удается получить исходные тексты длиной $2n$ символов, текст может быть расшифрован.

Эффективное кодирование

Учитывая статистические свойства источника сообщения, можно минимизировать количество символов, требующееся для выражения одного знака сообщения, что позволяет уменьшить время передачи или объем запоминающего устройства.

Основная теорема Шеннона о кодировании для канала без помех.

1. При любой производительности источника сообщений, меньшей пропускной способности канала: $\bar{I}(z) = C_д - \varepsilon$***; где*** ε ***— сколь угодно малая величина*** > 0***, существует способ кодирования, позволяющий передавать по каналу все сообщения источника.***

2. Не существует способа кодирования, обеспечивающего передачу сообщений без их неограниченного накопления, если: $\bar{I}(z) > C_д$.

Используется и другая формулировка: ***сообщения источника с энтропией H(z) всегда можно закодировать последовательностями символов с объемом алфавита m так, что среднее число символов на знак сообщения*** $l_{ср}$ ***будет сколь угодно близко к величине*** $\dfrac{H(z)}{\log m}$***, но не менее ее.***

Методы эффективного кодирования

Теорема не указывает способа кодирования. Первый из методов эффективного кодирования — *код Шеннона-Фано*:

Знаки алфавита сообщения выписывают в таблицу в порядке убывания вероятностей. Затем их разделяют на две группы так, чтобы суммы вероятностей

в каждой из групп были по возможности одинаковы. Всем знакам верхней половины в качестве единичного символа присваивают 0, а всем нижним — единицу. Каждую из полученных групп, в свою очередь, разбивают на две подгруппы с одинаковыми вероятностями и присваивают 2-й знак, и продолжают процесс до исчерпания символов.

Методика годится не для всех возможных алфавитов. Методика Хаффмана гарантирует однозначное построение кода с наименьшим средним числом знаков на букву.

Более удобна методика, использующая принцип префиксности кодов, при которой эффективный код строится так, чтобы ни одна комбинация кода не совпадала с началом более длинной комбинации.

Недостатки системы эффективного кодирования:

— различие в длине кодовых комбинаций. Это требует буферов на обеих сторонах канала;

— задержка в передаче и при декодировании;

— помехонезащищенность.

Кодирование при передаче по дискретному каналу с помехами

Основная теорема Шеннона о кодировании для канала с помехами.

1. При любой производительности источника сообщений, меньшей, чем пропускная способность канала, существует такой способ кодирования, который позволяет обеспечить передачу всей информации, создаваемой источником, со сколь угодно малой вероятностью ошибки.

2. Не существует способа кодирования, позволяющего вести передачу информации со сколь угодно малой вероятностью ошибки, если производительность источника сообщений больше пропускной способности канала.

Теорема устанавливает теоретический предел возможной эффективности системы при достоверной передаче информации. Из теоремы следует, что помехи в канале не накладывают ограничений на точность передачи (накладывают на скорость). Она мобилизовала ученых на разработку помехоустойчивых кодов.

Разновидности помехоустойчивых кодов

Высокие требования к достоверности передачи, обработки и хранения информации в современных вычислительных системах и сетях требовали использования помехоустойчивого кодирования с возможностями обнаружения и исправления ошибок.

Это достигается ценой введения при кодировании избыточности, позволяющей обнаруживать и исправлять ошибки. Коды, обладающие такими свойствами, называют *помехоустойчивыми*. Они делятся на *блоковые* и *непрерывные*.

Для блоковых кодов процедура кодирования заключается в сопоставлении каждой букве сообщения (кодируемой k символами) блока из n символов. В операциях по преобразованию принимают только k символов, и выходная по-

следовательность не зависит от других символов в сообщении. Блоковый код называют *равномерным*, если n остается постоянным для всех букв сообщения.

Различают *разделимые* и *неразделимые* блоковые коды. Для разделимых кодов выходные последовательности состоят из раздельных информационных символов и проверочных символов. Неразделимые символы нельзя разделить информационные и проверочные.

Непрерывными (древовидными) называют коды, в которых введение избыточных символов в кодируемую последовательность осуществляется непрерывно, без разделения на независимые блоки. Непрерывные коды также могут быть разделимыми и неразделимыми.

Блоковые коды

Общие принципы использования избыточности. На вход двоичного кодирующего устройства поступает k символов. На выходе ему соответствует n ($n > k$) символов. Всего может быть 2^k входных (*разрешенных*) и 2^n выходных последовательностей, $2^n - 2^k$ являются *запрещенными комбинациями*.

Искажение информации в процессе передачи сводится к тому, что некоторые символы заменяются другими — неверными. Всего $2^n \cdot 2^k$ возможных вариантов:

— 2^k случаев безошибочной передачи;

— 2^k ($2^k - 1$) случаев перехода в другие разрешенные комбинации (необнаруживаемые ошибки);

— 2^k ($2^n - 2^k$) случаев перехода в неразрешенные комбинации, которые могут быть обнаружены.

Вероятность обнаружения ошибки: $2^k \dfrac{2^n - 2^k}{2^n \cdot 2^k} = 1 - \dfrac{2^k}{2^n}$.

Исправление ошибок. При получении запрещенной комбинации, принимают решение, что передавалась разрешенная комбинация A_i. Ошибка будет исправлена, если полученная комбинация действительно образовалась из A_i, т.е. в $2^n - 2^k$ случаях.

Большинство разработанных кодов предназначено для корректирования взаимно независимых ошибок определенной кратности и пачек ошибок. Степень отличия любых двух кодовых комбинаций назовем кодовым расстоянием d (число единиц в сумме по модулю 2).

$$\begin{array}{rl} & 1001111101 \\ \oplus & 1100001010 \\ \hline & 0101110111 \end{array}, \; d = 7.$$

При декодировании принятая комбинация отождествляется с той разрешенной, которая находится от нее на наименьшем кодовом расстоянии (метод максимального правдоподобия).

При $d = 1$ все кодовые комбинации являются разрешенными (при $n = 3$ подходят: 000, 001, 010 и т.д.). Если $d = 2$, ни одна из разрешенных комбинаций

при одиночной ошибке не переходит в другую разрешенную. Код обнаруживает одиночные ошибки и ошибки нечетной кратности (при $n = 3$ — тройные). В общем случае при необходимости обнаруживать ошибки кратностью до r включительно минимальное расстояние между разрешенными кодовыми комбинациями должно быть по крайней мере на единицу больше r. Для исправления одиночной ошибки каждой разрешенной комбинации необходимо сопоставить подмножество запрещенных кодовых комбинаций. Чтобы эти подмножества не пересекались, расстояние должно быть не менее 3. При $n = 3$ за разрешенные комбинации можно принять 000 или 111. Тогда (для 000) запрещенные комбинации 001, 010, 100 образуются в результате однократной ошибки.

Для разрешенной комбинации 111 запрещенные 110, 011, 101.

В общем случае для обеспечения возможности исправления всех ошибок кратностью до s включительно при декодировании по методу максимального правдоподобия каждая из ошибок должна приводить к запрещенной комбинации, относящейся к подмножеству исходной разрешенной комбинации и $d_{\min} \geq 2s + 1$.

Для пачек ошибок возможны меньшие расстояния, обеспечивающие исправление ошибок. Обычно же корректирующие коды предназначены для исправления комбинаций ошибок, наиболее вероятных для заданного канала и наиболее опасных по последствиям.

Линейные коды

Самый большой класс разделимых кодов составляют линейные коды, у которых значения проверочных символов определяются при проведении линейных операций над информационными символами. Они обязательно являются групповыми. Для двоичных кодов каждый проверочный символ выбирают таким образом, чтобы его сумма с определенными информационными символами была равна 0. При декодировании определяется справедливость проверочных равенств (например, проверка на четность). Совокупность проверок дает информацию о том, имеется ли ошибка и какие символы искажены.

Построение двоичного группового кода

Исходным является объем кода Q. Число информационных разрядов k, необходимое для передачи нужных кодов $2^k - 1 \geq Q$. Каждой из $2^k - 1$ ненулевых комбинаций k-разрядного безызбыточного кода необходимо поставить в соответствие комбинацию из n символов. Значения символов в $n - k$ проверочных разрядах такой комбинации устанавливаются в результате суммирования по модулю 2 значений символов в определенных информационных разрядах. Это множество разрешенных кодовых комбинаций и будет групповым кодом. Требуется лишь определить число проверочных разрядов и номера информационных разрядов для определения символов в проверочных разрядах. Из общего

числа $2^n - 1$ возможных ошибок групповой код может исправить всего $2^{n-k} - 1$ разновидностей ошибок.

Например, для передачи 15 команд (4 информационных разряда) и возможности исправления только единичных ошибок требуется 7-разрядный код. Дополнительным достоинством такого кода будет возможность обнаружения двойных ошибок.

Построение циклических кодов

Любой групповой код (n, k) может быть записан в виде матрицы, включающей k линейно независимых строк по n символов и, наоборот, любая совокупность k линейно независимых n разрядных кодовых комбинаций может рассматриваться как образующая матрица некоторого группового кода. Среди таких кодов можно выделить такие, у которых строки образующих матриц связаны дополнительным условием цикличности. Все строки образующей матрицы такого кода могут быть получены циклическим сдвигом одной комбинации, называемой образующей для данного кода (отсюда и название). Сдвиг осуществляется справа налево, причем крайний левый символ каждый раз переносится в конец комбинации:

$$G = \begin{vmatrix} 001011 \\ 010110 \\ 101100 \\ 011001 \\ 110010 \\ 100101 \end{vmatrix}.$$

При описании циклических кодов кодовые комбинации представляются в виде многочленов фиктивной переменной x. Например, комбинация 01011 запишется как $G(x) = 0 \cdot x^4 + 1 \cdot x^3 + 0 \cdot x^2 + 1 \cdot x + 1$. Получится: $G(x) = x^3 + x + 1$.

Любая разрешенная кодовая комбинация делится на образующий многочлен без остатка, а ни один многочлен, соответствующий запрещенной кодовой комбинации без остатка не делится. Это свойство позволяет обнаруживать и исправлять ошибки.

Итеративные коды

Для итеративных кодов характерно, что операции кодирования проводятся над совокупностью информационных символов, располагаемых по нескольким (q) координатам. Число информационных символов в кодовом векторе:

$$m = \sum_{y=1}^{q} m_y , \qquad (3.2)$$

где m_y — число символов по координате y. Последовательности информационных символов по каждой из координат кодируются каким-либо линейным кодом. В общем случае каждый символ входит одновременно в q различных кодовых векторов.

Классический итеративный код (код П. Элайеса): линейным кодом кодируется каждая из отдельных последовательностей символов по координате y_i (например, каждая строка).

В двухстепенном коде с проверкой на четность по строкам и столбцам (запись на магнитную ленту и т.п.) значения проверочных символов, располагающихся в крайнем правом (или другом) столбце и нижней строке, определяются уравнениями:

$$a_{jn} = \sum_{i=1}^{n-1} a_{ij} \,(\mathrm{mod}\, 2); \quad a_{li} = \sum_{j=1}^{l-1} a_{ji} \,(\mathrm{mod}\, 2); \quad a_{ln} = \sum_{j=1}^{l-1} a_{jn} \,(\mathrm{mod}\, 2). \qquad (3.3)$$

Передачу такого символа обычно осуществляют последовательно символ за символом, от строки к другой. Проверка справедливости (3.3) при декодировании позволяет исправить любое нечетное число искаженных символов, расположенных в одной строке или столбце.

	1	2	.	.	.	i	.	.	.	n–1	n
1	a_{11}	a_{12}				$a_{1\,i}$				$a_{1\,n-1}$	$a_{1\,n}$
2	a_{21}	a_{22}				$a_{2\,i}$				$a_{2\,n-1}$	$a_{2\,n}$
.											
j	a_{j1}	$a_{j\,2}$				$a_{j\,i}$				$a_{j\,n-1}$	$a_{j\,n}$
.											
l–1	$a_{l-1\,1}$	$a_{l-1\,2}$				$a_{l-1\,i}$				$a_{l-1\,n-1}$	$a_{l-1\,n}$
l	$a_{l\,1}$	$a_{l\,2}$				$a_{l\,i}$				$a_{1\,n-1}$	$a_{1\,n}$

Литература

1. Дмитриев В.И. Прикладная теория информации. —М.: Высш. шк., 1989.

2. Шеннон К. Работы по теории информации и кибернетике. —М.: ИЛ., 1963.

3. Хемминг Р.В. Теория кодирования и теория информации. —М.: Радио и связь, 1983.

4. Герасименко В.А. Защита информации в автоматизированных системах обработки данных. В 2 т. —М.: Энергоатомиздат, 1994.

5. Радиотехнические системы передачи информации / Под ред. В.В. Калмыкова. —М.: Радио и связь, 1990.

6. Куликовский Л.Ф., Мотов В.В.. Теоретические основы информационных процессов. —М.: Высш. шк., 1987.

Св. план 2002, поз. 54.

Учебное издание

Лукин Евгений Сергеевич

ПРИКЛАДНАЯ ТЕОРИЯ ИНФОРМАЦИИ

Учебное пособие
для студентов специальности «Информатика»

Редактор Е.Н. Батурчик
Компьютерная верстка Т.В. Шестакова

Подписано в печать 6.11.2002. Формат 60×84 $^{1}/_{16}$.
Бумага офсетная. Печать ризографическая. Гарнитура «Таймс». Усл. печ. л 2,67.
Уч.-изд. л. 2,3. Тираж 50 экз. Заказ 213.

Издатель и полиграфическое исполнение:
Учреждение образования
«Белорусский государственный университет информатики и радиоэлектроники».
Лицензия ЛП № 156 от 05.02.2001.
Лицензия ЛВ № 509 от 03.08.2001.
220013, Минск, П. Бровки, 6.