

Багаторівневий підхід до тонального аналізу повнотекстової інформації

Ольга Канищева

Національний технічний університет "Харківський політехнічний інститут",
21, вул. Фрунзе, Харків, 61002, Україна, E-mail: olya-kanisheva@rambler.ru

Анотація – У статті розглянуто сучасний стан тонального (емоційного) аналізу та рубрикування. Запропоновано використання багаторівневого підходу до тонального аналізу повнотекстової інформації. Для моделювання емоційно-насиченої лексики запропоновано використовувати алгебру скінчених предикатів.

Ключові слова – тональне рубрикування, емоційний аналіз, засоби масової інформації (ЗМІ), телебачення, емоції.

I. ВСТУП ТА АКТУАЛЬНІСТЬ ДОСЛІДЖЕННЯ

У сучасну епоху великим державним та комерційним структурам стає усе складніше стежити за інформаційним полем, яке дуже швидко змінюється та містить велику кількість новинних та оглядових матеріалів. Телебачення є важливим джерелом інформації про навколишній світ. Основною функцією телевізійної комунікації є інформаційна, завдяки якій телебачення здійснює процес соціальної інтеграції індивідів. Регулярне ознайомлення з публікаціями у засобах масової інформації для будь-якої серйозної діяльності є необхідним, але не завжди цього достатньо. Великі масиви інформації необхідно підлягати якісному аналізу.

У наш час виникла необхідність у спеціальному інструментарії для роботи з інформацією, під яким розуміють комп'ютерні програми, що допомагають збирати й сортувати матеріали засобів масової інформації. Зараз програми вже здатні на простий контент-аналіз, тобто на семантичну статистичну обробку. Але до якісного контент-аналізу ще далеко. Під механізмом якісного аналізу тексту передбачається оцінка таких нюансів, як емоції, погрози, характер відносин між об'єктами та т. і.

II. МЕТА РОБОТИ

Метою роботи є створення моделі тонального аналізу, а потім і тонального рубрикування повнотекстових документів, за допомогою алгебри скінчених предикатів. Модель буде реалізована за допомогою багаторівневого підходу, тобто на першому рівні – синтаксичний, на другому – лексичний, а на третьому – морфологічний аналіз.

III. ТОНАЛЬНЕ РУБРИКУВАННЯ

Ще однією не менш важливою задачею класифікації тексту є рубрикування тональності публікацій. Система

повинна визначати емоційне забарвлення повідомлень загальне, так і стосовно різних об'єктів документів. Нейромережева модель, яку застосовують у тематичному рубрикуванні, тут не працює [1, 2]. Тональне рубрикування дозволяє:

- аналізувати суспільну думку;
- автоматично складати статистичні звіти настрої преси стосовно об'єкта у кожному регіоні;
- досліджувати закономірності між висловлюваннями в пресі та рейтингом політиків;
- складати дайджести про об'єкти моніторингу, тобто вибирати речення, у яких згадується відповідний об'єкт;
- проводити оцінку Web-сторінок.

У даній роботі використовується багаторівневий підхід до проблеми емоційної оцінки повнотекстової інформації, якою є наприклад повідомлення ЗМІ та її подальша класифікація (рубрикація), що полягає у системному аналізі мовних засобів.

Для того, щоб здійснити системний підхід до вираження емоційного значення, була розроблена купність лексичних і граматичних засобів, які є компонентами мовної системи.

Опис лексичних засобів для вираження емоцій у українській і російській мовах також становить великий інтерес, оскільки дослідження певних груп слів, а також словотворчих зв'язків, може служити для виділення лексикони, сутністю якої є вираження почуттів людини [4]. Оцінці емоцій та їх класифікації присвячена велика кількість робіт [5, 6, 7]. Стилістичний підхід до вивчення експресивно-оцінних та емоційних явищ містить у собі аналіз виразних засобів мови, уживаних для особливого впливу на читача (слухача).

IV. ЕМОЦІЇ ТА ЇХ КЛАСИФІКАЦІЯ

Для якісного тонального аналізу необхідна класифікація емоцій. У своїй роботі ми використовували вже існуючу класифікацію, запропоновану у роботі [8].

При складанні цієї класифікації були відібрані всі слова з художньої літератури та фільмів, що відповідають визначенню емоційних станів.

Ті слова, які вживаються найчастіше були скомпоновані у 16 груп. Кожна група складалася з синонімічних термінів, які розташовувалися у порядку зростаючої емоційної насиченості: крайній ліворуч термін виражає слабкий ступінь емоційної насиченості, крайній із правого – сильний ступінь; відповідно терміни, що займають

... положення виражають середній ступінь ... насиченості.

... з цих груп може бути представлена словом, яке ... найчастіше. У результаті була створена ... схема класифікації емоцій:

Позитивні емоції: радість, ніжність, схвалення.

Негативні емоції: обурення, розпач, образа, гнів, сором, ... сум, презирство, підозра.

Виспані форми: подив, іронія.

... які входять до останньої групи, такі як подив, ... можуть бути використані як для вираження ... позитивних, так і негативних емоцій. Характер емоційних ... згаданих у списку, є доволі різним: деякі з ... збережених термінів є емоціями, інші – психічними ... емоційними конотаціям. Звичайно, існують і ... форми: докір із прикрістю, подив і обурення та ...

... представлена вище класифікаційна схема, не претендує ... універсальність. Труднощами класифікації емоцій є те, ... поняття "емоція" поєднує в собі різні явища.

... представлені синтаксичні та лексичні засоби, що ... беруть участь у вираженні емоційного значення.

Синтаксичні засоби: синтаксичний паралелізм, ... паралелізація, повтор, повторення звернення, емоційні ... частки, інверсія, еліпсис та інші.

Лексичні засоби: прикметники, іменники, дієслова, ... прислівники міри та ступеню, іменники-зуземізми.

Вони дуже різноманітні, але далеко не всі ... зустрічаються в текстах новин та ЗМІ.

Треба відзначити, що синтаксис не передає характер ... емоцій, але чітко реагує на їх наявність або відсутність. ... якщо треба визначити наявність емоції або їх ... відсутність у документі достатньо лише провести ... синтаксичний аналіз документу.

Для вираження різних емоцій мова має у своєму ... порядку багатьох різних лексичних засобів. Емоційні ... конотації виявляються в тексті завдяки оцінюючим емоції ... прикметникам, іменникам, дієсловам, які містять таку ... зв'язку у своїй семантиці.

Одним з важливих лексичних засобів вираження ... емоційних конотацій є прикметник, тому що крім свого ... категоріального значення – визначення ознаки предмета ... або особи – прикметники можуть також передавати ... відношення мовця, об'єкта або дії. Це, наприклад, такі ... прикметники як дурний, жадливий та інші.

Також для вираження позитивних і негативних емоцій ... використовуються прислівники міри й ступеня (жажливо, ... всім та інші), а також прислівники, які позначають стан ... тепло, треба та інші), іменники (наприклад, ... дурень, диявол, любий та інші), іменники-зуземізми ... (відка лань, та інші), дієслова (нехтувати, ненавидіти, ... зликувати та інші).

Таким чином, синтаксичні моделі й лексичні засоби, що ... беруть участь у вираженні емоційного значення, дуже ... різноманітні. Вибір слів, розташування їх, своєрідне ... групування речень і таке інше, використовуються ... автором для емоційного впливу на читача (слухача).

V. АЛГЕБРО-ЛОГІЧНІ МОДЕЛІ ЕМОЦІЙНО-НАСИЧЕНОЇ ЛЕКСИКИ

Для побудови алгебро-логічних моделей емоційно-насиченої лексики у даній роботі використовується апарат алгебри скінчених предикатів. Цей апарат розроблено школою Ю.П. Шабанова-Кушнарєнко [10, 11]. На наш погляд, алгебра скінчених предикатів може бути міцним фундаментом теорії інтелекту.

Алгебра скінчених предикатів є узагальненням апарата булевих функцій (перехід від двійкових знаків до букв) і апарата багатозначної логіки (перехід від функцій до відносин). Використання алгебри скінчених предикатів у теорії інтелекту приводить до виключення потенційної нескінченності з формальних описів інтелектуальних процесів, що забезпечує краще наближення отриманих моделей до дійсності, ніж ті, що може дати апарат теорії алгоритмів [12].

Опишемо за допомогою алгебри скінчених предикатів лексичні засоби, які беруть участь у вираженні емоційного значення.

Дієслово

Нас цікавить дієслово стосовно предмета (об'єкту). Тут, відповідно до правил граматики, відбувається поділ усіх на дві великі групи: перехідних і неперехідних дієслів [13]. До перехідних варто віднести дієслова, які позначають дію, яка безпосередньо переходить на предмет (прямий об'єкт), а неперехідні – позначають протилежну дію. З точки зору лексико-семантичних зв'язків слів у словосполученні перехідні дієслова можуть позначати:

- дії, за допомогою яких створюються які-небудь матеріальні або духовні цінності – *будувати міст, малювати триптих*;
- дії, за допомогою яких руйнуються (знищуються) предмети (об'єкти) – *розірвати ковдру, розбити вазу*;
- дії, пов'язані з різними видами сприйняття людиною проявів дійсності – *відчувати радість, сумувати, ніяковіти*.

Нас цікавить останній підвид перехідних дієслів. Вони можуть бути представлені у вигляді бінарних предикатів. Для доказу цього запишемо в предикативному виді словосполучення, яке містить у собі перехідне дієслово, наприклад, таке словосполучення, як "Вона злякалась публіки" містить у собі перехідне дієслово "зняковіти" та може бути представлено у вигляді – бінарний предикат *злякалась (вона(x), публіка(y))* або розшифруємо як *злякалась (x, y) – предмет x злякався предмета y'.*

$$(вона злякалась публіки)(x, y) = вона(x) \wedge злякалась(x, y) \wedge публіка(y).$$

Іменники й прикметники

Прикметник (як і іменник) формально можна представити у вигляді унарного предиката. Тому можна ускладнити просту конструкцію, що містить тільки якийсь об'єкт або суб'єкт (іменник) до словосполучення, переважно атрибутивного типу.

Наведемо конкретний приклад. Розглянемо словосполучення "жахлива подія", у якому предикат *подія* (*x*) розширюється до предиката (*жахлива подія*)(*x*). Тому лексема *жахливий* виконує роль деякої предикатної операції

$$\text{жахлива (подія } (x) \text{)} = (\text{жахлива подія})(x).$$

Дана операція однозначно представляється через предикат *жахливий* (*x*) як

$$\text{жахлива (подія } (x) \text{)} = \text{жахливий } (x) \wedge \text{ подія } (x).$$

Або: (*жахлива подія*)(*x*) = *жахливий* (*x*) \wedge *подія* (*x*).

Розглядаючи окремо предикат *жахливий* (*x*) можна зробити висновок про те, що він приймає значення 'істина', тобто *жахливий* (*x*)=1 тільки у випадку, коли розглянутий об'єкт подія *x* є жахливим.

Іменники-зусемізни

Це вид метафори, коли назва тварини використовується для визначення людини [8]. Покажемо, що цю частину мови можна представити як унарний предикат.

Наведемо конкретний приклад. Розглянемо словосполучення "працьовита бджілка", у якому предикат *бджілка* (*x*) розширюється до предиката (*працьовита бджілка*)(*x*). Тому лексема *працьовита* виконує роль деякої предикатної операції

$$\text{працьовита (бджілка } (x) \text{)} = (\text{працьовита бджілка})(x).$$

Дана операція однозначно представляється через предикат *працьовита* (*x*) як

$$\text{працьовита (бджілка } (x) \text{)} = \text{працьовита } (x) \wedge \text{ бджілка } (x).$$

Або: (*працьовита бджілка*)(*x*) = *працьовита* (*x*) \wedge *бджілка* (*x*).

Прислівники міри й ступеня

Ці прислівники вживаються не для визначення прикметника, а для вказівки дуже високого ступеня емоційної насиченості.

Прислівники, через свої лексикографічні особливості поки не піддаються формалізації за допомогою предикатів або предикатних операцій. Це предикатна операція над операцією. Пояснимо прикладом

$$\begin{aligned} & \text{Вивчив віри напам'ять} = \\ & [\text{напам'ять (вивчити)}] \text{ віри } (x) = [\text{напам'ять} \\ & \text{(вивчити)}] (x) \wedge \text{ віри } (x). \end{aligned}$$

Таким чином, завдяки апарату алгебри скінчених предикатів у нас є можливість переходу від алгоритмічного опису інформаційних процесів до опису їх у виді рівнянь.

ЛІТЕРАТУРА

- [1] Ильин Н., Киселев С., Рябышкин В., Танков С. Технологии извлечения знаний из текста. «Открытые системы», 2006, № 6, <http://www.DIRECTUM-Journal.ru/docs/1732678.html>.
- [2] Киселев С.Л. «Аналитический курьер» – современная система извлечения знаний из хранилищ документов. «Открытые системы», 2006, № 6, <http://www.DIRECTUM-Journal.ru/docs/1732678.html>.
- [3] Жельвис В.И. Эмотивный аспект речи. Психолингвистическая интерпретация речевого воздействия. Ярославль, 1990.

- [4] Контекстная семантизация лингвистических единиц. М.: Изд-во МПИ иностранных языков, 1984.
- [5] Мягкова Е.Ю. Когнитивная теория эмоций // исследование эмоциональности лексики. М.: Ин-т языкознания АН СССР, 1991.
- [6] Пиотровская Л.А. Эмотивность как языковая категория // Вестник Санкт-Петербургского Университета. Серия 2. История, языкознание, литературоведение. СПб.: СПГУ, 1993. С. 41-47.
- [7] Шаховский В.И. Категоризация эмоций в лексико-семантической системе языка. Воронеж: Изд-во ВГУ, 1987.
- [8] Нушкиня Э.А. Типология эмоциональной речи. Киев: Вища школа, 1986.
- [9] Чешко Л.А. Русский язык: Для подготовительных отделений вузов: Учеб. Пособие. – М.: Высш. Шк., 1990. – 333с.
- [10] Хайрова Н.Ф., Шаронова Н.В. Автоматизированные информационные системы: задачи обработки информации – Х.: Нар. укр. акад., 2002. – 120с.
- [11] Шабанов-Кушнаренко Ю.П., Шаронова Н.В. Компараторная идентификация лингвистических объектов. – К., ИСИО, 1993. – 116 с.
- [12] Шабанов – Кушнаренко Ю.П. Теория интеллекта. Проблемы и перспективы. Ю.П. Шабанов Кушнаренко – Х.: Вища шк., 1987. – 158с.
- [13] Розенталь Д.Э., Теленкова М.А. Словарь-справочник лингвистических терминов. Пособие для учителей. М.: Просвещение, 1976.

VI. ВИСНОВКИ

Запропонований підхід дозволить аналітикам, працюючи із новинними стрічками аналізувати суспільну думку, автоматично класифікувати (рубрикувати) новини та створювати статистичні звіти щодо настроїв преси стосовно об'єкта обговорення, досліджувати закономірності між висловленнями у пресі та рейтингами політиків та таке інше.

Також запропоновану технологію тонального рубрикування можна використовувати для оцінки Web-сторінок.

У результаті емоційного аналізу та рубрикування (класифікації) повнотекстової інформації, взятої відкритих джерел, можна не тільки аналізувати стан ситуації у будь-якій галузі, але й прогнозувати розвиток ситуації, що є життєво важливим моментом для прийняття відомих рішень.

Важливу роль у сучасному бізнесі грає й конкурентна розвідка, в основі якої, лежить маркетингова інформація, яку отримали з преси. На думку фахівців, близько 80% інформації, необхідної для підтримки преси стратегічного управління може братися з відкритих джерел, таких, як Інтернет, традиційні засоби масової інформації, інформаційно-рекламні матеріали компаній тощо.