УДК 004.912

АНАЛІЗ ТОНАЛЬНОСТІ ОБ'ЄДНАННЯ ТЕКСТОВИХ, АУДІО ТА ВІЗУАЛЬНИХ ДАНИХ

Орел А.В.
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

Оглядова стаття представляє новий підхід до виконання діагностики аналізу тональності новинних відеороликів, що базується на об'єднанні текстової, аудіо і візуальної інформаціях, що узяті з їх вмісту. Пропонований підхід має на меті сприяти семіо-дискурсивному вивченню щодо побудови ідентичності всесвітніх ЗМІ, що стали центральною частиною сьогоденного життя мільйонів людей. Для досягнення цієї мети застосовується найсучасніші обчислювальні методи для автоматичного розпізнавання емоцій з міміки, витяги модуляцій виступів учасників та аналіз тональності із субтитру відео, що відповідають інтересам. Більш детально обчислимо такі функції, як, наприклад, зорові інтенсивності розпізнаних емоцій, розміри обчислювальних областей учасників, голосова ймовірність, гучність звуку, фундаментальні частоти мови та оцінки (полярності) тональності з текстових речень у субтитрах. Експериментальні результати з набору даних, що містять 520 анотованих відео новин популярних телевізійних каналів показують, що цей підхід досягає точності до 84% у завданнях класифікації тональностей (рівні напруженості), що демонструє його високий потенціал для використання мультимедійними аналітиками у кількох напрямках, особливо у журналістській області.

Ключові слова: аудіо тональність, візуальна тональність, тональність тексту, аналіз мультимодальної тональності, розпізнавання емоцій.

Постановка проблеми. Випуски новин є основою усіх телевізійних мереж у світі. У порядку довіри телевізійному випуску новин, часто можна спостерігати за спектаклізацією інформації журналістами як спосіб підготувати ментальну глядацьку модель (Goffman, 1981).

Аналіз випусків новин є важливим для медіа аналітиків деяких сфер, особливо в області журналістики (Stegmeier, 2012). Оскільки випуски новин є специфічним видом дискурсу, а також соціокультурною практикою, техніки аналізу дискурсу (Charaudeau, 2002) були застосовані для аналізу структури випусків новин на багатьох рівнях опису, щодо деяких властивостей, такі як їх загальна тематика, схеми висловів та стиль дискурсу як масштабність випуску новин (Cheng, 2012).

Зазвичай мови аналізуються без підтримки обчислювальних засобів, такі як автоматизовані анотації програмного забезпечення і аналітичні відео програми. Тільки нещодавно, з розвитком таких сфер, як аналіз тональності, обчислювальна (математична) лінгвістика, мультимедійні системи та комп'ютерне бачення, були запропоновані нові методи підтримки аналізу дискурсу, особливо в мультимедійному контенті телевізійних новини (Pantti, 2010). Однак, в міру наших знань, немає попередніх результатів, що намагалися використовувати мультимодальні функції (наприклад, аудіо, текстові та візуальні функції) для того, щоб виміряти рівень напруженості новин. Також підходи, щоб зробити висновок, що напруженість може мати здатність для важливості новин, а також допомоги в організації і підсумуванні новин.

Мета статті. Головною метою цієї статті є представлення обчислювальний підхід, аби підтримати вивчення рівня напруженості новин з мультимодальних функцій, доступних у їх відео. Запропонований підхід дозволяє дослідникам виконувати семіо-дискурсивний аналіз словесного та невербального мовлення, які проявляються через міміку і жести журналістів перед камерою, що є візуальним способом виразити свої ідеї (Eisenstein, Barzilay, і Davis, 2008). У експериментах показано ефективність запропонованого підходу, а також важливість аналізу тональності для висновку про новини з рівнем високої напруженості.

Аналіз останніх досліджень і публікацій.

– напруженість та аналіз тональності у новинах;

Багато людей читають онлайн-новини з вебсайтів комунікаційних порталів. Ці новинні вебсайти повинні створювати ефективні стратегії, щоб привернути увагу людей до цього змісту. У цьому контексті (Reis et al. (2015)) розгляд стратегій, що використовуються організаціями онлайн-новин при розробці їх заголовків. Було проаналізовано зміст 69 907 заголовків, випущених чотирма великими медіа компаніями протягом мінімум восьми місяців поспіль у 2014 році. Як результат, новини з негативною тональністю, як правило, генерують багато негативних поглядів і коментарів. В результаті аналізу звернемо увагу на те, що чим більша негативна напруженість новин, тим більша потреба користувача додати свою думку. Це може бути коментар, який суперечить думці іншого користувача та сприяє подальшим обговоренням посту.

Pantti (2010) вивчає емоційну цінність експресії журналістів у випусках новин за допомогою проявів публічних емоцій. Стаття дає свідчення того, як журналісти оцінюють роль та стан емоцій у висвітленні ЗМІ, а також емпатію, викликану новинами. Крім того, стаття вивчає, як емоційний дискурс журналістів пов'язаний з їх ідеєю хорошої журналістики та образом професіоналів.

– аналіз мультимодальної тональності;

Що стосується мультимедійних файлів, цього недостатньо для обробки однієї інформаційної модальності для примусового виконання аналізу

© Орел А.В., 2018

тональності у цьому змісті. У цьому контексті, (Poria et al. (2016)) представлено інноваційний підхід до аналізу мультимодальної тональності, який полягає у зборі тональності відео в Інтернеті через модель, яка поєднує звук, візуальну і текстову модальності як інформаційні ресурси. Функція вектору була створена в об'єднаному підході функції і рівня прийняття рішень, що становить приблизно 80%, представляючи збільшення на 20% точності при порівнянні на усі сучасні системи.

Maynard, Dupplaw, і Hare (2013) описують підхід для аналізу тональності на основі змісту соціальних мереж, поєднуючи мовлення в тексті та мультимедіа ресурси (наприклад, зображення та відео), зосереджуючись на об'єкті та події розпізнавання, щоб допомогти у виборі матеріалу для включення у соціальні мережі, щоб вирішити невизначеність і надавати більше контекстної інформації. Використовуються інструменти обробки природної мови (NLP) та підхід, заснований на правилах для тексту, що стосується питань, притаманних соціальним медіа, таких як граматично неправильний текст, використання профанації та сарказму.

– розпізнавання емоцій у аудіо та відео;

Ekman і Friesen (1978) показали, що емоції на обличчі можна визначити швидкими ознаками. Ці сигнали характеризуються зміною зовнішнього вигляду обличчя, це секунди або частки секунди. Таким чином, автори сформулювали модель основних емоцій, що називається Facial Action Coding System (FACS), заснована на шести виразах обличчя (щастя, сюрприз, відраза, гнів, страх і смуток), і які зустрічаються в багатьох культурах і представлені однаково, у дітей та літніх людях. Точки сингулярності були підключені до кожного типу виразу обличчя за допомогою тестів на великій базі даних зображень.

Що стосується визначення просодичних (мелодика, відносна сила вимови слів та їхніх частин, співвідношення відрізків мовлення по довготі, загальний темп мовлення, паузи, загальне темброве забарвлення тощо) особливостей в звукові модуляції аудіо сигналів, Eyben та співавтори (2013) показали розвиток openSMILE, рамки для детекції особливостей емоційної мови, музики та звуків з відео та аудіо сигналів. Детекція активності, голосовий моніторинг та детекція обличчя також є ресурсами, що запропоновані цією схемою.

Виклад основного матеріалу. Ми можемо спостерігати, що існує потреба в аналізі емоційного змісту відео та новин у багатьох видах медіа. У цьому контексті застосовуються надійні методи, які дозволяють автоматично визначати рівні емоційної напруги, використовуючи методи аналізу тональності у відео новинах для контент-аналізу новинних програм.

1. Пропонований підхід

У цьому розділі представлені мультимодальні функції та підхід до об'єднання текстової, візуальної та аудіо інформації для виконання обчислень рівнів напруженості у наративному змісті із подій, показаних у відеороликах.

– мультимодальні особливості;

У роботі мультимодальні особливості організовані у дві групи: аудіовізуальні та оцінки тональності текстових речень, отриманих із субтитрів (текстова інформація). Аудіовізуальна група представлена візуальною інтенсивністю розпізнаних емоцій, розмірами обчислювальних областей учасників та просодичними особливостями аудіо сигналу, що відповідають аспектам мовлення, які виходять за рамки фонем і що займаються якістю звуку: ймовірністю голосу, гучністю і основною частотою.

– візуальна інтенсивність;

Ця функція була виміряна на основі вихідного поля, іншими словами, відстань до поділу гіперплощини класифікаторів, що використовуються (Bartlett та співавтори (2006 р.)) під підхід "один до одного", де використовувався лінійний SVM класифікатор для кожної змодельованої емоції (щастя, сюрприз, відраза, презирство, гнів, страх і смуток). Візуальна інтенсивність емоції обчислюється для кожного кадру відео новини. У кадрах, де не було виявлено обличчя, система вираховує емоцію як неіснуючу.

– розмір обчислювальної області учасників;

Обчислюємо співвідношення між областю розпізнавання найбільшого масштабу обличчя, виявленого в кадрі та поточною роздільною здатністю кадру, для визначення пропорції значень обличчя учасників в межах розташування камери. У всесвіті телевізійних новин, коли є кілька осіб у різних обчислювальних розмірах областей відносно розміщення камери, фокус буде на людині з найближчим розміром обчислювальної області, тобто обличчя займає найбільшу площу у кадрі. Крім того, чим більша площа, яку обличчя займає у кадрі, тим більше емоційна інтенсивність, яка була застосована як комунікаційна стратегія програми в різноманітному використанні розміру області під час демонстрації (Gutmann, 2012).

– голосова ймовірність;

Показує ймовірність отримання тонової диференціації під час репліки в наступний момент.

– гучність звуку;

Гучність звуку відображає сприйняття гучності звукової хвилі людським вухом і вимірюється в децибелах (дБ).

– фундаментальна частота;

Відповідає першій збалансованій звуковій хвилі, це найвпливовіша частота для сприйняття певного звуку і одного з головних елементів для характеристики голосу (Eyben та співавтори, 2013).

– оцінки тональності;

Витягуються із субтитрів. Кожне речення формується з тексту субтитрів і аналізується 18 найновішими методами (Araujo та співавтори, 2014) для аналізу тональності тексту, що генерують вектор з 18 балів (-1, 0, або +1), по одному для кожного методу. Підраховуємо його значення, щоб отримати оцінку тональності для цього речення.

2. Обчислення рівня напруженості

У статті ми отримуємо дані з аудіо просодичних особливостей, візуальних особливостей з відео, і оцінок тональності, що витягуються з субтитрів. Після цього, ми визначаємо відповідні рівні напруженості. Рисунок 1 представляє огляд пропонованого підходу.

На кроці 1, розроблена система витягує мультимодальні ресурси з телевізійних новин, де кожен розміщений відеоролик містить різні новини. Можемо отримати список зображень з відео (кадрів), аудіо сигнал у форматі WAV і текст,

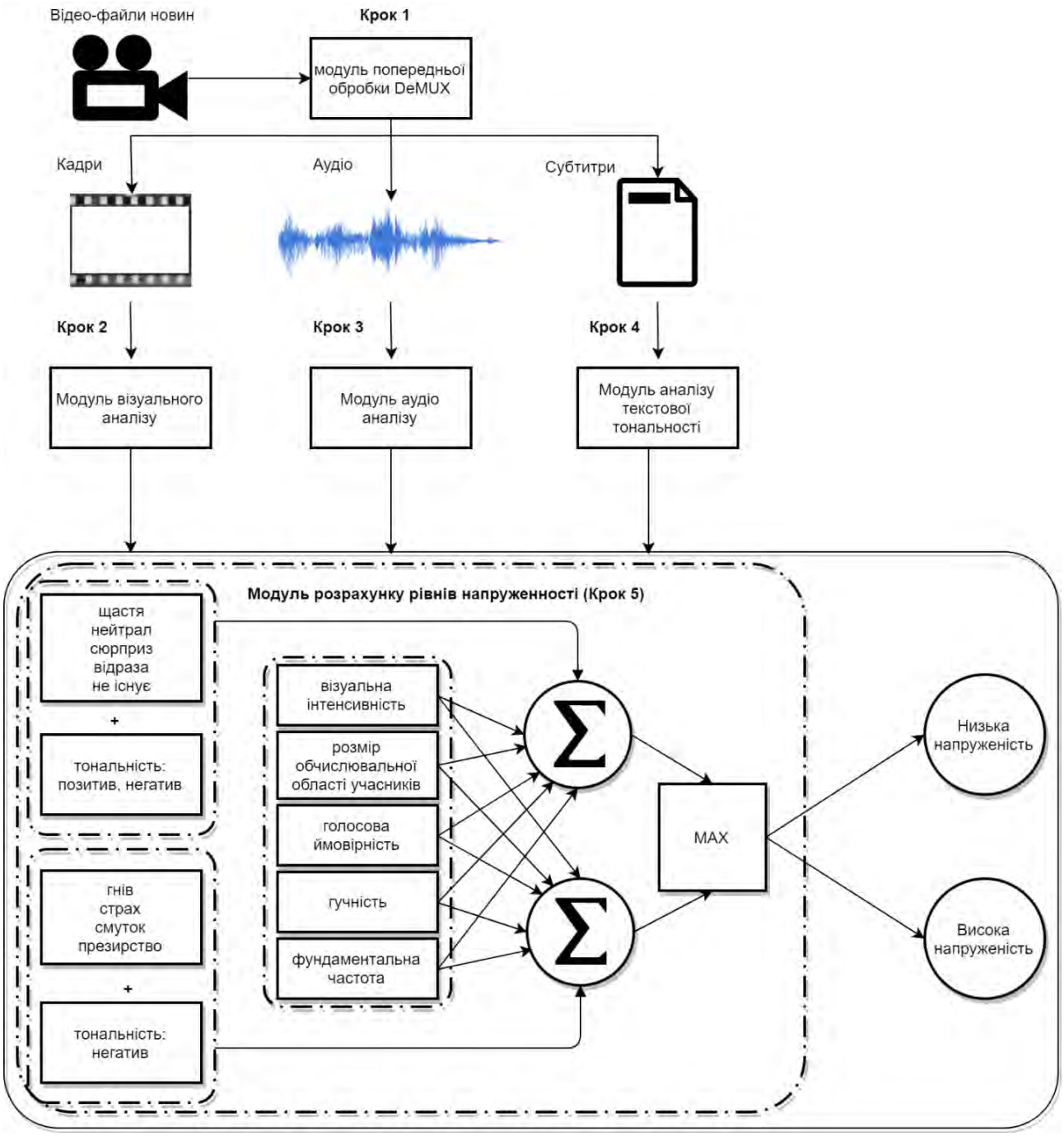

Рис. 1. Огляд запропонованого підходу в аналізі тональності для розрахунку рівнів напруженості

отриманий із субтитру. Ці модальності візуальної, звукової та текстової інформації організовані у трьох технологічних лініях, по одному для кожної модальності.

На кроці 2, використовуючи отримані кадри з попереднього кроку, застосовуємо методи вираження обличчя та розпізнані емоції (запропоновані Bartlett та співавторами (2006)), отримавши значення для візуальних особливостей (візуальна інтенсивність і розмір обчислювальної області зображення) на емоційному виразі обличчя.

На кроці 3, аудіо сигнал обробляється шляхом вилучення відповідних акустичних даних, також враховуючи мовні моменти людей, тобто, коли лунає мова. Для досягнення цього ми використовуємо фреймворк openSMILE (Eyben та співавтори, 2013) для вилучення аудіо компоненту спектру і отримання просодичних особливостей гучності, враховуючи ймовірність і основну частоту модуляції мови на кожну соту частину секунди цього аудіо сигналу.

На кроці 4 субтитр оброблюється для того, щоб витягнути лише текстовий зміст, що стосується транскрипції мови (текст субтитрів). Після цього кожний текст субтитрів перетворюється в одне речення. Кожне речення записується в текстовий файл, а потім текст аналізує його тональність. Для досягнення цієї мети використовується iFeel для визначення тональності полярності кожного речення, класифікуючи їх як позитивні, нейтральні чи негативні, для 18 методів (Ara'ujo та співавтори, 2014). Також звер-

немо увагу, що смайлики не використовувались, оскільки у субтитрі немає спеціального набору символів, що використовує запропонований метод смайликів для представлення емоцій (Park та співавтори, 2013).

Крок 5, показаний на малюнку 1, отримує рівень напруженості конкретного відео на основі суми з найвищим значенням серед двох рівнів напруженості: Низька напруженість і Висока напруженість. Суми емоцій, визначені у виразах обличчя та тональності, оцінки субтитрів для кожного рівня напруженості зважуються аудіовізуальною інформацією, розрахованій на відео.

Окрема відео новина класифікується за рівнем напруженості, емоції яких супроводжувалися найвищими оцінками мультимодальних функцій, що розраховані впродовж всього відео під час автоматичного розпізнання.

3. Заключні спостереження

Експресії обличчя – це форми невербального спілкування, що впливають на наші прояви подразників, в яких ми представлені і є елементами, що є частиною медіа новин. Для того, щоб легітимізувати повідомлений факт як комунікативну стратегію, телеканали висловлюють свої емоції, забезпечуючи докази напруженості мови, сформованого новинами. У цьому сенсі робота представлена як один з підходів до висновку про напруженість новинних відеороликів з урахуванням безлічі джерел показників.

Ці експерименти змогли показати, що аналіз тональності тексту, який вперше використовувався в цій галузі, може бути важливим для виявлення новин з високою напруженістю. Також було продемонстровано, що цей підхід може мати хороші результати для висновку про Високу напруженість новин. Однак, для ідентифікації напруженості новин, цей підхід гірше, ніж просто використання аналізу тональності тексту.

Висновки і пропозиції. У цій оглядовій статті було представлено огляд концепції та цілей аналізу мультимодальної тональності, а також обговорювалися проблеми та перспективи, пов'язані з цією галуззю. Огляд існуючої літератури свідчить про те, що аналіз мультимодальної тональності – це перспективний підхід до взаємодії каналів інформації для аналізу тональності і часто є кращими за унімодальні методи. Це також потенціал для посилення інших інструментів, які в даний час користуються аналізом унімодальних тональностей, такі як аналіз і розпізнавання суб'єктів. Цей огляд заохочує подальші крос-дисциплінарні зусилля у дослідженні цього нового питання.

Список літератури:

1. Ararujo, M.; Goncëalves, P.; Cha, M.; and Benevenuto, F. 2014. iFeel: A System that Compares and Combines Sentiment Analysis Methods. In Proc. of WWW'14.
2. Bartlett, M.S.; Littlewort, G.; Frank, M.; Lainscsek, C.; Fasel, I.; and Movellan, J. 2006. Fully Automatic Facial Action Recognition in Spontaneous Behavior. In Proc. Of FGR'06, 223–230. Southampton: IEEE.
3. Charaudeau, P. 2002. A Communicative Conception of Discourse. Discourse studies 4(3): 301–318.
4. Cheng, F. 2012. Connection between News Narrative Discourse and Ideology based on Narrative Perspective Analysis of News Probe. Asian Social Science 8: 75–79.
5. Eisenstein, J.; Barzilay, R.; and Davis, R. 2008. Discourse Topic and Gestural Form. In Proc. of AAAI'08, 836–841.
6. Ekman, P., and Friesen, W. 1978. Facial Action Coding System (FACS): Manual. Consulting Psychologists Press.
7. Eyben, F.; Weninger, F.; GroЯ, F.; and Schuller, B. 2013. Recent Developments in openSMILE, the Munich Open-Source Multimedia Feature Extractor. In Proc. of ACM MM'13, 835–838.
8. Goffman, E. 1981. The Lecture. In Forms of talk. Pennsylvania: University of Pennsylvania Press. 162–195.
9. Gutmann, J. 2012. What Does Video-Camera Framing Say during the News? A Look at Contemporary Forms of Visual Journalism. Brazilian Journalism Research 8(2): 64–79.
10. Maynard, D., Dupplaw, D., and Hare, J. 2013. Multimodal Sentiment Analysis of Social Media. Proc. of BCS SGAI SMA'13 44–55.
11. Pantti, M. 2010. The Value of Emotion: An Examination of Television Journalists' Notions on Emotionality. European Journal of Communication 25(2): 168–181.
12. Park, J., Barash, V., Fink, C., and Cha, M. 2013. Emoticon Style: Interpreting Differences in Emoticons Across Cultures. In Proc. of ICWSM'13.
13. Pereira, M.H.R., Pŕadua, F.L.C., and David-Silva, G. 2015. Multimodal Approach for Automatic Emotion Recognition Applied to the Tension Levels Study in TV Newscasts. Brazilian Journalism Research 11(2): 146–167.
14. Poria, S., Cambria, E., Howard, N., Huang, G.B., and Hussain, A. 2016. Fusing Audio, Visual and Textual Clues for Sentiment Analysis from Multimodal Content. Neurocomputing 174: 50–59.
15. Reis, J., Benevenuto, F., Vaz de Melo, P., Prates, R., Kwak, H., and An, J. 2015. Breaking the News: First Impressions Matter on Online News. In Proc. of ICWSM'15.
16. Stegmeier, J. 2012. Toward a computer-aided methodology for Discourse Analysis. Stellenbosch Papers in Linguistics 41: 91–114.

Орел А.В.
Национальный технический университет Украины
«Киевский политехнический институт имени Игоря Сикорского»

АНАЛИЗ ТОНАЛЬНОСТИ ОБЪЕДИНЕНИЯ ТЕКСТОВЫХ, АУДИО И ВИЗУАЛЬНЫХ ДАННЫХ

Аннотация
Статья представляет новый подход к выполнению диагностики анализа тональности новостных видеороликов, основанный на объединении текстовой, аудио и визуальной информации, взятые из их содержания. Предлагаемый подход имеет целью способствовать семи-дискурсивному изучению относительно построения идентичности всемирных СМИ, которые стали важной частью жизни миллионов людей. Для достижения этой цели применяем самые современные вычислительные методы для автоматического распознавания эмоций по мимике, выдержки модуляций выступлений участников и анализ тональности из субтитров видео, отвечающих интересам. Более подробно вычислим такие функции, как, например, зрительные интенсивности распознанных эмоций, размеры вычислительных областей участников, голосовая вероятность, громкость звука, фундаментальные частоты языка и оценки (полярности) тональности из текстовых предложений в субтитрах. Экспериментальные результаты из набора данных, содержащих 520 аннотированных видео новостей популярных телевизионных каналов показывают, что этот подход достигает точности до 84% в задачах классификации тональностей (уровни напряженности), что демонстрирует его высокий потенциал для использования мультимедийными аналитиками в нескольких направлениях, особенно в журналистской области.
Ключевые слова: аудио тональность, визуальная тональность, тональность текста, анализ мультимодальной тональности, распознавание эмоций.

Orel A.V.
National Technical University of Ukraine
“Igor Sikorsky Kyiv Polytechnic Institute”

SENTIMENT ANALYSIS OF AUDIO, VISUAL AND TEXTUAL DATA

Summary
This review article introduces a new method for sentiment analysis on an example of news videos based on a combination of text, audio and visual information. The method aimed at conducting research using modern methods of emotional recognition using facial expressions, audio diagnostics and video sentiment analysis. Experimental results show that the approach achieves accuracy up to 84% in the tasks of sentiment classification (tension levels). This allows us to say that it is very promising for multimedia analytics in the journalistic sphere and not only.
Keywords: audio sentiment analysis, visual sentiment analysis, sentiment analysis, multimodal sentiment analysis, emotions recognition.