

РЕФЕРАТ

Магістерська дисертація: 99 с., 15 рис., 26 табл., 2 додатки і 37 джерел.

Об'єкт дослідження – семантична орієнтація текстів новин.

Предмет дослідження – метод Наївного Байєса та згорткові нейронні мережі.

Мета роботи – дослідження семантичної орієнтації тексту, використовуючи різні підходи до побудови списку ознак та різні методи оцінки орієнтації (позитивної, негативної). Було проведено огляд існуючих моделей, що використовуються для побудови списку ознак та оцінки семантичної орієнтації, підібрані оптимальні.

Методи дослідження – нейронні мережі, методи обробки текстів.

Основний результат даного дослідження – це розробка алгоритмів класифікації тональності новин. Для цього вивчаються два методи. По-перше, це алгоритм Наївного Байєсу, який використовує репрезентативну групу для класифікації. Другий - це згорткова нейронна мережа, яка показала дуже хороші результати під час розв'язання задач з обробки зображень, але й може бути використана для пошуку семантичної орієнтації текстів. Були проведені експерименти та дослідження ефективності двох різних алгоритмів, що виявляють позитивні та негативні окраси тексту. Крім того, необхідно визначити алгоритм, що дає кращі результати; ще важливо вивчити, як точність алгоритмів може впливати на попередню обробку даних, вибір ознак та даних.

Основними джерелами даних - є платформа Twitter, експертні дані оцінки впливу новин на економіку США, заголовки новин з Австралійського джерела новин ABC. Аналіз проводився з використанням мови програмування Python та таких бібліотек для аналізу даних, як pandas, sklearn, tensorflow тощо.

СЕМАНТИЧНА ОРІЄНТАЦІЯ ТЕКСТУ, НАЇВНИЙ БАЙЄС, ЗГОРТКОВІ НЕЙРОННІ МЕРЕЖІ, ПОБУДОВА СПИСКУ ОЗНАК ТЕКСТУ, АЛГОРИТМИ КЛАСИФІКАЦІЇ