**Lec\_6**

**Оптимальне кодування**

**Алгоритм Шеннона-Фано**

**Алгоритм Хаффмана**

ОПТИМАЛЬНЕ КОДУВАННЯ

Код чи кодування може бути оптимальним лише за певних умов, за певними критеріями. Найбільш часто оцінюють оптимальність коду за показниками надійності передачі або за швидкістю передачі.

Поєднання двох умов в одному оптимальному коді не рекомендується, оскільки вони переважно є суперечливими. Наприклад, для збільшення швидкості передачі інформації необхідно усувати надлишковість, а для покращення показників надійності потрібно вводити надлишковість.

Одним з основних критеріїв, які характеризують код є середня довжина кодових слів, яка визначається за виразом:



p(ai) – ймовірність появи і-го символа первинного алфавіту

l(ai) – її довжина в коді (кількість розрядів при представленні).

Оптимальними без надлишковими називаються коди, які представляються кодовими словами мінімальної середньої довжини L. Ці коди є практично з нульовою надлишковістю. Верхня і нижня границі L визначаються нерівністю

H/log2m ≤ L ≤ (H/log2m) + 1

Н – ентропія первинного алфавіту; m – кількість якісних ознак вторинного алфавіту.

У випадку блокового кодування де кожен блок складається з М незалежних букв а1, а2, а3… ам. Мінімальна середня довжина кодового слова лежить в межах, які позначаються виразом:

(H\*M)/log2m ≤ LM ≤ ((H\*M)/log2m) + 1

З точки зору інформаційного навантаження на один символ повідомлення, блочне кодування є оптимальнішим ніж одиночне посимвольне кодування.

Основні властивості оптимальних кодів

1. Мінімальна середня довжина кодового слова для оптимального коду забезпечується у тому випадку, коли надлишковість кожного кодового слова зводиться до мінімуму, а в ідеальному випадку зводиться до 0;

2. Кодові слова оптимального коду повинні будуватись з рівноймовірних і взаємонезалежних символів.

З цих двох основних властивостей оптимальних кодів випливають принципи побудови оптимальних кодів:

1. принцип: вибір кожного кодового слова необхідно проводити так, щоб кількість вмістимої інформації була максимальною.

2. принцип: полягає в тому, що буквам вторинного алфавіту, які мають більшу ймовірність появи присвоюються короткі слова у вторинному алфавіті.

Знайти код, який був би оптимальним з усіх точок зору практично неможливо. Тому код може бути оптимальним тільки за певних умов (з точок зору швидкості передачі інформації, здатності виправляти помилки тощо).

У теорії інформації існує кілька методик побудови оптимальних з точки зору швидкості передачі інформації безнадмірних кодів.

До оптимальних безнадмірних кодів (з точки зору їх довжини, тобто швидкості передачі інформації) належать нерівномірні коди, які передають повідомлення комбінаціями мінімальної середньої довжини. Це зовсім не означає, що вони дійсно є абсолютно безнадмірними, оскільки такими вважаються коди, які задовольняють умову рівності обєму та кількості інформації. Ці коди все ж таки мають потенціальну надмірність через заборонені кодові комбінації, до яких належать комбінації, що доповнюють вершини неповного кодового дерева, яке відповідає оптимальному нерівномірному коду (ОНК), до повного утворення рівномірного коду.

Відповідно до оптимальним кодуванням називається процедура перетворення символів первинного алфавіту q1 на кодові комбінації вторинного алфавіту q2, при якій середня довжина повідомлення у вторинному алфавіті мінімальна.

Таким чином, основним завданням оптимального кодування є досягнення рівності між кількістю інформації І, що виробляється джерелом повідомлень, та об’ємом інформації Q на вході приймача повідомлень. Якщо І = Н, то збільшення швидкості передачі інформації завдяки поліпшенню процедури кодування стає неможливим.

Можна запропонувати кілька методик побудови ОНК для дискретних ансамблів повідомлень {Х,р(х)} із середньою довжиною кодових комбінацій

n(X) = H(X)/logq.

***Перша універсальна методика побудови ОНК ґрунтується на методиці Шеннона — Фано*** і передбачає цю побудову в кодовому алфавіті з кількістю якісних значень q. Згідно з цією методикою виконують такі процедури:

1. множину з N повідомлень, які кодуються, розташовують у порядку спадання ймовірностей;
2. впорядковані за ймовірностями повідомлення розбивають, по можливості, на q рівноймовірних груп;
3. кожній з груп завжди в одній і тій самій послідовності присвоюють символи алфавіту q (всім повідомленням першої групи — першу якісну ознаку цього алфавіту, всім повідомленням другої групи — другу якісну його ознаку тощо);
4. створені групи розбивають, по можливості, на рівноймовірні підгрупи, кількість яких дорівнює або менша ніж q (якщо після розбивання в групі залишається одне повідомлення, то подальший поділ стає неможливим);
5. кожній з утворених підгруп присвоюють якісні ознаки з алфавіту q за процедурою п.3;
6. розбивання та присвоєння ознак алфавіту q повторюють доти, поки після чергового поділу в утворених підгрупах залишиться не більш як одне повідомлення.

Для побудови ОНК за викладеною методикою слід ураховувати також відхилення від рівноймовірних значень, що утворюються при поділі на підгрупи. Вони враховуються згідно з правилами заліку остач ділення та середнього відхилення:

1. для того щоб повідомлення первинного джерела можна було поділити по можливості на якомога рівноймовірні підгрупи при побудові ОНК з алфавітом q, остача попереднього ділення додається за абсолютним значенням сумарної ймовірності чергового ділення [остачею ділення називається різниця між квантом ділення та реальним значенням сумарної ймовірності в групі (підгрупі), де квант ділення дорівнює 1/ q]
2. середнє відхилення має бути меншим або дорівнювати значенню ймовірності першого символу чергового ділення. Якщо середнє відхилення не дорівнює нулю, то середнє значення сумарної ймовірності в групі (підгрупі) при черговому діленні підраховується з додаванням значення середнього відхилення (середнім відхиленням називається абсолютне значення суми остач ділень на проміжних етапах побудови коду).

Розглянемо приклад побудови ОНК для передачі 16 повідомлень за допомогою четвіркового коду з алфавітом q = 4, якщо повідомлення на виході джерела з’являються з імовірностямир(х), як зазначено в табл. 5.5. Послідовність цієї побудови така:

Визначаємо квант поділу 1/q = 1/4 = 0,25.

Розбиваємо всі повідомлення, по можливості, на чотири рівномірні Групи:

І група – 1 повідомлення

ІІ група – 2-4 повідомлення

ІІІ група – 5-8 повідомлення

IV група – 9-16 повідомлення



За результатами першого поділу груп як перший символ кодових комбінацій присвоюємо послідовно якісні ознаки алфавіту q = 4 згідно з третьою процедурою першої універсальної методики побудови ОНК.

Утворені групи, крім першої, розбиваємо на підгрупи. Друга та третя групи мають до чотирьох повідомлень, тому як другий символ кодових комбінацій їм присвоюємо відповідно три та чотири якісні ознаки алфавіту q.

5. Четверта група має вісім повідомлень, тому розбиваємо її на рівноймовірні підгрупи. Квант поділу.

Поділ присвоєння ознак алфавіту q виконуємо доти, поки після чергового поділу в утворених підгрупах залишиться не більш як одне повідомлення.

6. Для того щоб перевірити оптимальність коду відносно довжини кодових комбінацій, визначаємо середню довжину nсер кодової комбінації ОНК. У разі оптимальності ця довжина не повинна перевищувати довжину рівномірного четвіркового коду, яким можна закодувати 16 повідомлень, тобто qn = 42= 16(n= 2):

nсер = 1,92 < 2.

Таким чином, утворений код дійсно оптимальний, оскільки nсер < n.

Оптимальність кодування можна визначити також порівнянням ентропії, що припадає на одне повідомлення, із середньою довжиною кодової комбінації, які мають бути дуже близькі за значеннями, причому ентропія має бути меншою від середньої довжини кодової комбінації або дорівнювати їй.

Отже,

 = 3.63 біт/символ

При цьому треба звернути увагу на те, що ентропія залежить від алфавіту оптимального коду, яким кодуються повідомлення, тобто слід урахувати кількість інформації, яка міститься в одному елементі кодової комбінації. Для оптимального коду q = 4 в одному елементі кодової комбінації буде 2 біти інформації.

Таким чином,

Н = 3,6354 біт/повідомлення < 2nсер =

= 3,84 біт/повідомлення < 2n = 2 • 2 = 4 біт/повідомлення,

тобто код є оптимальним.

***Друга універсальна методика побудови ОНК грунтується на відомій методиці Хаффмена.***

Вона, як і методика Шеннона - Фано, передбачає побудову ОНК у кодовому алфавіті з кількістю якісних значень q. Згідно з цією методикою виконують такі процедури:

1. множину з N повідомлень, що кодуються, розташовують у порядку спадання ймовірностей;
2. останні N0 (2 < N0 <=q). повідомлень об’єднують у нове повідомлення з імовірністю, що дорівнює сумі ймовірностей об’єднуваних повідомлень;
3. утворену множину (N - N0 + 1) повідомлень розташовують у порядку спадання ймовірностей;
4. об’єднують останні q повідомлень і впорядковують множину повідомлень у порядку спадання ймовірностей. Так діють доти, доки ймовірність чергового об’єднаного повідомлення не дорівнюватиме одиниці;
5. будують кодове дерево, починаючи з кореня, і гілкам цього дерева присвоюють якісні ознаки кодового алфавіту q.

Кодові комбінації ОНК — це послідовність якісних ознак, які зустрічаються на шляху від кореня до вершини кодового дерева.

Побудову ОНК за допомогою другої універсальної методики розглянемо стосовно передачі 16 повідомлень комбінаціями четвіркового коду, які задано в попередньому прикладі. Послідовність цієї побудови така:

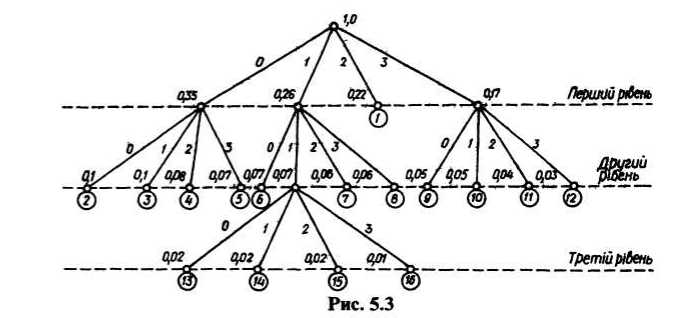
1. Множину з N = 16 повідомлень розташовуємо в порядку спадання ймовірностей.
2. Оскільки q = 4, об’єднуємо останні чотири повідомлення й утворюємо нове умовне повідомлення з імовірністю, що дорівнює сумі ймовірностей об’єднаних повідомлень.
3. Утворену множину з 16-4 + 1 = 13 повідомлень розташовуємо в порядку спадання ймовірностей.
4. Знову об’єднуємо останні чотири повідомлення та впорядковуємо множину повідомлень у порядку спадання ймовірностей. Цю процедуру повторюємо ще три рази, поки при останньому об’єднанні сумарна ймовірність не досягне значення одиниці (табл. 5.6).
5. Будуємо кодове дерево (рис. 5.3). Його гілкам присвоюємо якісні ознаки кодового алфавіту від 0 до 3.
6. Кодові комбінації ОНК заносимо до табл. 5.6. Вони визначаються послідовністю якісних ознак, які зустрічаються на шляху від кореня до певної вершини кодового дерева.
7. Для того щоб перевірити оптимальність коду відносно довжини кодових комбінацій, визначаємо середню довжину л кодової комбінації ОНК та ентропію, що припадає на одне повідомлення:

nсер=1,85<2

Н = 0,6354 біт/пов < 2nсер = 3,7 біт/пов< 2n =4 біт/пов

Отже, цей код також є оптимальним і має кращі показники, ніж оптимальний код, побудований за першою універсальною методикою, оскільки nсер2 <nсер1 (0.85 < 1,92).

Обидві універсальні методики мають неоднозначність, але перша з них дає змогу точніше будувати ОНК. До недоліків другої універсальної методики побудови ОНК слід віднести громіздкість (особливо зі збільшенням кількості повідомлень N та алфавіту q коду), що пояснюється необхідністю побудови кодового дерева.



Відзначимо, що переваги другої універсальної методики побудови ОНК з q>2 при N<qn будуть вагоміші при більш ретельному виборі кількості найменш імовірних повідомлень, що об’єднуються на першому етапі (2 < N0 <q). На всіх наступних етапах ця кількість має дорівнювати q.

## Оптимальні нерівномірні коди (онк)

Побудовані за даними, універсальними або подібними методиками, коди з нерівномірними розподілами символів, які мають мінімальну середню довжину кодового слова називаються оптимальними нерівномірними кодами (ОНК).

Рівномірні коди можуть бути оптимальними лише за умови використання їх для передачі повідомлень з рівномірним розподілом символів первинного алфавіту. При цьому кількість символів первинного алфавіту повинна бути рівною цілому степеню кількості якісних ознак вторинного алфавіту. Для двійкових кодів – це цілому степеню двійки. Максимально ефективними будуть ОНК для яких :

lcep = H

Для двійкових кодів повинна забезпечуватись наступна умова:

https://studfiles.net/html/2706/1080/html_p99rt0Dtat.dUAC/img-_m6R9D.png

При цьому довжина коду li –го у вторинному алфавіті повинна задовольняти наступну рівність.

li= -log2 pi = log2(1/pi)

Ефективність ОНК оцінюють з використанням коефіцієнта статистичного стиснення, яке характеризує зменшення кількості двійкових знаків на символ повідомлення у випадку застосування ОНК у порівнянні із застосуванням методів нестатистичного кодування.

Цей коефіцієнт визначається із наступного співвідношення.



Коефіцієнт за наступним виразом вказує ефективність використання статистичної надлишковості:

Kcc= H/lcep

Ентропію обчислюють в залежності від випадків:

1. нерівномірних і взаємозалежних – за ф-ю Шенона;

2. нерівномірних і взаємозалежних символів – за формулою повної умовної ентропії