

doi: <https://doi.org/10.15407/dopovidi2018.06.032>

УДК 517.58/5892

В.М. Заяць^{1,2}, М.М. Заяць³

¹ Університет технологічно-природничий, Бидгощ, Польща

² Національний університет водного господарства та природокористування, Рівне

³ Національний університет “Львівська політехніка”

E-mail: v.m.zaiats@utp.edu.pl, v.m.zaiats@nuwm.edu.ua, marijazajats@gmail.com

Образний підхід до кількісної оцінки цінності інформації

Представлено членом-кореспондентом НАН України А.М. Довбнею

На основі теорії образів запропоновано новий підхід, який дає змогу кількісно оцінити цінність інформації. Розглянуто різні підходи до визначення та обчислення основних понять теорії інформації, зокрема, кількості інформації та оцінки її цінності, виходячи із статистичних міркувань (класичний підхід), теорії алгоритмів (алгоритмічний підхід) та теорії розпізнавання образів (образний підхід). Проведено їх аналіз, відзначено межі їх використання та області ефективного застосування.

Ключові слова: теорія інформації, кількість інформації, цінність інформації, ймовірність, алгоритм, образ.

Точні формулювання наукових термінів та визначень ґрунтуються на двох основних підходах. Перший підхід (теоретичний), як правило, застосовується у математичних дисциплінах, де визначення ґрунтуються на формулюванні основних постулатів або гіпотез. Всі інші складніші сутності виводяться із цих введених тверджень і не порушують логіку міркувань. Другий підхід до формулювання визначень характерний для прикладних наук і часто називається операційним (практичним). За цим підходом вважається доцільним вводити у наукознавство ті величини (поняття, параметри), які можуть бути визначені експериментально або оцінені опосередковано через їх вплив на поведінку досліджуваної системи. Нововведені терміни, які не піддаються операційному визначенню і не мають практичної цінності з плином часу виключаються з наукового словника. Зокрема, так було з поняттям ефіру, якого теорія відносності залишила будь-якого змісту і він вийшов з ужитку в науковій мові, незважаючи на чисельні публікації по цій проблемі.

Мета даної роботи — запропонувати новий підхід до кількісної оцінки цінності інформації на основі теорії образів, порівняти його з існуючими методами та підходами до оцінки цінності інформації, вказати межі їх застосувань, переваги та недоліки, визначити області доцільного застосування.

© В.М. Заяць, М.М. Заяць, 2018

Будь-яка фізична система не є повністю визначеною. Ми маємо дані про деякі макроскопічні змінні системи, але не можемо точно визначити стани і швидкості всіх неподільних елементів (атомів), що утворюють систему. Мірою невизначеності станів системи є поняття ентропії. Ентропія виражає кількість відсутньої інформації про ультрамікроскопічну структуру системи. Ці два поняття кількості інформації і ентропії слід розглядати і трактувати сумісно. Дана точка зору сформульована як негентропійний принцип інформації у роботі [1]. Суть цього принципу зводиться до твердження: отримання інформації про систему пов'язане із зменшенням її ентропії. Наскільки зменшиться ентропія системи, настільки зросте інформація про неї. З цієї точки зору інформацію можна визначити як негативну (від'ємну) ентропію системи.

На основі теорії розпізнавання образів у роботі запропоновано підхід до оцінки цінності інформації, який розширює область його прикладних застосувань.

Класичний підхід. У класичній теорії інформації, яка розроблена в фундаментальних роботах Р. Хартлі [2] та К. Шеннона [3], основні поняття ґрунтуються на теорії ймовірностей. Згідно з теорією Р. Хартлі, якщо система може перебувати в N станах, то повна кількість інформації про неї визначається як двійковий логарифм від числа цих станів:

$$I = \text{Log}_2 N. \quad (1)$$

При цьому одиницею вимірювання кількості інформації є один біт. Таким чином, біт інформації можна визначити як кількість інформації, що міститься у повідомленні з двома можливими станами. Зауважимо, що ймовірності появи цих повідомлень можуть бути різними. Якщо априорі невідомо число можливих станів системи, але кожен із них формується за допомогою n різноякісних ознак, можна з яких може бути на одному з k розрядів, то скориставшись формулою розміщень C_k^n та застосувавши формулу Стірлінга для обчислення логарифма від факторіала

$$\text{Ln } X! \approx X(\text{Ln } X - 1),$$

яка є справедливою для великих значень X , отримуємо формулу Шеннона [3] для вирахування кількості інформації:

$$I = -k \sum_i^n p_i \text{Log}_2 p_i, \quad (2)$$

де p_i — ймовірність появи i -ї якісної ознаки. У випадку, якщо ймовірності появи кожної із ознак є рівноімовірними, то $p_i = \frac{1}{n}$ і з (2) ми отримуємо

$$I = -k \text{Log}_2 \frac{1}{n} = k \text{Log}_2 n, \quad (3)$$

що відповідає формулі (1), оскільки $N = n^k$.

Таким чином, формули (1)–(3) дозволяють обчислювати кількість інформації як при рівномірній, так і при нерівномірній ймовірності появи якісних ознак незалежно від якості інформації, що отримана (повідомлення, звук, зображення чи сигнал будь-якої природи). Але при цьому не враховується цінність інформації, яка є значною мірою суб'єктивною характеристикою і залежить від цілей та уподобань користувача. По суті справи це є та плата

за можливість обчислення кількості інформації, виходячи із статистичних міркувань, незалежно від її якості і цінності.

Зазначимо, що кількість корисної інформації є завжди величиною додатною. У випадку взаємного впливу однієї системи на іншу з'являється умовна невизначеність станів системи, яка може призвести до зміни знака у визначенні кількості інформації. Тоді маємо справу з іншою якістю інформації – дезінформацією або хибною інформацією.

Алгоритмічний підхід. Принципово інший підхід до визначення кількості інформації запропонований російським вченим А.М. Колмогоровим [4]. Його алгоритмічна теорія інформації ґрунтується на понятті складності алгоритму перетворення одного об'єкта в інший. За такого підходу принциповим є встановлення взаємного зв'язку між об'єктами, які досліджуються та довжиною програми, що їх опрацьовує. Кількість інформації за теорією алгоритмів перетворення одних об'єктів в інші визначається як довжина програми, що забезпечує можливість перетворення об'єкта A в об'єкт B :

$$I = f[G(A, B)], \quad (4)$$

де G – програма перетворення об'єкта A в об'єкт B ; f – функція, що визначає довжину програми перетворення машинних кодів в біти.

Зазначимо, що кількість інформації за такого підходу має значну залежність від вибору структурного елемента перетворення. Кількість інформації максимальна, якщо в якості елемента перетворення вибрано піксель, а мінімальна при виборі в ролі елемента цілої літери, а також буде проміжною при виборі частини літери елементом перетворення. Обчислити ж кількість інформації для окремо взятої літери цілком неможливо, оскільки для реалізації алгоритму (8) необхідно два об'єкти. Порівнювати ж літеру з нею самою не має сенсу, оскільки довжина програми перетворення у цьому випадку буде рівною нулю.

Таким чином, при алгоритмічного підході маємо такі недоліки:

а) при обчисленні кількості інформації як параметр використовується довжина програми перетворення, яка суттєво залежить від структури елементів, що дають можливість перетворити один об'єкт в інший. Чим дрібнішою є вибрана структура елементів, тим довшою стає програма перетворення для одних і тих же об'єктів;

б) одна і та ж сама програма перетворення може бути використана для опрацювання цілого набору об'єктів, аналіз яких потребує виконання різної кількості комп'ютерних команд, в той час як довжина програми залишається незмінною;

в) не визначений метод для вимірювання кількості інформації при розгляді окремого об'єкта, оскільки згідно з формулою (4) їх потрібно два;

г) втрачається властивість адитивності інформації при розгляді взаємодіючих систем, що значно ускладнює їх аналіз.

Ці недоліки характерні і для семантичного підходу при обчисленні інформації, який ґрунтується на опрацюванні логічних тверджень:

$$I = \text{Log}_2 L(O), \quad (5)$$

де L – функція, що залежить від кількості станів у логічних твердженнях (змінних, фактах, правилах); O – логічне твердження або предикат.

Обчислення кількості інформації в окремих літерах чи словах при підході (5) неможливе в принципі, оскільки ні окремі літери, ні окремі слова не є логічними твердженнями.

Зазначені недоліки усунуті у підході, який ґрунтується на застосуванні методів теорії розпізнавання і оперує із поняттям інформації та кількості інформації, які відмінні від класичних означень, отриманих на основі теорії ймовірності.

Образний підхід. Класична теорія інформації, розроблена К. Шенноном, на сьогодні є досконалим універсальним апаратом при розв'язанні задач, пов'язаних з кодуванням інформації, її перетворенням та оптимальною передачею каналами зв'язку на великі відстані. Обмеження цієї теорії в тому, що вона ніяким чином не враховує семантику (спосіб утворення) інформації та цілком ігнорує людський фактор у формуванні інформації, тобто цілком ігнорується поняття цінності інформації.

Алгоритмічна теорія, що ґрунтується на понятті складності алгоритму, зробила крок у напрямі врахування способу утворення інформації. Незважаючи на недоліки цієї теорії, які вказані у попередньому розділі, її основні положення використані в образній концепції теорії інформації [5, 6]. Згідно з цією концепцією під інформацією слід розуміти розпізнані образи, які зберігаються у пам'яті комп'ютера або будь-якої іншої кібернетичної машини. Образом вважається сигнал, який записаний у сенсорну пам'ять скануючих пристроїв кібернетичної машини. Таким чином, для отримання інформації слід реалізувати процедуру розпізнавання вхідного об'єкта — образу на основі його відображення Q — певного еталону цього образу, який створений на основі домовленостей між відправником та приймачем повідомлень.

Кількість інформації I_o , що міститься в деякому образі, який отриманий і успішно розпізнаний кібернетичною машиною, можна визначити за формулою

$$I_o = q^{-1}F[G(Q)], \quad (6)$$

де q — ймовірність правильного розпізнавання образу; F — функція отримання довжини розгорнутої (з урахуванням циклів) програми розпізнавання образу; G — довжина програми розпізнавання образу, яка виражена в машинних командах; Q — відображення реального об'єкта.

Очевидно, при підході (6) з'являється можливість обчислення кількості інформації складно структурованих образів, які можуть бути окремими словами, реченнями або текстами чи малюнками. При цьому кількість інформації суттєво залежатиме від довжини розгорнутої програми розпізнавання, ймовірності правильного розпізнавання та функціональних можливостей кібернетичної машини, яка реалізує процедуру розпізнавання. Для об'єктивного обчислення кількості інформації згідно з формулою (6) слід враховувати таке:

а) програми розпізнавання образів мають бути оптимальними щодо власного розміру, швидкодії та функціональних можливостей;

б) збільшення кількості операцій (команд), які потрібно виконати машині для успішного розпізнавання, призводить до збільшення інформації, яку отримуємо від образу;

в) чим меншою є ймовірність правильного розпізнавання образу, тим більшу кількість інформації він має;

г) кількість інформації у незнаковому повідомленні дорівнює довжині цього повідомлення, вираженого в бітах, помноженій на довжину програми розпізнавання одного біта цього повідомлення і розділеній на ймовірність правильного розпізнавання повідомлення.

Перші експерименти [6] підтвердили право на існування такого підходу до обчислення кількості інформації, хоча його ефективність може виявитися лише в процесі постановки численних експериментів з проведенням розпізнавання різноструктурованих об'єктів. Недоліки такого підходу такі:

а) значною мірою суб'єктивна оцінка кількості інформації, яку несе у собі образ, що зумовлено як технічними можливостями кібернетичної машини, так і якістю програми розпізнавання;

б) втрачена властивість адитивності кількості інформації, яка присутня у класичній теорії інформації;

в) надмірна громіздкість формули (6), що ускладнює проведення аналітичних оцінок.

Для усунення другого недоліку видається доцільним обчислення кількості семантичної (образної) інформації проводити у логарифмічному масштабі:

$$I_o = -\text{Log}_2 q + \text{Log}_2 f(O)], \quad (7)$$

де f – довжина програми розпізнавання образів, виражена у кількості операцій (машинних команд), необхідних для проведення успішного розпізнавання образу.

Новий підхід до вирахування цінності інформації. Стосовно цінності інформації, то про неї можна говорити за потреби досягнення певної мети після одержання інформації користувачем, тобто забезпечення реалізації цільової функції. У роботі [5] А.А. Харкевичем запропоновано цінність інформації обчислювати за формулою

$$F = \text{Log}_2 \frac{p_0}{p_i}, \quad (8)$$

де p_0 – ймовірність правильного розв'язання проблеми до отримання інформації; p_1 – ймовірність правильного розв'язання проблеми після отримання інформації. Такий підхід має право на існування, хоча і викликає ряд сумнівів щодо його ефективності та доцільності застосування. По-перше, одиницею вимірювання цінності інформації при такому підході є біт, як і в випадку вирахування кількості інформації. Очевидно, при введенні нової величини повинна з'явитися і нова розмірність або обчислення слід проводити в безрозмірних одиницях. По-друге, формула (8) не може претендувати на об'єктивність, оскільки оцінки p_0 і p_1 будуть проводитися користувачем. По-третє, методика оцінки цих ймовірностей не є очевидною. По-четверте, цінність інформації є динамічною величиною [7] і в міру надходження інформації мала би змінюватися. Очевидно, при вирахуванні цінності інформації в ролі цільової функції для досягнення поставленої мети слід вибрати задоволення певних потреб користувача (матеріальних, духовних, естетичних, смакових, пізнавальних та інших) або виконання певних дій.

Найбільш доцільним з урахуванням розмаїття потреб користувача видається підхід при якому цінність інформації будемо обчислювати у відсотках: 100 % – за умови цінності інформації; 0 % – за умови, якщо цільова функція не досягнута. Таким чином, цінність інфор-

мації F для сформованого поточного значення функції цілі Z та досягнутої ефективності E після отримання повідомлення на даний момент часу можна задати у вигляді правила:

$$F = \frac{100\%, \text{ якщо } Z - E = 0}{0\%, \text{ якщо } Z - E > 0}. \quad (9)$$

Для підвищення точності обчислення цінності інформації доцільно мати сформовану цільову функцію і на наступні моменти часу та розширити шкалу обчислення цінності інформації з певним кроком. Перша умова значно підвищує цінність інформації після отримання всього повідомлення, а друга може бути реалізована за формулою:

$$F = \frac{E}{Z} \cdot 100\%. \quad (10)$$

Формула (10) є уточненням (9), що підтверджено розглядом конкретних прикладів з розгляду систем масового обслуговування та теорії ігор [9]. Отримані результати є свідченням доцільності прикладного використання запропонованого підходу кількісної оцінки цінності інформації незалежно ні від її змісту, ні від способу подання. На прикладі системи масового обслуговування показано, що якщо формула (9) дає змогу оцінити обіг товару і його вартість за рік, то (10) забезпечує масштабування процесу з довільним кроком, що дозволяє оптимізувати процес випуску товару і здійснювати контроль фінансових витрат практично після кожної фінансової операції.

Очевидно, для реалізації описаного підходу щодо обчислення цінності інформації доцільно використовувати декларативні мови програмування (Лісп, Пролог або їх модифікації залежно від конкретної предметної задачі) [7], які найбільш вдало пристосовані для реалізації функцій вигляду (6)–(10), які можуть бути як аналітичними, так і описовими (функціонали, правила логіки, нечіткі множини), що дозволяє розв'язувати задачі, пов'язані з якісним розпізнаванням та аналізом об'єктів складної структури (розпізнавання почерку, рукописного тексту, психофізіологічного стану особи, побудова та аналіз систем зберігання, опрацювання, захисту інформації, автоматизованого доведення теорем, екологічного моніторингу та прийняття рішень) та реалізації відповідних до конкретної прикладної чи наукової задачі цільових функцій.

Таким чином, авторами запропоновано новий підхід до оцінки цінності інформації на основі теорії розпізнавання образів, який розширює область його застосування та може бути успішно реалізований при використанні декларативних мов програмування [7] або універсальних мов моделювання (UML) [8]. Відзначені особливості його застосування та напрями подальшого удосконалення.

У роботі авторами висвітлено три основні підходи до оцінки кількості інформації та оцінки її цінності: класичний, алгоритмічний та образний. Наведена порівняльна характеристика наведених підходів, вказані переваги та обмеження кожного з них та визначені перспективи їх використання.

Подальшого розвитку запропонованого підходу можна досягнути при проведенні статистичних досліджень конкретних прикладних задач, пов'язаних з необхідністю оцінки як кількості, так і цінності інформації, отриманої від їх застосування.

Запропонований підхід доцільно використовувати в прикладних задачах, математичне описання яких складне або й цілком неможливе. Такий підхід сприятиме розвитку як власне методів теорії розпізнавання та ідентифікації, так і теорії інформації та кодування.

ЦИТОВАНА ЛІТЕРАТУРА

1. Бриллюэн Л. Наука и теория информации. Москва: Госиздат, 1960. 392 с.
2. Хартли Р.В. Теория информации и её приложения. Москва: Физматгиз, 1959. 356 с.
3. Шеннон К. Работы по теории информации и кибернетике. Москва: Изд-во иностран. л-ры, 1963. 286 с.
4. Колмогоров А.Н. Три подхода к определению понятия “количество информации”. *Проблемы передачи информации*. 1965. 1, вып. 1. С. 63–67.
5. Харкевич А.А. О ценности информации. *Проблемы кибернетики*. 1960. Вып. 4. С. 53–57.
6. Партико З.В. Образна концепція теорії інформації. Львів: Вид-во ЛНУ ім. І.Франка. 2001. 98 с.
7. Заяць В.М., Заяць М.М. Логічне та функціональне програмування. Навчальний посібник. Кам'янець-Подільський: Рута. 2016. 400 с.
8. Erich Gamma, Richard Helm, Ralih Johnson, John Vlissides. *Wzorce projektowe*. 2012. 270 с.
9. Заяць В.М., Рибицька О.М. Приховані можливості математики для статистичної обробки інформації. Матеріали Міжнар. наук.-техн. конф. “Системний аналіз та інформаційні технології” (SAIT-2013), 27-31 травня 2013 р., НТТУ “КПІ”. Київ. 2013. С. 317–318.

Надійшло до редакції 09.11.2017

REFERENCES

1. Brillouin, L. (1960). *Science and Information Theory*. Moscow: Gosizdat (in Russian).
2. Hartley, R. V. (1959). *Information Theory and its Applications*. Moscow: Fismatgiz (in Russian).
3. Shannon, K. (1963). *The Works on the Theory of Information and Cybernetics*. Moscow (in Russian).
4. Kolmogorov, A. N. (1965). The Three Approach to the Definition of the Concept of “amount of information.” *Problems of Information Transfer*. 1. Iss. 1, pp. 63-67 (in Russian).
5. Kharkevich, A. A. (1960). About the value of information. *Problems of cybernetics*. Iss. 4, pp. 53-57.
6. Partiko, Z. V. (2001). The figurative concept of information theory. Lviv: LNU ім. І. Franko. 98 p. (in Ukrainian).
7. Zayats, V. M. & Zayats, M. M. (2016). Logical and functional programming. Tutorial. Kamyanets-Podilsky: Ruta (in Ukrainian).
8. Erich, Gamma, Richard, Helm, Ralih, Johnson & John, Vlissides. (2012). *Wzorce projektowe* (in Poland).
9. Zaiats, V. M. & Rybitskay, O. M (2013). The hidden possibilities of mathematics for the statistical processing of information. *Materials Int. Sci. Technical Conf. “System Analysis and Information Technologies” (SAIT-2013)*, May 27-31, 2013, NTUU “KPI”. Kiev (in Ukrainian).

Received 09.11.2017

В.М. Заяць^{1,2}, М.М. Заяць³

¹ Университет технологий и науки, Быдгощ, Польша

² Национальный университет водного хозяйства и природопользования, Ровно

³ Национальный университет “Львовская политехника”

E-mail: v.m.zaiats@utp.edu.pl, v.m.zaiats@nuwm.edu.ua, marijazajats@gmail.com

ОБРАЗНЫЙ ПОДХОД К КОЛИЧЕСТВЕННОЙ ОЦЕНКЕ ЦЕННОСТИ ИНФОРМАЦИИ

На основе теории образов предложен новый подход, который позволяет количественно оценить ценность информации. Рассмотрены различные подходы к определению и вычислению основных понятий теории информации, в частности, количества информации и оценки ее ценности, исходя из статистических соображений (классический подход), теории алгоритмов (алгоритмический подход) и теории распознавания образов (образный подход). Проведен их анализ, отмечены пределы их использования и области эффективного применения.

Ключевые слова: теорія інформації, кількість інформації, цінність інформації, вероятність, алгоритм, образ.

V.M. Zaiats^{1,2}, M.M. Zaiats³

¹ University of Technology and Science, Bydgoszcz, Poland

² National University of Water Management and Nature Management, Rivne

³ National University "Lviv Polytechnic"

E-mail: v.m.zaiats@utp.edu.pl, v.m.zaiats@nuwm.edu.ua, marijazajats@gmail.com

FIGURATIVE APPROACH TO THE QUANTITATIVE EVALUATION OF THE VALUE OF INFORMATION

On the basis of the theory of pattern recognition, a new approach is proposed that allows one to quantify the value of information. Different approaches to the determination and calculation of basic concepts of information theory are considered, in particular, the amount of information and evaluation of its value, based on statistical considerations (the classical approach), the theory of algorithms (algorithmic approach), and the theory of pattern recognition (figurative approach). Their analysis is carried out, and the limits of their use and the fields of effective application are marked.

Keywords: information theory, information quantity, information value, probability, algorithm, pattern recognition.