УДК 519.876.5

Артем Ульянко, аспірант, Юрій Дорофєєв, д.т.н., проф.

МУЛЬТИМОДАЛЬНЕ РОЗПІЗНАВАННЯ ЕМОЦІЙНОГО СТАНУ ЛЮДИНИ НА ОСНОВІ ВІДЕО, ПРОМАРКОВАНОГО СУБТИТРАМИ

Мультимодальне розпізнавання емоцій – це відносно новий напрям в сфері автоматичного розпізнавання емоцій людини, який передбачає використання одразу декілька джерел інформації для отримання результату класифікації. Цей напрям отримав стрімкий розвиток разом із ростом кількості даних в соціальних мережах. Основною мотивацією для розробки подібних систем є можливість ідентифікації емоцій у контексті, що робить розпізнавання більш об'єктивним на відміну від одномодальних систем, які, зазвичай, використовують лише відеозображення.

Метою роботи є розробка автоматичної системи для розпізнавання емоційного стану людини на основі використання трьох джерел інформації: відеозображення, аудіофайла та тексту у вигляді субтитрів.

На рис. 1 представлено схематичне зображення системи, до складу якої включено чотири типи класифікаторів, а вхідними даними є відео, промарковане субтитрами.

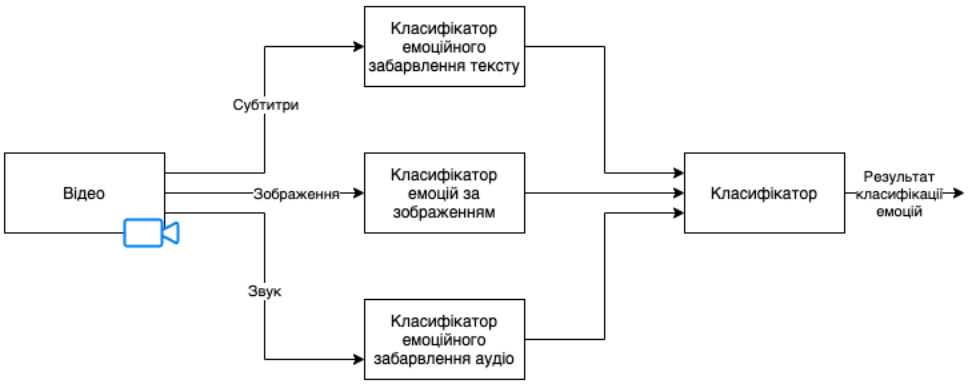

Рис. 1 – Мультимодальна система розпізнавання емоцій

Першим є класифікатор емоційного забарвлення тексту, для побудови якого використовується архітектура рекурентних нейронних мереж (RNN). Вважається, що текст може мати три тональності: позитивну, нейтральну та негативну, де нейтральна тональність відповідає відсутності емоцій [1].

Другий – класифікатор емоцій на основі зображення – має архітектуру згорткових нейронних мереж (CNN) та приймає рішення, яке відноситься до одного з семи базових класів емоцій [2].

Для розпізнання емоційного забарвлення аудіопотоку використовується класифікатор, побудований на основі RNN [3], який відносить емоцію людини до одного з чотирьох класів: сум, злість, радість та нейтральний.

Вихідний класифікатор отримує вектор, що включає результати розпізнавання усіх трьох модальностей, та на виході визначає найбільш вірогідну емоцію.

Запропонована архітектура системи мультимодального розпізнавання емоцій на основі відео, що промарковане субтитрами, визначає емоції з урахуванням контексту та дозволяє покращити якість розпізнавання емоційного стану людини у порівнянні з одномодальними системами розпізнавання. В подальшому архітектура системи може бути розширена завдяки додаванню нових модальностей та автоматизації генерування субтитрів.

Література

1. Pang B. Opinion Mining and Sentiment Analysis / B. Pang, L. Lee.// Foundations and Trends in Information Retrieval – 2008. – № 2, с. 1-135.
2. Ekman P. Universals and cultural differences in the judgments of facial expressions of emotion // Journal of personality and social psychology. – 1987. – Т. 53, № 4, с. 712-714.
3. Yoon S. Multimodal Speech Emotion Recognition Using Audio and Text / S. Yoon, S. Byun, K. Jung. // IEEE Spoken Language Technology Workshop (SLT) , Athens, Greece – 2018. – с. 112–118.