УДК 004.7 **doi: 10.26906/SUNZ.2022.3.115**

А. І. Кулягін

Національний аерокосмічний університет імені М. Є. Жуковського «ХАІ», Харків, Україна

ВИКОРИСТАННЯ РОЗПІЗНАНОЇ ЕМОЦІЇ ЯК НЕЯВНОГО ФІДБЕКУ ДЛЯ РЕКОМЕНДАЦІЙНОЇ СИСТЕМИ

Анотація. Актуальність. Через зростання цифровізації мистецтва постають задачі покращення імерсивності під час взаємодії користувача з системами extended reality мистецтвом. **Методи дослідження.** Глибока нейронна мережа з шарами занурення, 3D згорткова нейронна мережа. **Мета статті:** Покращення відбору найбільш релевантних відео за допомогою використання розпізнаних емоцій користувача як неявного фідбеку в рекомендаційній системі віртуальних арт-композицій. **Отримані результати.** Було розроблено систему для класифікації емоцій користувача на відео, подальшої калькуляції емоційного скорингу та використанні отриманого значення в якості неявного фідбеку для рекомендаційної системи відбору найбільш релевантних відео для створення віртуальних арт композицій. Поєднання представлених методів дозволить покращити персоналізації рекомендацій та її збільшити імерсивність час взаємодії користувача з віртуальними арт-композиціями. **Висновок.** Розроблений у роботі підхід може бути використаний для покращення імерсивності та персоналізації рекомендацій під час взаємодії користувача з системами extended reality мистецтвом.

Ключові слова: глибока нейронна мережа з шарами занурення, extended reality, імерсивність, 3D згорткова нейронна мережа.

Вступ

У результаті зростаючої цифровізації мистецтва, новітніх способів взаємодії користувача з творами мистецтва, виникають проблеми з покращенням ефекту занурення під час взаємодії користувача з системами розширеної реальності (XR). Імерсивність, або ступінь занурення в середовище, є ключовим елементом систем XR мистецтва, особливо коли мова йде про мобільні додатки.

Перспективним напрямком є поєднання живопису та технологій VR/AR/MR в рамках XR [1-3].

Поєднуючи традиційні техніки малювання з технологіями віртуальної, доповненої або змішаної реальності, художники та розробники можуть створювати імерсивні та захопливі враження для користувачів. Це злиття дозволяє користувачам зануритися у віртуальний світ, де вони можуть взаємодіяти з цифровими творами мистецтва, досліджувати віртуальні галереї або навіть брати участь в інтерактивному мистецтві. Завдяки використанню XR межі між фізичним і цифровим мистецтвом стираються, відкриваючи нові можливості для творчості та художнього вираження [4].

Одним із аспектів покращення ефекту занурення в мистецькі системи є адаптація контенту та інтерфейсу до потреб і вподобань користувача. Зокрема, використання неявного зворотного зв’язку з користувачем може сприяти підвищенню ефекту занурення та спрощенню інтерфейсу користувача. Між тим, перенасичення системи явним зворотним зв'язком може знизити рівень занурення та перевантажити інтерфейс [5-7].

Наша ціль — розробити мобільний додаток, який би дозволяв користувачеві сканувати зображення-маркер за допомогою камери, яке б слугувало AR-якорем і відображало відео, завантажене з сервера, на місці даного якоря.

Картини в музейній композиції виконуватимуть роль маркерних зображень. Сукупність зображення та доданого до нього відео будемо називати віртуальною художньою композицією. Розпізнавання зображення маркера відбувається на сервері. Після розпізнавання система повинна вибрати відео, яке найбільше відповідає вподобанням користувача (за жанром, колірною гамою, композицією тощо). Відбір здійснюється з урахуванням явних (сподобалася композиція чи ні, рейтингова оцінка, додаткова анкета) і неявних (чи користувач додивився відео до кінця, як довго користувач зосереджувався на композиції тощо) фідбеків користувачів. Для цього ми використовуємо гібридну рекомендаційну систему [8].

Початкове визначення вподобань користувача здійснюється за допомогою анкетування, яке користувач може заповнити на початку використання програми. За результатами цього анкетування до профілю користувача додаються дані, які можуть вказувати на його переваги, такі як: стать, вік, улюблений колір, улюблені жанри живопису та інші [9]. Задача вибору відеоролика, який найбільше відповідає вподобанням користувача, буде вирішена за допомогою рекомендаційних систем.

У цій роботі ми припускаємо, що покращення відбору найбільш релевантних відео для створення віртуальної арт-композиції за допомогою рекомендаційних систем можна досягти шляхом використання розпізнаних емоцій користувача як неявного зворотного фідбеку користувача.

Мета статті: ідея даного дослідження полягає в наступному: покращити вибір найбільш релевантних відео за допомогою використання розпізнаних емоцій користувача як неявного фідбеку в рекомендаційній системі віртуальних арт-композицій.

Виклад основного матеріалу

Неявні фідбеки користувачів — це цінна інформація про поведінку користувачів, яка дозволяє визначити їхні вподобання та інтереси. Вони можуть включати різноманітні показники, такі як тривалість відео, його завершеність перегляду, частота натискань на ті чи інші рекомендації, аналіз відгуків і багато іншого [10].

Ефективне використання неявних відгуків користувачів може революціонізувати роботу рекомендаційної системи. Надаючи більше інформації про вподобання користувача, ніж явні відгуки, наприклад оцінки чи відгуки, вони створюють вищий рівень персоналізації. Ці додаткові дані дозволяють системі отримати глибше розуміння вподобань користувачів і, отже, надавати більш точні та цілеспрямовані рекомендації.

Під час впровадження збору неявних користувацьких фідбеків необхідно ретельно розглянути деякі важливі аспекти. Давайте розглянемо деякі з них.

Контекст. Неявний фідбек може залежати від контексту, в якому користувач взаємодіє з продуктом або контентом. Час доби, день тижня або поточні події можуть по-різному впливати на інтереси користувачів.

Тип взаємодії. Різні типи неявного фідбеку можуть мати різне значення для рекомендаційної системи. Наприклад, додавання продукту в кошик може свідчити про більший інтерес користувачів, ніж просто перегляд сторінки продукту.

Нормалізація фідбеку. Користувачі можуть взаємодіяти із системою по-різному, і це слід враховувати під час оцінки значення неявного фідбеку. Системи рекомендацій можуть нормалізувати неявний зворотний зв'язок, порівнюючи поведінку окремого користувача із сукупною статистикою або порівняно з іншими користувачами.

Моніторинг змін у поведінці користувачів. Важливість неявного фідбеку може змінюватися з часом залежно від поточних інтересів користувачів.

Ваги для різних джерел фідбеку. У гібридних рекомендаційних системах, які використовують як явний, так і неявний фідбеки, може бути важливо надавати різні ваги різним типам фідбеків.

Y. Hu, Y. Koren і C. Volinsky у своїй статті «Collaborative Filtering for Implicit Feedback Datasets» (2008) [11] досліджують цю тему, розробляють модель, яка використовує неявний фідбек для визначення вподобань користувача, і представляють нові методи оцінки. для таких систем.

Неявні відгуки користувачів є потужним інструментом для вдосконалення систем рекомендацій, але їх слід ретельно проаналізувати, враховуючи контекст, різні типи взаємодії, необхідність нормалізації відгуків, зміни в поведінці користувачів і потребу у вагових коефіцієнтах для різних джерел відгуків. Це дозволить налаштувати систему таким чином, щоб надавати користувачам найбільш актуальні рекомендації. Ці принципи відіграють ключову роль у впровадженні та оптимізації використання неявних відгуків користувачів у рекомендаційних системах.

У нашому дослідженні ми обмежуємося використанням відеоданих для розпізнавання емоцій, хоча пристрої для сканування обличчя, такі як системи розпізнавання обличчя (TrueDepth) на смартфонах, стають все більш поширеними в наш час. Ці пристрої можуть надати більш детальну інформацію про міміку та емоційний стан користувача.

Однак для наших конкретних цілей, рекомендацій AR сесій для мистецтва, лише відеоданих недостатньо. Наша система рекомендацій спрямована на вибір і рекомендацію AR сесій, які найкраще відповідають емоційному стану користувача під час перегляду арт-відео.

Таким чином, хоча пристрої для сканування обличчя можуть бути потенційно корисними для отримання детальніших даних про емоційний стан користувача, вони не потрібні для наших конкретних цілей. Замість цього ми зосереджуємось на використанні 3D згорткової нейронної мережі для аналізу відеоданих і класифікації емоцій, які вже надають достатньо інформації для системи рекомендацій AR-сесій мистецтва.

Вираз обличчя сильно корелює з рухом обличчя. Залежно від того, чи використовується часова інформація про рух обличчя, риси обличчя можна класифікувати як статичні або динамічні. Перший, який в основному включає геометричні об’єкти та особливості зовнішнього вигляду, можна отримати за допомогою згортки або інших фільтрів навчання; останні, які спрямовані на моделювання динамічних властивостей руху обличчя, можуть бути розраховані відповідно за допомогою оптичного потоку або інших методів.

Коли вводяться тривимірні згорткові нейронні мережі (3D CNN), вилучення двох різних типів ознак, згаданих вище, стає легшим [12].

Наша мета — адаптувати тривимірну згорткову нейронну мережу для класифікації відео, відому зі статті «A Closer Look at Spatiotemporal Convolutions for Action Recognition» [13], щоб виявляти емоції та інтерпретувати їхню послідовність як неявний відгук користувача в нашій системі рекомендацій.

На рис. 1 показані шари оригінальної 3D моделі CNN, яку ми прагнемо адаптувати для розпізнавання емоцій.

Ця модель працює з відеопотоком, який дискретизується на окремі кадри для подальшого аналізу. Кожне відео обробляється за допомогою тривимірної згорткової нейронної мережі (3D CNN) для класифікації емоцій користувача та визначення важливості класифікованих емоцій. Використовуючи 3D CNN, можна використовувати тривимірний фільтр для виконання згорток і просторово-часової обробки відеоданих.

На етапі попередньої обробки ми налаштували частоту кадрів відео, розміри (змінивши ширину та висоту відео до 224 пікселів) і сегментували його на відрізки по 2 секунди. Вибір оптимальної частоти кадрів відіграє вирішальну роль у нашій системі, оскільки це допомагає мінімізувати затримку розпізнавання та покращити загальну продуктивність. При цьому частота кадрів повинна бути достатньою, щоб оцінити зміни в обличчі користувача.

Обрано 5 кадрів в секунду як задовільну для нашої моделі.

Потрібно адаптувати 3D CNN для аналізу послідовностей кадрів, де кожен кадр вважається одним «часовим кроком». Це дозволяє моделі розпізнавати інформацію, яка поширюється з часом, наприклад динаміку виразу обличчя, що є критично важливим для розпізнавання емоцій.

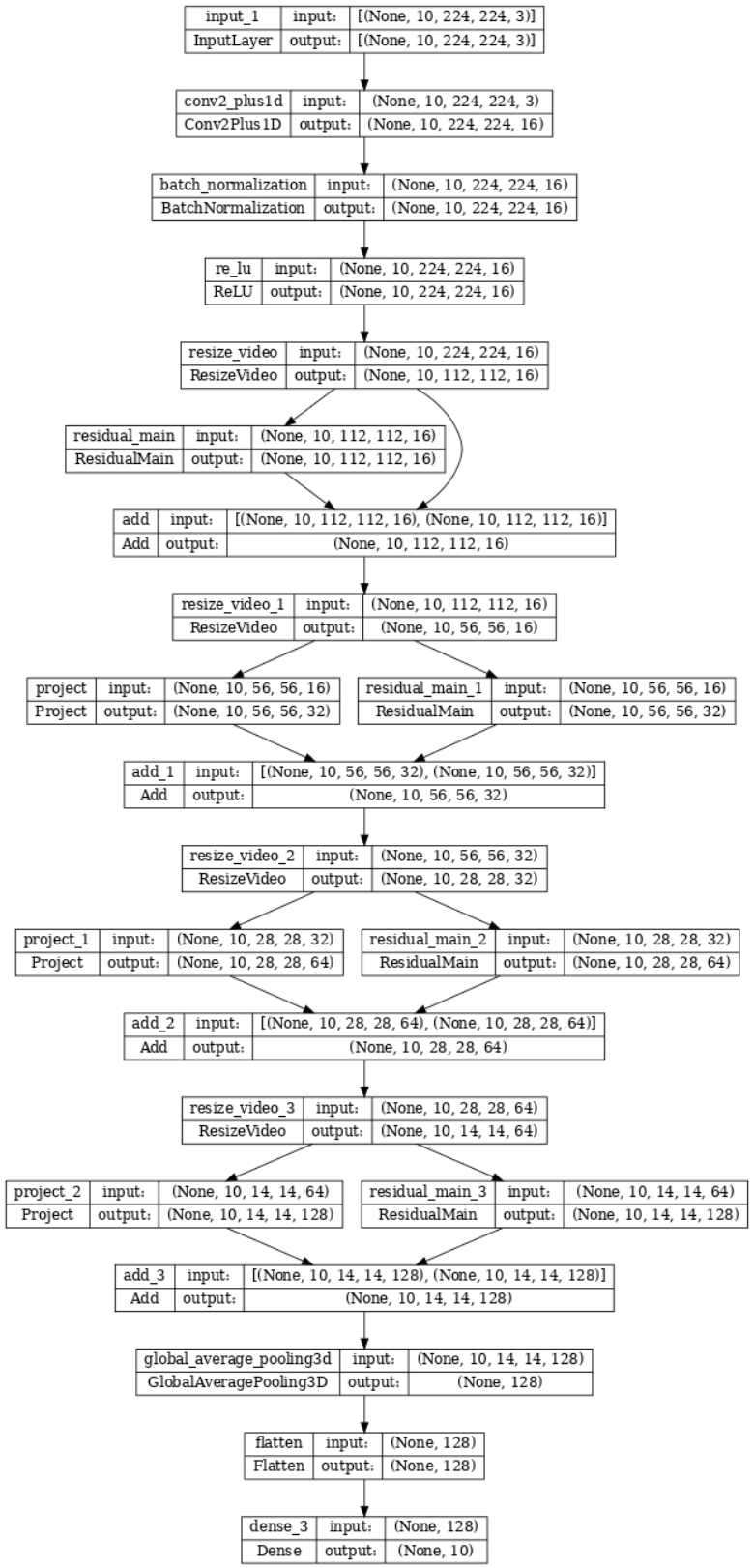

Рис. 1. Шари оригінальної 3D CNN

Набір даних FER (Facial Expression Recognition) – це набір даних, який широко використовується для навчання та оцінювання моделей для завдань розпізнавання виразу обличчя. Він складається з колекції зображень у відтінках сірого з виразами облич людей. Основні проблеми, які виникають під час розробки системи FER, стосуються неявних особливостей і зміщень, спричинених різними культурами та умовами збору. Поточний набір даних має сильне вбудоване зміщення, а відповідні запропоновані методи показують, що умовний розподіл ймовірностей між навчальними та тестовими наборами даних відрізняється [14].

Обличчя на відео аналізується 3D CNN, який навчений розпізнавати емоції за допомогою CMU-MOSI. Набір даних Multimodal Corpus of Sentiment Intensity (CMU-MOSI) — це колекція 2199 відеокліпів, що висловлюють думки. Кожне відео з думкою додається в діапазоні [-3,3].

Оскільки ми адаптували оригінальний 3D CNN до моделі, яка може класифікувати емоції, ми використовуємо його для обробки коротких сегментів відео (2 секунди кожен) для підвищення точності, окремо алгоритмічно обробляючи результати класифікації кожного сегмента. Ми виконуємо цю дію для відео з передньої камери користувача під час кожного сеансу AR.

Набір даних суворо анотований мітками для суб’єктивності, інтенсивності настроїв, анотованих візуальних характеристик за кадром і кожною думкою, а також анотованих звукових характеристик за мілісекунди [15].

Результатом цього кроку є визначення емоції та її інтенсивності для обличчя користувача в кадрі. Щоб перетворити ці дані на неявний зворотний зв’язок, ми розробляємо емоційну оцінку для кожного сеансу AR, у якому ми підсумовуємо вимірювання емоцій користувача в одне числове значення для сеансу, яке може бути позитивним, нейтральним або негативним. Модель, що представляє вхідні дані, показана на рис. 2.

Щоб перетворити вихідні дані моделі в емоційну оцінку (ES) сеансу AR з N кадрами, ми можемо використати таку формулу:

$$ES = (\sum_{i=1}^{N} e_i * s_i)/N, \qquad (1)$$

де e_i — значення емоцій (-1 для негативних, 0 для нейтральних, 1 для позитивних) для кадру i, s_i — важливість емоцій для кадру i (від 0 до 1).

Рис. 2. Модель вхідної послідовності

Наш процес базується на розділенні відеопотоку на окремі кадри, які обробляються за допомогою 3D

CNN для класифікації емоцій обличчя користувача за допомогою набору даних CMU-MOSI. За результатами цього етапу ми отримуємо визначення емоції та її інтенсивності для обличчя користувача в кадрі. Щоб перетворити ці дані на відгуки, ми будемо використовувати систему підрахунку емоційних балів для кожної сесії AR, у якій ми підсумовуємо вимірювання емоцій користувача в одне числове значення за сеанс, яке може бути позитивним, нейтральним або негативним. Отже, на основі цієї частини ми можемо сформувати набір емоційних вимірювань, який може слугувати ефективним інструментом для забезпечення неявного зворотного зв'язку в системі рекомендацій. Це, у свою чергу, може збільшити здатність системи краще реагувати на вподобання та потреби користувачів. Повний процес показано на рис. 3.

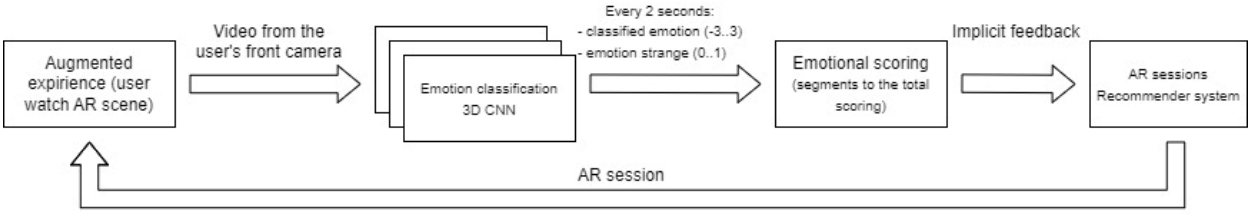

Рис. 3. Повна послідовність системи

Нейронні мережі, які аналізують відеодані користувача та розпізнають емоції, можуть покращити результати рекомендаційних систем. Використовуючи 3D CNN для виявлення емоцій у відео, ми створюємо основу для генерування неявних користувацьких фідбеків. Такий підхід допомагає зібрати важливу інформацію про емоційну реакцію користувачів на різний контент. За допомогою формули емоційного скорингу для кожної AR сесії ми можемо кількісно оцінити емоційну реакцію користувача на кожну рекомендовану AR сесію.

Ми модифікували нашу стару модель гібридної рекомендаційної системи, яка поєднує в собі метод спільної фільтрації, метод рекомендацій, заснований на знаннях, і глибоку нейронну мережу з тонучими шарами [8], включивши значення емоційного скорингу як неявний фідбек у вхідні дані.

Завдяки гнучкості сучасного апарату нейронних мереж стало можливим поєднати різні підходи до створення рекомендаційних систем в рамках однієї моделі. Ми додали дві окремі глибокі нейронні мережі до моделі Neural Collaborative Filtering, які використовують інформацію про властивості AR сесії та інформацію профілю користувача для вибору відео. Це дозволило нам покращити якість рекомендацій, які ми отримуватимемо від системи.

Шари старої гібридної моделі, яку ми адаптуємо, показані на рис. 4.

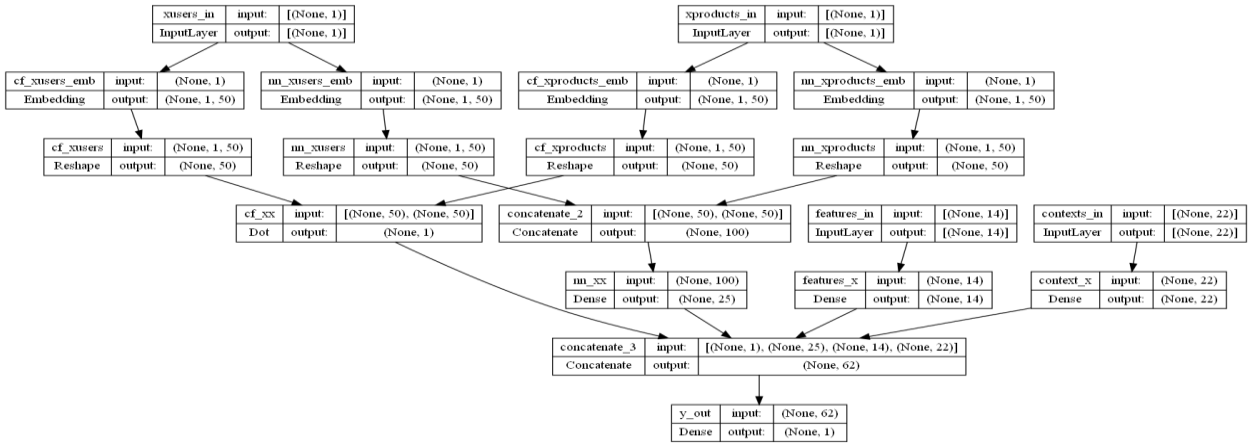

Рис. 4. Шари гібридної моделі

Таким чином, після обробки результатів 3D CNN для розпізнавання емоцій у відео з оцінкою емоційності для кожної сесії AR ми додаємо результат як неявний фідбек до вхідної моделі нашої рекомендаційної системи. Наша мета — розширити модель, додавши новий вхідний параметр емоційного скорингу, щоб використовувати його як неявний фідбек.

Використання трьох можливих варіантів емоційного фідбеку (негативний, нейтральний, позитивний) на відміну від двох (негативний, позитивний) дасть нам більше гнучкості при визначенні ваги цього відгуку. Адже нейтральний емоційний фідбек виключає помилкові оцінки під час невизначеності чи відсутності емоцій, що, у свою чергу, не вимагатиме значного коливання ваги фідбеку, а також загрози встановлення великих ваг.

Ми очікуємо, що в результаті наших досліджень і оновлення системи рекомендацій вдасться отримати підвищення точності рекомендацій. Використання оцінки розпізнаних емоцій на відео як неявного зворотного зв'язку також дозволить системі краще зрозуміти вподобання користувачів і реакцію на вміст. Це може допомогти системі рекомендацій створити більш детальний профіль користувача та надати більш персоналізовані рекомендації, які відповідають його емоційному стану.

Висновки

Було запропоновано використовувати результати тривимірної згорткової нейронної мережі (CNN) для аналізу відео та розпізнавання емоцій користувача як даних для алгоритму оцінки емоційної сесії.

Подальше застосування емоційної оцінки як неявного фідбеку для гібридної рекомендаційної системи.

Також пропонується використовувати три можливі варіанти емоційного зворотного зв'язку (негативний, нейтральний, позитивний) замість двох (негативний, позитивний), що дасть більшу гнучкість при визначенні ваги цього відгуку.

Розроблений підхід може бути використаний для покращення імерсивності та персоналізації рекомендацій під час взаємодії користувача з системами extended reality мистецтвом.

Подальші дослідження доцільно проводити у напрямі поєднання існуючої системи з генерацією віртуальних арт-композицій з метою збільшення вибірки рекомендацій та нівелювання обмеженої кількості мануально створених віртуальних арт-композицій.

СПИСОК ЛІТЕРАТУРИ

1. F. Ye, "Image Art Innovation based on Extended Reality Technology," 2022 7th IEEE International Conference on Data Science in Cyberspace (DSC), Guilin, China, 2022, pp. 584-587, doi: 10.1109/DSC55868.2022.00087.
2. Gironacci, Irene. (2021). State of the Art of Extended Reality Tools and Applications in Business. 10.4018/978-1-7998-4339-9.ch008.
3. Caarls, Jurjen & Jonker, Pieter & Kolstee, Yolande & Rotteveel, Joachim & Eck, Wim. (2009). Augmented Reality for Art, Design and Cultural Heritage—System Design and Evaluation. EURASIP J. Image and Video Processing. 2009. 10.1155/2009/716160.
4. Lai, Chi-Hui & Chen, Chun-Chih & Wu, Shu-Ming. (2023). Analysis of Key Factors for XR Extended Reality Immersive Art Experience. International Journal of Social Sciences and Artistic Innovations. 3. 24-36. 10.35745/ijssai2023v03.01.0004.
5. Wang, Fei. (2023). Research on the application of immersive art in digital technology scene. Advances in Education, Humanities and Social Science Research. 5. 88. 10.56028/aehssr.5.1.88.2023.
6. Zhang, Ying. (2023). Immersive Multimedia Art Design Based on Deep Learning Intelligent VR Technology. Wireless Communications and Mobile Computing. 2023. 1-8. 10.1155/2023/9266522.
7. Ha, Taejin & Kim, Yeongmi & Ryu, Jeha & Woo, Woontack. (2006). Enhancing Immersiveness in AR-Based Product Design. 4282. 207-216. 10.1007/11941354_22.
8. Kuliahin, A. & Narozhnyi, V. & Tkachov, V. & Kuchuk, H.. (2022). Study of methods of building recommendation system for solving the problem of selecting the most relevant video when creating virtual art compositions. Control, navigation and communication systems. Collection of scientific papers. 4. 94-99. 10.26906/SUNZ.2022.4.094.
9. Zhao, Qian & Harper, Franklin & Adomavicius, Gediminas & Konstan, Joseph. (2018). Explicit or implicit feedback? engagement or satisfaction?: a field experiment on machine-learning-based recommender systems. SAC '18: Proceedings of the 33rd Annual ACM Symposium on Applied Computing. 1331-1340. 10.1145/3167132.3167275.
10. Yang, Zhen. (2022). Research on Personalized Product Recommendation Algorithm for User Implicit Behavior Feedback. 10.1007/978-981-19-6901-0_149.
11. Hu, Yifan & Koren, Yehuda & Volinsky, Chris. (2008). Collaborative Filtering for Implicit Feedback Datasets. Proceedings - IEEE International Conference on Data Mining, ICDM. 263-272. 10.1109/ICDM.2008.22.
12. Zhao, Jianfeng & Mao, Xia & Zhang, Jian. (2018). Learning deep facial expression features from image and optical flow sequences using 3D CNN. The Visual Computer. 34. 10.1007/s00371-018-1477-y.
13. D. Tran, H. Wang, L. Torresani, J. Ray, Y. LeCun and M. Paluri, "A Closer Look at Spatiotemporal Convolutions for Action Recognition," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018, pp. 6450-6459, doi: 10.1109/CVPR.2018.00675.
14. Umer, Saiyed & Rout, Ranjeet & Hossain, Sanoar & Asari, Vijayan. (2021). A Unified Framework of Deep Learning-Based Facial Expression Recognition System for Diversified Applications. Applied Sciences. 11. 10.3390/app11199174.
15. Zadeh, A., Liang, P.P., Poria, S., Vij, P., Cambria, E., & Morency, L.P. (2016). CMU-MOSI Dataset (Version 1.0) [Data set]. CMU Multimodal SDK. http://multicomp.cs.cmu.edu/resources/cmu-mosi-dataset/.

Received (Надійшла) 24.04.2022
Accepted for publication (Прийнята до друку) 23.08.2023

Using recognized emotion as implicit feedback for a recommender system

A. Kuliahin

Abstract. Topicality. Due to the growing of digitalization of art, the tasks of improving immersiveness during user interaction with extended reality art systems arise. **Research methods.** Deep Neural Network with Immersion Layers, 3D Convolutional Neural Network. **The purpose of the article**: Improving the selection of the most relevant videos by using recognized user emotions as implicit feedback in the recommender system of virtual art compositions. **The results obtained.** A system was developed for classifying the user's emotions in the video, further calculating the emotional scoring and using the obtained value as implicit feedback for the recommender system for selecting the most relevant videos for creating virtual art compositions. The combination of the presented methods will allow to improve the personalization of recommendations and increase its immersiveness during user interaction with virtual art compositions. **Conclusion.** The approach developed in the work can be used to improve the immersiveness and personalization of recommendations during user interaction with extended reality art systems.

Keywords: deep neural network with immersion layers, extended reality, immersiveness, 3D convolutional neural network.