Collaborative Filtering with Preferences Inferred from Brain Signals

Keith M. Davis III
keith.davis@helsinki.fi
University of Helsinki, Finland

Michiel Spapé
michiel.spape@helsinki.fi
University of Helsinki, Finland

Tuukka Ruotsalo
tuukka.ruotsalo@helsinki.fi
University of Helsinki, Finland and
University of Copenhagen, Denmark

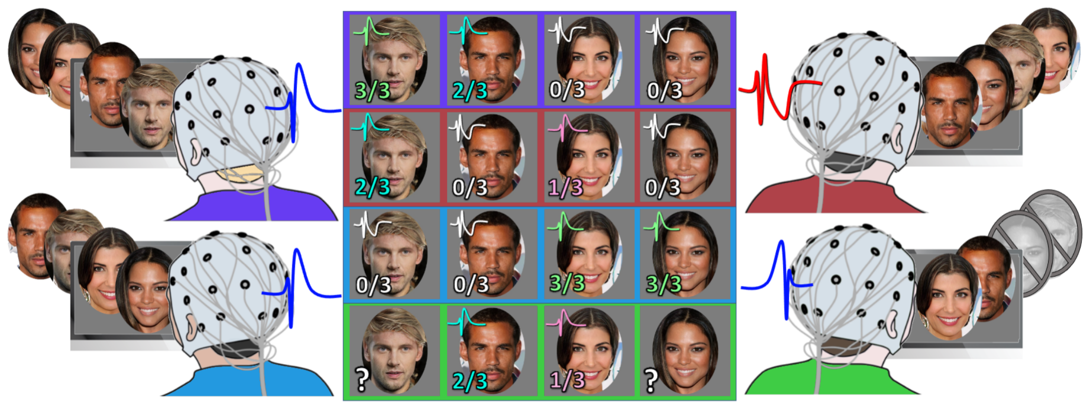

Figure 1: Applying collaborative filtering to preferences inferred from the brains of individuals to estimate preferences for unseen samples.

ABSTRACT

Collaborative filtering is a common technique in which interaction data from a large number of users are used to recommend items to an individual that the individual may prefer but has not interacted with. Previous approaches have achieved this using a variety of behavioral signals, from dwell time and clickthrough rates to self-reported ratings. However, such signals are mere estimations of the real underlying preferences of the users. Here, we use brain-computer interfacing to infer preferences directly from the human brain. We then utilize these preferences in a collaborative filtering setting and report results from an experiment where brain inferred preferences are used in a neural collaborative filtering framework. Our results demonstrate, for the first time, that brain-computer interfacing can provide a viable alternative for behavioral and self-reported preferences in realistic recommendation scenarios. We also discuss the broader implications of our findings for personalization systems and user privacy.

CCS CONCEPTS

• **Information systems** → **Collaborative filtering**; *Personalization*; • **Human-centered computing** → *Interaction paradigms.*

WWW '21, April 19–23, 2021, Ljubljana, Slovenia

ACM ISBN 978-1-4503-8312-7/21/04.
https://doi.org/10.1145/3442381.3450031

KEYWORDS

Brain-computer interface, collaborative filtering, brain signals, eeg

ACM Reference Format:
Keith M. Davis III, Michiel Spapé, and Tuukka Ruotsalo. 2021. Collaborative Filtering with Preferences Inferred from Brain Signals. In *Proceedings of the Web Conference 2021 (WWW '21), April 19–23, 2021, Ljubljana, Slovenia.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3442381.3450031

1 INTRODUCTION

The use of collaborative filtering to infer user preferences and recommend content according to an individual's needs has become ubiquitous across the web. It is widely adopted by industries to contend with information overload, from entertainment providers creating personal music [59] or movie [6] recommendations, to news aggregation services recommending articles of preferred topics [14] and online retailers suggesting products for customers [37].

Recent years have seen extensive research in developing new techniques to improve collaborative filtering results and better utilize the ever-increasing amount of user and content data. In particular, a wide variety of neural architectures have been introduced that may offer more robust estimations of user behavior and interests [10, 22, 25, 69].

Despite these advancements, it has become clear that the performance of collaborative filtering is not solely dependent on the exact computational method [12], but relies more on the quality of the signal capturing user preferences. That is, significant potential for advancement lies within novel user signals and their ability to represent the "true" underlying user preference.

Previous work has exploited a variety of explicit and implicit user signals to model the users, for example by using reviews and ratings [38], click behavior [39], time spent attending to certain content [70], content saving or sharing [47], explicit interactions [53, 54], and affective information recorded from the users [2, 5]. However, explicit and implicit signals are just indirect probes of the real, underlying preferences of the users inferred from behavior. Consequently, they are prone to bias and error [1].

Even explicit ratings are not always indicators of direct preference, but may instead reflect polarized opinions of the users – users may be selectively rating items to express support or opposition to various aspects of the rated content. Implicit measures relying on behavioral probes can also be unreliable [27, 44], as users may also click items by mistake or out of curiosity to find out what the actual content is like. Furthermore, users might interact with content they otherwise do not prefer because they find it outrageous or particularly unusual, or simply because they a victim of link baiting – a phenomenon used by content providers to entice visitors.

Here, we propose a novel alternative signal to capture the real preferences of users: direct measurement of the human brain. Such an approach is not dependent on implicit or explicit behavior. Instead, it only requires the user to perceive the content. We reveal neurophysiological markers of graded preference in electroencephalography (EEG) from a realistic preference experiment and use the inferred preferences in a collaborative filtering setting, as illustrated in Figure 1.

In detail, we ask the following research questions:

RQ1: Can brain responses be used to predict user preferences?

RQ2: Can preferences inferred from the brain be used for collaborative filtering?

Our results demonstrate, for the first time, the decoding and predicting of preferences directly from data collected via EEG. We then show that matrix factorization and neural collaborative filtering models can effectively incorporate this information and yield pragmatic recommendation performance by using only brain signals as input.

2 BACKGROUND

2.1 Collaborative Filtering

Collaborative filtering is a technique in which data from many users are used to inform personalization systems and tailor content recommendation to better match an individual's needs and interests [24, 56]. Since its early applications in news and email filtering, as described by [24], collaborative filtering has become a widely researched topic in academia [59] and technique used across a variety of industries [6, 14, 37]. The cornerstone of collaborative filtering is the availability of data that correlates with the underlying user preferences. Research has mainly relied on two types of data: explicit ratings available from the users and implicit behavioral signals, such as clicks or dwell time [40].

These data are exploited by a wide variety of collaborative filtering approaches, from classic matrix factorization [26, 34] and k-nearest neighbor [67] approaches to more specialized models. Many neural architectures have been proposed, and an extensive survey of these approaches is provided in [71]. These vary from restricted Boltzmann machines and autoencoders [55, 57] to more recent neural matrix factorization and graph-based collaborative filtering [25, 68].

Recently, many of the novel algorithmic approaches have been called into question. In some instances, new approaches to collaborative filtering are just different architectures that do not necessarily show improved performance apart from specific datasets or selected baselines [13]. As it is unclear whether or not the newest methods are genuinely better than well-established baselines, a more promising approach to improving the performance of collaborative filtering approaches is to obtain more information from users.

In addition to collecting implicit interaction data from users to further enhance information retrieval and personalization systems, physiological information and affective features related to an individual user have also been used. Early work in this area includes the use of affective features in a small collaborative filtering scenario [43], as well as successful attempts in using facial expressions to enhance search results [3]. Affective metadata has also been used to retrieve and label image content [62, 63], while emotional responses inferred from a user's physiology have been used successfully to improve music recommendations [4]. In this work, we utilize data beyond explicit ratings, implicit behavioral measures, and peripheral physiological signals, and derive users' real preferences by directly measuring brain responses to content.

2.2 Brain Imaging for Information Retrieval

Brain imaging and brain-computer interfacing refers to techniques and approaches to measuring and utilizing brain signals for purposes of understanding and interacting with digital systems. For example, [46] utilized brain imaging methods to reveal how information needs arise, and how they can be predicted from fMRI data. Functional near-infrared spectroscopy (fNIRS) relies on the same signal captured by fMRI but presents a more affordable, wearable alternative that shows also potential for information filtering in a recommendation context [49]. Magnetoencephalography (MEG), which measures the magnetic field produced by synchronously firing neurons, has also been used, such as in [32], to infer the relevance of images. While MEG's popularity as an imaging method for brain-computer interfaces (BCIs) is currently inhibited by its cumbersome user experience, lack of portability, and considerable cost, steps towards developing more user-friendly wearable MEG are being undertaken [8].

A long history of research and relative affordability has seen electroencephalography (EEG) as the most widely applied neuroimaging tool used in information retrieval and human-computer interaction. Synchronized firing of areas of similarly oriented neurons produce changes in electrical potential that can be measured by electrodes on the scalp. By classifying temporal changes in the potential across the different electrodes, EEG-based BCIs detect associated mental activities like imagining hand movements and use this classification to control an interface [36].

Another type of EEG-based BCIs use psychophysiological knowledge of EEG activity synchronized towards known events, commonly referred to as the event-related potential (ERP), to make inferences on underlying mental events. Early research showed that ERPs respond to the degree that a stimulus reduces ambiguity, or in other words, how much information the stimulus delivers [61]. Later work showed that uncommon, relevant, attended stimuli reliably amplify a late, parietal part of the ERP [51], while the latency of this component responds to the degree the stimulus is processed [18], or *needs processing* [65]. Given its close relationship with information needs and cognitive processing, this P3 potential is clearly of particular importance with regards to detecting relevance.

The use of ERPs for inferring informational preference may have two distinct advantages over classic indicators. First, since ERPs inform both on explicit and implicit forms of cognition, they provide more comprehensive measurement for gauging preferences [33]. Classic ERP-based BCIs are known to improve their reliability by requiring users to mentally count relevant stimuli, thus amplifying P3s by increasing the reliance on explicit cognition [23]. EEG has been shown to be sensitive to implicit memory recall, showing activity related to forgotten memories [66]. Therefore, EEG measurements may provide more information than what is given by a user's physical interaction with a computing system.

Secondly, EEG systems provide more sensitive measurements for inferring preferences as they have been shown to predict graded relevance of stimuli as opposed to mere binary relevance. For example, while early research used simple term detection [21], later studies found ERPs can likewise indicate multiple relevance levels [50] and a deeper sense of topical relevance [28, 29]. Such information can also be used beyond simple tagging of relevant material or annotation towards recommendation systems, such as material matching the underlying interest of a user who is reading encyclopedic texts [20]. Further novel augmentations of EEG-based information retrieval include application of scalable crowd-sourcing platforms [15] and generative adversarial models that enable production as well retrieval of information [17, 30].

Despite these numerous applications in which EEG has been applied, the use of EEG in a collaborative filtering context has not been reported. Here, we report on findings that demonstrate the feasibility of a complete collaborative filtering pipeline in which user preferences are inferred directly from EEG responses.

3 NEUROPHYSIOLOGICAL EXPERIMENT

In this section, we report a neurophysiological experiment where images of artificially generated faces were presented to participants, their evoked brain responses were obtained via EEG, and the collected EEG data were associated with personal preference ratings provided by each participant. These data were used to train classifiers for each participant. The predicted outputs from these classifiers were then used as inputs for various collaborative filtering models.

3.1 Participants

Thirty-one participants were recruited from the University of Helsinki. All participants were fully informed of the nature of the study and signed a statement of informed consent to acknowledge understanding of their rights as participants in accordance with the Declaration of Helsinki, including their right to withdraw from the experiment at any time and for any reason. Complete data was obtained for 31 participants, 13 of which self-reported as female and 18 as male. The mean reported age was 28 years (SD = 7.14, range 18-45). All participants received compensation for their participation in the form of two vouchers to the local cinema.

3.2 Stimuli

A primary challenge in designing a subjective preference experiment was selecting stimuli that adequately reflect a real-world application and where preference assessments can be made by users in a reasonable amount of time. In other words, we needed to select stimuli such that many preference assessments could be made within the time frame of a typical EEG experiment (30 to 90 minutes, including setup time). Since humans have a natural and rapid preference reaction in response to seeing a face [42], and this reaction may naturally vary between high preference and low preference [48], we designed our experiment as a human facial attractiveness experiment, intended to reflect the visual assessments users commonly make on internet dating platforms.

In selecting facial stimuli, it was also important that participants did not recognize any of the stimuli, so that their personal assessment of each stimulus would be based on perceived preferences rather than a participant's confounding knowledge of the content appearing in the image. Using images of celebrities or other publicly-available images could introduce additional confounding factors that are difficult to account for. For example, viewing the face of a person one recognizes can produce a measurable brain response whether or not one finds the face personally attractive. Additionally, using relatively homogeneous stimuli (oval cutouts of faces placed on a grey background) reduces the likelihood of confounds related to high variation within the stimuli. Segments from films, musical pieces, or images of products may contain a high degree of internal variation which may produce neurophysiological effects unrelated to preference assessment (such as bright colors or distinct patterns). To avoid this, we used artificially created images of faces as stimuli. This allowed us to keep the stimuli relatively homogeneous and ensure that they did not depict any known individual.

Stimuli were generated using a GAN architecture trained from 30,000 images of celebrity faces[1], using a random process by sampling from 70,000 latent vectors from a 512-dimensional multivariate normal distribution [31]. To ensure there were no errors in the generated images, a human assessor manually screened each stimulus for potential artifacts (such as faces being unnaturally distorted or having jewelry for eyes). The human assessor did not screen the generated images for any other physical attributes. These images were then sorted into male- and female-appearing groups, and the first 120 images of each group were selected, resulting in a total of 240 unique images. These images were then resized to a resolution of 512 x 512 pixels, and an elliptic grey frame was applied to mask the background and non-facial features.

3.3 Apparatus

Stimuli were presented using an LCD display positioned approximately 60 cm from the participant, running at 60 Hz and with a resolution of 1680 x 1050 pixels. Psychology Software Tools E-Prime 3.0.3.60 stimulus presentation software was used to optimize the timing of the display and the EEG amplifier trigger control.

[1] https://github.com/tkarras/progressive_growing_of_gans

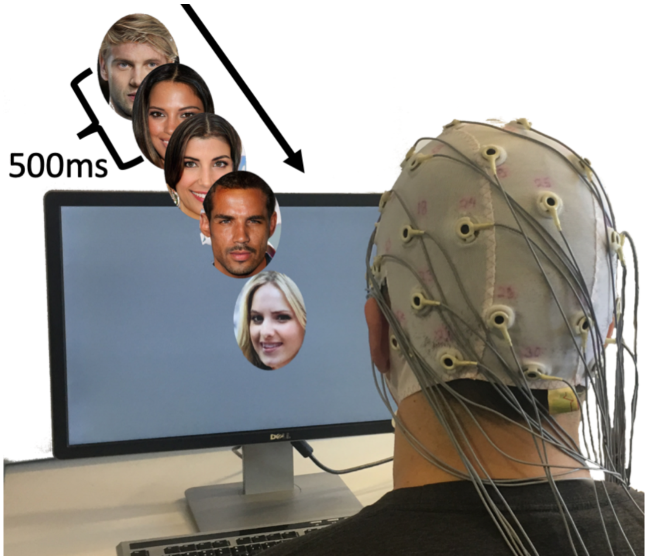

Figure 2: A visual representation of the RSVP task. A participant views stimuli images depicting faces on a computer screen, with a new stimulus presented to the participant every 500ms. The participant visually assesses each face to determine if they find it personally attractive while EEG data are collected. Aside from the collection of EEG data, the participant does not have to physically interact with the system during the RSVP task.

EEG data were recorded from 32 Ag/AgCl electrodes, positioned at equidistant locations of the 10-20 system. A QuickAmp USB (BrainProducts GmbH, Gilching, Germany) amplifier running at 2000 Hz was used to filter and digitize the data. The electro-oculogram (EOG), used for artifact removal during the data preprocessing step, was collected using two pairs of bipolar electrodes, placed 1 cm lateral to the outer canthi of the left and right eyes, and 2 cm inferior and superior to the right pupil.

3.4 Task

Participants were asked to view faces presented to them on a screen and visually assess each face for personally attractiveness. The experiment was divided into 3 blocks, with 8 rapid serial visual presentation (RSVP) trials per block. Each RSVP trial consisted of 60 images, for a total for 480 stimulus presentations per block. Prior to each RSVP trial, participants were instructed to observe the presented faces and make a mental note whenever they saw a face they found personally attractive. The RSVP then started, with 60 images presented at 2 Hz (500 ms stimulus onset asynchrony) against a grey background. After viewing the last stimulus, participants were asked by the software to give an estimation of how many times they had made a mental note. This number was then used in the confirmation task at the end of the block. During the confirmation task, participants were presented with smaller versions of all images presented during the RSVP trials in groups of 60; the participants were requested to click on the images which they had found attractive. The experiment took approximately 25 minutes, not including two self-timed breaks placed between the three blocks.

3.5 Data Cleaning and Labeling

To improve the signal-to-noise ratio while maintaining the paradigm of a real-world BCI application, EEG data were preprocessed using simple heuristics that can be applied in real-time and without human input [41]. First, the EEG data were digitized and referenced to the common average. Next, a band-pass filter was applied between 0.2 Hz and 35 Hz to remove line noise and slow fluctuations in the signal. Then, the data were time-locked to stimulus-onset and split into 1100 ms long segments (commonly referred to as "epochs") including 200 ms of baseline electrical activity. The average of this baseline activity was subtracted from the rest of the epoch. Finally, epochs containing artifacts (e.g. eye blinks) were excluded from the data set by applying individually tailored thresholds to the absolute maximum voltage. After preprocessing, approximately 12 percent of all epochs were removed. The final dataset consisted of 1265 epochs per participant (SD = 109), out of a maximum possible 1440.

In order to assign ground truth labels of attractiveness for the stimuli images, we used the answers provided by participants during the confirmation task. Thus, for each participant, epochs and their associated stimuli were assigned an integer rating from 0 to 3, based on the number of times that particular participant had reported the image as personally attractive during the confirmation tasks.

3.6 Neurophysiological Results

ERP plots were produced by averaging the voltages for epoched data across participants, grouped by individual ratings. A clear graded response was found prominently at the Fz and Pz electrodes (See Figure 3).

Averaged ERPs evoked by images were analyzed to estimate the effect of attractiveness. Visual inspection of the effect of attractiveness across the scalp suggested two main components. A somewhat earlier, more frontal component was found in particular over the Fz, while a later, more parietal component peaked over Pz. Two separate repeated measures ANOVAs were conducted to determine whether attractiveness had dissociable effects across these two sites. A first repeated measures ANOVA used the average amplitude between 250-350 as measure and attractiveness (0 vs 1 vs 2 vs 3) as factor. A significant effect of attractiveness was observed, F (3, 90) = 15.13, $p < 0.0001$, $\eta^2 = 0.34$. Attractiveness had a near linear effect of average amplitude from 0.23 to 1.22 μV, with attractiveness 1 and 2 in between (0.80 and 1.11 μV). Bonferroni corrected post-comparisons suggested almost all comparisons showed significant differences, ps < 0.005, apart from 0 vs 1, 0 vs 2, and 2 vs 3. A second repeated measures ANOVA used the average amplitude between 350-500 as measure and attractiveness (0 vs 1 vs 2 vs 3) as factor to test the later, parietal effect. Here, attractiveness also had a significant effect, F (3, 90) = 34.93, $p < 0.0001$, $\eta^2 = 0.54$, characterized by a monotonically increasing amplitude evoked the higher an image's rating. Post-hoc comparisons showed significant effects (ps<0.01) between all attractiveness rating combinations except for 1 vs 2.

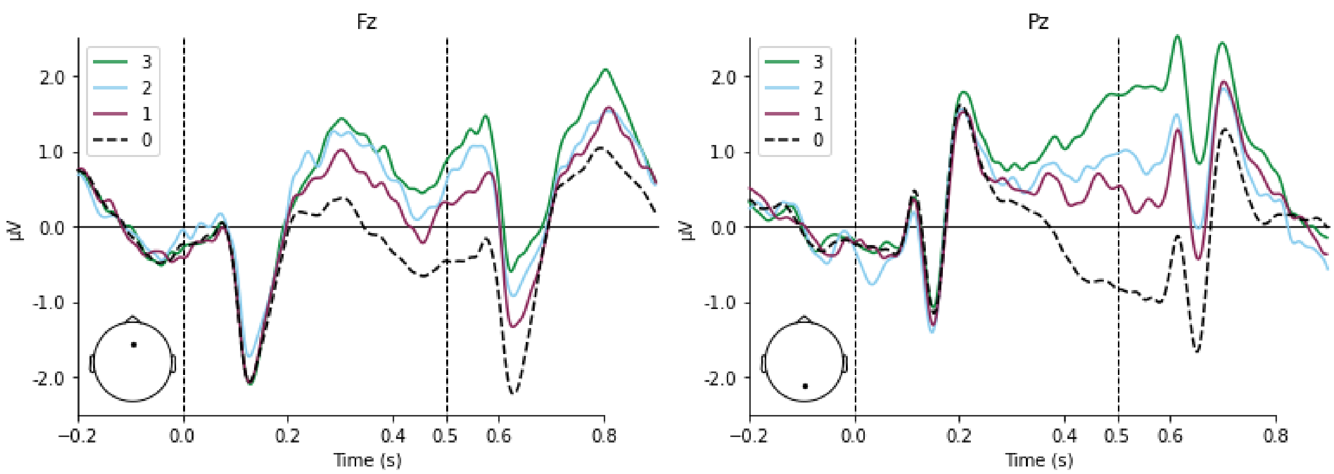

Figure 3: ERP plots for the Fz (left) and Pz (right) channels, grouped by class label and averaged across all participants. Stimuli onset are indicated with the dashed vertical line at 0ms and 500ms. Visual inspection of the ERPs indicates a strong late positivity correlated with preference. This effect is particularly noticeable at the Pz channel.

4 INFERRING PREFERENCES FROM BRAIN SIGNALS

4.1 Problem Formulation

Here we formalize the problem outlined in the preceding section as an information retrieval task. Given a set of brain signals from multiple users and known preference interactions, we learn a mapping $G : B, U \longrightarrow Y$ to estimate the preferences of a given user P for unseen items X. This formulation can be broken down into two sub-problems, formalized in detail below.

4.1.1 *Preference Estimation.*

First, we learn the associations between an individual user's brain signals and their expressed preferences for particular items. Given an individual user's brain responses and self-reported preferences, we learn a mapping between users' brain signals and their preferences in order to predict new preferences for content they have seen but for which preference information is not available.

Given a recorded brain signal B_i consisting of 32 channels of time series voltage data and an associated preference rating y_i for an item x_i, we learn a mapping $G : B \longrightarrow Y$ such that for a new brain signal $\hat{B}_x$, we can approximate a preference rating $\hat{y}_x$. Intuitively, this problem presents itself as a supervised classification task.

4.1.2 *User-Item Interaction.*

Next, we learn the relationships between users and items such that, given an item for which there are not any recorded brain signals for a particular user, we may still infer their preference responses using the available preferences inferred from other users.

Given predicted preferences $\hat{Y}$ of users U for items I, we learn a mapping between users and items $G : U, I \longrightarrow \hat{Y}$ such that given an item i unseen by user u, we can estimate their preference response $\hat{y}_{u_i}$. This problem can be generalized as a matrix factorization task, where we seek to learn interactions between users and items by estimating their latent representations.

4.2 Machine Learning Experiments

We model the brain signals with a neural architecture representing a combination of a neural EEG classifier [11] and neural collaborative filtering approach [25]. The architecture is depicted graphically in Figure 4.

To decode individual preferences using a neural architecture, we designed a multi-layer perceptron (MLP) that takes as inputs vectorized epochs and labeling information, and produces estimations of class probabilities as outputs. The MLP architecture consists of 3 densely connected layers of 256, 128, and 64 neurons respectively, using ReLU activations. Batch normalization and 25% dropout is applied between each layer. The final layer consists of four neurons corresponding to different preferences levels and a softmax activation. The MLP model was trained for 10 epochs using leave-one-out cross-validation for each participant, Adam optimization, and sparse categorical crossentropy as the loss function.

For the neural collaborative filtering step (NCF), we replicate the architecture presented in [25]. The NCF model was trained for 10 epochs using leave-one-out cross-validation for each participant, Adam optimization, and root mean squared error (RMSE) as the loss function.

4.3 Data Featurization

Recall that our epoch dataset contains time series data 1100ms in length, across 32 channels, and with integer labels from 0 to 3 for each epoch indicating how much a participant preferred its associated stimulus image. These epochs were shortened from 1100 millisecond (ms) long segments to omit the 200ms of pre-stimulus baseline activity, leaving epochs 900ms in length. From these, we resampled using a sliding window of 750ms in length, with a starting position of 0 ms to a maximum of 150ms in increments of 2ms, for a total of 75 resampled epochs produced from each original epoch.

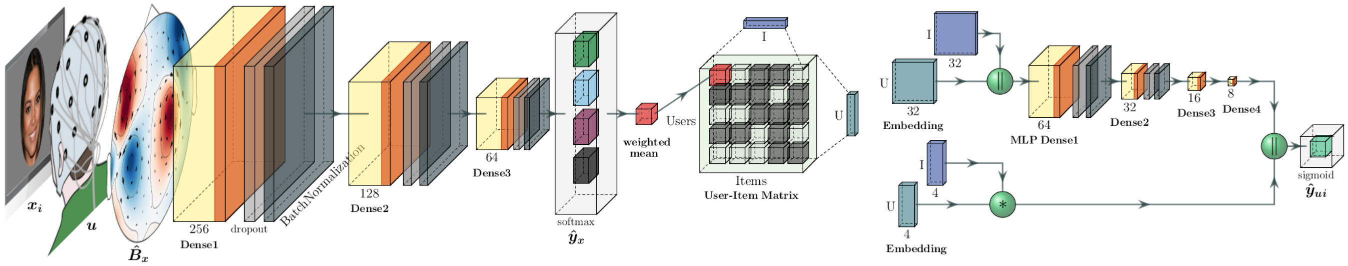

Figure 4: A diagram of the neural architecture. For all architectures, the machine learning and collaborative filtering are conducted in two phases. In the first phase, an individual classifier model associating a brain response B_x with an explicit preference rating y_x, is trained for participant u . Then, given a new stimulus x_i and brain response $\hat{B}_x$, the classifier estimates a preference rating $\hat{y}_x$. This process is repeated for all participants to predict preferences for many stimuli for which explicit ratings are not available. In the second phase, a collaborative filtering approach (in this example, neural collaborative filtering) is used to estimate preferences for missing user-item interactions.

For each individual participant, class imbalances were corrected using downsampling.

A vectorized representation of the EEG data was then constructed by splitting each epoch into 25 equidistant time windows and averaging the measured voltages for each time window and for each channel. This resulted in a data matrix $X^{n \times m \cdot t'}$, where $m = 32$ channels and $t' = 25$ time windows, for a total of 800 data points for each response n.

4.4 Prediction Setup and Control Conditions

In order to reveal the effect of brain input for collaborative filtering, we designed an experiment with controls for classification and collaborative filtering, rating data distributions, and explicit upper bound performance. It is noteworthy that the purpose of the experiment was *not* to conduct a conventional computational method comparison, but rather to study whether the brain input can be a reliable alternative to explicit feedback and exclude the possibility that the effect would be due to a specific choice of a computational architecture.

Control methods: We compared the neural architecture with two well-established approaches: A linear discriminant analysis (LDA) classifier step and singular value decomposition (SVD) collaborative filtering step, and LDA with k-nearest neighbors (kNN). LDA with shrinkage, which was selected automatically using the Ledoit-Wolf lemma [35], is found to be robust for single-trial EEG/ERP classification [7]. In this experiment, the LDA output is a probability estimation for each of the four possible classes. Results from the LDA classifiers were then converted to float values from 0 to 1 using a weighted arithmetic mean of the estimated probabilities and then aggregated across participants to create a user-item rating matrix.

With these data representations as inputs, personalized classifiers were trained separately for each participant using a leave-one-out strategy. These results were then combined to produce a single user-item rating matrix. Using this matrix, we run SVD and kNN, which represent well-known matrix factorization and nearest neighbor approaches of collaborative filtering [60]. We refer to these methods as LDA+SVD and LDA+kNN respectively.

In the experiments, kNN was run with max k=40 and min k=1, and SVD with 10 factors (we did not find that changing these parameters had any significant affect on performance). Matrix factorization using SVD was achieved by following the approaches outlined in [34] and [52]. Both kNN and SVD were achieved using implementations from the Surprise[2] library.

For all methods, test and control, the collaborative filtering step was conducted via a standard procedure where random entries of the user-item rating matrix were removed in various proportions to simulate different sparsity levels (25%. 50% and 75%), in addition to the dataset's natural sparsity of 10% from pre-processing and artifact removal. The ground-truth explicit ratings of removed entries were then predicted for using collaborative filtering.

Random performance control: As the data was obtained in a neurophysiological experiment specifically for the present study, we also wanted to control for potential bias in the data and included a random baseline in which a random rating is predicted based on the distribution of the training set using Maximum Likelihood Estimation – with the assumption that the training data is normally distributed. We also used label permutation of classifier outputs as a random baseline, and found it achieved similar results.

Upper-bound performance control: In addition, we computed an upper bound for all models by using explicit ratings, rather than ratings inferred from EEG data, as the collaborative filter input. This approach reveals the ideal upper-bound performance of the models for this dataset and allows for a better comparison of the performance of the EEG-derived preferences to the performance of explicitly reported ratings.

5 RESULTS

5.1 Classifier Models

For inferring preferences from brain signals, the MLP classifier achieved a mean RMSE of 0.41 (SD=0.05) across participants. The LDA classifier performed significantly better than the MLP classifier ($p<0.01$), with a mean RMSE of 0.35 (SD=0.06) across participants.

[2]https://github.com/NicolasHug/Surprise

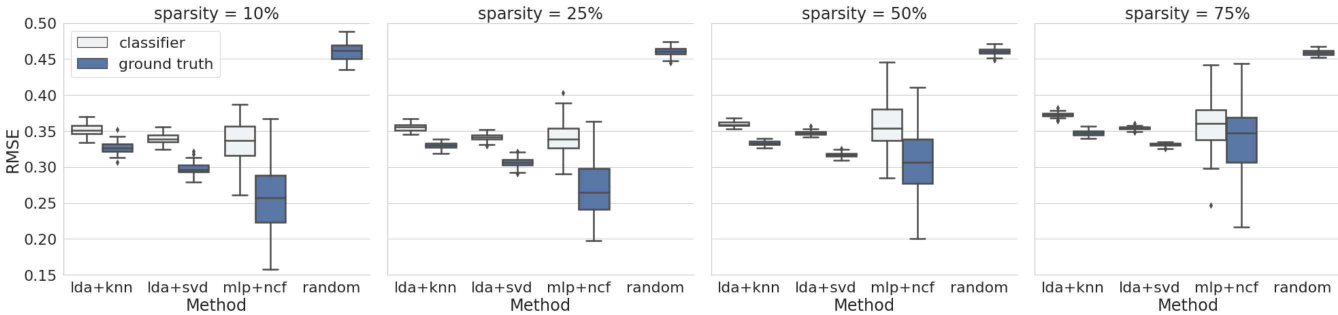

Figure 5: Comparison of RMSE values (lower is better) for various collaborative filtering methods, averaged across participants. The performance of random baselines for the various methods, which did not vary significantly, has been averaged into a single random baseline to keep the plots readable.

The range in RMSE across participants was similar for both methods, with the LDA classifier ranging from a minimum of 0.24 to a maximum of 0.50, and the MLP ranging from a minimum of 0.31 to a maximum of 0.52. Both methods performed significantly better ($p<0.001$) than random baselines (RMSE of 0.53, SD=0.06). Investigating per-participant performance, the LDA classifier was better than random for 28 out of 31 participants, while the MLP classifier was better than random for 27 out of 31 participants. Per-participant performance for the MLP classifier is shown in the top of Figure 6.

5.2 Collaborative Filtering Models

For all sparsity levels and input types (EEG classifier and ground truth labels), all three methods performed significantly better than random ($p<0.001$). Testing our collaborative filtering methods with leave-one-out cross validation at 10% sparsity, the MLP+NCF approach achieved a mean RMSE of 0.29 (SD=0.05), compared to 0.34 (SD=0.02) for LDA+SVD and 0.35 (SD=0.02) for LDA+KNN. The performance of all models was negatively correlated with sparsity. When comparing collaborative filtering methods across sparsity levels using the classifier outputs (see Figure 5), there were no statistically significant differences in mean performance. Statistically significant differences ($p<0.05$) were found when comparing performance between 10% and 75% sparsity ($p<0.05$) using ground-truth labels as collaborative filtering inputs.

Collaborative filtering results across participants, shown at the bottom of Figure 6), suggest that meaningful information is recovered. Not only does the predictive performance improve for most participants (29 out of 31, $p<0.01$) when compared to the MLP classifier, but for the 4 subjects where the MLP classifier could *not* produce statistically significant results better than random, collaborative filtering *does* successfully predict their responses better than the random baselines.

It is worth noting that the differences in performance related to sparsity may also be due to the relatively small number training examples available (240 items * 31 users = 7440 total data points at 0% sparsity). Thus, it cannot be ruled out that the varying results for the neural MLP+NCF approach are largely influenced by the number of training examples used.

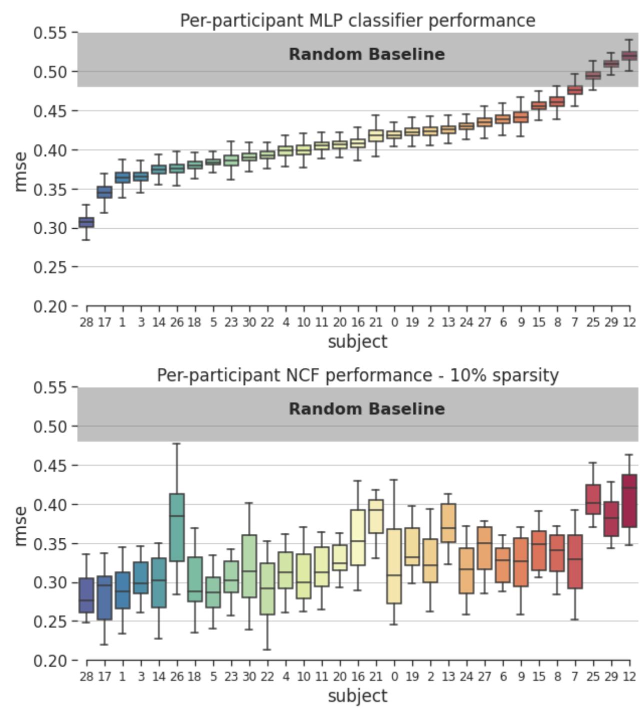

Figure 6: Comparison of RMSE values (lower is better) across participants for the MLP classifier (top), and the neural collaborative filter (bottom). MLP classifier performance was better than random ($p<0.01$) for 27 out of 31 participants, and neural collaborative filtering performance was better than random ($p<0.01$) for all participants.

6 DISCUSSION AND CONCLUSIONS

6.1 Summary of Contributions

In this work, we explored preferences inferred from human brain signals using collaborative filtering. We devised an approach for capturing preferences via EEG and evaluated the approach using two modeling instantiations — a neural architecture and a conventional combination of classification and matrix factorization. To this end, we set out to study two research questions that we answer below.

RQ1: Can brain responses be used to predict user preferences? Our results 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. Our results are also in line with previous findings that the evoked ERPs follow a pattern according to the graded level of the self-reported preferences [50].

RQ2: Can preferences inferred from the brain be used for collaborative filtering? Our results show the feasibility of brain input as an information source for collaborative filtering – thus demonstrating collaborative filtering where preferences are obtained directly from human brain responses evoked by digital content. While the neural architecture yielded the highest performance, all collaborative filtering architectures produced results close to the performance achieved by using the explicit ratings. Furthermore, all architectures for brain-based collaborative filtering achieved results significantly better than random baselines with large effect sizes.

6.2 Limitations

We evaluated the approach in a task of visual preferences toward human faces and demonstrated — for the first time — the feasibility of using brain signals for collaborative filtering. The models and experiments presented here demonstrate that brain signals are feasible sources of preference information, and can be used to complement or replace the conventional behavioral signals that are widely exploited by recommender systems. Nonetheless, it is not without its limitations.

Our approach is based on capturing event-related potentials via EEG evoked in response to perception of content. This limits the approach to information that can be matched with perception. However, the same limitation applies to any behavioral signal that is used as implicit feedback, as implicit signals also require pairing with the content.

While our approach theoretically relies upon user perception and brain responses alone, we nonetheless required the user to still provide some explicit ratings in order to properly train the EEG classification models and establish an upper bound for the performance. Ideally, explicit ratings could be omitted entirely and a completely unsupervised approach be used instead, however such an approach would still require some ground-truth labels in order to quantitatively assess performance.

We limited our dataset to faces and presented several machine learning architectures to infer and predict preferences from brain signals, however our approach is not restricted to the particular type of stimuli or computational method used in the present experiment. Rather, it is meant to serve as a general guideline for developing physiological and brain-computer interfacing to capture user preferences for recommendation. As the methodology 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.

6.3 Future Work

While our models show high performance on both preference estimation and collaborative filtering, we foresee some straightforward approaches to improve the model architectures. First, the time series nature and the known component structure of EEG/ERP data could be exploited by recurrent and attention-based neural architectures [64]. Moreover, rather than using outputs from a classifier model as the primary feature for the collaborative filtering step, a more exciting, albeit challenging approach would be to use the brain signals as input features without any supervision signal, and to instead recommend content based on similar cognitive profiles.

6.4 Ethical considerations

The opportunities of utilizing brain responses for recommender systems come with great risks for privacy and present unique ethical challenges, especially with regard to how data collected from BCIs may be used and abused. Brain-computer interfacing with online services may be particularly vulnerable to malicious third-parties and misuse or mishandling of sensitive information. For all intents and purposes, raw data collected via EEG should be treated as private information, as such data can potentially be used to identify certain physiological disorders and cognitive mental states [9, 19, 58]. EEG can also can be used to identify an individual [16, 45], so measures must be taken to ensure the data is properly anonymized. Future neuroimaging techniques may provide less noisy and higher resolution data from which greater amounts of information could be obtained.

The misuse of such data presents a highly disturbing vision of the future. From monitoring citizens and subliminally probing them for private information, influencing their political opinions via partisan selective exposure, to explicitly targeting users with neural profiles that suggest susceptibility to addiction and binge behavior, our community at large must carefully consider data privacy, ownership, and ethically sustainable utilization of these novel user signals.

ACKNOWLEDGMENTS

The research was partially funded by the Academy of Finland (Decision numbers: 322653, 328875, 336085). Computing resources were provided by the Finnish Grid and Cloud Infrastructure (persistent id: urn:nbn:fi:research-infras-2016072533).

REFERENCES

[1] Xavier Amatriain, Josep M Pujol, and Nuria Oliver. 2009. I like it... i like it not: Evaluating user ratings noise in recommender systems. In *International Conference on User Modeling, Adaptation, and Personalization*. Springer, 247–258.

[2] Ioannis Arapakis, Joemon M. Jose, and Philip D. Gray. 2008. Affective Feedback: An Investigation into the Role of Emotions in the Information Seeking Process. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (Singapore, Singapore) *(SIGIR '08)*. Association for Computing Machinery, New York, NY, USA, 395–402.

[3] Ioannis Arapakis, Yashar Moshfeghi, Hideo Joho, Reede Ren, David Hannah, and Joemon M Jose. 2009. Enriching user profiling with affective features for the improvement of a multimodal recommender system. In *Proceedings of the ACM International Conference on Image and Video Retrieval*. 1–8.

[4] Deger Ayata, Yusuf Yaslan, and Mustafa E Kamasak. 2018. Emotion based music recommendation system using wearable physiological sensors. *IEEE transactions on consumer electronics* 64, 2 (2018), 196–203.

[5] Oswald Barral, Ilkka Kosunen, Tuukka Ruotsalo, Michiel M Spapé, Manuel JA Eugster, Niklas Ravaja, Samuel Kaski, and Giulio Jacucci. 2016. Extracting relevance and affect information from physiological text annotation. *User Modeling and User-Adapted Interaction* 26, 5 (2016), 493–520.

[6] James Bennett, Stan Lanning, et al. 2007. The netflix prize. In *Proceedings of KDD cup and workshop*, Vol. 2007. New York, 35.

[7] Benjamin Blankertz, Steven Lemm, Matthias Treder, Stefan Haufe, and Klaus-Robert Müller. 2011. Single-trial analysis and classification of ERP components—a tutorial. *NeuroImage* 56, 2 (2011), 814–825.

[8] Elena Boto, Niall Holmes, James Leggett, Gillian Roberts, Vishal Shah, Sofie S Meyer, Leonardo Duque Muñoz, Karen J Mullinger, Tim M Tierney, Sven Bestmann, et al. 2018. Moving magnetoencephalography towards real-world applications with a wearable system. *Nature* 555, 7698 (2018), 657–661.

[9] S. G. Boyd, A. Harden, and M. A. Patton. 1988. The EEG in early diagnosis of the Angelman (Happy Puppet) syndrome. *European Journal of Pediatrics* 147, 5 (01 Jun 1988), 508–513.

[10] Wanyu Chen, Fei Cai, Honghui Chen, and Maarten De Rijke. 2019. Joint Neural Collaborative Filtering for Recommender Systems. *ACM Trans. Inf. Syst.* 37, 4, Article 39 (Aug. 2019), 30 pages.

[11] Alexander Craik, Yongtian He, and Jose L Contreras-Vidal. 2019. Deep learning for electroencephalogram (EEG) classification tasks: a review. *Journal of neural engineering* 16, 3 (2019), 031001.

[12] Maurizio Ferrari Dacrema, Paolo Cremonesi, and Dietmar Jannach. 2019. Are We Really Making Much Progress? A Worrying Analysis of Recent Neural Recommendation Approaches. In *Proceedings of the 13th ACM Conference on Recommender Systems* (Copenhagen, Denmark) *(RecSys '19)*. Association for Computing Machinery, New York, NY, USA, 101–109.

[13] Maurizio Ferrari Dacrema, Paolo Cremonesi, and Dietmar Jannach. 2019. Are we really making much progress? A worrying analysis of recent neural recommendation approaches. In *Proceedings of the 13th ACM Conference on Recommender Systems*. 101–109.

[14] Abhinandan S Das, Mayur Datar, Ashutosh Garg, and Shyam Rajaram. 2007. Google news personalization: scalable online collaborative filtering. In *Proceedings of the 16th international conference on World Wide Web*. 271–280.

[15] Keith M Davis III, Lauri Kangassalo, Michiel Spapé, and Tuukka Ruotsalo. 2020. Brainsourcing: Crowdsourcing Recognition Tasks via Collaborative Brain-Computer Interfacing. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–14.

[16] Luigi De Gennaro, Cristina Marzano, Fabiana Fratello, Fabio Moroni, Maria Concetta Pellicciari, Fabio Ferlazzo, Stefania Costa, Alessandro Couyoumdjian, Giuseppe Curcio, Emilia Sforza, Alain Malafosse, Luca A. Finelli, Patrizio Pasqualetti, Michele Ferrara, Mario Bertini, and Paolo Maria Rossini. 2008. The electroencephalographic fingerprint of sleep is genetically determined: A twin study. *Annals of Neurology* 64, 4 (2008), 455–460.

[17] Carlos de la Torre-Ortiz, Michiel M. Spapé, Lauri Kangassalo, and Tuukka Ruotsalo. 2020. Brain Relevance Feedback for Interactive Image Generation. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '20)*. Association for Computing Machinery, New York, NY, USA, 1060–1070.

[18] Connie C Duncan-Johnson. 1981. Young Psychophysiologist Award Address, 1980: P300 Latency: A New Metric of Information Processing. *Psychophysiology* 18, 3 (1981), 207–215.

[19] RJ Ellingson. 1954. The incidence of EEG abnormality among patients with mental disorders of apparently nonorganic origin: a critical review. *American Journal of Psychiatry* 111, 4 (1954), 263–275.

[20] Manuel JA Eugster, Tuukka Ruotsalo, Michiel M Spapé, Oswald Barral, Niklas Ravaja, Giulio Jacucci, and Samuel Kaski. 2016. Natural brain-information interfaces: Recommending information by relevance inferred from human brain signals. *Scientific Reports* 6 (2016), 38580.

[21] Manuel JA Eugster, Tuukka Ruotsalo, Michiel M Spapé, Ilkka Kosunen, Oswald Barral, Niklas Ravaja, Giulio Jacucci, and Samuel Kaski. 2014. Predicting term-relevance from brain signals. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*. 425–434.

[22] Wenqi Fan, Yao Ma, Qing Li, Yuan He, Eric Zhao, Jiliang Tang, and Dawei Yin. 2019. Graph neural networks for social recommendation. In *The World Wide Web Conference*. 417–426.

[23] Lawrence Ashley Farwell and Emanuel Donchin. 1988. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and clinical Neurophysiology* 70, 6 (1988), 510–523.

[24] David Goldberg, David Nichols, Brian M Oki, and Douglas Terry. 1992. Using collaborative filtering to weave an information tapestry. *Commun. ACM* 35, 12 (1992), 61–70.

[25] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. 2017. Neural collaborative filtering. In *Proceedings of the 26th international conference on world wide web*. 173–182.

[26] Thomas Hofmann. 2004. Latent semantic models for collaborative filtering. *ACM Transactions on Information Systems (TOIS)* 22, 1 (2004), 89–115.

[27] Thorsten Joachims, Laura Granka, Bing Pan, Helene Hembrooke, and Geri Gay. 2005. Accurately interpreting clickthrough data as implicit feedback. In *28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2005*. 154–161.

[28] Lauri Kangassalo, Michiel Spapé, Giulio Jacucci, and Tuukka Ruotsalo. 2019. Why do users issue good queries? neural correlates of term specificity. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 375–384.

[29] Lauri Kangassalo, Michiel Spapé, Niklas Ravaja, and Tuukka Ruotsalo. 2020. information gain modulates brain activity evoked by reading. *Scientific reports* 10, 1 (2020), 1–10.

[30] Lauri Kangassalo, Michiel Spapé, and Tuukka Ruotsalo. 2020. Neuroadaptive modelling for generating images matching perceptual categories. *Scientific Reports* 10, 1 (2020), 1–10.

[31] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. 2017. Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196* (2017).

[32] Jukka-Pekka Kauppi, Melih Kandemir, Veli-Matti Saarinen, Lotta Hirvenkari, Lauri Parkkonen, Arto Klami, Riitta Hari, and Samuel Kaski. 2015. Towards brain-activity-controlled information retrieval: Decoding image relevance from MEG signals. *NeuroImage* 112 (2015), 288–298.

[33] Sander Koelstra and Ioannis Patras. 2013. Fusion of facial expressions and EEG for implicit affective tagging. *Image and Vision Computing* 31, 2 (2013), 164–174.

[34] Yehuda Koren, Robert Bell, and Chris Volinsky. 2009. Matrix factorization techniques for recommender systems. *Computer* 42, 8 (2009), 30–37.

[35] Olivier Ledoit and Michael Wolf. 2004. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of multivariate analysis* 88, 2 (2004), 365–411.

[36] Steven Lemm, Christin Schafer, and Gabriel Curio. 2004. BCI competition 2003-data set III: probabilistic modeling of sensorimotor/spl mu/rhythms for classification of imaginary hand movements. *IEEE Transactions on Biomedical Engineering* 51, 6 (2004), 1077–1080.

[37] Greg Linden, Brent Smith, and Jeremy York. 2003. Amazon. com recommendations: Item-to-item collaborative filtering. *IEEE Internet computing* 7, 1 (2003), 76–80.

[38] Guang Ling, Michael R. Lyu, and Irwin King. 2014. Ratings Meet Reviews, a Combined Approach to Recommend. In *Proceedings of the 8th ACM Conference on Recommender Systems* (Foster City, Silicon Valley, California, USA) *(RecSys '14)*. Association for Computing Machinery, New York, NY, USA, 105–112.

[39] Jiahui Liu, Peter Dolan, and Elin Rønby Pedersen. 2010. Personalized News Recommendation Based on Click Behavior. In *Proceedings of the 15th International Conference on Intelligent User Interfaces* (Hong Kong, China) *(IUI '10)*. Association for Computing Machinery, New York, NY, USA, 31–40.

[40] Nathan N Liu, Evan W Xiang, Min Zhao, and Qiang Yang. 2010. Unifying explicit and implicit feedback for collaborative filtering. In *Proceedings of the 19th ACM international conference on Information and knowledge management*. 1445–1448.

[41] Steven J Luck. 2014. *An introduction to the event-related potential technique*. MIT press.

[42] Qiuling Luo, Bruno Rossion, and Milena Dzhelyova. 2019. A robust implicit measure of facial attractiveness discrimination. *Social cognitive and affective neuroscience* 14, 7 (2019), 737–746.

[43] Judith Masthoff and Albert Gatt. 2006. In pursuit of satisfaction and the prevention of embarrassment: affective state in group recommender systems. *User Modeling and User-Adapted Interaction* 16, 3-4 (2006), 281–319.

[44] Bradley N Miller, John T Ried, and Joseph A Konstan. 2003. GroupLens for Usenet: Experiences in applying collaborative filtering to a social information system. In *From Usenet to CoWebs*. Springer, 206–231.

[45] Chisei Miyamoto, Sadanao Baba, and Isao Nakanishi. 2009. Biometric person authentication using new spectral features of electroencephalogram (EEG). In *2008 international symposium on intelligent signal processing and communications*

systems. IEEE, 1–4.

[46] Yashar Moshfeghi, Peter Triantafillou, and Frank E Pollick. 2016. Understanding information need: An fMRI study. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*. 335–344.

[47] Douglas W Oard, Jinmook Kim, et al. 1998. Implicit feedback for recommender systems. In *Proceedings of the AAAI workshop on recommender systems*, Vol. 83. WoUongong.

[48] Ingrid R Olson and Christy Marshuetz. 2005. Facial attractiveness is appraised in a glance. *Emotion* 5, 4 (2005), 498.

[49] Evan M Peck, Daniel Afergan, and Robert JK Jacob. 2013. Investigation of fNIRS brain sensing as input to information filtering systems. In *Proceedings of the 4th Augmented Human International Conference*. 142–149.

[50] Zuzana Pinkosova, William J McGeown, and Yashar Moshfeghi. 2020. The cortical activity of graded relevance. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 299–308.

[51] John Polich. 2007. Updating P300: an integrative theory of P3a and P3b. *Clinical neurophysiology* 118, 10 (2007), 2128–2148.

[52] Francesco Ricci, Lior Rokach, and Bracha Shapira. 2011. Introduction to recommender systems handbook. In *Recommender systems handbook*. Springer.

[53] Tuukka Ruotsalo, Giulio Jacucci, Petri Myllymäki, and Samuel Kaski. 2014. Interactive intent modeling: Information discovery beyond search. *Commun. ACM* 58, 1 (2014), 86–92.

[54] Tuukka Ruotsalo, Jaakko Peltonen, Manuel JA Eugster, Dorota Głowacka, Patrik Floréen, Petri Myllymäki, Giulio Jacucci, and Samuel Kaski. 2018. Interactive intent modeling for exploratory search. *ACM Transactions on Information Systems (TOIS)* 36, 4 (2018), 1–46.

[55] Ruslan Salakhutdinov, Andriy Mnih, and Geoffrey Hinton. 2007. Restricted Boltzmann machines for collaborative filtering. In *Proceedings of the 24th international conference on Machine learning*. 791–798.

[56] J Ben Schafer, Dan Frankowski, Jon Herlocker, and Shilad Sen. 2007. Collaborative filtering recommender systems. In *The adaptive web*. Springer, 291–324.

[57] Suvash Sedhain, Aditya Krishna Menon, Scott Sanner, and Lexing Xie. 2015. Autorec: Autoencoders meet collaborative filtering. In *Proceedings of the 24th International Conference on World Wide Web*. 111–112.

[58] SJM Smith. 2005. EEG in the diagnosis, classification, and management of patients with epilepsy. *Journal of Neurology, Neurosurgery & Psychiatry* 76, suppl 2 (2005), ii2–ii7.

[59] Yading Song, Simon Dixon, and Marcus Pearce. 2012. A survey of music recommendation systems and future perspectives. In *9th International Symposium on Computer Music Modeling and Retrieval*, Vol. 4. 395–410.

[60] Xiaoyuan Su and Taghi M Khoshgoftaar. 2009. A survey of collaborative filtering techniques. *Advances in artificial intelligence* 2009 (2009).

[61] Samuel Sutton, Patricia Tueting, Joseph Zubin, and E. Roy John. 1967. Information delivery and the sensory evoked potential. *Science* 155, 3768 (1967), 1436–1439.

[62] Marko Tkalčič, Urban Burnik, and Andrej Košir. 2010. Using affective parameters in a content-based recommender system for images. *User Modeling and User-Adapted Interaction* 20, 4 (2010), 279–311.

[63] Marko Tkalcic, Ante Odic, Andrej Kosir, and Jurij Tasic. 2012. Affective labeling in a content-based recommender system for images. *IEEE transactions on multimedia* 15, 2 (2012), 391–400.

[64] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in neural information processing systems*. 5998–6008.

[65] Rolf Verleger. 1997. On the utility of P3 latency as an index of mental chronometry. *Psychophysiology* 34, 2 (1997), 131–156.

[66] Joel L Voss and Ken A Paller. 2008. Brain substrates of implicit and explicit memory: The importance of concurrently acquired neural signals of both memory types. *Neuropsychologia* 46, 13 (2008), 3021–3029.

[67] Jun Wang, Arjen P De Vries, and Marcel JT Reinders. 2006. Unifying user-based and item-based collaborative filtering approaches by similarity fusion. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*. 501–508.

[68] Xiang Wang, Xiangnan He, Meng Wang, Fuli Feng, and Tat-Seng Chua. 2019. Neural graph collaborative filtering. In *Proceedings of the 42nd international ACM SIGIR conference on Research and development in Information Retrieval*. 165–174.

[69] Hong-Jian Xue, Xinyu Dai, Jianbing Zhang, Shujian Huang, and Jiajun Chen. 2017. Deep Matrix Factorization Models for Recommender Systems.. In *IJCAI*, Vol. 17. Melbourne, Australia, 3203–3209.

[70] Xing Yi, Liangjie Hong, Erheng Zhong, Nanthan Nan Liu, and Suju Rajan. 2014. Beyond Clicks: Dwell Time for Personalization. In *Proceedings of the 8th ACM Conference on Recommender Systems* (Foster City, Silicon Valley, California, USA) *(RecSys '14)*. Association for Computing Machinery, New York, NY, USA, 113–120.

[71] Shuai Zhang, Lina Yao, Aixin Sun, and Yi Tay. 2019. Deep learning based recommender system: A survey and new perspectives. *ACM Computing Surveys (CSUR)* 52, 1 (2019), 1–38.