

## ВИЗНАЧЕННЯ СЕНСУ ПРИРОДНО-МОВНИХ КОНСТРУКЦІЙ НА ОСНОВІ МОДЕЛІ РОЗПОВСЮДЖЕННЯ ОБМЕЖЕНЬ

О. В. Бісікало, канд. техн. наук; І. О. Назаров, студент,  
Вінницький національний технічний університет  
obisikalo@gmail.com  
igor.nazarov.1991@gmail.com

Комп'ютерне розуміння природно-мовних текстів є одним з провідних напрямків теорії штучного інтелекту, який знаходиться на межі інтересів як гуманітарних (педагогіка, лінгвістика, психологія спілкування тощо), так і технічних дисциплін (обчислювальна техніка, програмування, теорія представлення знань). Помітно невеликі успіхи у розв'язанні даної проблеми пояснюються, перш за все, складністю формалізації природної мови, а також відсутністю єдиного підходу у розумінні проблеми різними дослідниками. На даному етапі більшість з них визнають необхідність розгляду тексту не тільки з позицій морфології і синтаксису, скільки з погляду семантики [1].

Автоматичне розуміння текстової інформації дозволить вирішити цілий ряд досить складних задач. До них, у першу чергу, відноситься створення систем повнотекстових баз даних та інтелектуального пошуку, автоматичного індексування, гіпертекстових систем, систем автоматичного анотування і реферування, систем машинного перекладу, генерації тексту тощо [2]. Наразі очевидною є серйозна недосконалість наведених вище систем, які значно поступаються у своїй функціональності природним можливостям людини.

Метою даної роботи є формалізація актуальної задачі визначення сенсу природно-мовних конструкцій на основі моделі розповсюдження обмежень, яка зводиться до оптимізації функцій від великої кількості дискретних аргументів. Специфіка підходу полягає у тому [3], що потрібну функцію представлено у вигляді суми великої кількості складових, кожна з яких залежить від малої кількості аргументів, у частковому випадку – від двох аргументів. Такі задачі відомі як *(max,+)*-задачі розмітки.

Нехай  $T = \{x_1, x_2, \dots, x_n\}$  – текст на природній мові, тоді  $x_i \in T$  ( $i = \overline{1, n}$ ) – речення, з яких побудовано цей текст. У свою чергу, кожне  $i$ -те речення  $x_i = \{s_{i1}, s_{i2}, \dots, s_{im}\}$  складається з лексичних одиниць (слів)  $s_{ij}$  ( $j = \overline{1, m}$ ). Внаслідок мовного явища полісемії кожному слову може ставитися у відповідність сукупність певних мовних образів  $s_{ij} \rightarrow o = \{o_{ij1}, o_{ij2}, \dots, o_{ijk}\}$  ( $k = \overline{1, l}$ ), які фактично визначають сенси слова [4]. Для забезпечення найбільш якісного визначення сенсу покладемо виконання таких умов: максимізація відповідності слова мовному образу та асоціативному зв'язку для пари мовних образів. Тому введемо функції відповідності слова мовному образу  $F_{so}(s_{ij}, o_{ijk})$  і зв'язку між парою синтаксично поєднаних у реченні образів  $F_{oo}(o_{ijk}, o_{i',j',k'})$ . Значення цих функцій – це числа  $\alpha \in [0, 1]$ , які показують силу відповідних зв'язків та отримуються з матриць суміжності, попередньо складених для зазначених вище наборів даних. Отже, задача зводиться до максимізації суми функцій відповідності слова мовному образу і зв'язку між парою образів:

$$W = \max \left( \sum_{i,j,k} F_{so}(s_{ij}, o_{ijk}) + \sum_{i,j,k} F_{oo}(o_{ijk}, o_{i',j',k'}) \right). \quad (1)$$

Запропонована модель отримує вхідні дані з побудованого для певної мови нечіткого простору сенсу образних конструкцій [4] та може бути розв'язана одним з відомих методів [5]. У запропонованому підході зроблено акцент у обробленні природно-мовних текстів саме на визначенні сенсу, тобто на семантичному аналізі, а не, як у традиційних підходах, на морфологічному та синтаксичному. На відміну від більшості існуючих, даний підхід забезпечує кількісну оцінку семантичних зв'язків між лексичними одиницями речення.

1. Бісікало О. В. Концептуальні основи моделювання образного мислення людини / О. В. Бісікало. – Вінниця : ПП Балюк І. Б., ВДАУ, 2009. – 163 с.
2. Леонтьева Н. Н. Автоматическое понимание текстов: системы, модели, ресурсы / Н. Н. Леонтьева. – М. : Издательский центр «Академия», 2006. – 304 с.
3. Шлезингер М. И. Десять лекций по статистическому и структурному распознаванию / М. И. Шлезингер, В. Главач. – К. : Наукова думка, 2004. – 545 с.
4. Бісікало О. В. Побудова нечітких відношень і простору сенсу образних конструкцій / О. В. Бісікало // Вісник КНУ ім. Тараса Шевченка. Серія : Фізико-математичні науки. – 2011. – Вип. № 1. – С. 70–73.
5. Werner T. A Linear Programming Approach to Max-sum Problem: A Review. IEEE Trans. on Pattern Recognition and Machine Intelligence. – 29(5.7). – July 2007.

