EEG Signal Processing for Recommender Systems

Diego Mendieta
PUC
Santiago, Chile
damendieta@uc.cl

Nicolás Sumonte
PUC
Santiago, Chile
nisumonte@uc.cl

Diego Iruretagoyena
PUC
Santiago, Chile
diiru@uc.cl

Abstract

Covert aspects of ongoing user mental states provide key context information for user-aware human computer interactions. Leveraging affective state estimation coming directly from users brains could result in bias-free item recommendations, expanding and enhancing current state-of-the-art methods. A consumer item that has one of the strongest relationship with human emotions is music. Therefore, our research focuses in music recommendation based on images generated by brain signal time series. We propose a Transformer approach for brain signal processing. Experimental results demonstrate that emotion based recommendation surpasses effectiveness of baseline methods, validating the use of sentiments in recommendation tasks.

CSS Concepts

Information systems → Recommender Systems
Human computer interaction → EEG
Continuous mathematics → Transformers, Time Series

1 Introduction

An emotion is defined as a short, intense human reaction that's triggered by an stimulus. This stimulus acts as an input to the human brain, which in turn makes a person feel a specific feeling [1]. As we all know, emotions are the fundamental pillar that holds our social relationships and our psychological state [2].

In fact, researchers in the area of psychology and behavioral neuroscience agree that emotions influence our thinking, our decision making, well-being and our mental and physical health [3]. Technological advances in the last decades have allowed us to brag about the understanding of the brain.

One of these inventions corresponds to Human-Computer Interfaces, which are based in brain signal acquisition, processing and interpretation through time series. An example of these artifacts are EEGs, which capture our brains cognitive dynamic by measuring electric signals coming from our brain, transforming them into readable time series. This is performed by measuring electric potential deltas and signal frequency. EEGs market is estimated to be in the order of USD 2 billion by 2026 [4].

2 State of the Art

Recently, uses of emotion-based recommender systems have been explored. FaceFetch [5] , for example, leverages facial expression in order to comprehend users emotional state to recommend items, such as videos and music. EmoWare [6], utilizes emotions under an implicit feedback system to recommend videos. Lastly, Moodplay [7] uses emotions described by users while consuming music in order to recommend artists that evoke similar feelings.

Databases and experiments that intend to generate a correlation between user explicit and EGG based preference measurements include DEAP [8], SEED [9], DREAMER [10] and Emotify [11].

3 Solution

This paper focuses in music recommendation based on EEG signals. We propose a Transformer approach for brain signal sequences processing.

The process consists in an initial preprocessing of labels, mapping scalar ratings with a domain between 1 to 10 into a binary classification label indicating user preference (like/dislike). The average rating of items has been used as threshold.

Then, a temporary-visual representation is generated from the time series given by the electroencephalogram nodes. An image of thirty frames with a fixed lenght of 1 second duration of readings is generated for each item. Then, these images are processed by a **DenseNet121** Convolutional Neural Network to obtain representative sequence features. These sequence features are given as input to a **Transformer's Decoder**, followed by a fully connected layer that generates a binary classification representing like or dislike for an item. To measure the effectiveness of the transformer, measures as accuracy, recall y precision were tracked and improved.

4 Dataset

For our experiment we have chosen the DEAP dataset. This dataset was released by a group of researchers following an experiment conducted in 2012 that sought to interpret mood states of individuals when consuming multimodal content. It collects data from 32 subjects who were exposed to 40 music videos.

The data obtained are presented in time series obtained by an electroencephalogram that measures brain activity in 32 brain channels and 8 channels of physiological activity such as heart rate, temperature, breathing rate, among others.

The selection of music videos was obtained from a platform that allows users to tag songs with keywords. Therefore, 120 music videos were chosen based on the most popular tags of the moment.

In order to reduce the number of stimuli and find the most stimulating minute of each video, the J. Kierkels, et. al. model was used, which analyzes the most stimulating seconds of a video based on volume, auditory signal, visual arousal, and other variables.

After applying this algorithm, the number of videos was reduced to 40, considering popularity and the aforementioned model. Finally, the users participating in this study were subjected to the electroencephalogram test and were asked to rate the emotionality generated by each stimulus by rating on a scale of 0-9 the valence, arousal, dominance, and pleasantness that each video provoked in them.

To reduce the source of errors, the signals were preprocessed by reducing the sampling frequency and applying some filters. The data set used in this project is the one obtained after processing.

5 Methodology

In this section, we report the pipeline followed in order to take advantage of the usable information coming from the database of EGG Signals. The neurophysiological experiment consisted of multimodal content which was presented to participants, their evoked brain responses were obtained via EEG, and the collected EEG data was associated with personal preference ratings provided by each participant. This data was used to train classifiers for each participant. The predicted outputs from these classifiers were then used as inputs for various collaborative filtering models.

5.1 Pipeline

In order to formalize the inference of preferences using brain signals, we define a set of brain signals **S**, a user set **U** and an item set **I**, searching to create a function **A**, which estimates user u preferences for an unseen item i. A is defined as seen in Equation 1.

$$A : S, U, I \rightarrow Y \tag{1}$$

For this objective, our architecture is defined using three steps. First, a transformation of time series into a temporary visual representation of the EEG. Second, a preference estimation utilizing machine learning methodologies and third, a preference prediction based on user-item interactions. These three steps are explained as follows.

5.2 Temporary-Visual Representation

DEAP dataset is based in electroencephalogram readings that use 32 nodes in order to acquire the brain wave signals expressed while a user consumes items, in this particular case, musical videos. These signals can be represented in a unidimensional way in a given time, with

$$s_t = \left[c_t^1, \cdots, c_t^n\right]^T \in R^n \tag{2}$$

$$t \in (0, ..., n-1) \tag{3}$$

where T represents transposition, n the number of electrodes (1...32) and c_i^t represents the data of the i-term electrode at specific time t, Each of these vectors s_t can be represented in a topographical map ($Tp_t(s_t)$), which illustrates the cerebral activation on a given moment and represents s_t signal frequency graphically.

With the objective of generating a notion of the current mind state and being able to take advantage of the rich visual information generated by brain deltas, a vector is created using 30 seconds of the original 60 seconds total duration of an item. Then, an image is created using these vectors, creating a matrix represented as Equation 4.

$$S_{ui} = [Tp_0, Tp_1, \cdots, Tp_{29}]^T \tag{4}$$

This matrix contains the frames of the topographic map of half of the total duration of an stimulus for user u and item i. Then, it is possible to obtain 40×32 matrices S_{ui} corresponding to each user and each item.

5.3 Preference Estimation

First, our algorithm learns existing relationship between the images generated by the brain signals, and the preferences of user expressed like / dislike.

This enables the model to predict new preferences for unseen items of the given user, where the information of the preference given by that user to that item being hidden manually in the database for the experiment, so the model has no information of previous behavior.

Given a topographical map $S_u i$ for an item i, a user u and a associated preference y_i, a mapping

$$A : S_{ui} \to y_j \tag{5}$$

is learned where given a new set of frames S_{uj} for an item $j \neq i$, we can approximate a new preference y_j.

5.4 User Interaction

The algorithm learns relationships between users and items in a way where given a new item where no brain signal information of a user is provided, it is possible to infer the new preference based on explicitly given preference (like/dislike) from other users. For this, the preference predictions $\hat{y}$ for a given user u and an item i, we create a mapping function

$$A : U, I \to \hat{y_{ui}} \tag{6}$$

where given an unseen item i and a user u it is possible to estimate preference y_{ui}. This problem was resolved by using matrix factorization, where user-item interactions are elucidated based on a collaborative filtering using implicit feedback, as described in [11].

6 Experiments, Parameter Analysis and Control Conditions

6.1 Machine Learning Experiments

At the beggining of the experiment we have the readings of 32 electrodes of an encephalogram. These time series are then use to condense this information into images, taking 30 frames of 1 second each.

A DenseNet 121 CNN is then utilized to process the brain signal generated images. From this we obtain image features sequences. These sequences are given to a Transformer's Encoder and a fully connected layer to obtain the binary classification that represents the user preference. This classification was then measured and tuned analyzing its accuracy, recall and precision in order to optimize the Transformer's effectiveness.

This described method is used to create a general model with all users, and also to create individual, custom models for each user.

6.2 Recommendation Experiments

In order to obtain the Users and Items profiles, we utilize this previous classification to generate a sparse matrix of dimension 32x10 corresponding Users-Items. These matrices are then processed by BPR and ALS methods. To measure BPR and ALS effectiveness, we use nDCG and mAP metrics, varying and tuning their hyperparameters in order to maximize gain. We utilize these profiles to create a recommendation over the test set.

6.3 Parameter Analysis

The first preprocessing decision was to map the user given explicit rating into a binary variable indicating if the user liked the item or not.

$$p_{ui} = \begin{cases} 1 & r_{ui} \geq \bar{r_{ui}} \\ 0 & r_{ui} < \bar{r_{ui}} \end{cases} \tag{7}$$

Then, images were generated taking snapshots of 30 frames of 1 second of duration. This length was set in order to capture at least half of the stimuli information obtained for any given input. This resulted in a mapping between time series and images. A DenseNet 121 was used in order to obtain feature sequences from the generated images. For the Transformer, 2 blocks of a dimension of 2048 were used.

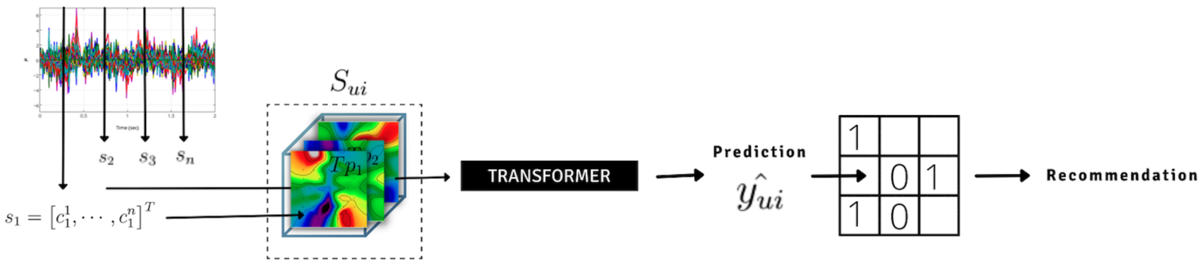

Figure 1: Step by step predictive model architecture

7 Results

7.1 Control Conditions

7.1.1 Random performance control

It is worth to mention that the objective of this study is not to perform a comparison between prediction performances between architectures, but to study if brain signals are an alternative of explicit retroalimentation for recommender systems. Therefore, we exclude the time and memory analysis of different architectures. Following this approach, an experiment where label permutation in the classifier's output is presented, both as baseline and as potential data bias control, to then proceed and recommend items with the architecture previously described.

7.2 Classifier Model

The results of the training for each model are presented in Figure 2. As we can observe, highest score surrounded 0.9 of accuracy. This proves the Transformer plus the fully connected layer to be good classifiers.

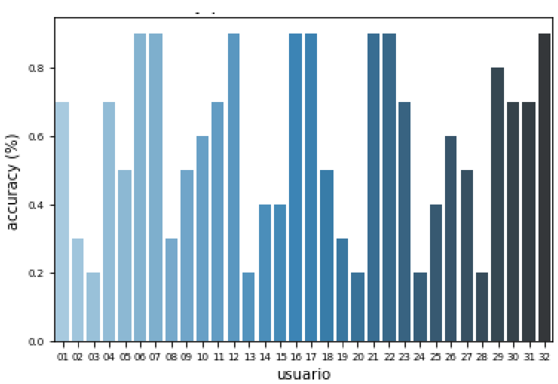

Figure 2: Results for each model training

7.3 Recommendation Model

The results for BPR and ALS are presented in Table 1 and 2. It is observed that the best results for ALS were mAP=0.465 using 5 factors and nDCG=0.487 of ALS@200 using 7 factors. For BPR, best results were BPR@50 mAP=0.496 and BPR@100 using 7 factors.

ALS	mAP			nDCG		
factors	3	5	7	3	5	7
50	0.465	**0.465**	0.465	0.439	0.458	0.486
100	0.437	0.437	0.437	0.419	0.436	0.465
200	0.465	0.465	0.437	0.451	0.458	**0.487**
500	0.436	0.436	0.437	0.417	0.436	0.465
1000	0.441	0.441	0.441	0.434	0.439	0.468

Table 1: ALS results using top 3, 5, 7 and varying factors

BPR	mAP			nDCG		
factors	3	5	7	3	5	7
50	0.496	0.496	**0.496**	0.489	0.499	0.515
100	0.495	0.495	0.495	0.485	0.500	**0.516**
200	0.431	0.431	0.431	0.429	0.448	0.457
500	0.494	0.494	0.494	0.482	0.497	0.514
1000	0.494	0.494	0.494	0.476	0.502	0.515

Table 2: BPR results using top 3, 5, 7 and varying factors

The matrix that was previously obtained through the Transformer and the fully connected layer recommendation is now replaced by a matrix with random binary values. This was to enable us to measure the effectiveness of the Transformer approach in the recommendation task. The results obtained after ALS and BPR method applications are presented in the following table.

	mAP			nDCG		
	3	5	7	3	5	6
BPR	0.390	0.390	0.390	0.348	0.394	0.423
ALS	0.428	0.428	0.428	0.413	0.426	0.457

Table 3: ALS and BPR on Random Sparse Matrix

8 Challenges and Design Decisions

8.0.1 Volume and diversity of data

While the research topic is interesting and novel, it also presents various challenges. First, it is clear that there is a significant lack of volume and quality of information. To put it in perspective, of the most commonly used databases for these studies, we find that DREAMER has samples from 23 participants, SEED contains 15 and DEAP has 32, while Emotify has 400 song excerpts.

In addition, the experiments presented are difficult to reproduce in their entirety, as some of the machines used to measure these signals are expensive and are usually found in specialized research centers such as universities or laboratories, which adds a significant barrier to entry. Finally, getting people who are willing to undergo the study multiple times with different items can be expensive in time and money.

8.0.2 Human emotion mapping

One of the first approaches in the study was to perform a correct mapping between the EEG time series and the values of valence and arousal to human emotions. Emotions are the fundamental pillar that sustains social and psychological relationships [14]. In the area of psychology and behavioral neuroscience, it is claimed that emotions influence thinking, decision making, well-being, actions, and mental and physical health [15].

There are multiple investigations that have attempted to construct a function that connects these domains. A notable study conducted by Mariano Scandar in 2019 [13] shows that there is indeed a mild interaction between content type and emotional valence, where it is evident that participants under positive valence performed better in tasks involving memory challenges than those who had reduced negative valence induction (always with stable activation and dominance at high levels).

Furthermore, it is established that valence does indeed play a role in content comprehension. Another widely referenced study in this attempt to spatially represent emotions comes from the work of James A. Russell, who produced the well-known circumflex model of affect. In this work [4] it is established that while analytic evidence has led most psychologists to describe affect as a set of dimensions, such as disgust, distress, depression, and arousal, it is possible to show evidence that, with each dimension varying independently of the others, rather than being independent, these affective dimensions are interrelated in a systematic way.

The evidence suggests that these interrelationships can be represented by a spatial model in which affective concepts obtain positions within a circle, associated with an angle. This mapping establishes: pleasure (0), excitement (45), arousal (90), distress (135), disgust (180), depression (225), sleepiness (270), and relaxation (315). This model was offered both as a way for psychologists to represent the structure of affective experience and as a representation of the cognitive structure for conceptualizing affect.

Supporting evidence was obtained by scaling 28 adjectives denoting emotions in 4 different ways: the 1938 RT Ross technique [2] for a circular ordering of variables, a multidimensional scaling procedure based on perceived similarity between terms, a unidimensional scale on the pleasure-disgust hypothesis, and arousal degree dimensions.

This study refrained from setting the recommendation task into a quadrant focused objective, and chose to exploit image processing and explicit user rating relationship in order to make a recommendation. This was mainly due to lack of sufficient data volume to support the mapping analysis. Instead, the research focused on analysis like / dislike of a user for any given item present in the database, where no previous data existed.

8.0.3 Absence of neutral groups

One of the main limitations of the candidate databases analyzed is the absence of a neutral condition within each experiment. Taking into account the doubts that have arisen during the analysis of the data obtained, it becomes evident that, if we had had a group in which no emotional state had been induced, we could have greater certainty about the role of emotional induction in the consumption of content, not only by comparing one dimension against its opposite but with a control group, which could be relevant when evaluating the relevance of inducing emotions as a way of improving the cognitive capacity of the subjects.

8.0.4 Data processing

There is a subjacent challenge while using multiple EEG nodes, such as correctly interpreting data for embedding generation. Which channels of the encephalogram to use? Which ones to discard? If you have a mapping of items to measures such as valence and arousal, how do you translate these values into an emotion map? In order to make informed decisions, the project builds on existing research in order to choose the most representable readout channels, thus using data filtering similar to that used in DEAP [10].

8.0.5 Handling of time series peaks

There are certain error spaces in the measurements provided by EEG signals. When an individual performs a movement or blinks, an action potential occurs that affects the EEG measurements producing signal peaks. This is why brain signals require some attention in their post-processing, where to reduce sources of errors, corrections are applied to the database using filters and frequency reductions. One way to deal with this is to use the convention used in our training database and reduce the sampling frequency along with employing a bandpass filter at the frequency. This managed to overcome described anomaly peaks.

8.0.6 Ground truth and recommendation design decision

The first approach taken was to utilize the emotion mapping to enclose the recommendation objective as trying to induce an specific emotion to the user, depending to their previous mood, and then measure whether or not the recommended song was of the user's liking. This approach was discarded due to lack of a reasonable amount of data that would enable a correct mapping.

In addition, contact was made with the UC Laboratory of Psychophysiology to request the use of their encephalogram equipment, in order to generate a ground truth dataset of our own. This was not possible given strict timelines and COVID procedures, therefore is presented as a future work possibility.

9 Conclusions

In this work, we explored preferences inferred from human brain signals by creating a temporary-visual representation of signals, preference estimation and matrix factorization. The results enable the conclusion that user biometric information can help augment models to generate a more effective recommendation. Our results and the ones from cited papers also show that brain responses can be reliably associated with self-reported preferences and that preferences can be predicted from EEG/ERP responses in a single-trial setting.

As the methodology followed in this paper and its references relies on ERPs, in principle it allows to operate on any perception, whether it is visual, auditory, somatosensory, or even olfactory or gustatory. This opens new avenues for capturing user preferences and interests in real-time as users experience the physical and digital world around them.

9.1 Future Work

An in-depth effort is required in the availability of open-source databases with these type of measurements, hopefully with a wider range of activities beyond visual or musical consumables. Furthermore, extensive research on human emotion mapping was explored but not developed beyond theory, utilizing Russel's circumplex model [6] in order to focus the recommendation problem into maintaining or changing a given user emotional quadrant.

10 Ethical considerations

The use of brain signals in recommendation has a great ethical underlying challenge regarding the storage and use of BCI data, since this data contains sensitive biometric information about users and could be used evilly. Because of this, researchers must be careful of data handling and sharing.

References

Izard, C. E. (2010). The many meanings/aspects of emotion: Definitions, functions, activation, and regulation. Emotion Review, 2(4), 363-370.

Widen, S.C.; Russell, J.A. Descriptive and Prescriptive Definitions of Emotion. Emot. Rev. 2010, 2, 377–378.

Kong, T., Shao, J., Hu, J., Yang, X., Yang, S., and Malekian, R. (2021). EEG-Based Emotion Recognition Using an Improved Weighted Horizontal Visibility Graph. Sensors, 21(5), 1870.

Grand View Research (2019). Electroencephalography Devices Market Size, Share and Trends Analysis Report By Product (pp.81-11). GVR.

M.B. Mariappan, M. Suk, B. Prabhakaran, "FaceFetch: A User Driven Multimedia Content Recommendation System Based on Facial Expression Recognition", 2012 IEEE International Symposium on Multimedia, vol.1, pp. 84-87.

A. Tripathi, T.S. Ashwin, R.M. R. Guddeti, "EmoWare: A Context-Aware Framework for Personalized Video Recommendation Using Affective Video Sequences", IEEE Access, vol. 7, pp. 51185-51200, 2019.

Andjelkovic, I., Parra, D., and O'Donovan, J. (2019). Moodplay: Interactive music recommendation based on Artists' mood similarity. International Journal of Human-Computer Studies, 121, 142-159.

S. Koelstra, C. Muehl, M. Soleymani, J.-S. Lee and A. Yazdani. DEAP: A Database for Emotion Analysis using Physiological Signals. EEE Transactions on Affective Computing, vol. 3, no. 1, pp. 18-31. (2012).

Girdhar, R., Carreira, J., Doersch, C., and Zisserman, A. Video action transformer network. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 244-253). (2019).

Gábor Takács and Domonkos Tikk. Alternating least squares for personalized ranking. In Proceedings of the sixth ACM conference on Recommender systems (RecSys '12). (2012).

Rendle, S., Freudenthaler, C., Gantner, Z., and Schmidt-Thieme, L. BPR: Bayesian personalized ranking from implicit feedback. arXiv preprint arXiv:1205.2618. (2012).

Keith M. Davis III, Michiel Spapé, and Tuukka Ruotsalo. Collaborative Filtering with Preferences Inferred from Brain Signals. In Proceedings of the Web Conference 2021 (WWW '21). Association for Computing Machinery, New York, NY,USA,602–611. doi.org/10.1145/3442381.3450031. (2021).

Russell, J. A. A circumplex model of affect. Journal of Personality and Social Psychology, 39(6), 1161–1178. doi.org/10.1037/h0077714. (1980).

S. Katsigiannis, N. Ramzan, "DREAMER: A Database for Emotion Recognition Th- rough EEG and ECG Signals from Wireless Low-cost Off-the-Shelf Devices," IEEE Jour- nal of Biomedical and Health Informatics, vol. 22, no. 1, pp. 98-107, Jan. 2018. doi: 10.1109/JBHI.2017.2688239.

Zheng, Wei-Long Lu, Bao-Liang. (2016). A Multimodal Approach to Estimating Vigi- lance Using EEG and Forehead EOG. Journal of Neural Engineering. 14. 10.1088/1741- 2552/aa5a98.

Aljanaki, Anna, Frans Wiering, and Remco Veltkamp. "Collecting annotations for indu- ced musical emotion via online game with a purpose Emotify", Technical Report Series 2014, UU-CS-2014-015 (2014).