

# Застосування поняття ентропії для чисельної оцінки образного сенсу вербальних конструкцій

Бісікало О. В., Кондратюк Н.В.

ВНТУ, obisikalo@gmail.com, <http://aivt.inaeksu.vntu.edu.ua/ksklad/1153.html>

*The associative network of linguistic images as formal model of knowledge base of verbal constructions is examined in the article. The approach to a numerical information estimation of the notion of figurative sense is suggested. The upper limit of correlation of units of figurative sense and entropy is substantiated.*

## ВСТУП

Поняття інформації, як і поняття знань не мають однозначного трактування, що підтверджується існуванням значної кількості різних визначень. Цінність застосування цих понять у сучасних інформаційних технологіях базується на формальних обмеженнях і, головне, кількісних оцінках існуючих баз даних та баз знань. Зрозуміло, що найбільш загальний характер має класична міра інформації К. Шеннона, в основу якої покладено поняття ентропії [1]. Проте оцінка знань у вигляді наукового тексту або бази знань в одиницях інформації виглядає неінформативно чи, навіть, незрозуміло. Тому, в залежності від типу бази знань, використовують такі показники, як кількість аксіом, правил на зразок ЯКЩО-ТО, вузлів семантичної мережі, фреймів тощо.

Шлях від загального поняття інформації до більш складного поняття знань проходить через накладення певних формальних обмежень. На основі моделі асоціативного образного мислення людини [2] було формалізовано поняття образного сенсу з визначенням відповідної одиниці *Sav* (Синтагматичної асоціації вага) [3]. Мета дослідження полягає в отриманні чисельної оцінки образного сенсу вербальних конструкцій на основі поняття ентропії.

## АСОЦІАТИВНА МЕРЕЖА МОВНИХ ОБРАЗІВ

Вважатимемо, що деяка система  $S$  здатна розпізнавати окремі образи з нескінченної множини  $I = \{i_1, i_2, \dots, i_n, \dots\}$  аналогічно тому, як людина розпізнає гештальт.  $S$  також може сприймати асоціативні зв'язки між парами образів як елементи множини  $\omega \in \Omega$ , де  $\Omega \subseteq I \times I$  – довільна множина упорядкованих пар. Образною конструкцією (ОК) будемо вважати будь-яку підмножину  $\gamma \subseteq \Omega$ , яка є елементом  $\mathbf{F}$  –  $\sigma$ -алгебри підмножин з  $\Omega$ .

Нехай система  $S$  сприймає інформацію з зовнішнього світу виключно у вигляді ОК, з яких розрізнятимемо послідовність вхідних подій  $X = \{x_1, x_2, \dots\}$ , де  $x_i \in \mathbf{F}$ . Внаслідок цього формується база знань системи як семантична мережа, що задається матрицею  $A_{\mathcal{O}}$ . Надалі вона, у зв'язку з вербальним характером [2] вхідної інформації системи, називатиметься асоціативною мережею мовних образів (АММО).

Задамо деяку АММО на певний момент часу такими параметрами:  $k_{lg}$  – кількість виявлених системою зв'язків між  $l$ -м та  $g$ -м образами,  $m$  – кількість ненульових елементів матриці  $A_{\mathcal{O}}$ . Маємо статистичну оцінку математичного сподівання кількості повторень одного зв'язку як  $\lambda = k_{\Sigma} / m$ , де  $k_{\Sigma} = \sum_{l=1}^n \sum_{g=1}^n k_{lg}$ . Тоді образний сенс пари  $(l, g)$  нормується

сигмоїдальною функцією як  $\mu_Q(<i_l, i_g >) = 1/(1 + e^{-k_{lg} + \lambda})$  [4], що дозволяє знайти оцінку його середнього значення для всієї АММО

$$\overline{\mu_Q} = \frac{1}{m} \sum_{j=1}^m \mu_{Qj} = 0,5 \text{ [Cав]}. \quad (1)$$

#### ІНФОРМАЦІЙНА ОЦІНКА ОБРАЗНОГО СЕНСУ

Одиниця образного сенсу розміром один *Cав* відповідає  $\mu_Q(<i_l, i_g >) = 1$ . У той же час факт появи на вході *S* кожної *j*-ї пари мовних образів з імовірністю  $p_j$  дозволяє оцінити ентропію системи. Для отримання верхньої межі ентропії будемо вважати, що ОК складається з незалежних пар образів, хоча в реальних природно-мовних конструкціях це не зовсім так. Відомо, що в цьому випадку загальна ентропія або кількість інформації [1] системи *S* дорівнює

$$H = - \sum_{j=1}^m n_j \cdot \log p_j, \quad (2)$$

де значення  $n_j$  відповідає  $k_{lg}$  як кількості зв'язків між *l*-м та *g*-м образами.

Також можна визначити середню ентропію, що припадає на одну пару. З цією метою розділимо (2) на  $k_{\Sigma}$ :

$$H_1 = - \sum_{j=1}^m \frac{n_j}{k_{\Sigma}} \cdot \log p_j.$$

Врахуємо, що для великих значень  $n_j$  та  $k_{\Sigma}$  імовірність *j*-ї пари мовних образів

$p_j = \lim_{k_{\Sigma} \rightarrow \infty} \frac{n_j}{k_{\Sigma}}$ . Тоді середня ентропія однієї пари дорівнює

$$H = - \sum_{j=1}^m p_j \cdot \log p_j. \quad (3)$$

Якщо поява однієї з *m* різних пар мовних образів на вході *S* рівноімовірна, то середня ентропія пари (3) досягає максимального значення:

$$\overline{H_1} = \log_2 m \text{ [Bin]} \quad (4)$$

Зрозуміло, що побудова матриці  $A_Q$  на основі реального текстового матеріалу не призведе до максимального значення ентропії (4). З суто формальною точки зору кількісні оцінки  $\overline{\mu_Q}$  та  $\overline{H_1}$  є різними інтерпретаціями тієї ж самої характеристики однієї АММО. Отже, з урахуванням (1), можна отримати верхню оцінку співвідношення одиниць образного сенсу та інформації як логарифмічну згортку

$$1 \text{ [Cав]} = 2 \log_2 m \text{ [Bin]}.$$

Зауважимо, що визначення нижньої межі та інших властивостей функції згортки інформації в образний сенс потребує подальших досліджень.

#### ВИСНОВКИ

За ознакою максимальної ентропії розглянуту чисельну міру образного сенсу 1 *Сав* можна вважати логарифмічною згорткою інформації з конструкцій мовних образів. Врахування обмежень АММО у подальшому дозволить визначити нижню межу ущільнення вербальної інформації на основі умовної ентропії.

#### ЛІТЕРАТУРА

- [1] Кузьмин И.В. Основы теории информации и кодирования / И.В. Кузьмин, В.А. Кедрус. – К.: Вища школа, 1986. – 238 с.
- Бісікало О.В. Концептуальні основи моделювання образного мислення людини / Бісікало О.В. – Вінниця: ПП Балюк І.Б., ВДАУ, 2009. – 163 с.
- Бісікало О.В. Суб'єктивна одиниця смисла образних конструкцій / О.В. Бісікало // Nauka: teoria i praktyka – 2009: materialy V miedzynar. naukowi-praktycznej konf., (Przemysl, 7–15 sierpnia 2009). – Przemysl: Nauka i studia, 2009. – Vol. 6. – P. 9–12.
- Бісікало О.В. Онтогенетичний метод побудови нечіткого відношення сенсу / О.В. Бісікало // Штучний інтелект. – 2011. – № 1. – С. 134–140.













