Journal of Innovations in Computer Science and Engineering
Shahid Beheshti Unversity

January 2024, Volume 1, Issue 2

Improving Social Image Recommendations through Emotion Analysis

Somaye Ahmadkhani a, ✉, **ORCID:**
Mohsen Ebrahimi Moghaddam b, **ORCID:** : 0000-0002-7391-508X
a Shahid Beheshti University, Faculty of Computer Science and Engineering, Tehran, Iran, s_ahmadkhani@sbu.ac.ir
b Shahid Beheshti University, Faculty of Computer Science and Engineering, Tehran, Iran, m_moghadam@sbu.ac.ir

Abstract

With the rapid expansion of the Internet and social networks, the volume of images available online has increased dramatically, making it challenging for users to find relevant content. To address this problem, we propose incorporating emotion analysis as a key factor in understanding user preferences, thereby creating a more personalized and effective image recommendation system. In this article, we examine two approaches to utilizing emotional features in image recommendation. The first approach integrates emotional features directly into the feature vector used for training the recommendation model. The second approach refines recommendations through emotion-based postprocessing, where emotional proximity between users and images is used to re-rank recommendations. This study emphasizes the value of emotion analysis in advancing the personalization and efficacy of social image recommendation systems. Experimental results indicate that both approaches significantly improve recommendation performance, achieving higher metrics such as Recall@k and Precision@k. These findings demonstrate that emotional analysis enhances personalization and effectiveness in social image recommendation systems.

KEYWORDS

Recommender system, emotion analysis, deep learning, social network.

1. INTRODUCTION

In recent times, with the vast increase in image content on social networks, social image recommender systems have become increasingly vital. Researchers have focused on enhancing these systems, exploring various aspects to better understand user preferences [1-3]. Lovato et al. [4] examined users' liked images to delve into their personal aesthetic interests, positing that everyone possesses a unique aesthetic evaluation framework. By analyzing high-level and low-level features from users' preferred images, they identified distinctive characteristics for each user through regression. In some research, the problem of image recommendation has been studied from the perspective of user modeling. For this purpose, use the set of images that have been tagged as favorites by the user. Guntuku et al. [5], added high-level comprehensible features from favored images to those in [4], aiming to improve personality recognition and thus generate more accurate recommendations.

The application of deep learning in recommender systems has recently expanded considerably. A comprehensive review in [6], explored recommender systems utilizing deep learning across multiple media types, such as text, image, audio, and video. Lei et al. [7] proposed a hybrid representation that combines deep learning with user preferences. Their approach utilizes a dual deep network, comprising two subnets that map input images and user preferences into a shared hidden semantic space. In this space, the distance between images and user preferences is computed to facilitate decision-making. In this article, we proposed emotional analysis of images tagged as favorites by the user to more effectively understand user preferences and generate tailored recommendations.

In sentiment analysis, there are three primary areas: text-based, image-based, and audio-based sentiment analysis. Most sentiment analysis systems focus on text. In the text domain, sentiment analysis—often referred to as opinion mining or opinion extraction—is widely used in recommender systems. In [8], Zhang et al. first extract the explicit features of the products (i.e., aspects) and user opinions through phrase-level sentiment analysis on user reviews and emotional analysis of expression levels in user comments. Recommendations are then generated based on the specific features of the product and user interests. Also, in [9], research in this area has been furthered. Beyond text-based sentiment analysis, the surge of images on social media has increasingly directed attention towards sentiment analysis based on social images. Sentiment analysis of social media images encompasses approaches based on

Submit Date: 2024-11-10
Revise Date: 2025-01-06
Accept Date: 2025-03-12
✉ Corresponding author

low-level visual features [10, 11], mid-level visual features [12], and Deep learning [13, 14]. Additionally, some methods utilize only image features, while others combine image and text features. One prominent feature-based approach by Katsurai et al.[15], utilizes visual, textual, and sentiment features from training images. After applying dimensionality reduction to optimize these feature sets, they are mapped to a target space, where classification is then performed. Another method, known as Sentribute, was proposed by Yuan et al. [16]. This method begins by extracting low-level image features, such as gist descriptors, gradient histograms, and color histograms. These features are then used to categorize the image within one of 800 categories from the SUN dataset, a major resource for scene and object recognition. Following this, classification is performed using the SVM method. The image's class label, along with its low-level features, is subsequently provided as input to another SVM classifier to determine the image's sentiment class. Another method that only utilizes low-level image features for image sentiment detection is presented by Siersdorferet al. [17]. This approach uses features such as 1) Histogram Color Global: GCH, 2) Histogram Color Local: LCH, and 3) SIFT descriptor. Methods based solely on low-level image features face limitations in detecting emotions across large-scale images. To address this issue, Borth et al. [12] developed a large-scale sentiment ontology. The key innovation in their work is the introduction of a middle-level feature, Adjective Noun Pairs (ANP), which exhibits a strong association with sentiment. Yuan et al. [16] introduced a method for predicting image sentiment using mid-level attributes. Cao et al. [18] developed a visual sentiment topic model, which leverages Visual Sentiment Ontology (VSO) to extract visual sentiment features. This model constructs a Visual Sentiment Topic Model by grouping all images within the same topic and then selects the most suitable visual sentiment features based on the distribution of visual sentiment features within that topic [19]. Yu et al. [13] implemented a deep learning approach, achieving high accuracy and performance across various criteria. However, this method is limited by its slow processing speed due to the complexity of multiple layers and the depth of the neural network, necessitating highly powerful hardware. Additionally, [20] introduced a novel Bi-Directional Multi-Level Attention (BDMLA) model to leverage complementary information from image and text modalities for joint visual-textual sentiment classification. This model incorporates a visual attention network and a semantic attention network to emphasize emotional regions and words within image-text pairs.

The structure of the article is as follows: In Section 2, we discuss the proposed approaches for incorporating emotional analysis into image recommendation, focusing on two methods: Emotion-Integrated Feature Vector and Emotion-Based Postprocessing. Section 3 presents the implementation results and evaluates the performance of the proposed methods. Finally, the article concludes with a summary of findings and future directions in Section 4.

2. PROPOSED METHOD

In this section, we describe the methods used for image recommendation based on visual emotion analysis. For emotion analysis, we adopt the method proposed in [21]. The emotion analysis method from [21] uses the EfficientNetB1 model, fine-tuned for emotion recognition, which has been shown to achieve superior results in detecting emotions in social images. By leveraging this method, we extract emotion vectors for all images, which are then integrated into our recommendation framework.

Next, we propose incorporating emotion analysis into the recommendation system in two distinct ways, as illustrated in Figure 1 and Figure 2. These approaches are designed to analyze the effect of emotional features on the performance of the image recommender system. Below, we describe the workflow and components of these two approaches in detail.

A. IMAGE RECOMMENDATION BASED ON VISUAL EMOTION ANALYSIS

In this section, we examine two methods for incorporating emotion analysis into the image recommendation process, as illustrated in Figure 1 and Figure 2. Both methods aim to evaluate the impact of emotional features on the performance of the recommender system. While we provide a brief overview of the method shown in Figure 1, the primary focus of this section is on the method outlined in Figure 2, which we describe in greater detail.

In the first approach, shown in **Figure 1**, referred to as the **Emotion-Integrated Feature Vector**, each image's feature vector is constructed by combining **visual features** (see **Table I**) with **emotional features** (see **Table II).** This composite feature vector is then used as input for training the recommendation model. The visual features in Table I are based on those used in [4], while the emotional features are detailed in Table II. After extracting these features, we apply two learning methods, Lasso [22] and FGM [23, 24], to predict which images are liked by each user and recommend those to him/her.

Lasso is a widely-used regression analysis method that incorporates feature selection and regularization to improve predictive accuracy. Similarly, FGM is a feature selection strategy focused on identifying the most influential features to explain variations in user preferences. Sparse SVM formulation, as described in Eq. (1), is employed to select these informative features. In the FGM method, d denotes the indices of the selected features, while w indicates their associated weights [23, 24].

$$\min_{\mathbf{d}} \min_{\mathrm{w},\varepsilon,\rho} \frac{1}{2}\|\mathbf{w}\|_2^2 + \frac{C}{2}\sum_{i=1}^{n}\varepsilon_i^{\,2} - \rho$$
$$s.t.\ \ y_i \mathbf{w}^{T}(\mathbf{x}_i \bullet \mathbf{d}) \geq \rho - \varepsilon_i \tag{1}$$

Where y_i indicates a binary value, if image *i* is liked by the user y_i value equals +1, otherwise it is -1. The variable **x** denotes the feature vector, and **w** indicates the weight vector, which shows the importance of the features. $\mathbf{d} \in D = \{\|\mathbf{d}\|_0 \leq B, \mathbf{d}_j \in \{0,1\}\}$

denotes whether the feature is selected or not. ρ is the bias, ε_i denotes the i^{th} instances loss incurred by classifiers, B is the number of features to be selected, and C indicates the trade-off parameter.

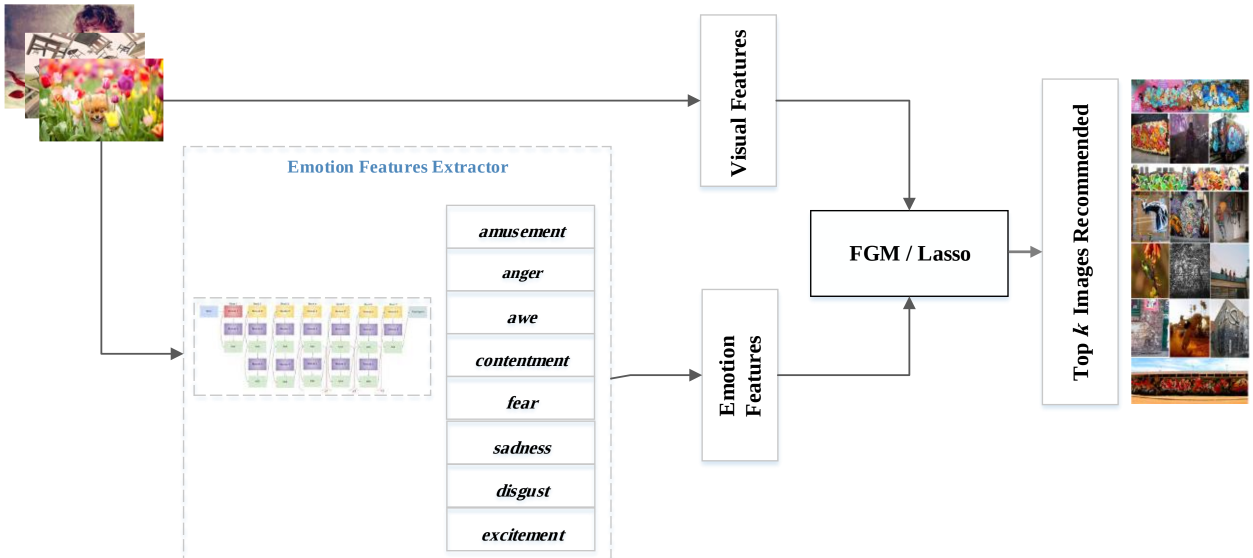


Figure 1. The first approach for Image Recommendation based on visual emotion analysis

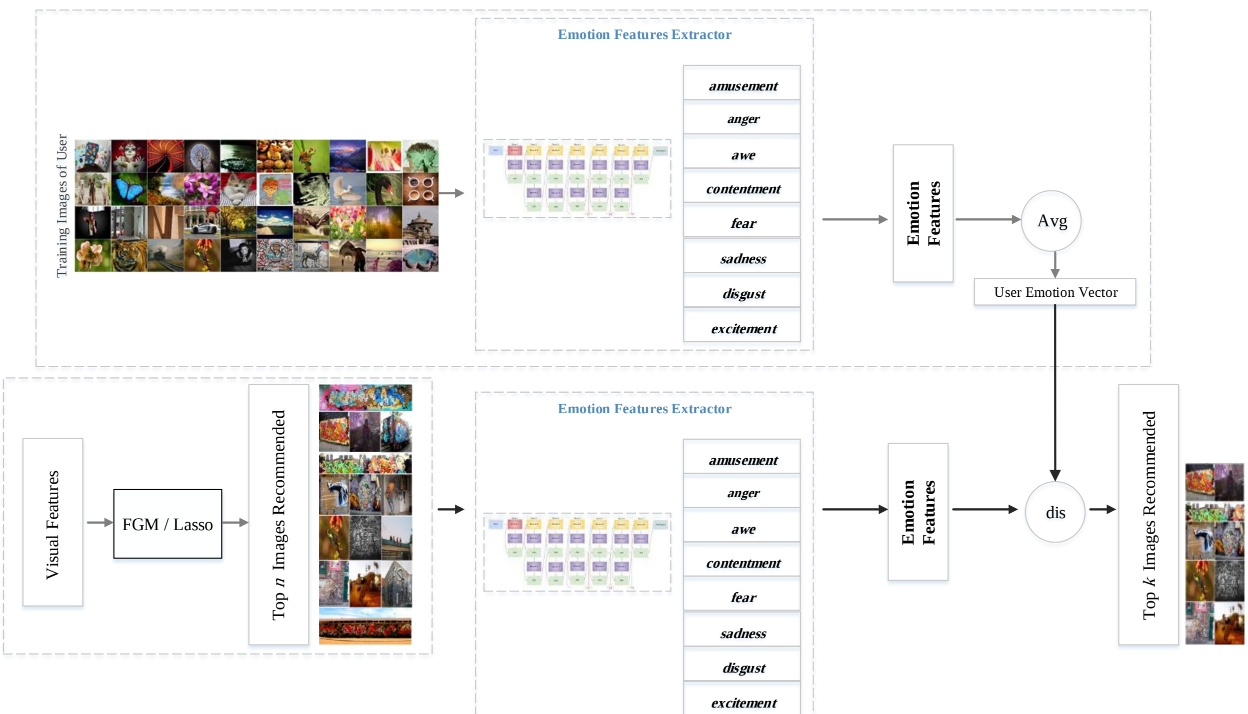


Figure 2. The second approach for Image Recommendation based on visual emotion analysis

Table 1. List of visual features

Name	Feature Dimensions
Use of light	1
HSV statistics	3
Emotion based	3
Hue Circular Variance	1
Colorfulness	1
Color name	11
Entropy	1
Wavelet textures	12
Tamura	3
GCLM features	12
Edges	1
Level of detail	1
Regions	1
Low depth of field (DOF)	3
Faces	1
GIST descriptors	24
Total	**79**

In the second approach, referred to as **Emotion-Based Postprocessing** (Figure 2), we emphasize the role of emotional proximity in refining recommendations. First, a user's emotion vector is computed by averaging the emotional features of their training images. Then, for each image recommended by the system, the distance between the image's emotion vector and the user's emotion vector is calculated. The system re-ranks the recommendations based on this distance, prioritizing images that are emotionally closer to the user. This method allows for a more personalized and emotion-centric refinement of the recommendation process.
The method shown in Figure 2 comprises three main steps:

1. User Emotion Vector Calculation: For each user, the system calculates an emotion vector by averaging the emotional features of the images they have liked in the training set. This vector represents the user's overall emotional profile and serves as the baseline for comparing new images.
2. Initial Recommendation: The initial set of n candidate images is generated using traditional visual feature-based methods (e.g., LASSO or FGM [23, 24]). These images are ranked based on their visual similarity to the user's preferences, as determined by the training data.
3. Emotion-Based Refinement: For each of the n candidate images, the system calculates the distance between the image's emotion vector and the user's emotion vector. This distance acts as a measure of emotional alignment between the user and the image. The system then re-ranks the images based on emotional proximity, selecting the top k images (where $k < n$) that are closest to the user's emotion vector. These top k images are presented as the final recommendations.

By focusing on emotional proximity in the postprocessing phase, this method enhances the personalization of recommendations, ensuring that the images align not only with the user's visual preferences but also with their emotional profile.

3. EXPERIMENT

A. DATA SET

In this study, we utilized two datasets. The first dataset, referenced in [25], was used to detect emotions in images. It includes images associated with eight emotions, as outlined in Table 2: amusement, anger, awe, contentment, disgust, excitement, fear, and sadness. However, due to some missing images, the dataset used in this study contains fewer images than the original. Table 2 provides the number of images for each emotion, and Figure 3 presents eight sample images along with their corresponding emotions.
The second dataset, used to assess the impact of emotions on image recommendations, is a subset of the dataset from [4]. This dataset consists of 4000 images from 20 Flickr users. Figure 4 shows random samples of images marked as favorites by these users.

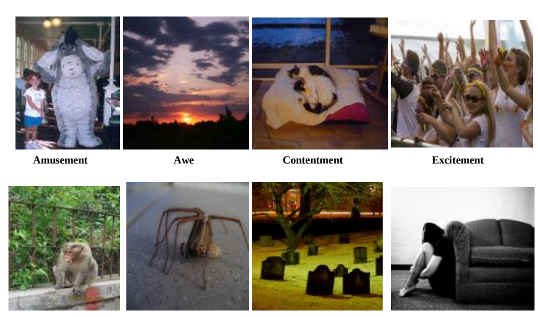


Figure 3. Sample images representing the eight emotion categories. The top row displays four positive emotions, while the bottom row presents four negative emotions [25].

Table 2: Statistics of the current labeled image data set [21, 25].

Sadness	Fear	Excitement	Disgust	Contentment	Awe	Anger	Amusement	Sum
2577	969	2725	1592	5130	2881	1176	4724	**21774**


Figure 4. random samples of the dataset (These images are from reference [4]).

B. EXPERIMENTAL RESULTS

In this section, we evaluate the performance of the proposed approaches for emotion-based image recommendation. To develop the recommender system, we first identified the emotions associated with each image. The methodology for emotion recognition is detailed in [21], where a model was designed to classify images into emotional categories. This process involves utilizing the EfficientNetB1 model, initialized with pre-trained weights from the ImageNet dataset and fine-tuned for the emotion recognition task. The final fully connected layer of the network was modified to match the number of emotional classes, producing a probability distribution over the emotion labels.

The extracted emotional features from this process serve as the foundation for the two proposed recommendation approaches: Emotion-Integrated Feature Vector and Emotion-Based Postprocessing. By leveraging emotional analysis, these methods aim to improve the personalization and quality of the recommendations.

C. IMAGE RECOMMENDATION BASED ON VISUAL EMOTION ANALYSIS

In this section, we examine how emotion analysis impacts the recommendation process. We utilize the FinetuneFMEfficientNetB1 method, which was introduced in the previous section for emotion recognition, to extract emotional features. Experiments were conducted with 20 users, each having 200 images marked as favorites. These images were input into the trained emotion recognition model to determine the emotions associated with each image, and the resulting output for each image is treated as an emotion feature vector.

In the first approach, to investigate the effect of emotion on the recommender system, we assess the integration of the emotion feature vector as part of the overall image feature vector (Figure 1). For each experiment, the dataset is split 50%/50% for training and testing. We require two sets of samples—positive and negative—for each user. Since we only have data on which images were marked as favorites by the users, identifying negative samples is not straightforward. To address this, we selected 100 out of the 200 images "liked" by each user as positive samples for training. For the negative samples, we first calculated the similarity between users and then selected 100 images from the 200 liked by the ten users with the least similarity (most distant) to serve as negative samples. The remaining images for each user were used as the test set.

After training, the test images for each user are ranked according to the scores generated by the classifier. We implemented various methods and report the top k recommended images, with k being a configurable parameter. Since we only have information on relevant images (those marked as "liked" by the user), we use Precision@k and Recall@k to assess the performance of the methods. Precision@k refers to the proportion of the top k images that are relevant, while Recall@k represents the proportion of relevant images included within the top k. The formulas for Recall@k and Precision@k are as follows:

$$\text{Recall} = \frac{\text{Relevant\_Items\_Recommended in topk}}{\text{Relevant\_Items}} \qquad (2)$$

$$\text{Precision} = \frac{\text{Relevant\_Items\_Recommended in topk}}{\text{k\_Items\_Recommended}} \qquad (3)$$

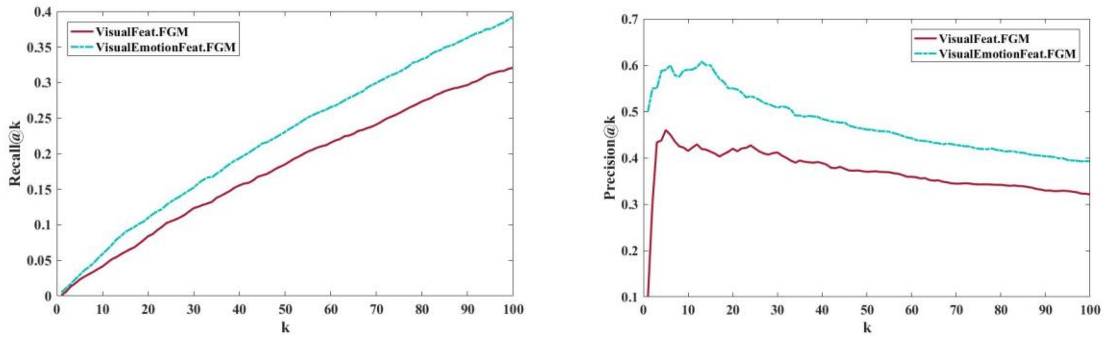


Figure 5. Precision@k for different k values (Top), Recall@k for different k values (Down).

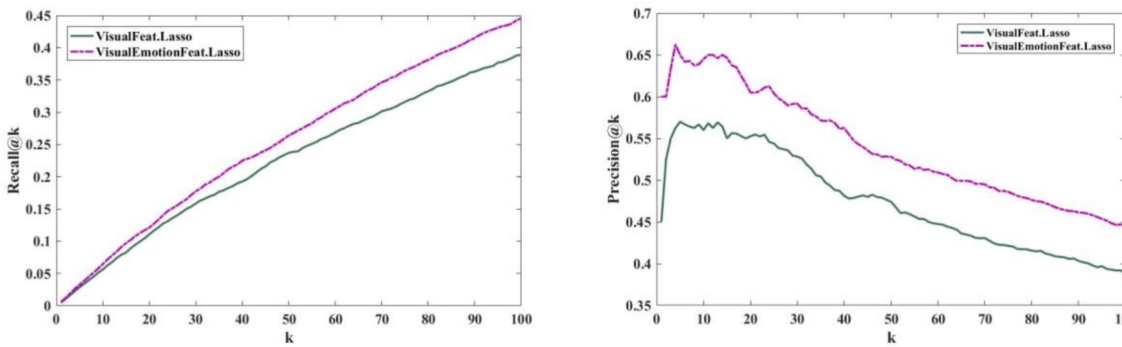


Figure 6. Precision@k for different k values (Top), Recall@k for different k values (Down).

In Figures 5 and 6, we compute Recall@K and Precision@K based on Equations 2 and 3. Recall@K measures the proportion of relevant items included in the top-K recommendations. Precision@K, on the other hand, evaluates the proportion of items in the top-K that are relevant (i.e., liked by the user). Precision reflects the system's ability to exclude irrelevant items from the recommendation list, and is influenced by the number of false positives, where the system suggests non-relevant items that shouldn't have been recommended.

In our experiments, the number of relevant items is fixed at 100. As a result, in recall, the denominator remains constant, and as K increases, it becomes more likely that relevant items will be recommended, causing recall to rise. For precision, however, as K increases, the denominator grows, but the numerator does not increase proportionally, leading to a decrease in precision.

Figure 5 presents a comparison between VisualFeat.FGM and VisualEmotinFeat.FGM. VisualFeat.FGM utilizes a visual feature vector, followed by FGM feature selection and classification as described in [26]. In contrast, the proposed approach, VisualEmotinFeat.FGM, incorporates both emotion and visual feature vectors before applying FGM feature selection and classification.

The comparison between these two methods demonstrates that incorporating emotion into the feature set enhances the performance of the recommender system.

Figure 6 also presents a comparison between VisualFeat.Lasso [4] and VisualEmotinFeat.Lasso. The findings from this comparison demonstrate that incorporating emotion features positively influences the performance of the recommender system.

In the second approach (proposed method in Figure 2), we consider the visual features as a feature vector and use the Lasso and FGM methods to recommend images to each user. Additionally, for each user, we calculate the average of the emotion vectors of the user's training images, which we call the user's emotion vector. For each image recommended to each user, we calculate the distance between the image and the user's emotion vector, and then rank them based on the closest distance. The recommended images are updated according to how close each image is to the user's emotion vector.

Figures 7 and 8 present the Precision@k and Recall@k results for this approach. The findings demonstrate that incorporating emotion in the postprocessing phase has a positive impact on the quality of image recommendations.

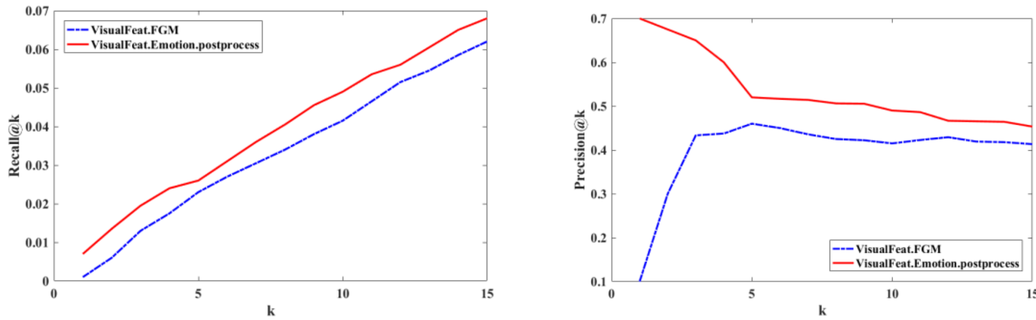


Figure 7. Precision@k for different k values (Left), Recall@k for different k values (Right).

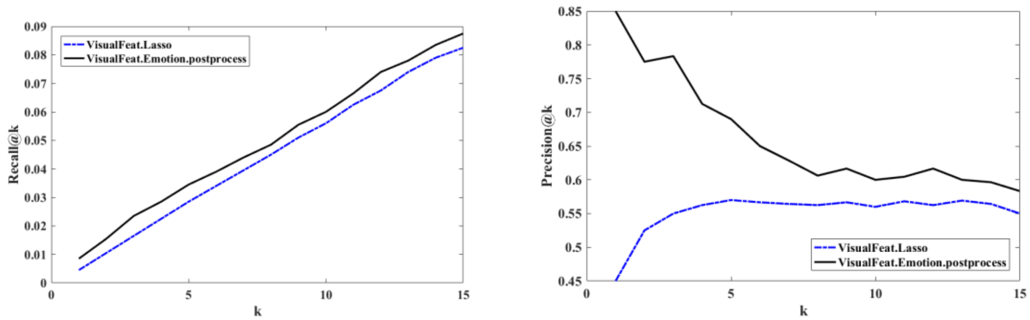
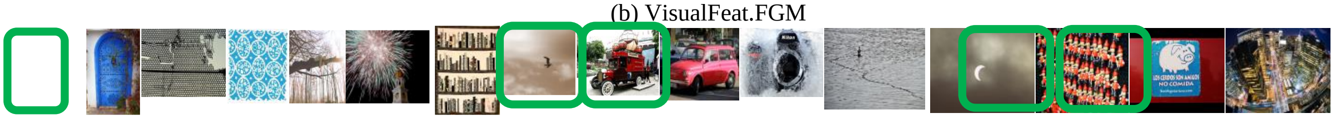


Figure 8. Precision@k for different k values (Left), Recall@k for different k values (Right).

Figure 9 illustrates the top 15 images recommended to a user, both with and without emotion-based postprocessing. To recommend 15 images, we first retrieve 20 candidate images using the recommender system. Then, we apply emotion postprocessing to select the best 15 images from the 20 candidates. This method was evaluated on 20 users, where emotion features were employed to refine the final recommendations

(a) Sample of random user training images

(b) VisualFeat.FGM



(c) VisualFeat.FGM.Emotionpostproces

(d) VisualFeat.Lasso

(e) VisualFeat.Lasso.Emotionpostprocess

Figure 9: (a) Sample training images for a user; (b)–(e) show the top 15 images recommended to the user using different methods, with correct recommendations highlighted in green.

The emotion-based postprocessing approach provides several advantages. It allows for a dynamic adjustment of recommendations based on the user's evolving emotional preferences and incorporates emotional closeness explicitly in the ranking process, resulting in more personalized recommendations. By separating the initial recommendation phase from the emotion-based refinement phase, it ensures that both visual and emotional aspects are given appropriate weight. This detailed focus on the method highlights how emotional features can be effectively utilized to refine and improve the quality of image recommendations.

4. CONCLUSION

Recommender systems aim to provide users with personalized and relevant suggestions, making it essential to understand their preferences. In this article, we explored the role of emotions conveyed by images in enhancing the recommendation process within a social image recommender system. We proposed two approaches to incorporate emotional analysis: Emotion-Integrated Feature Vector, which embeds emotional features directly into the model, and Emotion-Based Postprocessing, which refines recommendations based on emotional proximity. The experimental results demonstrate that incorporating the emotional features significantly improves the quality of image recommendations, as evidenced by enhanced metrics such as Recall@k and Precision@k. For future research, we suggest exploring advanced methods to improve the accuracy of emotion detection in images, as this could further enhance recommendation performance. Additionally, investigating the integration of emotional analysis with other behavioral traits, such as personality insights or sentiment analysis, could open new avenues for creating even more personalized and effective recommendation systems.

5. REFERENCES

1. Ahmadkhani, S. and M.E. Moghaddam, *An image recommendation technique based on fuzzy inference system.* Soft Computing, 2021. **25**(8): p. 6427-6443.
2. Ahmadkhani, S. and M.E. Moghaddam, *Image recommender system based on compact convolutional transformer image style recognition.* Journal of Electronic Imaging, 2022. **31**(4): p. 043054.
3. Chen, X., et al., *Deep reinforcement learning in recommender systems: A survey and new perspectives.* Knowledge-Based Systems, 2023. **264**: p. 110335.
4. Lovato, P., et al., *Faved! biometrics: Tell me which image you like and I'll tell you who you are.* IEEE Transactions on Information Forensics and Security, 2014. **9**(3): p. 364-374.
5. Guntuku, S.C., S. Roy, and L. Weisi. *Personality modeling based image recommendation*. in *MultiMedia Modeling: 21st International Conference, MMM 2015, Sydney, NSW, Australia, January 5-7, 2015, Proceedings, Part II 21*. 2015. Springer.
6. Zhang, S., et al., *Deep learning based recommender system: A survey and new perspectives.* ACM computing surveys (CSUR), 2019. **52**(1): p. 1-38.
7. Lei, C., et al. *Comparative deep learning of hybrid representations for image recommendations*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
8. Zhang, Y., et al. *Explicit factor models for explainable recommendation based on phrase-level sentiment analysis*. in *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*. 2014.
9. Contratres, F.G., et al. *Sentiment analysis of social network data for cold-start relief in recommender systems*. in *Trends and Advances in Information Systems and Technologies: Volume 2 6*. 2018. Springer.
10. Yang, Y., et al. *How do your friends on social media disclose your emotions?* in *Proceedings of the AAAI Conference on Artificial Intelligence*. 2014.
11. Jia, J., et al. *Can we understand van gogh's mood? learning to infer affects from images in social networks*. in *Proceedings of the 20th ACM international conference on Multimedia*. 2012.
12. Borth, D., et al. *Large-scale visual sentiment ontology and detectors using adjective noun pairs*. in *Proceedings of the 21st ACM international conference on Multimedia*. 2013.
13. You, Q., et al. *Robust image sentiment analysis using progressively trained and domain transferred deep networks*. in *Proceedings of the AAAI conference on Artificial Intelligence*. 2015.
14. Bakariya, B., et al., *Facial emotion recognition and music recommendation system using CNN-based deep learning techniques.* Evolving Systems, 2024. **15**(2): p. 641-658.
15. Katsurai, M. and S.i. Satoh. *Image sentiment analysis using latent correlations among visual, textual, and sentiment views*. in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2016. IEEE.
16. Yuan, J., et al. *Sentribute: image sentiment analysis from a mid-level perspective*. in *Proceedings of the second international workshop on issues of sentiment discovery and opinion mining*. 2013.
17. Siersdorfer, S., et al. *Analyzing and predicting sentiment of images on the social web*. in *Proceedings of the 18th ACM international conference on Multimedia*. 2010.

18. Cao, D., et al., *Visual sentiment topic model based microblog image sentiment analysis.* Multimedia Tools and Applications, 2016. **75**: p. 8955-8968.
19. Ji, R., et al., *Survey of visual sentiment prediction for social media analysis.* Frontiers of Computer Science, 2016. **10**: p. 602-611.
20. Xu, J., et al., *Visual-textual sentiment classification with bi-directional multi-level attention networks.* Knowledge-Based Systems, 2019. **178**: p. 61-73.
21. Ahmadkhani, S. and M.E. Moghaddam, *A social image recommendation system based on deep reinforcement learning.* Plos one, 2024. **19**(4): p. e0300059.
22. Flexeder, C., *Generalized lasso regularization for regression models*. 2010, Institut für Statistik.
23. Tan, M., I.W. Tsang, and L. Wang, *Towards ultrahigh dimensional feature selection for big data.* Journal of Machine Learning Research, 2014.
24. Tan, M., L. Wang, and I.W. Tsang. *Learning sparse svm for feature selection on very high dimensional datasets*. in *Proceedings of the 27th international conference on machine learning (ICML-10)*. 2010.
25. You, Q., et al. *Building a large scale dataset for image emotion recognition: The fine print and the benchmark*. in *Proceedings of the AAAI conference on artificial intelligence*. 2016.
26. Guntuku, S.C., et al., *Understanding deep representations learned in modeling users likes.* IEEE Transactions on Image Processing, 2016. **25**(8): p. 3762-3774.