

ОГЛЯД СУЧАСНИХ ПІДХОДІВ ДО АВТОМАТИЧНОГО РОЗПІЗНАВАННЯ ГУМОРУ

Вінницький національний технічний університет

Анотація

Розглянуто основні методи автоматичного розпізнавання гумору, оцінено їх переваги та недоліки, що дозволило визначити загальні тенденції досліджень та сформулювати подальший напрямок розвитку у даній області.

Ключові слова: розпізнавання гумору, гумористичний якір, автоматична класифікація, комп'ютерна лінгвістика, розпізнавання природної мови.

Abstract

The main methods of automatic recognition of humor are considered, their advantages and disadvantages are estimated, which allowed to determine the general tendencies of research and to formulate further direction of development in this area.

Keywords: humor recognition, humor anchor, automatic classification, computer linguistics, natural language recognition.

Вступ

У сучасних умовах розвитку інформаційних технологій обробка об'єктів образного характеру займає досить важливе місце, проте неоднозначність трактування людиною результатів образного мислення призводить до ускладнення постановки задачі для комп'ютера.

Автоматична обробка гумору за допомогою обчислювальної техніки є важливим кроком для покращення розуміння комп'ютером сенсу людської мови та адаптації поведінки інформаційних систем відповідно до дій користувача.

Виявлення гумору у заданому контексті є складною проблемою в обробці природної мови за рахунок неоднозначності трактування висловлювань та необхідності розглядання гумору в більш широкому, іноді загальнолюдському контексті [1]. Загальною тенденцією більшості робіт у даному напрямку є відсутність систематичного підходу і виявлення гумору в дуже обмеженому контексті, тому актуальним є формування більш формалізованого підходу з врахуванням сучасних досягнень.

Метою роботи є вивчення та аналіз сучасних підходів до автоматичного розпізнавання гумористичного контенту.

Результати дослідження

Серед інструментів утворення гумору при його автоматичному розпізнаванні та обробці в першу чергу виділяють неоднозначність, алітерацію, порівняння та гру слів [2]. Попри прогнози дослідників, з точки зору обчислювальної техніки розпізнавання гумору за даними критеріями, виявленими як найбільш очевидні, виявилось більш складним, ніж його генерація, що спиралась на аналогічні теорії формування смішного або іронічного контенту.

Помітного поширення набуло застосування методів автоматичної класифікації для розрізнення гумористичних та негумористичних текстів. Головною перевагою ранніх досліджень є їх застосування на досить великих об'ємах даних, проте вони концентруються на аналізі так званих одноліній [3]. За рахунок простого синтаксису і використання типових засобів формування гумористичного ефекту їх автоматична класифікація давала досить високий відсоток точності, що, проте, дозволило лише підтвердити життєздатність даного підходу у галузі розпізнавання гумору. Основним недоліком даного підходу є обмеженість вхідного контенту та абстрагування від мовного оточення, що, з одного боку, значно полегшувало задачу розпізнавання, але, з іншого боку, робило цей підхід нежиттєздатним для більшості реальних даних.

Перехід до аналізу реального гумористичного контенту, такого як дані з соціальних медіа, дозволив відійти від аналізу прототипних структур та зосередитися на більш глибоких ознаках гумористи-

чного контенту, таких як неоднозначність, невідповідність, полярність та наявність емоційного сценарію у текстах [4]. Одночасно було підтверджено, що визначення гумору є більш простою задачею, ніж визначення іронії через більш глибокий зміст та вплив контексту у випадку другої. Перевагою даного етапу досліджень є виділення більш чіткого набору ознак, таких як структурна, морфосинтаксична та семантична неоднозначність, полярність, що позначає позитивний або негативний ефект конкретних слів, несподіваність, що виражається дисбалансом значень, та орієнтовані на психологічний контекст емоційні сценарії. Головним недоліком даного підходу є те, що класифікатори опираються на негативні зразки з доменів, відмінних від доменів позитивних прикладів гумору, що дозволяє показувати більш високу ефективність на штучно підібраних даних, проте погану – на реальних.

Останні дослідження опираються на виявлення гумористичних якорів – слів або фраз, що є генераторами гумористичного ефекту. Такий підхід дозволяє забезпечити універсалізацію процесів розпізнавання гумору, а також застосовувати його до більш широкого контексту та речень загального характеру [5].

Формування поняття латентних семантичних структур дозволило більш чітко формалізувати процедуру розпізнавання гумористичного контенту та гумористичних якорів, зокрема в останніх дослідженнях було створено набори функцій, що відслідковують індикатори гумору. Попри значні переваги даного методу у вигляді відходу від типового набору даних, недолік врахування малого контексту (10-30 слів) зберігається.

Загальною тенденцією більшості сучасних підходів до автоматичної обробки та розпізнавання гумору є використання автоматичної класифікації та машинного навчання, а розвиток області спирається на виявлення загальних ознак гумористичного контенту.

Висновки

Виконано огляд сучасних тенденцій в області автоматичного розпізнавання гумору, визначено основні напрямки розвитку та наявні технічні рішення.

Подальшого дослідження потребує питання розробки формальної лінгвістичної моделі на основі класифікації гумористичних ефектів. Перспективним напрямком моделювання може бути врахування асоціативних зв'язків між словоформами у текстах. Наявні методи прогнозування гумору потребують адаптування для застосування їх в загальних ситуаціях для більш широкого контексту.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Attardo S., Linguistic theories of humor / S. Attardo – Walter de Gruyter, 1994. - 426 p.
2. Mihalcea, R. Learning to Laugh (Automatically): Computational Models for Humor Recognition / R. Mihalcea, C. Strapparava. - Journal of Computational Intelligence 22 (2), 2006 - p. 126–142.
3. Mihalcea, R. Making Computers Laugh: Investigations in Automatic Humor Recognition / R. Mihalcea, C. Strapparava, - HLT/EMNLP, 2005.
4. Reyes, A. From humor recognition to irony detection: The figurative language of social media / A. Reyes, P. Rosso, D. Buscaldi. - Data Knowl. Eng., 74, 2012 – p. 1-12.
5. Humor Recognition and Humor Anchor Extraction / D. Yang, A. Lavie, C. Dyer, E. H. Hovy, - Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, 2015. – p. 2367–2376.

Шолота Владислава Владиславівна — студент групи ІКТ-16б, факультет комп'ютерних систем і автоматики, Вінницький національний технічний університет, Вінниця, e-mail: Vladislava.Sholota@gmail.com

Науковий керівник: *Бісікало Олег Володимирович* – д-р техн. наук, професор, декан факультету КСА, Вінницький національний технічний університет, м. Вінниця

Sholota Vladyslava V. — student group ICT-16b, Faculty for Computer Systems and Automatic, Vinnytsia National Technical University, Vinnytsia, email : Vladislava.Sholota@gmail.com

Supervisor: *Bisikalo Oleg V.* – Dr.Sc. (Eng.), Professor, Dean of the Faculty for Computer Systems and Automatic, Vinnytsia National Technical University, Vinnytsia