

DOI: 10.34185/1991-7848.itmm.2023.01.078

ІДЕНТИФІКАЦІЯ ЕМОЦІЙНИХ СЛІВ У ТЕКСТОВОМУ СПІЛКУВАННІ

Дмитрієва І.С., Бімалов Д.В.

Український державний університету науки і технологій, Україна

В даний час існує безліч програмних продуктів призначених для способів спілкування на далеких відстанях. Але в умовах соціальної потреби тих, що відбуваються у світі, виникає необхідність безперервного спостереження за різними інформаційними діями. Величезна кількість сфер діяльності людини зумовлює появу інформаційних ресурсів, що відображають соціальні спілкування. Повідомлення користувачів, коментаторів стрічок тощо. Їх висловлювання містять інформацію про особисті стосунки до того, що відбувається у суспільному житті [1].

Дослідження ідентифікації емоцій у текстовому спілкуванні є актуальним напрямком досліджень в галузі обробки природної мови та машинного навчання. Основна мета дослідження полягає в розробці алгоритмів та моделей, які можуть автоматично визначати емоційний стан людини на основі текстових повідомлень. Дана робота присвячена огляду деяких моделей та алгоритму для поліпшення обробки даних в середині текстового спілкування користувачів [2].

В обробці природної мови для збору та обробки даних, помічених людиною, потрібно значну кількість часу та ресурсів. Зі швидким зростанням обсягу текстових даних, виникла потреба в методах для аналізу цих текстів.

Важливість завдань та їх широкі застосування започаткували великий інтерес до вивчення, але поки що не існує методу, який міг би наблизитися до людської продуктивності за її вирішенням. Відомо, що вони взаємовигідні, а це означає, що дискурс інформації може бути використаний для покращення аналізу емоцій співрозмовника.

При розробці систему стояло завдання враховувати технологічні процеси аналітичної моделі опису природної мови. Завдання у більшості випадків є вузькоспеціалізованими та складними з погляду адаптації до конкретних видів завдань обробки текстової інформації відкритих мереж. Для підвищення якості обробки тексту у предметній галузі необхідно вирішити питання про формалізації семантичної складової. Одним із підходів, який може бути застосований для обробки текстових повідомлень є модель, у якій виділяється три рівні: морфологічний, семантико-синтаксичний, семантичний.

Також слід звернути увагу на дискурс - це завдання побудови ієрархічної

деревоподібної структури, листям якої є пропозиції. Вузли відповідають зіставленню дочірніх текстів. В роботі було висловлено припущення, що інформація, витягнута з дерев дискурсу, може допомогти з аналізом настроїв та знаючи настрій двох творів, аналогічним чином, текст може допомогти у виявленні дискурсивних відносин між ними.

Дерево дискурсу представляється як набір складових $R[i; j]$ де $i < j$. У цьому поданні R відноситься до риторичного відношення, яке має місце між одиницею дискурсу, від i до j , та одиниця, що містить $i+1$ до j .

Розглянемо на прикладі повідомлення: «Я ходив сьогодні в кіно. Мені дуже сподобалося – раджу».

Дискурс - є поняття філософії, лінгвістики, соціальної комунікації та описується як сукупність текстів, які присвячені деякій темі. Тема цього обговорення - кіно.

Побудуємо ієрархічно деревоподібну структуру (рисунок 1).

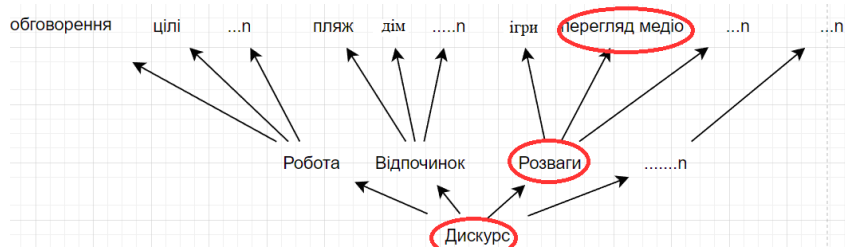


Рисунок 1 - Ієрархічна деревоподібна структура

Семантична модель може бути проілюстрована графічно за допомогою ієрархічної діаграми абстракцій, яка показує типи даних у вигляді полів та їх зв'язки у вигляді рядків. Це робиться ієрархічно, щоб типи, на які посилаються інші типи, завжди були перелічені вище того, на що вони посилаються. Це полегшує читання та розуміння. Абстракції, що використовуються в семантичній моделі даних, це:

- класифікація – “екземпляр_зв’язку відносин
- агрегація - відносини “має_а”
- узагальнення - відносини “є”

Додавання системи функцій для позначення дій до ієрархії класів дозволяє перекладати конструкції семантичною мовою.

$$S = \langle W, Sh, C \rangle,$$

де W - безліч словоформ;

Sh - безліч семантичних шаблонів;

C - клас слова, що приєднується.

Розглянемо на нашому прикладі: «Я ходив сьогодні в кіно. Мені дуже

сподобалося – раджу».

Абстракції, що використовуються в семантичній моделі даних представлена на рисунку 2. Семантична модель допомагає визначити емоцію.

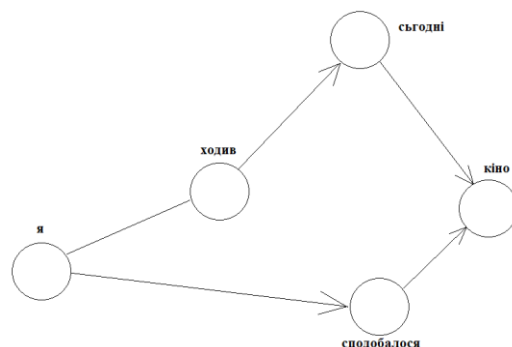


Рисунок 2 - Семантична модель першого речення

В роботі для пошуку емоційних слів у повідомленнях запропоновано використання векторного алгоритму пошуку. Цей алгоритм є класичним алгебраїчним. Він заснований на векторній моделі інформаційного масиву, в якій для визначення міри близькості результатів і запитів використовується значення косинуса кута між векторами в багатовимірному просторі інформаційного масиву. Опис запиту, відповідної користувачеві тематики, також є вектором у тому ж евклідовому просторі термів.

Провівши аналіз роботи алгоритму пошуку, використовуючи текстовий файл з листуванням співрозмовників в кількості 100 штук. Середній показник точності виявлення емоційних слів дорівнює 74.6%. Для порівняння також було проведено аналіз експертом тих самих повідомлень. Середня точність ідентифікації емоційних слів експертом склала 92,5% із існуючих слів у даній базі повідомлень.

В результаті проведеної роботи пропонується за допомогою дискурсу та семантичної моделі будувати деревоподібну структура пропозицій. А за допомогою векторного алгоритму пошуку знаходити емоційні слова у текстових повідомленнях.

Література

1. Andreichuk N. I., Babeliuk O.A. Contrastive Lexicology of English and Ukrainian languages: theory and practice: textbook/ N.I. Andreichuk, O.A. Babeliuk. – Kherson: Publishing House “Helvetica”, 2019. – 236 p.
2. Гнезділова Я.В. Емоційність та емотивність сучасного англомовного дискурсу: структурний, семантичний і прагматичний аспекти: дис. ... канд. філолг. наук. – К.: КНЛУ, 2007 – 269 с.

IDENTIFICATION OF EMOTIONAL WORDS IN TEXT COMMUNICATION

Dmytriieva Iryna, Bimalov Dmytro

Abstract. The study of the identification of emotions in text communication is an actual direction of research in the field of natural language processing and machine learning. The main goal of the research is to develop algorithms and models that can automatically determine a person's emotional state based on text messages. This work is devoted to the review of some models and an algorithm for improving data processing in the middle of text communication of users.

Keywords: emotion identification, discourse tree, semantic model, vector algorithm

Reference

1. Andreichuk N. I., Babeliuk O.A. Contrastive Lexicology of English and Ukrainian languages: theory and practice: textbook/ N.I. Andreichuk, O.A. Babeliuk. – Kherson: Publishing House “Helvetica”, 2019. – 236 p.
2. Hnezdilova Ya.V. Emotsiinist ta emotyvnist suchasnoho anhlo-movnoho dyskursu: strukturnyi, semantychnyi i prahmatychnyi aspekty: dys. ... kand. filolh. nauk. – K.: KNLU, 2007 – 269 s.