Brain Topography Adaptive Network for Satisfaction Modeling in Interactive Information Access System

Ziyi Ye
yeziyi1998@gmail.com
BNRist,DCST,Tsinghua University
Beijing, China

Xiaohui Xie
xiexh_thu@163.com
BNRist,DCST,Tsinghua University
Beijing, China

Yiqun Liu∗
yiqunliu@tsinghua.edu.cn
BNRist,DCST,Tsinghua University
Beijing, China

Zhihong Wang
wangzhh629@mail.tsinghua.edu.cn
BNRist,DCST,Tsinghua University
Beijing, China

Xuesong Chen
chenxuesong1128@163.com
BNRist,DCST,Tsinghua University
Beijing, China

Min Zhang
z-m@tsinghua.edu.cn
BNRist,DCST,Tsinghua University
Beijing, China

Shaoping Ma
msp@tsinghua.edu.cn
BNRist,DCST,Tsinghua University
Beijing, China

ABSTRACT

With the growth of information on the Web, most users heavily rely on information access systems (e.g., search engines, recommender systems, etc.) in their daily lives. During this procedure, modeling users' satisfaction status plays an essential part in improving their experiences with the systems. In this paper, we aim to explore the benefits of using Electroencephalography (EEG) signals for satisfaction modeling in interactive information access system design. Different from existing EEG classification tasks, the arisen of satisfaction involves multiple brain functions, such as arousal, prototypicality, and appraisals, which are related to different brain topographical areas. Thus modeling user satisfaction raises great challenges to existing solutions. To address this challenge, we propose BTA, a **B**rain **T**opography **A**daptive network with a multi-centrality encoding module and a spatial attention mechanism module to capture cognitive connectives in different spatial distances. We explore the effectiveness of BTA for satisfaction modeling in two popular information access scenarios, i.e., search and recommendation. Extensive experiments on two real-world datasets verify the effectiveness of introducing brain topography adaptive strategy in satisfaction modeling. Furthermore, we also conduct search result re-ranking task and video rating prediction task based on the satisfaction inferred from brain signals on search and recommendation scenarios, respectively. Experimental results show that brain signals extracted with BTA help improve the performance of interactive information access systems significantly.

CCS CONCEPTS

• **Human-centered computing** → **HCI design and evaluation methods**; • **Computing methodologies** → **Artificial intelligence**; • **Information systems** → **Users and interactive retrieval**.

KEYWORDS

Satisfaction Modeling, EEG signal processing, Search, Recommendation

1 INTRODUCTION

Web-based information access systems (e.g., search engines and recommender systems) heavily rely on implicit feedback (e.g., clicks, dwell time, and eye-tracking) to improve user experience [12, 54]. However, implicit feedback acts as an indirect probe of user feelings and thus is sometimes biased and misleading [3, 32, 52]. Therefore, advancement is still required with novel user signals to model the information access process. In recent years, the developments of electroencephalogram (EEG) devices make it feasible to collect brain signals in almost real-time. EEG directly captures brain activities, which can potentially reveal the true underlying user satisfaction while users are accessing information on the Web. Based on the estimated user satisfaction, the user's intent or preference can be better understood and successfully used to improve user experience.

However, to the best of our knowledge, few studies have thoroughly investigated the methods of utilizing brain signals for satisfaction modeling in the information access procedure. Little is known about to what extent search and recommendation system can be benefited by brain signals. In Figure 1, we provide a possible interactive system that re-ranks the search results upon the

real-time satisfaction modeling. A user examines two search results corresponding to the query "Jaguar", and our system estimates the user satisfaction with brain signals. Based on the satisfaction modeling, we reestimate the user intent and find that the user intent is related to the "Jaguar animal". Then unseen search results (i.e., search results on the next screen/page) are re-ranked, and the ones with the subtopic of "Jaguar animal" are placed higher than those of "Jaguar browser" and "Jaguar car".

Different from existing EEG classification tasks, modeling user satisfaction is challenging since satisfaction involve both cognitive processes (e.g., prototypicality [33] and appraisals [46]) and affective states (e.g., arousal [5]). This combined process starts at the occipital region and engages frontal-parietal attentional circuits, which involve brain functions in dorsolateral frontal region and medial temporal region [6, 37]. To connect and aggregate the signals in various brain regions, we believe capturing the topographical relations of brain signals is necessary in satisfaction modeling.
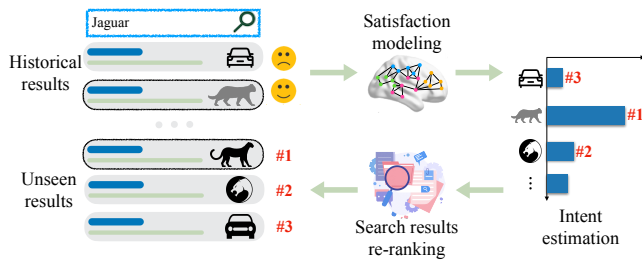

Figure 1: An example of EEG-enhanced search engine. By modeling user satisfaction, we interactively predict user intent and provide potentially more satisfactory results.

Existing studies have achieved many compelling outcomes by introducing the topographical information into EEG classification models. Among these models, convolutional neural network (CNN) [25, 28] or graph convolutional network (GCN) [20, 29, 47, 61] are most widely used. They have achieved high classification performance in motion imaginary [28] and emotion recognition tasks [29]. However, most of them fail to capture the adaptive topography connections. For example, EEGNet [28] utilizes convolutions in local brain regions yet ignores the ulterior topographical relations. DGCNN [47] and RGNN [61] apply a public adjacency matrix for graph convolutions, and thus the topography connections are fixed among different data samples. Since satisfaction is associated with several cognitive processes and affective states, the topographical relations of brain signals is changing and sophisticated. Thus it is challenging yet valuable to capture the topographical relations in a data-dependent way.

To tackle above challenges, we design a novel architecture for EEG-based satisfaction modeling named Brain Topography Adaptive network (BTA). The network applies a multi-centrality encoding module to generate encodings with 3D topographical information. Then, motivated by the success of the attention mechanism [50], we adopt a spatial attention module to capture the brain's cognitive connectivity in a data-dependent manner. The effectiveness of BTA is verified in two typical interactive information access scenarios, i.e., search and recommendation. Experimental results

in two public datasets (i.e., one for search and one for recommendation) demonstrate that BTA outperforms various baselines in terms of satisfaction estimation. Furthermore, we explore to what extent we can utilize the satisfaction estimated with brain signals to improve search result re-ranking and video rating prediction performance for search and recommendation scenarios, respectively. We demonstrate that the performance can be significantly improved with the satisfaction inferred from brain signals.

In summary, our contributions are three-fold. 1) We identify and tackle a novel problem of satisfaction modeling for interactive information access systems using brain signals. 2) We propose the Brain Topography Adaptive Network (BTA), which adaptively captures the topographical information to solve the EEG-based satisfaction prediction problem. Empirical experiments show that our method outperforms state-of-the-art EEG classification models. 3) We explore the possible advancements in interactive information systems brought by the estimated user satisfaction inferred from brain signals. As far as we know, this is the first time EEG-based satisfaction modeling is successfully applied to interactive search (i.e., search result re-ranking) and recommendation (i.e., video rating prediction) tasks.

2 RELATED WORK

2.1 User Satisfaction in Information Access

Satisfaction measures users' subjective feelings about the system, which can be treated as the fulfillment of their information requirement [22]. It has been noticed that modeling user satisfaction is valuable for performance improvement and evaluation in information systems [2, 31].

Researchers have focused on modeling users' satisfaction with users' implicit feedback signals (click, dwell time, scroll, etc.) [12, 16]. Recent years have witnessed much research introducing novel additional user signals and corresponding strategies to estimate user satisfaction, e.g., mouse movements [8] and eye-tracking [54]. However, implicit feedback signals are just indirect probes of real user satisfaction and thus are often incorrect [3]. In search scenarios, Liu et al. [32] reveal that a large proportion (45.8%) of eye fixations are irrelevant to relevance estimation. In news recommendation, Wang et al. [52] find that users' click behaviors can not be simply treated as positive signals due to the "clickbait" effect.

In this study, we explore satisfaction modeling with brain signals with our proposed BTA. Besides, we conduct search and recommendation tasks with the satisfaction inferred from brain signals. We aim to demonstrate the benefits of brain signals as effective user feedback for designing an interactive information access system.

2.2 Neuroscience & Information Access System

Recent research has applied brain imaging tools to study aspects of the information access process from a neuroscience perspective. For example, Moshfeghi et al. [35, 36] conduct a series of studies to unravel the neural basis of information need. Apart from neurological-based analysis, researchers also utilize brain signals to predict user judgments in the information access process. For example, Gwizdka et al. [15], Kim and Kim [23] conduct extensive studies to judge text relevance and topical relevance of visual shots using EEG signals. Davis III et al. [10] conduct an interesting study

to estimate facial preference with brain signals and then utilizes the estimated preference for collaborative filtering. However, few studies have thoroughly investigated the satisfaction modeling in information systems and explored to what extent the system can be improved with it.

What we add on top of prior works is that we design special EEG classification models and conduct interactive information access tasks to demonstrate the benefits of introducing brain signals into user satisfaction modeling.

2.3 EEG-based Classification

EEG has the advantages of high temporal resolution, non-invasiveness, and relatively low financial cost. It is widely used in researches involving motion imaginary [25, 28], emotion recognition [20, 29, 61], and sleep stage scoring [21, 48]. Recently trends of EEG-based classification have been developed substantially from topology-invariant algorithms (e.g., support vector machines (SVM) [53], Gradient Boosting Decision Tree (DT) [14], and Recurrent Neural Network (RNN) [30]) to topology-aware methods such as CNN and GCN. The topology-aware models take the topological structure of EEG features into account when learning the representations. For example, Lawhern et al. [28] and Kostas et al. [25] map the 3D EEG topographical information into 2D representations and adopt CNN-based architecture to aggregate channel informations. Additionally, there are some studies [47, 59, 61] applying GCN-based methods, and a public adjacency matrix is adopted to automatically learn the aggregation weight between EEG channels. And Li et al. [29] exploit multi-domain information to build the adjacency matrix for their GCN-based model.

However, as satisfaction is related to various cognitive processes and affective states [5, 33, 46], the brain connectivities might be different due to different stimulus factors [17, 57]. Few studies adaptively learn the brain connectivities in a completely data-dependent way. One exception is the HetEmotionNet [20], which utilizes mutual information [26] to construct the adjacency matrix for each data sample and then applies the graph convolution with it. However, the mutual information indicates the similarity of EEG features. Simply aggregating EEG channels with similar features is not always reasonable. For example, previous studies have suggested that the asymmetry in neuronal activities between the left and right hemispheres is informative [45]. Our study differs from previous work in using a completely data-dependent aggregation strategy with the specially designed multi-centrality encoding and spatial attention mechanism.

3 EEG-BASED SATISFACTION MODELING

3.1 Problem statement

In this paper, we define the temporal features of EEG signals as $X^t = \{x_1^t, x_2^t, ..., x_N^t\} \in \mathbb{R}^{N \times E}$, where N is the length of temporal features, E is the number of EEG channels. The spectral features of EEG signals are denoted as $X^s = \{x_1^s, x_2^s, ..., x_B^s\} \in \mathbb{R}^{B \times E}$, where B is the length of spectral features, E is the number of EEG channels. The spectral features are the differential entropy (DE) [11] extracted from B frequency bands (e.g., $\delta, \theta, \alpha, \beta$, and γ). Then the EEG-based satisfaction modeling problem is to learn a mapping function F_s
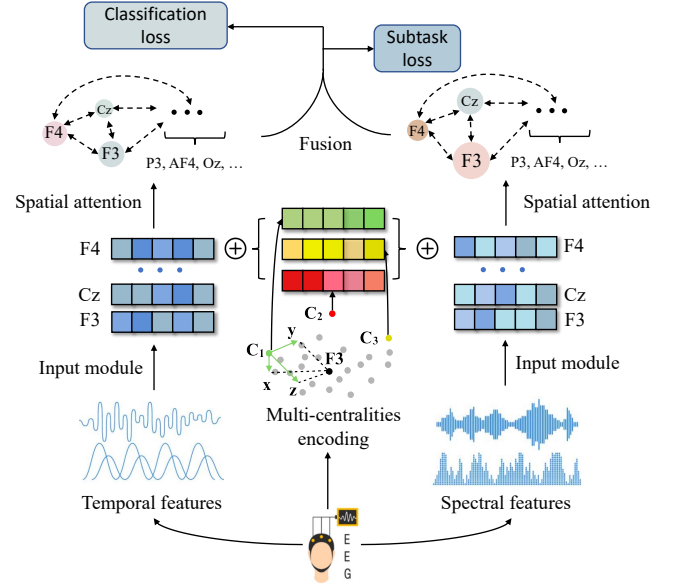

Figure 2: The architecture of proposed BTA.

based on the proposed model, which can be formulated as:

$$\hat{y} = F_s(X^t, X^s)$$

where $\hat{y}$ denotes the estimated satisfaction score. With the predicted satisfaction score on historical items, information system can interactively provide more satisfying unseen items. In this paper, we explore two common information access scenarios, i.e., search and recommendation, which are detailed in Section 4.

3.2 Brain Topography Adaptive Network

3.2.1 Model overview. An overview of the proposed satisfaction modeling method is illustrated in Figure 2. We propose a Brain Topography Adaptive Network (BTA), which consists of a temporal data stream and a spectral data stream, and these two data streams have similar structures. The model is composed of five components: (1) *Input encoding module.* (2) *Multi-centrality encoding module.* (3) *Spatial attention module.* (4) *Fusion and classification module.* (5) *Subtask module.* Firstly, the spectral features or temporal features are linearly projected into a latent space with the *input module.* Then, the *multi-centrality encoding module* generates a spatial encoding for each channel according to its topographical relation to various spatial centralities. After that, the *spatial attention module* aggregates the channel information with attention mechanisms to capture the brain's cognitive connectivity. Finally, the *fusion and classification module* fuses the temporal data stream and spectral data stream to classify user satisfaction in information access tasks. Additionally, to initialize the centrality embedding vectors, we replace the *fusion and classification module* with a *subtask module* before the model training.

3.2.2 Input encoding module. The spectral features or temporal features are linearly projected into a latent space, which can be

formulated as:

$$H^t = W^t * X^t + B^t, H^s = W^s * X^s + B^s$$

where $X^t \in \mathbb{R}^{E \times N}, X^s \in \mathbb{R}^{B \times N}$ are temporal and spectral EEG features, respectively, $W^t \in \mathbb{R}^{H \times N}, W^s \in \mathbb{R}^{H \times B}, B^t \in \mathbb{R}^{H \times E}, B^s \in \mathbb{R}^{H \times E}$ are learnable parameters, $H^t \in \mathbb{R}^{H \times E}, H^s \in \mathbb{R}^{H \times E}$ are the input vectors. We transform X^t and X^s into latent spaces with the same dimension since the multi-centrality encoding module, which adds topographical information to the input vectors (see in Section 3.2.3), is shared between them.

For simplicity, in Section 3.2.3 and Section 3.2.4, we omit the subscripts t and s (e.g., $H \in \{H^t, H^s\}$) since the temporal and spectral data stream share the same architectures.

3.2.3 Multi-centrality encoding module. To map each channel with a meaningful spatial encoding, we adopt multi-centrality encoding as an additional signal to the neural network. Specifically, we select M spatial points $C_1, C_2, ..., C_M$ in the brain topography space as the centralities. For each centrality C_j, we build a spherical coordinate system $\mathscr{F}_j$ with C_j being regraded as the origin. And the zenith direction and the azimuth direction of $\mathscr{F}_j$ are defined as straight above and ahead of the human brain, respectively. Then, we develop a centrality encoding $p_{i,j}$ which assigns each EEG channel E_i three learnable embedding vectors according to its spherical coordinate in $\mathscr{F}_j$. The centrality encoding $p_{i,j} \in \mathbb{R}^H$ of a EEG channel E_i according to a centrality C_j is formulated as:

$$p_{i,j} = \rho_{i,j} * c_{j,\rho} + \theta_{i,j} * c_{j,\theta} + \varphi_{i,j} * c_{j,\varphi}$$

where $(\rho_{i,j}, \theta_{i,j}, \varphi_{i,j}) \in \mathbb{R}^3$ is the spherical coordinate of E_i in $\mathscr{F}_j$, $(c_{j,\rho}, c_{j,\theta}, c_{j,\varphi}) \in \mathbb{R}^{3 \times H}$ are the centrality embedding vectors related to C_j. To enrich the spatial representation of EEG channels and increase their distinction, we obtain M centrality encodings for an EEG channel E_i, i.e., $p_{i,j}, j \in \{1, ..., M\}$. The centrality encodings are then combined with the input vector by:

$$z_i = h_i \oplus \sum_{1 \le j \le M} p_{i,j}$$

where $h_i \in \mathbb{R}^H$ is the subvector of $H^\top = \{h_1, ..., h_e\}$, $H^\top$ is the transposed matrix of $H \in \{H^t, H^s\}$, $\oplus$ is an element-wise operator. Then the input vector after adding the multi-centrality encoding can be expressed as $Z = \{z_1, ..., z_e\}^\top \in \mathbb{R}^{E \times H}$. We adopt a direct addition as $\oplus$ since it is simple to implement and adds minimal overhead to training time and model size. We also tried more complex interaction operators, but the results are similar, so we omit them. By using the multi-centrality encoding in the input, the input encoding vectors obtain additional topographical information. Therefore, the model can capture both the topographical correlation and the spectral or temporal information.

3.2.4 Spatial attention module. The brain's cognitive connectivity between EEG channels should not be neglected in EEG-based classification tasks. Therefore, we adopt attention mechanisms to capture the adaptive channel correlations for each data sample. We first apply a multihead attention layer to calculate the interacted sequence:

$$Z_1 = MultiHead(Z^\top, Z^\top, Z^\top)$$
$$= Concat(head_1, head_2, ..., head_D)W^O$$

where $head_i = Attention(Z^\top W^Q, Z^\top W^K, Z^\top W^V), W^Q, W^K, W^V \in \mathbb{R}^{H \times H/D}$ are trainable parameters, D is the number of heads, and $Attention(Q, K, V)$ is the scaled dot-product attention mechanism [50]. $Concat(\cdot)$ denotes the concatenation operation in the second dimension and the concatenated vector is fed into a linear matrix $W^O \in \mathbb{R}^{H \times H}$ to obtain the spatial interacted vectors $Z_1 \in \mathbb{R}^{E \times H}$.

Next, we apply a batch normalization layer BN to accelerate the training procedure and obtain the output vector $Z_2 = BN(Z_1) \in \mathbb{R}^{E \times H}$. In spite of using the layer normalization as Transformer [50] does, we adopt batch normalization since it performs better than layer normalization. We suggest that using batch normalization can mitigate the effect of instability and outline values related to EEG signals [58].

3.2.5 Fusion and classification module. With the above modules, the input temporal features X_t and spectral features X_s are transformed into output vectors Z_2^t and Z_2^s, respectively. We concatenate the output vectors to obtain $Z_3 = Concat(z_{2,1}, ..., z_{2,e}) \in \mathbb{R}^{EH}$, where $Z_2 = \{z_{2,1}, ..., z_{2,e}\} \in \mathbb{R}^{E \times H}$, $Z_2 \in \{Z_2^t, Z_2^s\}$, and $Z_3 \in \{Z_3^t, Z_3^s\}$. Then to fuse Z_3^t and Z_3^s, we adopt a fully connected layer, an activation function, and a softmax function, which can be formulated as:

$$\hat{y} = softmax(W\sigma(Concat(Z_3^t, Z_3^s)) + B)$$

where σ denotes as activation function (*Gelu* in our experiments), $W \in \mathbb{R}^{EH \times 1}$ and $B \in \mathbb{R}^1$ are trainable parameters, $\hat{y}$ is the estimated satisfaction score. Finally, the classification cross entropy is used as the loss function, which is defined as follows:

$$L = -y \lg(\hat{y}) - (1 - y) \lg(1 - \hat{y})$$

where y is the true label.

3.2.6 Subtask module. Besides, to initialize the centrality embedding vectors $c_{j,\rho}, c_{j,\theta}$, and $c_{j,\varphi}$ for each j in $\{1, 2, ...M\}$, we adopt an unsupervised subtask prior to the satisfaction modeling task. The unsupervised subtask is a reconstruction task to predict some mask EEG features, which share the same architecture except for the classification layer. Specially, randomized binary noise masks $W_{t,mask} \in \mathbb{R}^{H \times N}$ and $W_{s,mask} \in \mathbb{R}^{H \times N}$ are generated for each data sample, and the input X_t and X_s are masked by $\widetilde{X} = W_{mask} \odot X, X \in \{X_t, X_s\}$. Then, we replace X with $\hat{X}$ to generate the output vector $\hat{Z}_2$ with BTA. After that, X' is reconstructed with $\hat{Z}_2$ using a linear connection layer: $X' = linear(\hat{Z}_2)$. Finally, we utilize the mean squared error loss to present the reconstruction loss, which can be formulated as:

$$L_{MSE} = \sum_{(i,j) \in Q} (X'(i,j) - X(i,j))^2$$

where we utilize $Q = \{(i,j)|W_{mask}(i,j) = 0\}$ to obtain the loss related to the predictions on the masked values. After the subtask, the centrality vectors of $c_{j,\rho}, c_{j,\theta}$, and $c_{j,\varphi}$ for each j in $\{1, 2, ...M\}$ are set as the initialized parameters for the satisfaction modeling task. The training procedures of our proposed BTA model are summarized in Algorithm 1. Note that the unsupervised subtask aims to initialize a better mutli-centrality encoding which reflects the spatial relations of EEG channels. We find that the initialization process is effective when supervised labels are limited while not when using a larger dataset, which is elaborated in Section 5.1.

4 EXPERIMENTAL SETUP

This section details the datasets and experimental settings. The Search-Brainwave [56] dataset and the AMIGOS [34] dataset are applied for the satisfaction modeling experiments (detailed in 8.1). Then, we explore the performance of EEG-enhanced interactive search and recommendation systems with a search result re-ranking task on the Search-Brainwave dataset and a rating prediction task on the AMIGOS dataset, respectively. The implementation code of our experiment is based on PyTorch [1] and is publicly available in https://github.com/YeZiyi1998/DL4EEG-Classification.

4.1 Satisfaction prediction

4.1.1 Baselines. We exploit three types of EEG classification models as baselines, topology-invariant models, CNN-based models, and GCN-based models.

Topology-invariant models include SVM [49], DT [44], and multilayer perceptron (MLP) [10]. We implement these models with the scikit-learn library [38].

CNN-based models include EEGNet [28] and the recently proposed BENDR [25]. EEGNet stacks several CNN layers to boost its performance. BENDR applies CNN to extract EEG features and then uses transformers to capture temporal patterns of EEG signals, which achieves state-of-the-art performances in various brain–computer interface (BCI) tasks. It also adopts a cross-dataset pre-training task to initialize the parameters. We implement EEGNet and BENDR with their open-sourced code and use the public pre-trained weights to initialize the BENDR model.

GCN-based models include DGCNN [47], RGNN [61], and Het-EmotionNet [20]. DGCNN applies a public adjacency matrix to aggregate multichannel EEG information dynamically. RGNN uses two regularizers to improve the robustness. Het-EmotionNet uses mutual information to model the topographical information and fuse the temporal and spectral information together. Here we implement RGNN and Het-EmotionNet with open-sourced codes and DGCNN by ourselves since their code is not available.

4.1.2 Parameter Setups. We train BTA with the Adam optimizer [24]. The centralities' number M is selected from $\{1, 3, 5, 7, 9, 14\}$. Results show that 3 is a proper setting and achieves the best performance. Besides, we find that selecting centralities in different spatial positions (e.g., randomly selecting three channels as the centralities) leads to only a marginal difference. Therefore, we choose these three centralities (i.e., C_1, C_2, C_3) for simplicity and representative of the spatial distribution of EEG channels. Among them, C_1 is selected as a center point in the international 10-20 EEG system. C_2 and C_3 are selected as the left and right mastoid points since they are widely used as the reference channels in existing EEG studies [55]. The initialize learning rate, min-batch size, and the hidden dimension H (as denoted in Section 3.2) are tuned and selected from $\{0.01, 0.05\}$, $\{8, 32\}$ and $\{8, 16, 32\}$, respectively. The head number in the multihead attention mechanism is set to eight. Besides, in the unsupervised subtask, we set the random mask ratio as 15%, the same as prior work in time sequence prediction [58].

4.1.3 Protocols. We apply Area Under Curve (AUC) and F1-score as evaluation metrics in our experiments. For Search-Brainwave

---

[1]https://pytorch.org/

dataset, we conduct experiments on each subject and evaluate the models with a *task-independent ten-fold cross validation*: search tasks are partitioned into ten folds, and we leave each fold for evaluation after training with the remaining folds. For AMIGOS dataset, following existing studies [13, 18], we randomly split the data and apply a *ten-fold cross validation* for each subject.

4.2 Downstream task1: Search result re-ranking

4.2.1 Task definition. In the search result re-ranking task, unseen search results are re-ranked by the estimated satisfaction of historical search results. We present a search task within a query Q whose search results is a list $D = \langle d^1, ..., d^N \rangle$. We assume that a user historically examines the top several results $D_{hi} = \langle d^1, ..., d^{N_{hi}} \rangle$, hence the interaction on the i-th search result is formulated as $I_i = \{\hat{y}_i, d_i\}$, where $\hat{y}_i$ is the estimated satisfaction score to search result d_i. Then we would like to place the relevant search results in the unseen search results list $D_{un} = \langle d^{N_{hi}+1}, ..., d^N \rangle$ as high as possible. Therefore, the goal of the search result re-ranking task is:

$$\max \sum_Q \pi(\hat{D}_{un}, R_{un}), \hat{D}_{un} = F_{search}(Q, D_{un}, I_1, ..., I_{N_{hi}})$$

where π denotes the evaluation metric, e.g., Normalized Discounted Cumulative Gain (NDCG) [19] or Mean Average Precision (MAP) [62], R_{un} is the true relevance label of D_{un}, $\hat{D}_{un}$ is the returned re-ranked list, F_{search} denotes the re-ranking strategy that re-ranks the unseen search results according to historical interactions, which is elaborated in Section 4.2.2.

4.2.2 Methods. To generate a better re-ranked list for unseen results $\hat{D}_{un}$, we build a language model (LM) [27] and rewrite the query for a better intent estimation. The language model estimates the relevance of the words in historical search results $d_i = \{w_{i,1}, w_{i,2}, ...\} \in D_{hi}$ to find the most relevant words related to the query $Q = \{q_1, ..., q_k\}$. Then, the query is rewritten with l most relevant words as $Q' = \{q_1, ..., q_k, w_1, ..., w_l\}$. In the language model, the word relevance $R(w)$ can be formulated as:

$$R(w) = LM(D_{hi}, Q, F_{D_{hi}})$$

where $F_{D_{hi}}$ is a probability distribution of D_{hi}. Generally, a word will have higher relevance if it appears in search results with higher distribution probabilities. Convention language model sets $F_{D_{hi}}$ as a uniform distribution since we have no prior knowledge, which can be denoted as *uniformed language model* (ULM). With the estimated satisfaction inferred from brain signals, we adjust $F_{D_{hi}}$ to assign a higher probability $P(d|D_{hi})$ for satisfying search results:

$$P(d|D_{hi}) = (\lambda + \hat{y}(d))/(\lambda|D_{hi}| + \sum_{d \in D_{hi}} \hat{y}(d))$$

where λ is a hyperparameter to smooth the estimated satisfaction score $\hat{y}(d)$. Here we denote our method as *satisfaction-enhanced language model* (SLM). After extending the query Q to Q' with the language model, we adopt ranking algorithms (e.g., BM25 [43]) with Q' to generate the re-ranked search results list $\hat{D}_{un}$.

In our experiment, we use BM25 [43] as the ranking algorithm and exploit the performance of the original ranking list generated by BM25 and the re-ranked lists with ULM or SLM rewriting the query.

For parameter settings, we empirically set the length of rewritten words l and the hyperparameter λ as 5 and 2, respectively.

4.2.3 Protocols. For each search task in the Search-Brainwave dataset, if the user examines cu search results, we will re-rank the remanent $N - cu$ search results and evaluate the re-ranked list with the true relevance label of search results. Note that cu varies with search tasks since the user can stop her search at any time. Finally, ranking evaluation metrics NDCG and MAP are adopted to evaluate the averaged performance among all search tasks.

4.3 Downstream task2: Rating prediction

4.3.1 Task definition. In the rating prediction task, we predict the ratings of unseen user-item pairs based on historical user-item interactions. Specially, we denote the user and item set as U and V, respectively. The historical user-item interactions are denoted as $I_{hi} = \{(u, v, \hat{y}_{u,v}) | u \in U, v \in V\}$, where $\hat{y}_{u,v}$ is the satisfaction score of item v for user u estimated with brain signals. Note that in conventional recommendation tasks, $\hat{y}_{u,v}$ is replaced by explicit ratings or implicit feedback (e.g., click, liked). Then the rating prediction task is to better estimate the true label $y_{u,v}$ in the unseen user-item pairs $I_{un} = \{(u, v, y_{u,v}) | u \in U, v \in V\}$. Therefore, the task is formulated as:

$$\max \pi(I_{un}, F_{rec}(I_{hi}, \{(u, v) | (u, v, y_{u,v}) \in I_{un}\}))$$

where π denotes the evaluation metric, i.e., AUC , F_{rec} is the rating prediction strategy elaborated in Section 4.3.2, and we use the explicit annotation as true label $y_{u,v}$.

4.3.2 Methods. The personalized recommendation aims to learn a mapping function F that maps the rating of a given user-item pair with their embeddings:

$$y_{(u,v)} = F(e_u, e_v)$$

where e_u, e_v indicate the user embeddings and item embeddings, respectively, $y_{(u,v)}$ is the rating of the user-item pair (u, v). For a recommendation system, the true labels $y_{(u,v)}$ are difficult to obtain [52]. Therefore, in the training process, we hypothesize that we can obtain the true labels $y_{(u,v)}$ for a α ratio of user-item pairs. And we use the estimated satisfaction score $\hat{y}_{(u,v)}$ for other user-item pairs to introduce more training data. Finally, the true label $y_{(u,v)}$ is applied to evaluate the model performance.

We denote function F using α ratio of true labels T as $F^{T(\alpha)}$, and F using α ratio of true labels T and the estimated satisfaction S as $F^{T(\alpha),S}$. We exploit prevalent personalized recommendation methods as F, including LR [42], FM [41], and Wide&Deep [9]. All recommendation methods are implemented with the open-sourced code of Recbole [60] and are applied with its default settings.

4.3.3 Protocols. We treat each video segment in the AMIGOS dataset as an item and assign it with one-hot encoding as item embedding. For user embedding, the collected users' demographic information, personality profiles, and the mood (PANAS) files are used and expressed as 71-dimensional embedding vectors. We randomly split the user-item pairs into training, validating, and testing sets with a ratio of 8:1:1. In the training set, we replace the true labels with the satisfactions inferred from brain signals, and we

Table 1: The performance of satisfaction estimation for each model. ∗ indicates difference compared to the BTA is significant with p-value < 0.01.

Model	Search-Brainwave		AMIGOS	
	F1	AUC	F1	AUC
Topography-invariant				
DT	0.5642*	0.5205*	0.5608*	0.6245*
MLP	0.6196*	0.5204*	0.5629*	0.6123*
SVM	0.6227*	0.5189*	0.5580*	0.5892*
CNN-based				
BENDR	0.7118*	0.7291*	0.5580*	0.5869*
EEGNet	0.7254*	0.7614*	0.6025*	0.6920*
GCN-based				
DGCNN	0.7170*	0.7374*	0.6630*	0.7663*
HetEmotionNet	0.7362*	0.7717*	0.6428*	0.7405*
RGNN	0.7440*	0.7663*	0.6694*	0.7782*
BTA (ours)	**0.7837**	**0.8278**	**0.7143**	**0.8353**

evaluate rating prediction performance with the true labels in the validating and testing set.

5 RESULTS AND DISCUSSIONS

We empirically evaluate BTA in the satisfaction prediction task and utilize the estimated satisfaction to boost interactive information access performance to address the following research questions:

- **RQ1** Can we effectively estimate satisfaction feedback signals (e.g., feedback on search results and preference of items) with BTA from brain signals?
- **RQ2** Can we interactively provide satisfying information (e.g., helpful search results and users' preferred items) with the satisfaction inferred from the brain signals?

To address **RQ1**, we compare BTA with prevalent EEG classification baselines and delve into BTA's components in Section 5.1. Then, we summarize the performance of the search result re-ranking task and the rating predicting tasks to answer **RQ2** in Section 5.2 and Section 5.3, respectively.

5.1 Satisfaction Prediction

5.1.1 Overall performance. Table 1 presents the satisfaction estimation performance in terms of F1-score and AUC [39] of different models. From Table 1, we have the following observations: (1) All CNN-based models and GCN-based models outperform topography-invariant models in most of the evaluation metrics. The topography-invariant models directly concatenate all the EEG channels' features together. Thus the topographical information is omitted, which leads to limited performance. (2) In general, CNN-based models perform worse than GCN-based models. CNN-based architectures compress the 3D topographical information into 2D representation and aggregate the information in adjacent EEG channels. Conversely, GCN-based models utilize learnable adjacency matrixes to learn a more flexible aggregation strategy, which better exploits the topographical information than CNN-based models. (3) The proposed BTA performs the best among all models. On

the one hand, BTA introduces multi-centrality encoding to exploit the 3D spatial relations. , while previous models don't take the 3D information into account On the other hand, the spatial attention mechanism is applied to capture the topographical information adaptively. Thus we can learn different aggregation strategies in different data samples. Conversely, previous CNN-based and GCN-based models utilize common aggregation weights shared by all data samples. One exception is HetEmotionNet, which uses mutual information between EEG channels to learn the aggregation weight. More detailed comparison on the exploitation of topographical information between BTA and HetEmotionNet is conducted in Section 5.1.3.

5.1.2 Ablation study. To explore the effectiveness of different components in BTA, we compare it to a series of baselines without considering a variety of model structures or training strategies. Figure 3 shows the results of BTA and its variants. w/o A, w/o M, and w/o S indicate the BTA model that masks the spatial attention module, the multi-centrality encoding module, and the subtask module, respectively. As shown in Figure 3, there are different degrees of performance degradation in both tasks when masking one of the three components. This implies that all of them facilitate the model performance. Among these components, the spatial attention module plays the most important role, which suggests the effectiveness of adaptively aggregating the channel information. Besides, the subtask module leads to the least performance improvement. Especially in the AMIGOS dataset, the performance improvement of BTA is merely 0.003 in terms of F1 compared to BTA w/o S. We suggest that the AMIGOS dataset contains more training samples and supervised labels, and thus the unsupervised procedure to initialize the parameters is less necessary.

5.1.3 Brain topography analysis. In this subsection, we explore the brain topographical information in our proposed model and the baseline HetEmotionNet by visualizing their aggregation weight. Although BTA and HetEmotionNet have different architectures, both of their aggregation weights indicate how the model aggregate the information from other EEG channels to a certain channel. For BTA, the aggregation weight denotes the averaged attention weight between EEG channels among all data samples. For HetEmotionNet, the aggregation weight indicates the averaged edge weight [20] between EEG channels among all data samples.

Figure 4 presents the visualization of the aggregation weight of various EEG channels to channel F4 in the Search-Brainwave
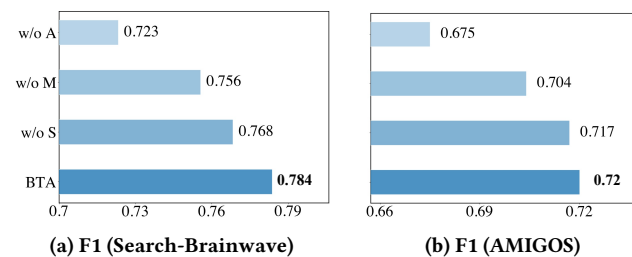

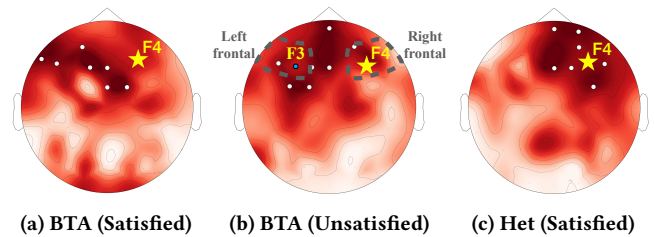**(a) BTA (Satisfied)**   **(b) BTA (Unsatisfied)**   **(c) Het (Satisfied)**

Figure 4: The visualization of the aggregation weights in BTA/HetEmotionNet and satisfied/unsatisfied data samples. The darker color indicates higher aggregation weight to channel F4 and the highlighted channels have the highest weights among all channels. Het indicates HetEmotionNet.

dataset. Here we select channel F4 because previous neurological studies [1, 45] suggest that the frontal alpha asymmetry (i.e., the α band difference between left and right frontal) indicates motivation, desire, and positive/negative feelings. From Figure 4, we can observe that the aggregation weight to channel F4 is higher in brain regions of right frontal and left frontal for HetEmotionNet and BTA, respectively. HetEmotionNet adopts mutual information [26] to obtain the edge weight. Since adjacent channels usually share higher mutual information scores, the aggregation process is just aggregating some of the most adjacent channels. Conversely, BTA adaptively captures the topographical relationships between EEG channels with attention mechanism and multi-centrality encoding. Thus it aggregates information in a flexible manner. Interestingly, the aggregation weights of BTA are higher in left frontal. BTA utilizes a higher weight to capture the relationship between F4 and channels in left frontal, including F3, which implies BTA has good coherence to existing neurological studies [1]. Besides, the satisfied and unsatisfied samples in BTA have different topographical relationships, which also demonstrate the data-dependent modeling strategy can be adaptive to different user statuses.

5.2 Search result re-ranking performance

Table 3 summarizes the performance of the search result re-ranking task. We can observe that the proposed SLM outperforms BM25 and ULM. It demonstrates the effectiveness of SLM, which rewrites the query with additional relevant words inferred from the search

(a) F1 (Search-Brainwave)   **(b) F1 (AMIGOS)**

Figure 3: Performance comparisons between BTA and its variants. A: spatial attention; M: multi-centrality encoding; S: unsupervised subtask.
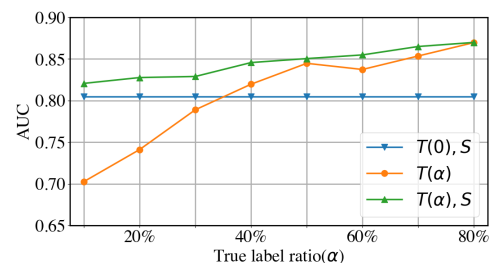

Figure 5: The performance of Wide&Deep using only the estimated satisfaction $F^{T(0),S}$, different ratios of true labels $F^{T(\alpha)}$, and their combination $F^{T(\alpha),S}$.

Table 2: A case study to investigate the rewriting performance. The bold words are related to the user intent.

Query	Search result/Satisfaction	ULM rewriting	SLM rewriting
permanent teeth	(1) ..., online medical advice: dentist Mr.Song /☹ (2) how **old** does a **child** grow its permanent teeth, ... /☺	permanent, teeth, dentist, know, online, **child**, **kid**	permanent, teeth, **child**, **old**, **when**, know, **kid**

Table 3: The performance of search result re-ranking for each model. ∗ indicates difference compared to SLM is significant with p-value < 0.01.

Model	NDCG@1	NDCG@5	NDCG@10	MAP@10
BM25	0.6881*	0.7397*	0.8164*	0.7333*
ULM	0.7237*	0.7620*	0.8309*	0.7687*
SLM	**0.7351**	**0.7767**	**0.8337**	**0.7741**

Table 4: The AUC performance of personalized rating prediction for each model. $F^{T(0.1),S}$ indicates models using 0.1 ratio of true labels T and the estimated satisfaction S.

Model	LR	FM	Wide&Deep
$F^{T(0.1)}$	0.6798	0.5010	0.7027
$F^{T(0.1),S}$	**0.7537**	**0.8056**	**0.8207**

results with higher estimated user satisfaction. The significant improvements of SLM also demonstrate the benefits of introducing the estimated satisfaction into interactive search scenarios.

To investigate the rewriting performance of SLM and ULM, we conduct a case study to analyze their difference. Table 2 shows a search task that requires the user to explore information about when the period of growing permanent teeth is. The ULM rewrites the query with words appeared in both search result 1 (e.g., "online", "dentist") and search result 2 (e.g., "child", "kid"). On the other hand, with the help of the estimated satisfaction, the SLM tends to rewrite the query with words that appeared in satisfying search results, i.e., search result two. Therefore, words such as "child", "old", and "when" are used for rewriting in SLM, and the rewritten query can better present user intent. Then, the re-ranking performance can be improved with the rewritten query.

5.3 Rating prediction performance

Table 4 summarizes the rating prediction performances of different models using a true label ratio of 0.1 (see in Section 4.3.2). As shown in Table 4, all recommendation models improve a large margin with the enhancement of the estimated satisfaction S. This suggests the effectiveness of introducing brain signals into recommendation by utilizing the estimated satisfaction to train the model.

Additionally, to better understand to what extent brain signals can improve the interactive recommendation system, we explore the performance of Wide&Deep using different ratios of true labels T and whether using the estimated satisfaction S or not. Figure 5 presents their performance comparison. We observe that Wide&Deep using only the satisfaction estimated with brain signals S (i.e., $F^{T(0),S}$) are as effective as Wide&Deep using about 30%-40% true labels T (i.e., $F^{T(\alpha)}$). Besides, $F^{T(\alpha),S}$ outperforms

$F^{T(\alpha)}$ at the same true label ratio α, which indicates introducing the estimated satisfaction S can boost the performance stably. Note that in recommendation scenarios, true user satisfaction is often scarce [51]. Thus, the estimated satisfaction is valuable and here we demonstrate to what extent the satisfaction estimated with brain signals can improve the performance.

6 CONCLUSION

In this paper, we addresss a problem of utilizing brain signals for satisfaction modeling in interactive information systems. Compared to conventional user signals, brain signals can directly present user status and thus contain less bias. We then propose a Brain Topography Adaptive network (BTA) to estimate user satisfaction with EEG signals. BTA exploits the 3D topographical information of EEG channels by multi-centrality encoding module and adaptively learns data-dependent aggregation strategies with spatial attention mechanisms. Extensive experiments on the Search-Brainwave and the AMIGOS datasets demonstrate the outstanding performance of our model in comparison with various competitive baselines. Moreover, to verify that the estimated satisfaction can help the interactive information access procedures, we conduct two downstream tasks in search and recommendation scenarios. Experimental results show that brain signals can boost the system performance of result re-ranking in search and rating prediction in recommendation.

With wearable devices becoming more portable and cheaper, information systems are possible to collect users' psychophysiological signals in situations such as Virtual Reality (VR) applications [4] and the disabled service [7]. As BCIs become more prevalent and cheaper, we believe that more application scenarios will soon emerge. Besides directly controlling information systems, we reveal another benefit of BCI for information systems. We suggest utilizing brain signals to automatically predict user satisfaction with well-designed models, and then, the system can better understand the user and interactively provide useful information. Future studies may include interactive systems in broad situations, an overall framework for EEG-enhanced search systems, and online learning algorithms for real-life EEG based satisfaction estimation.

7 ACKNOWLEDGEMENT

REFERENCES

[1] Mashael Aldayel, Mourad Ykhlef, and Abeer Al-Nafjan. 2020. Deep learning for EEG-based preference classification in neuromarketing. *Applied Sciences* 10, 4 (2020), 1525.

[2] Rashid Ali and MM Sufyan Beg. 2011. An overview of Web search evaluation methods. *Computers & Electrical Engineering* 37, 6 (2011), 835–848.

[3] 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.

[4] Thomas Baumgartner, Lilian Valko, Michaela Esslen, and Lutz Jäncke. 2006. Neural correlate of spatial presence in an arousing and noninteractive virtual reality: an EEG and psychophysiology study. *CyberPsychology & Behavior* 9, 1 (2006), 30–45.

[5] Daniel E Berlyne. 1970. Novelty, complexity, and hedonic value. *Perception & psychophysics* 8, 5 (1970), 279–286.

[6] Anjan Chatterjee. 2003. Prospects for a cognitive neuroscience of visual aesthetics. (2003).

[7] Xuesong Chen, Ziyi Ye, Xiaohui Xie, Yiqun Liu, Xiaorong Gao, Weihang Su, Shuqi Zhu, Yike Sun, Min Zhang, and Shaoping Ma. 2022. Web Search via an Efficient and Effective Brain-Machine Interface. In *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*. 1569–1572.

[8] Ye Chen, Yiqun Liu, Min Zhang, and Shaoping Ma. 2017. User satisfaction prediction with mouse movement information in heterogeneous search environment. *IEEE Transactions on Knowledge and Data Engineering* 29, 11 (2017), 2470–2483.

[9] Heng-Tze Cheng, Levent Koc, Jeremiah Harmsen, Tal Shaked, Tushar Chandra, Hrishi Aradhye, Glen Anderson, Greg Corrado, Wei Chai, Mustafa Ispir, et al. 2016. Wide & deep learning for recommender systems. In *Proceedings of the 1st workshop on deep learning for recommender systems*. 7–10.

[10] 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*. 602–611.

[11] Ruo-Nan Duan, Jia-Yi Zhu, and Bao-Liang Lu. 2013. Differential entropy feature for EEG-based emotion classification. In *6th International IEEE/EMBS Conference on Neural Engineering (NER)*. IEEE, 81–84.

[12] Steve Fox, Kuldeep Karnawat, Mark Mydland, Susan Dumais, and Thomas White. 2005. Evaluating implicit measures to improve web search. *ACM Transactions on Information Systems (TOIS)* 23, 2 (2005), 147–168.

[13] Martin Gjoreski, Blagoj Mitrevski, Mitja Luštrek, and Matjaž Gams. 2018. An inter-domain study for arousal recognition from physiological signals. *Informatica* 42, 1 (2018).

[14] Shan Guan, Kai Zhao, and Shuning Yang. 2019. Motor imagery EEG classification based on decision tree framework and Riemannian geometry. *Computational intelligence and neuroscience* 2019 (2019).

[15] Jacek Gwizdka, Rahilsadat Hosseini, Michael Cole, and Shouyi Wang. 2017. Temporal dynamics of eye-tracking and EEG during reading and relevance decisions. *Journal of the Association for Information Science and Technology* 68, 10 (2017), 2299–2312.

[16] Ahmed Hassan, Ryen W White, Susan T Dumais, and Yi-Min Wang. 2014. Struggling or exploring? Disambiguating long search sessions. In *Proceedings of the 7th ACM international conference on Web search and data mining*. 53–62.

[17] Tiffany C Ho, Colm G Connolly, Eva Henje Blom, Kaja Z LeWinn, Irina A Strigo, Martin P Paulus, Guido Frank, Jeffrey E Max, Jing Wu, Melanie Chan, et al. 2015. Emotion-dependent functional connectivity of the default mode network in adolescent depression. *Biological psychiatry* 78, 9 (2015), 635–646.

[18] Jingzhao Hu, Chen Wang, Qiaomei Jia, Qirong Bu, Richard Sutcliffe, and Jun Feng. 2021. ScalingNet: extracting features from raw EEG data for emotion recognition. *Neurocomputing* 463 (2021), 177–184.

[19] Kalervo Järvelin and Jaana Kekäläinen. 2017. IR evaluation methods for retrieving highly relevant documents. In *ACM SIGIR Forum*, Vol. 51. ACM New York, NY, USA, 243–250.

[20] Ziyu Jia, Youfang Lin, Jing Wang, Zhiyang Feng, Xiangheng Xie, and Caijie Chen. 2021. HetEmotionNet: two-stream heterogeneous graph recurrent neural network for multi-modal emotion recognition. In *Proceedings of the 29th ACM International Conference on Multimedia*. 1047–1056.

[21] Ziyu Jia, Youfang Lin, Jing Wang, Ronghao Zhou, Xiaojun Ning, Yuanlai He, and Yaoshuai Zhao. 2020. GraphSleepNet: Adaptive Spatial-Temporal Graph Convolutional Networks for Sleep Stage Classification.. In *IJCAI*. 1324–1330.

[22] Diane Kelly. 2009. *Methods for evaluating interactive information retrieval systems with users*. Now Publishers Inc.

[23] Hyun Hee Kim and Yong Ho Kim. 2019. ERP/MMR algorithm for classifying topic-relevant and topic-irrelevant visual shots of documentary videos. *Journal of the Association for Information Science and Technology* 70, 9 (2019), 931–941.

[24] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[25] Demetres Kostas, Stephane Aroca-Ouellette, and Frank Rudzicz. 2021. BENDR: using transformers and a contrastive self-supervised learning task to learn from massive amounts of EEG data. *Frontiers in Human Neuroscience* 15 (2021).

[26] Alexander Kraskov, Harald Stögbauer, and Peter Grassberger. 2004. Estimating mutual information. *Physical review E* 69, 6 (2004), 066138.

[27] Victor Lavrenko and W Bruce Croft. 2017. Relevance-based language models. In *ACM SIGIR Forum*, Vol. 51. ACM New York, NY, USA, 260–267.

[28] Vernon J Lawhern, Amelia J Solon, Nicholas R Waytowich, Stephen M Gordon, Chou P Hung, and Brent J Lance. 2018. EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces. *Journal of neural engineering* 15, 5 (2018), 056013.

[29] Rui Li, Yiting Wang, and Bao-Liang Lu. 2021. A Multi-Domain Adaptive Graph Convolutional Network for EEG-based Emotion Recognition. In *Proceedings of the 29th ACM International Conference on Multimedia*. 5565–5573.

[30] Xiang Li, Dawei Song, Peng Zhang, Guangliang Yu, Yuexian Hou, and Bin Hu. 2016. Emotion recognition from multi-channel EEG data through convolutional recurrent neural network. In *2016 IEEE international conference on bioinformatics and biomedicine (BIBM)*. IEEE, 352–359.

[31] Mengyang Liu, Yiqun Liu, Jiaxin Mao, Cheng Luo, Min Zhang, and Shaoping Ma. 2018. " Satisfaction with Failure" or" Unsatisfied Success" Investigating the Relationship between Search Success and User Satisfaction. In *Proceedings of the 2018 world wide web conference*. 1533–1542.

[32] Yiqun Liu, Chao Wang, Ke Zhou, Jianyun Nie, Min Zhang, and Shaoping Ma. 2014. From skimming to reading: A two-stage examination model for web search. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*. 849–858.

[33] Colin Martindale and Kathleen Moore. 1988. Priming, prototypicality, and preference. *Journal of Experimental Psychology: Human Perception and Performance* 14, 4 (1988), 661.

[34] Juan Abdon Miranda-Correa, Mojtaba Khomami Abadi, Nicu Sebe, and Ioannis Patras. 2018. Amigos: A dataset for affect, personality and mood research on individuals and groups. *IEEE Transactions on Affective Computing* 12, 2 (2018), 479–493.

[35] Yashar Moshfeghi, Peter Triantafillou, and Frank Pollick. 2019. Towards predicting a realisation of an information need based on brain signals. In *The World Wide Web Conference*. 1300–1309.

[36] 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.

[37] Marcos Nadal, Enric Munar, Miquel Angel Capo, Jaume Rossello, and Camilo Jose Cela-Conde. 2008. Towards a framework for the study of the neural correlates of aesthetic preference. *Spatial vision* 21, 3 (2008), 379.

[38] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. 2011. Scikit-learn: Machine learning in Python. *the Journal of machine Learning research* 12 (2011), 2825–2830.

[39] David MW Powers. 2020. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *arXiv preprint arXiv:2010.16061* (2020).

[40] B Pyakillya, N Kazachenko, and N Mikhailovsky. 2017. Deep learning for ECG classification. In *Journal of physics: conference series*, Vol. 913. IOP Publishing, 012004.

[41] Steffen Rendle. 2010. Factorization machines. In *2010 IEEE International conference on data mining*. IEEE, 995–1000.

[42] Matthew Richardson, Ewa Dominowska, and Robert Ragno. 2007. Predicting clicks: estimating the click-through rate for new ads. In *Proceedings of the 16th international conference on World Wide Web*. 521–530.

[43] Stephen Robertson and Hugo Zaragoza. 2009. *The probabilistic relevance framework: BM25 and beyond*. Now Publishers Inc.

[44] S Rasoul Safavian and David Landgrebe. 1991. A survey of decision tree classifier methodology. *IEEE transactions on systems, man, and cybernetics* 21, 3 (1991), 660–674.

[45] Louis A Schmidt and Laurel J Trainor. 2001. Frontal brain electrical activity (EEG) distinguishes valence and intensity of musical emotions. *Cognition & Emotion* 15, 4 (2001), 487–500.

[46] Paul J Silvia. 2005. Cognitive appraisals and interest in visual art: Exploring an appraisal theory of aesthetic emotions. *Empirical studies of the arts* 23, 2 (2005), 119–133.

[47] Tengfei Song, Wenming Zheng, Peng Song, and Zhen Cui. 2018. EEG emotion recognition using dynamical graph convolutional neural networks. *IEEE Transactions on Affective Computing* 11, 3 (2018), 532–541.

[48] Akara Supratak and Yike Guo. 2020. TinySleepNet: An efficient deep learning model for sleep stage scoring based on raw single-channel EEG. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 641–644.

[49] Johan AK Suykens and Joos Vandewalle. 1999. Least squares support vector machine classifiers. *Neural processing letters* 9, 3 (1999), 293–300.

[50] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems* 30 (2017).

[51] Wenjie Wang, Fuli Feng, Xiangnan He, Liqiang Nie, and Tat-Seng Chua. 2021. Denoising implicit feedback for recommendation. In *Proceedings of the 14th ACM international conference on web search and data mining*. 373–381.

[52] Wenjie Wang, Fuli Feng, Xiangnan He, Hanwang Zhang, and Tat-Seng Chua. 2021. Clicks can be cheating: Counterfactual recommendation for mitigating clickbait issue. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1288–1297.

[53] Xiao-Wei Wang, Dan Nie, and Bao-Liang Lu. 2014. Emotional state classification from EEG data using machine learning approach. *Neurocomputing* 129 (2014), 94–106.

[54] Yingying Wu, Yiqun Liu, Yen-Hsi Richard Tsai, and Shing-Tung Yau. 2019. Investigating the role of eye movements and physiological signals in search satisfaction prediction using geometric analysis. *Journal of the Association for Information Science and Technology* 70, 9 (2019), 981–999.

[55] Dezhong Yao, Yun Qin, Shiang Hu, Li Dong, Maria L Bringas Vega, and Pedro A Valdés Sosa. 2019. Which reference should we use for EEG and ERP practice? *Brain topography* 32, 4 (2019), 530–549.

[56] Ziyi Ye, Xiaohui Xie, Yiqun Liu, Zhihong Wang, Xuancheng Li, Jiaji Li, Xuesong Chen, Min Zhang, and Shaoping Ma. 2022. Why Don't You Click: Understanding Non-Click Results in Web Search with Brain Signals. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval.*

[57] Zhongliang Yin, Jun Li, Yun Zhang, Aifeng Ren, Karen M Von Meneen, and Liyu Huang. 2017. Functional brain network analysis of schizophrenic patients with positive and negative syndrome based on mutual information of EEG time series. *Biomedical Signal Processing and Control* 31 (2017), 331–338.

[58] George Zerveas, Srideepika Jayaraman, Dhaval Patel, Anuradha Bhamidipaty, and Carsten Eickhoff. 2021. A transformer-based framework for multivariate time series representation learning. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining.* 2114–2124.

[59] Guanhua Zhang, Minjing Yu, Yong-Jin Liu, Guozhen Zhao, Dan Zhang, and Wenming Zheng. 2021. SparseDGCNN: recognizing emotion from multichannel EEG signals. *IEEE Transactions on Affective Computing* (2021).

[60] Wayne Xin Zhao, Shanlei Mu, Yupeng Hou, Zihan Lin, Yushuo Chen, Xingyu Pan, Kaiyuan Li, Yujie Lu, Hui Wang, Changxin Tian, et al. 2021. Recbole: Towards a unified, comprehensive and efficient framework for recommendation algorithms. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management.* 4653–4664.

[61] Peixiang Zhong, Di Wang, and Chunyan Miao. 2020. EEG-based emotion recognition using regularized graph neural networks. *IEEE Transactions on Affective Computing* (2020).

[62] Mu Zhu. 2004. Recall, precision and average precision. *Department of Statistics and Actuarial Science, University of Waterloo, Waterloo* 2, 30 (2004), 6.

8 SUPPLEMENTARY MATERIAL

8.1 Dataset

8.1.1 Search-Brainwave dataset. The Search-Brainwave dataset [56] records the brain signals generated by 18 participants when performing search tasks. Each participant averagely accomplishes 69.6 search tasks and examines 3.6 corresponding search results in a search task. During the procedures, their explicit feedback (i.e., satisfied or dissatisfied) to search results are collected. And the dataset also provides the true relevance label of each search result for performance evaluation. We choose this dataset because it is a public benchmark specially designed for interactive search tasks.

For feature extraction, we utilize the officially preprocessed temporal and spectral features (i.e., raw signals and DE [11] features, respectively) and treat the brain signals in response to each search result as a data sample.

8.1.2 AMIGOS dataset. The AMIGOS dataset [34] is a public dataset that includes EEG, electrocardiogram (ECG), and other psychological signals generated by 40 participants under video stimulations. Each participant watches 16 short videos and 4 long videos in two experiments. They rate each video in "valence", "arousal", "dominance", "familiarity", and "liking" from 1 to 9. Although the AMIGOS dataset is not specially designed for recommendation scenarios, we choose this dataset for recommendation scenario because it contains rich participant profiles (anonymized participants' data, personality profiles, and mood (PANAS) profiles). The user profile can be utilized as user embeddings for the presonalized rating prediction task (see in Section 4.3.1). Conducting a specially designed user study on interactive recommendation scenarios is left as future work.

In our experiments, we adopt the data in short videos experiment since the long videos experiment is conducted in different configurations. We divide the "liking" annotation into satisfied and unsatisfied with a threshold of five for our satisfaction estimating tasks. For feature extraction, we utilize the officially preprocessed EEG signals and apply a non-overlapping window with a length of one second to divide each video into several segments. The window length is different from the original paper [34] but agrees with a lot of existing EEG classification tasks [20, 47]. We utilize a different time window since the original paper also considers other physiological signals, such as ECG, which requires a longer time window than EEG [40]. For brain signals in response to each video segment, the raw signals are used as temporal features, and we use Fourier transform over four frequency bands (i.e., $\theta, \alpha, \beta, \gamma$) to extract the DE [11] features as spectral features.

8.2 Parameter Setups of Satisfaction Prediction Baselines

For the satisfaction prediction baselines of various EEG classification models (see in Section 4.1.1), we use the original parameter settings according to their original papers and released codes [20, 25, 28, 47, 61]. However, there exist some implementation details as follows. For EEGNet [28], the original parameter settings are not suitable for AMIGOS dataset due to the difference in input data length. Thus we decrease the second convolution layer's kernel size to 7 and the last pooling layer's kernel size to $(2, 1)$ for experiments on AMIGOS dataset. For RGNN [61], the released code excludes the regularized parameters, and we empirically set it as 0.001. The implementation code of our experiment is publicly available in https://github.com/YeZiyi1998/DL4EEG-Classification.

Algorithm 1: The Training Procedures of BTA.

Input: User's brain signals in response to search results $\langle X^1, ..., X^N \rangle$; Usefulness of search results $\langle Y^1, ..., Y^N \rangle$; The Initialize BTA model Φ.

1 Generate time domain and spectral domain features $X^t = \langle x^{1,t}, ..., x^{N,t} \rangle$ and $X^s = \langle x^{1,s}, ..., x^{N,s} \rangle$.

2 $\Phi' = Copy(\Phi)$.

3 **for** *iteration=1,2,...* **do**

4   **for** *all X in $\langle X^1, ..., X^N \rangle$* **do**

5   Generate randomized binary noise masks $W_{t,mask}$ and $W_{s,mask}$.

6          $\widetilde{X} = W_{mask} \odot X, X \in \{X_t, X_s\}$.

7   Compute reconstruction loss $L_{MSE} = L_{MSE}(\Phi', \widetilde{X}, X)$.

8   Update Φ' with L_{MSE}.

9   **end**

10 **end**

11 Replace the centrality embedding vectors of Φ with Φ'.

12 **for** *iteration=1,2,...* **do**

13   **for** *all X in $\langle X^1, ..., X^N \rangle$ and Y in $\langle Y^1, ..., Y^N \rangle$* **do**

14   Compute classification loss $L = L(\Phi, X, Y)$.

15   Update Φ with L.

16   **end**

17 **end**

18 return Φ;