

Міністерство внутрішніх справ України
Харківський національний університет внутрішніх справ
кафедра інформаційних технологій, факультет № 4

ТЕКСТ ЛЕКЦІЇ

з навчальної дисципліни Теорія інформації та кодування
обов'язковий компонент
освітньої програми першого (бакалаврського) рівня вищої освіти
125 – Кібербезпека (безпека інформаційних та комунікаційних систем)

за темою – «Інформаційні характеристики сукупності дискретних
немарковських джерел інформації»

ЗАТВЕРДЖЕНО

Науково-методичною радою
Харківського національного
університету внутрішніх справ
Протокол від _____ № ____

СХВАЛЕНО

Вченою радою факультету № 4
Протокол від _____ № ____

ПОГОДЖЕНО

Секцією науково-методичної ради
ХНУВС з технічних дисциплін
Протокол від _____ № ____

Розглянуто на засіданні кафедри інформаційних технологій протокол
від _____ № ____

Рецензенти: *Носов В.В., професор кафедри кібербезпеки факультету №4
Харківського національного університету внутрішніх справ к.т.н., доцент.*

Розробник: *Тулупов Володимир Володимирович к.т.н., доцент – м. Харків:
Харківський національний університет внутрішніх справ*

План лекції

Вступ

1. Характеристики дискретного джерела інформації
2. Характеристики двох дискретних немарковських джерел інформації

Висновки

Література:

1. Духин, Александр Александрович. Теория информации: Учебное пособие / А.А.Духин. - М.: Гелиос АРВ, 2007.
2. Стариченко Б. Е. Теоретические основы информатики: Учебное пособие для вузов. - 2-е изд. перераб. и доп. - М.: Горячая линия - Телеком, 2003.
3. Дмитриев В.И. Прикладная теория информации. Учебник для студентов ВУЗов по специальности
4. «Автоматизированные системы обработки информации и управления». – М.: Высшая школа, 1989 – 320 с.
5. К.Шеннон. Работы по теории информации и кибернетике. Издательство иностранной литературы, Москва, 1963.

Текст лекції

Вступ

У більшості літератури з теорії інформації замість поняття **ймовірнісної схеми** використовується поняття **джерело інформації**. Спочатку розглянемо коротко характеристики **дискретного джерела інформації**, а потім інформаційні характеристики сукупності дискретних немарковських джерел інформації.

1. Характеристики дискретного джерела інформації

Дискретне джерело інформації – це таке джерело, яке може виробити (згенерувати) за скінчений відрізок часу тільки скінчену множину повідомлень. Кожному такому повідомленню можна співставити відповідне число, та передавати ці числа замість повідомлень.

Дискретне джерело інформації є достатньо адекватною інформаційною моделлю дискретних систем, а також неперервних систем, інформаційні сигнали про стан яких піддають аналого-цифровому перетворенню; таке перетворення виконується в більшості сучасних автоматизованих систем управління.

Первинні характеристики дискретного джерела інформації – це

- алфавіт,
- сукупність ймовірностей появи символів алфавіту на виході дискретного джерела
- та тривалості символів.

Алфавіт – множина $X = \{x_1, x_2, x_3, \dots, x_M\}$ символів, які можуть з'явитися на виході дискретного джерела; M – потужність, тобто кількість різноманітних символів алфавіту.

Якщо всі ймовірності, які визначають виникнення символів на виході джерела, не залежать від часу, джерело називають **стаціонарним**. Ми будемо розглядати тільки стаціонарні джерела та для скорочення замість "**стаціонарне джерело**" будемо всюди використовувати "джерело".

Для опису джерел, які не мають пам'яті, достатньо мати значення безумовних ймовірностей $p(x_i)$ виникнення символів x_i , $i = 1, 2, 3, \dots, M$ на його виході.

Більшість реальних джерел інформації є джерелами з **пам'яттю**. Розподіл ймовірностей виникнення чергового символу на виході дискретного джерела з пам'яттю залежить від того, які символи були попередніми. Таке джерело інформації називають **марковським**, оскільки процес появи символів на його виході адекватний ланцюгам Маркова; останні в свою чергу

отримали таку назву на честь російського математика Маркова (старшого) Андрія Андрійовича (1856 – 1922), який заклав основи розділу теорії випадкових процесів.

Будемо говорити, що **глибина пам'яті** марковського дискретного джерела інформації дорівнює h , ($h \geq 0$), якщо ймовірність появи чергового символу залежить тільки від h попередніх символів на виході цього джерела.

Кількість інформації – одне із основних понять теорії інформації, яка розглядає технічні аспекти інформаційних проблем, тобто вона дає відповіді на запитання такого типу: якою повинна бути ємність запам'ятовуючого пристрою для запису даних про стан деякої системи, якими повинні бути характеристики каналу зв'язку для передачі певного повідомлення тощо.

Кількісна оцінка інформації пов'язана з поняттям ентропії. **Ентропія є мірою невизначеності, непрогнозованості ситуації.** Зменшення ентропії, що відбулось завдяки деякому повідомленню, точно збігається з кількістю інформації, яка міститься в цьому повідомленні.

Для дискретного немарковського (без пам'яті) джерела інформації ентропія H визначається за таким виразом:

$$H = - \sum_{i=1}^M p(x_i) \log p(x_i). \quad (2.1)$$

Зазначимо, що H не залежить від того, якими є випадкові події або величини (якщо x_i – випадкова величина), а визначається тільки значеннями ймовірностей. Це означає, що **ентропія є характеристикою розподілу ймовірностей.**

Значення H показує, яку кількість інформації в середньому дає поява одного символу на виході дискретного джерела інформації. Ця міра запропонована американським математиком і інженером Клодом Шенноном.

Якщо основа логарифма в (2.1) дорівнює двом, то одиниці вимірювання H , а також кількості інформації називають **бітами** або двійковими одиницями.

Ентропія дискретного розподілу ймовірностей завжди невід'ємна і набуває максимального значення H_{\max} , коли всі $p(x_i)$ мають однакові значення:

$$p(x_i) = \frac{1}{M} \quad (2.2)$$

В цьому разі маємо міру кількості інформації, яку ще до Шеннона було запропоновано англійським математиком Р.Хартлі. Підставимо (2.2) в (2.1), отримаємо

$$H = H_{\max} = \log_2 M. \quad (2.3)$$

Значення H_{\max} збігається з кількістю двійкових комірок пам'яті, які необхідно мати, щоб зафіксувати за допомогою двійкового коду інформацію про один із M можливих станів системи, або про символ, що з'явиться на виході дискретного джерела інформації.

Ентропія дорівнює нулю, якщо ймовірність появи одного з символів є одиниця (при цьому, звичайно, ймовірність появи будь-якого іншого символу буде дорівнювати нулю); в такій ситуації невизначеність відсутня.

Продуктивність \bar{H} джерела інформації – це кількість інформації, що виробляється джерелом за одиницю часу:

$$\bar{H} = \frac{H}{\tau}, \quad (2.4)$$

де $\tau = \sum_{i=1}^M p(x_i) \tau_i$ – середня тривалість символу, τ_i – тривалість символу x_i .

Надмірність (надлишок) R дискретного джерела інформації дає відносну оцінку використання потенційних можливостей джерела з алфавітом заданої потужності M :

$$R = \frac{H_{\max} - H}{H_{\max}} = \frac{\log_2 M - H}{\log_2 M} = 1 - \frac{H}{\log_2 M} \quad (2.5)$$

Надмірність може приймати значення від 0 до 1. Вона дорівнює нулю, якщо $H = H_{\max}$; в цьому випадку дискретне джерело інформації буде виробляти максимально можливий інформаційний потік. Із першої теореми Шеннона виходить, що при застосуванні ефективного кодування надмірність може бути зведена практично до нуля, внаслідок чого об'єм повідомлення буде зменшено майже в $1/(1-R)$ разів.

Ентропія, продуктивність та надмірність – інтегральні інформаційні характеристики дискретного джерела інформації.

2. Характеристики сукупності дискретних немарковських джерел інформації

Розглянемо сукупність двох дискретних немарковських джерел інформації з алфавітами $X = \{x_1, x_2, x_3, \dots, x_M\}$ та $Y = \{y_1, y_2, y_3, \dots, y_N\}$, ентропії яких позначимо відповідно як $H(X)$ та $H(Y)$. Для спрощення будемо вважати, що всі символи мають однакові тривалості, а зміни символів на виходах обох джерел відбуваються одночасно. Ентропії кожного з джерел знаходяться за виразом (2.1) через безумовні ймовірності появи символів x_i та y_k . Якщо джерела є статистично залежними, то поява, наприклад, символу x_1 на виході першого джерела дасть розподіл умовних ймовірностей $p(y_k / x_1)$, $k=1, 2, \dots, N$ виникнення символів y_k на виході другого джерела, який у загальному випадку буде відрізнятися від розподілу умовних ймовірностей $p(y_k / x_2)$ при умові появи на виході першого джерела символу x_2 . Звичайно, в такій ситуації ентропія другого джерела буде залежити від того, який символ з'явився на виході першого джерела. Знаходиться ця ентропія через умовні ймовірності:

$$H(Y / x_i) = - \sum_{k=1}^N p(y_k / x_i) \cdot \log_2 p(y_k / x_i) . \quad (2.6)$$

Вона має назву **умовної частинної ентропії** та характеризує невизначеність символів на виході другого джерела при умові, що на виході першого з'явився символ x_i . Якщо

$H(Y / x_i)$ усереднити по всіх x_i , то отримаємо **середню** або **повну умовну ентропію**:

$$\begin{aligned} H(Y / X) &= \sum_{i=1}^M p(x_i) \cdot H(Y / x_i) = \\ &= - \sum_{i=1}^M \sum_{k=1}^N p(x_i) \cdot p(y_k / x_i) \cdot \log_2 p(y_k / x_i) = \\ &= - \sum_{i=1}^M \sum_{k=1}^N p(x_i, y_k) \cdot \log_2 p(y_k / x_i) , \end{aligned} \quad (2.7)$$

де $p(x_i, y_k) = p(x_i) \cdot p(y_k / x_i)$ – ймовірність сумісної появи символів x_i та y_k на виходах відповідно першого та другого джерела.

Ця ентропія характеризує в середньому невизначеність символів на виході другого джерела, якщо є можливість спостерігати за появою символів на виході першого джерела.

Аналогічно визначається частинна $H(X / y_k)$ та середня (повна) $H(X / Y)$ умовні ентропії для першого джерела.

Якщо дискретні джерела статистично незалежні, то

$$\begin{aligned} p(y_k / x_i) &= p(y_k) , \\ p(x_i, y_k) &= p(x_i) \cdot p(y_k) . \end{aligned} \quad (2.8)$$

В цьому випадку

$$\begin{aligned} H(Y / X) &= H(Y / x_i) = H(Y) ; \\ H(X / Y) &= H(X / y_k) = H(X) . \end{aligned} \quad (2.9)$$

Середня умовна ентропія $H(Y / X)$ не може перевищувати безумовну ентропію $H(Y)$; частинна умовна ентропія $H(Y / x_i)$ може бути більша, ніж $H(Y)$. Середня умовна ентропія буде дорівнювати нулю, якщо поява будь якого символу x_i , $i = 1, 2, 3, \dots, M$ на виході першого джерела однозначно визначає символ на виході другого джерела. Тобто, спостерігаючи за виникненням символів на виході першого джерела, будемо мати повну інформацію про послідовність символів на виході другого джерела навіть за умов недоступності цього виходу для спостереження.

Таким чином для середньої умовної $H(Y / X)$ та безумовної $H(Y)$ ентропій мають місце співвідношення:

$$0 \leq H(Y / X) \leq H(Y) . \quad (2.10)$$

Для двох визначених вище дискретних джерел можна розрахувати **сумісну ентропію** $H(X, Y)$ сукупності символів x_i , y_k або ентропію об'єднання ансамблей x та y (рис. 2.1):

$$H(Y, X) = - \sum_{i=1}^M \sum_{k=1}^N p(x_i, y_k) \log_2 p(x_i, y_k) \quad (2.11)$$

Рис. 2.1.

Для обчислення $H(X, Y)$ слід мати набір або матрицю ймовірностей $p(x_i, y_k)$ сумісної появи x_i та y_k

$$\begin{bmatrix} p(x_1, y_1) & p(x_2, y_1) & \dots & p(x_M, y_1) \\ p(x_1, y_2) & p(x_2, y_2) & \dots & p(x_M, y_2) \\ \vdots & \vdots & \ddots & \vdots \\ p(x_1, y_N) & p(x_2, y_N) & \dots & p(x_M, y_N) \end{bmatrix}. \quad (2.12)$$

Сума елементів k -го рядка цієї матриці дорівнює безумовній ймовірності $p(y_k)$ появи символу y_k на виході другого джерела, а сума елементів i -го стовпця – безумовній ймовірності $p(x_i)$ появи символу x_i на виході першого джерела:

$$p(y_k) = \sum_{i=1}^M p(x_i, y_k); \quad p(x_i) = \sum_{k=1}^N p(x_i, y_k) \quad (2.13)$$

Маючи безумовні ймовірності $p(x_i)$ та $p(y_k)$ появи символів x_i та y_k на виході кожного з джерел, а також ймовірності $p(x_i, y_k)$ сумісної їх появи, можна обчислити умовні ймовірності $p(x_i / y_k)$ та $p(y_k / x_i)$, користуючись виразом:

$$p(x_i, y_k) = p(x_i) \cdot p(y_k / x_i) = p(y_k) \cdot p(x_i / y_k), \quad (2.14)$$

а далі за виразом (2.7) знайти умовні ентропії $H(Y / X)$ та $H(X / Y)$. Тобто матриця (2.12) дає необхідні дані для обчислення ентропій $H(X)$, $H(Y)$ кожного з джерел та умовних ентропій $H(X / Y)$, $H(Y / X)$.

Всі вище перелічені ентропії можна також отримати із матриць умовних ймовірностей $p(x_i / y_k)$ або $p(y_k / x_i)$ та необхідної кількості безумовних ймовірностей $p(x_i)$ та $p(y_k)$.

Ентропію $H(X, Y)$ системи двох джерел, користуючись виразом (2.14), подамо у вигляді:

$$\begin{aligned} H(X, Y) = & - \sum_{i=1}^M p(x_i) \cdot \log_2 p(x_i) \cdot \sum_{k=1}^N p(y_k / x_i) - \\ & - \sum_{i=1}^M \sum_{k=1}^N p(x_i) \cdot p(y_k / x_i) \cdot \log_2 p(y_k / x_i). \end{aligned} \quad (2.15)$$

Перша складова з урахуванням того, що

$$\sum_{k=1}^N p(y_k / x_i) = 1,$$

є ентропією $H(X)$ першого джерела, друга збігається з виразом (2.7) для умовної ентропії. Тобто

$$H(X, Y) = H(X) + H\left(\frac{Y}{X}\right) \quad (2.16)$$

Аналогічно можна показати, що

$$H(X, Y) = H(Y) + H\left(\frac{X}{Y}\right) \quad (2.17)$$

Якщо джерела статистично незалежні, то із виразу (2.9) виходить

$$H(X, Y) = H(Y) + H(X) \quad (2.18)$$

У загальному випадку

$$H(X, Y) \leq H(Y) + H(X) \quad (2.19)$$

Для системи, що складається з s джерел з алфавітами $X_1, X_2, X_3, \dots, X_s$, ентропія визначається так:

$$\begin{aligned} H(X_1, X_2, \dots, X_s) = & H(X_1) + H(X_2 / X_1) + \\ & + H(X_3 / X_2, X_1) + \dots + H(X_s / X_{s-1}, X_{s-2}, \dots, X_2, X_1). \end{aligned} \quad (2.20)$$

Звичайно, як і для двох джерел, має місце співвідношення

$$H(X_1, X_2, \dots, X_s) \leq H(X_1) + H(X_2) + \dots + H(X_s), \quad (2.21)$$

де рівність має місце, коли всі джерела статистично незалежні.

У підсумку, **умовна ентропія** має наступні властивості:

- $0 < H\left(\frac{Y}{X}\right) < H(Y)$;
- $H\left(\frac{Y}{X}\right) = 0$, коли з реалізації ансамблю X можна точно встановити реалізацію ансамблю Y ;
- $H\left(\frac{Y}{X}\right) = H(Y)$, коли ансамблі X і Y незалежні і знання реалізації X не додає інформації про Y ;
- $H(Y) > H\left(\frac{Y}{X}\right)$ – загальний випадок, коли знання реалізації X знижує первісну невизначеність Y .

Звернемось знову до системи двох дискретних джерел. Спостерігаючи за виникненням символів на виході одного із джерел, наприклад першого, в загальному випадку будемо отримувати певну **кількість інформації про появу символів на виході другого джерела**. Ця інформація $I(X, Y)$ в розрахунку на один символ буде дорівнювати зменшенню ентропії другого джерела. Оскільки початкова або апіорна ентропія другого джерела (тобто ентропія, яка мала місце до спроби, де під спробою будемо розуміти появу символу на виході першого джерела, який є доступним) дорівнює $H(Y)$, а залишкова або апостеріорна (після спроби) ентропія буде $H(Y/X)$, то

$$I(X, Y) = H(Y) - H(Y/X). \quad (2.22)$$

Ця величина показує, яка кількість інформації в середньому міститься в одному символі першого джерела про виникнення символів на виході другого джерела.

Користуючись виразами для безумовної та умовної ентропії, після деяких перетворень можна отримати:

$$I(Y, X) = - \sum_{i=1}^M \sum_{k=1}^N p(x_i, y_k) \log_2 \frac{p(x_i, y_k)}{p(x_i)p(y_k)} \quad (2.23)$$

Крім того, враховуючи (2.16) та (2.17), будемо мати такі інтерпретації для $I(X, Y)$:

$$I(Y, X) = H(X) + H(Y) - H(X, Y) = H(X) + H\left(\frac{Y}{X}\right) \quad (2.24)$$

Тобто кількість інформації, що містить в середньому символ на виході першого джерела про виникнення символів на виході другого джерела, дорівнює кількості інформації, яка міститься в середньому в символі на виході другого джерела про виникнення символів на виході першого. Через це $I(X, Y)$ має назву **повної взаємної інформації**. Аналіз виразів (2.22), (2.23), (2.24) показує, що рівність повної взаємної інформації нулю є необхідною і достатньою умовою статистичної незалежності джерел.

Пояснимо графічно поняття **власної ентропії, умовної ентропії, сумісної ентропії і повної взаємної інформації**.

На рис. 2.2 умовно показана **власні ентропії $H(X)$ і $H(Y)$, умовні ентропії $H\left(\frac{Y}{X}\right)$ і $H\left(\frac{X}{Y}\right)$ і сумісна ентропія $H(X, Y)$** .

Рис. 2.2

На рис. 2.2 заштрихована частина - **повна взаємна інформація $I(X, Y)$** , що міститься в ансамблях X і Y . Вона показує, яка (в середньому) кількість інформації містить повідомлення X про повідомлення Y (або навпаки, повідомлення Y про повідомлення X).

Як впливає з рис. 2.2

$$I(X, Y) = H(X) - H\left(\frac{X}{Y}\right) = H(Y) - H\left(\frac{Y}{X}\right) = H(X, Y) - H\left(\frac{Y}{X}\right) - H\left(\frac{X}{Y}\right)$$

Якщо повідомлення X і Y повністю незалежні (рис. 2.3), то взаємна інформація відсутня і $I(X, Y) = 0$, $H(X, Y) = H(X) + H(Y)$.

Рис. 2.3

Якщо повідомлення X і Y частково залежні (рис. 2.2), то **взаємна інформація $I(X, Y) \neq 0$**

$$H(X, Y) = H(X) + H(Y) - I(X, Y)$$

Якщо X і Y повністю залежні (X і Y містять одну і ту ж інформацію) і $H(X) = H(Y)$, то $I(X, Y) = H(X) = H(Y)$,

$$H(X, Y) = I(X, Y) = H(X) = H(Y)$$

Поняття **взаємної інформації** дуже широко використовується в теорії передачі інформації. Вимоги до взаємної інформації різні залежно від того, з якою інформацією ми маємо справу.

Наприклад, якщо X і Y це повідомлення, що публікуються різними газетами, то для отримання можливо більшої сумарної (спільної) інформації взаємна (тобто однакова в даному випадку) інформація повинна бути мінімальною.

Якщо ж X і Y це повідомлення на вході і на виході каналу зв'язку з перешкодами, то для одержання можливо більшої інформації її отримувачем необхідно, щоб взаємна інформація була найбільшою. Тоді умовна ентропія $H\left(\frac{Y}{X}\right)$ це втрати інформації в каналі зв'язку (ненадійність каналу), $H\left(\frac{Y}{X}\right)$ це інформація про перешкоди (ентропія джерела перешкод $H(n)$), що надходить в канал ззовні або створювана внутрішніми перешкодами в каналі (схематично цей процес показаний на рис. 2.4).

Рис. 4

Властивості взаємної інформації.

1. $I(X, Y) = I(Y, X)$. Взаємна інформація симетрична.
2. $I(X, Y) \geq 0$. Взаємна інформація завжди позитивна.
3. $I(X, Y) = 0$ тоді і тільки тоді, коли ансамблі X і Y незалежні.
4. $I(X, Y) = H(X) - H\left(\frac{Y}{X}\right) = H(Y) - H\left(\frac{X}{Y}\right) = H(X) + H(Y) - H(X, Y)$, тобто у разі настання спільної події $H(X) + H(Y) = H(X, Y)$ взаємна інформація відсутня.
5. $I(X, Y) \leq \min\{H(X), H(Y)\}$. Взаємна інформація не може бути більше інформації про кожний ансамбль окремо.
6. $I(X, Y) \leq \min\{\log|X|, \log|Y|\}$. Логарифмічна міра кожного з ансамблів окремо більше або дорівнює взаємної інформації.
7. Взаємна інформація $I(X, Y)$ має максимум (є опуклою функцією розподілу ймовірностей).

Контрольні питання

1. Як визначається умовна ентропія $H\left(\frac{Y}{X}\right)$, яка характеризує невизначеність символів на виході джерела Y при умові, що на виході джерела X з'явилися всі можливі символи?
2. Як співвідносяться умовна $H\left(\frac{Y}{X}\right)$ та безумовна $H(Y)$ ентропії для статистично залежних джерел?
3. Як визначається взаємна ентропія $H(X, Y)$ об'єднання ансамблів X та Y ?
4. Як співвідносяться взаємна $H(X, Y)$ та безумовні $H(X)$, $H(Y)$ ентропії для статистично залежних джерел?
5. Як визначається взаємна інформація $I(X, Y)$ статистично залежних джерел X та Y ?
6. Якими є властивості взаємної інформації $I(X, Y)$ статистично залежних джерел X та Y ?
7. Якими є властивості умовної ентропії $H\left(\frac{Y}{X}\right)$ для статистично залежних джерел?

