МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»

«ТЕОРІЯ ІНФОРМАЦІЇ ТА ОБРОБКА СИГНАЛІВ-1»

КОНСПЕКТ ЛЕКЦІЙ

*Рекомендовано Методичною радою КПІ ім. Ігоря Сікорського
як навчальний посібник для студентів,
які навчаються за спеціальністю 171 «Електроніка»,
освітньою програмою «Електронні компоненти і системи»*

Київ
КПІ ім. Ігоря Сікорського
2020

«Теорія інформації та обробка сигналів-1»: конспект лекцій [Електронний ресурс]: навч. посіб. для студ. спеціальності 171 «Електроніка», освітньої програми «Електронні компоненти і системи» / КПІ ім. Ігоря Сікорського ; уклад.: Ю.С. Ямненко, К. С. Клен. – Електронні текстові данні (1 файл: 2,107 Мбайт). – Київ: КПІ ім. Ігоря Сікорського, 2019. – 120 с.

Гриф надано Методичною радою КПІ ім. Ігоря Сікорського (протокол № 6 від 31.01.2020 р.)
за поданням Вченої ради факультету електроніки (протокол № 01/2020 від 27.01.2020 р.)

Електронне мережне навчальне видання

«ТЕОРІЯ ІНФОРМАЦІЇ ТА ОБРОБКА СИГНАЛІВ-1»
КОНСПЕКТ ЛЕКЦІЙ

Укладачі: *Ямненко Юлія Сергіївна,* д-р техн. наук, проф.
Клен Катерина Сергіївна, канд. техн. наук, доц.

Відповідальний редактор: *Терещенко Т.О.*, професор кафедри промислової електроніки, д-р техн. наук, проф.

Рецензент: *Найда С.А.,* професор кафедри акустики та акустоелектроніки, д-р техн. наук, проф.

При підготовці бакалаврів за спеціальністю 171 Електроніка, освітньою програмою «Електронні системи» однією з важливих дисциплін, що є компонентом циклу професійної підготовки, є дисципліна «Теорія інформації та обробка сигналів-1».

Метою розробки конспекту лекцій є набуття ґрунтовних знань в галузі представлення інформації, її вираження та перетворення, кодування, а також моделювання аналогових, дискретних та цифрових сигналів, що використовуються в інформаційних електронних системах; отримання навичок подальшого застосування набутих знань в процесі навчання, при виконанні курсових та дипломної роботи, у практичній діяльності та наукових дослідженнях за фахом.

Конспект лекцій містить теоретичні відомості до 16 лекцій та список рекомендованої літератури.

ЗМІСТ

ВСТУП

З кожним роком зростають обсяги інформації, які потрібно обробляти, передавати та зберігати для ефективного використання її при вирішенні наукових, економічних та інших завдань.

Без знання теорії інформації, кодування та обробки сигналів неможливе створення нових сучасних інформаційних систем. Тому її вивчення є невід'ємною частиною теоретичної підготовки фахівців у галузі інформаційних електронних систем, комп'ютерної інженерії, телекомунікацій тощо.

Вивчення дисципліни «Теорія інформації та обробка сигналів (ТІОС)» базується на змісті таких дисциплін: Інформатика; Теорія електричних кіл; Імовірнісні основи обробки даних. Курс ТІОС є базовим для наступних дисциплін: Цифрові інформаційні системи; Електронні системи; Мікропроцесорні пристрої.

У конспекті лекцій розглянуто поняття інформації та інформаційні параметри; передавання інформації по каналах зв’язку з шумами; кодування інформації; основи цифрової обробки сигналів; диференціально-інтегральні методи обробки сигналів; цифрову обробку сигналів; цифрову фільтрацію.

Лекція 1. Інформація, її форми. Інформаційні характеристики дискретних сигналів. Дискретизація, квантування та цифрове подання неперервних сигналів. Теорема Котельникова (Найквіста, Шеннона)

Діяльність людей пов’язана з переробкою та використанням матеріалів, енергії та інформації. Відповідно розвиваються наукові технічні дисципліни, що відображують питання технології, енергетики та інформатики.

З передаванням та обробкою інформації пов’язані дії будь-якого автоматичного пристрою, поведінка живих істот, творча діяльність людини, розвиток науки і техніки, економічні та соціальні перетворення у суспільстві та саме життя.

В природі існує два основних види взаємодії: обмін речовиною та енергією. Всі інші взаємодії відбуваються тільки через них. Ці види взаємодії підкорюються закону збереження. Скільки речовини та енергії один об’єкт передав іншому, стільки той і отримав, і навпаки.

Енергетична та речовинна взаємодія об’єктів є симетричною, тобто скільки один віддав, стільки ж другий отримав. Переходи між речовиною та енергією не впливають на загальний баланс, оскільки діють закони збереження константи їх співвідношення.

На основі комбінації фундаментальних взаємодій між об’єктами може відбуватися взаємодія більш високого порядку, при якій від одного до другого переходить деяка субстанція, і при цьому втрати одного не співпадають з набуттям другого. Така взаємодія є несиметричною. В граничному випадку несиметричної взаємодії при передаванні деякої субстанції між об’єктами один з них її набуває, а другий не втрачає. Зміни кількості енергії і речовини при цьому, природньо, будуть мати місце, оскільки даний акт взаємодії має в своїй основі комбінацію фундаментальних видів взаємодії, що забезпечують перенесення субстанції.

Тепер сформулюємо найбільш загальне визначення поняття інформації, від якого ми будемо відштовхуватися в подальшому.

Будь-яка взаємодія між об'єктами, в процесі якої один набуває деякої субстанції, а другий її не втрачає, називається інформаційною взаємодією. При цьому субстанція, що передається, називається інформацією.

З цього визначення витікають дві найбільш загальні властивості інформації. Перша - інформація не може існувати без взаємодії об'єктів. Друга - інформація не втрачається жодним з об'єктів в процесі цієї взаємодії.

В основі теорії інформації лежить вимірювання кількості інформації, що міститься у якихось даних (повідомленнях).

Поняття інформації пов'язано з моделями реальних процесів, що відображають їх сутність у тому ступені, в якому це необхідно для конкретної практичної мети.

Під інформацією слід розуміти не самі предмети та процеси, а їх характеристики, відображення у вигляді чисел, формул, описів, креслень, символів, образів та інших абстрактних характеристик.

Сама по собі інформація є абстрактним поняттям, як, наприклад, математичні формули. Однак проявляється вона завжди у матеріально-енергетичній формі у вигляді ***сигналів***.

Сигнал-це змінна, яка передає або містить деякий вид інформації і яку можна переносити, виводити на екран або виконувати якісь дії. Розрізняють наступні типи сигналів:

- Мовлення(розмова по телефону, прослуховування радіо , спілкування)
- Біомедичні сигнали(наприклад , електроенцефалографія – сигнал мозку)
- Звуки та музика
- Відео- та телезображення
- Сигнали радарів ,які дослідження заданих діапазонів та пеленгують віддалені цілей

У циклі перетворення інформації можна виділити декілька фаз:

- *Сприйняття*, яке полягає в тому, що формується образ об’єкту, здійснюється його впізнання та оцінка. При сприйнятті треба відділити корисну інформацію від шумів. В результаті сприйняття отримується сигнал у формі, зручній для передавання або обробки. До фази сприйняття можуть включатися операції підготовки інформації, її нормалізації, квантування, кодування, модуляції та побудови моделей;

- *Передавання*, яке полягає у перенесенні інформації на відстань за допомогою сигналів різної фізичної природи по механічним, гідравлічним, пневматичним, акустичним, оптичним, електричним або електромагнітним каналам. Приймання інформації на другому боці каналу має характер вторинного сприйняття та включає операції боротьби з шумами;

- *Обробка* полягає у перетворенні інформації за допомогою пристроїв, що здійснюють аналогові або цифрові перетворення величин та функцій. Проміжним етапом обробки може бути зберігання у запам’ятовувальних пристроях. Вилучення інформації із запам’ятовувальних пристроїв також має характер сприйняття та пов’язано із боротьбою з шумами. З пристрою обробки інформація може виводитися людині (оператору) або безпосередньо впливати на об’єкт керування;

- *Представлення* інформації потрібно тоді, коли у циклі перетворення інформації приймає участь людина. Представлення полягає у демонстрації перед людиною умовних зображень, що містять якісні та кількісні характеристики інформації. Для цього використовуються пристрої, здатні впливати та органи чуття людини;

- *Вплив* полягає у тому, що сигнали, які несуть інформацію, виконують регулюючі або захисні дії, викликаючи зміни у об’єкті керування.

Види інформації. Інформацію можна розрізнити за галузями знань (біологічна, технічна, економічна), фізичною природою сприйняття (зорова, слухова, смакова), а також за структурно-метричними властивостями.

Для технічних застосувань найбільш придатною є класифікація за структурно-метричними властивостями (табл.1).

Вид інформації	Форми представлення інформації		
	Топологічна	Абстрактна	Лінгвістична
Подія	Точка	Судження	Знак
Величина	Лінія	Поняття	Буква
Функція	Поверхня	Образ	Слово
Комплекс	Об’єм	Система	Речення
…	…	…	…
Поле	Простір	Універсум	Фонд

До *топологічної* інформації відносяться геометричні образи, мапи місцевості, різні пласкі зображення та об’ємні об’єкти. Топологічною інформацією зручно виражати образи та ситуації, що підлягають розпізнанню.

Абстрактну інформацію використовують у дослідженнях на високому теоретичному рівні, коли потрібні узагальнення та символізація.

До *параметричної* інформації відносяться набори чисельних оцінок значень деяких параметрів (величин, що вимірюються), кількісні результати дослідження, аналізу, контролю та обліку. Параметричною інформацією найчастіше користуються у науці та техніці для вираження результатів вимірювань.

Найчастіше параметрична інформація оперує ***параметрами***, під якими будемо розуміти показники реальних фізичних процесів, подій або явищ. Параметри в залежності від задачі контролю можуть бути представлені у вигляді неперервних або дискретних функцій часу.

Класифікація інформаційних параметрів. За характером зміни у часі параметри поділяються на функціональні та сигнальні.

Функціональні параметри $\lambda_Ф$ (рис.1,а) є неперервними функціями часу, кількість градацій цих параметрів за рівнем нескінченна. Плавно змінюються у часі багато фізичних параметрів, наприклад, температура, тиск, вологість і т.ін.

Сигнальні параметри $\lambda_С$ (рис.1,б) характеризуються стрибкоподібною зміною у часі значення фізичної величини. До них відносяться, наприклад, параметри «увімкнено-вимкнено», «норма-не норма», «так-ні».

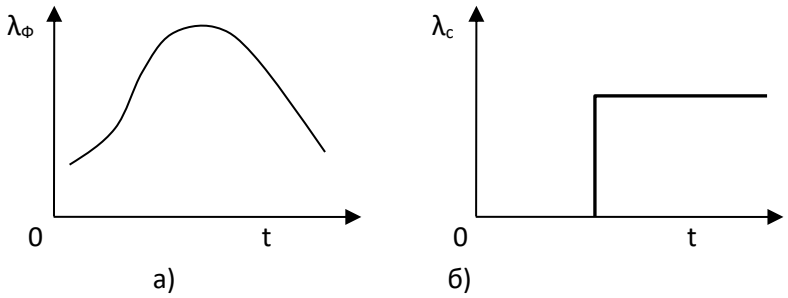


Рис.1. Функціональні та сигнальні параметри

В залежності від швидкості зміни у часі параметри поділяються на *повільно змінні* та *швидкозмінні*. Перші характеризуються шириною спектру від 0 до 20-50 Гц, а другі мають верхню границю спектра від одиниць до десятків і сотень кГц. До повільно змінних параметрів відносяться температура рідких та твердих тіл і газів, тиск, механічні та кутові переміщення, швидкість, прискорення. До швидкозмінних параметрів відносяться вібрації, акустичні шуми та перехідні процеси.

Незалежно від фізичної природи у більшості випадків зміни параметрів у часі є нестаціонарними випадковими процесами, однак можна визначити ділянки їх стаціонарності. Як будь-який випадковий процес, залежність $\lambda(t)$ можна охарактеризувати числовими характеристиками:

- Одновимірною та багатовимірними функціями розподілення;
- Одновимірною та багатовимірною густиною імовірності випадкового процесу;

- Розподіленням імовірностей випадкових дискретних величин;
- Середнім значенням або математичним чеканням;
- Дисперсією;
- Функцією кореляції (авто кореляційною та взаємно кореляційною);
- Спектральною густиною.

Інформаційні повідомлення. У реальних системах інформація передається у вигляді *повідомлень* S(t), що приймаються системою та несуть інформацію про контрольовані події і процеси, а також службову інформацію.

Для сучасних інформаційних та телеметричних систем характерні три види повідомлень:

1. Повідомлення про події типу «увімкнено-вимкнено», «так-ні», «працює-не працює». Показником такої події є стан об'єкту або системи на даний момент часу, в який відбулася подія.

2. Повідомлення, що несуть інформацію про величини параметрів у визначений момент часу. Такі повідомлення містять відомості про окремі виміри фізичних величин.

3. Повідомлення про процеси кінцевої тривалості, представлені у неперервній або дискретній формі.

Способи представлення параметрів. Неперервні сигнали описуються неперервними функціями x(t). Перехід від аналогового представлення сигналу до цифрового, який в ряді випадків дає значні переваги при передаванні, зберіганні та обробці інформації, пов'язаний з квантуванням (дискретизацією) сигналу x(t) по часу та рівню.

Існують наступні види сигналу, який описується функцією x(t):

- Неперервний – неперервна функція неперервного аргументу;
- Дискретно-неперервний (одна координата дискретна, друга – неперервна) – неперервна функція дискретного аргументу та дискретна функція неперервного аргументу;

- Дискретний - дискретна функція дискретного аргументу.

Неперервна функція неперервного аргументу (рис.2,а) характеризується тим, що значення, які може приймати функція x(t) та аргумент t, заповнюють кінцеві (або нескінченні) проміжки $[x_{min},x_{max}]$ та $[0,T]$ відповідно.

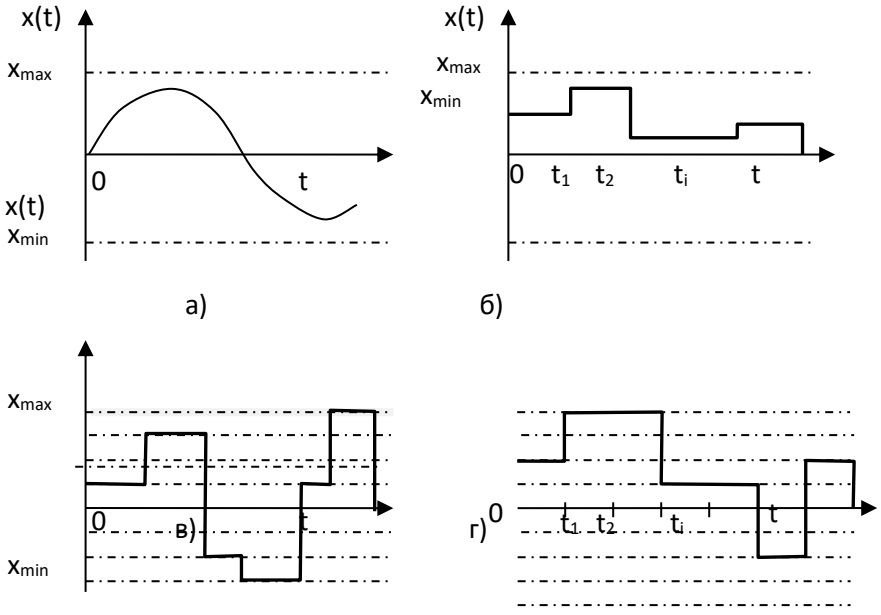


Рис.2. Види функцій (сигналів):

а – неперервна функція неперервного аргументу;

б – неперервна функція дискретного аргументу;

в – дискретна функція неперервного аргументу;

г– дискретна функція дискретного аргументу

Значення *неперервної функції дискретного аргументу* (рис.2,б) визначаються лише на дискретній множині значень аргументу t_i, $i = 0,1,2,...; t \in [0,T]$. Величина $x(t_i)$ може приймати будь-яке значення на відрізку $[x_{min},x_{max}]$.

Значення *дискретної функції неперервного аргументу* (рис.2,в) утворюють дискретний ряд чисел $x_1,x_2,x_3,\ldots, x_k,\ldots$, тобто такий кінцевий або нескінченний ряд, в якому кожному числу x_k, можна поставити у відповідність інтервал (a_k,b_k),

всередині якого інших чисел даного ряду немає. Значення аргументу t може бути будь-яким на відрізку [0,T].

Значення *дискретної функції дискретного аргументу* (рис.2,г) та її аргумент t, утворюють дискретні ряди чисел $x_1,x_2,x_3,\ldots, x_k,\ldots$ та $t_1,t_2,t_3,\ldots, t_i,\ldots$, які заповнюють відрізки $[x_{min},x_{max}]$ та [0,T] відповідно.

Звичайно операцію, яка переводить неперервний сигнал у другий вид, називають *квантуванням за часом*, або *дискретизацією*. Отже, дискретизація полягає у перетворенні сигналу x(t) неперервного аргументу t у сигнал $x(t_i)$ дискретного аргументу t_i.

Квантування за рівнем полягає у перетворенні неперервної множини значень сигналу $x(t_i)$ у дискретну множину значень x_k, k=0,…,m-1, $x_k \in [x_{min}, x_{max}]$ (третій вид сигналу).

Сумісне застосування операцій дискретизації та квантування дозволяє перетворити неперервний сигнал x(t) у дискретний по координатам x та t (четвертий вид сигналу).

Класифікація методів дискретизації. Методи дискретизації можна розділити на декілька груп у відповідності від обраних ознак класифікації. На рис.3 наведено ці ознаки та відповідні їм групи методів.

Рис.3. Класифікація методів дискретизації

Вибір частоти відліків за теоремою Котельникова. Вченим В.А.Котельниковим доведено теорему для сигналів з обмеженим спектром, згідно якої сигнал повністю визначається дискретною множиною значень, якщо його дискретизація здійснюється з частотою

$$F_0 \geq 2 f_{\max},$$

де f_{max} – максимальна частота у спектрі, або найбільш високочастотний компонент сигналу x(t). В цьому випадку сигнал x(t) може бути відновлений без похибок за точними значеннями $x(t_i)$.

Також має місце і наступне твердження: Якщо максимальна частота в сигналі перевищує половину частоти дискретизації, то способу відновити сигнал з дискретного в аналоговий без спотворень не існує.

Наведена теорема має назви *теорема Котельникова*, *теорема про дискретне представлення*, *теорема Найквіста*, *теорема відліків*.

Отже, якщо максимальна частота аналогового сигналу складає 4 кГц, то для того, щоб зібрати або зберегти всю інформацію, яка міститься у сигналі, його дискретизація повинна здійснюватися з частотою 8 кГц або більше. Дискретизація з частотою, меншою за ту, яку дає теорема Котельникова, призведе до появи перегинів або накладанні дзеркальних частот у частотній області. Отже, якщо потрібно буде перетворити отриманий дискретизований сигнал назад у аналоговий, початковий сигнал буде вже неможливо відновити. Важливо пам’ятати про те, що часто значна доля енергії сигналу може потрапляти за межі частотної області, що розглядається, та/або сигнал може містити шум, ширина смуги якого буде завжди великою. Наприклад, у телефонному зв’язку найвища частота складає приблизно 3,4 кГц, але частоти мовного сигналу можуть перевищувати 10 кГц. Тому, якщо не видалити зайвий сигнал або шум за межі смуги частот, що розглядається, умова теореми Котельникова не буде виконуватися. На практиці це досягається шляхом попереднього пропускання сигналу через аналоговий фільтр захисту від накладання спектрів.

Лекція 2. Ентропія як міра невизначеності процесу. Ентропія бінарних повідомлень

Невизначеність кожної ситуації характеризується величиною, що називається ентропією.

Ентропія у термодинаміці означає ймовірність теплового стану речовини, в математиці - ступінь невизначеності завдання, в інформатиці вона характеризує здатність джерела віддавати інформацію. Ентропія широко застосовується для опису стану механічних та термодинамічних систем, для вивчення властивостей алфавітів різних мов, при дослідженні економічних систем.

Ентропія (інформаційна) - міра хаотичності інформації, невизначеність появи будь-якого символу алфавіту. За відсутності інформаційних втрат ентропія чисельно дорівнює кількості інформації на символ переданого повідомлення.

Так, візьмемо, наприклад, послідовність символів, що складають якесь речення. Кожен символ з'являється з різною частотою, отже, невизначеність появи для деяких символів більше, ніж для інших. Якщо ж врахувати, що деякі поєднання символів зустрічаються дуже рідко, то невизначеність ще більше зменшується.

Інформаційна ентропія для незалежних випадкових подій x з n можливими станами (номер стану i змінюється від 1 до n) розраховується за формулою Шеннона:

$$H(x) = -\sum_{i=1}^{n} p(i) \log_2 p(i)$$

Ця величина також називається середньою ентропією повідомлення. Величина

$$\log_2 \frac{1}{p(i)}$$

називається частковою ентропією, що характеризує тільки і-й стан.

Таким чином, ентропія події x є сумою з протилежним знаком всіх добутків відносних частот появи події i, помножених на їх двійкові логарифми (основу логарифму 2 вибрано тільки для зручності роботи з інформацією, представленою в двійковій формі).

Якщо всі стани системи є рівноімовірними, то для обчислення ентропії використовується формула Хартлі, яка визначає кількість інформації, що міститься в повідомленні довжиною *n*.

Для розуміння формули Хартлі розглянемо алфавіт A, з літер якого складене повідомлення:

$|A|=m$.

Кількість модливих варіантів різних повідомлень:

$N=m^n$,

де N – можлива кількість різних повідомлень, m – кількість літер в алфавіті, n – кількість літер в повідомленні.

Приклад. Алфавіт складається з двох літер «В» и «Х», довжина повідомлення – 3 літери. Таким чином, m = 2, n = 3. При обраних нами алфавіті та довжині повідомлення можна скласти $N=m^n=2^3=8$ різних повідомлень: «ВВВ», «ВВХ», «ВХВ», «ВХХ», «ХВВ», «ХВХ», «ХХВ», «ХХХ» – інших варіантів немає.

Формула Хартлі визначається:

$$I=\log_2 N=n*\log_2 m,$$

де I – кількість інформації в бітах.

При рівноімовірності символів $p=1/m$, $m=1/p$ формула Хартлі переходить у власну інформацію.

Формула Хартлі була запропонована Ральфом Хартлі в 1928 році як один з наукових підходів до оцінки повідомлень.

Приклад. Чому дорівнює ентропія системи, яка складається:

а) з двох елементів, кожний з яких може з рівною імовірністю знаходитися у двох станах?

б) з трьох елементів, кожний з яких може з рівною імовірністю знаходитися у чотирьох станах?

За формулою Хартлі:

$H = \log_2 M$

$M = m^n$

де m-кількість елементів

n- кількість станів

Розв'язок.

а) кількість елементів системи m=2, кількість станів n=2. Ентропія

$H = \log_2 m^n = n \log_2 m = 2 \log_2 2 = 2$ *біт / стан*;

б) кількість елементів системи m=3, кількість станів n=4. Ентропія

$H = \log_2 m^n = n \log_2 m = 4 \log_2 3 = 6,32$ *біт / стан*.

В даному прикладі розглядалися рівноімовірні стани, тому для обчислення ентропії використовувалась формула Хартлі.

Якщо між елементами системи не спостерігається ніяких кореляційних зв'язків, а стани системи нерівноімовірні, то ентропія такої системи обчислюється за формулою Шеннона.

Приклад. Чому дорівнює ентропія системи, стан якої описується дискретною величиною з наступним розподіленням імовірностей станів

a_i	a_1	a_2	a_3	a_4
p_i	0,1	0,2	0,3	0,4

Розв'язок.

Ентропія

$$H = -\sum_{i=1}^{m} p_i \log_2 p_i = -(0,1 \log_2 0,1 + 0,2 \log_2 0,2 + 0,3 \log_2 0,3 +$$

$$+0,4 \log_2 0,4) = 1,84644 \text{ біт / стан}$$

В розглянутих двох прикладах стани елементів умовних джерел повідомлень не залежать один від одного. Ентропія джерел, у яких елементи ансамблів повідомлень взаємозалежні, називається *ентропією нульового порядку*.

При дослідженні властивостей ентропії найбільший інтерес представляє її залежність від кількості m можливих ознак (якостей) та імовірності p_i появи у повідомленні елементу з і-ю ознакою.

Ентропія характеризує міру невизначеності сукупності подій, що складають повну групу (сума імовірностей появи окремих подій повинна бути рівною 1, тобто $\sum_{i=1}^{m} p_i = 1$. Якщо m=i=1, тобто передається повідомлення з однією і-ю ознакою, імовірність її появи $p_i = 1$, то

$$H = \sum_{i=1}^{m} p_i \log_2 \frac{1}{p_i} = 1 \log 1 = 0 .$$

Це очевидно, оскільки заздалегідь відомо, що буде передано повідомлення з і-ю ознакою і при його отриманні нічого нового ми не дізнаємось, тобто отримаємо нульову інформацію. Якщо повідомлення заздалегідь відомо, то ентропія мінімальна і дорівнює 0.

Якщо імовірність появи і-ї ознаки у ансамблі повідомлень дорівнює нулю, то доданок з цією ознакою приймає вид невизначеності типу нуль, помножений на нескінченність. Дійсно,

$$p_i \log_2 \frac{1}{p_i} = 0 \bullet \infty .$$

Розкриємо цю невизначеність, скориставшись правилом Лопіталя. Для цього перш за все невизначеність типу $0 \bullet \infty$ приведемо до вигляду ∞ / ∞:

$$\lim_{p_i \to 0}\left(p_i \log_2 \frac{1}{p_i} \right) = \lim_{p_i \to 0}\left(\frac{\log \frac{1}{p_i}}{\frac{1}{p_i}} \right).$$

Позначимо $\frac{1}{p_i} = k$ (при $p_i \to 0 \;\; k \to \infty$). Тоді можна записати

$$\lim_{p_i \to 0}\left(\frac{\log \frac{1}{p_i}}{\frac{1}{p_i}} \right) = \lim_{k \to \infty}\left(\frac{\log k}{k} \right) .$$

Згідно правилу Лопіталя,

$$\lim_{k \to \infty} \frac{\log k}{k} = \frac{d}{dk}\left(\frac{\log k}{k} \right),$$

але похідна $\left(\log k\right)' = \frac{1}{k}\log e$, тобто

$$\lim_{k \to \infty} \frac{\log k}{k} = \frac{\frac{1}{k}\log e}{1} = 0.$$

Відомо, що якщо окремі доданки прямують до нуля, то до нуля прямує і сума, тобто остаточно можна записати

$$\lim_{p_i \to 0} \sum_{i=1}^{m} p_i \log_2 \frac{1}{p_i} = 0 \; .$$

Таким чином, знайдемо дві граничних умови, при яких ентропія мінімальна і дорівнює 0.

Дослідимо тепер вираз для ентропії на екстремум. Для цього, як відомо, необхідно знайти точку, в якій похідна функції змінює знак, тобто

$$\frac{d}{dp_i}(-p_i \log_2 p_i) = -(1 \cdot \log_2 p_i + p_i \cdot \frac{1}{p_i \ln 2}) = -(\log_2 p_i + \log_2 e) = \log_2 p_i \cdot e$$

$$\frac{d}{dp_i}(-p_i \log_2 p_i) = 0 \Leftrightarrow -\log_2 p_i \cdot e = 0 \Rightarrow p_i \cdot e = 1 \Rightarrow p_i = \frac{1}{e} = 0,37 \quad ,$$

$$-p_i \log_2 p_i\big|_{p_i=\frac{1}{e}} = \frac{1}{e} \cdot \log_2 e = 0,531$$

Отже, ентропія є величиною дійсною та має екстремум. Оскільки відомо, що логарифми правильних дробів є від'ємними, то ентропія досліду з кінцевою кількістю виходів є завжди позитивною.

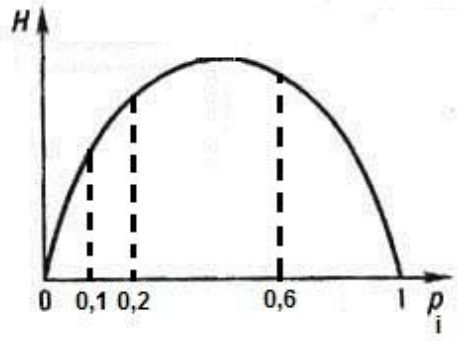


Рис.1. Графік функції $f(p_i) = -p_i \log p_i$

Графік функції $f(p_i) = -p_i \log p_i$ зображено на рис.1. Цей графік являє цікавість з тієї точки зору, що дозволяє оцінити вплив імовірності появи окремого символу на величину виразу ентропії для повідомлення в цілому. Як видно з графіка, при $p_i < 0,1$, величина $-p_i \log p_i$ зростає круто. Це означає, що на даній ділянці навіть незначне зменшення імовірності p_i веде до різкого зменшення доданку $-p_i \log p_i$, тобто при малих значеннях імовірності p_i відповідні доданки у виразі для ентропії не відіграють суттєвої ролі, часто ними можна знехтувати.

З рис.1 також видно, що найбільші значення доданків вигляду $-p_i \log p_i$ наявні при імовірностях появи символів з і-ю ознакою, які лежать в області від 0,2 до 0,6. Це зрозуміло, оскільки при малих імовірностях появи і-ї ознаки легко передбачити її відсутність у повідомленні, і навпаки. В обох випадках величина

невизначеності, яка існувала до отримання повідомлення, буде малою. Відповідно малою є і кількість інформації при знятті цієї невизначеності.

Особливий інтерес представляють бінарні повідомлення, що використовують алфавіт з двох знаків: (0,1). При m=2 сума імовірностей знаків алфавіту $P_1 = P_2 = 1$. Можна виразити $P_1 = P$,тоді $P_2 = 1 - P$.

Ентропію можна визначити по формулі:

$$H = -p_1 \log p_1 - p_2 \log p_2 = -p \log p - (1-p)\log(1-p)$$

Ентропія бінарних повідомлень досягає максимального значення, рівного 1 біту, коли знаки алфавіту повідомлень рівноімовірні, тобто при Р=0,5,та її графік симетричний відносно цього значення (рис.2).

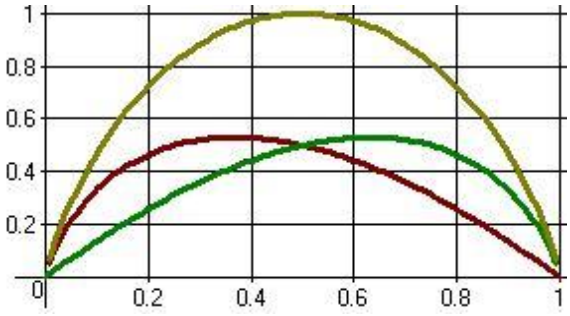


Рис.2. Графік залежності ентропії Н двійкових повідомлень та її складові:

$$-(1-p)\log_2(1-p) \text{ та } -p\log_2 p \text{ від Р}$$

Для будь-якої кількості символів середня кількість інформації на один символ досягає максимуму в тому випадку, колі всі символи використовуються з рівними ймовірностями.

Ще одна властивість ентропії - ентропія повідомлення, що складається з деяких часткових повідомлень, дорівнює сумі повідомлень, що його складають.

Лекція 3. Умовна ентропія

На практиці найчастіше всього повідомлення та символи у повідомленнях є взаємозалежними, якщо передавати не просто окремі літери алфавіту, а змістовні повідомлення. Одні літери зустрічаються частіше, інші рідше, одні літери та слова часто йдуть за іншими і т.д. Наприклад, у англійській мові найбільш часто зустрічається літера *e*; у французькій мові після літери *q* майже завжди стоїть літера *u*, якщо *q*, звичайно, не стоїть в кінці слова; в українському тексті після слова «ветеран» найчастіше стоїть слово «праці» або «війни»; поява у повідомленні слів «ветеран праці» дає, в свою чергу, інформацію про характер повідомлення, наприклад, що це повідомлення ближче до суспільно-політичної, ніж до театральної тематики і т.д.

Для взаємопов'язаних систем стан однієї впливає на стан іншої. В таких випадках ентропія не може бути визначена тільки на базі безумовних імовірностей.

При підрахунку середньої кількості інформації на символ повідомлення взаємозалежність враховують через умовні імовірності появи одних подій відносно інших, а отриману при цьому ентропію називають ***умовною ентропією***. У всіх випадках при обчисленні умовної ентропії в тому або іншому вигляді використовуються умовні імовірності.

Якщо при передаванні *k* повідомлень символ А з'явився *m* разів, символ В – *l* разів, а символ А разом з символом В – *n* разів, то:

- імовірність появи символа А $p(A)=\frac{m}{k}$,
- імовірність появи символа В $p(B)=\frac{l}{k}$,
- імовірність сумісної появи символів А та В $p(AB)=\frac{n}{k}$,

- умовна імовірність появи символа А відносно символа В

$$p(A/B)=\frac{p(AB)}{p(B)}=\frac{n}{l},$$

- умовна імовірність появи символа В відносно символа А

$$p(B/A)=\frac{p(AB)}{p(A)}=\frac{k}{m}.$$

Якщо відома умовна імовірність, то можна легко визначити і імовірність сумісної появи символів А та В наступним чином:

$$p(AB)=p(B)p(A/B)=p(A)p(B/A).$$

Від класичного виразу формула умовної ентропії відрізняється тим, що в ній використовуються умовні імовірності:

$$H(b_j/a_i)=-\sum_j p(b_j/a_i)\cdot\log p(b_j/a_i)$$

$$H(a_i/b_j)=-\sum_i p(a_i/b_j)\cdot\log p(a_i/b_j)$$

і-індекс, що характеризує стан джерела А.

j- індекс, що характеризує стан джерела В.

Умовні ентропії $H(b_j/a_i)$ та $H(a_i/b_j)$ - часткові.

Загальна умовна ентропія повідомлення В відносно повідомлення А характеризує кількість інформації, що міститься в будь-якому символі алфавіту, визначається усередненням по всім символам, тобто по всім станам з урахуванням імовірності появи кожного стану, і дорівнює сумі добутків безумовних імовірностей появи символів алфавіту на невизначеність, яка заміщується після того, як адресат прийняв сигнал:

$$H(B/A)=\sum_i p(a_i)\cdot H(b_j/a_i)=-\sum_i\sum_j p(a_i)\cdot p(b_j/a_i)\cdot\log p(b_j/a_i)$$

Оскільки добуток безумовної імовірності $p(a_i)$ та умовної імовірності $p(b_j/a_i)$ являє собою імовірність сумісної появи двох подій:

$$p(a_i)\cdot p(b_j/a_i)=p(a_i,b_j),$$

то можна записати:

$$H(B/A) = -\sum_i \sum_j p(a_i, b_j) \cdot \log p(b_j/a_i).$$

Поняття загальної та часткової умовних ентропій використовується при обчисленні інформаційних втрат в каналах у зв'язку з шумами.

В загальному випадку, якщо m сигналів А передаються і очікується отримання m сигналів В, вплив завад у каналі зв'язку повністю визначається канальною матрицею:

B A	b_1	b_2	…….	b_m
a_1	$p(b_1/a_1)$	$p(b_2/a_1)$	…….	$p(b_m/a_1)$
a_2	$p(b_1/a_2)$	$p(b_2/a_2)$	…….	$p(b_m/a_2)$
…...	……..	……..	…….	……..
a_m	$p(b_1/a_m)$	$p(b_2/a_m)$	…….	$p(b_m/a_m)$

Імовірності по головній діагоналі визначають правильне приймання, інші – невірне. Значення цифр є звичайно менші по мірі віддалення від головної діагоналі. За повної відсутності завад всі, крім значень головної діагоналі, нульові.

Для першого рядка $p(b_1/a_1) + p(b_2/a_1) + \ldots + p(b_m/a_1) = 1$ - цей вираз описує проходження сигналу a_1 в даному каналі.

Втрати інформації на долю сигналу a_1

$$H(b_j/a_1) = -\sum_{j=1} p(b_j/a_1) \cdot \log p(b_j/a_1)$$

Щоб урахувати втрати при передаванні всіх сигналів по даному каналу, треба додати всі часткові умовні ентропії, тобто провести подвійне додавання по і та по j. При цьому у випадку рівномірної появи сигналів на виході джерела:

$$H(B/A) = -\frac{1}{m}\sum_i \sum_j p(b_j/a_i)\cdot \log_2 p(b_j/a_i)$$

(множник 1/m з'явився, оскільки ентропія обчислюється на 1 символ).

У випадку нерівноімовірної появи символів необхідно врахувати імовірність появи кожного символу, помноживши її на відповідну часткову умовну ентропію.

При цьому загальна умовна ентропія

$$H(B/A) = -\sum_i p(a_i)\sum_j p(b_j/a_i)\cdot \log p(b_j/a_i)$$

Оскільки $p(a_i)p(b_j/a_i) = p(a_i,b_j)$, то $H(B/A) = -\sum_i \sum_j p(a_i,b_j)\cdot \log p(b_j/a_i)$.

Якщо ми досліджуємо канал з боку приймача, то, отримавши сигнал b_j, ми вважаємо, що був надісланий якийсь з сигналів $a_1, a_2, \ldots, a_i, \ldots, a_m$. При цьому канальна матриця має вигляд:

B A	b_1	b_2	…….	b_m
a_1	$p(a_1/b_1)$	$p(a_1/b_2)$	…….	$p(a_1/b_m)$
a_2	$p(a_2/b_1)$	$p(a_2/b_2)$	…….	$p(a_2/b_m)$
…...	……..	……..	…….	……..
a_m	$p(a_m/b_1)$	$p(a_m/b_2)$	…….	$p(a_m/b_m)$

Тут суми умовних імовірностей мають =1 не по рядках, а по колонках:

$$p(a_1/b_j) + p(a_2/b_j) + \ldots + p(a_i/b_j) + \ldots + p(a_m/b_j) = 1$$

Часткова умовна ентропія $H(a_i/b_j) = -\sum_{i=1}^{m} p(a_i/b_j)\cdot \log p(a_i/b_j)$, а загальна ентропія $H(A/B) = -\sum_j p(b_j)\sum_j p(a_i/b_j)\cdot \log p(a_i/b_j)$.

Безумовні імовірності джерела та приймача пов'язані співвідношенням:

$$p(b_j) = \sum_i p(a_i) \cdot p(b_j / a_i).$$

Тобто, якщо задані безумовні імовірності джерела та канальна матриці з боку джерела, то можно обчислити ентропію приймача

$$H(B) = -\sum_j p(b_j) \log p(b_j).$$

І навпаки, якщо задані імовірності вигляду $p(b_j)$ та канальна матриця, що описує канал зв'язку з боку приймача повідомлень, то:

$$p(a_i) = \sum_j p(b_j) \cdot p(a_i / b_j),$$

А значить, можна визначити ентропію джерела

$$H(A) = -\sum_i p(a_i) \log p(a_i).$$

Якщо взаємно пов'язані 3 елементи a_i, a_j, a_k, то умовна ентропія обчислюється за формулою:

$$H(A/B) = -\sum_i \sum_j \sum_k p(a_i, a_j, a_k) \log p(a_i, a_j, a_k).$$

Аналогічно для 4,5,…, n елементів.

Лекція 4. Ентропія об'єднання

Використовується для обчислення невизначеності сумісної появи статично залежних повідомлень.

Наприклад, передавши 100 разів цифру 5 по каналу зв'язку із завадами, помітимо, що цифра 5 була прийнята 90 разів, цифра 6 була прийнята 8 разів, цифра 4 була прийнята 2 рази.

Невизначеність виникнення комбінацій 5-4, 5-5,5-6 при передаванні цифри 5 може бути описана за допомогою ентропії об'єднання.

$H(A,B)$ - невизначеність того, що передано А, а прийнято В.

Для ансамблів переданих повідомлень А та прийнятих повідомлень В ентропія об'єднання являє собою суму вигляду

$$H(A,B) = -\sum_i \sum_j p(a_i, b_j) \log p(a_i, b_j) \text{ біт/ 2 символи}$$

Ентропія об'єднання та умовна ентропія пов'язані між собою наступними співвідношеннями:

$$H(A,B) = H(A) + H(B/A) = H(B) + H(A/B)$$

$$H(B/A) = H(A,B) - H(A)$$

$$H(A/B) = H(A,B) - H(B)$$

Для незалежних подій $p(b_j, a_i) = p(b_j)$, тому $H(B/A) = H(B)$, $H(A,B) = H(A) + H(B)$, якщо А та В повністю залежні , тобто при появі a_i обов'язково з'явиться b_j, то $p(b_j/a_i)$=1 при $i = j$ і нулю при $i \neq j$. Тому $H(B/A) = 0$, отже $H(A,B) = H(A)$, тобто при повній залежності двох ансамблів жоден не вносить ніякої додаткової інформації

Ентропія об'єднання може бути розрахована за допомогою матриці вигляду

$$p(a_i, b_j) = \begin{vmatrix} p(a_1,b_1) & p(a_1,b_2) & \dots & p(a_1,b_m) \\ p(a_2,b_1) & p(a_2,b_2) & \dots & p(a_2,b_m) \\ \dots & \dots & \dots & \dots \\ p(a_m,b_1) & p(a_m,b_2) & \dots & p(a_m,b_m) \end{vmatrix}$$

Для цієї матриці $\sum_i p(a_i,b_j) = p(b_j)$, $\sum_j p(a_i,b_j) = p(a_i)$.

При цьому $\sum_i p(a_i) = \sum_j p(b_j) = 1$.

Це дозволяє обчислити ентропію як джерела , так і приймача безпосередньо за матрицею:

$$H(A) = -\sum_i \sum_j p(a_i,b_j) \log \sum_j p(a_i,b_j)$$

$$H(B) = -\sum_j \sum_i p(a_i,b_j) \log \sum_i p(a_i,b_j)$$

Додавання здійснюється по і та по j, оскільки для розрахунку безумовних імовірностей треба додавати по одній координаті, а для знаходження Н - по іншій.

Умовні імовірності обчислюються так:

$$p(a_i/b_j) = \frac{p(a_i,b_j)}{p(b_j)} \;;\; p(b_j/a_i) = \frac{p(a_i,b_j)}{p(a_i)}$$

Кількість інформації на символ повідомлення, що передається по каналу зв'язку, в якому вплив завад описується за допомогою ентропії об'єднання, розраховується наступним чином:

$$I(A,B) = H(A) + H(B) - H(B,A)$$

Приклад 1. Канал зв'язку описується наступною канальною матрицею

$$p(b/a) = \begin{vmatrix} 0,98 & 0,01 & 0,01 \\ 0,1 & 0,75 & 0,15 \\ 0,2 & 0,3 & 0,5 \end{vmatrix}$$

Обчислити середню кількість інформації, яка переноситься одним символом повідомлення, якщо імовірності появи символів джерела повідомлень дорівнюють $p(a_1) = 0,7$, $p(a_2) = 0,2$, $p(a_3) = 0,1$. Чому дорівнюють інформаційні втрати при передаванні повідомлень з 400 символів алфавіту a_1, a_2, a_3? Чому дорівнює кількість прийнятої інформації?

Розв'язок. Ентропія джерела повідомлень:

$$H(A) = -\sum_{i=1}^{m} p_i \cdot \log_2 p_i = -(0,7 \cdot \log_2 0,7 + 0,2 \cdot \log_2 0,2 + 0,1 \cdot \log_2 0,1) = 0,36 + 0,46 + 0,33 =$$
$$= 1,16 \quad \textit{біт / символ}$$

Загальна умовна ентропія:

$$H(B/A) = -\sum_i p(a_i) \sum_j p(b_j/a_i) \log_2 p(b_j/a_i) = -[0,7 \cdot (0,98 \cdot \log_2 0,98 + 2 \cdot 0,01 \cdot \log_2 0,01) +$$
$$+0,2 \cdot (0,1 \cdot \log_2 0,1 + 0,75 \cdot \log_2 0,75 + 0,15 \cdot \log_2 0,15) + 0,1 \cdot (0,2 \cdot \log_2 0,2 + 0,3 \cdot \log_2 0,3 +$$
$$+0,5 \cdot \log_2 0,5)] = 0,7 \cdot (0,03 + 2 \cdot 0,07) + 0,2 \cdot (0,31 + 0,32 + 0,41) + 0,1 \cdot (0,46 + 0,52 + 0,5) =$$
$$= 0,473 \quad \textit{біт / символ.}$$

Втрати у каналі зв'язку:

$$\Delta I = kH(B/A) = 400 \cdot 0,473 = 189,5 \textit{ біт}$$

Ентропія приймача:

$$H(B) = -\sum_{j=1}^{m} p(b_j) \cdot \log_2 p(b_j)$$

$$p(b_1) = \sum_i p(a_i) \cdot p(b_1/a_i) = p(a_1) \cdot p(b_1/a_1) + p(a_2) \cdot p(b_1/a_2) + p(a_3) \cdot p(b_1/a_3) =$$
$$= 0,7 \cdot 0,98 + 0,2 \cdot 0,1 + 0,1 \cdot 0,2 = 0,726$$

$$p(b_2) = \sum_i p(a_i) \cdot p(b_2/a_i) = p(a_1) \cdot p(b_2/a_1) + p(a_2) \cdot p(b_2/a_2) + p(a_3) \cdot p(b_2/a_3) =$$
$$= 0,7 \cdot 0,01 + 0,2 \cdot 0,1 + 0,1 \cdot 0,2 = 0,187$$

$$p(b_3) = \sum_i p(a_i) \cdot p(b_3/a_i) = p(a_1) \cdot p(b_3/a_1) + p(a_2) \cdot p(b_3/a_2) + p(a_3) \cdot p(b_3/a_3) =$$
$$= 0,7 \cdot 0,01 + 0,2 \cdot 0,15 + 0,1 \cdot 0,5 = 0,087$$

Сума безумовних імовірностей символів вторинного алфавіту (підрахуємо для перевірки)

$$p(b_1) + p(b_2) + p(b_3) = 0,726 + 0,187 + 0,087 = 1$$

Тобто $\sum_{j=1}^{3} p(b_j) = 1$

Тоді ентропія (безумовна) приймача:

$$H(B) = -(0,726 \cdot \log_2 0,726 + 0,187 \cdot \log_2 0,187 + 0,726 \cdot \log_2 0,726 + 0,087 \cdot \log_2 0,087) =$$
$$= 1,094 \quad \textit{біт / символ.}$$

Середня кількість отриманої інформації

$$I = k[H(B) - H(B/A)] = k \cdot H(B) - \Delta I = 400 \cdot 1,094 - 189,5 = 248,1 \quad бит$$

Приклад 2. Визначити середню кількість інформації, що міститься у прийнятому ансамблі повідомлень відносно переданого, якщо повідомлення складаються з алфавіту A,B,C. Імовірність появи літер алфавіту на виході джерела повідомлень $p(A) = p(B) = 0,25$; $p(C) = 0,5$.

Умовні імовірності виникнення пар вигляду b_j/a_i наступні:

$$p(A/A) = 0,97 \qquad p(A/B) = 0,02 \qquad p(A/C) = 0,01$$
$$p(B/A) = 0,015 \qquad p(B/B) = 0,97 \qquad p(B/C) = 0,01$$
$$p(C/A) = 0,015 \qquad p(C/B) = 0,01 \qquad p(C/C) = 0,98$$

Перевірити правильність результату всіма відомими способами.

Розв'язок. Імовірність сумісних подій - об'єднання

$$p(A,A) = p(A) \cdot p(A/A) = 0,25 \cdot 0,97 = 0,24$$
$$p(B,A) = p(A) \cdot p(B/A) = 0,25 \cdot 0,015 = 0,004$$
$$p(C,A) = p(A) \cdot p(C/A) = 0,25 \cdot 0,015 = 0,004$$

$$p(A,B) = p(B) \cdot p(A/B) = 0,25 \cdot 0,02 = 0,005$$
$$p(B,B) = p(B) \cdot p(B/B) = 0,25 \cdot 0,97 = 0,24$$
$$p(C,B) = p(B) \cdot p(C/B) = 0,25 \cdot 0,01 = 0,0025$$

$$p(A,C) = p(C) \cdot p(A/C) = 0,5 \cdot 0,01 = 0,005$$
$$p(B,C) = p(C) \cdot p(B/C) = 0,5 \cdot 0,01 = 0,005$$
$$p(C,C) = p(C) \cdot p(C/C) = 0,5 \cdot 0,98 = 0,49$$

$$\sum_A \sum_B \sum_C = 1$$

Перевірка. За знайденими імовірностями будуємо матрицю сумісних подій

$$\begin{vmatrix} 0,24 & 0,005 & 0,005 \\ 0,004 & 0,24 & 0,005 \\ 0,004 & 0,0025 & 0,49 \end{vmatrix} \qquad \begin{matrix} p(a_1) = 0,25 \\ p(a_2) = 0,249 \\ p(a_3) = 0,4965 \end{matrix}$$

$$p(b_1) = 0,248 \quad p(b_2) = 0,2475 \quad p(b_3) = 0,5$$

$$\sum_j p(b_j) = 1 \qquad \sum_i p(a_i) = 1$$

Імовірність появи літер A,B,C з боку приймача

$$p(A) = p(A)\cdot p(A/A) + p(B)\cdot p(A/B) + p(C)\cdot p(A/C) = p(a_1) = 0,25$$
$$p(B) = p(A)\cdot p(B/A) + p(B)\cdot p(B/B) + p(C)\cdot p(C/B) = p(a_2) = 0,249$$
$$p(C) = p(A)\cdot p(C/A) + p(B)\cdot p(C/B) + p(C)\cdot p(C/C) = p(a_3) = 0,4965$$

Середня кількість інформації в прийнятому ансамблі:

$$I(B,A) = \sum_i p(a_i)\sum_j [\, p(b_j/a_i)\log_2 \frac{p(b_j/a_i)}{p(b_j)}] =$$
$$= p(a_1)[\, p(b_1/a_1)\log_2 \frac{p(b_1/a_1)}{p(b_1)} + p(b_2/a_1)\log_2 \frac{p(b_2/a_1)}{p(b_2)} + p(b_3/a_1)\log_2 \frac{p(b_3/a_1)}{p(b_3)}] +$$
$$+ p(a_2)[\, p(b_1/a_2)\log_2 \frac{p(b_1/a_2)}{p(b_1)} + p(b_2/a_2)\log_2 \frac{p(b_2/a_2)}{p(b_2)} + p(b_3/a_2)\log_2 \frac{p(b_3/a_2)}{p(b_3)}] +$$
$$+ p(a_3)[\, p(b_1/a_3)\log_2 \frac{p(b_1/a_3)}{p(b_1)} + p(b_2/a_3)\log_2 \frac{p(b_2/a_3)}{p(b_2)} + p(b_3/a_3)\log_2 \frac{p(b_3/a_3)}{p(b_3)}] =$$
$$= 0,25\cdot[0,97\cdot\log_2 \frac{0,97}{0,248} + 0,015\cdot\log_2 \frac{0,015}{0,2475} + 0,015\cdot\log_2 \frac{0,015}{0,5}] + 0,25\cdot[0,02\cdot\log_2 \frac{0,02}{0,248} +$$
$$+ 0,97\cdot\log_2 \frac{0,02}{0,2475} + 0,01\cdot\log_2 \frac{0,01}{0,5}] + 0,5\cdot[0,01\cdot\log_2 \frac{0,01}{0,248} + 0,01\cdot\log_2 \frac{0,01}{0,2475} +$$
$$+ 0,98\cdot\log_2 \frac{0,98}{0,5}] \approx 1,3 \; біт$$

Лекція 5. Розрахунок втрат інформації у каналі зв'язку під впливом завад. Використання умовної ентропії та ентропії об'єднання для розрахунку втрат та кількості переданої інформації

Кількість інформації при передаванні повідомлень визначається як добуток кількості повідомлень k на ентропію одного повідомлення:

$$I = kH \text{ біт.}$$

Для рівноімовірних повідомлень ентропія визначається наступним чином:

$$H_1 = \log_2 m_1 = n \log_2 m_2 \text{ біт/символ,}$$

де m_1 та m_2 відповідно кількість якісних ознак первинного та вторинного алфавітів.

Для нерівноімовірних незалежних повідомлень ентропія дорівнює:

$$H_2 = -\sum_i p_i \log_2 p_i \text{ біт/символ.}$$

Для нерівноімовірних взаємопов'язаних повідомлень ентропія може бути розрахована за допомогою наступних виразів:

$$H_3(B/A) = -\sum_i \sum_j p(a_i)\, p(b_i / a_i) \log_2 p(b_i / a_i) \text{ біт/символ;}$$

$$H_4(A/B) = -\sum_i \sum_j p(b_i)\, p(a_i / b_i) \log_2 p(a_i / b_i) \text{ біт/символ;}$$

$$H_5(A,B) = H_5(B,A) = -\sum_i \sum_j p(a_i, b_i) \log_2 p(a_i, b_i) \text{ біт/2 символи.}$$

Обчислення інформаційних втрат при передачі повідомлень по каналам зв'язку з шумами. Втрати інформації в каналах зв'язку з шумами звичайно описується за допомогою умовної ентропії та ентропії об'єднання.

Якщо завад немає або їх рівень настільки низький, що вони не в змозі знищити корисний сигнал або імітувати корисний сигнал за відсутності передавання, то при передаванні a_i ми будемо точно знати, що отримаємо b_j - сигнал, що відповідає переданому сигналу a_i. Події А та В статистично жорстко

пов'язані між собою, умовна імовірність максимальна $p(b_j/a_i)=1$, а умовна ентропія

$$H(b_j/a_i) = -\sum_{i=1}^{m} p(b_j/a_i)\log p(b_j/a_i) = 0.$$

Оскільки $\log p(b_j/a_i) = \log p(1) = 0$. В цьому випадку кількість інформації, що міститься у прийнятому ансамблі повідомлень В, пропорційна ентропії повідомлень ансамблю А, що передаються, тобто $I(B,A) = kH(A)$.

Якщо рівень завад високий, будь-який з прийнятих сигналів b_j може відповідати будь-якому переданому сигналу a_i, статистичний зв'язок між переданими та прийнятими сигналами відсутній. В цьому випадку імовірності $p(a_i)$ та $p(b_j)$ є ймовірностями незалежних подій, і $p(b_j/a_i) = p(b_j)$; $p(a_i/b_j) = p(a_i)$ При цьому

$$H(A/B) = -\sum_i \sum_j p(b_j)\,p(a_i/b_j)\log p(a_i/b_j) =$$
$$= -\sum_i \sum_j p(b_j)\,p(a_i)\log p(a_i) = \sum_j p(b_j)H(A) = H(A),$$

Оскільки $\sum_j p(b_j) = 1$, тобто умовна ентропія дорівнює безумовній, а кількість інформації, що міститься в В відносно А, дорівнює нулю:

$$I(A,B) = k[H(A) - H(A/B)] = 0$$

Інформаційні характеристики реальних каналів зв'язку лежать між цими двома граничними випадками. При цьому втрати інформації при передеванні k символів по даному каналу зв'язку

$$\Delta I = kH(A/B)$$

Незважаючи на те, що частина інформації пошкоджується завадами, між прийнятими та переданими повідомленнями існує статистичний взаємозв'язок. Це дозволяє описувати інформаційні характеристики реальних каналів зв'язку за допомогою ентропії об'єднання статистично залежних подій. Оскільки

$$H(A,B) = H(A) + H(B/A) = H(B) + H(A/B),$$

то втрати у каналі зв'язку можуть бути враховані за допомогою ентропії об'єднання наступним чином:

$$I(B,A) = k[H(A) + H(B) - H(B,A)],$$

а з використанням умовної ентропії кількість отриманої інформації

$$I(B,A) = k[H(A) - H(A/B)] = k[H(B) - H(B/A)].$$

Для обчислення середньої кількості інформації, що міститься у прийнятому ансамблі повідомлень B відносно переданого ансамблю повідомлень A за умови дії завад, користуються наступними виразами:

$$I(B,A) = -\sum_i \sum_j p(a_i)\, p(b_j/a_i) \log \frac{p(b_j/a_i)}{p(b_j)},$$

$$I(A,B) = -\sum_i \sum_j p(b_j)\, p(a_i/b_j) \log \frac{p(a_i/b_j)}{p(a_i)},$$

$$I(B,A) = I(A,B) = -\sum_i \sum_j p(a_i,b_j) \log \frac{p(b_j/a_i)}{p(b_j)} = -\sum_i \sum_j p(b_j,a_i) \log \frac{p(a_i/b_j)}{p(a_i)} =$$

$$= -\sum_i \sum_j p(a_i,b_j) \log \frac{p(a_i,b_j)}{p(a_i)\, p(b_j)} =$$

$$= -\sum_i \sum_j p(a_i,b_j) \log p(a_i,b_j) - \sum_i \sum_j p(a_i,b_j) \log\big(p(a_i) \cdot p(b_j)\big)$$

Для обчислення часто зручно застосовувати ці вирази у наступному вигляді:

$$I(B,A) = \sum_i p(a_i) \sum_j [p(b_j/a_i) \log p(b_j/a_i) - p(b_j/a_i) \log p(b_j)]$$

$$I(A,B) = \sum_j p(b_j) \sum_i [p(a_i/b_j) \log p(a_i/b_j) - p(a_i/b_j) \log p(a_i)],$$

$$I(A,B) = I(B,A) = \sum_i \sum_j p(a_i,b_j) \log p(a_i,b_j) - \sum_i \sum_j p(a_i,b_j) \log(p(a_i)\, p(b_j))$$

Для повного і всебічного описання каналу зв'язку необхідно задати:

- канальну матрицю вигляду $p(a_i/b_j)$

- безумовні імовірності вигляду $p(b_j)$

або:

- канальну матрицю вигляду $p(b_j/a_i)$
- безумовні імовірності вигляду $p(a_i)$

або:

- канальну матрицю вигляду $p(a_i, b_j)$

В останньому випадку сума значень матриці по колонкам дає безумовні імовірності вигляду $p(b_j)[\sum_j p(b_j) = 1]$, а сума по рядках дає безумовні імовірності вигляду $p(a_i)[\sum_i p(a_i) = 1]$. Умовні імовірності можуть бути знайдені з виразів

$$p(a_i/b_j) = \frac{p(b_j, a_i)}{p(b_j)};\ p(b_j/a_i) = \frac{p(a_i, b_j)}{p(a_i)}$$

Знаючи умовні та безумовні імовірності, можна знайти $H(A), H(B), H(A/B), H(B/A)$.

Якщо рівень завад настільки високий, що з рівною імовірністю можна очікувати перехід будь-якого символу джерела у довільний символ, то ентропія каналу зв'язку буде дорівнювати $\log_2(m)$, а кількість інформації

$$I = H(A) - \log_2 m \le 0$$

Значення I може бути від'ємним- це означає, що канал вносить дезінформацію.

Приклад 1. Алфавіт складається з 2 елементів: 0 та 1. Якщо ці елементи рівноімовірні, то кількість інформації на 1 елемент повідомлення

$$H_1 = \log_2 m = \log_2 2 = 1\ біт/символ.$$

Якщо не рівноімовірні, наприклад: $p(0) = \frac{3}{4}$, $p(1) = \frac{1}{4}$, то

$$H_2 = -\sum_{i=1}^{m} p_i \log_2 p_i = -p(0)\log_2 p(0) - p(1)\log p(1) = -(\frac{3}{4}\cdot\log_2\frac{3}{4} + \frac{1}{4}\cdot\log_2\frac{1}{4}) = 0,815 \; \textit{біт / симв.}$$

Якщо ж є взаємна залежність елементів, яка визначається умовними імовірностями $\left.\begin{matrix} p(0/0)=\frac{2}{3} \\ p(1/0)=\frac{1}{3} \end{matrix}\right\} = 1 \quad \left.\begin{matrix} p(0/1)=0 \\ p(1/1)=1 \end{matrix}\right\} = 1,$

то умовна ентропія:

$$H_3 = -p(0)[\,p(0/0)\log_2 p(0/0) + p(1/0)\log_2 p(1/0)] - p(1)[\,p(0/1)\log_2 p(0/1) + p(1/1)\log_2 p(1/1)] =$$

$$= -\frac{3}{4}\cdot[\frac{2}{3}\cdot\log_2\frac{2}{3} + \frac{1}{3}\cdot\log_2\frac{1}{3}] - \underbrace{\frac{1}{4}\cdot[1\cdot\log_2 1 + 0\cdot\log_2 0]}_{=0} = 0,685 \quad \textit{біт / символ}$$

Ентропія при взаємно залежних елементах завжди менше, ніж при незалежних.

Приклад 2. Розглянемо канальну матрицю імовірностей об'єднання $p(a_i, b_j)$:

A\B	4	5
3	0,17	0,1
10	0,13	0,3
12	0,25	0,05

Знайти безумовні імовірності об'єднання величин А та В.

Розв'язок. Додамо імовірності по рядках:

$$p(3) = 0,17 + 0,1 = 0,27$$
$$p(10) = 0,13 + 0,3 = 0,43$$
$$p(12) = 0,25 + 0,05 = 0,3$$

Тоді для величини А:

a_i	3	10	12
$p(a_i)$	0,27	0,43	0,3

Контроль: 0,27+0,43+0,3=1

Додамо імовірності по колонках:

$p(4) = 0,17 + 0,13 + 0,25 = 0,55$

$p(5) = 0,1 + 0,3 + 0,05 = 0,45$

Розподіл імовірностей для величини B:

b_i	4	5
$p(b_i)$	0,55	0,45

Контроль: 0,55+0,45=1

Приклад 3. Задана канальна матриця $p(a_i, b_j)$

A\B	$b_1 = 0,4$	$b_2 = 0,8$	
$a_1 = 2$	0,15	0,05	$p(a_1)$
$a_2 = 5$	0,3	0,12	$p(a_2)$
$a_3 = 8$	0,35	0,03	$p(a_3)$
	$p(b_1)$	$p(b_2)$	

Знайти:

1) безумовні імовірності $p(a_i)$, $p(b_j)$
2) умовні імовірності $p(a_i / b_1)$
3) умовні імовірності $p(b_j / a_2)$

Розв'язок. Додамо імовірності по рядкам і знайдемо $p(a_i)$

a_i	2	5	8
$p(a_i)$	0,2	0,42	0,38

Контроль: $0,2 + 0,42 + 0,38 = 1$

Додамо імовірності по колонкам и знайдемо $p(b_j)$

b_j	0,4	0,8
$p(b_j)$	0,8	0,2

Контроль: $0,2+0,8=1$

Знайдемо умовні імовірності $p(a_i/b_1)$:

$$a_1=2: p(a_1/b_1)=\frac{p(a_1,b_1)}{p(b_1)}=\frac{0,15}{0,8}=\frac{3}{16}$$

$$a_2=5: p(a_2/b_1)=\frac{p(a_2,b_1)}{p(b_1)}=\frac{0,3}{0,8}=\frac{3}{8}$$

$$a_3=8: p(a_3/b_1)=\frac{p(a_3,b_1)}{p(b_1)}=\frac{0,35}{0,8}=\frac{7}{16}$$

Отже, розподіл умовних імовірностей:

a_i	2	5	8
$p(a_i/b_1)$	3/16	3/8	7/16

Контроль: $3/16+3/8+7/16=1$

Аналогічно знайдемо умовні імовірності $p(b_j/a_2)$

$$b_1=0,4: p(b_1/a_2)=\frac{p(a_2,b_1)}{p(a_2)}=\frac{0,3}{0,42}=0,71$$

$$b_2=0,8: p(b_2/a_2)=\frac{p(a_2,b_2)}{p(a_2)}=\frac{0,12}{0,42}=0,29$$

Розподіл умовних імовірностей:

b_j	0,4	0,8
$p(b_j/a_2)$	0,71	0,29

Контроль: $0,71+0,29=1$

Приклад 4. Задана матриця імовірностей об'єднання. Визначити ентропії $H(A), H(B), H(A/B), H(B/A)$.

A\B	b_1	b_2	b_3
a_1	0,4	0,1	0
a_2	0	0,2	0,1
a_3	0	0	0,2

Розв'язок. Обчислимо безумовні імовірності $p(a_i)$ та $p(b_j)$:

1) додамо значення по рядках:

$p(a_1) = 0,4 + 0,1 = 0,5$

$p(a_2) = 0,3$

$p(a_3) = 0,2$

2) додамо значення по колонках:

$p(b_1) = 0,4$

$p(b_2) = 0,3$

$p(b_3) = 0,3$

3) ентропія джерела:

$$H(A) = -[p(a_1)\log_2 p(a_1) + p(a_2)\log_2 p(a_2) + p(a_3)\log_2 p(a_3)] = -[0,5\cdot\log_2 0,5 + 0,3\cdot\log_2 0,3 + +0,2\cdot\log_2 0,2] = 1,485 \quad \textit{біт / символ}$$

4) ентропія приймача:

$$H(B) = -[p(b_1)\log_2 p(b_1) + p(b_2)\log_2 p(b_2) + p(b_3)\log_2 p(b_3)] = -[0,4\cdot\log_2 0,4 + 2\cdot 0,3\cdot\log_2 0,3] = = 1,57 \quad \textit{біт / символ}$$

5) умовна ентропія

$$H(B/A) = -\sum_i p(a_i)\sum_j p(b_j/a_i)\log_2 p(b_j/a_i)$$

$$p(b_j/a_i) = \frac{p(a_i, b_j)}{p(a_i)}$$

Знайдемо умовні імовірності:

$$p(b_1/a_1)=\frac{p(a_1,b_1)}{p(a_1)}=\frac{0,4}{0,5}=0,8 \qquad p(b_1/a_2)=\frac{p(a_2,b_1)}{p(a_2)}=0$$

для a_1: $p(b_2/a_1)=\frac{p(a_1,b_2)}{p(a_1)}=\frac{0,1}{0,5}=0,2$ для a_2: $p(b_2/a_2)=\frac{p(a_2,b_2)}{p(a_2)}=\frac{0,2}{0,3}=0,67$

$$p(b_3/a_1)=\frac{p(a_1,b_3)}{p(a_1)}=0 \qquad p(b_3/a_2)=\frac{p(a_2,b_3)}{p(a_2)}=\frac{0,1}{0,3}=0,33$$

$$p(b_1/a_3)=\frac{p(a_3,b_1)}{p(a_3)}=0$$

для a_3: $p(b_2/a_3)=\frac{p(a_2,b_2)}{p(a_3)}=0$

$$p(b_3/a_3)=\frac{p(a_2,b_3)}{p(a_3)}=\frac{0,2}{0,2}=1$$

Тому

$$H(B/A)=-[0,5\cdot(0,8\cdot\log_2 0,8+0,2\cdot\log_2 0,2)+0,3(0,67\cdot\log_2 0,67+0,33\cdot\log_2 0,33)+0,2\cdot(1\cdot\log_2 1)]=$$
$$=0,635 \quad \textit{біт / символ}$$

$$H(A/B)=-\sum_j p(b_j)\sum_i p(a_i/b_j)\log_2 p(a_i/b_j)$$

Аналогічно для $p(a_i/b_j)=\frac{p(a_i,b_j)}{p(b_j)}$

$$p(a_1/b_1)=\frac{p(a_1,b_1)}{p(b_1)}=$$

для b_1: $p(a_2/b_1)=\frac{p(a_1,b_2)}{p(b_1)}=$ для b_2: $p(a_1/b_2)=$ $p(a_2/b_2)=$ $p(a_3/b_2)=$

$$p(a_3/b_1)=\frac{p(a_1,b_3)}{p(b_1)}=$$

$p(a_1/b_3)=$

для b_3: $p(a_2/b_3)=$

$p(a_3/b_3)=$

Приклад 5. У результаті статичних випробувань встановлено, що при переданні кожних 100 повідомлень довжиною по 5 символів у повідомлені символ k зустрічається 50 разів; символ T зустрічається 30 разів. Разом з символом k символ E зустрічається 10 разів.

Визначити умовні ентропії $H(k/T)$ та $H(T/k)$.

Розв'язок. Загальна кількість переданих символів

n=100*5=500

Імовірність появи символа k:

$$p(k) = \frac{50}{500} = 0,1$$

Імовірність появи символа T:

$$p(T) = \frac{30}{500} = 0,06$$

Імовірність сумісної появи символів k та T:

$$p(kT) = \frac{10}{500} = 0,02$$

Оскільки $p(kT) = p(T) \cdot p(k/T) = p(k) \cdot p(T/k)$, то умовна імовірність появи символа k відносно символа T:

$$p(k/T) = \frac{p(kT)}{p(T)} = \frac{0,02}{0,06} = 0,33$$

Умовна імовірність появи символа T відносно символа k:

$$p(T/k) = \frac{p(kT)}{p(k)} = \frac{0,02}{0,1} = 0,2$$

Умовна ентропія символа k відносно символа T:

$$H(k/T) = -\sum_i p(b_j/a_i) \log_2 p(b_j/a_i) = -\{p(k/T) \log_2 p(k/T) + (1 - p(k/T)) \log_2 [1 - p(k/T]\} =$$

$$= -(0,33 \cdot \log_2 0,33 + 0,67 \cdot \log_2 0,67) = 0,915 \quad \textit{біт / символ}$$

Умовна ентропія символа T відносно символа k:

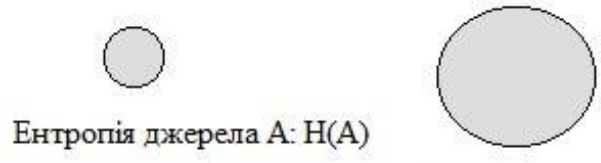
$$H(T/k) = -(0,2 \cdot \log_2 0,2 + 0,8 \cdot \log_2 0,8) = 0,722 \quad \textit{біт / символ}$$

У неідеальному каналі зв'язку частина інформації пошкоджується завадами. Але незважаючи на це, між прийнятим та переданим повідомленнями існує статистичний взаємозв'язок. Це дозволяє описувати інформаційні характеристики реальних каналів зв'язку за допомогою ентропії об'єднання статистично залежних подій:

$$H(A,B)=H(A)+H(B/A)=H(B)+H(A/B)$$

Приклад 6. Відомі ентропії статистично пов'язаного джерела та приймача H(A) = 5 біт/символ; H(B) = 10 біт/символ. Визначити, в яких межах буде змінюватися умовна ентропія H(B/A).

Розв'язок. З'ясувати відношення між ентропіями допомагає їх графічне відображення. За відсутності взаємозв'язків між джерелом і приймачем:

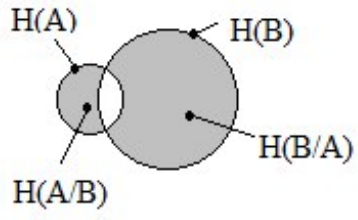


Якщо джерело та приймач незалежні, то H(B/A)=H(B)=10 біт/символ, а H(A/B)=H(A)=5 біт/символ, тому

H(A,B)=H(A)+H(B/A)=H(A)+H(B)=5+10=15 біт/символ.

Таким чином, за відсутності статистичного зв'язку H(B/A)=H(B)=10 біт/символ і приймає максимальне значення.

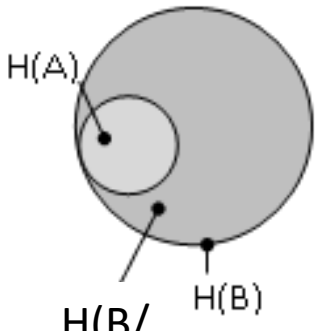
По мірі збільшення взаємозв'язку значення H (B/A) та H (A/B) будуть зменшуватися:



При повній залежності джерела та приймача один з них не вносить ніякої інформації, тобто при появі a_i однозначно виникає b_j , т.я. $p(a_i,b_j)$ рівно одиниці при i=j та нулю при i≠j. Тому

H(A/B)=0,

H(A,B)=H(B)+H(A/B)=H(B).



При цьому H(B/A)=H(B)-H(A)=10-5=5 біт/символ. Тому H(B/A) буде змінюватися від 10 біт/символ до 5 біт/символ при максимально можливій зміні H(A/B) від 5 біт/символ до 0 біт/символ.

Лекція 6. Коди. Представлення кодів. Поняття про кодування

Передавання інформації від джерела до приймача відбувається з використанням сигналів. Для, того, щоб сигнали можна було одночасно зрозуміти, їх необхідно складати за правилом, яке строго фіксовано протягом всього часу передавання даної групи повідомлень.

Правило (алгоритм), яке співставляє кожному конкретному повідомленню строго визначену комбінацію різних символів (або відповідних їм сигналів), називається кодом, а процес перетворення повідомлення у комбінацію різних символів або відповідних їм сигналів — кодування.

Процес відновлення змісту повідомлення за даним кодом називається декодуванням.

Послідовність символів, яка в процесі кодування присвоюється кожній з множин символів, які передаються, називається кодовим словом.

Символи, за допомогою яких записується передане повідомлення, складають первинні алфавіт, а символи, за допомогою яких повідомлення трансформується в код — вторинний алфавіт.

Коди, в яких повідомлення представлені комбінаціями з нерівною кількістю символів, називаються нерівномірними або некомплектними. Коди, в яких повідомлення представлені комбінаціями з рівною кількістю символів, називаються рівномірними або комплектними.

Прикладом нерівномірного коду може слугувати простий двійковий код, який, як відомо, являє собою степеневий ряд двійки

$$2^0 + 2^1 + 2^2 + \ldots + 2^n$$

Однак якщо комбінації двійкового коду доповнити такою кількістю нулів, щоб число символів в кожному кодовому слові дорівнювало числу символів послідовного кодового слова, то такий двійковий код буде рівномірним:

Нерівномірний двійковий код	Рівномірний двійковий код
1	0001
10	0010
11	0011
100	0100
101	0101
110	0110
111	0111
1000	1000

Для однозначного декодування кодових комбінацій на боці приймача імпульс в каналі зв'язку мають бути розділені так, щоб кожний символ повідомлення міг бути прийнятий самостійно.

Розділення імпульсів може бути просторове, часове або якісне.

Просторове розділення передбачає багатоканальний зв'язок і не потребує спеціальних методів.

При **якісному розділенні** передбачається наявність мінімум двох якісних ознак. При цьому якісні ознаки, присвоєні певними символам, мають легко розрізнятися на приймальному блоці. Якісне розділення дає можливість одночасного передавання інформації від різних об'єктів по одному каналу зв'язку.

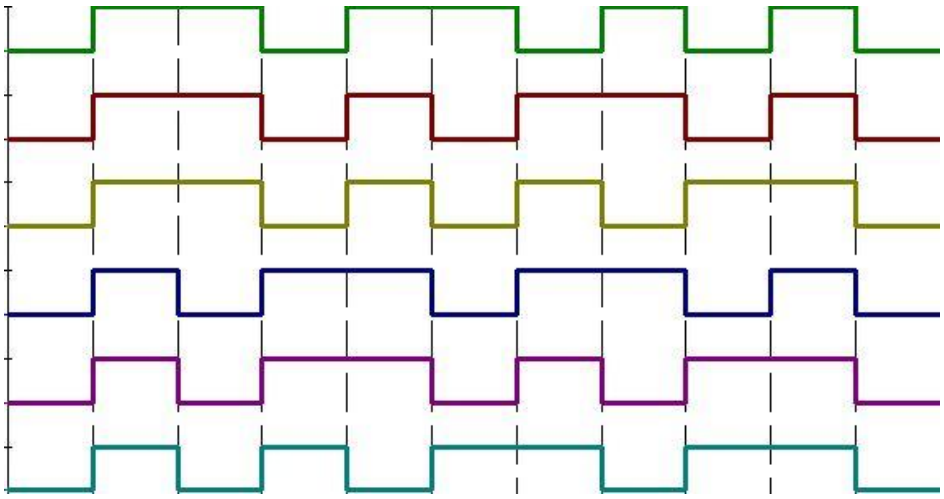
Найбільш поширеним видом якісного розділення символів повідомлення при побудові кодів з кількістю якісних ознак $m>2$ є **частотне розділення**. Тому в подальшому при вивченні кодів, які містять три або більше якісних ознак, більшу увагу приділятимемо частотним кодам.

При часовому розділенні повідомлення можуть бути передані за допомогою однієї якісної ознаки. Оскільки тривалості імпульсу та паузи теж є якісними ознаками, то під часовим розділенням розуміють звичайно розділення у часі повідомлень від різних об'єктів, що передаються однією лінією зв'язку.

Паралельне передавання повідомлень за відсутності якісного розділення повністю виключається.

Часове розділення звичайно реалізується за допомогою синхронізованих комутуючих пристроїв, які знаходяться на передавальному та приймальному боках та почергово з'єднують джерела з відповідними приймачем.

При побудові кодів часто використовують комбінацію часових та частотних якісних ознак. Частотно-часовий код утворюється шляхом сумісного застосування частотної та часової якісних ознак. В таблиці представлені комбінації частотно-часового коду на одне сполучення з посиланням повної серії імпульсів при кількості якісних ознак m=2, кількості імпульсів у коді (часових позицій) $n_B = 4$.

Символьний запис коду	**Частотно-часовий код**
a a b b	
a b a b	
a b b a	
b a a b	
b a b a	
b b a a	

Способи представлення кодів базуються як на застосуванні теорії сполучень, так і на алгебраїчних перетвореннях та геометричних побудовах. Коди можуть бути представлені формулою, геометричною фігурою, таблицею, графом, поліномом, матрицею і т.д.

Представлення коду у вигляді поліному. Для будь-якої системи числення з основою x за наявності _ Різних цифрових знаків _ представлення коду у вигляді поліному має вигляд

$$F(x) = a_0 + a_1 \cdot x + \ldots a_{n-2} \cdot x^{n-2} + a_{n-1} \cdot x^{n-1} = \sum_{i=0}^{n-1} a_i \cdot x^i,$$

де *i* – показник степеню основи системи числення і порядковий номер чергового розряду.

Наприклад, в десятковій системі числення число 435 можна записати у вигляді

$$435 = F(10) = 5 \cdot 10^0 + 3 \cdot 10^1 + 4 \cdot 10^2$$

В даному випадку $x = 10, a_0 = 5, a_1 = 3, a_2 = 4, n = 3$

У двійковій системі число 73 записується у вигляді полінома з основою 2:

$$73 = F(2) = 1 \cdot 2^0 + 0 \cdot 2^1 + 0 \cdot 2^2 + 1 \cdot 2^3 + 0 \cdot 2^4 + 0 \cdot 2^5 + 1 \cdot 2^6 = 1 + 8 + 64$$

В двійковому коді це число має вигляд: 1001001.

На використанні властивостей послідовностей двійкових чисел базується методика побудови багатьох кодів.

Особливий інтерес представляють властивості двійкових кодів, які представляються при додаванні, множенні та діленні за модулем 2.

Правила додавання за модулем 2 визначаються наступними рівностями

$0 \oplus 0 = 0$
$1 \oplus 1 = 0$
$0 \oplus 1 = 1$
$1 \oplus 0 = 1$

В якості прикладу додамо за модулем 2 двійкові числа 10111011 та 100010:

$$\begin{array}{r} 10111011 \\ \oplus \quad 100010 \\ \hline 10011001 \end{array}$$

Відмінність операції додавання за модулем 2 від звичайного арифметичного додавання двійкових чисел полягає в тому, що при додаванні за

модулем 2 кожного разу розглядають конкретну пару двійкових знаків без зв'язку з усім числом.

Тому результати попередніх операцій при додаванні чергової пари двійкових знаків не враховуються, тоді як при арифметичному додаванні двійкових чисел цей результат обов'язково враховується. Наприклад, при додаванні двох одиниць записують 0, а 1 переноситься у старший розряд. Так, для нашого прикладу:

$$\begin{array}{r} 10\overleftarrow{1}11011 \\ +\quad 100010 \\ \hline 11011101 \end{array}$$

Множення та ділення двійкових чисел за модулем 2 зводиться до додавання за модулем 2, але зсув чисел відбувається немовби з протилежного боку. Перший множник зсувають в бік старшого розряду стільки разів, скільки розрядів у другому множнику. При цьому 1-й множник виписують тільки в тому випадку, якщо у 2-му множнику є 1. Якщо у 2-му множнику 0, то черговий зсув відбувається без виписування множника 1:

$$\begin{array}{r} 1011 \\ \otimes 1101 \\ \hline 1011 \\ \oplus 1011 \\ \oplus\ 1011 \\ \hline 1111111 \end{array}$$

Іноді зручно множення починати зі старшого розряду, тоді рядки 1-го множника записуються один під одним зі зсувом, що відповідає наявності одиниць у 2-му множнику, зліва направо:

$$\begin{array}{r} 1011 \\ \otimes 1101 \\ \hline 1011 \\ \oplus\ 1011 \\ \oplus\quad 1011 \\ \hline 1111111 \end{array}$$

При діленні за модулем 2 дільник підписують під діленим так, щоб співпадали старші розряди. Якщо кількість розрядів діленого більше або дорівнює кількості розрядів в дільника, то у частку переносять 1. Потім виконують додавання за модулем 2, після чого до залишку приписують праворуч чергову цифру діленого. Якщо кількість розрядів залишку разом з приписаною цифрою дорівнює кількості розрядів дільника, то у частку записують ще одну одиницю. Інакше до частки записують нулі доти, доки не зрівняються розряди залишку та дільника. Потім знову виконують додавання за модулем 2 і операцію повторюють доти, доки всі розряди діленого не перенесуться до залишку:

$$\begin{array}{r|l} 10001 & 101 \\ \oplus\,101 & \hline 101 \\ \hline 101 & \\ \oplus\ 101 & \\ \hline 0 & \end{array} \qquad \begin{array}{r|l} 10111011 & 100010 \\ \oplus\,100010 & \hline 101+\dfrac{10001}{100010} \\ \hline 110011 & \\ \oplus\ 100010 & \\ \hline 10001 & \end{array} \qquad \begin{array}{r|l} 11010011 & 10011 \\ \oplus\,10011 & \hline 1100+\dfrac{111}{10011} \\ \hline 10010 & \\ \oplus\ 10011 & \\ \hline 111 & \end{array}$$

Лекція 7. Поняття надмірності

Для визначення максимальної пропускної здатності системи зв'язку необхідно вміти визначати максимальну кількість інформації, яка може бути передана за допомогою символів даного алфавіту за одиницю часу.

Ми знаємо, що максимальну кількість інформації на символ повідомлення H=log m можна отримати тільки у випадку рівноімовірних та незалежних символів. Реальні коди рідко повністю задовольняють цій умові, тому інформаційне навантаження на кожний їх елемент звичайно менше за те, яке б вони могли переносити.

Ентропія повідомлень, що представляються такими кодами, менша за максимальну.

Якщо елементи кодів, що представляють повідомлення, недовантажені, то саме повідомлення має інформаційну надмірність. Поняття надмірності в теорії інформації та кодування введене для кількісного описання інформаційного резерву коду, з якого складено повідомлення.

Розрізняють надмірність *природну* і *штучну*. **Природна надмірність** є характерною для первинних алфавітів, а **штучна** – для вторинних.

Природна надмірність може бути поділена на *семантичну* та *статистичну* надмірності.

Семантична надмірність полягає в тому, що зміст, переданий у повідомленні, може бути вираженим коротше. Якщо повідомлення можна скоротити без зміни змісту, а потім відновити зміст, то воно має семантичну надмірність. Семантичну надмірність можна усунути різними способами. Наприклад, стандартні повідомлення, що часто повторюються, можна замінити на умовні позначення; повідомлення, що містить різні характеристики одних і тих самих елементів, можна представити у вигляді таблиць; застосувати згортання інформації, абревіатури, тощо. Спільним при цьому є те, що *всі дії по усуненню семантичної надмірності виконуються у первинному алфавіті*.

Статистична надмірність обумовлюється нерівноімовірним розподілом якісних ознак первинного алфавіту та їх взаємопов'язаністю.

Наприклад, для англійського алфавіту, в якому 26 літер, максимальне значення ентропії:

$$H_{max} = \log_2 m = \log_2 26 = 4{,}7 \text{ біт/літеру.}$$

Якщо дослідити частоту появи різних літеру а англійських текстах, то можна переконатися, що імовірності появи літер англійського алфавіту далеко не рівні, а отже, ентропія англійської мови менша, ніж 4,7 біт/літеру.

Дійсно, дослідження показали, що при врахуванні частоти розподілу 8-літерних сполучень, тобто взаємозалежності між символами, ентропія англійської мови зменшується до 2,35 біт/літеру. Якщо ж врахувати статистику появи слів у англійських текстах, то ентропія англійської мови не перевищує 2 біт/літеру.

Отже, як бачимо, надмірність закладена у самій природі англійського алфавіту (так само, як і для інших мов світу).

При врахуванні частоти появи літер у текстах, появи літер у різних сполученнях, та слів у різних повідомленнях інформацію, що передається, можна значно стиснути, скоротити.

Чим більшою є ентропія, тим більшу кількість інформації містить в середньому кожний елемент повідомлення.

При передаванні однакової кількості інформації повідомлення тим довше, чим менше його ентропія та більше надмірність.

Ентропія може бути визначена як інформаційне навантаження на символ повідомлення. Надмірність визначає недовантаженість символів. Якщо $H = H_0 = H_{max}$, то недовантаженості не існує.

Позначимо

$$\mu = \frac{n_0}{n_1} = \frac{H_1}{H_0},$$

де n_0 та n_1 – відповідно кількість символів у переданому та прийнятому повідомленнях.

Величина μ, яка називається ***<u>коефіцієнтом стиснення</u>***, або ***<u>відносною ентропією</u>***, характеризує степінь скорочення повідомлення при переході до кодування станів елементів, які характеризуються більшою ентропією.

При цьому доля надлишкових елементів оцінюється ***<u>коефіцієнтом надмірності:</u>***

$$D = \frac{H_0 - H_1}{H_0} = 1 - \frac{H_1}{H_0} = 1 - \mu$$

Крім загального поняття статистичної надмірності, існують часткові поняття:

1) часткова надмірність D_p, яка залежить від розподілу та обумовлена нерівноімовірністю появи символів у повідомленні:

$$D_p = 1 - \frac{H_1}{H_0},$$

де $H_0 = H_{\max} = \log m, \quad H_1 = -\sum_i p_i \log p_i$;

2) часткова надмірність D_s, обумовлена статистичним зв'язком між символами повідомлення:

$$D_s = 1 - \frac{H_2}{H_1},$$

де $H_2 = -\sum_i \sum_j p(a_i)\, p(b_j / a_i) \log p(b_j / a_i)$;

3) повна надмірність

$$D = 1 - \frac{H_2}{H_0}$$

Надмірність D_p характеризує інформаційний резерв повідомлень з нерівноімовірними символами відносно повідомлень, символи яких рівноімовірні.

Надмірність D_S характеризує інформаційний резерв повідомлень зі взаємно залежними символами, які характеризуються статистичним зв'язком, по відношенню до незалежних повідомлень.

Ці три величини пов'язані залежністю

$$D = D_s + D_p - D_p D_s.$$

Приклад. Український алфавіт містить 33 літери. При аналізі текстів слід враховувати також пробіли між словами. Отже, при однакових імовірностях появи всіх 34 елементів алфавіту, невизначеність, що припадає на один елемент, складає

$$H_0 = H_{max} = \log 34 = 5{,}087 \text{ біт/літеру}$$

Аналіз показує, що з урахуванням нерівномірної появи різних літер алфавіту

$$H_1 = 4{,}42 \text{ біт/літеру},$$

а з урахуванням залежності дволітерних сполучень

$$H_2 = 3{,}52 \text{ біт/літеру},$$

тобто $H_2 < H_1 < H_0$

Внаслідок залежності між сполученнями, що містять дві та більше літер, а також змістовної залежності між словами, надмірність української мови (як і інших європейських мов) не перевищує 50% ($D = 1 - \frac{H_2}{H_0} = 1 - \frac{3{,}52}{5{,}087} = 0{,}31$).

Надмірність відіграє позитивну роль, оскільки завдяки їй повідомлення захищені від завад. Це використовують при завадостійкому кодуванні.

Надмірність усувають побудовою оптимальних кодів, які вкорочують повідомлення в порівнянні з рівномірними кодами. Це використовують при архівації даних. Дія засобів архівації заснована на використанні алгоритмів

стиснення, що мають досить довгу історію розвитку, яка розпочалася задовго до появи першого комп'ютера – ще в 40-х рр. XX століття. Група вчених-математиків, які працювали в області електротехніки, зацікавилася можливістю створення технології зберігання даних, що забезпечує більш економне витрачання простору. Одним з них був **Клод Елвуд Шеннон**, основоположник сучасної теорії інформації. З розробок того часу пізніше практичне застосування знайшли алгоритми стиснення **Хаффмана** і **Шеннона-Фано**. Суть роботи архіваторів: вони знаходять в файлах надлишкову інформацію (повторювані ділянки і прогалини), кодують їх, а потім при розпакуванні відновлюють вихідні файли за особливими відмітками.

При архівації треба мати на увазі, що якість стиснення файлів сильно залежить від ступеня надмірності даних, що зберігаються, яка визначається їх типом. Наприклад, ступінь надмірності у відеоданих зазвичай в кілька разів більше, ніж у графічних, а ступінь надмірності графічних даних в кілька разів більше, ніж текстових. На практиці це означає, що, скажімо, зображення форматів **BMP** і **TIFF**, будучи поміщеними в архів, як правило, стискаються в рази сильніше, ніж документи MS Word. А ось малюнки JPEG вже заздалегідь компресовані, тому навіть найкращий архіватор для них буде малоефективним. Також вкрай незначно стискаються виконувані файли програм і архіви.

Принцип роботи архіваторів заснований на пошуку в файлі «надлишкової» інформації та наступному її кодуванні з метою отримання мінімального об’єму. Найвідомішим методом архівації файлів є стиснення послідовностей однакових символів. Наприклад, всередині файлу знаходяться послідовності байтів, які часто повторюються. Замість того, щоб зберігати кожен байт, фіксується кількість повторюваних символів і їх позиція. Наприклад, файл, що архівується, займає 15 байт і складається з наступних символів:

B B B B B L L L L L A A A A A

В рамках таблиці кодування символів **ASCII** (American Standard Code for Information Interchange) в шістнадцятковій системі числення цю послідовність можна представити так:

42 42 42 42 42 4C 4C 4C 4C 4C 41 41 41 41 41

Архіватор може представити цей файл в наступному вигляді (шістнадцядковому):

01 05 42 06 05 4C 0B 05 41

Це означає: з першої позиції п’ять разів повторюється символ «B», з позиції 6 – п’ять разів повторюється символ «L» і з позиції 11 – п’ять разів повторюється символ «A». Для зберігання файлу в такій формі буде потрібно всього 9 байт, що на 6 байт менше вихідного об’єму.

Описаний метод є простим і дуже ефективним способом стиснення файлів. Однак він не забезпечує великої економії об’єму, якщо опрацьований текст містить невелику кількість послідовностей повторюваних символів.

Більш ефективний метод стиснення даних, який використовується в тому чи іншому вигляді практично будь-яким архіватором, – це так званий оптимальний префіксний код і, зокрема, кодування символами змінної довжини (алгоритм Хаффмана).

Код змінної довжини дозволяє записувати символи, що найбільш часто зустрічаються, і групи символів всього лише декількома бітами, в той час як символи, що зустрічаються рідко, і фрази будуть записані довшими бітовими рядками. Наприклад, в будь-якому англійському тексті літера E зустрічається частіше, ніж Z, а X і Q відносяться до тих, що найменш часто зустрічаються. Таким чином, використовуючи спеціальну таблицю відповідності, можна закодувати кожну букву E меншим числом бітів і використовувати довший код для більш рідкісних літер.

Надмірність може бути закладена у самому коді. Наприклад, трьома двійковими розрядами можна передати і цифру 5, і цифру 7, тобто для

передавання п'яти повідомлень треба будувати код такої самої довжини, у для передавання семи повідомлень.

Надмірність, закладену у природі коду, повністю виключити не можна.

Природна надмірність характерна для первинного алфавіту, а штучна — для вторинного, причому природна надмірність присутня у повідомленні до того, як воно трансформується у код.

Штучна надмірність необхідна для підвищення завадозахищеності кодів, її вводять у вигляді n_k додаткових символів. Якщо в коді всього розрядів, з них n_i несуть інформаційне навантаження. То

$$n = n_k + n_i$$

При цьому n_k - абсолютна корегуюча надмірність, а величина $D_k = \frac{n - n_i}{n_i}$ називається відносною корегуючою надмірністю

Основна теорема кодування для каналу зв'язку без шумів. У випадку рівномірного та взаємозалежного алфавіту його ентропія виражається наступним чином:

$$H = \log N = n \log m \text{, т.к } N = m^n$$

При цьому довжина повідомлення у вторинному алфавіті

$$n = \frac{\log N}{\log m} = \frac{H_{\max}(\textit{первинного})}{H_{\max}(\textit{вторинного})}$$

Для випадку нерівноімовірного та незалежного первинного алфавіту

$$n = -\frac{1}{\log m} \sum_{i=1}^{N} p_i \log p_i$$

Для нерівноімовірного алфавіту з урахуванням взаємозалежності між символами:

$$n = -\frac{1}{\log m} \sum_{i=1}^{N} \sum_{o=1}^{N} p_i \cdot p(i/j) \log p(i/j)$$

Для m- значного рівноімовірного алфавіту довжина кодового слова $L \geq \frac{\log N}{\log m}$.

Надмірність - не завжди небажане явище. Для підвищення завадостійкості кодів надмірність необхідна, її вводять штучно за рахунок додаткових корегуючих символів.

Інформаційна надмірність звичайно явище природнє, закладене у первинному алфавіті. Корегуюча надмірність явище штучне, воно закладене у кодах, представлених у вторинному алфавіті.

Найбільш ефективним способом зменшення надмірності є побудова оптимальних кодів.

Оптимальні коди - коди з практично нульовою надмірністю. Оптимальні коди мають мінімальну середню довжину кодових слів Верхня та нижня границі визначаються з нерівності

$$\frac{H}{\log m} \le L \le \frac{H}{\log m} + 1,$$

де H - ентропія первинного алфавіту, кількість якісних ознак вторинного алфавіту.

У випадку поблочного кодування, де кожний з блоків складається з M незалежних літер мінімальна середня довжина кодового блоку лежить у межах:

$$\frac{MH}{\log m} \le L_m \le \frac{MH}{\log m} + 1,$$

де L_m - середня кількість символів у блоці.

Кількість елементарних символів на літеру повідомлення при блочному кодуванні $L = \frac{L_m}{M}$ - середня довжина блоку L_m, поділена на кількість літер у блоці M.

Оскільки кожна літера складається з L елементарних символів, тоді, розділивши всі частини нерівності на M , отримаємо:

$$\frac{H}{\log m} \le L_m \le \frac{H}{\log m} + \frac{1}{M}$$

З цього видно, що при $M \to \infty$ середня кількість елементарних символів, що витрачаються на передавання однієї літери, необмежено наближається до величини $\frac{H}{\log m}$.

Наведені нерівності є основними виразами фундаментальної теореми кодування за відсутності шумів.

Сама теорема формулюється наступним чином:

– при кодуванні множники сигналів з ентропією Н у алфавіті, який налічує символів, за умови відсутності шумів середня довжина кодового слова не може бути меншою, ніж $\frac{H}{\log m}$. При кодуванні достатньо довгими блоками до цієї межі можна скільки завгодно наближатися.

Для двійкових кодів основну теорему кодування можна сформулювати так:

– при кодуванні повідомлень у двійковому алфавіті зі зростанням кількості кодових слів середня кількість символів на літеру повідомлення наближається до ентропії джерела повідомлень.

Лекція 8. Оптимальне кодування

Оптимальним кодуванням називається процедура перетворення символів первинного алфавіту $\mathbf{m_1}$ у кодові слова у вторинному алфавіті $\mathbf{m_2}$, при якій середня довжина повідомлень у вторинному алфавіті має мінімально можливе для даного $\mathbf{m_2}$ значення.

У повідомленнях, складених з кодових слів оптимального коду, статистична надмірність зведена до мінімуму, в ідеальному випадку — до нуля.

Основна теорема кодування для каналу зв'язку без шумів доводить принципову можливість побудови оптимальних кодів, З неї однозначно витікають методика побудови і властивості оптимальних кодів.

Одним з основних положень теореми є те, що при кодуванні повідомлення, розбитого на М-літерні блоки, можна, обравши М достатньо великим, досягти того, щоб середня кількість елементарних символів на одну літеру повідомлення, була скільки завгодно близькою до $\frac{H}{\log m}$.

Різниця $L - \frac{H}{\log m}$ буде тим меншою, чим більше Н, а Н має максимум при рівноімовірних та взаємозалежних символах, звідси витікають основні ***властивості оптимальних кодів***:

- мінімальна середня довжина кодового слова оптимального коду забезпечується у тому випадку, коли надмірність кожного кодового слова зведена до мінімуму (в ідеальному випадку — до нуля);
- кодові слова оптимального коду повинні будуватися з рівноімовірних та взаємонезалежних символів.

З властивості оптимальних кодів витікають принципи їх побудови.

Перший принцип оптимального кодування:

вибір кожного кодового слова необхідно здійснювати так, щоб воно містило максимальну кількість інформації.

Другий принцип оптимального кодування:

Літерам первинного алфавіту, які мають більшу імовірність, присвоюються більш короткі кодові слова у вторинному алфавіті.

Принципи оптимального кодуванні визначають ***методику побудови оптимальних кодів.***

Побудова оптимального коду за ***методом Шеннона-Фано*** для ансамбля з M повідомлень:

1) множину з M повідомлень (символів) розташовують у порядку зменшення імовірностей;
2) впорядковану за п.1 множину далі розбивають на дві групи таким чином, щоб сумарні імовірності повідомлень обох груп були по можливості рівні;
3) першій групі присвоюють символ 0, другій – символ 1;
4) кожну з груп ділять на дві підгрупи так, щоб їх сумарні імовірності були по можливості рівні;
5) першим підгрупам кожної з підгруп знову присвоюють символ 0, другим групам – 1, в результаті чого отримують другі цифри коду. Потім кожну з чотирьох груп знову ділять на рівні (з точки зору сумарної імовірності) частини і т.д. доти, доки в кожній групі не залишаться по одному повідомленню (символу).

Приклад 1. Побудуємо оптимальний код повідомлення, що складається з 8 рівноімовірних літер $A_1 - A_8$.

Розв'язок: оскільки імовірності цього ансамблю рівні $p_1 = p_2 = \ldots = p_8 = \frac{1}{2^3} = 2^{-3}$, а порядок їх розташування не має значення, то розташуємо їх так, як показано у таблиці.

Літера	Кодове слово після розбиття		
	Першого	Другого	Третього
A_1	0	0	0
A_2	0	0	1
A_3	0	1	0
A_4	0	1	1
A_5	1	0	0
A_6	1	0	1
A_7	1	1	0
A_8	1	1	1

Розбиваємо цю множину повідомлень на дві рівноімовірні групи. Першій групі в якості першого символу кодового слова присвоюємо 0, другій – 1. У другій колонці таблиці записуємо чотири нуля і чотири одиниці. Після цього розбиваємо кожну з груп ще на дві рівноімовірні підгрупи. Кожній першій підгрупі присвоюємо 0, другій - 1, і записуємо це у третю колонку таблиці. Далі кожну з чотирьох підгруп розбиваємо на дві рівноімовірні частини, першій з них присвоюємо 0, другій – 1. Таким чином, у четвертій колонці таблиці з'являється значення третього символу кодових слів.

Перевірка оптимальності коду здійснюється шляхом порівняння ентропії кодованого (первинного) алфавіту з середньою довжиною кодового слова у вторинному алфавіті.

Ентропія джерела: $H = \log_2 N = \log_2 8 = 3$ *біт / символ*, а середня кількість двійкових знаків на літеру коду:

$$L = \sum_i^N l(i) \cdot p_i = 3 \cdot \frac{1}{8} \cdot 8 = 3,$$

де $l(i)$ - довжина i-ї кодової комбінації,

p_i - імовірність появи *i*-го символу комбінації довжиною $l(i)$.

Таким чином, $H = L$, тобто код є оптимальним для даного ансамблю повідомлень.

Висновок: для ансамблю рівноімовірних повідомлень оптимальним є рівномірний код. Якщо кількість елементів ансамблю є цілим степенем двох, то завжди $H = L$.

Приклад 2. Побудуємо оптимальний код для передачі повідомлення, в якому імовірності появи літер підкорюються закону $p_i = (\frac{1}{2})^i$, тобто літери даного повідомлення можуть бути розташовані таким чином, що імовірність появи кожної з них буде вдвічі меншою за імовірність появи попередньої.

Літера	Імовірність появи літери	Кодове слово після розбиття						К-ть знаків *l(i)*	$l(i) \cdot p_i$
		1-го	2-го	3-го	4-го	5-го	6-го		
A_1	1/2	0	-	-	-	-	-	1	0,5
A_2	1/4	1	0	-	-	-	-	2	0,5
A_3	1/8	1	1	0	-	-	-	3	0,375
A_4	1/16	1	1	1	0	-	-	4	0,25
A_5	1/32	1	1	1	1	0	-	5	0,15625
A_6	1/64	1	1	1	1	1	0	6	0,09175

Середня кількість двійкових знаків на літеру коду:

$$L = \sum_{i}^{N} l(i)\, p_i = 0.5 + 0.5 + 0.375 + 0.25 + 0.15625 + 0.09175 = 1.873,$$

а ентропія джерела повідомлень

$$H = -(p_1 \cdot \log_2 p_1 + p_2 \cdot \log_2 p_2 + \ldots + p_6 \cdot \log_2 p_6) =$$
$$= 0.5 + 0.5 + 0.375 + 0.2487 + 0.1554 + 0.0909 = 1.87 \quad \textit{біт / символ}$$

Деяке розходження у третій цифрі після коми пояснюється тим, що у даному коді $\sum_i p_i \neq 1$, тобто даний ансамбль повідомлень не є повним ($\sum_i p_i \approx 0,984$). Однак чим довший буде обраний ряд значень A_i ,тим ближче $\sum_i p_i$ буде до 1, тим ближче L буде до H. Таким чином, $L \approx H$, тобто код є оптимальним для даного ансамблю повідомлень.

Висновок: число елементарних символів на літеру повідомлення з розподіленням імовірностей літер за законом $p_i = (\frac{1}{2})^i$ зростає в порядку зменшення імовірності як натуральний ряд чисел (1,2,3,…, M), якщо i=1,2,3,…,M.

Код, розглянутий у цьому прикладі, є зручним для декодування, оскільки кожне кодове слово закінчується нулем, який відділяє кодові слова одне від іншого.

<u>Приклад 3.</u> Побудувати оптимальний нерівномірний код (ОНК) для передавання повідомлень, в яких імовірності появи літер первинного алфавіту дорівнюють:

Літера	p_i	Кодове слово						К-ть знаків $l(i)$	$l(i) \cdot p_i$
		0	-	-	-	-	-		
A_1	0,5	0	-	-	-	-	-	1	0,5
A_2	0,25	1	0	-	-	-	-	2	0,5
A_3	0,098	1	1	0	0	-	-	4	0,392
A_4	0,052	1	1	0	1	-	-	4	0,208
A_5	0,04	1	1	1	0	-	-	4	0,16
A_6	0,03	1	1	1	1	0	-	5	0,15
A_7	0,019	1	1	1	1	1	0	6	0,14
A_8	0,011	1	1	1	1	1	1	6	0,066

$$L = 0.5 + 0.5 + 0.392 + 0.208 + 0.16 + 0.15 + 0.14 + 0.066 = 2.216,$$

$$H = -\sum_{i=1} p_i \log_2 p_i = -(0.5\log_2 0.5 + 0.25\log_2 0.25 + \ldots + 0.011\log_2 0.011) =$$

$$= 0.5 + 0.5 + 0.3284 + 0.2217 + 0.1875 + 0.1517 + 0.1086 + 0.0715 = 2.0676 \quad \textit{біт / символ}$$

На основі розглянутих прикладів можна сказати, що оптимальним є ті коди, у яких середня довжина кодової комбінації є мінімальною і мало відрізняється від ентропії джерела повідомлень.

Максимально ефективними є ті ОНК, у яких

$$\log_2 m \sum_{i=1}^{N} l(i) \cdot p_i = l_{сер} = H$$

де m та N – символи відповідно вторинного та первинного алфавітів.

Для двійкових кодів

$$l_{сер} = \sum_{i=1}^{N} l(i) \cdot p_i = -\sum_{i=1}^{N} p_i \log_2 p_i$$

за умови $l(i) = -\log_2 p_i = \log_2 \frac{1}{p_i}$

Величина $l_{сер}$ точно дорівнює H, якщо $p_i = \frac{1}{m^{n_i}}$, де n_i - ціле число. Якщо n_i не є цілим числом для всіх літер первинного алфавіту, то $l_{сер} > H$ і згідно основній теоремі кодування, наближається до H із збільшенням довжини кодованих блоків.

Ефективність ОНК оцінюють коефіцієнтом статистичного стиснення

$$K_{сс} = \frac{H_{\max}}{l_{сер}} = \frac{\log_2 N}{\log_2 m \sum_{i=1}^{n} l(i) \cdot p_i},$$

який характеризує зменшення кількості елементарних символів на літеру повідомлення в порівнянні з нестатистичним кодуванням.

Коефіцієнт відносної ефективності:

$$K_{в.е.} = \frac{H}{l_{сер}}$$

Для загального випадку $K_{в.е.} = \dfrac{-\sum_{i=1}^{n} p_i \log_2 p_i}{\log_2 m \sum_{i=1}^{n} l(i) \cdot p_i}$

Для прикладу 3: $K_{сс} = \dfrac{H_{\max}}{l_{сер}} = \dfrac{\log_2 8}{\sum_{i=1}^{8} l(i) \cdot p_i} = \dfrac{3}{2.216} = 1.45$

$$K_{в.е.} = \frac{H}{l_{сер}} = \frac{2.0676}{2.216} \approx 0.93$$

Якщо первинний алфавіт складається з рівноімовірних символів, імовірності яких дорівнюють цілим від'ємним степеням двох, то $K_{в.е.} = 1$.

Методика Хаффмена побудови оптимальних рівномірних кодів (ОНК). Методика використовує для побудови кодів кодові дерева.

Хаффмен показав, що для отримання мінімально можливої довжини кода з кількістю взаємозалежних літер первинного алфавіту N

$$L_{сер} = \log_2 m \sum_{i=1}^{N} l(i) \cdot p_i$$

необхідно і достатньо виконання наступних умов:

1) якщо виписати символи в порядку зменшення імовірностей $p_i \geq p_j$, то при $i < j$ $l(i) \leq l(j)$, тобто якщо імовірність більше, довжина кодової комбінації менше.

2) Декілька (n_0) останніх, але не більше m? кодових слів рівні по довжині і відрізняються менше останнім символом, при цьому

$$2 \leq n_0 \leq m,$$

де m - кількість якісних ознак вторинного алфавіту, а n_0 - кількість найменш імовірних повідомлень , які об'єднуються на першому етапі побудови кодового дерева. Крім того, $\dfrac{N - n_0}{m - 1} = a,$ де а - ціле позитивне число.

3) будь-яка можлива послідовність $l_n - 1$ кодових слів повинна або сама бути кодовою комбінацією, або мати своїм префіксом дозволену кодову комбінацію

Методика Хаффмена наступна:

1) Символи первинного алфавіту виписують у порядку зменшення імовірностей. Останні n_0 символів, де $2 \le n_0 \le m$ та $\frac{N - n_0}{m-1} = a$ об'єднуються у деякий новий символ з імовірністю , рівною сумі імовірностей цих символів.

2) Символи, що залишилися, разом з утвореним символом знову об'єднують і отримують новий допоміжний символ.

3) Знову виписують символи в порядку зменшення імовірностей з урахуванням допоміжного символу і т.д. доти, доки імовірності m символів , що залишилися після $\frac{N - n_0}{m-1}$ -го виписувань не дадуть в сумі 1.

4) При кожному об'єднанні верхньому символу приписують значення m , наступному m-1, і т.д., нижньому – 0.

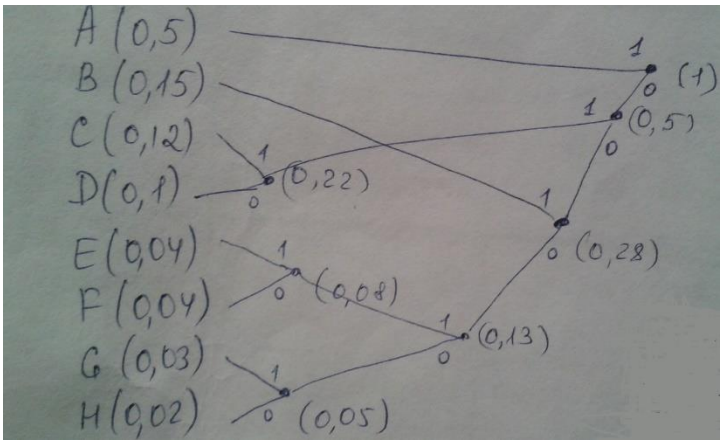
5) Для кожного символу виписують код комбінації як послідовність коду знака на шляху від 1 до цього символу.

На практиці звичайно не виконують багаторазового виписування, а обходяться елементарними геометричними побудовами, суть яких для кодів з кількістю якісних ознак m=2 зводиться до того, що символи попарно об'єднуються у нові символи, починаючи з тих, що мають найменшу імовірність, а потім, з урахуванням імовірностей утворених символів, знову проводять попарне об'єднання символів з найменшими імовірностями, і таким чином будують двійкове кодове дерево, у вершині якого стоїть символ з імовірністю 1.

Приклад 1. Побудувати ОНК для передавання повідомлень, що складаються з літер алфавіту, які мають наступні імовірності появи:

Літера	A	B	C	D	E	F	G	H
p_i	0.5	0.15	0.12	0.1	0.04	0.04	0.03	0.02

Розв’язок.



1) Спочатку знаходять літери з найменшими імовірностями H(0,02) та G(0,03) , проводять від них лінією до точки , де імовірність появи літери H або літери G дорівнює 0,05.

2) Потім беруть дві найменші імовірності F(0,04) та E(0,04) і отримують нову точку з імовірністю 0,08.

3) Тепер найменші імовіріності мають точки, що відповідають допоміжним символам (0,05) та (0,08). З’єднуємо їх лініями та отримуємо точку (0,13).

4) Продовжуємо об'єднання, доки лінії від основних та допоміжних символів не об’єднуються у точці, яка дає сумарну імовірність, рівну 1.

5) Позначимо кожну верхню лінію з пари цифрою 1, нижню - ∅

Отримаємо ОНК, який для кожної літери являє собою послідовність нулів та одиниць, які зустрічаються на шляху до даної літери від кінцевої точки (1).

Літера	A	B	C	D	E	F	G	H
p_i	1	001	011	010	00011	00010	00001	00000
К-ть знаків $l(i)$	1	3	3	3	5	5	5	5
$l(i) \cdot p_i$	0,5	0,45	0,36	0,3	0,2	0,2	0,15	0,1

Отримані коди можуть бути однозначно декодовані на приймальному боці завдяки тому, що жоден код не є початком іншого. Більше того, такий код може автоматично відновити правильний зміст навіть у тому випадку, коли декодування почалося не з початку повідомлення або код був прийнятий з помилкою.

Ємність каналу звязку для побудованого ОНК визначається як

$$C = \frac{L_{сер}}{\tau} = \frac{\sum l(i) \cdot p_i}{\tau}, \quad \text{[знак/сек]}$$

де τ - час передавання однієї літери (величина, зворотна швидкості появи знаків на виході джерела повідомлень).

Лекція 9. Ідея корекції помилок

Для того, щоб у прийнятому повідомленні можна було знайти помилки, воно повинне мати деяку надмірність, яка дозволяє відрізнити помилковий код від правильного. Наприклад, якщо передане повідомлення складається з трьох абсолютно однакових частин, то у прийнятому повідомленні відділення правильних символів від помилкових може бути здійснене за результатами накопичення однакових символів. Для двійкових кодів це можна проілюструвати так:

Передана комбінація:	10110	Кожного разу помилка в одному символі, але в різних символах 3 рази.
1-ша прийнята комбінація:	10010	
2-га----------:	10100	
3-тя----------:	00110	
Накопичена комбінація :	10110	

Надмірність - n_k символів.

Разом $n = n_i + n_k$ символів — довжина кодової комбінації

n_i символів ➔ кодуючий пристрій ➔ n символів

Якщо код двійковий (m=2), то всього є 2^{n_i} вхідних послідовностей, та 2^n вихідних послідовностей — дозволених кодових комбінацій.

Інші $2^n - 2^{n_i}$ комбінації не використовуються для передавання. Це заборонені кодові комбінації.

Оскільки будь-яка з 2^{n_i} дозволених комбінацій в результаті завад може трансформуватися в будь-яку іншу, то є $2^n \cdot 2^{n_i}$ можливих випадків передавання. До цього значення входять:

1) 2^{n_i} випадків безпомилкового передавання

2) $2^{n_i}(2^{n_i} - 1)$ випадків переходу у інші дозволені комбінації, що відповідає не знайденим помилкам.

3) $2^{n_i}(2^n - 2^{n_i})$ випадків переходу у недозволені комбінації, які можуть бути знайдені.

Отже, частина помилкових комбінацій, що можуть бути знайдені, від загальної кількості можливих випадків передавання складає:

$$\frac{2^{n_i}(2^n - 2^{n_i})}{2^{n_i} \cdot 2^n} = 1 - \frac{2^{n_i}}{2^n}$$

Приклад. Визначити корегуючу властивість коду, кожна комбінація якого містить один надлишковий символ:

$$n = n_i + 1.$$

Розв'язок. Загальна кількість вихідних комбінацій - 2^{n_i+1}, вхідних комбінацій - 2^{n_i}, тобто в 2 рази менша.

За підмножину дозволених комбінацій можна прийняти, наприклад, 2^{n_i} комбінацій, що містять парну кількість нулів (або одиниць).

При кодуванні до кожної послідовності символів додають один символ (0 або 1) такий, щоб кількість нулів у комбінації була парною. Наявність непарної кількості нулів переводить дозволену комбінацію у підмножину заборонених комбінацій, що виявляється на приймальному боці.

Частина виявлених випадків складає:

$$1 - \frac{2^{n_i}}{2^n} = 1 - \frac{2^{n_i}}{2^{n_i+1}} = \frac{1}{2}.$$

Всього випадків переходу у заборонені комбінації:

$$2^{n_i}(2^n - 2^{n_i}).$$

Таким чином, за умови надмірності будь-який код здатний виявляти та виправляти помилки.

Коди без надмірності знаходити, а тим більше виправляти помилки не можуть.

Мінімальну кількість символів, в яких дві кодові комбінації відрізняються одна від одної, називають кодовою відстанню.

Мінімальну кількість символів, в яких всі кодові комбінації одного коду відрізняються, називають мінімальною кодовою відстанню.

Це параметр, який визначає завадостійкість коду та закладену у коді надмірність, а також корегуючі властивості коду.

Зв'язок корегуючої властивості коду з кодовою відстанню. В загальному випадку для знаходження r помилок кодова відстань має бути:

$$d_0 = r + 1.$$

Для одночасного знаходження та виправлення помилок:

$$d_0 = r + s + 1,$$

де s – кількість помилок, що виправляються.

Для кодів, які тільки виправляють помилки:

$$d_0 = 2s + 1.$$

Щоб визначити кодову відстань між двома комбінаціями двійкового коду, треба додати їх за модулем 2 та підрахувати кількість одиниць у результаті.

Наприклад, кодова відстань між комбінаціями 0001 та 0001 (однаковими):

$$\begin{array}{l} 0001 \\ \oplus \quad\quad\quad \Rightarrow d = 0 \\ \underline{0001} \\ 0000 \end{array}$$

Між комбінаціями 11000111001 та 10000011101

$$\begin{array}{l} 11000111001 \\ \oplus \quad\quad\quad\quad\quad \Rightarrow d = 3 \\ \underline{10000011101} \\ 01000100100 \end{array}$$

Більш повне уявлення про властивості коду дає матриця відстаней D, елементи якої $d_{ij}(i, j = 1,2,...,m)$ дорівнюють кодовим відстаням між кожною парою з усіх дозволених комбінацій.

Приклад. Представити матрицею відстаней код

$$x_1 = 000; x_2 = 001; x_3 = 010; x_4 = 111$$

Розв'язок. min кодова відстань для коду $d_0 = 1$

Матриця 4-го порядку має вигляд

	x_1 000	x_2 001	x_3 010	x_4 111
x_1 000	0	1	1	3
x_2 001	1	0	2	2
x_3 010	1	2	0	2
x_4 111	3	2	2	0

Найбільш імовірним є перехід у кодову комбінацію, яка відрізняється у найменшій кількості символів.

Декодування після приймання відбувається таким чином, що прийнята комбінація утотожнюється з тією дозволеною комбінацією, яка знаходиться від неї на найменшій кодовій відстані.

Таке декодування називається декодуванням за методом максимальної правдоподібності.

Очевидно, що при $d_0 = 1$ всі комбінації є дозволеними. Наприклад, при n=3 дозволені комбінації утворюють множину:

000

001

010

011

100

101

110

111

Будь-яка одинична помилка трансформує комбінацію у дозволену. Це випадок безнадмірного коду, який не має корегуючої властивості.

Якщо $d_0 = 2$, то жодна з дозволених кодових комбінацій при одиничній помилці не переходить в іншу дозволену комбінацію. Наприклад, підмножина дозволених комбінацій може бути утворена за принципом парності одиниць.

Тоді для n=3:

000,011,101,110- дозволені комбінації

001,010,100,111- заборонені комбінації

Такий код може виявляти одиничні помилки (r=1)

Для виправлення одиничної помилки необхідно проаналізувати підмножину заборонених комбінацій.

d_0 повинно бути не менше 3.

При n=3, $d_0 = 3$ можна прийняти за дозволені комбінації, наприклад, 000 та 111.

Тоді дозволеній комбінації 000 треба приписати підмножину заборонених комбінацій 001,010,100, які утворюються в результаті одиничної помилки в комбінації 000.

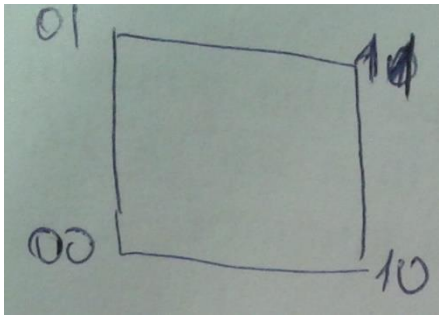
Комбінації 111 приписується підмножина заборонених комбінацій 110,101,011, що утворюються в результаті одиничної помилки в комбінації 111.

Дозволені комбінації $000 \Rightarrow 100$, $111 \Rightarrow 110$ — заборонені комбінації: 001, 010, 011, 101

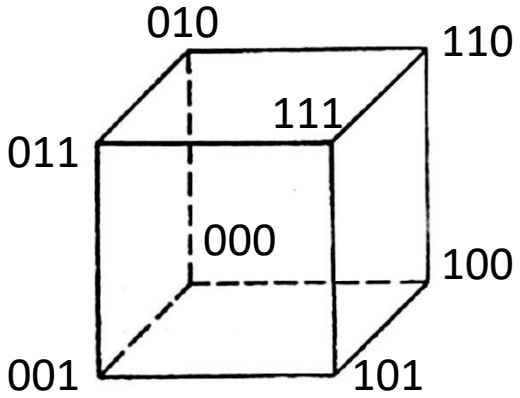
В загальному випадку для забезпечення можливості виправлення всіх помилок кратності до ***s*** включно при декодуванні за методом максимальної правдоподібності кожна з помилок повинна приводити до забороненої комбінації, що відноситься до підмножини початкової дозволеної комбінації.

Будь-яка *n*-розрядна двійкова комбінація може бути інтерпретована як вершина *n*-вимірного одиничного куба, тобто куба з одиничною довжиною ребра.

При *n*=2 кодові комбінації розташовуються у вершинах квадрату:



При n=3 – у вершинах куба:



В загальному випадку *n*-вимірний куб має 2^n вершин, що дорівнює максимально можливій кількості кодових комбінацій.

Така модель дає просту геометричну інтерпретацію кодової відстані між комбінаціями. d_0 відповідає кількості ребер куба, які необхідно пройти, щоб досягти однієї комбінації від іншої.

Проілюструємо побудову корегуючого коду на наступному прикладі. Хай початковий алфавіт складається з 4-х літер і закодований наступним кодом:

$x_1 = 00$

$x_2 = 01$

$x_3 = 10$

$x_4 = 11$

Цей код використовує всі можливі комбінації довжиною 2, і тому не може виявляти помилки (оскільки $d_0 = 1$)

Приписуємо до кожної комбінації 0 або 1 так, щоб була парна кількість одиниць:

$x_1 = 000$

$x_2 = 011$

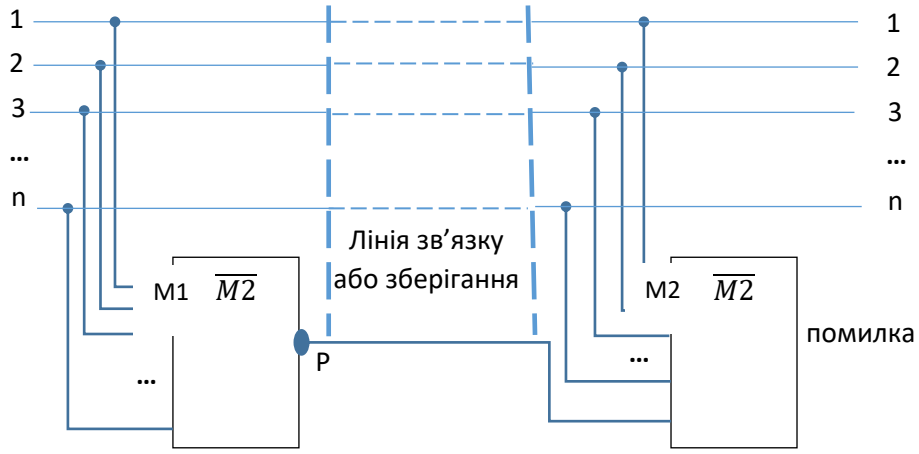
$x_3 = 101$

$x_4 = 110$

Для цього коду $d_0 = 2$, отже, він здатний виявляти всі однократні помилки.

Оскільки будь-яка заборонена комбінація містить непарну кількість одиниць, то для знаходження помилки достатньо перевірити комбінацію на парність (наприклад, додаванням за модулем 2 цифр комбінації). Якщо кількість одиниць парна, то сума за модулем 2 буде 0, якщо непарна – 1. Ознакою парності називають інверсію цієї суми.

Розглянемо загальну схему організації контролю парності (або контролю непарності).



На інвертуючому виході n-входового елемента «М1» (логічна функція «виключаюче АБО») формується ознака парності числа Р, яка в якості додаткового (n+1)-го контрольного розряду відправляється разом з переданою послідовністю n інформаційних символів до лінії зв'язку або запам'ятовувального пристрою. (n+1)-розрядне слово має завжди гарантовану непарну кількість одиниць. Якщо у початковому слові вона була непарною, то інверсія функції М1 від такого слова дорівнює 0, і нульове значення контрольного розряду не змінює кількість одиниць при передаванні слова.

Якщо ж кількість одиниць у початковому слові була парною, то контрольний розряд Р для такого числа буде дорівнювати 1, і результуюча кількість одиниць у переданому (n+1)-му розряді стане непарною.

На приймальному боці або з пам'яті відбувається контроль парності отриманого (n+1)-розрядного слова (блок «М2» з логічною функцією «Виключаюче АБО»). Якщо результат дорівнює 1, то або у переданому слові, або у контрольному розряді при передаванні або зберіганні з'явилася помилка.

Такий простий контроль не дозволяє виправити помилку, але принаймні дає можливість виявляти помилки, виключити помилкові дані, запросити повторне передавання і т.д.

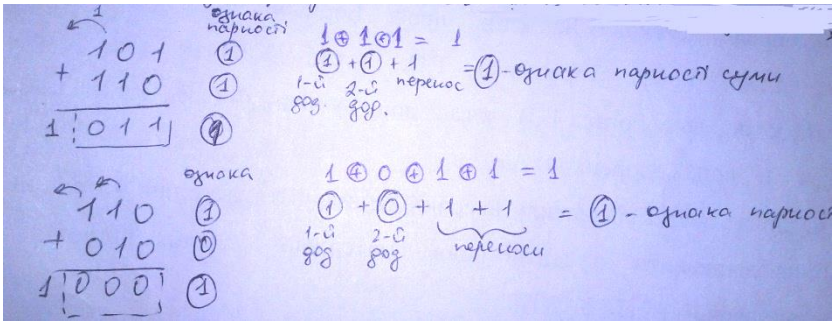
Систему контролю можна побудувати не тільки на інверсії функцій М1, М2, але і на простих М1,М2. Однак в цьому випадку, якщо початковий код буде складатися з усіх нулів, контрольний розряд теж буде дорівнювати 0. Тоді в лінію буде надіслано посилку з усіх нулів, і на приймальному боці її не можна буде відрізнити від досить небезпечної несправності - повного пропадання зв'язку. Тому найчастіше використовують саме інвертуючу логічну функцію «Виключаюче АБО».

Контроль по парності будується на тому, що одинична помилка (не має значення — пропадання одиниці або поява зайвої) інвертує ознаку парності. Однак дві помилки проінвертують її двічі, тобто залишать незмінною. Таким чином, подвійну помилку контроль по парності не виявить.

Контроль по парності може виявляти всі непарні помилки і не зреагує на парні. Це викликало надто малою надмірністю, що дорівнює лише одному розряду.

Для кращого контролю необхідна і більша надмірність.

Ознаку парності можна використовувати для контролю лише незмінних даних. При виконанні над даними будь-яких логічних операцій ознаки парності в загальному випадку змінюються. Виключенням є операція арифметичного додавання: сума за модулем 2 ознак парності двійкових доданків і всіх переносів, які відбулися в процесі додавання, дорівнює ознаці парності коду арифметичної суми цих доданків, наприклад:



Контроль за парністю - найдешевший з точки зору апаратних витрат вид контролю, він широко застосовується. Практично будь-який канал передавання цифрових даних або запам'ятовувальний пристрій, якщо вони не мають більш сильних методів контролю, захищені контролем парності.

Для того, щоб код був здатний і виправляти одиничні помилки, необхідно додати ще не менше 2-х розрядів. Це можна зробити різними способами, наприклад, повторити перші 2 цифри в кінці кожної кодової комбінації.

Для коду, розглянутого раніше:

$$\begin{aligned} x_1 &= 000 \qquad x_1 = 000:00 \\ x_2 &= 011 \quad \Rightarrow \quad x_2 = 011:01 \\ x_3 &= 101 \qquad x_3 = 101:10 \\ x_4 &= 110 \qquad x_4 = 110:11 \end{aligned}$$

	x_1	x_2	x_3	x_4
x_1	0	3	3	4
x_2	3	0	4	3
x_3	3	4	0	3
x_4	4	3	3	0

Видно, що $d \geq 3$, що відповідає нерівності $d \geq r + S + 1$

Для знаходження та виправлення одиничної помилки співвідношення між кількістю інформаційних n_i та корегуючих розрядів n_k має задовольняти наступним умовам

$$2^{n_k} \geq n + 1 \quad (1)$$

$$2^{n_i} \le \frac{2n}{n+1} \quad (2)$$

При цьому загальна довжина кодової комбінації

$$n = n_i + n_k \quad (3).$$

Лекція 10. Систематичні коди

Систематичні коди – це коди, в яких інформаційні та корегуючі символи розташовані за строго визначеною системою і завжди займають певні місця у кодових комбінаціях. Систематичні коди є рівномірними, тобто всі комбінації коду із заданими корегуючими властивостями завжди мають однакову довжину.

Типовий приклад систематичного коду – **код Хемінга**. Для знаходження коду задається значення n_i або кількість повідомлень $N = m^{n_i}$. За допомогою виразів (1), (2) та (3) обчислюється n_k та n. При цьому можна користуватися таблицею для коду Хемінга

n	1	2	3	4	5	6	7	8	…
n_i	0	0	1	1	2	3	4	4	…
n_k	1	2	2	3	3	3	3	4	…

Номери позицій контрольних символів зручно обирати за законом 2^i, де і =0,1,2,… , тобто номери корегуючих символів: 1,2,4,8,16,…

Потім визначаються **перевірочні позиції.** Для цього складається таблиця двійкових кодів натуральних чисел. Кількість колонок у таблиці $n = n_i + n_k$:

n	1	2	3	4	5	6	7	8	…
Двійк.код	0001	0010	0011	0100	0101	0110	0111	1000	…
Перевір.коеф	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	…

Числу 1 відповідає перевірочний коефіцієнт a_1,числу 2 - a_2 і т.д. Ці коефіцієнти далі виписують за наступним принципом:

Перша перевірка: всі перевірочні коефіцієнти, що містять 1 у молодшому розряді:

$$a_1, a_3, a_5, a_7, ...$$

Друга перевірка: всі перевірочні коефіцієнти, що містять 1 у другому розряді:

$$a_2, a_3, a_6, a_7, \ldots$$

і т.д.

Номери перевірок коефіцієнтів відповідають номерам позицій, які входять до загальної таблиці перевірок:

№ перевірки	Перевірочні позиції	№ контрольного символу(розряду)
1	1,3,5,7,…	1
2	2,3,6,7,…	2
3	4,5,6,7,…	3
4	8,9,10,11,…	4

Потім визначають значення корегуючих символів за правилом: сума одиниць на перевірочних позиціях у всіх перевірках має бути парною. Якщо вона парна, то значення корегуючого символу=0, інакше =1.

Якщо у прийнятому коді є помилка, то результати перевірок по контрольних позиціях утворюють двійкове число, що вказує номер помилкової позиції. Помилку виправляють, змінюючи символ помилкової позиції на протилежний.

Приклад 1. Побудувати код Хемінга для комбінації інформаційних символів 0101.

Розв'язок. $n_i = 4$. Знаходимо $n = 7, n_k = 3$. Корегуючі символи будуть на позиціях 1,2,4. Код має вигляд:

$$\overset{1\ \ 2\ \ 3\ \ 4\ \ 5\ 6\ 7}{k_1 k_2 0 k_3 101}$$

Користуючись таблицею перевірочних позицій, визначаємо k_1, k_2, k_3:

1-ша перевірка: $П_1 \oplus П_3 \oplus П_5 \oplus П_7$ має бути парною

$$k_1 \oplus 0 \oplus 1 \oplus 1 = 0, \text{ якщо } k_1 = 0$$

2-га перевірка: $П_2 \oplus П_3 \oplus П_6 \oplus П_7$ має бути парною

$$k_2 \oplus 0 \oplus 0 \oplus 1 = 0, \text{якщо } k_2 = 1$$

3-тя перевірка: $П_4 \oplus П_5 \oplus П_6 \oplus П_7$ має бути парною

$$k_3 \oplus 1 \oplus 0 \oplus 1 = 0, \text{ якщо } k_3 = 0$$

Остаточно корегуючий код:

0100101.

Припустимо, в результаті спотворень в каналі замість 0100101 було прийнято 0100111 (помилка у шостому розряді).

Для знаходження помилки виконаємо перевірки на парність, аналогічні перевіркам при виборі корегуючих символів.

1-ша перевірка: $П_1 \oplus П_3 \oplus П_5 \oplus П_7 = 0 \oplus 0 \oplus 1 \oplus 1 = 0$ парна.

молодший розряд номеру помилкової позиції $= \varnothing$

2-га перевірка: $П_2 \oplus П_3 \oplus П_6 \oplus П_7 = 1 \oplus 0 \oplus 1 \oplus 1 = 1$ - непарна

другий розряд номеру помилкової позиції $= 1$.

3-тя перевірка: $П_4 \oplus П_5 \oplus П_6 \oplus П_7 = 0 \oplus 1 \oplus 1 \oplus 1 = 1$ - непарна

третій розряд номеру помилкової позиції $= 1$.

Номер помилкової позиції : 110=6. Отже, символ 6-ї позиції треба замінити на зворотній.

Для виправлення одиничної та знаходження подвійної помилки, крім перевірок по перевірочним позиціям, треба зробити ще одну перевірку на парність для кожного коду. Щоб це зробити, треба до кожного коду в кінці додати контрольний символ таким чином, щоб сума одиниць в отриманій комбінації була парною. Тоді у випадку однієї помилки перевірки по позиціях вкажуть номер помилкової позиції, а перевірка на парність вкаже наявність помилки. Якщо перевірки позиції вкажуть на наявність помилки, а перевірка на парність не фіксує помилки, значить, у коді дві помилки.

Так, для розглянутого прикладу:

- для правильної комбінації 8-й символ контролю парності дорівнює 1;

- для комбінації з помилкою у 6-му розряді 8-й символ дорівнює 0;
- для комбінації з двома помилками у 6-му та у 7-му розрядах 8-й символ дорівнює 1 (не фіксує помилки).

n=7 розрядів коду Хемінга							8-й символ контролю парності
0	1	0	0	1	0	1	1
0	1	0	0	1	**1**	1	0
0	1	0	0	1	**1**	**0**	1

Якщо по вказаних правилах будувати корегуючий код Хемінга з восьмим додатковим символом контролю парності для знаходження та виправлення однократної помилки, виявлення двократної помилки, то перші 8 комбінацій такого коду матимуть вигляд, наведений у таблиці.

№ комбінації	n=7 розрядів коду Хемінга							8-й символ контролю парності
0	0	0	0	0	0	0	0	0
1	1	1	0	1	0	0	1	0
2	0	1	0	1	0	1	0	1
3	1	0	0	0	0	1	1	1
4	1	0	0	1	1	0	0	1
5	0	1	0	0	1	0	1	1
6	1	1	0	0	1	1	0	0
7	0	0	0	1	1	1	1	0
8	1	1	1	0	0	0	0	1

Лекція 11. Циклічні коди

Циклічні коди є частковим випадком систематичних лінійних (n, k)-кодів. Назву циклічні коди отримали через свою основну властивість: циклічна перестановка (зсув) символів дозволеної кодової комбінації дає також дозволену кодову комбінацію. При циклічній перестановці символи кодового слова переміщуються зліва направо на одну позицію, причому крайній правий символ переноситься на місце крайнього лівого.

Якщо, наприклад, А1 – 101100 є дозволеною комбінацією, то дозволеною кодовою комбінацією буде і А2 – 010110, що отримана циклічною перестановкою. Відмітимо, що перестановка виконується разом з перевірочними символами, і за правилами лінійних кодів сума за модулем 2 дозволених кодових комбінацій дає також дозволену кодову комбінацію.

Опис циклічних кодів пов'язаний з представленням кодових комбінацій у вигляді поліномів (многочленів) фіктивної змінної *x*. n-розрядну комбінацію можна описати поліномом (n-1)-го степеню у вигляді:

$$A_{n-1}(x) = a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + ... + a_1x + a_0,$$

де $a_i = \{0,1\}$, причому $a_i = 0$ відповідають нулям у комбінації, а $a_i = 1$ - одиницям.

Для **прикладу** переведемо кодове слово А= 001011 в поліноміальний вигляд:

Номер розряду *i*	6	5	4	3	2	1
Код	0	0	1	0	1	1

При цьому $A(x)=0x^5+0x^4+1x^3+0x^2+1x^1+1x^0 = x^3+x+1$.

Зсув ліворуч на один розряд дає: $0x^6+0x^5+1x^4+0x^3+1x^2+1x^1$ (001011).

Для отримання циклічного зсуву потрібно лівий доданок перенести в крайню праву позицію: $0x^5+1x^4+0x^3+1x^2+1x^1+0x^6$ (010110) і прийняти $x^6 \equiv x^0$.

Тоді $A2(x)= 0x^5+1x^4+0x^3+1x^2+1x^1+0x^0$ (А2=010110), або $A2(x)= x^4+x^2+x$.

Звідки слідує: $x^6=x^n=x^0=1$.

Дії з кодовими векторами, представленими у вигляді поліномів, виконуються за правилами арифметики «по модулю 2», в якій віднімання рівносильне додаванню. Тому $x^k(+)x^k = 0$, і при подальших операціях з поліномами необхідно викреслювати пари фіктивних змінних x з однаковими ступенями.

Утворюючий поліном. Ідея побудови циклічних кодів базується на використанні неподільних поліномів. Неподільним називається поліном, який не можна представити у вигляді добутку поліномів нижчих степенів, тобто такий поліном ділиться тільки на самого себе та на 1 і не ділиться на жодний інший поліном.

На такий поліном ділиться без залишку двочлен x^n+1.

Утворююча матриця циклічного коду має вигляд:

$$V = \begin{vmatrix} g(x) \\ g(x)\cdot x \oplus C_2(x^n+1) \\ g(x)\cdot x^2 \oplus C_3(x^n+1) \\ \dots \\ g(x)\cdot x^{m-1} \oplus C_m(x^n+1) \end{vmatrix},$$

де g(x) - початкова кодова комбінація, якій відповідає утворюючий поліном і на базі якої утворено всі інші (m-1) комбінацій, $C_i=0$, якщо результуюча степінь полінома $g(x)\cdot x^i$ не перевищує (n-1), та $C_i=1$, якщо перевищує.

Двійкова комбінація, що відповідає утворюючому поліному g(x), називається утворюючою комбінацією. Для побудови циклічного коду із заданими властивостями достатньо вірно обрати g(x).

Утворюючий поліном g(x) повинен задовольняти наступні вимоги:

1. g(x) має бути ненульовим.

2. Вага g(x), тобто кількість його ненульових коефіцієнтів, має бути не менше мінімальної кодової відстані:

$$v\big(g(x)\big) \geq d_0 .$$

3. g(x) повинен мати максимальний степінь n_k, де n_k – кількість додаткових (надмірних, корегуючих, перевірочних) розрядів у кожній кодовій комбінації.

4. g(x) має бути дільником поліному x^n+1.

Виконання умови 4 призводить до того, що всі робочі кодові комбінації циклічного колу набувають властивості ділитися без залишку на g(x). Враховуючи це, можна сказати, що у циклічному коді всі комбінації діляться на утворюючий поліном без залишку.

Для визначення степеню утворюючого поліному можна скористатися виразом:

$$n_k \geq \log_2(n+1),$$

де n – загальна довжина кодової комбінації. Таким чином, утворюючий поліном слід обирати якомога коротким, але з дотриманням вищенаведених умов.

В залежності від корегуючих властивостей (кількості помилок, що виправляються кодом) співвідношення між кількістю інформаційних n_i та корегуючих n_k розрядів визначається наступною таблицею.

n	7	15	15	15	31	31	31	31	31	…
n_i	4	11	7	5	26	21	16	11	6	…
n_k	3	4	8	10	5	10	15	20	25	…
s	1	1	2	3	1	2	3	5	7	…

Утворення кодових комбінацій циклічного коду. Всі комбінації циклічного коду можна записати як рядки матриці. Поліном кожного наступного рядка утворюється з попереднього шляхом множення на x. При цьому, якщо крайній лівий символ відрізняється від нуля, для реалізації операції переносу одиниці в кінець комбінації з результату необхідно відняти (додати за модулем 2) поліном x^n+1.

№	Комбінації циклічного коду	Пояснення	Відповідний поліном	Пояснення
1	001011	Початкова комбінація	$x^3 + x + 1$	
2	010110	Циклічний зсув попередньої комбінації на 1 розряд ліворуч	$x^4 + x^2 + x = \\ = \left(x^3 + x + 1\right) \cdot x$	Множення попередньої комбінації на x
3	101100	--//--	$x^5 + x^3 + x^2 = \\ = \left(x^4 + x^2 + x\right) \cdot x$	--//--
4	011001	--//--	$x^4 + x^3 + 1 = \\ = \left(x^5 + x^3 + x\right) \cdot x \oplus (x^6 + 1)$	Множення попередньої комбінації на x; якщо у старшому розряді попередньої комбінації 1, то до результату додається за модулем 2 значення x^n+1 (n – максимальна степінь полінома після множення на x)

Результуюча комбінація у випадку, коли старший розряд дорівнює 1 (множення на x попередньої комбінації та наступне додавання за модулем 2 двочлену (x^n+1)) визначається також альтернативним шляхом як залишок від ділення попередньої комбінації, помноженої на x, на двочлен (x^n+1):

$$\begin{array}{l|l} \left(x^{n-1} + x^{n-2} + \ldots + x + 1\right) \cdot x = x^n + x^{n-1} + \ldots + x^2 + x & x^n + 1 \\ \hline \oplus \quad x^n + 1 & 1 \\ \hline x^{n-1} + \ldots + x^2 + x + 1 & \end{array}$$

Лекція 12. Спосіб утворення кодового многочлена

Інформаційний многочлен A(x) множиться на x^{nk}, де n_k - старший ступінь утворюючого многочлена g(x), і отриманий вираз ділиться на g(x). В результаті вийде частка Q(x) і залишок R(x) ступеня менше n_k:

$$x^{nk}A(x) = g(x)Q(x) + R(x) .$$

Перенесемо R(x) в ліву частину:

$$x^{nk}A(x) + R(x) = g(x)Q(x) .$$

Права частина останньої рівності кратна g(x) і, відповідно, є кодовим многочленом S(x), тобто

$$x^{nk}A(x) + R(x) = S(x).$$

Саме так отримуються кодові многочлени. Перший доданок кодового многочлена має нульові коефіцієнти в n_k молодших членах. Цей многочлен відповідає зсуву вліво на n_k розрядів інформаційної частини A(x). Ступінь многочлена R(x) менше n_k, тому його коефіцієнти не змінюють нульові коефіцієнти першого многочлена при їх розкладанні. Таким чином, інформаційні елементи в кодовій комбінації зберігаються, а перевірочними елементами є коефіцієнти залишку R(x), число яких дорівнює ступеню породжувального многочлена.

Приклад. Визначити кодову комбінацію, якщо інформаційна частина 101011, а породжувальний многочлен $g(x) = x^2 + x + 1$, $n_k=2$

Очевидно, що

$$A(x) = x^5 + x^3 + x + 1,$$

$$S(x) = x^2A(x) + R(x) = x^7 + x^5 + x^3 + x^2 + x + 1,$$

чому відповідає кодова комбінація 10101111. Шість перших елементів – інформаційні, решта два – перевірні, які відповідають залишку:

$$R(x) = x + 1.$$

Теорема. Многочлен g(x) є породжувальним многочленом лінійного циклічного коду довжини n тоді і тільки тоді, коли g(x) ділиться на $1 + x^n$ без залишку.

З теореми випливає, що для отримання породжувального многочлена g(x) нам необхідно розкласти на множники $x^n + 1$ і виділити многочлен такого ступеня, який відповідає довжині кодового слова.

У табл. 1 наведено розкладання полінома $x^n + 1$ на многочлени для значень $n \leq 31$. Таблиця дозволяє вибирати породжувальні поліноми g(x) в залежності від числа n символів в коді і n_i інформаційних символів, оскільки ступінь g(x) дорівнює n_k.

Таблиця 1. Розкладання полінома x^n+1 на дільники

n	**Дільники полінома** x^n+1
7	$(x+1)(x^3+x+1)(x^3+x^2+1)$
9	$(x+1)(x^2+x+1)(x^6+x^3+1)$
15	$(x+1)(x^2+x+1)(x^4+x+1)(x^4+x^3+1)(x^4+x^3+x^2+x+1)$
17	$(x+1)(x^8+x^5+x^4+x^3+x+1)(x^8+x^7+x^6+x^4+x^2+x+1)$
21	$(x+1)(x^2+x+1)(x^3+x+1)(x^3+x^2+1)(x^6+x^4+x^2+x+1)(x^6+x^5+x^4+x^2+1)$
23	$(x+1)(x^{11}+x^{10}+x^6+x^5+x^4+x^2+1)(x^{11}+x^9+x^7+x^6+x^5+x+1)$
25	$(x+1)(x^4+x^3+x^2+x+1)(x^{20}+x^{15}+x^{10}+x^5+1)$
27	$(x+1)(x^2+x+1)(x^6+x^3+1)(x^{18}+x^9+1)$
31	$(x+1)(x^5+x^3+1)(x^5+x^2+1)(x^5+x^4+x^3+x^2+1)(x^5+x^4+x^3+x+1)\times$ $\times(x^5+x^4+x^2+x+1)(x^5+x^3+x^2+x+1)$

Будь-який дільник полінома x^n+1 або будь-який добуток може бути використаний як породжувальний поліном.. Наприклад, при $n=15$ $g(x)=x^4+x+1$ та $g(x)=x^4+x^3+1$ породжують код (15,11,3) – код Хеммінга з кодовою відстанню $d_0=3$; код (15,7,5) породжується поліномом $g(x)=(x^4+x+1)(x^4+x^3+x^2+x+1)$; код (15,5,7) –

поліномом $g(x)=(x^2+x+1)(x^4+x+1)(x^4+x^3+x^2+x+1)$; код (15,4,8) – поліномом $g(x)=(x+1)(x^2+x+1)(x^4+x+1)(x^4+x^3+x^2+x+1)$ і т.д.

Поліноміальний код (n,n_i,d_0) - це множина всіх многочленів ступеню n-1 або нижче, що діляться на g(x), обто $S(x)=g(x)A(x)$.

Для $n=3,5,11,13,19,29$ розкладання має вигляд: $x^n+1=(x+1)(x^{n-1}+x^{n-2}+\ldots+x+1)$, що визначає лише два типи кодів цієї довжини.

1. Коди $(n,n-1,2)$ з простою перевіркою на парність, кодовою відстанню $d_0=2$ і породжувальним поліномом $g(x)=x+1$, що дозволяє лише знаходити помилки непарної кратності.

2. Коди $(n,1,n)$ непарної довжини з повторенням, $d_0=n$, $g(x)=x^{n-1}+x^{n-2}+\ldots+x+1$.

У таблиці відсутні розкладання для парних n. Але оскільки $(x^m+1)(x^m+1)=x^{2m}+x^m+x^m+1=(x^{2m}+1)$, то при парному n поліном $(x^n+1)=\left(x^{n/2}+1\right)\left(x^{n/2}+1\right)$ і може бути зведений до добутку поліномів непарних ступенів. Наприклад, $(x^6+1)=(x^3+1)(x^3+1)=(x+1)(x^2+x+1)(x+1)(x^2+x+1)$.

Використання породжувальної або перевірочної матриці. Породжувальна матриця G має n_i рядків і n стовпців, містить n_i базисних лінійно незалежних кодових комбінацій. Найбільш зручна для користування канонічна форма породжувальної матриці. Її рядки в своїй інформаційній частині утворюють квадратну n_i x n_i одиничну матрицю:

$$G=\begin{vmatrix} S_{б1} \\ S_{б2} \\ \ldots \\ S_{бk} \end{vmatrix}=\begin{vmatrix} 1 & 0 & \ldots & 0 & b_{11} & b_{12} & \ldots & b_{1n-k} \\ 0 & 1 & \ldots & 0 & b_{21} & b_{22} & \ldots & b_{2n-k} \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & 1 & b_{k1} & b_{k2} & \ldots & b_{kn-k} \end{vmatrix}$$

Значення додаткових розрядів першого рядка $b_{11},b_{12},\ldots, b_{1nk}$ визначаються за залишками від ділення першого рядка одиничної інформаційної матриці

(10…0), доповненого праворуч n_k нулями, на утворюючий поліном g(x). Додаткові розряди другого рядка $b_{21}, b_{22}, \ldots, b_{2nk}$ визначаються за залишками від ділення другого рядка одиничної інформаційної матриці (010…0), доповненого праворуч n_k нулями, на утворюючий поліном g(x), і т.д.

Утворююча матриця G дає n_i комбінацій циклічного коду. Будь-яка інша комбінація може бути отримана як урівноважена (з вагами a_i від джерела повідомлень) сума рядків породжувальної матриці:

$$S = a_1 S_{\delta 1} + a_2 S_{\delta 2} + \ldots + a_k S_{\delta k},$$

Що відповідає добутку матриць

$$S = AG.$$

Так, для двійкового коду (7, 4, 3) n=7, n_i=4, n_k=3. Отже, необхідно в якості утворюючого обрати поліном g(x) степеня n_k=3. З таблиці незвідних поліномів обираємо $g(x)=x^3+x+1$, або у двійковій формі 1011.

Далі доповнюємо трьома (n_k=3) нулями кожний рядок одиничної інформаційної матриці, і виконуємо ділення кожного рядка на утворюючу комбінацію 1011:

Отже, породжувальна матриця має вигляд:

$$G_{7\times 4} = \begin{vmatrix} 1000101 \\ 0100111 \\ 0010110 \\ 0001011 \end{vmatrix}.$$

Перевірочна матриця H будується шляхом транспонування правої частини утворюючої матриці G розмірністю n x n_k та дописування до неї праворуч одиничної матриці розмірністю n_k.

Для двійкового коду (7, 4, 3) після дописування праворуч одиничної матриці розмірністю 3x3:

$$H = \begin{vmatrix} 1110100 \\ 0111010 \\ 1101001 \end{vmatrix}.$$

Алгоритм декодування.

1. Прийнята комбінація r(x) ділиться на утворюючий поліном g(x). Якщо прийнята комбінація не містить помилок, тобто співпадає з надісланою інформаційною комбінацією a(x), а отже, r(x)=a(x), то залишок від ділення буде нульовим.

2. Підраховуємо вагу залишку (кількість одиниць) v. Якщо v≤s, де s – припустима кількість помилок, які можуть бути виправлені (для коду (7, 4, 3) кількість помилок, що виправляються, дорівнює s=1), то прийняту комбінацію додають за модулем 2 з отриманим залишком. Сума дає виправлену комбінацію.

3. Виконують циклічний зсув прийнятої комбінації ліворуч на 1 розряд. Комбінацію, отриману в результаті цього зсуву, ділять на утворюючий поліном g(x). Якщо в результаті цього повторного ділення v≤s, то ділене додають із залишком.

4. Виконують циклічний зсув праворуч на 1 розряд. Отримана комбінація вже не містить помилок. Якщо після першого циклічного зсуву та наступного ділення залишок має вагу v>s, то

5. Повторюють операцію п.3 доти, доки не буде виконана умова v≤s. Тоді комбінацію, отриману в результаті останнього циклічного зсуву, додають до залишку від ділення цієї комбінації на утворюючий поліном.

6. Виконують циклічний зсув праворуч на стільки розрядів, на скільки була зсунута остання комбінація відносно прийнятої. В результаті отримується правильна комбінація.

Лекція 13. Кореляція. Застосування кореляції

Часто необхідно визначити степінь незалежності одного процесу від іншого або встановити подібність одного набору даних до іншого. Іншими словами, потрібно визначити кореляцію процесів або даних у вигляді величини, яку можна описати математично та виміряти.

Така задача виникає, наприклад, при розборці систем комп'ютерного зору та при обробці зображень, отриманих з супутника, при визначенні місцезнаходження (пеленгації) об'єктів, коли порівнюються передані та відбиті сигнали.

Кореляція є також невід'ємною частинкою операції згортки, тобто для обчислення кореляції та згортки можна використовувати ті самі алгоритми, тільки при згортці одна з послідовностей є оберненою у часі.

Розглянемо 2 послідовності даних, що складаються зі значень двох сигналів. Якщо обидва сигнали схоже змінюються при переході між значеннями, то міру їх кореляції можна визначити як суму добутків відповідних пар значень.

Якщо взяти дві незалежних випадкових послідовності, то сума добутків прямує до нуля при збільшенні кількості пар значень. Це пояснюється тим, що всі числа (додатні чи від'ємні) є рівномірними, тому добутки компенсуються при додаванні. Тому кореляції немає.

Від'ємна сума вказує на від'ємну кореляцію, тобто збільшення одного сигналу веде до зменшення другого.

Взаємна кореляція $r_{12}(n)$ двох послідовностей даних $x_1(n)$ та $x_2(n)$, які містять по N значень, записується як:

$$r_{12}(n) = \sum_{n=0}^{N-1} x_1(n)\ x_2(n)\ .$$

Результат в цьому випадку залежить від кількості значень N. Щоб це виправити, результат нормується на N. Отже:

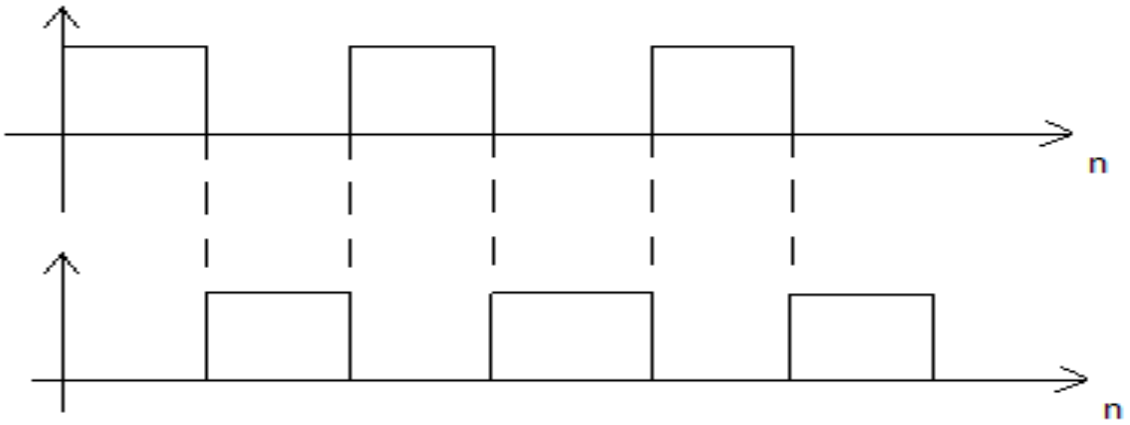
$$r_{12}(n) = \frac{1}{N} \sum_{n=0}^{N-1} x_1(n)\ x_2(n)$$

Приклад 1.

n	0	1	2	3	4	5	6	7	8
$X_1(n)$	4	2	-1	3	-2	-6	-5	4	5
$X_2(n)$	4	1	3	7	4	-2	-8	-2	1

$$r_{12}(n) = \frac{1}{9} \sum_{n=0} x_1(n)\ x_2(n)\ = \frac{1}{9}\ [4(-4) + 2 * 1 - 1 * 3 + 3 * 7 - 2 * 6 * 2 + 5 * 8 - 4 * 2 + 5 * 1] = 5$$

Проте таке визначення теж потребує модифікації, оскільки в деяких випадках обчислена таким чином кореляція може бути мінімальною, хоча дві послідовності корелюють на 100%. Це може бути, наприклад, коли два сигнали йдуть не в фазі.



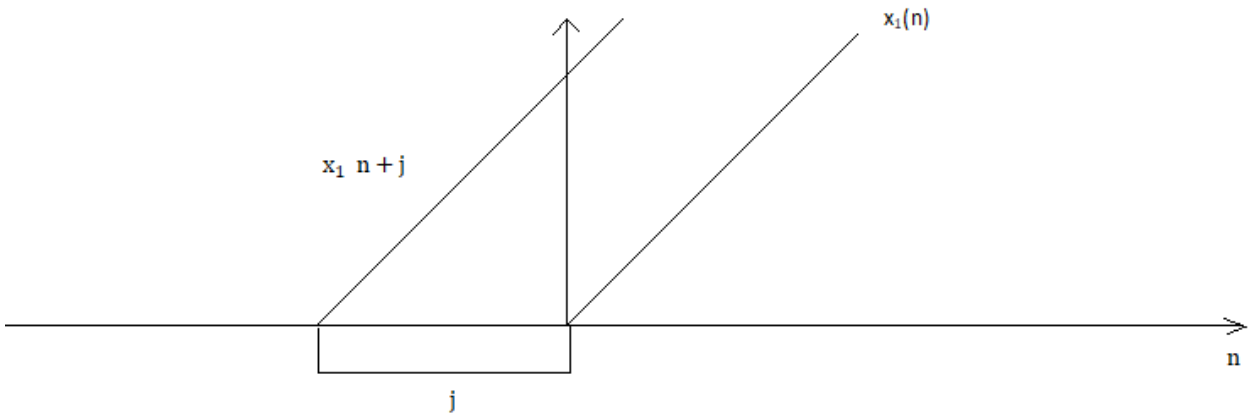
Видно, що в цьому випадку кожний добуток дорівнює нулю, отже, кореляція нульова, оскільки одне зі значень $x_1(n)$ або $x_2(n)$ завжди дорівнюють нулю. При цьому очевидно, що сигнали сильно корелюють, хоча й ідуть не в фазі. Це може бути, наприклад, при запізненні.

Щоб вирівняти зсув фаз, треба зсунути, наприклад, $x_2(n)$ вліво. Це еквівалентно заміні $x_2(n)$ на $X_2(n+j)$, де j – величина затримки. Альтернатива – зсув $x_1(n)$ вправо. В результаті:

$$r_{12}(j) = \frac{1}{N}\sum_{n=0}^{N-1} x_1(n)\, x_2(n+j) = r_{12}(-j) = \frac{1}{N}\sum_{n=0}^{N-1} x_2(n)\, x_1(n-j)\,.$$

На практиці фазовий зсув, як правило, невідомий, так що кореляцію треба знаходити для деяких різних затримок, щоб встановити таке значення кореляції, яке потім вважається вірним.

Сигнал $x_1(n)$, зсунутий на j значень вліво:



Приклад 2.

Розглянемо взаємну кореляцію сигналів $x_1(n)$ та $x_2(n)$ з прикладу 1, із затримкою 3, тобто $r_{12}(3)$. Отже, використовуємо послідовності:

n	0	1	2	3	4	5	6	7	8
$X_2(n)$	4	2	-1	3	-2	-6	-5	4	5
$X_2(n+3)$	7	4	-2	-8	-2	1	-	-	-

$$r_{12}(3) = \frac{1}{9}\,[4*7 + 2*4 + 1*2 - 3*8 + 2*2 - 6*1] = 1.33$$

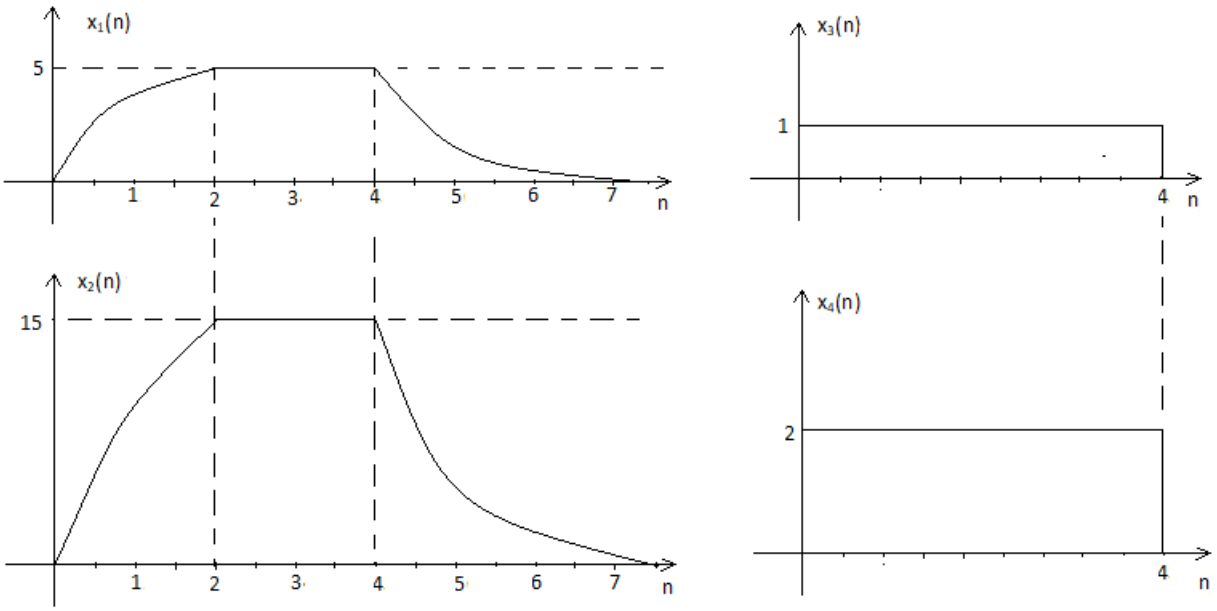
Кореляція зменшилась в результаті крайового ефекту. Оскільки зменшилась кількість пар значень з 9 до 6. Щоб цього уникнути, можна зробити довжину однієї з послідовностей вдвічі більше за довжину іншої.

Для непреривних функцій $x_1(t)$ та $x_2(t)$ кінцевої довжини T кореляція при заримці на $\boldsymbol{\tau}$:

$$r_{12}(\boldsymbol{\tau}) = \frac{1}{T}\int_0^T x_1(t)\, x_2(t+\boldsymbol{\tau})dt.$$

Часто кореляцію вимірюють у фіксованому масштабі між -1 та +1. Для цього треба нормувати дані.

Наприклад, розглянемо дві пари сигналів $x_1(n)$ та $x_2(n)$ та $x_3(n)$, $x_4(n)$



n	0	1	2	3	4	5	6	7	8
$x_1(n)$	0	3	5	5	5	2	0.5	0.25	0
$X_2(n)$	0	9	15	15	15	6	1.5	0.75	0
$X_3(n)$	1	1	1	1	1	0	0	0	0
$X_4(n)$	2	2	2	2	2	0	0	0	0
	(3)	(3)	(3)	(3)	(3)	(0)	(0)	(0)	(0)

Як видно з графіків, сигнали $x_1(n)$ та $x_2(n)$ подібні і відрізняються лише амплітудою. Це ж справедливо для пари сигналів $x_3(n)$ та $x_4(n)$. Проте, значення кореляції для цих пар різні, а має бути однакова кореляція.

Нормуємо взаємну кореляцію $r_{12}(j)$ на коефіцієнт

$$\left[\frac{1}{N}\sum_{n=0}^{N-1} x_1^2(n) * \frac{1}{N}\sum_{n=0}^{N-1} x_2^2(n)\right]^{\frac{1}{2}} = \frac{1}{N}\left[\sum_{n=0}^{N-1} x_1^2(n) * \sum_{n=0}^{N-1} x_2^2(n)\right]^{\frac{1}{2}}.$$

Подібним чином нормуємо також $r_{34}(j)$. В результаті нормований вираз для $r_{12}(j)$ має вигляд:

$$\rho_{12}(j) = \frac{r_{12}(j)}{\frac{1}{N}[\sum_{n=0}^{N-1} x_1^2(n) * \sum_{n=0}^{N-1} x_2^2(n)]^{\frac{1}{2}}},$$

$\rho_{12}(j)$ – це коефіцієнт взаємної кореляції. Його значення завжди лежить між – 1 та +1 , причому «+1» означає 100% кореляцію в прямому сенсі, «-1» - 100% кореляцію в протилежному сенсі, наприклад, для сигналів у протифазі. Значення «0» вказує на нульову кореляцію. Це значить, що сигнали абсолютно незалежні. Малі близькі до нуля значення вказують на малу кореляцію.

Нормувальний коефіцієнт для $r_{12}(j) = \frac{1}{9}[\sum_{n=0}^{N-1} x_1^2(n) * \sum_{n=0}^{N-1} x_2^2(n)]^{\frac{1}{2}} =$

$\frac{1}{9}[(9 + 2.5 * 3 + 4 + 0.25 + 0.0625) * (81 + 225 * 3 + 36 + 2.25 + 0.0625)]^{\frac{1}{2}} =$ 29.4375.

Для $r_{34}(j)$: $\frac{1}{9}[\sum_{n=0}^{N-1} x_3^2(n) * \sum_{n=0}^{N-1} x_4^2(n)]^{\frac{1}{2}} = \frac{1}{9}[(1 + 1 + 1 + 1 + 1)(4 + 4 + 4 + 4 + 4)]^{\frac{1}{2}} = 1.11.$

Якщо $x_4(n)$ в 3 рази більший за $x_3(n)$:

$$\frac{1}{9}[(1 + 1 + 1 + 1 + 1)(9 * 5)]^{\frac{1}{2}} = \frac{5}{3}$$

Знайдемо коефіцієнти взаємної кореляції:

$$\rho_{12}(j) = \frac{r_{12}(j)}{\frac{1}{9}[\sum_{n=0}^{N-1} x_1^2(n) * \sum_{n=0}^{N-1} x_2^2(n)]^{\frac{1}{2}}} = \frac{r_{12}(j)}{29.4375}$$

$r_{12}(1) = \frac{1}{9} \sum_{n=0}^{8} x_1(n)\ x_2(n+1) = \frac{1}{9}(3 * 15 + 5 * 15 + 5 * 15 + 5 * 6 + 2 * 1.5 + 0.5 * 0.75) = 25.375$

$\rho_{12}(1) = \frac{25.375}{29.375} = 0.862$

$r_{12}(0) = \frac{1}{9}(3 * 9 + 5 * 15 * 3 + 2 * 6 + 0.5 * 1.5 + 0.25 * 0.75) = 29.4375$

$\rho_{12}(0) = \frac{29.4375}{29.4375} = 1$

Для сигналів $x_3(n)$ та $x_4(n)$:

$$\rho_{34}(j) = \frac{r_{34}(j)}{\frac{1}{9}\left[\sum_{n=0}^{8} x_3^2(n) * \sum_{n=0}^{8} x_4^2(n)\right]^{\frac{1}{2}}} = \frac{r_{34}(j)}{1.11}$$

При j = 0 :

$$r_{34}(0) = \frac{1}{9}\sum_{n=0}^{8} x_3(n)\ x_4(n) = \frac{1}{9}\ [1 * 2 * 5] = 1.11$$

$$\rho_{34}(0) = \frac{1.11}{1.11} = 1$$

При j = 1 :

$$r_{34}(1) = \frac{1}{9}\sum_{n=0}^{8} x_3(n)\ x_4(n+1) = \frac{1}{9}\ [1 * 2 * 4] = 0.889$$

$$\rho_{34}(1) = \frac{0.889}{1.11} = 0.8$$

Бачимо, що $\rho_{12}(0) = \rho_{34}(0) = 1$, а $\rho_{12}(1)$ приблизно рівна $\rho_{34}(1)$, тобто процес нормування дозволяє незалежно порівнювати взаємні кореляції абсолютних значень даних.

Якщо порівнювати $x_1(n)$ з $x_3(n)$, а $x_2(n)$ з $x_4(n)$, то отримаємо :

$$\rho_{13}(1) = 0.57\ , \quad \rho_{24}(1) = 0.58$$

Розглянемо частковий випадок $x_1(n) = x_2(n)$, тобто знайдемо кореляцію сигнала з самим собою – автокореляцію. Автокореляційна функція визначається як :

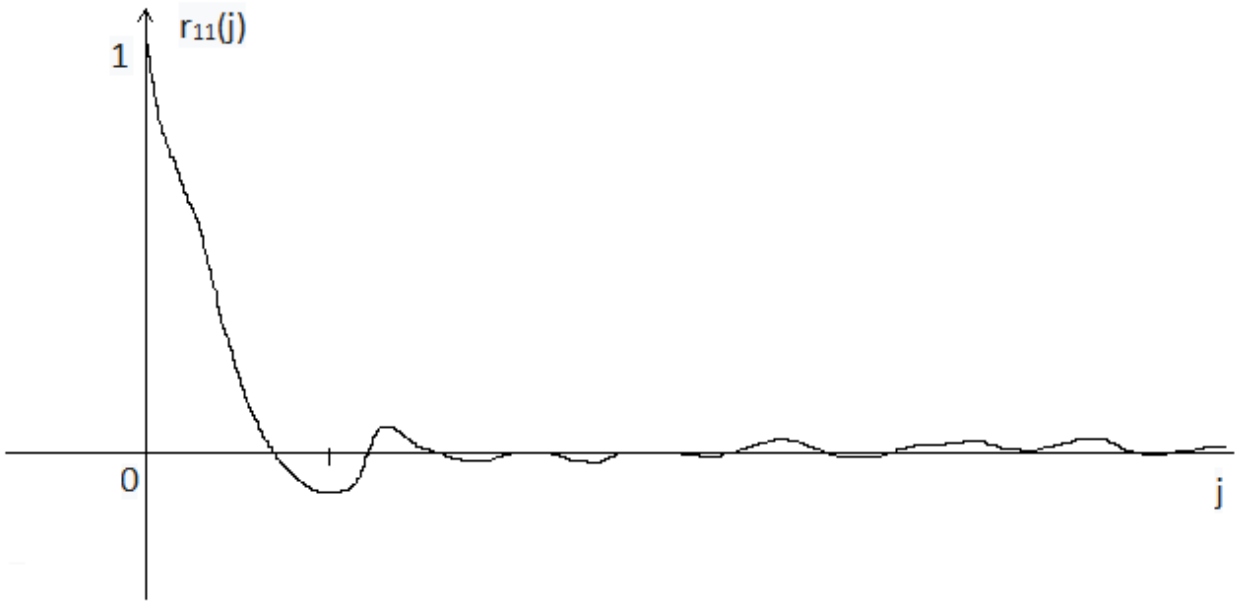
$$r_{11}(j) = \frac{1}{N}\sum_{n=0}^{N-1} x_1(n)\ x_1(n+j)$$

Властивість автокореляційної функції :

$$r_{11}(0) = \frac{1}{N}\ \sum_{n=0}^{N-1} x_1^2(n) = S,$$

де S – нормована енергія сигналу. Якщо сигнал випадковий, наприклад, білий гаусівський шум, його автокореляція буде максимальною при нульовій затримці і зменшується до випадкових флуктуацій малої амплітуди біля нуля для затримок > 1. Крім того, завжди

$$r_{11}(0) > r_{11}(j).$$



При визначенні кореляції періодичних послідовностей нерівної довжини треба користуватися наступним правилом.

Якщо є дві періодичні послідовності з періодами N, та N_2 відповідно, то для знаходження взаємної кореляції треба доповнити кожну з них нулями до значення N_1+ N_2-1 (тобто додати N_2-1 нулів до 1-ої послідовності з періодом N_1 та N_1-1 нулів до 2-ої послідовності з періодом N_2) і проводити розрахунок $r_{12}(j)$ для j = 0 , N_1+ N_2-1. Кореляція в одному випадку є періодичною з періодом N_1+ N_2-1

Наприклад, для знаходження взаємної кореляції двох періодичних послідовностей {4,3,1,6} та {5,2,3}, N_1 = 4, N_2 = 3

Записуємо $N_1 + N_2 - 1 = 4 + 3 - 1 = 6$

Послідовність	Затримка, j	$r_{12}(j)$
4 3 1 6 0 0 5 2 3 0 0 0	0	29/6
2 3 0 0 0 5	1	17/6
3 0 0 0 5 2	2	12/6
0 0 0 5 2 3	3	30/6
0 0 5 2 3 0	4	17/6
0 5 2 3 0 0	5	35/6
5 2 3 0 0 0	6	29/6

Взаємна кореляція :

$$r_{12}(j) = \frac{1}{6}\{29, 17, 12, 30, 17, 35\}$$

Взаємна кореляція дозволяє виявити властивості сигналів, знижуючи їх шумову складову. Якщо є 2 зашумлених сигналами $\{S_1(t)+q_1(t)\}$ та $\{S_2(t)+q_2(t)\}$, де $S_1(t)$ та $S_2(t)$ – корисні сигнали, а $q_1(t)$ та $q_2(t)$ – шуми, то

$$r_{12}(j) = r_{s1s2}(j) + r_{s1q2}(j) + r_{s2q1}(j) + r_{q1q2}(j),$$

причому останні 3 складові прямують до нуля при збільшенні j.

При збільшенні N, $r_{12}(j) \rightarrow r_{s1s2}(j)$, тобто знижується вплив шумів.

Застосування кореляції. Можна показати, що

$$F\{r_{11}(\boldsymbol{\tau})\} = G_E(f),$$

де $F\{r_{11}(\boldsymbol{\tau})\}$ – Фур'є зображення автокореляційної функції, $G_E(f)$ – спектральна густина енергії сигналу.

Крім того, $r_{11}(0) = S$ – загальна енергія сигналу.

Якщо є 2 різних сигнали $x_1(n)$ та $x_2(n)$, а їх сума $x(n) = x_1(n) + x_2(n)$, то автокореляція суми :

$r_x(0) = E_x = \frac{1}{N} \sum_{n=0}^{N-1} x^2(n) = \frac{1}{N} \sum_{n=0}^{N-1} [x_1(n) + x_2(n)]^2 = \frac{1}{N} \sum_{n=0}^{N-1} [x_1^2(n) + x_2^2 + 2x_1(n) * x_2(n)]^2 = \frac{1}{N} \sum_{n=0}^{N-1} x_1^2(n) + \frac{1}{N} \sum_{n=0}^{N-1} x_2^2(n) + 2\frac{1}{N} \sum_{n=0}^{N-1} x_1(n) * x_2(n)$

Отже, $E_x = r_{x1}(0) + r_{x2}(0) + 2\, r_{x1x2}(0)$

Інакше можна записати : $E_x = E_{x1} + E_{x2} + 2\, r_{x1x2}(0)$

Таким чином, енергія сумарного сигналу дорівнює сумі енергій його компонентів плюс $2\, r_{x1x2}(0)$, тобто плюс подвоєна функція взаємної кореляція компонентів з нульовою затримкою.

Якщо $x_1(n)$ та $x_2(n)$ не корелюють, то загальна енергія є просто сумою енергій компонентів.

Якщо сигнали $x_1(n)$ та $x_2(n)$ зашумлені, тобто

$x_1(n) = x_1'(n) + q_1(n)$, $x_2(n) = x_2'(n) + q_2(n)$, то

$E_x = E_{x1'} + E_{x2'} + E_{q1} + E_{q2} + E_{x1'x2'}(0)$

Лекція 14. Автокореляційна функція

Автокореляційна функція - це характеристика сигналу, яка допомагає знаходити повторювані ділянки сигналу або визначати несучу частоту сигналу, приховану через накладання шуму і коливань на інших частотах. Вона вводиться для кількісної та якісної оцінки збігу двох сигналів $U(t)$ і $U(t-\tau)$ (його зміщеної копії) .Автокореляційна функція характеризує міру залежності між двома сигналами, і дорівнює скалярному добутку цих двох сигналів.

$$B_U(\tau) = \int_{-\infty}^{\infty} U(t)U(t-\tau)dt \qquad (1)$$

Рівність (1) означає згортку двох однакових сигналів, зсунутих за часом один щодо одного на величину, рівну τ. Цю величину можна називати розладом двох сигналів за часом τ.

На рис.1, а зображен пакет, що складається з трьох однакових відеоімпульсів прямокутної форми. Також тут представлена його автокореляційна функція, обчислена за формулою (1) (рис. 1,б).

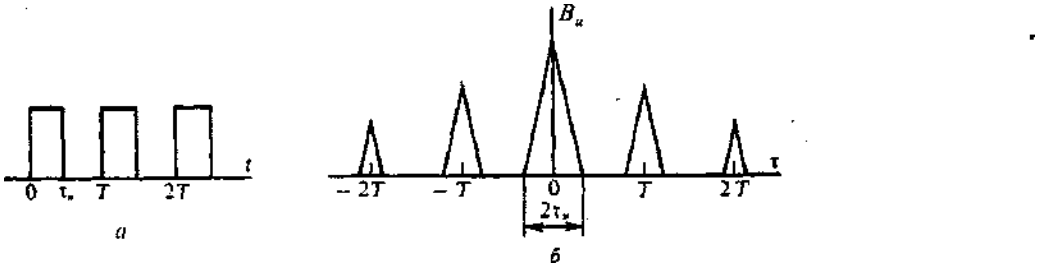


Рис.1. АКФ пакету з трьох однакових відеоімпульсів: а - пакет імпульсів; б - графік АКФ

Добре видно, що максимум АКФ досягається при τ = 0. Однак якщо затримка τ виявляється кратною періоду послідовності (при τ = ± T, ± 2T в нашому випадку), спостерігаються побічні пелюстки АКФ, які можна порівняти за висотою з головною пелюсткою. Тому можна говорити про відому недосконаліть кореляційної структури даного сигналу.

Приведемо формулу до такого вигляду, щоб можна було обчислювати дискретний аналог АКФ стосовно багатопозиційних сигналів. Для цього операцію інтегрування замінюємо сумою, а замість змінної τ використовуємо ціле число n (позитивне чи негативне), тим самим вказуючи, на скільки позицій зрушена копія щодо вихідного сигналу.

Запишемо дискретну АКФ в такому вигляді:

$$\hat{\hat{B}}_u(n) = \sum_{j=-\infty}^{\infty} u_j u_{j-n} \tag{2}$$

З даного виразу видно, що дискретна АКФ парна:

$$\hat{\hat{B}}_u(n) = \hat{\hat{B}}_u(-n) \tag{3}$$

При нульовому зсуві ця АКФ визначає енергію дискретного сигналу:

$$\hat{\hat{B}}_u(0) = \sum_{j=-\infty}^{\infty} u_j^2 = E_u \tag{4}$$

Властивості автокореляційної функції.

1. Якщо $\tau = 0$, то $B_U(0) = \int_{-\infty}^{\infty} U(t)U(t-0)dt = \int_{-\infty}^{\infty} U^2(t)dt = E_U$.

Автокореляційна функція при значенні τ = 0 приймає своє максимальне і позитивне значення, рівне енергії самого сигналу.

2.Автокореляційна функція $B_U(\tau)$ є парною функцією.

$$B_U(\tau) = B_U(-\tau)$$

Цей вираз легко довести, зробивши таку заміну виду $x = t - \tau$. Тоді:

$$B_U(\tau) = \int_{-\infty}^{\infty} U(t)U(t-\tau)dt = \int_{-\infty}^{\infty} U(x-\tau)U(x)dx$$

3. При будь-якому τ модуль автокореляційної функції не перевищує енергії сигналу: $|B_U(\tau)| \le |B_U(0)| = E_U$.

Цей вираз випливає з нерівності Коші: $(U, U\tau) \le |U| \cdot |U\tau| = E_U$.

Ця властивість ще раз показує, що зі збільшенням часу (τ) АКФ сигналу повинна зменшуватися. Її максимум знаходиться в точці $\tau = 0$ і дорівнює енергії сигналу.

4.Дана властивість випливає з попередніх. Якщо проаналізувати АКФ щодо енергії сигналу $B_N(\tau) = \frac{B_U(\tau)}{E_U}$, то максимальне значення АКФ дорівнює ± 1, а при $\boldsymbol{\tau > t_u}$ характеристика $\boldsymbol{\tau > t_u}$ буде приймати нульове значення.

Межі зміни АКФ від -1, до1 і зі зростанням τ вона затухає.

Зв'язок між енергетичним спектром сигналу і його автокореляційною функцією.

$$W_u(\omega) = \int_{-\infty}^{+\infty} B_U(\tau) e^{-j\omega\tau} d\tau .$$

Енергетичний спектр являє собою пряме перетворення Фур'є від автокореляційної функції сигналу.

Кореляційні властивості кодових послідовностей, використовуваних в широкосмугових системах, залежать від типу кодової послідовності, її довжини, частоти проходження її символів і від її посимвольної структури.

Вивчення АКФ грає важливу роль при виборі кодових послідовностей з точки зору найменшої імовірності встановлення помилкової синхронізації.

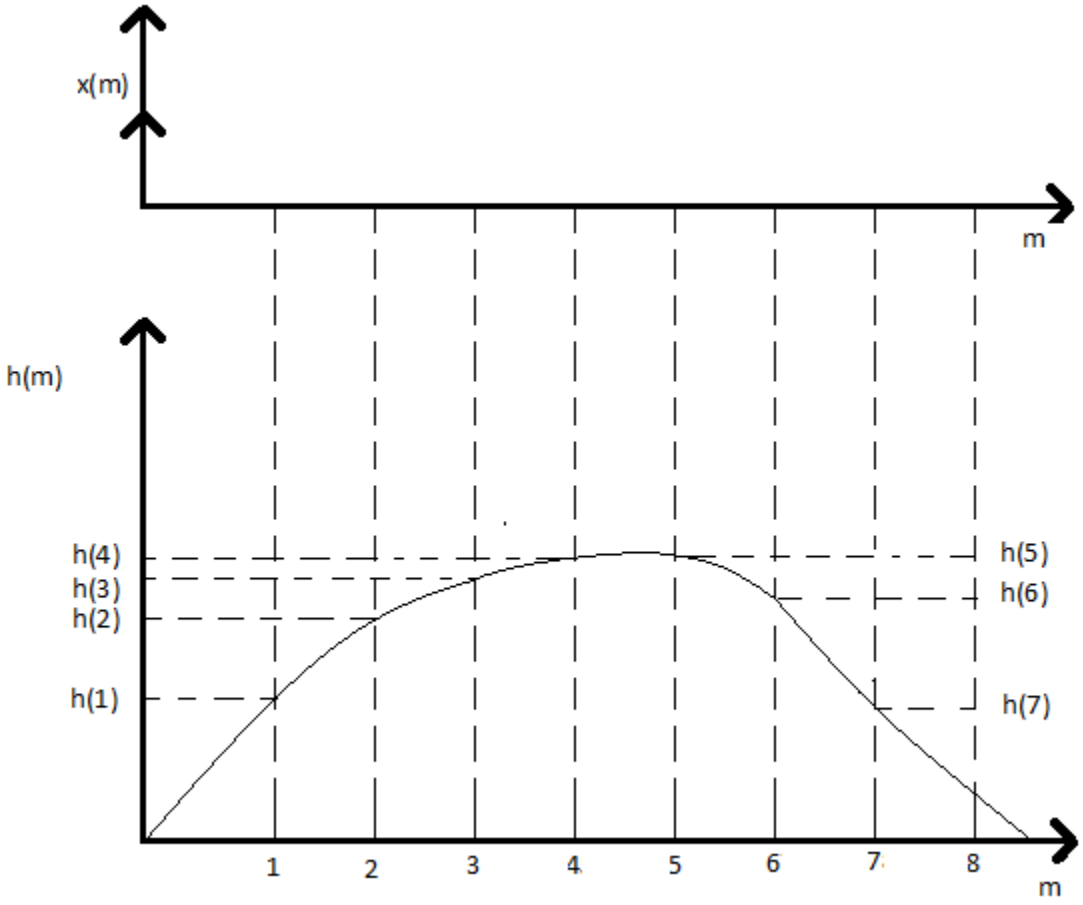
Лекція 15. Дискретна згортка

Згортка описує, як вихід систем визначається взаємодією входу з самою системою.

Розглянемо систему, на вхід якої подано одиничний імпульс

$$x(m) = \begin{cases} 1, & m = 0 \\ 0, & m \neq 0 \end{cases}$$

На виході спостерігатиметься деякий сигнал h(m)

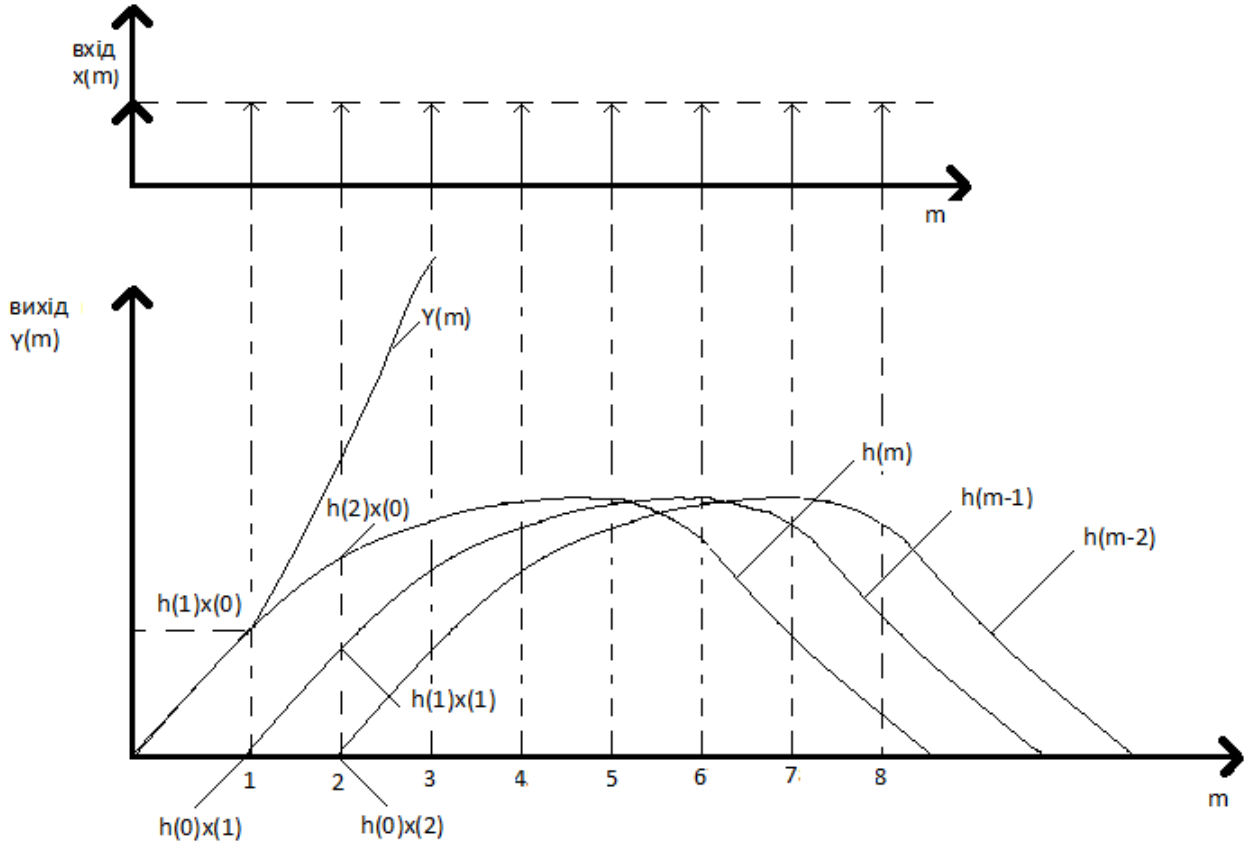


Величина h(m) – вихідний сигнал системи, на вихід якої подано одиничний імпульс

$$x(m) = \begin{cases} 1, & m = 0 \\ 0, & m \neq 0 \end{cases}$$

В момент m = 0, називається імпульсною характеристикою системи – реакція системи на одиничний імпульс.

Розглянемо тепер ситуацію, коли на вхід системи подається послідовність одиничних імпульсів. На виході буде спостерігатися множина сигналів, кожний з яких являє собою реакцію на одиничний відповідний імпульс. Якщо система лінійна, то вихід можна записати як суму реакцій на окремі одиничні імпульси.



В момент часу m = 0 на виході сигнал дорівнює y(0) = h(0) x(0). В дискретний момент m = 1 вихід дорівнює сумі h(0) x(0) (запізнений вплив вхідного імпульсу x(0)). Отже,

$$y(1) = h(1)\,x(0) + h(0)\,x(1) \qquad (1)$$

Наступні дискретні значення вихідного сигналу записуються так :

y(2) = h(2)x(0) + h(1)x(1) + h(0) x(2)

y(3) = h(3)x(0) + h(2)x(1) + h(1) x(2) + h(0) x(3)

…

y(n) = h(n)x(0) + h(n-1)x(1) + h(n-2) x(2) +…+ h(0) x(n)

Таким чином, вихідний сигнал дорівнює сумі добутків значень вхідного сигналу на відповідні значення оберненої у часі імпульсної характеристики.

Альтернативний варіант – можна переписати вираз для y(n) наступним чином :

$$y(n) = h(n)x(n) + h(1)x(n-1) +\ldots+ h(n)\, x(0). \quad (2)$$

Тоді можна розглядати вихідний сигнал як суму добутків значень імпульсної характеристики на відповідні значення оберненого у часі вхідного сигналу.

Таким чином, згортку можна розглядати як взаємну кореляцію однієї послідовності та оберненої у часі другої послідовності. Рівняння (1) та (2) можна записати компактно :

$$y(n) = \sum_{m=0}^{n} h(n-m) * x(m)$$

$$y(n) = \sum_{m=0}^{n} h(m) * x(n-m) \quad (3)$$

Функція y(n) називається згорткою входу x(m) з імпульсною характеристикою h(m).

Рівняння можна поширити на сигнали нескінченної тривалості :

$y(n) = \sum_{m=0}^{n} h(n-m) * x(m)$ = x(n) * h(n)

$y(n) = \sum_{m=0}^{n} h(m) * x(n-m)$ = h(n) * x(n), де * - символ згортки.

Якщо на вході сигнал неперервний, то сума замінюється на інтеграл :

$$y(t) = \int_{-\infty}^{\infty} x(\tau)\ h(t-\tau)\, d\tau = \int_{-\infty}^{\infty} h(\tau)\, x(t-\tau) d\tau \quad (4)$$

Цей інтеграл називається інтегралом згортки.

Описана згортка (дискретна чи неперервна) виконується у часі. При переході до спектральної області вихідний сигнал системи y(t) перетворюється на зображення y(p), причому

$$Y(p) = H(p) * X(p),$$

де H(p), X(p) – відповідно Фур'є зображення імпульсної характеристики та вхідного сигналу. Тоді

$$x(t)*h(t) = F^{-1}\{H(p) * X(p)\}$$

Таким чином, згортка у часовій області еквівалентна множенню у частотній. Можна показати і справедливість дуального твердження:

$$X(p) * H(p) = F\ \{\ x(t) * h(t)\ \}$$

Тобто, згортка у частотній області еквівалентна множенню у часовій.

Властивості згортки.

1. Комутативність

$$x_1(t) * x_2(t) = x_2(t) * x_1(t)$$

Визначимо, що даний вираз є евівалентним наступному:

$$\int_{-\infty}^{\infty} x_1(\tau)\, x_2(t-\tau)\mathrm{d}\tau = \int_{-\infty}^{\infty} x_2(\tau)\, x_1(t-\tau)\mathrm{d}\tau$$

2. Дистрибутивність

$$x_1(t) * [\ x_2(t) + x_3(t)\ ] = x_1(t) * x_2(t) + x_1(t) * x_2(t)$$

3. Асоциативність

$$x_1(t) * [\ x_2(t) + x_3(t)\ ] = [x_1(t) * x_2(t)] * x_3(t)$$

Лекція 16. Ідентифікація систем

Терміном ідентифікація позначають визначення імпульсної характеристики h(n). Якщо на вхід системи подати тестовий сигнал x(n), виміряти реакцію (вихідний сигнал) y(n), то можна визначити h(n).

З рівняння (2) $y(n) = h(n)x(n) + h(1)x(n-1) + \ldots + h(n)\,x(0)$

При n = 0, y(0) = h(0) * x(0), тому

$$h(0) = \frac{y(0)}{x(0)} \qquad (5)$$

Далі розриваючи та перегруповуючи другий з виразів (3), маємо

$$Y(n) = h(n) * x(0) + \sum_{m=0}^{n-1} h(m) * x(n - m),\ n \geq 1$$

Тому,

$$h(n) = \frac{y(n) - \sum_{m=0}^{n-1} h(m) * x(n - m)}{x(0)},\ n \geq 1,\ x(0) \neq 0 \qquad (6)$$

Приклад 1. Тестовий сигнал x(n) = {1;1;1} подається на вхід системи з невідомою імпульсною характеристикою h(n). На виході спостерігається послідовність y(n) = {1;4;8;10;8;4;1}. Провести ідентифікацію системи.

З рівняння (5)

$$h(0) = \frac{y(0)}{x(0)} = \frac{1}{1} = 1$$

Використовуючи (6), знаходимо :

$$h(n) = \frac{y(n) - \sum_{m=0}^{n-1} h(m) * x(n - m)}{x(0)}$$

n =1 :

$$h(1) = \frac{y(1) - \sum_{m=0}^{0} h(m) * x(n - m)}{x(0)} = \frac{y(1) - h(0) * x(1)}{x(0)} = \frac{4 - 1 * 1}{1} = 3$$

n = 2 :

$$h(2) = \frac{y(2) - \sum_{m=0}^{1} h(m) * x(2 - m)}{x(0)} = \frac{y(2) - h(0) * x(2) - h(1)x(1)}{x(0)} = \frac{8 - 1 * 1 - 3 * 1}{1} = 4$$

n =3 :

$$h(3) = \frac{y(3) - \sum_{m=0}^{2} h(m) * x(3 - m)}{x(0)} = \frac{y(3) - h(0) * x(3) - h(1)x(2) - h(2)x(1)}{x(0)} = \frac{10 - 1 * 0 - 3 * 1 - 4 * 1}{1} = 4$$

n =4 :

$$h(4) = \frac{y(4) - \sum_{m=0}^{3} h(m) * x(4-m)}{x(0)} = \frac{y(4) - h(0) * x(4) - h(1)x(3) - h(2)x(2) - h(3)x(1)}{x(0)} =$$

$$\frac{8 - 1*0 - 3*0 - 4*1 - 3*1}{1} = 1$$

n =5 :

$$h(5) = \frac{y(5) - \sum_{m=0}^{4} h(m) * x(5-m)}{x(0)} = \frac{y(5) - h(0) * x(5) - h(1)x(4) - h(2)x(3) - h(3)x(2) - h(4)x(1)}{x(0)} =$$

$$\frac{4 - 1*0 - 3*0 - 4*0 - 3*1 - 1*1}{1} = 0$$

n =6 :

$$h(6) = \frac{y(6) - \sum_{m=0}^{5} h(m) * x(6-m)}{x(0)} =$$

$$\frac{y(6) - h(0) * x(6) - h(1)x(5) - h(2)x(4) - h(3)x(3) - h(4)x(2) - h(5)x(1)}{x(0)}$$

$$= \frac{1 - 1*0 - 3*0 - 4*0 - 3*0 - 1*1 - 0*1}{1} = 0$$

Фактично, h(n) = 0, при n ≥ 5. Отже, h(n) = {1;3;4;3;1}

Обернення згортки. Якщо, відомі імпульсна характеристика та вихід системи, то для визначення входу використовується операція обернена згортці. Ця операція подібна до ідентифікації системи. Розпишемо та переформуємо другий з виразів (3), отримаємо

$$y(n) = h(0)x(0) + \sum_{m=1}^{n} h(m) * x(n - m)$$

При n = 0 y(0) = h(0) x(0), отже,

$$x(0) = \frac{y(0)}{h(0)} \quad (7)$$

Для n ≥ 1

$$x(n) = \frac{y(n) - \sum_{m=1}^{n} h(m) * x(n-m)}{x(0)} \quad (8)$$

Вирази для обернення згортки ідентичні вирази для ідентифікації, тому процедура обернення відбувається аналогічно.

Приклад 2. Для систем прикладу 1 обчислити вхід x(n) по даним h(n) = {1;3;4;3;1} та y(n) = {1;4;8;10;8;4;1}.

З рівняння (7)

$$x(0) = \frac{y(0)}{h(0)} = \frac{1}{1} = 1$$

З рівняння (8) :

n = 1 :

$$x(1) = \frac{y(1) - \sum_{m=1}^{1} h(m) * x(1-m)}{h(0)} = \frac{y(1) - h(1) * x(0)}{h(0)} = \frac{4 - 3*1}{1} = 1$$

n = 2 :

$$x(2) = \frac{y(2) - \sum_{m=1}^{2} h(m) * x(2-m)}{h(0)} = \frac{y(2) - h(1) * x(1) - h(2)x(0)}{h(0)} = \frac{8 - 3*1 - 4*1}{1} = 1$$

n = 3 :

$$x(3) = \frac{y(3) - \sum_{m=1}^{3} h(m) * x(3-m)}{h(0)} = \frac{y(3) - h(1) * x(2) - h(2)x(1) - h(3)x(0)}{h(0)} = \frac{10 - 3*1 - 4*1 - 3*1}{1} = 0$$

Для $n \geq 3$ $x(n) = 0$

Отже, x(n) = {1;1;1}, що узгоджується з даними прикладу 1.

Зв'язок між згорткою та кореляцією. При згортці вихідний сигнал визначається сумою

$$y(n) = \sum_{m=0}^{n} h(m) * x(n-m) = h(0)x(n) + h(1)x(n-1) + \cdots + h(n)\,x(0) \quad (9)$$

Значення функції взаємної кореляції для сигналів h(n) та x(n) для j – ї затримки можна записати наступним чином:

$$r_{hx}(j) = \frac{1}{N} \sum_{n=0}^{n-1} h(n) * x(n+j) = \frac{1}{N} [h(0)x(j) + h(1)x(1+j) + \cdots + h(N-1)x(N-1+j)]$$

Порівняти y(n) та $r_{hx}(j)$ простіше всього, якщо j = 0, тобто при нульовій затримці. В цьому випадку кореляція

$$r_{hx}(0) = \frac{1}{N} \sum_{n=0}^{n-1} h(n) * x(n) = \frac{1}{N} [h(0)x(0) + h(1)x(1) + \cdots + h(N-1)x(N-1)]$$

Порівнюючи вирази (9) та (10), бачимо, що вони мають схожу форму, але послідовність x(n) у кореляційній функції йде у зворотному порядку у порівнянні зі згорткою.

Отже, згортка еквівалентна функції взаємної кореляції двох сигналів, в якій одна з початкових послідовностей обернена у часі, а нормувальний коефіцієнт $\frac{1}{N}$ дорівнює 1.

Це означає, що згортку і кореляцію можна обчислювати за одним і тим самим алгоритмом, просто обертаючи у часі одну з двох послідовностей.

Метод накладання-додавання для обчислення згортки. На практиці сигнали, які підлягають згортці, не завжди мають кінцеву тривалість. Іноді вхідні дані мають нескінченну тривалість або тому, що надходять на вхід системи фактично неперервно або тому, що наявний обсяг пам'яті недостатньо великий, щоб вмістити їх усі. В таких випадках згортку або кореляцію обчислюють поетапно, розділяючи вхідні данні на окремі блоки, використовуючи необхідні обчислення для кожного блоку та потім об'єднуючи результати.

Один із методів, що використовується для цього - це метод накладання - додавання.

Хай послідовність вхідних даних x(n) поділено на сегменти с N_1 точок даних. Припустимо, що потрібно знайти згортку з послідовністю h(n), що складається з N_2 точок даних.

Для отримання правильного результату кожна послідовність повинна мати довжину $N = N_1 + N_2 - 1$.

Будемо розглядати сегменти послідовності x(n) з довжиною N, де останні $N_2 - 1$ точок даних замінені нулями, тобто розширемо початкові N_1 точок до значення N. Позначимо ці сегменти $x_1'(n)$, $x_2'(n)$, …

Послідовність h(n) при цьому доповнимо $N_1 - 1$ нулями до значення N. Позначимо її $h'(n)$.

Після цього знаходимо згортки $x'(n) \cdot h'(n), x'_2(n) \cdot h'(n)$ і тд.

Оскільки останні $N_2 - 1$ точок сегментів послідовності x(n) були замінені нулями, отримані функції будуть хибними у перших $N_2 - 1$ та останніх $N_2 - 1$ точках кожної згортки, але всі ці точки додаються і дають правильний результат згортки.

Отже, спочатку для усунення крайових ефектів до сегментів вхідної послідовності додаються нулі, а потім результати згорток накладаються один на один, і знаходяться їх сума. Звідси витікає і назва методу - «накладання-додавання».

Приклад 3. Знайти за допомогою методу накладання - додавання згортку двох послідовностей:

$$x(n)= \{1,3,2,-3,0,2,-1,0,-2,3,-2,1,\ldots\}$$

$$h(n) = \{1,0,1\}$$

Хай послідовність x(n) розбита на сегменти довжиною $N_1 = 6$. Для другої послідовності $N_2 = 3$. При цьому $N = N_1 + N_2 - 1 = 6 + 3 - 1 = 8$

Доповнюючи h(n) нумерація($N_1 - 1 = 5$ нулів)отримуємо h'(n):

$$h`(n) = \{1,0,1,0,0,0,0,0\}$$

Розглянемо розширені до N=8 сегменти послідовності x(n):

$$x'_1(n) = \{1,3,2,-3,0,2,0,0\}$$

$$x'_2(n) = \{-1,0,-2,3,-2,1,0,0\}$$

Знайдемо складові згорткової суми $x'_1(n) \cdot h'(n)$:

$$y_{10} = h'_o x'_{10} = 1 \cdot 1 = 1$$

$$y_{11} = h'_0 x'_{11} + h'_1 x'_{10} = 1 \cdot 3 + 0 \cdot 1 = 3$$

$$y_{12} = h'_o x'_{12} + h'_1 x'_{11} + h'_2 x'_{10} = 1 \cdot 2 + 0 \cdot 3 + 1 \cdot 1 = 3$$

$$y_{13} = h'_0 x'_{13} + h'_1 x'_{12} + h'_2 x'_{11} + h'_3 x'_{10} = 1 \cdot (-3) + 0 \cdot 2 + 1 \cdot 3 = 3$$

$$y_{14} = h'_o x'_{14} + h'_1 x'_{13} + h'_2 x'_{12} + = 1 \cdot 0 + 0 \cdot (-3) + 1 \cdot 2 = 2$$

$$y_{15} = h'_0 x'_{15} + h'_1 x'_{14} + h'_2 x'_{13} + = 1 \cdot 2 + 0 \cdot 0 + 1 \cdot (-3) = -1$$

$$y_{16} = h'_o x'_{16} + h'_1 x'_{15} + h'_2 x'_{14} + = 1 \cdot 0 + 0 \cdot 2 + 1 \cdot 0 = 0$$

$$y_{17} = h'_o x'_{17} + h'_1 x'_{16} + h'_2 x'_{15} + = 1 \cdot 0 + 0 \cdot 0 + 1 \cdot 2 = 2$$

Складові згорткової суми $x'_2(n) \cdot h'(n)$:

$$y_{20} = h'_o x'_{20} + = 1 \cdot (-1) = 1$$

$$y_{21} = h'_o x'_{21} + h'_1 x'_{20} + = 1 \cdot 0 + 0 \cdot 1 = 0$$

$$y_{22} = h'_o x_{22} + h'_1 x'_{21} + h'_2 x'_{20} + = 1 \cdot (-2) + 0 \cdot 0 + 1 \cdot (-1) = -3$$

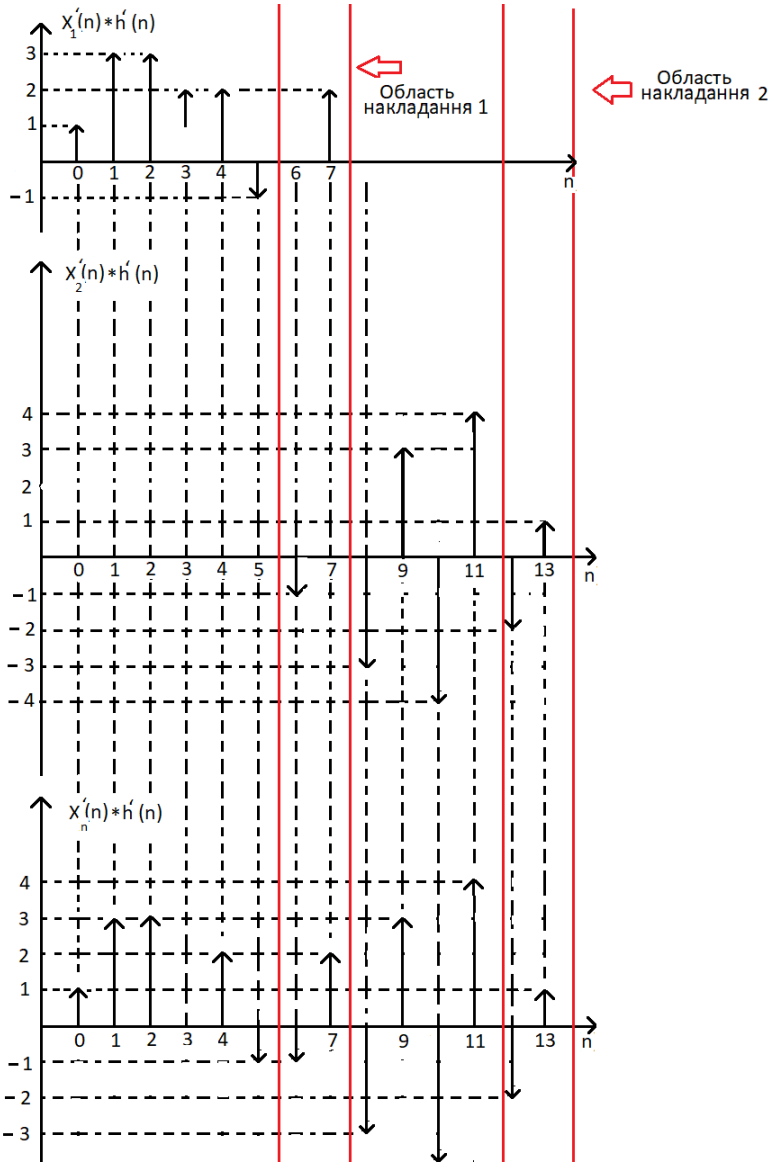
$$y_{23} = h'_o x_{23} + h'_1 x'_{22} + h'_2 x'_{21} + = 1 \cdot 3 + 0 \cdot (-2) + 1 \cdot 0 = 3$$

$$y_{24} = h'_o x_{24} + h'_1 x'_{23} + h'_2 x'_{22} + = 1 \cdot (-2) + 0 \cdot 3 + 1 \cdot (-2) = -4$$

$$y_{25} = h'_o x_{25} + h'_1 x'_{24} + h'_2 x'_{23} + = 1 \cdot 1 + 0 \cdot (-2) + 1 \cdot 3 = 4$$

$$y_{26} = h'_o x_{26} + h'_1 x'_{25} + h'_2 x'_{24} + = 1 \cdot 0 + 0 \cdot 1 + 1 \cdot (-2) = -2$$

$$y_{27} = h'_o x_{27} + h'_1 x'_{26} + h'_2 x'_{25} + = 1 \cdot 0 + 0 \cdot 0 + 1 \cdot 1 = 1$$



Якщо перші N_2 -1 = 2 точки даних $x_2'(n)$ накласти на останні N_2 -1 = 2 точки даних $x_1'(n)$ і додати значення, отримаємо перших 12 точок даних згорткового сигналу:

$$x(n) * h(n) = \{1,3,3,0,2,-1,-1,2,-3,3,-4,4\}$$

Можна показати, що наведений вище результат ідентичний тому, який отримується при прямому обчисленні згортки. Початкова послідовність x(n) містить 12 значень, а h(n) – 3 точки даних. Щоб отримати згортку цих послідовностей, їх треба доповнити нулями, щоб довжина обох склала

N_1+ N_2 - 1 = 12+3-1=14 точок. Тоді маємо :

$$h'(n) = \{1,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0\}$$

$$x'(n) = \{1,3,2,-3,0,2,-1,0,-2,3,-2,1,0,0\}$$

Знайдемо складові згорткової суми $x'(n) * h'(n)$:

$y_0 = h_0'x_0' = 1 * 1 = 1$

$y_1 = h_0'x_1' + h_1'x_0' = 1 * 3 + 0 * 1 = 3$

$y_2 = h_0'x_2' + h_1'x_1' + h_2'x_0' = 1 * 2 + 0 * 3 + 1 * 1 = 3$

$y_3 = h_0'x_3' + h_1'x_2' + h_2'x_1' = 1(-3) + 0(2) + 1*3 = 0$

$y_4 = h_0'x_4' + h_1'x_3' + h_2'x_ч' = 1*0 + 0(-3) + 1*2 = 2$

$y_5 = h_0'x_5' + h_1'x_4' + h_2'x_3' = 1*2 + 0 + 1(-3) = -1$

$y_6 = h_0'x_6' + h_1'x_5' + h_2'x_4' = 1(-1) + 0*2 + 1*0 = -1$

$y_7 = h_0'x_7' + h_1'x_6' + h_2'x_5' = 1*0 + 0\ (-1) + 1*2 = 2$

$y_8 = h_0'x_8' + h_1'x_7' + h_2'x_6' = 1(-2) + 0 + 1(-1) = -3$

$y_9 = h_0'x_9' + h_1'x_8' + h_2'x_7' = 1*3 + 0\ *(-2) + 1*\ 0 = 3$

$y_{10} = h_0'x_{10}' + h_1'x_9' + h_2'x_8' = 1(-2) + 0*3 + 1(-2) = -4$

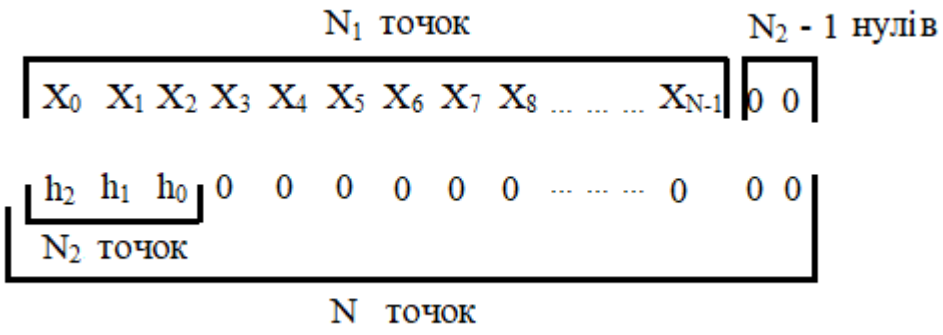
$y_{11} = h_0'x_{11}' + h_1'x_{10}' + h_2'x_9' = 1*1 + 0(-2) + 1*3 = 4$

Ці складові дійсно ідентичні величинами, отриманими методом накладання – додавання, зображеними на рис.

На основі викладеного можна вивести наступну процедуру накладання-додавання для швидкої згортки (або кореляції) за допомогою сегментації:

1. Обрати число N, даних послідовності x(n) порядку числа N_2 послідовності $h(n)$ (причому $N_1 > N_2$) і число значень N у вигляді $N = 2^k k$, де k – ціле (це зручно для застосування швидкого перетворення фур'є), причому $N = N_1 + N_2 - 1$. Для виконання цих умов послідовності даних доповнюються необхідною кількістю нулів.
2. Зсунути розширені сегменти даних x(n) у початок координат.
3. Для кожного розширеного сегменту даних $x'(n)$ знайти розгортку $x'(n) * h'(n)$. Можна застосовувати ШПФ, обчислити $X'(k) * H'(k)$, а потім виконати зворотнє перетворення.
4. Послідовно накласти отриманні згортки, суміщуючи останні N_2 - 1 точок одного результату з N_2 - 1 точок іншого, і додати згортки.

Метод накладання – запису. Розглянемо ще раз згортку $x'(n) * h'(n)$, проілюстровану на схемі



На цій схемі до h(n) додано N_1-1 нулів, так що обидві послідовності мають довжину $N = N_1 + N_2 - 1$.

Щоб отримати згортку, h(n) можна послідовно (по одній точці) зсувати праворуч, проводити перехресне множення відповідних значень і додавати результати.

Оскільки жодна з послідовностей не має довжину N, то перші $N_2 - 1$ значень згорткової суми будуть хибними. Отже, згортка $x'(n) * h'(n)$ буде містити періодичні пропуски довжиною $N_2 - 1$. Ці пропуски можна коректно заповнити,

наклавши останні $N_2 - 1$ точок кожної послідовності довжиною N_1, на перші $N_2 - 1$ точок наступної послідовності, а потім відкинувши ці $N_2 - 1$ перших точок. Ця процедура має назву накладання – запису.

Приклад. Умови ті ж самі, що і в попередньому прикладі

$$h(n) = \{1,0,1\}$$

$$x(n) = \{1,3,2,-3,0,2,-1,0,-2,3,-2,1\}$$

Оскільки для $h(n)$ $N_2 = 3$, величина накладання складає $N_2 - 1 = 2$ точки. Накладання відбувається згідно схеми :

$h(n)$	1	0	1									
x(n)	1	3	2	-3	0	2	-1	0	-2	3	-2	1
Ділянка 1	1	3	2	-3								
Ділянка 2			2	-3	0	2						
Ділянка 3					0	2	-1	0				
Ділянка 4							-1	0	-2	3		
Ділянка 5									-2	3	-2	1

Для сегмента 1 :

$y_{10} = h'_0 x'_{10} = 1 * 1 = 1$

$y_{11} = h'_0 x'_{11} + h'_1 x'_{10} = 1 * 3 + 0 * 1 = 3$

$y_{12} = h'_0 x'_{12} + h'_1 x'_1 + h'_2 x'_{10} = 1 * 2 + 0 * 3 + 1 * 1 = 3$

$y_{13} = h'_0 x'_{13} + h'_1 x'_{12} + h'_2 x'_{11} = 1(-3) + 0(2) + 1*3 = 0$

Отже : $y_1 = \{1,3,3,0\}$

Для сегменту 2 :

$y_{20} = h'_0 x'_{20} = 1 * 2 = 2$

$y_{21} = h'_0 x'_{21} + h'_1 x'_{20} = 1 * (-3) + 0 * 2 = -3$

$y_{22} = h'_0 x'_{22} + h'_1 x'_{21} + h'_2 x'_{20} = 1 * 0 + 0 * (-3) + 1 * 2 = 2$

$y_{23} = h'_0 x'_{23} + h'_1 x'_{22} + h'_2 x'_{21} = 1 * 2 + 0 * 0 + 1*(-3) = -1$

Отже : $y_2 = \{2,-3,2,-1\}$

Подібним чином для сегементів 3,4 та 5 отримаємо :

$y_3 = \{0,2,-1,2\}$

$y_4 = \{-1,0,-3,3\}$

$y_5 = \{-2,3,-4,4\}$

Дані результати ілюструються наступною таблицею

Сегмент 1	y_1	1 3 3 0 відкл
Сегмент 2	y_2	2 3 2 -1 відкл
Сегмент 3	y_3	0 2 -1 2 відкл
Сегмент 4	y_4	-1 0 -3 3 відкл
Сегмент 5	y_5	-2 3 -4 4 відкл
$x'(n) * h'(n)$		1 3 3 0 2 -1 -1 2 -3 3 -4 4 відкл

З таблиці видно, що перші $N_2 - 1 = 2$ точки кожного сегменту відкидаються. Останній рядок містить правильне значення згортки.

Таким чином, отримуємо наступну процедуру накладання запису :

1. Обрати число точок даних послідовності на початку координат.
2. Розмістити обидві послідовності на початку координат.
3. Для кожної послідовності обчислити відповідні зображення x(k) та H(k) за допомогою ШПФ.
4. Обчислити X(k) * H(k) і зворотнє до нього, які дорівнюють згортці відповідного по сегменту з n(n);
5. Розташувати всі згортки з накладанням по $N_2 - 1$ точкам
6. Відкинути $N_2 - 1$ значень кожної згортки.

Список рекомендованої літератури

Основна література

1. Душин В.К. Теоретические основы информационных процессов и систем.--М.: Дашков и К, 2006. –348 с.
2. Жураковський Ю.П., Полторак В.П. Теорія інформації та кодування: Підручник. –К.: Вища школа, 2001, -255 с.
3. Цимбал В. П. Теория информации и кодирование. –К.: Вища школа, 1992. -263 с.
4. Баскаков С.И. Радиотехнические цепи и сигналы: Учебник. –М.: Высш.школа, 1983. -536 с.
5. Зиновьев А.Л., Филиппов Л.И. Введение в теорию сигналов и цепей –М.: Высш.школа, 1975. -264 с.
6. Тутевич В.Н. Телемеханика. –М.: Высш.школа, 1985. -423 с.
7. Рабинер Л., Гоулд Б. Теория и применение цифровой обработки сигналов. –М.: Мир, 1978. -848 с.
8. Айфичер Э.С., Джервис Б.У. Цифровая обработка сигналов: практический подход, 2-е изд.: пер. с англ. –М.: Изд.дом «Вильямс», 2004. -992 с.
9. Сергиенко А.Б. Цифровая обработка сигналов: Учебник для вузов. 2-е изд. –Спб.: Питер, 2007. -751 с.
10. Фесечко В.О. методи перетворення сигналів: Навч. посіб. –К.: ІВЦ Політехніка, 2005. -128 с.

Додаткова література

11. Гоноровський И.С., Демин М.П. Радиотехнические цепи и сигналы: Учеб. Для вузов. –М.: Радио и связь, 1994. – 620 с.
12. Дьяконов В.П. Вейвлеты. От теории к практике. –М.: Солон-Р, 2002.
13. Оппенгейм А., Шафер Р. Цифровая обработка сигналов. –М.: Техносфера, 2006. -856 с.

14. Сиберт У.М. Цепи, сигналы, системы: в двух част. / пер.с англ. – М.: Мир, 1988.

15. Скляр Б. Цифровая связь. / Пер. с англ. –М.: Изд. Дом «Вильямс» , 2003. – 1104 с.

16. Смит С. Цифровая обработка сигналов. –М.: Додека XX1, 2008.- 720 с.

17. Денисенко А.Н. Сигналы. –М.: Горячая линия –Телеком. 2005.—704 с.