Лекція 03-04.

ТЕОРІЯ ІНФОРМАЦІЇ В ЛІНГВІСТИЦІ

1 ФОРМИ АДЕКВАТНОСТІ ТА МІРИ ІНФОРМАЦІЇ

Важливою характеристикою інформації є її *адекватність*, тобто певний рівень відповідності створюваного за допомогою отриманої інформації образу реальному об'єкту, процесу, явищу і т.п. Адекватність інформації може виражатися в трьох формах: ***прагматичній, семантичній*** та ***синтаксичній.***

Прагматична (споживацька) адекватність відображає відповідність інформації меті управління, яка на її основі реалізується. Ця форма адекватності безпосередньо пов'язана з практичним використанням інформації.

Семантична (смислова) адекватність визначає ступінь відповідності образу об'єкту і самого об'єкту. Семантичний аспект припускає облік смислового змісту інформації На цьому рівні аналізуються ті відомості, які відображає інформація, розглядаються смислові зв'язки. Ця форма служить для формування понять і уявлень, виявлення значення, змісту інформації та її узагальнення.

Синтаксична адекватність відображає формально-структурні характеристики інформації і не зачіпає її смислового змісту. На синтаксичному рівні враховуються тип носія і спосіб передачі інформації, швидкість передачі і обробки, розміри кодів представлення інформації, надійність і точність перетворення цих кодів і т.п. Ця форма сприяє сприйняттю структурних зовнішніх характеристик, тобто синтаксичної сторони інформації.

Якщо розглядається дослід, що може закінчитись одним із N можливих результатів, то необхідно оцінити такий результат. Такою оцінкою може стати міра інформації (або події). Міра загального об'єднання подій дорівнює сумі мір кожної події.

Кожній формі адекватності відповідає своя міра кількості інформації та об'єм даних.

Прагматична міра визначає корисність інформації (цінність) для досягнення користувачем поставленої мети. Ця міра також є відносною величиною, що зумовлено особливостями використання цієї інформації в тій чи іншій системі.

В якості **семантичної міри** інформації для вимірювання смислового змісту інформації, тобто її кількості на семантичному рівні, найбільше визнання отримала тезаурусна міра, яка пов'язує семантичні властивості інформації із здатністю користувача приймати повідомлення, що поступило. Для цього використовується поняття тезаурусу користувача.

Тезаурус - це сукупність відомостей, які має в своєму розпорядженні користувач.

Залежно від співвідношень між смисловим змістом інформації $\boldsymbol{S}$ і тезаурусом користувача Θ змінюється кількість семантичної інформації I_C, яка сприймається користувачем і включається ним надалі в свій тезаурус. Характер такої залежності показаний на рис.1.

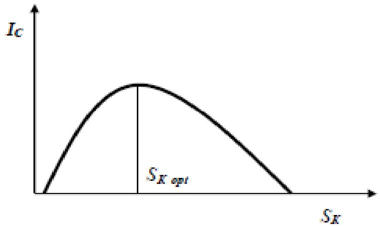


Рис. 1. Залежність кількості семантичної інформації, що сприймається користувачем, від його тезауруса: $I_c = f(\Theta)$

Розглянувши граничні випадки, коли кількість семантичної інформації рівна 0 ***(I_C=0),*** побачимо, що:

при $\Theta \approx 0$ користувач не сприймає, не розуміє інформацію, яка поступає;

при $\Theta \to \infty$ користувач усе знає, отже інформація, яка поступає, йому не потрібна.

Максимальну кількість семантичної інформації користувач отримує при узгодженні її смислового змісту $\boldsymbol{S}$ зі своїм тезаурусом Θ $(\Theta = \Theta_{опт})$, коли інформація, що поступає, зрозуміла користувачу і несе дані невідомі йому раніше (відсутні в його тезаурусі).

Отже, кількість семантичної інформації в повідомленні, кількість нових знань, одержуваних користувачем, є величиною відносною. Одне і те ж повідомлення може мати смисловий зміст для компетентного користувача і бути безглуздим (семантичний шум) для користувача некомпетентного. При оцінці семантичного (змістовного) аспекту інформації необхідно прагнути до узгодження величин $\boldsymbol{S}$ і Θ.

Відносною мірою кількості семантичної інформації може служити коефіцієнт змістовності $\boldsymbol{Z}$, який визначається як відношення кількості семантичної інформації до її об'єму:

$$Z = {}^{I_C}\!/_{V_Д} . \tag{1}$$

Синтаксична міра інформації оперує із знеособленою інформацією, що не виражає смислового відношення до об'єкту.

Об'єм даних $V_Д$ у повідомленні вимірюється кількістю символів (розрядів) в цьому повідомленні. В різних системах числення один розряд має різну вагу і відповідно міняється одиниця вимірювання даних.

Кількість інформації I на синтаксичному рівні неможливо визначити без розгляду поняття невизначеності стану системи (ентропії системи). Дійсно, отримання інформації про яку-небудь систему завжди пов'язане із зміною ступеня непоінформованості одержувача про стан цієї системи, тобто кількість інформації вимірюється зміною (зменшенням) невизначеності стану системи. Далі детальніше буде розглянуто існуючі кількісні міри інформації.

Коефіцієнт чи ступінь інформативності (лаконічність) повідомлення визначається відношенням кількості інформації до об'єму даних:

$$Y = {}^{I}\!/_{V_Д}, \tag{2}$$

причому $0 < Y < 1$.

Із збільшенням Y зменшуються об'єми роботи по перетворенню інформації (даних) в системі. Тому природним є прагнення підвищення інформативності, для чого розробляються спеціальні методи оптимального кодування інформації.

2 КІЛЬКІСНІ МІРИ ІНФОРМАЦІЇ

2.1 Ентропія як міра невизначеності лінгвістичної події

Кількісні виміри інформації можна здійснити, спираючись на два початкових поняття – ймовірності випадкової лінгвістичної події і невизначеності, яка є перед виконанням експерименту, результатом якого є вказана подія. Поняття невизначеності та її міри вимагає спеціального роз'яснення.

Якщо множина елементів, з яких здійснюється вибір, складається з одного єдиного елемента, то його вибір приречений, тобто ніякої невизначеності вибору немає. Таким чином, якщо ми взнаємо, що вибраний цей єдиний елемент, то не отримуємо ніякої нової інформації, тобто отримуємо нульову кількість інформації. Якщо множина складається з двох елементів, то невизначеність вибору існує, але її значення мінімальне. У цьому випадку мінімальна і кількість інформації, яку ми одержуємо, дізнавшись про вибір одного з елементів. Із збільшенням кількості елементів у множині збільшується невизначеність вибору, а отже ми отримуємо більшу кількість інформації, дізнавшись про те, який елемент був вибраний.

Кожний лінгвістичний експеримент (дослід) має деяку невизначеність результату. Якщо наш дослід полягає в послідовному вгадуванні букв невідомого слова, то вгадування кожної букви від початку слова має свою невизначеність. Що більше альтернатив при виборі можливого результату експерименту, то більша його невизначеність; що менше таких альтернатив, то менша невизначеність у результаті досліду. Між невизначеністю досліду і кількістю рівноможливих результатів є такі дві очевидні залежності:

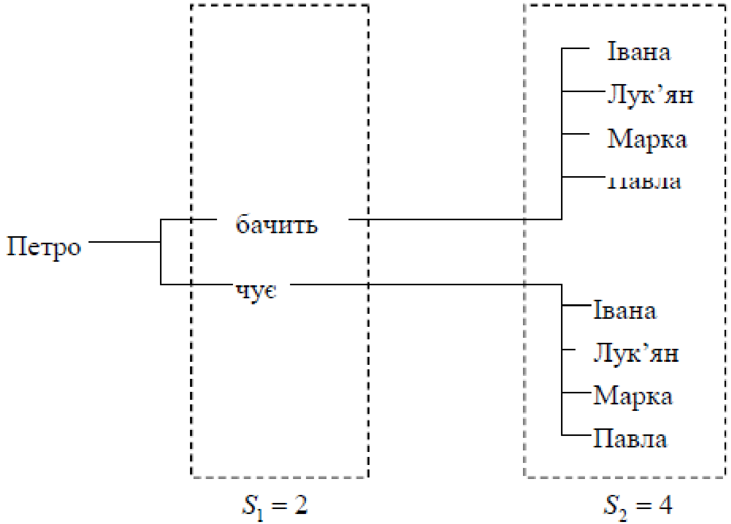
1) якщо кількість результатів $S=1$, то невизначеність $f(S)=0$;

2) якщо є два досліди, причому $S_1 > S_2$, то $f(S_1) > f(S_2)$.

Для того, щоб остаточно означити функцію $f(S)$, яка характеризує міру невизначеності, розглянемо ще один лінгвістичний експеримент.

Будуватимемо випадковим чином речення з трьох слів. Нехай перша позиція зайнята власною назвою *Петро*. Другу позицію потрібно зайняти однією з двох дієслівних словоформ *бачить* або *чує* ($S_1 = 2$), які навмання витягаються з урни.

Кінцева позиція заміщається однією з чотирьох словоформ – *Івана*, *Лук'яна*, *Марка*, *Павла* ($S_2 = 4$), котрі також навмання витягаються з урни. Цю побудову можна зобразити у вигляді такої схеми:



Невизначеність досліду, який полягає у виборі дієслівної форми, дорівнює $f(S_1)=f(2)$; невизначеність випробування, яке полягає у виборі власної назви, характеризується величиною $f(S_2)=f(4)$.

Розглянемо складний дослід, який полягає у комбінованому виборі з двох урн однієї з $S_1 \cdot S_2 = 2 \cdot 4 = 8$ послідовностей з двох слів для початкової словоформи *Петро*. Нехай перший крок полягає в утворенні словосполучення з власної назви *Петро* та однієї з двох дієслівних форм *бачить* або *чує*, що навмання витягуються з урни. Тоді кількість результатів такого досліду $S_1 = 2$. Наступний крок – доповнити утворене вже словосполучення прямим додатком – кінцева позиція речення з трьох слів заміщається однією з чотирьох словоформ, $S_2 = 4$; ці слова також навмання витягуються з урни.

Невизначеність цього складного досліду є сумою невизначеностей двох простих дослідів і характеризується рівністю $f(S_1 \cdot S_2) = f(S_1) + f(S_2)$. Ця рівність є третьою залежністю, яка характеризує співвідношення між невизначеністю досліду і кількістю його рівноможливих результатів.

Існує тільки одна функція аргументу S, яка задовольняє трьом сформульованим умовам: 1) $f(1) = 0$; 2) якщо $S_1 > S_2$, то $f(S_1) > f(S_2)$; 3) $f(S_1 \cdot S_2) = f(S_1) + f(S_2)$. Цією функцією є логарифмічна залежність

$$H = \log S, \tag{3}$$

за допомогою якої ми будемо оцінювати міру невизначеності, або *ентропію*, досліду.

Особливістю такої формули є відстороненість від семантичних, якісних, індивідуальних властивостей інформації.

Основа логарифма впливає лише на зручність обчислення.

У випадку оцінки ***ентропії:***

а) в двійкових одиницях

$$H = \log_2 S \quad \text{біт/символ}$$

б) в десяткових одиницях

$$H = \lg S \quad \text{діт/символ},$$

де $H = \log_2 S = 3{,}321 \lg S$, 1 біт≈0,3 діт

У лінгвістичних застосуваннях ентропії, здебільшого, використовують логарифми за основою 2, у зв'язку з чим вираз (3) набуває вигляду

$$H = \log_2 S\ . \tag{4}$$

Одиницею виміру ентропії є невизначеність, яку містить дослід з двома рівноймовірними результатами. Це двійкова одиниця, або ***біт***:

$$1\ \boldsymbol{біт} = \log_2 2.$$

Повернемось до розглянутого вище лінгвістичного експерименту з вибором продовжень для власної назви *Петро*. Тут невизначеність вибору дієслівної форми мови

$$\log_2 2 = 1\ \textit{біт},$$

а ентропія вибору власної назви в третій позиції складає

$$\log_2 4 = 2\ \textit{біт}.$$

Невизначеність складного досліду, який полягає в одночасному виборі присудка і прямого додатка, повинна скласти

$$\log_2 2 + \log_2 4 = 1 + 2 = 3\ \textit{біт}.$$

Дійсно,

$$\log_2 (2 \cdot 4) = \log_2 8 = 3\ \textit{біт}.$$

2.2 Комбінаторний підхід до визначення кількості інформації

Введення поняття ентропії дає можливість проводити кількісне вимірювання інформації. Дійсно, в результаті проведення досліду A ми отримаємо нові відомості, тобто деяку інформацію. Одночасно знання результату досліду знімає повністю або частково ту невизначеність, яка була до його проведення. Тому правильно припустити, що знята в результаті досліду A ентропія дорівнює кількості одержаної інформації, тобто

$$H(A) = I(A). \qquad (5)$$

З (4) та (5) випливає, що кількість інформації, отриманої від випробування з S рівноможливими результатами, визначається рівністю

$$I_0 = \log_2 S. \qquad (6)$$

Стосовно задач мовознавства? множина M, $|M| = S$, називається лінгвістичним алфавітом, а величини I_0 та H_0 – відповідно? *інформацією* та *ентропією* алфавіту.

Кількість рівноможливих результатів визначається звичайно шляхом дослідження комбінаторики елементів і зв'язків, які характеризують дане лінгвістичне явище. У зв'язку з цим уся щойно описана методика є комбінаторним підходом до визначення кількості інформації.

2.3 Вимірювання обмежень, що накладаються на застосування лінгвістичних одиниць системою і нормою мови

Хоча комбінаторний підхід дає, як правило, завищені дані про ентропію та інформацію досліду, він може бути використаний для одержання наближених оцінок тих обмежень, які накладають на застосування лінгвістичних одиниць система і норма мови. Розглянемо методику одержання таких оцінок на прикладі ланцюжків з двох букв.

Виходячи з формул розміщення з повтореннями та (6), можна стверджувати, що інформація, яка одержується з українського алфавіту (33 букви) за умови, що ніяких обмежень на утворення ланцюжків (з двох букв) не накладається і всі такі ланцюжки є рівноймовірними, дорівнює

$$I_0 = \log_2 \widetilde{A}_{33}^2 = \log_2 1089 \approx 10.089 \text{ біт}.$$

Якщо врахувати обмеження, яке полягає в тому, що наші ланцюжки не повинні містити м'якого знаку, то інформація, що міститься в одному ланцюжку з двох букв, дорівнює

$$I' = \log_2 \widetilde{A}_{32}^2 = \log_2 1024 = 10 \text{ біт}.$$

Якщо ж скласти всі ланцюжки з двох букв без повторень то за формулами розміщення та (6), то інформація від вибору одного такого ланцюжка складе

$$I'' = \log_2 A_{33}^2 = \log_2 1056 \approx 10.044 \text{ біт}.$$

Легко зауважити, що введення тих або інших обмежень призводить до зменшення кількості інформації при виборі одного ланцюжка. Ці обмеження, які ми будемо називати *структурними контекстними обмеженнями*, можна кількісно оцінити за допомогою різниці

$$I_0 - I = K , \qquad (7)$$

де I_0 – інформація алфавіту або, іншими словами, кількість інформації, котра одержується з досліду за відсутності будь-яких обмежень у комбінаториці лінгвістичних елементів і зв'язків, I – інформація за наявності обмежень, які нас цікавлять, а K – контекстна обумовленість.

Використовуючи вираз (7), неважко оцінити величину структурних обмежень, які накладаються на алфавіт українських ланцюжків з двох букв. У першому випадку ці обмеження складають

$$K_2' = \log_2 \widetilde{A}_{33}^2 - \log_2 \widetilde{A}_{32}^2 = 10.089 - 10 = 0.089 \ \textit{біт},$$

у другому випадку

$$K_2'' = 10.089 - 10.044 = 0.045 \ \textit{біт}.$$

Комбінаторні вимірювання інформації можуть бути з успіхом застосовані для оцінки „гнучкості мови”, тобто для вимірювання розгалуженості продовження тексту при заданому словнику і заданих правилах побудови речень.

2.4 Імовірнісний підхід до визначення кількості інформації

При описі комбінаторного методу для обчислення кількості інформації та ентропії ми використовували спрощення, за яким всі закінчення досліду вважались рівноймовірними. При реальних дослідженнях така ситуація практично ніколи не зустрічається. Норма мови приписує кожному лінгвістичному елементу певну ймовірність. Якщо лінгвістичне випробування передбачає нерівноймовірні результати, то, очевидно, ентропія такого досліду і отримана від нього кількість інформації будуть відрізнятись від аналогічних величин для досліду з рівноймовірними результатами.

Перехід від оцінки невизначеності і інформації досліду з рівноймовірними закінченнями до обчислення ентропії та інформації випробування з нерівноймовірними закінченнями здійснюється на основі таких міркувань.

Використовуючи відомі правила логарифмування, перепишемо (6) вигляді

$$I_0 = -\log_2(1/S). \tag{8}$$

Тут величина $1/S$ – це ймовірність p кожного закінчення досліду.

Припустимо тепер, що закінчення досліду нерівноймовірні і кожне закінчення має свою ймовірність p_i. Тоді індивідуальна кількість інформації, яка дається закінченням i при його окремій появі, дорівнює

$$I_i = -\log_2 p_i .$$

При багатократному виконанні досліду закінчення i буде відбуватись з імовірністю p_i. Тому середня кількість інформації, яка подається закінченням i при багатократному здійсненні випробування, складе

$$\tilde{I}_i = -p_i \cdot \log_2 p_i .$$

Величина $\tilde{I}_i$ визначає той вклад, котрий вносить результат i у загальну кількість інформації, яка отримується при багатократному проведенні досліду A. Що стосується загальної інформації, то вона є сумою вкладів усіх S можливих результатів і визначається наступною рівністю, яка дає оцінку інформації незалежно від її змісту.

$$I = -\sum p_i \cdot \log_2 p_i = \sum \tilde{I}_i . \tag{9}$$

Проте, інформаційні вимірювання, які ґрунтуються на обробці розподілів безумовних ймовірностей, мають у мовознавстві обмежене застосування. Справа полягає в тому, що мовні одиниці виступають у тексті як залежні лінгвістичні події, що обумовлені контекстом, а їхні ймовірності є умовними. Розподіл таких ймовірностей визначається тим положенням, яке займає дана лінгвістична одиниця в тексті. Так, наприклад, розподіл ймовірностей букв на початку слова сильно відрізняється від спектру їхніх безумовних ймовірностей.

Щодо інформації, яка одержується з даної ділянки тексту, то вона дорівнює ентропії, яка характеризує цю ділянку.

Розглянемо тепер методику обчислення інформації, яка одержується від деякого лінгвістичного досліду L, який має S результатів і здійснюється на n-й ділянці тексту за умови, що відомий ланцюжок b^{n-1} лінгвістичних елементів, який розташований перед цією ділянкою. Ланцюжок b^{n-1} розглядається як випадкова подія, яка набуває частковий вигляд i. Поява того чи іншого елементу в позиції n також розглядається як випадкова величина, яка набуває значення j_k $(1 \le k \le S)$. Для кожного значення i, яке може набути b^{n-1} відома умовна ймовірність $p\left(j_k / b_i^{n-1}\right)$ того, що L_n одержить значення j_k.

Середня умовна ентропія H_n, яка кількісно дорівнює інформації I_n, одержується в результаті усереднення ентропії, підрахованої по всіх значеннях b_i^{n-1} з вагами, які відповідають імовірностям ланцюжка b^{n-1}.

Таким чином, маємо

$$H_n = I_n = -\sum_{b_i^{n-1}} \left[p\left(b_i^{n-1}\right) \sum_{k=1}^{S} p\left(j_k / b_i^{n-1}\right) \log_2 p\left(j_k / b_i^{n-1}\right) \right]. \tag{10}$$

Рівність (10) показує, якою є в середньому міра невизначеності і кількість інформації від вибору лінгвістичного елементу в позиції n, коли відомий ланцюжок b^{n-1}.

2.5 Приріст інформації

Припустімо, що до початку процесу дослід може закінчуватись p_1 рівноімовірними результатами, жоден з яких не має переваги над іншим, а після закінчення процесу – p_2 результатами. Зміна інформації при цьому визначатиметься так:

$$\Delta I = k \ln(p_1 / p_2) = k(\ln p_1 - \ln p_2). \tag{11}$$

Якщо $\Delta I > 0$ – отримуємо приріст інформації, тобто відомості про дослід стали більш визначеними, при $\Delta I < 0$ – менш визначеними. Притому важливо, що не використовувалась явно структура досліду (механізм протікання процесу). Величина ΔI може бути інтерпретована як кількість інформації, що необхідна для переходу від одного рівня організації системи до іншого ($\Delta I > 0$ – вищий рівень, $\Delta I < 0$ – нижчий рівень).

2.6 Інформаційні виміри кодування інформації

Представимо описані вище характеристики інформації у термінах кодування інформації.

Комбінаторний підхід до визначення кількості інформації. Загальна кількість повідомлень, що не повторюються, яка може бути складена з алфавіту з m символів шляхом комбінування по n символів в повідомленні, визначається за формулою

$$N = m^n . \qquad (12)$$

Невизначеність, яка припадає на символ первинного (того, що кодується) алфавіту, складеного з рівноймовірних й взаємонезалежних символів

H=log m. $H = \log m$

13)

Введемо наступні означення.

Первинний алфавіт складається з m_1 *символів (якісні ознаки), за допомогою яких записується повідомлення, що передається.*

Вторинний алфавіт складається з m_2 *символів, за допомогою яких повідомлення трансформується в код.*

Оскільки інформація є невизначеністю, яка знімається при отриманні повідомлення, то ***кількість інформації*** може бути представлена, як добуток загальної кількості повідомлень ***k*** на середню ентропію **H,** яка припадає на одне повідомлення:

$$I = k \cdot H \ \textit{біт}.$$

Для випадків рівноймовірних та взаємонезалежних символів первинного алфавіту кількість інформації в ***k*** повідомленнях алфавіту з ***m*** символів дорівнює

$$I = k \cdot \log_2 m \ \textit{біт}.$$

Імовірнісний підхід до визначення кількості інформації. При описі комбінаторного методу для обчислення кількості інформації та ентропії ми використовували спрощення, за яким всі закінчення досліду вважались рівноймовірними. При реальних дослідженнях така ситуація практично ніколи не зустрічається. (Норма мови приписує кожному лінгвістичному елементу певну ймовірність). Якщо випробування передбачає нерівноймовірні результати, то, очевидно, ***ентропія*** такого досліду і отримана від нього кількість інформації будуть відрізнятись від аналогічних величин для досліду з рівноймовірними результатами.

Для нерівноймовірних результатів ***ентропія на символ*** алфавіту

$$H = \sum_{i=1}^{m} p_i \cdot \log_2 \frac{1}{p_i} = -\sum_{i=1}^{m} p_i \cdot \log_2 p_i \text{ біт/символ,} \tag{14}$$

а ***кількість інформації*** в повідомленні, що складається з ***k*** нерівноймовірних символів,

$$I = -k\sum_{i=1}^{m} p_i \cdot \log_2 p_i \quad \text{біт.} \tag{15}$$

При розв'язуванні задач, в яких ентропія визначається як сума добутків ймовірностей на їх логарифми, ймовірності завжди представляють групу повних подій, незалежно від того, є ті події безумовними $p(a_i)$, умовними $p(a_i/b_i)$ чи ймовірностями сумісних подій $p(a_i, b_i)$.

Кількість інформації визначається виключно характеристиками первинного алфавіту, ***об'єм*** – характеристиками вторинного алфавіту. ***Об'єм інформації*** (к-сть елементарних символів в прийнятому повідомленні) $Q = kl_{сер}$, де $l_{сер}$ – середня довжина кодових слів вторинного алфавіту.

Для рівномірних кодів (всі комбінації коду мають однакову кількість розрядів): ***Q=kn***, де ***n*** – довжина кода (к-сть елементарних посилок в коді).

Згідно (11), ***об'єм*** дорівнює ***кількості інформації*** , якщо $l_{сер} = H$, тобто у випадку максимального інформаційного навантаження на символ повідомлення. У всіх інших випадках ***I<Q*** .

ЛЕКЦІЯ 7. ІНФОРМАЦІЙНІ ВИМІРИ КОДУВАННЯ ІНФОРМАЦІЇ

1.1 УМОВНА ЕНТРОПІЯ ТА ЕНТРОПІЯ ОБ'ЄДНАННЯ

1.2 ЗАГАЛЬНА І ЧАСТКОВА УМОВНА ЕНТРОПІЯ

Поняття умовної ентропії в теорії інформації використовується при визначенні взаємозалежності між символами алфавіту, що кодується, для визначення втрат при передачі інформації по каналах зв'язку, при обчисленні ентропії об'єднання. (Суть ***взаємозалежності*** символів букв алфавіту полягає в тому, що ймовірність появи i-ї букви в будь-якому місці повідомлення залежить від того, які букви стоять перед нею і після неї, і буде відрізнятись від безумовної ймовірності p_i, яка відома зі статистичних властивостей даного алфавіту)

У всіх випадках при обчисленні умовної ентропії в тому чи іншому вигляді використовуються умовні ймовірності.

Якщо при передачі ***n*** повідомлень символ ***A*** з'явиться ***m*** разів, символ ***B*** з'явиться ***l*** разів, а символ ***A*** разом з символом ***B*** – ***k*** разів, то ймовірність появи символу ***A*** $p(A)=\frac{m}{n}$, ймовірність появи символу ***B*** $p(B)=\frac{l}{n}$; ймовірність сумісної появи символів ***A*** і ***B*** $p(AB)=\frac{k}{n}$; умовна ймовірність появи символу ***A*** відносно символу ***B*** та умовна ймовірність появи символу ***B*** відносно символу ***A***

$$p(A/B)=\frac{p(AB)}{p(B)}=\frac{k}{l}; \qquad p(B/A)=\frac{p(AB)}{p(A)}=\frac{k/n}{m/n}=\frac{k}{m}. \tag{1}$$

Якщо відома умовна ймовірність, то можна легко визначити також імовірність спільної появи символів ***A*** і ***B*,** використовуючи вираз (16)

$$p(AB)=p(B)p(A/B)=p(A)p(B/A). \tag{2}$$

Від класичного виразу формула ***умовної ентропії*** відрізняється тим, що в ній ймовірності – умовні:

$$H\left(b_j / a_i\right) = -\sum_j p\left(b_j / a_i\right) \cdot \log \; p\left(b_j / a_i\right), \tag{3}$$

$$H\left(a_i / b_j\right) = -\sum_i p\left(a_i / b_j\right) \cdot \log p \; \left(a_i / b_j\right), \tag{4}$$

де індекс ***i*** вибраний для характеристики довільного стану джерела повідомлень ***A,*** а індекс ***j*** вибраний для характеристики довільного стану адресата ***B.***

Розрізняють поняття часткової та загальної умовної ентропії. Вирази (3) та (4) представляють часткові умовні ентропії.

Загальна умовна ентропія повідомлення ***В*** відносно повідомлення ***А*** характеризує кількість інформації, що міститься в будь-якому символі алфавіту, і визначається усередненням по всіх символах, тобто по всіх станах з врахуванням ймовірності появи кожного зі станів і дорівнює сумі ймовірностей появи символів алфавіту на невизначеність, яка залишається після того, як адресат прийняв сигнал

$$H(B/A) = -\sum_i p(a_i) \cdot H(b_j/a_i) = -\sum_i \sum_j p(a_i) \cdot \; p(b_j/a_i) \log p(b_j/a_i). \tag{5}$$

Вираз (5) є загальним виразом для визначення кількості інформації на один символ повідомлення для випадку нерівномірних та взаємозалежних символів.

Оскільки $p(a_i)p(b_j/a_i)$ являє собою ймовірність спільної появи двох подій $p(a_i/b_j)$, то формулу (20) можна записати наступним чином

$$H(B/A) = -\sum_i \sum_j p(a_i/b_j) \cdot \; \log p(b_j/a_i). \tag{6}$$

Якщо ми досліджуємо канал зв'язку з боку приймача повідомлень, то з отриманням сигналу b_j припускаємо, що був відісланий якийсь із сигналів $a_1, a_2, ..., a_i, ..., a_m$. При цьому канальна матриця буде мати вигляд:

B / A	b_1	b_2	$\cdots$	b_j	$\cdots$	b_m
a_1	$p(a_1/b_1)$	$p(a_1/b_2)$	$\cdots$	$p(a_1/b_j)$	$\cdots$	$p(a_1/b_m)$
a_2	$p(a_2/b_1)$	$p(a_2/b_2)$	$\cdots$	$p(a_2/b_j)$	$\cdots$	$p(a_2/b_m)$
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
a_i	$p(a_i/b_1)$	$p(a_i/b_2)$	$\cdots$	$p(a_i/b_j)$	$\cdots$	$p(a_i/b_m)$
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
a_m	$p(a_m/b_1)$	$p(a_m/b_2)$	$\cdots$	$p(a_m/b_j)$	$\cdots$	$p(a_m/b_m)$

В цьому випадку часткова умовна ентропія

$$H\left(a_i / b_j\right) = -\sum_{i=1}^{m} p\left(a_i / b_j\right) \log p\left(a_i / b_j\right), \tag{7}$$

а загальна умовна ентропія

$$H\left(A / B\right) = -\sum_{j} p\left(b_j\right) \sum_{i} p\left(a_i / b_j\right) \log p\left(a_i / b_j\right). \tag{8}$$

Якщо задані канальна матриця вигляду $p\left(b_j / a_i\right)$ та безумовні ймовірності вигляду $p\left(a_i\right)$, то безумовні ймовірності приймача $p\left(b_j\right)$ знаходимо як $\sum_i p\left(a_i\right) p\left(b_j / a_i\right)$, тобто якщо задані безумовні ймовірності джерела та канальної матриці, то може бути розрахована ентропія приймача $H(B)=-\sum_j p\left(b_j\right)\log p\left(b_j\right)$, і навпаки, якщо задані ймовірності вигляду $p\left(b_j\right)$ та канальна матриця, яка описує канал зв'язку з боку приймача повідомлень, то $p\left(a_i\right)=\sum_j p\left(b_j\right) p\left(a_i / b_j\right)$, а значить може бути визначена ентропія джерела повідомлень $H(A)=-\sum_i p\left(a_i\right)\log p\left(a_i\right)$.

Якщо взаємозалежність пов'язує 3 елементи a_i, a_j, a_k, то умовна ентропія розраховується за формулою

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

аналогічно і для $4, 5, \ldots, n$ елементів.

1.3 СУМІСНА ПОЯВА СТАТИЧНО ЗАЛЕЖНИХ ПОВІДОМЛЕНЬ

Ентропія об'єднання використовується для розрахунку ентропії сумісної появи статично залежних повідомлень. Наприклад, передавши сто разів цифру 5 по каналу зв'язку із перешкодами, зауважимо, що цифра 5 була прийнята 90 разів, цифра 6 – 8 разів та цифра 4 – 2 рази. Невизначеність виникнення комбінацій вигляду 5 – 4, 5 – 5, 5 – 6 при передачі цифри 5 може бути описана за допомогою ентропії об'єднання. $H(A,B)$ – невизначеність того, що буде відіслано A, а прийнято B. Для набору переданих повідомлень A та прийнятих повідомлень B ентропія об'єднання представляє собою суму вигляду

$$H(A,B) = -\sum_{i}\sum_{j} p(a_i,b_j)\log_2 p(a_i,b_j) \quad \textit{біт/два} \quad \textit{символа}. \tag{9}$$

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

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

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

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

Ентропія об'єднання може бути обчислена за допомогою матриці вигляду:

$$p(a_i,b_j)=\left|\begin{matrix} p(a_1,b_1)\, p(a_1,b_2) \ldots p(a_1,b_m) \\ p(a_2,b_1)\, p(a_2,b_2) \ldots p(a_2,b_m) \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ p(a_m,b_1)\, p(a_m,b_2) \ldots p(a_m,b_m) \end{matrix}\right|.$$

Така матриця володіє наступною властивістю:

$$\sum_i p\left(a_i, b_j\right) = p\left(b_j\right),$$

$$\sum_j p\left(a_i, b_j\right) = p\left(a_i\right),$$

при цьому $\sum_i p\left(a_i\right) = \sum_j p\left(b_j\right) = 1$. Ця властивість, в свою чергу, дозволяє розраховувати ентропію як джерела, так і приймача повідомлень безпосередньо по канальній матриці:

$$H\left(A\right) = -\sum_i \sum_j p\left(a_i, b_j\right) \log \sum_j p\left(a_i, b_j\right). \tag{10}$$

$$H\left(B\right) = -\sum_i \sum_j p\left(b_j, a_i\right) \log \sum_i p\left(b_j, a_i\right). \tag{11}$$

Сумування проводимо по i та j, так як для того, щоб знайти безумовні ймовірності, необхідно сумувати їх за одною координатою (маючи на увазі матричне представлення ймовірностей), а для знаходження H сумування проводиться по іншій координаті.

Умовні ймовірності вигляду $p\left(a_i / b_j\right)$ та $p\left(b_j / a_i\right)$ розраховуються як:

$$p\left(a_i / b_j\right)=\frac{p\left(a_i, b_j\right)}{\sum_i p\left(a_i, b_j\right)},\ p\left(b_j / a_i\right)=\frac{p\left(a_i, b_j\right)}{\sum_j p\left(a_i, b_j\right)}.$$

Кількість інформації на символ повідомлення, переданого по каналу зв'язку, в якому вплив перешкод описується за допомогою ентропії об'єднання, підраховується наступним чином:

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

1.4 ІНФОРМАЦІЙНІ ВТРАТИ В КАНАЛАХ ІЗ ШУМОМ

Поняття ***загальної і часткової умовної ентропії*** широко використовується для обчислення інформаційних втрат в каналах зв'язку з шумом.

В загальному випадку, якщо ми передаємо ***m*** сигналів ***A*** і очікуємо отримати ***m*** сигналів ***B***, вплив перешкод в каналі зв'язку повністю описується ***канальною матрицею***, яка наведена нижче

B \ A	b_1	b_2	…	b_j	…	b_m
a_1	$p(b_1/a_1)$	$p(b_2/a_1)$	…	$p(b_j/a_1)$	…	$p(b_m/a_1)$
a_2	$p(b_1/a_2)$	$p(b_2/a_2)$	…	$p(b_j/a_2)$	…	$p(b_m/a_2)$
…	…	…	…	…	…	…
a_i	$p(b_1/a_i)$	$p(b_2/a_i)$	…	$p(b_j/a_i)$	…	$p(b_m/a_i)$
…	…	…	…	…	…	…
a_m	$p(b_1/a_m)$	$p(b_2/a_m)$	…	$p(b_j/a_m)$	…	$p(b_m/a_m)$

Ймовірності, які розміщені по діагоналі, визначають правильний прийом, решта – хибний. Значення цифр, які заповнюють колонки канальної матриці, звичайно зменшуються по мірі віддалення від головної діагоналі, і при повній відсутності перешкод, всі, крім цифр на головній діагоналі, дорівнюють нулю.

Втрати інформації, які припадають на частку сигналу a_1, описуються за допомогою ***часткової умовної ентропії*** виду

$$H(b_j / a_1) = -\sum_j p(b_j / a_1) \cdot \log p(b_j / a_1). \tag{12}$$

Сумування відбувається по *j*, оскільки *i*-ий стан (в даному випадку перший) залишається постійним.

Щоб врахувати втрати при передачі всіх сигналів по даному каналу зв'язку, потрібно просумувати всі часткові умовні ентропії (тобто провести подвійне сумування по i та по j). При цьому в випадку рівноймовірних появ сигналів на виході джерела повідомлень

$$H(B/A) = -\frac{1}{m}\sum_i \sum_j p(b_j/a_i)\log p(b_j/a_i), \tag{13}$$

(ділимо на m, оскільки ентропія є невизначеність на один символ).

У випадку нерівноймовірної появи символів джерела повідомлень потрібно врахувати ймовірність появи кожного символа, помноживши на неї відповідну часткову умовну ентропію.

При цьому загальна умовна ентропія:

$$H(B/A) = -\sum_i p(a_i)\cdot\sum_j p(b_j/a_i)\log p(b_j/a_i). \tag{14}$$

Якщо перешкод немає або їх рівень настільки низький, що вони не в стані знищити сигнал чи імітувати корисний сигнал при відсутності передачі, то при передачі a_i ми будемо твердо впевнені, що отримаємо b_j – сигнал, який відповідає переданому a_i-му сигналу. Події A та B статистично жорстко пов'язані, умовна ймовірність максимальна $p\left(b_j / a_i\right)=1$, а умовна ентропія

$$H\left(A / B\right)=-\sum_{i=1}^{m} p\left(b_j / a_i\right)\log_2 p\left(b_j / a_i\right)=0, \tag{15}$$

оскільки $\log_2 p\left(b_j / a_i\right)=\log_2 1=0$.

В цьому випадку кількість інформації, яка вміщується в прийнятому наборі повідомлень B, рівна ентропії переданих повідомлень набору A, тобто $I\left(B, A\right)=H\left(A\right)$

При високому рівні перешкод будь-який із прийнятих сигналів b_j може відповідати будь-якому переданому сигналу a_i, статистичний зв'язок між переданими та прийнятими сигналами відсутній. В цьому випадку ймовірності $p(a_i)$ та $p(b_j)$ є ймовірностями незалежних подій і $p(b_j/a_i) = p(b_j)$, $p(a_i/b_j) = p(a_i)$.

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

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

Оскільки $\sum_j p(b_j)=1$, тобто умовна ентропія рівна безумовній, а кількість інформації, яка вміщується в B, відносно A рівна нулю:

$I(A,B)=H(A)-H(A/B)=0$.

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

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

Не дивлячись на те, що частина інформації змінюється через перешкоди, між прийнятими та переданими повідомленнями існує статична взаємозалежність. Це дозволяє описувати інформаційні характеристики реальних каналів зв'язку за допомогою ентропії об'єднання статистично залежних подій. Так як

$$H(A,B)=H(A)+H(B/A)=H(B)+H(A/B), \qquad (16)$$

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

$$I(B,A)=H(A)+H(B)-H(B,A), \qquad (17)$$

а з використанням умовної ентропії

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

Для розрахунку середньої кількості інформації, що вміщується в прийнятому наборі повідомлень B відносно переданого набору повідомлень A в умовах дії перешкод, користуються наступними виразами, виділеними безпосередньо із виразу (17):

$$I(B,A)=\sum_i\sum_j p(a_i)p(b_j/a_i)\log_2\frac{p(b_j/a_i)}{p(b_j)}, \tag{18}$$

$$I(A,B)=\sum_i\sum_j p(b_j)p(a_i/b_j)\log_2\frac{p(a_i/b_j)}{p(a_i)}, \tag{19}$$

$$I(B,A)=I(A,B)=\sum_i\sum_j p(a_i,b_j)\log_2\frac{p(b_j/a_i)}{p(b_j)}= \tag{20}$$

$$=\sum_i\sum_j p(b_j,a_i)\log_2\frac{p(a_i/b_j)}{p(a_i)}=\sum_i\sum_j p(a_i,b_j)\log_2\frac{p(a_i,b_j)}{p(a_i)p(b_j)},$$

Для розрахунків часто зручно застосовувати вирази (18-20) у вигляді

$$I(A,B)=\sum_j p(b_j)\sum_i\left[p(a_i/b_j)\log_2 p(a_i/b_j)-p(a_i/b_j)\log_2 p(a_i)\right],$$

$$I(B,A)=\sum_i p(a_i)\sum_j\left[p(b_j/a_i)\log_2 p(b_j/a_i)-p(b_j/a_i)\log_2 p(b_j)\right],$$

$$I(A,B)=I(B,A)=\sum_i\sum_j p(a_i,b_j)\log_2 p(a_i,b_j)-$$

$$-\sum_i\sum_j p(a_i,b_j)\log_2 p(a_i)\,p(b_j)$$

Для повного та всебічного опису каналу зв'язку необхідно задати: канальну матрицю виду $p(a_i / b_j)$ та безумовні ймовірності виду $p(b_j)$ або канальну матрицю виду $p(b_j / a_i)$ та безумовні ймовірності виду $p(a_i)$, або канальну матрицю виду $p(a_i, b_j)$. В останньому випадку сума значень матриці по стовпцях дає безумовні ймовірності виду $p(b_j)\left(\sum_j p(b_j) = 1\right)$, а сума по рядках дає безумовні ймовірності виду $p(a_i)\left(\sum_i p(a_i) = 1\right)$. Умовні ймовірності можуть бути знайдені із 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 ймовірності, можна знайти $H(A)$, $H(B)$, $H(A/B)$ та $H(B/A)$.

Якщо рівень перешкод настільки великий, що з рівною ймовірністю можна очікувати перехід будь-якого символу джерела повідомлення в довільний символ первинного алфавіту, то ентропія каналу зв'язку буде рівна $\log_2 m$, а кількість інформації $I = H(A) - \log_2 m \leq 0$, при цьому значення I може бути від'ємною величиною, що означає, що канал зв'язку вносить дезінформацію.