Intelligent information systems

UDC 004.89 **doi:** https://doi.org/10.20998/2522-9052.2024.1.09

Heorhii Kuchuk[1], Andrii Kuliahin[2]

[1] National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine
[2] National Aerospace University "Kharkiv Aviation Institute", Kharkiv, Ukraine

HYBRID RECOMMENDER FOR VIRTUAL ART COMPOSITIONS WITH VIDEO SENTIMENTS ANALYSIS

Abstract. Topicality. Recent studies confirm the growing trend to implement emotional feedback and sentiment analysis to improve the performance of recommender systems. In this way, a deeper personalization and current emotional relevance of the user experience is ensured. **The subject of study** in the article is a hybrid recommender system with a component of video sentiment analysis. **The purpose of the article** is to investigate the possibilities of improving the effectiveness of the results of the hybrid recommender system of virtual art compositions by implementing a component of video sentiment analysis. **Used methods**: matrix factorization methods, collaborative filtering method, content-based method, knowledge-based method, video sentiment analysis method. **The following results** were obtained. A new model has been created that combines a hybrid recommender system and a video sentiment analysis component. The average absolute error of the system has been significantly reduced. Added system reaction to emotional feedback in the context of user interaction with virtual art compositions. **Conclusion.** Thus, the system can not only select the most suitable virtual art compositions, but also create adaptive and dynamic content, which will increase user satisfaction and improve the immersive aspects of the system. A **promising direction** of further research may be the addition of a subsystem with a generative neural network, which will create new virtual art compositions based on the conclusions of the developed recommendation system.

Keywords: hybrid recommender system; collaborative filtering; matrix factorization; convolutional neural network; emotional feedback; video sentiment analysis.

Introduction

As the study conducted in the work "Study of methods for building recommendation system to solve the problem of selecting the most relevant video when creating virtual art compositions" [1] showed, the most effective approach to solving the problem of building a recommendation system of virtual art compositions is the hybridization approach. It consists in the combination within one model of different methods of building recommender systems, namely, the collaborative filtering method, the content-based method, and the knowledge-based method. The hybrid model, which combines all three methods, showed better results compared to models that implement each method separately. This is due to the fact that an additional deep neural network added to the hybrid of matrix factorization and collaborative filtering methods takes into account user characteristics when determining the video rating in the virtual art composition.

The next step was the task of increasing the efficiency of the hybrid recommender system of virtual art compositions. We hypothesized that by analyzing the emotions or moods of users and their intensity while viewing virtual art compositions, it is possible to obtain additional information about their preferences [2]. This can be achieved through the use of a recognized emotion, as implicit feedback of the user to one or another art composition. In this way, emotional feedback is added to the parameters of the hybrid recommender system, which characterizes user's interaction with the virtual art composition.

In this paper, we want to create a model for recognizing emotions and their intensity in videos and add it as a subsystem to our hybrid recommender system [1]. To conduct computational experiments that will allow testing the assumption that taking into account information about the emotional feedback of users during interaction with virtual art compositions will increase the efficiency of the hybrid recommender system of virtual art compositions.

1. Literature review

To achieve our goal, it was decided to implement a recommender system of virtual art compositions capable of recognizing the emotions of users on video while viewing virtual art compositions (Hybrid Recommender with Video Sentiments Analysis, HR-VSA). For this purpose, it was decided to add a subsystem to the hybrid model of the recommender system (Hybrid Recommender, HR) presented in the article [1], which is responsible for recognizing the emotional feedback of users and their intensity in the video (VSA).

1.1. Research on recommender systems in multimedia and art domains. We reviewed various current research related to recommender systems in multimedia and art domains to analyze trends and new approaches to improve personalization and user immersion. Identified trends include the growing integration of deep learning and artificial intelligence for detailed personalization, the increasing impact of contextual analysis on recommendations. As well as the development of interactive and controlled recommendation systems that increase user engagement and satisfaction.

For example, in the article "Multimedia Recommender Systems: Algorithms and Challenges" [3], the authors consider modern research related to multimedia recommender systems (MMRS). They focus on methods that integrate multimedia content as additional information to various recommendation

models. The authors claim that multimedia features can be used by MMRS to recommend either media items from which the features were derived, or non-media items using features derived from a proxy multimedia representation of the item (e.g., an image of clothing).

In the article "The Effects of Controllability and Explainability in a Social Recommender System" [4], the authors highlight the critical role of controllability and explainability in social recommender systems. They argue that giving users the ability to control and understand recommendations increases their trust and satisfaction with the system. The work analyzes various methods and approaches that allow users to change recommendation parameters and gain insight into the logic of recommendations, which, in turn, improves user interaction with the system and provides more accurate recommendations.

"Content-based Artwork Recommendation: Integrating Painting Metadata with Neural and Manually-Engineered Visual Features" [5] authors propose an innovative method of recommending works of art, emphasizing the integration of detailed metadata of paintings with visual characteristics processed using neural networks and careful manual feature extraction. This not only highlights the depth of analysis of visual art, but also shows the potential to create more dynamic and relevant recommendations that match users' aesthetic preferences.

The authors of the publication "Hybrid Recommendations and Dynamic Authoring for AR Knowledge Capture and Re-Use in Diagnosis Applications" [6] explore the potential of hybrid recommendation systems and dynamic authoring in the context of augmented reality for knowledge capture and its reuse in diagnostic applications. They point to the expansion of the capabilities of recommender systems beyond their traditional uses, emphasizing integration with complex technological solutions to improve diagnostic applications.

In addition to the specifics of application in specific domains, we were also interested in the specifics of using emotional feedback or user mood analysis as input parameters of a recommender system. Currently, there is an increase in the use of the emotional component and the analysis of the user's mood to improve recommendations. "Recommendation of Micro Teaching Video Resources Based on Topic Mining and Sentiment Analysis" [7] explores approaches to recommending educational video resources through topic and sentiment analysis. They demonstrate how a deep understanding of the context and emotional state of users can improve learning effectiveness and personalize educational content for better engagement.

The authors of "EARS: Emotion-Aware Recommender System Based on Hybrid Information Fusion" [8] describe the use of emotional feedback in connection with hybrid data integration to create more accurate recommendations. Thus, the authors emphasize the importance of data that adapts to the emotional component and their potential to increase user satisfaction from interaction with recommender systems.

In the article "Evaluating Facial Recognition Services as Interaction Technique for Recommender Systems" [9], the authors investigate the use of facial recognition technologies as an innovative way of interaction in recommender systems. They analyze how this technology can increase the accuracy of recommendations by interpreting users' emotional states and reactions in real time. This research points to the potential of facial recognition to create more personalized and emotionally resonant recommendations, improving user-system interaction. This approach has great potential to change traditional recommender systems by offering more dynamic and intuitive ways of interacting.

In "Emotion-Based Movie Recommendations" [10], the concept of using emotional reactions of users to create more accurate and personalized movie recommendations is considered. The authors explore how emotion analysis can help identify user preferences, increasing satisfaction with viewing recommendations and personalizing the experience. This study emphasizes the importance of the emotional component in recommender systems. Thus, making the selection of movies more relevant to the emotional state and current preferences of the user.

These studies confirm the growing trend of using emotional feedback and sentiment analysis to enrich and refine recommendations, providing deeper personalization and ongoing emotional relevance of the user experience. Our task is to create a spatiotemporal neural network model that will analyze emotions in a video, which will represent the user's emotional feedback from viewing a virtual art composition. We want to include this model as a subsystem in our hybrid recommender system [1]. Thus, the user's emotional feedback on the video will be used as the implicit feedback of the recommender system. This will make it possible to make more similar recommendations of virtual art compositions and improve immersiveness during user interaction with the system.

1.2. Research on video sentiment analysis. Let's analyze the latest publications related to VSA. Some current research focuses on sentiment analysis in YouTube comments. For example, in the article "Sentiment Analysis on Online Videos by Time-Sync Comments. Entropy" [11] authors use the methods of sentiment analysis and topic clustering to study educational content on YouTube. In particular, the authors consider how these techniques can be used to analyze comments to determine popular themes, emotional connection to material, and the overall effectiveness of educational content. The article presents different approaches to sentiment analysis, such as machine learning and deep learning, and their application to identify positive, negative, or neutral emotions in the textual content of comments.

Another study, "Learning Analytics on YouTube Educational Videos: Exploring Sentiment Analysis Methods and Topic Clustering" [12] describes the use of sentiment analysis for time-synchronized comments on videos. The authors focus on identifying and analyzing the emotional reactions of viewers at specific moments of the video, which allows for a dynamic understanding of content perception. The research findings show that this approach can be used to improve engagement and

optimize video content, and provide clues for video content editors about how video affects the emotional state of viewers.

But the specificity of our task allows us to use some channels in a very limited way during the analysis of emotional feedback. In augmented reality environments, where we will present virtual art compositions, often only the video component is available or most reliable. Audio may be noisy or absent, and text data may be absent altogether. Thus, a video-oriented approach can optimize performance and resource utilization in AR applications [2]. Thus, in the article "Video-Based Cross-Modal Auxiliary Network for Multimodal Sentiment Analysis" [13], the authors note that previous works are more focused on the study of effective joint representations, but they rarely consider the insufficient extraction of unimodal characteristics and the redundancy of multimodal fusion data. They offer a video-based cross-modal auxiliary network (VCAN) consisting of an audio characteristic map module and a cross-modal selection module. Extensive experimental results from the RAVDESS, CMU-MOSI, and CMU-MOSEI benchmarks show that VCAN significantly outperforms the state-of-the-art methods for improving the classification accuracy of multimodal sentiment analysis. The use of deep learning is becoming increasingly common in multimodal sentiment analysis as it has proven to be a powerful technique for solving this problem. Recently, numerous deep learning models and various algorithms have been proposed in the field of multimodal sentiment analysis. The article "SI: Multimodal Corpus of Sentiment Intensity and Subjectivity Analysis in Online Opinion Videos" [14] offers an overview covering the latest trends and developments in this area.

***2.3. Datasets comparisons*.** The effectiveness and efficiency of models are quite often evaluated on two widely used datasets in the field of multimodal sentiment analysis: CMU-MOSI [14] and CMU-MOSEI [15].

CMU-MOSI (Carnegie Mellon University Multimodal Opinion Sentiment and Emotion Intensity) is a dataset of 93 videos divided into 2199 utterances, each containing one speaker expressing an opinion from a range of topics The dataset contains annotations for sentiment, emotion, and intensity. Each statement in the set is labeled with one of seven moods, ranging from strongly positive to strongly negative, with corresponding numerical labels from +3 to -3. The CMU-MOSI suite is widely used to explore multimodal sentiment analysis techniques.

As defined by the authors, the key challenges of multimodal video sentiment analysis include:

- *Intramodal dynamics*. This problem concerns interaction within a particular modality, regardless of other modalities. For example, analyzing the changing nature of expressed opinions, where the correct structure of language is often ignored, makes sentiment analysis difficult.
- *Intermodal dynamics*. This problem involves interactions between different modalities, which fall into synchronous and asynchronous categories. Synchronous dynamics involves the simultaneous appearance of visual and textual cues, while asynchronous dynamics involves the delayed appearance of cues in different modalities.
- *Combination of characteristics*. Choosing the best approach to combine features from different modalities is a major challenge in multimodal sentiment analysis.

The CMU-MOSEI data set is different from the previous one. It contains more than 23,500 video fragments, covers a wider range of topics, contains annotations of six basic emotions with their intensities, and is one of the largest multimodal datasets for emotion analysis. This is a fairly large data set, well suited for complex analysis of a wide range of emotions.

It is also worth noting that a previously quite popular dataset for recognition was FER+ [16], which is an extension of the original FER (Facial Expression Recognition) dataset, which was improved by adding the quantity and quality of annotations. Although it contains 35887 reduced images, there are several reasons why it is not always suitable for modern applications, especially when it comes to the analysis of emotional feedback in videos: outdated data, low diversity, accuracy and reliability issues, focus on static recognition. Given our task - recognition of emotional feedback on video, taking into account temporal and spatial data, this dataset is not suitable for us.

***2.4. Research on multimodal video analysis*.** In the [14], the authors provide a comprehensive overview of recent trends and research directions in the field of multimodal sentiment analysis in video. The state-of-the-art models presented in this review use a multimodal combination of audio, visual, and textual information to analyze video sentiment in a variety of ways.

One common approach is to extract features from each modality separately and then combine them at different levels of the model. For example, some models use early fusion, where features from each modality are combined and fed into the model as a single input. Other models use late fusion, where the features of each modality are processed separately and then combined at the output layer. Another approach is to use attentional mechanisms to selectively focus on the most informative features of each modality. For example, some models use visual attention mechanisms to focus on specific regions of video frames, while others use audio and text attention to focus on specific segments of input audio and text information.

In addition, some models use multi-utterance fusion to capture the contextual relationship between adjacent utterances in a video. This leads to the identification of relevant and important information from the pool of utterances, making the model more reliable and accurate.

Overall, the state-of-the-art models presented in this review use different techniques to combine information from different modalities and capture the complex dynamics of multimodal sentiment analysis in video.

From this review, it becomes clear that modern developments in the field of multimodal analysis of moods in video are aimed at the simultaneous use of information from three modalities, namely visual, acoustic and textual components, but the specificity of the task of determining the preferences of users of virtual art compositions allows effective use of only one modality, namely the visual component [17, 18]. Therefore, we decided to consider the task of recognizing emotions and their intensity in videos as a regression task.

As a basis for the implementation of the video emotion recognition model, we chose the video classification model, namely the video action recognition presented in the article “A Closer Look at Spatiotemporal Convolutions for Action Recognition” [19]. In this paper, the authors discuss different forms of spatiotemporal convolutions for video analysis and their impact on action recognition. They also demonstrate the accuracy advantages of 3D convolutional neural networks over 2D convolutional neural networks and show how decomposing 3D convolutional filters into separate spatial and temporal components yields significant gains in accuracy. The authors' empirical study led to the development of a new spatiotemporal convolutional block that demonstrates high performance on multiple datasets.

Spatiotemporal convolutions are a type of convolutional operation used in video analysis that takes into account both spatial and temporal information. They differ from 2D convolutions, which consider only spatial information, and 3D convolutions, which consider both spatial and temporal information, but can be computationally expensive. Spatiotemporal convolutions can be implemented in a variety of ways, such as using 3D convolutions, 2D convolutions over frames, or a combination of 3D and 2D convolutions. The authors of this article investigate different forms of spatiotemporal convolutions and their impact on the performance of action recognition.

Decomposing 3D convolutional filters into separate spatial and temporal components improves the accuracy of action recognition for several reasons. First, it allows spatiotemporal characteristics to be modeled more efficiently, as the spatial and temporal aspects of the data can be captured independently and then combined. This can lead to a better representation of complex spatiotemporal patterns in video. Second, the optimization process becomes easier when the 3D convolution is decomposed, resulting in lower training error compared to traditional 3D convolutional networks of the same capacity. This indicates that filter decomposition can lead to better convergence during training, ultimately improving model accuracy.

In addition, decomposition can lead to more efficient use of network bandwidth, allowing for improved generalization and discrimination between different classes of actions. In general, splitting 3D convolutional filters into separate spatial and temporal components provides a more efficient and effective way to model spatiotemporal information, leading to increased accuracy in action recognition tasks.

The new spatiotemporal convolution unit "R(2+1)D" is designed to decompose 3D convolutions into separate spatial and temporal components, which helps to increase the accuracy of action recognition. This block consists of a 2D spatial convolution followed by a 1D temporal convolution, effectively decomposing the spatiotemporal information processing.

The authors demonstrate that the use of the R(2+1)D convolution in the ResNet architecture leads to high results on four different action recognition tests. In particular, the R(2+1)D model achieves results comparable to or better than the state-of-the-art on the Sports-1M, Kinetics, UCF101, and HMDB51 datasets. The authors' empirical study also demonstrates the effectiveness of the R(2+1)D block compared to other spatiotemporal convolutional methods such as 3D convolutions and pseudo-3D blocks. This indicates that the R(2+1)D convolutional block is effective in capturing spatiotemporal features and outperforms existing approaches for various action recognition tasks.

The design of the R(2+1)D block and its performance on different datasets highlight its effectiveness in modeling spatiotemporal information and its potential to improve the accuracy of action recognition in a variety of video datasets. That is why this model was chosen by us as the basis for the development of a subsystem for recognizing user emotions and their intensity in videos for a hybrid recommender system of virtual art compositions.

2. Description of the experiment with the VSA model

Taking into account that the task of recognizing emotions in videos and their intensity is a regression problem [20, 21], we made appropriate modifications to the model from [19, 22] so that it is able to recognize user emotions in videos (negative, neutral, positive) and mark videos with appropriate digital labels from - 3 to +3. Thus, we obtained a new model that performs spatiotemporal analysis of the video stream [23, 24] and described it in detail in blocks in the article Regression neural model for video sentiments analysis [25]. The general scheme of the modified CNN model is presented in Fig. 1.

2.1. VSA Experiment results. We chose the CMU-MOSI dataset to evaluate the performance and effectiveness of the user emotion recognition model for several reasons:

- *Size and diversity*: This dataset is large enough and covers a diverse range of videos and speakers, providing a rich and diverse source of visual data for analysis. This diversity allows the model to be reliably evaluated in different contexts.
- *Annotation of moods*: the data set contains annotations for a wide range of moods, which allows a comprehensive assessment of the model's ability to capture and classify the nuances of emotional expressions.
- *Widespread use*: CMU-MOSI have become the de facto standard benchmarks in the field of multimodal sentiment analysis, and many researchers use these datasets to evaluate the performance of their models. This widespread adoption provides comparison with state-of-the-art models.
- *Real-world relevance*: The videos in this dataset are sourced from online platforms such as YouTube and reflect real-world scenarios and natural expressions of sentiment. This real-world relevance improves the practical applicability of models trained and evaluated on this dataset.

To conduct the experiment, we divided the data set into two subsets: training and test, respectively, in the ratio of 80% to 20%. 20% of the training set was used for model validation during training.

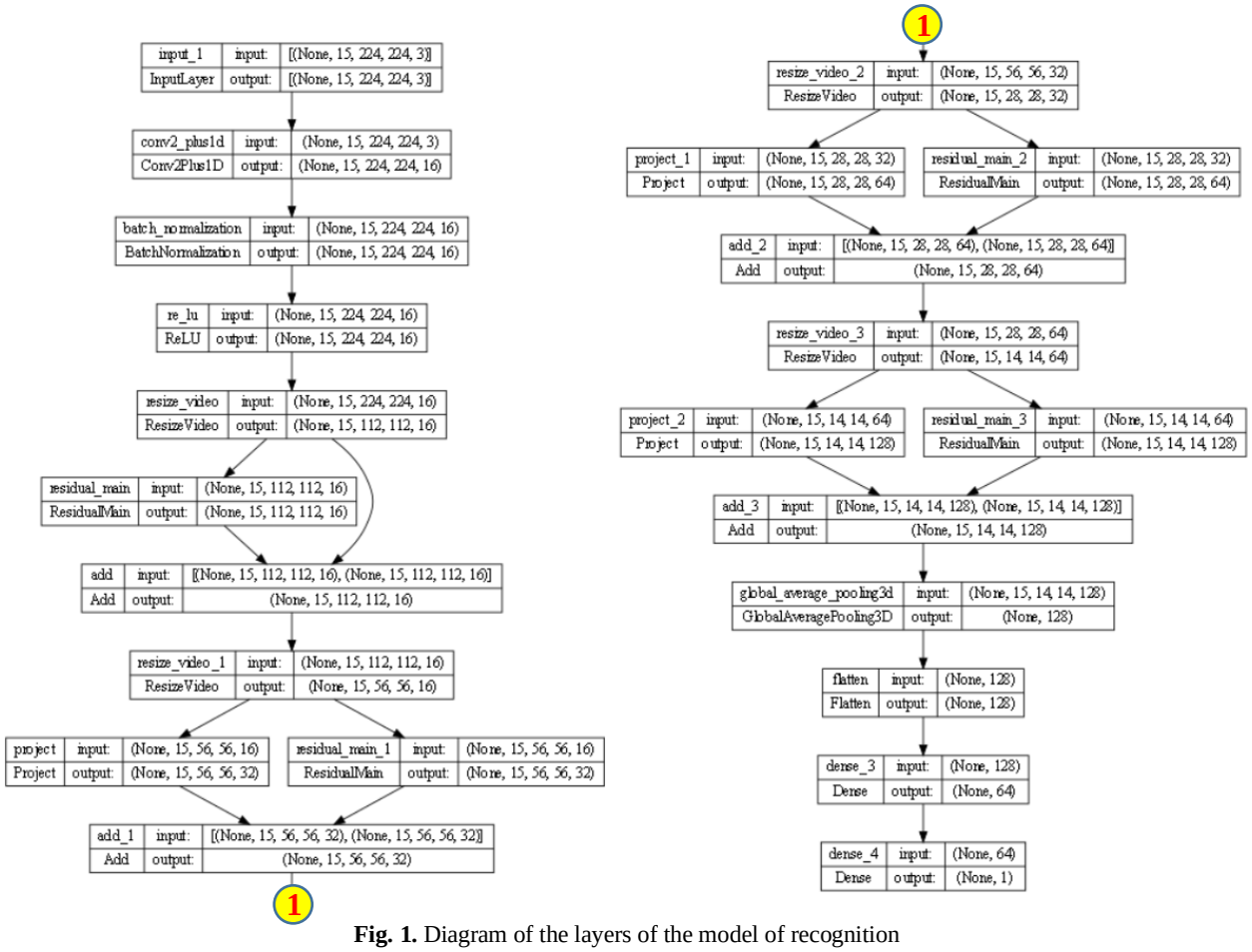

Fig. 1. Diagram of the layers of the model of recognition of emotions and their intensity in video (VSA), which was presented and described in detail in the article [6]

Fig. 2 shows graphs of changes in the values of the loss function that were obtained in the article [6], as well as metrics obtained during training of the model on training data and its verification on validation data.

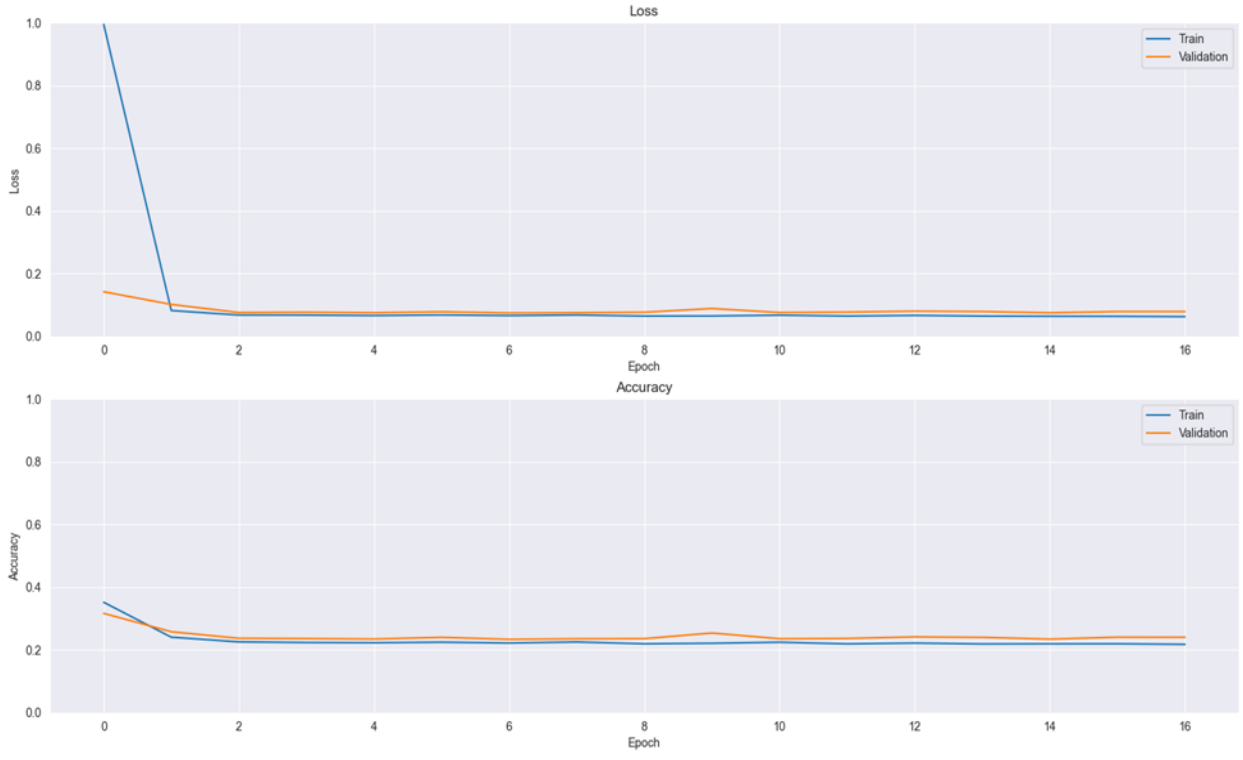

Fig. 2. The graph of changes in the values of the loss function and the metric used to evaluate the VSA model, which was presented in the results of the article [6]

The mean squared error was used as a loss function to evaluate the effectiveness of the learning process. The mean absolute error was used as a metric to evaluate the model during training and testing.

3.2. HR-VSA experiment description. The next step was to add the VSA model as a subsystem to the hybrid HR recommendation system in order to take into account the emotions of users to obtain more accurate recommendations for virtual art compositions. The general scheme of the created model is presented in Fig. 3.

To analyze the scheme of the model in more detail, we will consider its main parts. We added two separate deep neural networks to the Neural Collaborative Filtering model, which use information about the AR session properties of the virtual art composition and user profile information for video selection [1], Fig. 4.

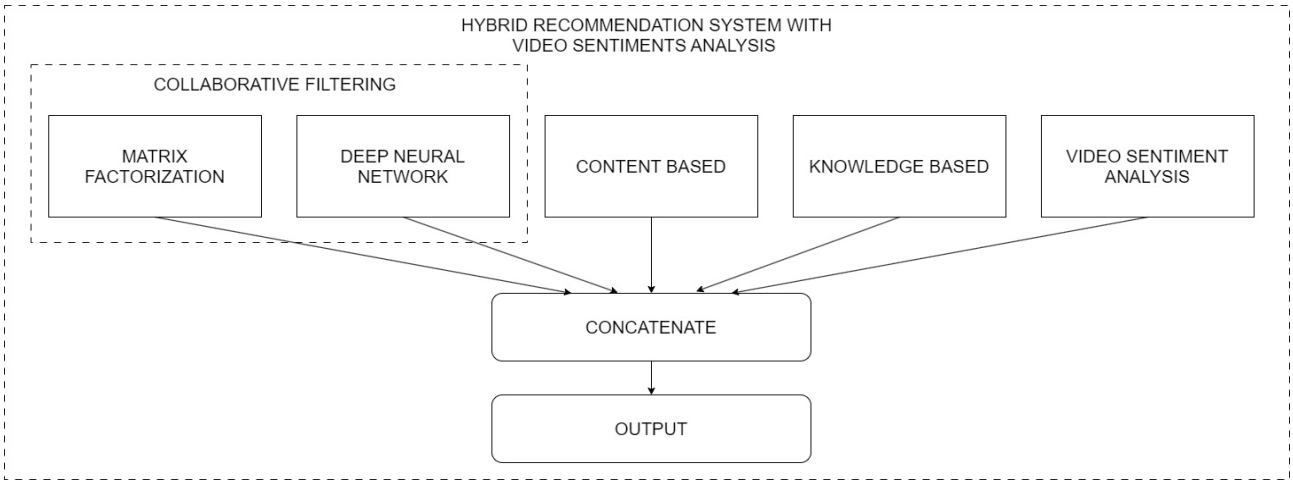

Fig. 3. General scheme of layers of the HR-VSA model

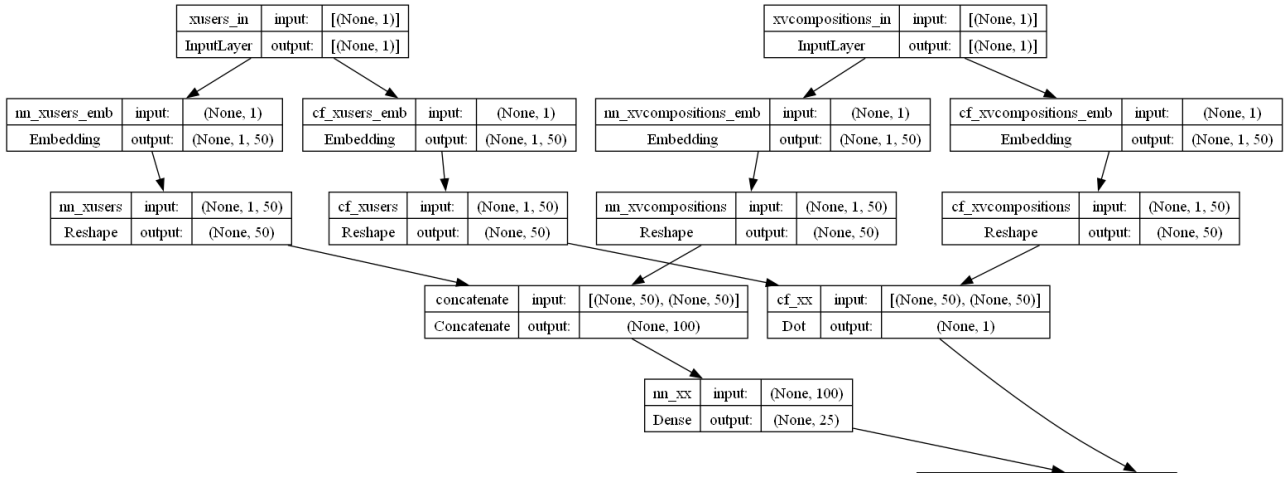

Fig. 4. Scheme of the layers of the collaborative filtering method based on matrix factorization and a deep neural network

The content-based method analyzes the characteristics of virtual art compositions to calculate their proximity to the user's interests. And the knowledge-based method uses contextual user data to refine recommendations. The layers of these methods are presented in Fig. 5. The system is complemented by the added Video Sentiments Analysis, which is presented in Fig. 1 and is described in detail in the article Regression neural model for video sentiments analysis [25].

Finally, all method layers are combined in the final Concatenate layer.

The final prediction is obtained after the Dense layer, which predicts the user's interest. The general scheme of the system is presented in Fig. 6.

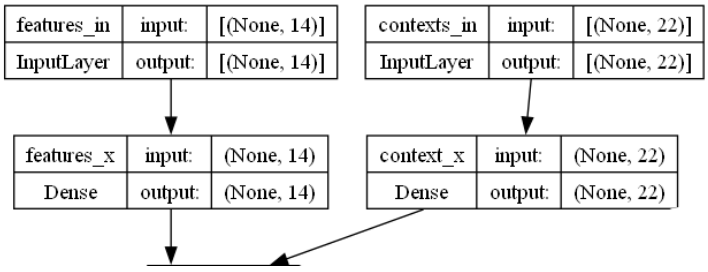

Fig. 5. Content-based and knowledge-based methods

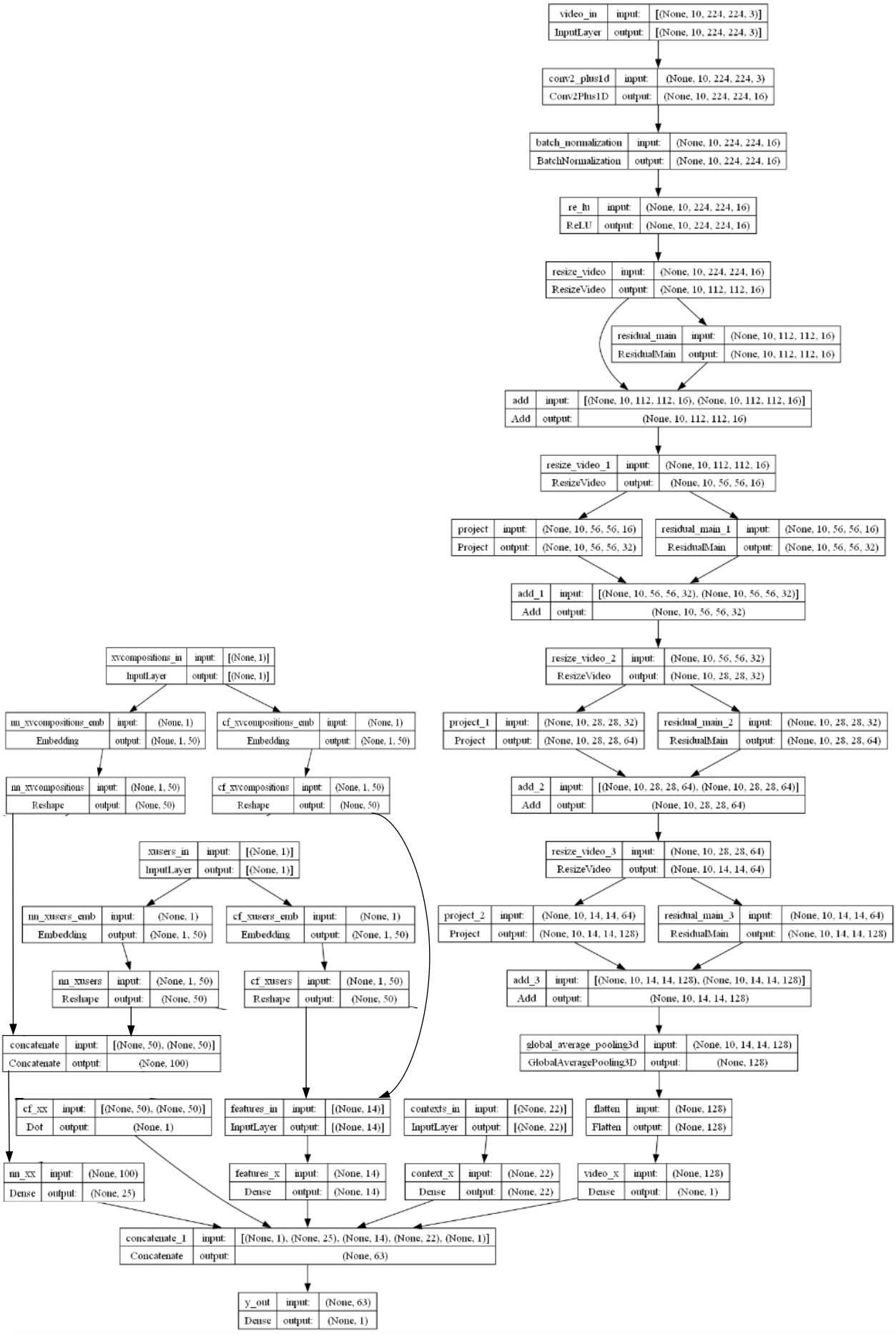

Fig. 6. Detailed diagram of the layers of the HR-VSA model

3.3. A dataset for model training. The data set obtained by consolidating the training set from [1] with the CMU-MOSI set, which was used to train the VSA model, was used for the computational experiment.

3.4. HR-VSA experiment results. Below are the results of a computational experiment conducted on the HR-VSA model using a consolidated data set.

For research, the data set was divided into two sets: training and test in a ratio of 80 to 20, respectively. 20% of the training set was used to validate the model in the training process.

Fig. 7 shows graphs of changes in the values of the loss function and metrics obtained during model training on training data and verification on validation data.

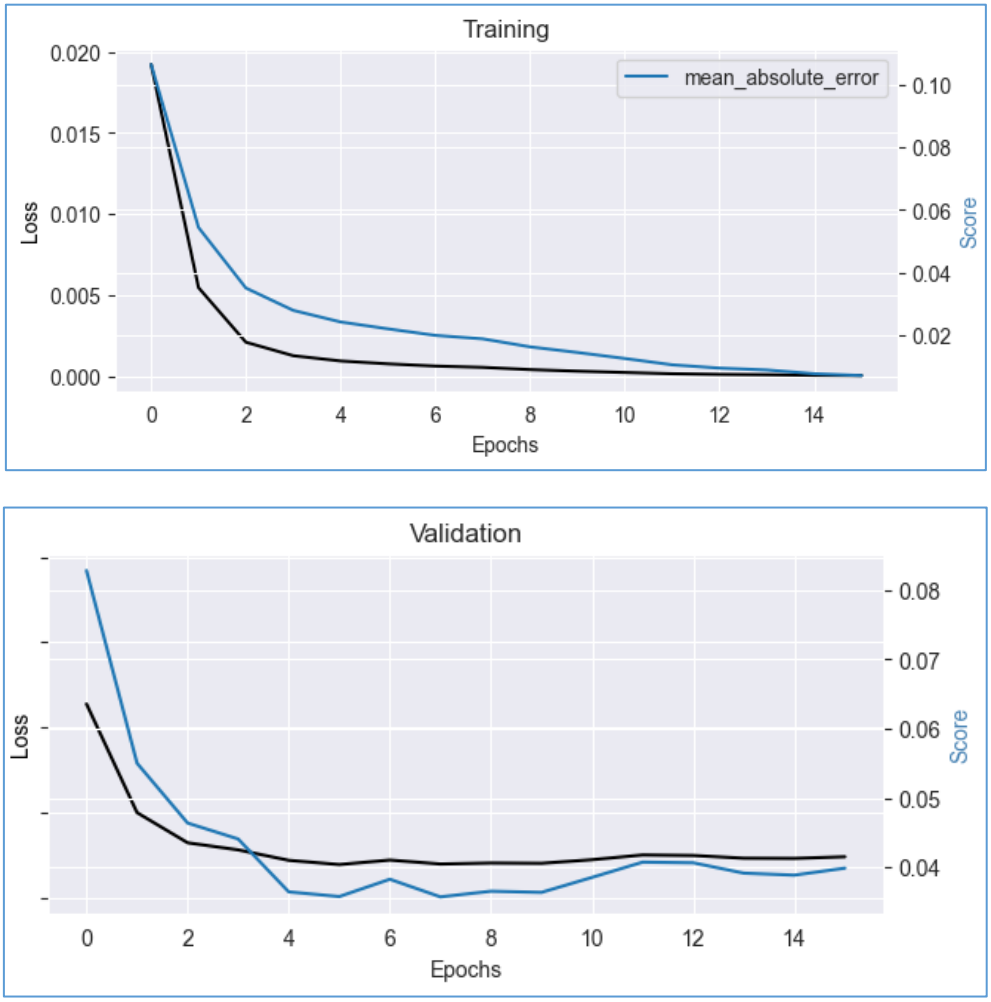

Fig. 7. Graph of changes in metric and loss function values used to evaluate the HR-VSA model

To evaluate the quality of the model training process, the mean squared error (mean squared error) was used as a loss function, and the mean absolute error (mean absolute error) was used as a metric for evaluating the model during training and testing.

During the testing of the HR-VSA model, the value of Mean Absolute Error (Σ|y-pred|/n) was obtained: 0.0360.

During the testing of the hybrid model of the HR recommendation system without the VSA subsystem, which was carried out in [1], the Mean Absolute Error value (Σ|y-pred|/n): 0.0922 was obtained.

As can be seen from the results of the experiments, the use of the video emotion recognition subsystem in the hybrid recommender system of virtual art compositions made it possible to achieve a significant reduction in the error of the system by 2.56 times.

From the results of testing the new model, it is clear that the inclusion of the VSA subsystem in the hybrid recommendation system for virtual art compositions has led to the creation of a more sensitive and accurate model.

This indicates that the system now responds much better to emotional feedback in the context of user interaction with virtual art compositions and can potentially improve the user experience and immersiveness of the system.

Conclusion and Future Work

Recent studies confirm the growing trend to implement emotional feedback and sentiment analysis to improve the performance of recommender systems [26, 27]. In this way, a deeper personalization and current emotional relevance of the user experience is ensured.

In this article, we aimed to compare the effectiveness of the recommender system of virtual art compositions with and without the emotion recognition subsystem on video.

A hybrid model from [1] was chosen for the research, to which was added a subsystem for recognizing emotions on video, as which the model from [19] was used. As the results of the computational experiment showed, thanks to the addition of the VSA subsystem, it was possible to significantly reduce the average absolute error of the system by 2.56 times.

This proves that by taking into account the additional information about user preferences, which can be obtained by analyzing the videos of users while they are viewing virtual artworks, it is possible to significantly reduce errors and improve the performance of virtual artwork recommender systems.

Future research may include the prospect of adding a subsystem with a generative neural network that, based on the findings of the developed recommender system, will create new virtual art compositions [28, 29].

Thus, the system can not only select the most suitable virtual art compositions, but also create adaptive and dynamic content, which will increase user satisfaction and improve the immersive aspects of the system.

Acknowledgements

The study was funded by the National Research Foundation of Ukraine in the framework of the research project 2020.02/0404 on the topic "Development of intelligent technologies for assessing the epidemic situation to support decision-making within the population biosafety management".

REFERENCES

1. Kuliahin, A., Narozhnyi, V., Tkachov, V. and 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*, No. 4(70), PNTU, Poltava, pp. 94–99, doi: https://doi.org/10.26906/SUNZ.2022.4.094
2. Kuliahin, A. (2023), "Using recognized emotion as implicit feedback for a recommender system", *Control, Navigation and Communication Systems*, No. 3(73), PNTU, Poltava, pp. 115–119, doi: https://doi.org/10.26906/SUNZ.2023.3.115
3. Deldjoo, Ya., Schedl, M., Hidasi, B., Wei, Yi. and He, X. (2022), "Multimedia Recommender Systems: Algorithms and Challenges", *Recommender Systems Handbook*, Springer, New York, doi: https://doi.org/10.1007/978-1-0716-2197-4_25
4. Tsai, Ch. and Brusilovsky, P. (2021), "The effects of controllability and explainability in a social recommender system", *User Modeling and User-Adapted Interaction*, Vol. 31, pp. 591–627, doi: https://doi.org/10.1007/s11257-020-09281-5
5. Messina, P., Dominguez, V., Parra, D., Trattner, Ch. and Soto, A. (2019), "Content-based artwork recommendation: integrating painting metadata with neural and manually-engineered visual features", *User Modeling and User-Adapted Interaction*, vol. 29, pp. 251–290, doi: https://doi.org/10.1007/s11257-018-9206-9
6. Fernández del Amo Blanco, Iñigo & Erkoyuncu, John & Farsi, M. and Ariansyah, D. (2021), "Hybrid recommendations and dynamic authoring for AR knowledge capture and re-use in diagnosis applications", *Knowledge-Based Systems*, Vol. 239, 107954, doi: https://doi.org/10.1016/j.knosys.2021.107954
7. Liu, J. and Lv, H. (2022), "Recommendation of Micro Teaching Video Resources Based on Topic Mining and Sentiment Analysis", *Int. Journal of Emerging Techn. in Learning* (IJET), vol. 17, pp. 243–256, doi: https://doi.org/10.3991/ijet.v17i06.30011
8. Qian, Y., Zhang, Y., Ma, X., Yu, H. and Peng, L. (2018), "EARS: Emotion-Aware Recommender System Based on Hybrid Information Fusion", *Information Fusion*, vol. 46, pp. 141–146, doi: https://doi.org/10.1016/j.inffus.2018.06.004
9. Pessemier, T., Coppens, I. and Martens, L. (2020), "Evaluating facial recognition services as interaction technique for recommender systems", *Multimedia Tools and Appl.*, Vol. 79, pp. 47–70, doi: https://doi.org/10.1007/s11042-020-09061-8
10. Berkovsky, S. (2015), "Emotion-Based Movie Recommendations", *EMPIRE '15: Proceedings of the 3rd Workshop on Emotions and Personality in Personalized Systems 2015*, doi: https://doi.org/10.1145/2809643.2815362
11. Li, J., Li, Z., Ma, X., Zhao, Q., Zhang, Ch. and Yu, G. (2023), "Sentiment Analysis on Online Videos by Time-Sync Comments", *Entropy*, vol. 25 (7), 1016, doi: https://doi.org/10.3390/e25071016
12. Chalkias, I., Tzafilkou, K., Karapiperis, D. and Tjortjis, C. (2023), "Learning Analytics on YouTube Educational Videos: Exploring Sentiment Analysis Methods and Topic Clustering", *Electronics*, vol. 12, 3949, doi: https://doi.org/10.3390/electronics12183949
13. Chen, R., Zhou, W., Li, Y. and Zhou, H. (2022), "Video-Based Cross-Modal Auxiliary Network for Multimodal Sentiment Analysis", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, is. 12, pp. 8703–8716, doi: https://doi.org/10.1109/TCSVT.2022.3197420
14. Zadeh, A., Zellers, R, Pincus, E. and Morency, L-P. (2016), "MOSI: Multimodal Corpus of Sentiment Intensity and Subjectivity Analysis in Online Opinion Videos", *Computation and Language*, arXiv:1606.06259, doi: https://doi.org/10.48550/arXiv.1606.06259
15. Zadeh, A.A. B., Liang P.P., Poria, S. Cambria, E. and Morency L.-P. (2018), "Multimodal Language Analysis in the Wild: CMU-MOSEI Dataset and Interpretable Dynamic Fusion Graph", ACL, *Association for Computational Linguistics*, Melbourne, Australia, pp. 2236–2246, doi: https://doi.org/10.18653/v1/P18-1208
16. Barsoum, E., Zhang, Ch., C. F., Cristian, and Zhang, Z. (2016), "Training Deep Networks for Facial Expression Recognition with Crowd-Sourced Label Distribution", *ACM International Conference on Multimodal Interaction* (ICMI), pp. 279–283, doi: https://doi.org/10.1145/2993148.2993165
17. Pu, H., Sun, Y., Song, R., Chen, X., Jiang, H., Liu, Y. and Cao, Z.(2024), "Going Beyond Closed Sets: A Multimodal Perspective for Video Emotion Analysis", *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 14430 LNCS, pp. 233–244, doi: https://doi.org/10.1007/978-981-99-8537-1_19
18. Yaloveha, V., Hlavcheva, D., Podorozhniak, A. and Kuchuk, H. (2019), "Fire hazard research of forest areas based on the use of convolutional and capsule neural networks", *2019 IEEE 2nd Ukraine Conference on Electrical and Computer Engineering*, UKRCON 2019 – Proceedings, doi: http://dx.doi.org/10.1109/UKRCON.2019.8879867
19. Tran, Du, Wang, Heng, Torresani, Lorenzo, Ray, Jamie, Le Cun, Yann and Paluri, Manohar (2017), "A Closer Look at Spatiotemporal Convolutions for Action Recognition", *Computer Vision and Pattern Recognition*, arXiv:1711.11248, doi: https://doi.org/10.48550/arXiv.1711.11248

20. Kovalenko, A. and Kuchuk, H. (2022), "Methods to Manage Data in Self-healing Systems", *Studies in Systems, Decision and Control*, Vol. 425, pp. 113–171, doi: https://doi.org/10.1007/978-3-030-96546-4_3
21. Dotsenko, N., Chumachenko, I., Galkin, A., Kuchuk, H. and Chumachenko, D. (2023), "Modeling the Transformation of Configuration Management Processes in a Multi-Project Environment", *Sustainability (Switzerland)*, Vol. 15(19), 14308, doi: https://doi.org/10.3390/su151914308
22. Ezzameli, K. and Mahersia, H. (2023), "Emotion recognition from unimodal to multimodal analysis: A review", Information Fusion, vol. 99, 101847, doi: https://doi.org/10.1016/j.inffus.2023.101847
23. Dun B., Zakovorotnyi, O. and Kuchuk, N. (2023), "Generating currency exchange rate data based on Quant-Gan model", *Advanced Information Systems*, Vol. 7, no. 2, pp. 68–74, doi: http://dx.doi.org/10.20998/2522-9052.2023.2.10
24. Yaloveha, V., Podorozhniak, A. and Kuchuk, H. (2022), "Convolutional neural network hyperparameter optimization applied to land cover classification", *Radioelectronic and Computer Systems*, No. 1(2022), pp. 115–128, DOI: https://doi.org/10.32620/reks.2022.1.09
25. Kuliahin, A. (2023), "Regression neural model for video sentiments analysis", *Global science: prospects and innovations*, Proceedings of the 5th International scientific and practical conference, Cognum Publishing House, Liverpool, United Kingdom, pp. 173–178, available at: https://sci-conf.com.ua/v-mizhnarodna-naukovo-praktichna-konferentsiya-global-science-prospects-and-innovations-28-30-12-2023-liverpul-velikobritaniya-arhiv/
26. Gomes, C.J., Gil-González, A.B., Luis-Reboredo, A., Sánchez-Moreno, D. and Moreno-García, M.N. (2022), "Song Recommender System Based on Emotional Aspects and Social Relations", *Lecture Notes in Networks and Systems*, vol. 327 LNNS, pp. 88–97, doi: https://doi.org/10.1007/978-3-030-86261-9_9
27. Xiong, W. and Zhang, Y. (2023), "An intelligent film recommender system based on emotional analysis", *PeerJ Computer Science*, vol. 9, pp. 1–15, doi: https://doi.org/10.7717/PEERJ-CS.1243
28. Zhou, S., Jia, J., Wang, Q., Dong, Y., Yin, Y. and Lei, K. (2018), "Inferring emotion from conversational voice data: A semi-supervised multi-path generative neural network approach", *32nd AAAI Conference on Artificial Intelligence*, AAAI 2018, pp. 579–586, doi: DOI: https://doi.org/10.1609/aaai.v32i1.11280
29. Kujani, T. and Kumar, V.D. (2022), "Emotion Understanding from Facial Expressions using Stacked Generative Adversarial Network (GAN) and Deep Convolution Neural Network (DCNN)", International Journal of Engineering Trends and Technology, 70(10), pp. 98–110, doi: https://doi.org/10.14445/22315381/IJETT-V70I10P212

Received (Надійшла) 20.10.2023
Accepted for publication (Прийнята до друку) 17.01.2024

ВІДОМОСТІ ПРО АВТОРІВ/ ABOUT THE AUTHORS

Кучук Георгій Анатолійович – доктор технічних наук, професор, професор кафедри комп'ютерної інженерії та програмування, Національний технічний університет "Харківський політехнічний інститут", Харків, Україна;
Heorhii Kuchuk – Doctor of Technical Sciences, Professor, Professor of Computer Engineering and Programming Department, National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine;
e-mail: kuchuk56@ukr.net; ORCID ID: http://orcid.org/0000-0002-2862-438X;
Scopus ID: https://www.scopus.com/authid/detail.uri?authorId=57057781300.

Кулягін Андрій Ігорович – аспірант, кафедра комп'ютерних систем, мереж та кібербезпеки, Національний аерокосмічний університет «Харківський авіаційний інститут», Харків, Україна;
Andrii Kuliahin – PhD Student of Department of Computer Systems, Networks and Cybersecurity, National Aerospace University "Kharkiv Aviation Institute", Kharkiv, Ukraine;
e-mail: kulagin.andrew38@gmail.com; ORCID ID: http://orcid.org/0000-0002-7465-351X;
Scopus ID: https://www.scopus.com/authid/detail.uri?authorId=58759297900.

Гібридна рекомендаційна система для віртуальних арт-композицій з аналізом настроїв на відео

Г. А. Кучук, А. І. Кулягін

Анотація. Актуальність. Останні дослідження підтверджують зростаючу тенденцію до впровадження емоційного фідбеку та аналізу настроїв для покращення результатів рекомендаційних систем. Таким чином забезпечується більш глибока індивідуалізація та поточна емоційна відповідність користувацького досвіду. **Предметом вивчення** в статті є гібридна рекомендаційна система з компонентом аналізу сентиментів на відео. **Метою статті** є дослідження можливостей покращення ефективності результатів гібридної рекомендаційної системи віртуальних арт-композицій шляхом впровадження компонента аналізу сентиментів на відео. **Використані методи**: метод матричної факторизації, метод колаборативної фільтрації, метод на основі контенту, метод на основі знань, метод аналізу сентиментів на відео. Отримано **наступні результати**. Створено нову модель, що поєднує гібридну рекомендаційну систему та компонент аналізу сентиментів на відео. Суттєво знижено середню абсолютну помилку системи. Додано реакцію системи на емоційний фідбек в контексті взаємодії користувача з віртуальними арт-композиціями. **Висновок.** Таким чином, система може не лише підбирати найбільш підходящі віртуальні арт-композиції, але й створювати адаптивний та динамічний контент, що дозволить підвищити задоволеність користувачів та покращити аспекти імерсивності системи. Перспективним **напрямком подальших досліджень** може бути додавання підсистеми з генеративною нейронною мережею, яка на основі висновків розробленої рекомендаційної системи буде створювати нові віртуальні арт-композиції.

Ключові слова: гібридна рекомендаційна система, колаборативна фільтрація, матрична факторизація, згорткова нейронна мережа, емоційний фідбек, аналіз сентиментів на відео.