

ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ КЛАСИФІКАЦІЇ ТОНАЛЬНОСТІ РЕЧЕНЬ

Вінницький національний технічний університет;

Анотація

Запропоновано інформаційну технологію класифікації тональності речень, яка побудована на основі згорткової нейронної мережі і дозволяє підвищити достовірність класифікації тональності речень щонайменше на 4,5%.

Ключові слова: тональність речень, класифікація, нейронна мережа, інформаційна технологія.

Abstract

Information technology of sentence tone classification is proposed, which is built on the basis of convolutional neural network and allows to increase the accuracy of sentence tone classification by at least 4.5%.

Keywords: sentence tone, judgment of thoughts, neural network, information technology.

Вступ

При стрімкому зростанні соціальних медіа (наприклад, огляди, форуми обговорення, блоги, мікроблоги, Twitter, коментарі та повідомлення в соціальних мережах) в Інтернеті, приватні особи та організації все частіше використовують вміст у цих середовищах для прийняття рішень. Сьогодні, якщо хтось бажає купити певний продукт, то майбутній покупець вже не обмежується лише опитуванням своїх друзів і родини, тому що є багато відгуків користувачів та обговорення на різних форумах в Інтернеті про вибраний цією людиною продукт. Тим не менш, постійний пошук і моніторинг сайтів в Інтернеті, відбір необхідної інформації, залишаються доволі складним завданням через значне поширення різноманітних таких сайтів. Кожен сайт, як правило, містить величезний обсяг тексту, який містить необхідну інформацію щодо певного продукту чи певної послуги, але цю інформацію не завжди легко виявити в довгих блогах, на форумах чи у великих повідомленнях. Середньостатистичний користувач може зіткнутися з труднощами із визначенням відповідних сайтів, видобуванням та узагальненням потрібної інформації на них. І це зумовило розвиток галузі автоматичний аналіз тональності текстів.

Метою роботи є розроблення інформаційної технології класифікації тональності речень із використанням згорткових нейронних мереж.

Результати дослідження

Для задачі класифікації тональності речень використано згорткову нейронну мережу (ЗНМ), яка має три згорткових шари з фільтрами. Кількість фільтрів – 16 для кожного шару. Відповідно до цього кожен згортковий шар утворює 16 ознак. В архітектурі ЗНМ є звичним вставляти агрегуювальний шар між послідовними згортковими шарами. Для задачі класифікації текстів кожен агрегуювальний шар перетворюватиме вихід згорткового шару у матрицю розмірністю 16 x 1, що значно знижуватиме розмірність, а відповідно і складність обчислень. Після кількох згорткових та максимізаційно агрегуювальних шарів, встановлено повноз'єднаний шар, нейрони якого з'єднуються з усіма нейронами попереднього шару. Після проведення операцій згортки та агрегування буде виділено по 16 ознак з кожної зв'язки згорткового та агрегуювального шарів, які будуть подані на вихідний повнозв'язний шар.

Структура інформаційної технології класифікації тональності речень на основі згорткової нейронної мережі, зображена на рис. 1.

Для роботи інформаційної технології класифікації тональності речень на основі згорткової нейронної мережі вхідною є інформація у вигляді тексту (ASCII коди символів). Потім здійснюється

процес векторизації слів тексту за методом word2vec [1]. Класифікація тональності речень буде виконуватись згортковою нейронною мережею на основі попереднього її навчання на основі відповідної навчальної вибірки. Тому відбувається процес створення цієї навчальної вибірки речень. Далі відбувається процес формування (ініціалізації) та навчання згорткової нейронної мережі, а потім процес подачі векторизованого речення на вхід згорткової нейромережі. Далі відбувається процес класифікації тональності речення згортковою нейронною мережею, а потім - процес виведення результату класифікації.

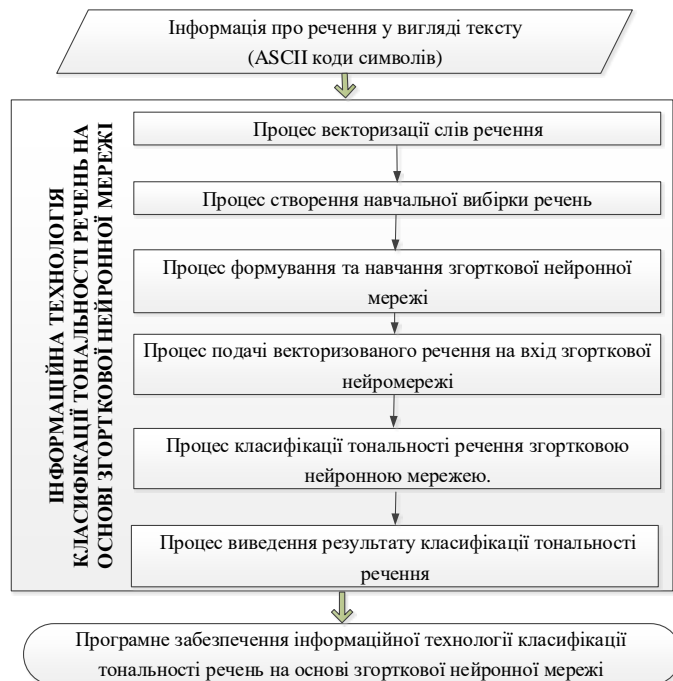


Рис. 1. Структура інформаційної технології класифікації тональності речень на основі згорткової нейронної мережі

Розроблена структура інформаційної технології класифікації тональності речень на основі згорткової нейронної мережі була використана для розробки програмних засобів.

Програмна реалізація інформаційної технології класифікації тональності речень здійснювалась на мовах програмування: для програми класифікації тональності речень – Python, для програмної реалізації клієнтського додатку для тестування роботи WebAPI та : web2py та scikit-learn. В результаті було розроблено програмне забезпечення класифікації тональності речень на основі згорткової нейронної мережі, яке порівняно з аналогом має кращу на 4,5% достовірність класифікації тональності речень.

Висновки

Встановлено, що запропонований підхід дозволяє підвищити достовірність класифікації тональності речень як мінімум на 4,5% яке порівняно з аналогічними програмами. Також пропонується у подальшому використовувати для класифікації тональності речень в реальному масштабі часу імпульсні нейронні мережі [2]. Це покращить точність класифікації. Крім того, імпульсні нейронні мережі мають гарні перспективи для апаратної реалізації [3] та найкраще підходять для побудови операційного ядра майбутніх нейрокомп'ютерів [4].

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Метод векторизації слів - word2vec – [Електронний ресурс]. – Режим доступу: <http://nlpx.net/archives/179>
2. В.Ф.Бардаченко, О.К.Колесницький, С.А.Василецький. Перспективи застосування імпульсних нейронних мереж з таймерним представленням інформації для розпізнавання динамічних образів// УСiМ.-2003-№6.- С. 73-82.

3. Колесницький О. К. Аналітичний огляд апаратних реалізацій спайкових нейронних мереж / О. К. Колесницький // Математичні машини і системи. – 2015. – №1, С.3-19. ISSN 1028-9763.

4. Колесницький О. К. Принципи побудови архітектури спайкових нейрокомп'ютерів / О. К. Колесницький // Вісник Вінницького політехнічного інституту. – Вінниця: УНІВЕРСУМ-Вінниця. – 2014. – №4 (115), С.70-78.

Барановський Владислав Сергійович — студент групи 2КН-18м, факультет інформаційних технологій та комп'ютерної інженерії, Вінницький національний технічний університет, м. Вінниця, e-mail: vladyn4uk@gmail.com

Колесницький Олег Костянтинович — доцент кафедри комп'ютерних наук ВНТУ, Вінницький національний технічний університет, м. Вінниця.

Денисов Ігор Костянтинович — викладач кафедри комп'ютерних наук ВНТУ, Вінницький національний технічний університет, м. Вінниця.

Baranovskyi Vladislav S. — Faculty of Information Technologies and Computer Engineering, Vinnytsia National Technical University, Vinnytsia, email : vladyn4uk@gmail.com

Kolesnytskyj Oleg K. — docent of the Computer Sciences Chair, Vinnytsia National Technical University, Vinnytsia

Denysov Ihor K. — lecturer of the Computer Sciences Chair, Vinnytsia National Technical University, Vinnytsia