Sequential Emotion-Aware Recommendation: A Preliminary Study on Data Acquisition and Modeling

Weihao Chen¹, Lin Zheng¹,²,* and Dazhi Jiang¹,²

¹Department of Computer Science, College of Engineering, Shantou University, China
²Key Laboratory of Intelligent Manufacturing Technology (Shantou University), Ministry of Education, China

*Corresponding author email: lzheng@stu.edu.cn

Abstract. The rapid development of recommender systems makes researchers pay more and more attention to the utilization of users’ subjective factors. Among various subjective factors, emotion plays an important role that can truly reflect the temporal changes of users’ preferences. However, the acquisition of user emotion is difficult, and is obtained passively in many methods. Moreover, there are few datasets suitable for continuous emotion modeling in existing recommendation research. To address these shortcomings, this work collects and organizes a dataset of users watching movies containing their temporal emotional data, which is appropriate for recommendation tasks based on sequential emotions. In addition, we propose a sequential emotion-aware recommendation method suitable for continuous emotion modeling, which outperforms other advanced sequential modeling approaches on our dataset. The contribution of this work is to conduct preliminary study and exploration on the modeling of users’ sequential emotion, so that this research direction can be further developed in future. The Sequential EMotion-Aware Recommendation (SEMAR) dataset we collected is available at: https://github.com/LinZheng666/Sequential-Emotion-Dataset-for-Recommendation.

Keywords: Emotion-aware recommendation; Sequential emotion; Emotion data acquisition; Emotion attention; Multiple emotion fusion.

1. Introduction

Nowadays, the research on recommender systems has made significant progress in several domains, and many researchers focus on enhancing recommender systems from the model perspective [1,2]. Other scholars improve recommendation performance by using additional information, such as text reviews or knowledge graphs. Most of the information is objective, for example, items’ attributes or users’ profile. Some of the additional information is subjective, such as user sentiment extracted from text reviews. However, sentiments from reviews generally have only three polarities: positive, neutral, and negative. Therefore, some researchers began to make use of users’ emotions as subjective information. Because individual emotions such as surprise, happiness, and sadness can more truly reflect users’ subjective thoughts, and therefore they can better reflect users’ preferences and affect their decision-making. In fact, as one of the most important subjective variables, emotion has been proven to be effective in different recommendation applications. Emotional information is first applied in recommender systems by Gonzale et al.[3], which is captured incrementally via small surveys. Based on this method, other work like [4,5] considering emotions as a context in emotion-aware recommender system are well-established. For example, Mizgajski et al. [6]
model emotions in multi-dimension for recommending news items, where the emotion is obtained from users’ explicit feedback, which is a kind of intrusive collection. Inspired by this work, Babanejad et al. [7] study the role of emotion features in news recommender systems by using implicit user feedback. Other applications such as movie recommendation and tourist destination recommendation have also been enriched by employing both user behavior and emotions in recommender systems [8, 9, 10].

For further exploring the role of emotions in relevant study, Zheng et al. [11] analyze the usage of emotional variables in the recommendation process and the influence of emotional information on rating prediction. More recently, Qian et al. [12] propose an emotion-aware recommender model, in which emotions are extracted from user reviews and hybrid information fusion strategy is employed. In 2020, Zheng et al. [13] proposed the Sentiment-Guided Sequential (SGS) recommendation model, which takes sentiments extracted from users’ reviews to provide a subjective-oriented sequential recommendation. This work proves that users’ subjective feelings after purchasing an item can affect their decision to purchase the next one.

Existing methods try to model user emotions to improve recommendation performance. However, emotion-aware recommendation still has the following challenges:

1) Existing emotion-aware methods rarely extract emotions from user faces, and can only be obtained passively, such as from reviews [12] or user questionnaire [11];
2) There are few publicly available emotion-aware recommendation datasets, and almost none include continuous emotional changes;
3) Existing emotion-aware models rarely model continuous emotions, which makes it difficult for models to capture user preference changes based on continuous emotions.

To address the above limitations, the main contributions of this work are as follows:

1) We collect a recommendation dataset of users watching movies containing their continuous emotions, which is suitable for emotion-aware recommendation tasks.
2) We try to achieve recommendation based on continuous emotions for the first time, and propose a sequential emotion-aware method suitable for continuous emotion modeling.

The rest of the paper is organized as follows. In the next section, we introduce the sequential emotion dataset we collect and preprocess. In Section 3, we present our method for how we model sequential emotions in detail. In Section 4, we conduct an experimental evaluation of our model and give some analysis of the results. Finally, we make a conclusion of the paper and directions for future work in Section 5.

2. Sequential Emotion-aware Recommendation (SEMAR) Dataset

Our first contribution is to establish a dataset that contains users’ continuous emotions for recommendation, we describe the dataset construction process in Figure 1 as follows.

![Figure 1. The main process of sequential emotion data collection.](image)

2.1. Collection of Users’ Continuous Emotions

We collected ten different types of short movies with no less than ten movies in each type. These ten categories include science fiction, swordsmen, costume, action, romance, drama, comedy, family, ethics, and war. We split each movie into clips, which are less than 4 minutes, and each of them are the most emotionally charged part of that movie. By doing this, we had generated 257 movie clips.
Then we developed an application and invited 159 users to help with the collection of the continuous emotional dataset. The application first iterates through ten types of movies from the sample material, randomly selects one of each type of movie clips, and then plays them to the user according to the sampling order. At the same time, a user's camera device is ready for recording the continuous emotional videos of that user's facial expressions in real-time; and it ended up with 1,590 videos. We finally obtained 151 users’ valid data, that is, 1,530 videos of the valid users' facial expressions when they were watching the randomly selected movie clips. In the end, all the users rated the movie clips they had watched in a range from 1 to 10.

2.2. Recognition of User Facial Emotions

The data we collected are video streams of faces. We first divide a video into multiple images at an interval of 5-7 frames. The images are then used to carry out face tracking and recognition. Traditional face recognition technology is mainly based on visual image, whose performance will decline sharply when the background environment changes greatly. Therefore, we adopt a face recognition library to obtain a 256*256 face image from the original image. Finally, we apply a deep model proposed in [17] to get emotional values of each face image corresponding to the user's facial expressions. Each emotion value contains 8 dimensions, including neutral, anger, disgust, happiness, contempt, sadness, fear, and surprise. Table 1 shows the statistics of the SEMAR dataset we established; Table 2 illustrates a sample of the sequential emotion values in the dataset; Table 3 and Table 4 show a sample of the ratings and movie information in this dataset, respectively.

**Table 1. The statistic of the SEMAR dataset.**

| Users | Ratings | Movie clips | Average Emotion Sequence Length |
|-------|---------|-------------|---------------------------------|
| 151   | 1530    | 257         | 1515                            |

**Table 2. A sample of the sequential emotion values in the SEMAR dataset.**

| Rating_id | neutral | happiness | surprise | sadness | anger | disgust | fear | contempt | time |
|-----------|---------|-----------|----------|---------|-------|---------|------|----------|------|
| 643       | 0.2664  | 0.1039    | 0.1046   | 0.1075  | 0.1049| 0.104   | 0.1039| 0.1044   | 0.1343|
| 643       | 0.2677  | 0.1038    | 0.1043   | 0.1070  | 0.1049| 0.1038  | 0.1038| 0.1042   | 0.2015|

**Table 3. A sample of ratings in the SEMAR dataset.**

| Rating_id | movie_clip_id | User_id | rating |
|-----------|---------------|---------|--------|
| 643       | 112           | 63      | 9      |

**Table 4. A sample of movie information in the SEMAR dataset.**

| movie_clip_id | cate_clip_id | movie_name | cate_genre |
|---------------|--------------|------------|------------|
| 112           | Cate0100001  | YeWen      | action     |
the user's final preferences, which is our preliminary study on how a recommender system performs continuous emotion modeling. The network architecture of MSEF is shown in Figure 2.

**Figure 2.** The architecture of the MSEF model.

First of all, since an emotion sequence is obtained by each user from diverse movies, we set a fixed input sequence length, and we will ensure that the length of each input sequence to MSEF is greater than this set length. There are lots of sequence modeling techniques, among them, the self-attention mechanism [16] is powerful and effective in long-sequence data modeling. The basic component of it is a scaled dot-product attention defined as follows:

$$
\text{Att}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V
$$

(1)

Where \( K, V, Q \) are keys, values, and queries, respectively. \( \sqrt{d} \) represents a scale factor that has different values according to the input dimensions. Given \( E \) as embeddings of the current sequence as input, the self-attention model is then defined as follows:

$$
\text{SelfAtt}(E) = \text{Att}(EW^Q, EW^K, EW^V)
$$

(2)

Here \( W^Q, W^K, W^V \in \mathbb{R}^{d \times d} \) indicate projection matrices. The self-attention mechanism considers the case of a single type of sequence input. However, in this case, there are as many as 8 emotion sequences, and self-attention cannot be used directly. To properly express the interaction between emotions, we establish a way to model emotions hierarchically to fuse multiple emotion sequences. In particular, we propose an Emotion-Attention (EmoAtt) mechanism that accepts two different types of emotion sequences as its input and output the fusion attention of the two emotion sequences as follows:

$$
\text{EmoAtt}(E_{m_1}, E_{m_2}) = \text{Att}(E_{m_1}W^Q, (E_{m_1} \odot E_{m_2})W^K, E_{m_2}W^V)
$$

(3)

Where the operator \( \odot \) denotes the element-wise multiplication. The first category of emotion \( E_{m_1} \) is employed as the query and the second category of emotion \( E_{m_2} \) is treated as the value. Their element-wise fusion \( (E_{m_1} \odot E_{m_2}) \) can be seen as the pairwise interaction of the two emotions, which plays a role as the key to connect them. The advantage of emotion-attention mechanism is that it can not only reduce the sequence number of emotions, but also well capture the interaction between emotion sequences. We can randomly select emotions for pairwise interaction, or perform a lot of pre-training to determine which two emotions are suitable for pairing, and get the pairwise output of 4 emotion interactions, denoted as EmoPair1, EmoPair2, EmoPair3, and EmoPair4. Then, we perform the second emotion interaction to implement the hierarchically emotion fusion. For example, by using EmoPair2 \( (E_{p_2}) \) and EmoPair3 \( (E_{p_3}) \) as query and value, the high-level emotion attention is defined as:

$$
\text{EmoAtt}_H(E_{p_2}, E_{p_3}) = \text{Att}(E_{p_2}W^Q, (E_{p_2} \odot E_{p_3})W^K, E_{p_3}W^V)
$$

(4)

The high-level emotion attention takes emotion interactions as input and generates emotion fusions as output. In particular, there are two emotion fusions denoted as \( E_{f_1} \) and \( E_{f_2} \) derived from \( \text{EmoAtt}_H \), which are used as input to a final MLP layer as follows:
\[ EmoPref = MLP(\text{ConCat}(E_{f1}, E_{f2})) \]  

The final output is the user preference derived from emotion, denoted as \( EmoPref \). We adopt such preference with a MSE loss function to implement rating prediction.

4. Experiments
For performance evaluation on the SEMAR dataset, we adopt the Root Mean Square Error (RMSE) metric that is commonly used in the rating prediction task. In method comparisons, we have selected three advanced sequential recommendation models in recent years. For example, GRU4Rec [14] is selected as the representative of traditional long sequence modeling, the SASRec [15] model is used as the representative model of the long sequence fusion attention mechanism, and SGS [13] is employed as the state-of-the-art representative study of modeling user sequential behavior and the relevant subjective factors. In experiments, we tried 4 different sequence lengths: 500, 800, 1000, 1500. The main hyper-parameters such as dropout_rate, learning_rate, batch_size, hidden_units, L2 are set to be 0.5, 0.05, 64, 16, 0.05, respectively. We chose AdamOptimizer to optimize the training process.

Table 5. The performance of the experiment comparisons.

| Length | GRU4Rec | SASRec | SGS | MSEF  |
|--------|---------|--------|-----|-------|
| 500    | 1.247   | 1.188  | 1.098| 1.012 |
| 800    | 1.252   | 1.152  | 1.117| 1.008 |
| 1000   | 1.274   | 1.209  | 1.075| 0.937 |
| 1500   | 1.276   | 1.138  | 1.124| 0.986 |

Figure 3. The Performance of the experiment comparison.

The experimental results are shown in Table 5 and Figure 3, where we use a light-red background and bold fonts to mark the best performances. In general, the MSEF model outperform other approaches throughout the experiments. Because GRU4Rec is more suitable for capturing the dependence between short sequences; though SASRec takes advantage of self-attention to solve long sequence dependence, but it does not take the influence of emotions into account. SGS does well in considering user subjective factors in long sequences, however, the proposed MSEF can better model sequential emotions by a hierarchical emotion fusion mechanism.

5. Conclusions and Future Work
In this paper, we collect and present a new dataset recording users’ temporal emotions to fill the gaps
in sequential emotion-aware recommendation. We then propose a hierarchical emotion fusion model to effectively capture the multiple sequential emotions interaction that outperforms other methods in the rating prediction task. For further research, we would like to carry out more complicated emotion modeling to deeply explore the potential abilities of different emotions in more recommendation tasks.

Acknowledgments
This research is funded by the National Natural Science Foundation of China (No. 61902231), the Guangdong Basic and Applied Basic Research Foundation (No. 2020A1515010531), and the Research Startup Fund Project of Shantou University (No. NTF18017). We thank Jianwei Chen for his great efforts in collecting the emotion-aware recommendation dataset.

References
[1] Zheng, Lin, et al. "Recommendation rating prediction based on attribute boosting with partial sampling." Chinese Journal of Computers. Jisuanji Xuebao 39.8 (2016): 1501-1514.
[2] Zheng, Lin, et al. "Memory reorganization: a symmetric memory network for reorganizing neighbors and topics to complete rating prediction." IEEE Access 8 (2020): 81876-81886.
[3] Gonzalez, Gustavo, et al. "Embedding emotional context in recommender systems." 2007 IEEE 23rd international conference on data engineering workshop. IEEE, 2007.
[4] Han, Byeong-jun, et al. "Music emotion classification and context-based music recommendation." Multimedia Tools and Applications 47.3 (2010): 433-460.
[5] Tkalčić, Marko, et al. "Emotion-aware recommender systems—a framework and a case study." International Conference on ICT Innovations. Springer, Berlin, Heidelberg, 2012.
[6] Mizzagksi, Jan, and Mikolaj Morzy. "Affective recommender systems in online news industry: how emotions influence reading choices." User Modeling and User-Adapted Interaction 29.2 (2019): 345-379.
[7] Babanejad, Nastaran, et al. "Leveraging Emotion Features in News Recommendations." INRA@ RecSys. 2019.
[8] Zheng, Lin, et al. "Context neighbor recommender: Integrating contexts via neighbors for recommendations." Information Sciences 414 (2017): 1-18.
[9] Zheng, Lin, Fuxi Zhu, and Alshahrani Mohammed. "Attribute and global boosting: A rating prediction method in context-aware recommendation." The Computer Journal 60.7 (2017): 957-968.
[10] Ishanka, Piumi, Takashi Yukawa, and Takashi Yukawa. "An analysis of emotion and user behavior for context-aware recommendation systems using pre-filtering and tensor factorization techniques." Global Journal of Computer Science and Technology (2018).
[11] Zheng, Yong, Bamshad Mobasher, and Robin D. Burke. "The Role of Emotions in Context-aware Recommendation." Decisions@ RecSys 2013 (2013): 21-28.
[12] Qian, Yongfeng, et al. "EARS: Emotion-aware recommender system based on hybrid information fusion." Information Fusion 46 (2019): 141-146.
[13] Zheng, Lin, et al. "Sentiment-guided sequential recommendation." Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. 2020.
[14] Hidasi, Balázs, et al. "Session-based recommendations with recurrent neural networks." arXiv preprint arXiv:1511.06939 (2015).
[15] Kang, Wang-Cheng, and Julian McAuley. "Self-attentive sequential recommendation." 2018 IEEE International Conference on Data Mining (ICDM). IEEE, 2018.
[16] Vaswani, Ashish, et al. "Attention is all you need." arXiv preprint arXiv:1706.03762 (2017).
[17] Barsoum, Emad, et al. "Training deep networks for facial expression recognition with crowdsourced label distribution." Proceedings of the 18th ACM International Conference on Multimodal Interaction. 2016.