Rethink Data-driven Human Behavior Prediction: A Psychology-powered Explainable Neural Network

Jiyao Wang, Chunxi Huang, Weiyin Xie, Dengbo He*, Ran Tu

ARTICLE INFO

Keywords:
Human behavior prediction
Psychological factor
Multi-task
Neural network

ABSTRACT

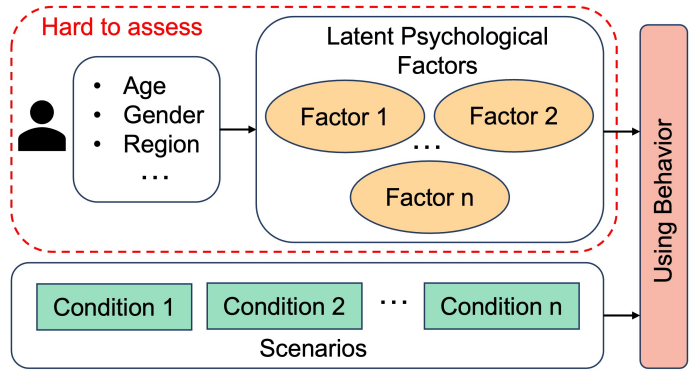
Human behavior prediction is a rapidly evolving field with numerous applications in various domains. The rapid development of machine learning facilitates research on data-driven behavior prediction. At the same time, in human factors domains, researchers still focus on modeling psychological decision-making processes from the psychological perspectives of view based on subjective data. However, it is difficult to obtain one's psychological states in real-world scenarios, given that the collection of subjective data in real-world settings is difficult and intrusive. Thus, psychology theories can hardly be applied to behavior prediction directly. To bridge the gap between data-driven behavior prediction and psychological decision-making models, we proposed a novel Psychology-powered Explainable Neural network (PEN). In PEN, for the first time, the psychological factors of individuals (i.e., human-side features such as attitudes to technologies) are modeled and recovered explicitly based on one's historical behaviors in certain scenarios. The multi-task optimization ensures each component in PEN to be optimized with proper gradient. Assessment results over three evaluation protocols demonstrate that, based on the historical behaviors of an individual only, the PEN can outperform the best existing behavior prediction models by 1.87% and 4.68% on two datasets when a 1:9 training-testing data partition was adopted. The framework in PEN has provided a new solution for explainable human behavior prediction when psychological information is not directly available.

1. Introduction

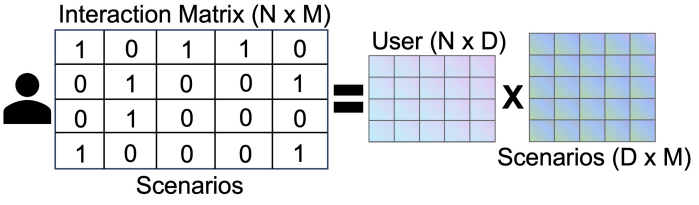
Understanding and accurately predicting human behavior is a crucial yet complex endeavor across many academic disciplines. Modeling human cognition and decision process allows researchers to gain insight into the underlying mental processes leading to human decisions (Ajzen, 1991; Ariely, Bracha and Meier, 2009). In recent years, the introduction of data-driven methods has revolutionized various fields such as transportation, healthcare, and social sciences by providing new and effective ways to predict human behavior (Xing, Chen, Zhang, Wang and Zhang, 2021; Wang, Kong, Xia and Sun, 2019; Ullah, Liu, Yamamoto, Zahid and Jamal, 2023). These approaches utilized vast quantities of data collected from the real world to identify patterns in human behavior automatically.

However, there are still issues in conventional data-driven methods (e.g., logistic regression (Hosmer Jr, Lemeshow and Sturdivant, 2013), LightGBM (Ke, Meng, Finley, Wang, Chen, Ma, Ye and Liu, 2017)). As shown in Figure 1(a), firstly, data collected directly from real-world environments tends to record only surface-level information, such as the ultimate decisions or behaviors of the user, environmental variables, or only some de-identified demographic information (e.g., region and gender). Deeper psychological or cognition-related information directly or indirectly related to the users' behaviors, for example, the attitudes to the technology and personality, are usually unavailable due to privacy and data acquisition issues. Thus, critical information that can influence individual decision-making or behavior patterns may not be available for models to learn, which can impair the predictive accuracy of the model. Additionally, most data-driven models for behavior prediction are in favor of the general data distribution (Leevy, Khoshgoftaar, Bauder and Seliya, 2018) but ignore individual heterogeneity. With the rapid development of deep learning (DL) and its strong capacity to learn non-linear correlations (Liu and Lang, 2019), existing studies (Wu, Li, Zhao and Qian, 2020; Wu, He, Wang, Zhang and Wang, 2022; Eke, Norman, Shuib and Nweke, 2019) have tried to implicitly model preference information from surface-level data through DL-based methods and achieved remarkable performance. For instance, in the recommendation field, Koren, Bell and Volinsky (2009) represented each distinct user's preference and characteristics of targeted scenarios by factorizing an individual-item interaction matrix to two learnable high-dimension vectors (Figure 1(b)). Nevertheless, the individual preferences represented through measures of vector similarity or probability may still not reveal the psychological states or true preferences of the users (Jung, Hong and Kim, 2005).

ORCID(s):



(a)



(b)

Figure 1: Illustration of example data-driven models. Subfigure (a) indicates the overall structure of the conventional data-driven model, where latent psychological factors are hard to access; (b) is the key principle of interaction matrix factorization, which is widely used in decision or preference prediction.

Being different from the research focusing on pure data-driven methods, scholars in the fields of psychology and cognitive science used different approaches: they investigate causal relations between influential psychological factors and human behaviors and model the cognitive processes behind human decisions or behaviors. Abundant studies illustrated that psychological factors are closely associated with real-world decisions or behaviors, such as range anxiety for electric vehicle charging behavior (Wang, Huang, He and Tu, 2023a; Tian, Sun, Hu and Du, 2023) and perceived benefits for the participation of shared parking (Wang, Guan, Guo, Han and Bian, 2022; Wang, Guan, Han and Cao, 2021a). However, to the best of our knowledge, no data-driven behavior prediction method explicitly modeled the psychological states of human decision-makers. Ignoring the intermediate cognitive or psychological states of human beings can increase the demands on the nonlinear modeling capacity of the models, making the models less explainable.

Though modeling psychological and cognitive states may increase the performance of machine learning models, due to the diverse concentration of different communities, most of the existing psychological or cognitive decision-making models are based on statistical methods (e.g., structured equation model for the theory of planned behaviors Cai, Jin and Chen (2022)), which focus on linear relationships among the variables of interests and can hardly be incorporated into data-driven models directly. Further, traditionally, the psychological states of human beings can only be reliably measured through subjective reports from participants using questionnaires or interviews and thus can be intrusive to be collected in real-world settings. Thus, we conclude the following three research questions: RQ1: *Would the introduction of theories from human factors and psychology domains enhance the performance and interpretability of data-driven models*; RQ2: *If effective, how to further promote the data-driven model's explainability by introducing theories from human factors and psychology domains*; RQ3: *Can the newly proposed framework maintain a good generalization ability despite the lack of trainable latent psychological data related to people*.

To bridge the gaps mentioned above, in this study, we proposed a novel end-to-end framework, named **P**sychology-powered **E**xplainable **N**eural network (**PEN**). The principle of this work is to recover one's latent psychological features (e.g., attitudes toward technologies) through individual feature learning based on one's historical behaviors or decisions in certain scenarios, which can increase the explainability of the data-driven neural networks. The latent psychological factors are selected based on previous research in human factors or psychology domains, given that using prior

knowledge can avoid the model fitting to some irrelevant features, which may cause the model to collapse locally onto some adverse features (Zhou, Khosla, Lapedriza, Oliva and Torralba, 2016). Specifically, the PEN framework leverages a human feature encoder to extract basic features from the learnable embedding bank for both decision prediction and auxiliary psychology-related feature modeling, which projects the basic human features to corresponding latent psychological features through several independent feature decoders. Then, the primary task is based on the human-scenario nonlinear interaction feature learning inspired by classic two-tower neural structure (He, Liao, Zhang, Nie, Hu and Chua, 2017). From the perspective of training, multiple task-specific auxiliary losses were utilized to assist the primary task through our designed joint-optimized framework. Finally, we built three evaluation protocols over two datasets to evaluate the generalizability of our framework. Given that in real-world applications, obtaining the psychological features of users is difficult and the models may have to be trained on a small dataset and applied to a large population, we evaluated the models on 9:1, 1:1, and 1:9 training-test data partitions. The main contributions of this work are as follows:

- To the best of our knowledge, this is the first attempt to involve prior research outcomes from the psychology field in the data-driven prediction model design, which bridges the cross-discipline gaps.
- The proposed multi-task learning framework (PEN) for human decision prediction explicitly models and outputs intermediate psychological variables during the decision-making process. This anthropomorphic reasoning process can make the DL-based models more explainable.
- Comparison experiments over two datasets and three evaluation protocols illustrated the outperformance and generalizability of PEN, and validated the correctness of the model design principles. Through visualization and statistical analysis, the validity of the modeled psychological states was verified, which further demonstrates the promising explainability of the PEN.

2. Related Work

2.1. Human Bahavior Prediction

Human behavior prediction is an interdisciplinary field that focuses on understanding and forecasting human actions, reactions, and decision-making in various contexts (Gulhane and Sajana, 2021). Human behavior prediction has numerous applications in various domains, such as driver behavior prediction (Ullah et al., 2023; Xing et al., 2021), item recommendation (Wang, Dong and Chen, 2021b; He et al., 2017), and marketing (Ahmed, Asif, Hina and Muzammil, 2017).

This area of research has gained significant attention in recent years due to advances in machine learning (ML) and artificial intelligence. For example, for the charging behavior prediction, there are an increasing number of studies that utilized ML methods to make prediction (Frendo, Graf, Gaertner and Stuckenschmidt, 2020; Chung, Khaki, Li, Chu and Gadh, 2019; Ullah et al., 2023). Recently, with the rapid development of deep learning (DL), researchers have started applying DL-based methods to learn nonlinear behavior patterns. As one of the most rapidly growing areas in recent years, the neural network is designed as a more powerful alternative method than previous data-driven matrix factorization (Klema and Laub, 1980; Koren et al., 2009) or ML methods (SUN, QU and REN, 2021; Adeniyi, Wei and Yongquan, 2016).

Nevertheless, restricted by the difficulty of collecting psychological and cognition-related data, existing prediction methods concentrate on exploring information from the scenario side. In brief, the model can only learn the direct correlations between human behaviors and multiple scenario conditions. In contrast, inspired by studies in human factors (Wang et al., 2023a; Wang, Zhang, Wen, He and Tu, 2023c), we wish to further model the intermediate role of psychological states between scenarios and human behavior levels, which may increase the generalizability and explainability of the DL models.

2.2. Auxiliary and Multi-task Learning

Auxiliary learning aims to enhance the model performance on a specific task by leveraging information from related auxiliary tasks (Chen, Wang, Guan, Liu and Zhu, 2022). To incorporate auxiliary tasks in the learning process of primary tasks, previous studies (Yu and Jiang, 2016; Cheng, Fang and Ostendorf, 2015) have examined the connection between tasks to select appropriate auxiliary tasks. Some other approaches incorporate auxiliary tasks by linearly blending auxiliary losses with the primary loss, then employing techniques such as grid search to adjust the linear

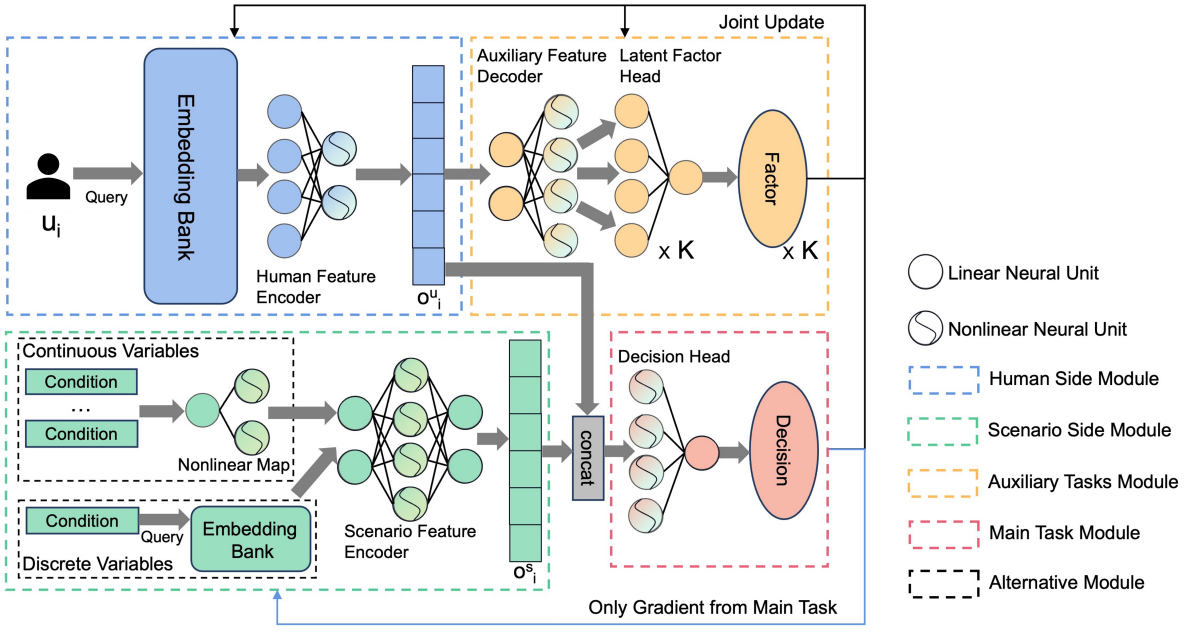


Figure 2: The overall architecture of the proposed PEN framework, which consists of one Human Side Module (labeled with blue); Scenario Side Module (contained with green dotted box); Auxiliary Tasks Module (contained with yellow dotted box); and the Main Task Module (contained with red dotted box). The Alternative Module (contained with black dotted box) is determined by the type of scenario inputs.

weights (Zhai, Oliver, Kolesnikov and Beyer, 2019). However, blending auxiliary losses, though straightforward and has been widely adopted, usually fails when dealing with a large number of auxiliary tasks due to the exponentially increasing complexity of the search space. To identify advantageous auxiliary tasks, recent studies (Lin, Baweja, Kantor and Held, 2019; Shi, Hoffman, Saenko, Darrell and Xu, 2020) adopted the parameter gradient similarity between the primary task and auxiliary tasks to weigh the loss of each task.

Besides, because both psychological factors and behavior are expected to be estimated, our work can also be considered as multi-task learning. In contrast to auxiliary learning which focuses solely on the performance of the main task, multi-task learning aims to develop a model that performs well on all assigned tasks (Chen et al., 2022). Current multi-task learning approaches primarily encompass multi-task architecture design, multi-task learning optimization, and multi-task relationship learning. Multi-task architecture design intends to create suitable architectures to determine which parameters to share across different tasks (Liu, Johns and Davison, 2019; Qin, Wang, Zhang and Zhu, 2021); multi-task optimization concentrates on optimizing architecture parameters using techniques like loss weighting (Kendall, Gal and Cipolla, 2018; Chen, Badrinarayanan, Lee and Rabinovich, 2018); and multi-task relationship learning focuses on acquiring explicit relationships among tasks (Zhao, Stretcu, Smola and Gordon, 2020).

In our study, according to existing research in the human factor and psychology domains, the relations between psychological factors and human decisions are usually linear (Wang et al., 2023a; Huang, Yan, Xie and He, 2024). Thus, we can linearly aggregate the gradient from all tasks with one dynamic-weight function. Further, because all auxiliary tasks concentrate on human-side feature (i.e., psychological features) modeling and are irrelevant to scenario-side features, we can utilize multi-task optimization methods to determine the gradient passing directions.

3. Methods and Materials

3.1. Architeture of Proposed Model

The proposed overall framework is shown in Figure 2, and the summary and descriptions of important notations in our framework are shown in Table 1. The framework of auxiliary-task learning for human decision or behavior prediction is composed of a **Human Side Module**, **Scenario Side Module**, **Auxiliary Tasks Module**, and **Main Task Module**, which can be trained in an end-to-end manner.

Table 1
A summary of symbols and descriptions

Symbol	Description
N	The number of individuals.
M	The number of different scenario conditions.
K	The number of psychological factors that are identified as significant to human decision.
U, u_i	The individual set.
S, s_i	The scenario condition set.
Y, y_i	The ground-truth set of human decision/behavior.
A, a_i	The set of significant psychological factors.
P, p_i	The predicted probability set.
E^u, e^u_i	The embedding vector of individuals.
E^s, e^s_{ij}	The embedding vector of discrete scenario conditions.
O^u, o^u_i	Encoded dense human-side representation.
O^s, o^s_i	Encoded dense scenario-side representation.
$O^{u'}, O^{u'}_i$	Decoded psychology-related individual representation.
O, o_i	Concatenated human-scenario interaction representation.
$f^u(U, \theta^u)$	The learnable human feature encoder function.
$f^{ad}(O^u, \theta^{ad})$	The learnable auxiliary feature decoders.
$f^s(s, \theta^s)$	The learnable scenario feature encoder function.
$f^{ah}_i(O^{u'}_i, \theta^{ah}_i)$	The learnable latent psychological feature head function.
$f^m(O, \theta^m)$	The learnable decision head function.

In general, the Human Side Module restores the original human features through a learnable embedding memory bank, which codes each human into a dense feature space. At the same time, the Scenario Side Module codes the scenario-related features into a dense representation. Following the Human Side Module, we further map the human features to multiple latent psychological factors in the Auxiliary Tasks Module. Then, we combine the dense feature from the Human Side Module and the Scenario Side Module to predict the decisions of the given subject in given scenarios. The Human Side module, Auxiliary Tasks Module, and Main Task Module were trained jointly based on the accuracies of behavior and psychological factor estimation; while the Scenario Side Module was updated based on the gradient from optimizing behavior prediction only, given that the features of the scenario are not affected by the subject features. With this structure, when subjects come, we can label these subjects (with the Human Side Module), estimate their psychological features (with the Auxiliary Tasks Module), and predict their decisions (with the Main Task Module), as long as we know some of their historical decisions in certain scenarios.

Specifically, for the main task: *how to predict one's decision/behavior in scenarios*, we want to learn a function $f(U, S; \theta)$, where $U = \{u_i\}_{i=1}^{N}$ is the individual set consists of N distinct individual identification, $S = \{s_i\}_{i=1}^{M}$ means there are M scenarios the individuals will encounter, and θ is the parameter of the model to optimize. Specifically, in practice, a scenario S is composed of M scenario conditions $S = \{s_i\}_{i=1}^{M}$, and each condition can be continuous or categorical. The final prediction objective is $Y = \{y_i\}_{i=1}^{N*M}$.

3.1.1. Human and Scenario Side Module

The backbone of each feature encoder is substitutable, and all main-stream feature encoders (e.g., convolutional neural network (Gu, Wang, Kuen, Ma, Shahroudy, Shuai, Liu, Wang, Wang, Cai et al., 2018), transformer (Vaswani, Shazeer, Parmar, Uszkoreit, Jones, Gomez, Kaiser and Polosukhin, 2017)) can be deployed in our framework. In our study, we instantiated it with the fundamental multi-layered perceptron (MLP) network (Rumelhart, Hinton, Williams et al., 1985). Its basic architecture contains three full-connected layers with activation functions.

Since different latent psychological feature learning tasks share similar fundamental information about human beings, we designed the human feature encoder $f^u(U; \theta^u)$ to extract general human-relevant features. Specifically, given the ID of one individual, u_i, we first queried the pre-initialized embedding bank $E^u \in \mathbb{R}^{N \times D}$ to get its corresponding untrained embedding vector $e^u_i \in \mathbb{R}^D$, where D is the dimension of hidden vector. The human feature extraction was formulated as an encoder-decoder network. Following Ng et al. (2011); Hinton and Salakhutdinov (2006), the human

feature extraction was formulated as an encoder-decoder network. At the **Human Side Module**, we utilized f^u to encode the original feature space to a dense feature space $o_i^u \in \mathbb{R}^{D/2}$.

The modeling of scenario conditions in **Scenario Side Module** runs in parallel with **Human Side Module**. Due to the varied data types of each s, we designed two processes corresponding to the categorical and continuous types of s. For each categorical condition, given one individual i, considering the nonlinear relations between different levels in one scenario condition s_j, we initialized an independent embedding bank to store the representation $e_{ij}^s \in \mathbb{R}^{D/M}$ of the condition s_j. Then, we concatenated M e_{ij}^s at the last dimension to one vector e_i^s. For M continuous conditions, we utilized the first layer of the scenario feature encoder $f^s(S;\theta^s)$ to map their original value to one united high-dimension vector $e_i^s \in \mathbb{R}^{D/2}$. Then, to convert tight patterns in the data to sparse points or lines in high-dimensional space to make the data linearly separable, inspired by the kernel function in SVM (Patle and Chouhan, 2013), the 1st layer of f^u mapped the e_i^u to a higher latent dimension D through the upsampling (Zuo, Fang, Yang, Shang and Wang, 2019). After the upsampling, we compress the high latent space to a dense representation $o_i^s \in \mathbb{R}^{D/2}$ with a lower dimension, to decrease computational complexity and alleviate nonlinear and non-normal situations in high-dimension space (Lee, Kwak, Tsui and Kim, 2019). Detailed structures varied in different datasets, and will be displayed in the Data Source section.

3.1.2. Auxiliary Tasks Module

The **Auxiliary Tasks Module** was built on top of the **Human Side Module** to exploit multi-view useful information in the decode stage, which leveraged auxiliary task annotations to refine individual features for primary task learning (Han, Zhang and Shan, 2020). Being different from previous data-driven methods, we explicitly inferred psychological factors and optimized θ^u through the auxiliary gradients.

Specifically, as there are K auxiliary tasks, K continuous psychological factors $A = \{a_i\}_{i=1}^K$ were identified as significant factors in human behaviors. Firstly, after receiving o_i^u, the auxiliary feature decoder $f^{ad}(O^u;\theta^{ad}) = \{f_i^{ad}(O^u;\theta_i^{ad})\}_{i=1}^K$ nonlinearly mapped the abstract human feature O^u to the multiple spaces $O^{u'} = \{O_i^{u'}\}_{i=1}^K, O_i^{u'} \in \mathbb{R}^{N\times D}$, which correspond to K significant psychological factors. Then, we initialized K independent latent factor heads, which took $O_i^{u'}$ as the input, respectively. Each head $f_i^{ah}(O_i^{u'};\theta_i^{ah})$ is in the format of a two-layered MLP without activation functions. Parameters of different latent factor heads and auxiliary feature decoders were only optimized by gradient calculated from their corresponding loss. The optimization procedure for each a_i is:

$$\begin{aligned} \hat{a}_i &= \mathrm{MLP}(o_i^{u'}), \\ \mathcal{L}_{ai}(\hat{a}_i, a_i) &= \begin{cases} 0.5 * (a_i - \hat{a}_i)^2, & if\ |a_i - \hat{a}_i| < 0.5 \\ |a_i - \hat{a}_i| - 0.5, & otherwise \end{cases} \\ , where\ i &= 1, 2, ..., K. \end{aligned} \tag{1}$$

3.1.3. Main Task Module and Joint Learning

Behavior prediction acts as the main task of our framework. To address the limited flexibility and non-linearity of the interaction learning by the fixed element-wise product in matrix factorization (MF), inspired by neural collaborative filtering (NCF) in recommendation (He et al., 2017), we aim to leverage MLP to express and generalize the user-scenario interaction MF.

In detail, after we got the representation from the human side o_i^u and scenario side o_i^s, we concatenated them into one representation space $O = \{o_i\}_{i=1}^{N*M}, O \in \mathbb{R}^{N*M\times D}$. Being different from the latent factor heads, we supplemented a nonlinear hidden layer with activation functions to the decision head $f^m(O;\theta^m)$, which extracted the class probability $P = \{p_i\}_{i=1}^{N*M}$ from nonlinear united representation space O. Further, we formulated the behavior prediction task as the classic binary classification, and the cross-entropy loss was utilized as the optimization goal. The loss function $\mathcal{L}_m$ is shown as:

$$\mathcal{L}_m(P, Y) = -\sum_{i=1}^{N*M} y_i \log p_i + (1 - y_i) \log(1 - p_i). \tag{2}$$

Compared to the NCF structure, we aim to not only learn the implicit features from the representation spaces of humans and scenarios, but can also effectively leverage the prior knowledge in the psychology domain through our

tailored explicit factor modeling. To achieve this goal, we jointly trained the **Human Side Module**, **Auxiliary Tasks Module**, and **Main Task Module** with full gradient as Eq.(3), and **Scenario Side Module** were optimized only through gradient from the primary task. The training process is shown in Algorithm 1.

$$\arg\min_{\theta^u,\theta^{ad},\theta^{ah},\theta^m} \mathcal{L}_m(P,Y) + \lambda(\sum_{i=1}^{K} w_i \mathcal{L}_{ai}(\hat{a}_i, a_i)). \tag{3}$$

To prevent irrelevant constraints on the network due to premature intervention of auxiliary tasks, we introduced an adaptation factor λ (Ganin and Lempitsky, 2015; Jia, Zhang, Shan and Chen, 2020). The overall loss also included K trade-off parameters (w_1 to w_K) to accelerate convergence speed and stabilize training.

Algorithm 1: **PEN**

Input:
Individual set: $U = \{u_i\}_{i=1}^{N}$
Scenario condition set: $S = \{s_i\}_{i=1}^{M}$
Ground-truth decision set: $Y = \{y_i\}_{i=1}^{N*M}$
Psychological factor set: $A = \{a_i\}_{i=1}^{K}$
Initialize the network $f(*)$.
do:
  Query embedding vector E^u.
  Encode E^u to dense representation O^u with $f^u(*)$.
  If the scenario condition is discrete:
    Query embedding vector E^s.
  Else if the scenario condition is continuous:
    Map the original value to nonlinear space E^s.
  Encode E^s to dense representation O^s with $f^s(*)$.
  Concatenate O^u and O^s to O.
  Decode O^u to K subspace $O_i^{u'}$ with $f_i^{ad}(*)$.
  For each auxiliary task:
    Estimate $\hat{a}_i$ with $O_i^{u'}$ through $f_i^{ah}(*)$.
    Compute $\mathcal{L}_{ai}$ with $\hat{a}_i$ and a_i as Eq.(1).
  Get predicted probability P with O through $f^m(*)$.
  Compute $\mathcal{L}_m$ with P and Y as Eq.(2).
  Update parameters $\theta^u, \theta^{ad}, \theta^{ah}$ as Eq.(3).
  Update parameters θ^s through gradient from $\mathcal{L}_m$.
until Converge

3.2. Data Source

Our proposed method was evaluated on two private datasets. To ensure there were adequate human-related psychological attributes, both of them were survey-based studies. Descriptive statistics of the two datasets are provided in Table 2.

3.2.1. Charging Behavior

In general, Wang et al. (2023a) conducted a survey aiming at gathering data on the charging preferences of battery electric vehicle (BEV) drivers in mainland China under various scenarios. These scenarios incorporated three within-subject factors (i.e., Waiting Time, Rest Battery, and Rest Trip, leading to 60 scenarios in total), in order to reveal how distance- and time-related anxieties can affect the charging behaviors of BEV users. Participants were recruited through online platforms. A total of 230 valid responses were obtained in the survey study, comprising 175 males and 55 females, leading to 13,800 charging decisions in different scenarios (230 participants * 60 scenarios). Details of the questionnaire can be found in Wang et al. (2023a,c). At the same time, some psychological factors of the respondents were collected, for example, their Comfort Range (i.e., the users' range comfort zone or range safety buffer), Comfort Time (i.e., BEV users' acceptance of waiting time at the charging stations), personality and attitudes towards the BEVs. In Wang et al. (2023a,c), we found that both distance and waiting time matter to BEV drivers' charging decisions. Thus, we used both **Comfort Range** and **Comfort Time** as range-anxiety-related psychological auxiliary factors.

Table 2
Descriptive Statistics of the Datasets.

Charging Behavior (Wang et al., 2023a)		**In-vehicle Interaction (Huang et al., 2024)**	
Attribute	Statistic/Description	Attribute	Statistic/Description
Sample Num	13800	Sample Num	8064
Participant	230	Participant	168
Gender	Female (24.0%) Male (76.0%)	Gender	Female (36.3%) Male (63.7%)
Age	Mean: 26.4 (SD: 4.7, min: 18, max: 40)	Age	Mean: 29.5 (SD: 5.1, min: 19, max: 46)
Comfort Range	mean: 47.3% (SD: 16.4, min: 25, max: 100)	Technology Attitude	mean: 8.9 (SD: 1.3, min: 3, max: 10)
Comfort Time	mean: 43.8 minutes (SD: 33.6, min: 0, max: 100)	Passenger Presence	1. driving alone 2. driving with a colleague
Waiting Time	mean: 37.1 minutes (SD: 26.0, min: 0, max: 99)	Roadway Type	1. urban road 2. traffic complexity is high
Rest Battery	mean: 49.9 % (SD: 19.4, min: 20, max: 79)	Task Type	1. adjust the temperature in the vehicle to a specific temperature 2. select a specific song in the favorite music list to play
Rest Trip	mean: 58.0 % (SD: 22.0, min: 15, max: 94)		3. make a phone call to a specific person in the contact list 4. compile and send a text message 5. navigate to a specific destination 6. open or close the window
		Task Complexity	1. traffic complexity is low 2. traffic complexity is high
Decision	Charge at the destination (55.7%) Charge at the upcoming rest area (44.3%)	Decision	VCS (37.1%) MIs (62.9%)

3.2.2. In-vehicle Interaction

The main focus of Huang et al. (2024) is to examine when drivers choose to use vehicle control systems (VCS) over manual interaction (MIs) when conducting six common in-vehicle tasks in different scenarios as defined by three within-subject factors (i.e., 2 types of passenger presence, 2 types of roadway, and 2 types of task complexity, leading to 8 scenarios in total). An online questionnaire was designed to gather drivers' demographic information, driving-related information, interaction-related information, and perceived distraction levels in six in-vehicle tasks. The respondents also indicated their preferred interaction methods in various scenarios. In total, 168 valid responses were collected, leading to 8064 choices (168 participants * 8 scenarios * 6 tasks). As demonstrated in (Lee, Mehler, Reimer and Coughlin, 2015; Chakraborty, 2010; Sokol, Chen and Donmez, 2017), drivers' adoption of in-vehicle VCS can be significantly influenced by their attitudes toward the technologies (i.e., perceived usefulness and perceived ease-of-use of the technology). Thus, the **Technology Attitude** was explicitly modeled as an auxiliary factor in our PEN framework to assist the primary task.

4. Experiment

4.1. Experiment Setting

4.1.1. Baselines

In this study, we compared our proposed framework with eight classic data-driven methods to elaborate on the superiority of PEN.

- Logistic Regression (LR) (Hosmer Jr et al., 2013) is a statistical modeling technique that utilizes a generalized linear regression model to estimate the probability of a sample belonging to a specific class. It is commonly adopted for classification tasks or statistical data analysis.
- Support Vector Machine (SVM) (Steinwart and Christmann, 2008) is a machine learning algorithm that analyzes data and constructs a decision boundary, known as a hyperplane, to classify or predict new instances. It aims to find the optimal hyperplane that maximally separates different classes or groups in the data space.
- Decision Tree (DT) (Safavian and Landgrebe, 1991) is a predictive modeling method that employs a tree-like structure to make predictions. Each internal node in the tree represents a decision based on a specific feature, while the leaf nodes correspond to the final outcomes or predictions.
- Random Forest (RF) (Breiman, 2001) is an ensemble learning technique that combines multiple decision tree classifiers to produce predictions. It aggregates the results from individual trees to make more accurate predictions.

- Adaboost (Freund and Schapire, 1997) is a prominent ensemble learning algorithm, which integrates multiple weak classifiers to establish a potent classifier. Its iterative process entails training weak classifiers on diverse subsets of the data, wherein subsequent classifiers assign increased emphasis to the misclassified samples identified by the preceding classifier.

- LightGBM (Ke et al., 2017) is a gradient-boosting framework that utilizes a tree-based learning algorithm. It incorporates innovative techniques such as Gradient-based Single-Side Sampling and Exclusive Feature Bundling to improve performance in various machine learning tasks.

- Multi-Layered Perception (MLP) (Jordan et al., 1995) is a deep learning algorithm inspired by the structure and function of the human brain. It consists of interconnected nodes, or neurons, organized into multiple layers. MLP processes and transforms input data through these layers to generate outputs.

- Neural Collaborative Filtering (NCF) (He et al., 2017) combines the strengths of both collaborative filtering and deep learning by incorporating a deep neural architecture. It typically consists of multiple layers of feedforward neural networks that capture the non-linear relationships between users and items. These networks are trained to predict user-item preferences or ratings based on historical data.

- NCRAE (Zhang and Liu, 2021) is a novel neural collaborative recommendation system that utilizes memory networks to learn attentional embedding from implicit interactions. It allows for the identification of the relative importance of various users and items through sequences of user-item interactions

- Neural Embedding Factorization Machine (NEFM) (Huang, 2023) efficiently initialized the embedding layers using an unsupervised pre-training method of probabilistic auto-encoder. The NEFM combines the strengths of the FM models in modeling second-order feature interactions and the neural networks in modeling non-linear feature interactions.

In practice, selected baselines are not able to model psychological states when predicting one's behaviors. To ensure the fair comparison, we directly input psychological factors into models as extra data attributes except for NCF, NCRAE, and NEFM. As they were restricted by their original structure, we can only take the individual-scenario interaction matrix as input. Additionally, compared to baselines which need extra psychological factors as input in the test stage, our framework no longer depends on psychological factors for the prediction of one's behaviors or decisions.

4.1.2. Implementation Details

Our model was implemented on the Pytorch framework in Python 3.8. The nonlinear activation functions were instantiated as Relu (Agarap, 2018). Additionally, the proper hidden dimension D size was selected from [64, 128, 256] according to the experimental results. Furthermore, we set hyper-parameters w_i to 0.01. The Adam optimizer (Kingma and Ba, 2014) was chosen to optimize parameters. The machine learning model construction, training, and hyper-parameter tuning were conducted using the Scikit-learn library (Pedregosa, Varoquaux, Gramfort, Michel, Thirion, Grisel, Blondel, Prettenhofer, Weiss, Dubourg et al., 2011). For evaluation metrics, we employed widely-used classification metrics, including Accuracy (ACC), F1 score (F1), and area under the receiver operating characteristic curve (AUC) score. We presented the average on 5-times independent tests with different random seeds for each model. Finally, the paired Student t-test was applied to check the significance of the performance difference between the best model and the second-best model.

4.2. Comparison Experiment

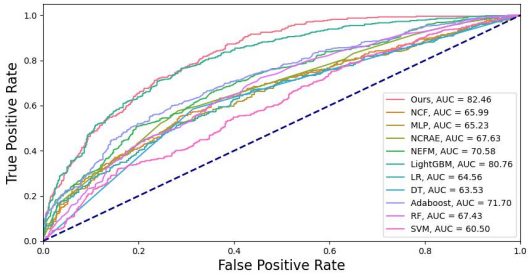
In previous studies (Ullah et al., 2023; Xing et al., 2021; Wang et al., 2019), researchers assumed the size of the test dataset to be smaller than the training data. However, the real-world scenario often violates this assumption. Particularly, considering the difficulty in collecting latent psychological characteristics, available training data will be limited. Therefore, we constructed three evaluation protocols according to the ratio of the training-test dataset partition: (1) 10% data for testing; (2) 50% data for testing; (3) 90% data for testing.

As shown in Table 3 and Figure 3, firstly, we found that LightGBM achieved the second-best performance in 10% and 50% testing protocols, while Adaboost was better in 90% testing protocols. Particularly, in Figure 3(c)(f), we found the AUC score of two classic DL-based baselines (MLP, NCF) was below 50%. It is possible that insufficient training under 90% testing protocols has led to the underfitting of non-linear learning modules. In contrast, Adaboost

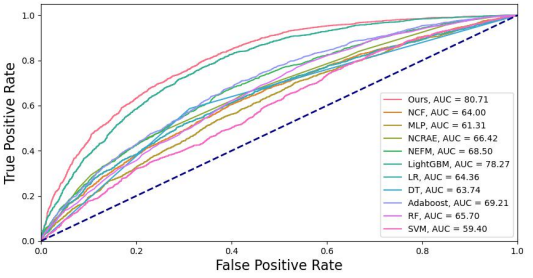
Table 3
Results of Comparison Experiments over Two Datasets with Three Evaluation Protocols.

	Charge Behavior (Wang et al., 2023a)						In-vehicle Interaction (Huang et al., 2024)					
	10%		50%		90%		10%		50%		90%	
Methods	**ACC(%)**	**F1(%)**	**ACC(%)**	**F1(%)**	**ACC(%)**	**F1(%)**	**ACC(%)**	**F1(%)**	**ACC(%)**	**F1(%)**	**ACC(%)**	**F1(%)**
LR	62.53	69.85	61.91	68.99	61.08	68.76	66.04	76.70	63.91	75.36	64.46	75.90
SVM	59.30	71.05	57.86	69.78	56.15	61.92	64.31	78.28	62.89	77.22	62.78	77.13
DT	64.10	68.01	64.34	68.35	59.41	63.51	72.24	78.37	71.40	77.07	62.27	69.73
RF	62.09	66.92	60.53	65.77	60.75	66.01	62.08	70.80	61.03	69.59	59.97	69.18
Adaboost	67.42	73.14	63.97	70.51	62.22	69.34	71.37	80.10	68.92	78.43	67.52	77.15
LightGBM	73.10	76.19	71.06	74.82	61.62	66.23	80.04	85.16	74.60	81.01	67.00	74.96
MLP	62.70	73.42	58.91	71.01	50.74	62.11	65.30	75.99	63.39	75.47	59.84	73.72
NCF	63.41	68.80	61.13	71.01	54.13	61.56	66.30	78.17	63.79	75.35	58.34	72.49
NCRAE	65.07	70.58	63.73	70.80	61.18	66.94	67.53	74.66	66.07	74.73	64.25	72.85
NEFM	66.81	72.66	63.24	70.01	62.05	69.54	70.76	79.73	67.26	77.64	66.88	76.99
Ours	**74.15***	**78.74***	**73.34***	**76.82***	**64.09***	**69.59***	**85.50***	**88.87***	**80.85***	**85.12***	**72.20***	**78.76***

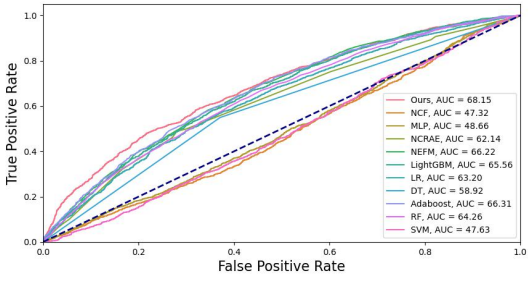
Notes: In this and following tables, * indicates the significant difference ($p < .05$) between our method and the best baseline, and the best results are bolded in **black**



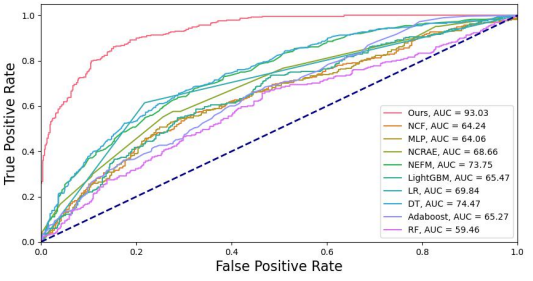
(a) Charge Behavior with 10% test data



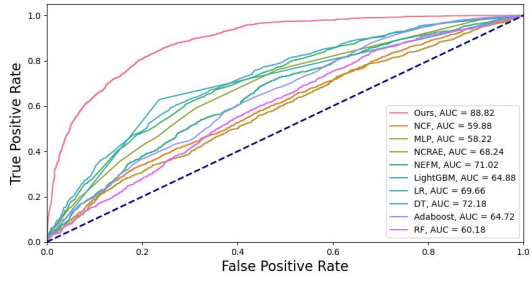
(b) Charge Behavior with 50% test data



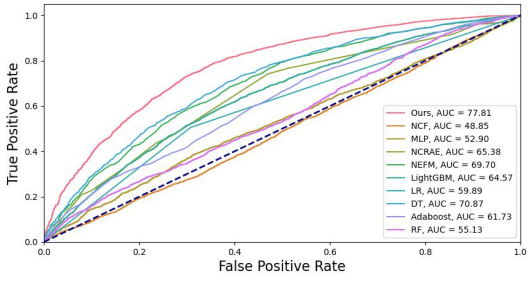
(c) Charge Behavior with 90% test data



(d) In-vehicle Interaction with 10% test data



(e) In-vehicle Interaction with 50% test data



(f) In-vehicle Interaction with 90% test data

Figure 3: The ROC curve and AUC score (%) of baselines and PEN under two datasets and three evaluation protocols.

and LightGBM are both ensembled tree-based boosting models. Their relatively better performance than generalized linear models (i.e., LR) or classic DL-based models (i.e., MLP, NCF) illustrates that the tree-based boosting models

Table 4
Results of Ablation Test on Charging Behavior Dataset.

Protocol	Variants	ACC(%)↑	F1(%)↑	MAE a_1↓	MAE a_2↓
10%	**Ours**	**74.15***	**78.74***	**12.05***	**27.15***
	w/o $\mathcal{L}_{a1}$	73.01	78.07	47.67	27.97
	w/o $\mathcal{L}_{a2}$	73.97	78.21	12.65	44.81
	Base	72.06	77.29	47.67	44.83
50%	**Ours**	**73.34***	**76.82***	**11.87**	**27.46***
	w/o $\mathcal{L}_{a1}$	73.07	76.15	47.09	27.83
	w/o $\mathcal{L}_{a2}$	72.76	75.84	11.87	44.35
	Base	72.57	74.72	47.10	44.34
90%	**Ours**	**64.09***	**69.59***	**11.70**	**26.86**
	w/o $\mathcal{L}_{a1}$	63.64	68.88	11.81	43.56
	w/o $\mathcal{L}_{a2}$	63.13	68.78	47.13	26.93
	Base	62.13	67.28	47.15	43.58

Notes: a_1 refers to the **Comfort Range** variable, and a_2 is the **Waiting Time** variable. In this and following tables, Base means the backbone network without $\mathcal{L}_{a1}$ and $\mathcal{L}_{a2}$.

Table 5
Results of Ablation Test on In-vehicle Interaction Dataset.

Protocol	Variants	ACC(%)↑	F1(%)↑	MAE a_1↓
10%	**Ours**	**85.50***	**88.87***	**1.02***
	w/o $\mathcal{L}_{a1}$	73.01	78.07	8.54
50%	**Ours**	**80.85***	**85.12***	**2.00***
	w/o $\mathcal{L}_{a1}$	78.10	83.21	8.56
90%	**Ours**	**72.20***	**78.76***	**4.44***
	w/o $\mathcal{L}_{a1}$	71.49	78.2	8.76

Notes: a_1 refers to the **Technology attitude** variable. In this and following tables, Base means the backbone network without $\mathcal{L}_{a1}$.

have better learning ability than conventional DL-based baselines on tabular data. It is worth noticing that, advanced DL-based models (i.e., NCRAE, NEFM) presented competitive performance compared to LightGBM and Adaboost, which illustrates their unique designs (embedding memory bank in NCRAE, hierarchy modeling in NEFM) brought notable performance gain.

Despite the higher score achieved by tree-based boosting models, the significantly superior performance of our proposed PEN framework was illustrated as well. Regardless of the disparities of datasets and train-test ratio, the PEN always achieved the highest accuracy, F1 score, and AUC score, which proves the effectiveness of our framework design. Particularly, compared to state-of-the-art DL-based baselines (i.e., NCRAE, NEFM), the proposed PEN significantly improved the performance, which suggests that the absence of explicit psychological factor inference in traditional data-driven methods leads to declined model performance. Additionally, with the increase in test data volume, all models showed a decreasing trend in their performance. Among them, the accuracies of DL-based baselines were even reduced to below 50%. By contrast, as a DL-based model, our proposed PEN still had the highest score in all metrics. This indicates that the explicit modeling of the intermediate psychological process can promote generalizability in small-sample learning scenarios, and achieve no dependency on additional input data (i.e., psychological factors) during the testing phase.

Finally, it should be noted that, all baseline models included psychological factors as inputs in our experiments, which is usually not the case in real-world scenarios. We did not include the baseline models without psychological factors, as removing psychological factors from the inputs leads to even worse performance under all testing protocols. Though removing the psychological factors from inputs of PEN leads to unfair comparisons between the PEN and other baselines, the superior performance of PEN highlights the importance of considering the heterogeneity of psychological states in human beings when predicting their behaviors.

4.3. Ablation Test

In two datasets, we instantiated varied psychological factors as auxiliary tasks. To further validate the key idea of the PEN framework, we conducted an ablation test to compare the performance of several variants without all or part of the auxiliary tasks and the full-version PEN.

As shown in Table 4,5, we first found that the full-version PEN outperformed other variants in both two datasets and three protocols, which illustrates explicit factor modeling of psychological factors is beneficial to decision or behavior pattern prediction. Specifically, refer to Table 4, for the Charging Behavior dataset, compared to the Base model (i.e., the backbone network without two auxiliary tasks), the performance of two variants was better but still worse than the full version. In the In-vehicle Interaction dataset (Table 5), there was only one psychological factor. Similarly, the variant without auxiliary tasks significantly performed worse than the full version. In general, it can be concluded that it is effective and necessary to consider explicit psychological latent factor inference in behavior prediction models.

Furthermore, we also present the factor inference precision in Table 4,5. The mean absolute error (MAE) was utilized for evaluation. Obviously, considering the loss of corresponding psychological factors as extra regularizations can promote the inference capacity of auxiliary tasks under different protocols - these estimated psychological factors can be used for decision interpretation.

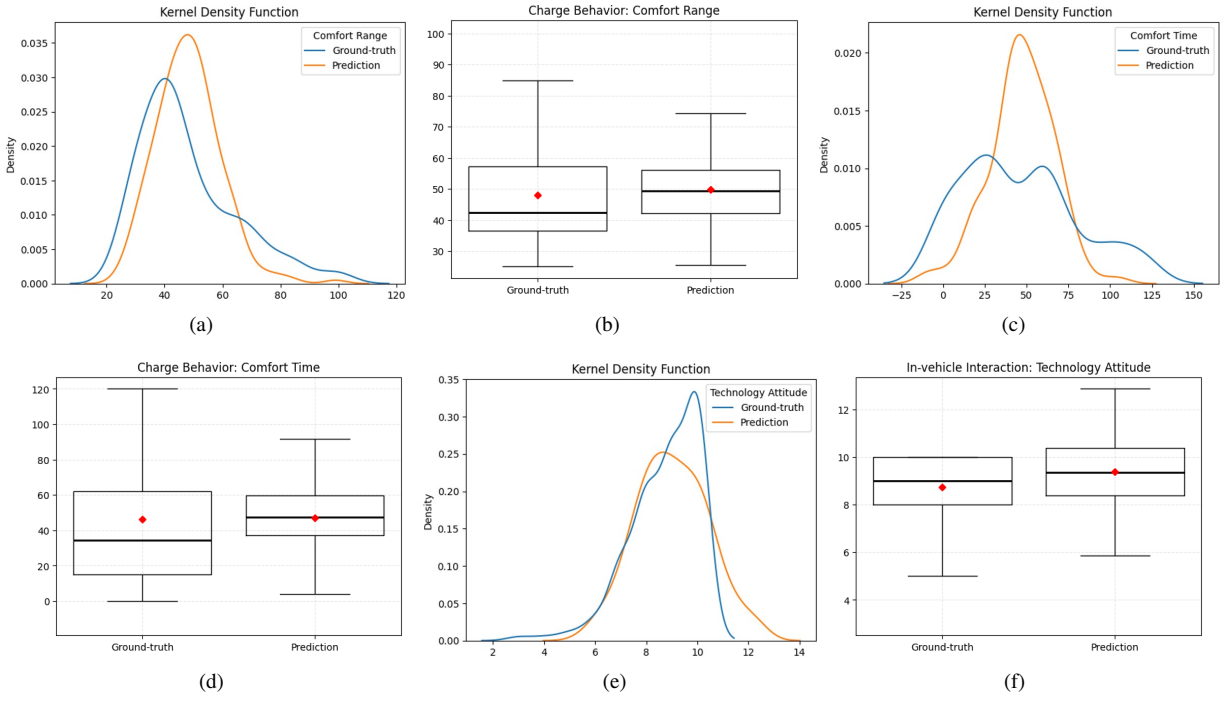


Figure 4: The descriptive statistics of the ground-truth and predicted psychological factor under 50% evaluation protocol. The kernel density estimation and box plot are both presented. Subfigure (a)(b) belong to the **Comfort Range** in the Charging Behavior dataset; (c)(d) belong to **Comfort Time** in the Charging Behavior dataset; and (e)(f) is **Technology Attitude** in the In-vehicle Interaction dataset.

4.4. Case Study: Factor Statistical Analysis

To evaluate whether the psychological variables learned by our model can recover the distribution of these psychological factors obtained from previous human factors research, we conducted a case study aiming to examine the distribution discrepancy between predicted psychological factors and the ground truth. Specifically, we normalized the range of the prediction of each psychological factor to the same numerical space as the ground truth. The case study was based on the 50% protocol, the kernel density estimation and the box plot were used to visualize the distribution of estimated and ground-truth psychological factors, and the paired T-test was leveraged to examine the discrepancy. The results are shown in Figure 4.

Specifically, a case study was conducted based on the 50% protocol. The kernel density estimation and the box plots were used to visualize the distribution of estimated and ground-truth psychological factors, and the paired t-test was conducted to examine the discrepancies. As shown in Figure 4, for the **Comfort Range**, we did not find a significant difference between the estimated value and the ground truth (Figure 4(b), p=.9), and the shape of the probability density curve is similar (Figure 4(a)). At the same time, the box plots show that the ground-truth distribution has a higher variance, which has also been observed in **Comfort Time** (Figure 4(c)(d), p=.9). This phenomenon illustrates that the present method can successfully recover the distribution of psychological variables for the general population, but the models are still not able to recover the long-tailed psychological state. For the **Technology Attitude**, as shown in Figure 4(e)(f), we found that although the shape of the predicted distribution is similar to the ground truth and the difference between their mean value is non-significant (p=.8), the distribution of the ground-truth value is denser compared to the ground truth variable. Particularly, when the **Technology Attitude** increases to a higher value, there were not sufficient samples in the ground truth, while the distribution shape of prediction still covers this area. This is due to the fact that the output of the model is continuous space, whereas the self-reported ground truth usually has numerical boundaries, and the range of predicted values will inevitably exceed the range of the ground truth. However, we believe that as long as the distributions are approximate, this will not matter in practice, since self-reported values are inherently relative. As shown in Table 5, the performance degradation after removing the auxiliary tasks in the In-vehicle Interaction dataset was higher than Charging Behavior, thus we believe the discrepancy in the value range was still acceptable.

5. Disscussion

Throughout our experiments, we successfully answered all three research questions. For RQ1, based on the results of the comparison experiment on two datasets, the superior performance of our proposed PEN framework proves the effectiveness of introducing prior psychology knowledge into data-driven human behavior prediction. At the same time, agreeing with previous research in other domains (Zhang and Liu, 2021; Ng et al., 2011; Hinton and Salakhutdinov, 2006), the embedding memory bank and encoder-decoder structure in the human-scenario side module increased the robustness of the models. The effectiveness of these structures has also been validated in our proposed PEN framework. The capability of the Auxiliary Module in recovering the distribution of the latent psychological models also proved the interpretability of the PEN framework, which is superior to previous data-driven only deep learning models.

For RQ2, the statistical analyses in psychological and human factors research revealed the importance of latent cognitive processes in shaping human behaviors. Through our proposed auxiliary task learning, key psychological factors were explicitly modeled. According to the results of our ablation study and case study, explicit psychological factor modeling can benefit the performance of human behavior prediction, potentially by mimicking the human cognitive process in the decision-making process. Therefore, compared to previous data-driven methods, PEN can enhance the interpretability of data-driven deep models by making personalized psychological inferences based on specific scenario conditions, and outputting the inferred mental states at the same time. Practically, the revealed psychological features may further be used for one's behavior prediction in other scenarios where similar psychological features may play a role.

Lastly, for RQ3, we found PEN still outperformed all baselines under 90% protocols, which illustrated the generalizability of our proposal. This is an important characteristic of a model, as psychological variables are usually unavailable in real-world application settings. Our research found that the PEN, with the Auxiliary Module, may still learn one's psychological features and be generalized to some extent to larger groups with limited historical behavioral data only. It is possible that human behavioral patterns are usually influenced by several key psychological variables (Wang et al., 2023a; Huang et al., 2024), and the statistically significant features are reliable and can optimize the learning process in a non-data-driven way (Wang et al., 2023a; Huang et al., 2024). Therefore, even learning from only a small number of samples, the models can still fit a larger population as long as the psychological theories apply to the population."

6. Limitation

It should be noted that, there are several limitations of our work, and they may be addressed by future research. First, restricted by the sample size of survey studies, some more advanced feature encoding networks (e.g., transformer (Vaswani et al., 2017) and convolutional neural network (Gu et al., 2018)) cannot be applied. Future research should

be conducted to validate our model and make structure improvements when large-scale datasets become available. At the same time, although explicitly estimating the psychological factors can improve the interpretability and precision of DL-based data-driven methods, we were not able to recover the long-tailed values of some psychological factors, indicating that the auxiliary module may not be able to cover the minorities in the population, which deserves further exploration.

7. Conclusion

While researchers have realized the role of psychological factors in affecting one's behavior or decisions under various scenarios, little attention has been paid to aggregating knowledge from both human factors or psychology and data-driven domains to enhance the predicting capacity and interpretability of the models. This article aims to tackle this issue by proposing a novel psychology-powered data-driven behavior prediction framework (**PEN**). In summary, based on the architecture of neural collaborative filtering (He et al., 2017), we incorporated multiple auxiliary feature decoders and auxiliary estimation heads to explicitly learn and further estimate psychological factors that are significantly associated with human behavior. Then, inspired by multi-task learning (Han et al., 2020), we precisely controlled the gradient passing path to a specific module to ensure the psychological factor modeling does not disturb the primary decision prediction, but enriches human-side latent feature modeling. Through experiments, we demonstrate that our method has yielded substantial improvements on two datasets and under three train-test data partition protocols. The results of the ablation test also validated the effectiveness of the proposed architecture and the principle of psychology-powered learning. Further, a case study was conducted to check the distribution discrepancy between the ground truth and the estimated psychological factors.

In conclusion, our work serves as a first step to combine deep learning and the psychology or the human factors domains. Future human behavior modeling or decision prediction work can consider combining prior knowledge with a data-driven approach to increase the interpretability and generalizability of the model. For example, when modeling drivers' behaviors in traffic, the human factors domain has revealed many factors that can shape one's decision (Wang et al., 2023c; Wang, Huang, Tu and He, 2023b), but the traffic domain mostly relies on data-driven models (e.g., (Wen, Jian and He, 2023)). With our framework, the knowledge from both can be combined, which may yield higher model prediction accuracies that may benefit autonomous vehicle design and traffic management.

References

Adeniyi, D.A., Wei, Z., Yongquan, Y., 2016. Automated web usage data mining and recommendation system using k-nearest neighbor (knn) classification method. Applied Computing and Informatics 12, 90–108.

Agarap, A.F., 2018. Deep learning using rectified linear units (relu). arXiv preprint arXiv:1803.08375 .

Ahmed, F., Asif, R., Hina, S., Muzammil, M., 2017. Financial market prediction using google trends. International Journal of Advanced Computer Science and Applications 8, 388–391.

Ajzen, I., 1991. The theory of planned behavior. Organizational behavior and human decision processes 50, 179–211.

Ariely, D., Bracha, A., Meier, S., 2009. Doing good or doing well? image motivation and monetary incentives in behaving prosocially. American economic review 99, 544–555.

Breiman, L., 2001. Random forests. Machine learning 45, 5–32.

Cai, W., Jin, Y., Chen, L., 2022. Impacts of personal characteristics on user trust in conversational recommender systems, in: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, pp. 1–14.

Chakraborty, J., 2010. An exploration of predictors of adoption and use of in-vehicle voice control systems. University of Toronto (Canada).

Chen, H., Wang, X., Guan, C., Liu, Y., Zhu, W., 2022. Auxiliary learning with joint task and data scheduling, in: International conference on machine learning, PMLR. pp. 3634–3647.

Chen, Z., Badrinarayanan, V., Lee, C.Y., Rabinovich, A., 2018. Gradnorm: Gradient normalization for adaptive loss balancing in deep multitask networks, in: International conference on machine learning, PMLR. pp. 794–803.

Cheng, H., Fang, H., Ostendorf, M., 2015. Open-domain name error detection using a multi-task rnn, in: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pp. 737–746.

Chung, Y.W., Khaki, B., Li, T., Chu, C., Gadh, R., 2019. Ensemble machine learning-based algorithm for electric vehicle user behavior prediction. Applied Energy 254, 113732.

Eke, C.I., Norman, A.A., Shuib, L., Nweke, H.F., 2019. A survey of user profiling: State-of-the-art, challenges, and solutions. IEEE Access 7, 144907–144924.

Frendo, O., Graf, J., Gaertner, N., Stuckenschmidt, H., 2020. Data-driven smart charging for heterogeneous electric vehicle fleets. Energy and AI 1, 100007.

Freund, Y., Schapire, R.E., 1997. A decision-theoretic generalization of on-line learning and an application to boosting. Journal of computer and system sciences 55, 119–139.

Ganin, Y., Lempitsky, V., 2015. Unsupervised domain adaptation by backpropagation, in: International conference on machine learning, PMLR. pp. 1180–1189.

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J., et al., 2018. Recent advances in convolutional neural networks. Pattern recognition 77, 354–377.

Gulhane, M., Sajana, T., 2021. Human behavior prediction and analysis using machine learning-a review. Turkish Journal of Computer and Mathematics Education (Turcomat) 12, 870–876.

Han, D., Zhang, J., Shan, S., 2020. Leveraging auxiliary tasks for height and weight estimation by multi task learning, in: 2020 IEEE International Joint Conference on Biometrics (IJCB), IEEE. pp. 1–7.

He, X., Liao, L., Zhang, H., Nie, L., Hu, X., Chua, T.S., 2017. Neural collaborative filtering, in: Proceedings of the 26th international conference on world wide web, pp. 173–182.

Hinton, G.E., Salakhutdinov, R.R., 2006. Reducing the dimensionality of data with neural networks. science 313, 504–507.

Hosmer Jr, D.W., Lemeshow, S., Sturdivant, R.X., 2013. Applied logistic regression. volume 398. John Wiley & Sons.

Huang, C., Yan, S., Xie, W., He, D., 2024. When do users prefer voice-controlled systems in vehicles? a survey of chinese drivers. Transportation Research Record .

Huang, T., 2023. Nefm: Neural embedding based factorization machines for user response prediction. Expert Systems with Applications 216, 119433. doi:https://doi.org/10.1016/j.eswa.2022.119433.

Jia, Y., Zhang, J., Shan, S., Chen, X., 2020. Single-side domain generalization for face anti-spoofing, in: Proc. CVPR, pp. 8484–8493.

Jordan, M.I., et al., 1995. Why the logistic function? a tutorial discussion on probabilities and neural networks.

Jung, S.Y., Hong, J.H., Kim, T.S., 2005. A statistical model for user preference. IEEE Transactions on Knowledge and Data Engineering 17, 834–843.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.Y., 2017. Lightgbm: A highly efficient gradient boosting decision tree. Advances in neural information processing systems 30.

Kendall, A., Gal, Y., Cipolla, R., 2018. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 7482–7491.

Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 .

Klema, V., Laub, A., 1980. The singular value decomposition: Its computation and some applications. IEEE Transactions on automatic control 25, 164–176.

Koren, Y., Bell, R., Volinsky, C., 2009. Matrix factorization techniques for recommender systems. Computer 42, 30–37.

Lee, C., Mehler, B., Reimer, B., Coughlin, J.F., 2015. User perceptions toward in-vehicle technologies: Relationships to age, health, preconceptions, and hands-on experience. International Journal of Human-Computer Interaction 31, 667–681.

Lee, S., Kwak, M., Tsui, K.L., Kim, S.B., 2019. Process monitoring using variational autoencoder for high-dimensional nonlinear processes. Engineering Applications of Artificial Intelligence 83, 13–27.

Leevy, J.L., Khoshgoftaar, T.M., Bauder, R.A., Seliya, N., 2018. A survey on addressing high-class imbalance in big data. Journal of Big Data 5, 1–30.

Lin, X., Baweja, H., Kantor, G., Held, D., 2019. Adaptive auxiliary task weighting for reinforcement learning. Advances in neural information processing systems 32.

Liu, H., Lang, B., 2019. Machine learning and deep learning methods for intrusion detection systems: A survey. applied sciences 9, 4396.

Liu, S., Johns, E., Davison, A.J., 2019. End-to-end multi-task learning with attention, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 1871–1880.

Ng, A., et al., 2011. Sparse autoencoder. CS294A Lecture notes 72, 1–19.

Patle, A., Chouhan, D.S., 2013. Svm kernel functions for classification, in: 2013 International conference on advances in technology and engineering (ICATE), IEEE. pp. 1–9.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al., 2011. Scikit-learn: Machine learning in python. the Journal of machine Learning research 12, 2825–2830.

Qin, Y., Wang, X., Zhang, Z., Zhu, W., 2021. Graph differentiable architecture search with structure learning. Advances in neural information processing systems 34, 16860–16872.

Rumelhart, D.E., Hinton, G.E., Williams, R.J., et al., 1985. Learning internal representations by error propagation.

Safavian, S.R., Landgrebe, D., 1991. A survey of decision tree classifier methodology. IEEE transactions on systems, man, and cybernetics 21, 660–674.

Shi, B., Hoffman, J., Saenko, K., Darrell, T., Xu, H., 2020. Auxiliary task reweighting for minimum-data learning. Advances in Neural Information Processing Systems 33, 7148–7160.

Sokol, N., Chen, E.Y., Donmez, B., 2017. Voice-controlled in-vehicle systems: Effects of voice-recognition accuracy in the presence of background noise, in: Driving Assesment Conference, University of Iowa.

Steinwart, I., Christmann, A., 2008. Support vector machines. Springer Science & Business Media.

SUN, Q.m., QU, Z.j., REN, C.g., 2021. Context-aware multi-modal transportation recommendation based on particle swarm optimization and lightgbm. Acta Electonica Sinica 49, 894.

Tian, H., Sun, Y., Hu, F., Du, J., 2023. Charging behavior analysis based on operation data of private bev customers in beijing. Electronics 12, 373.

Ullah, I., Liu, K., Yamamoto, T., Zahid, M., Jamal, A., 2023. Modeling of machine learning with shap approach for electric vehicle charging station choice behavior prediction. Travel Behaviour and Society 31, 78–92.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. Advances in neural information processing systems 30.

Wang, A., Guan, H., Guo, J., Han, Y., Bian, H., 2022. An empirical study on the segmentation of potential users of shared parking spaces considering individual heterogeneity. Journal of Advanced Transportation 2022, 1–18.

Wang, A., Guan, H., Han, Y., Cao, Y., 2021a. Private parking space sharing intention in china: An empirical study based on the mimic model. Discrete Dynamics in Nature and Society 2021, 1–13.

Wang, J., Dong, Z., Chen, L., 2021b. An multi-aspect attentional model to capture multistratal influence in social group, in: 2021 International Conference on Electronic Information Engineering and Computer Science (EIECS), IEEE. pp. 315–320.

Wang, J., Huang, C., He, D., Tu, R., 2023a. Range anxiety among battery electric vehicle users: Both distance and waiting time matter. arXiv preprint arXiv:2306.05768 .

Wang, J., Huang, C., Tu, R., He, D., 2023b. Influential factors of users' trust in the range estimation systems of battery electric vehicles–a survey study in china. arXiv preprint arXiv:2301.10076 .

Wang, J., Kong, X., Xia, F., Sun, L., 2019. Urban human mobility: Data-driven modeling and prediction. ACM SIGKDD explorations newsletter 21, 1–19.

Wang, J., Zhang, W., Wen, X.L., He, D., Tu, R., 2023c. Modeling battery electric vehicle users' charging decisions in scenarios with both time-related and distance-related anxiety. in processing .

Wen, X., Jian, S., He, D., 2023. Modeling the effects of autonomous vehicles on human driver car-following behaviors using inverse reinforcement learning. IEEE Transactions on Intelligent Transportation Systems .

Wu, L., He, X., Wang, X., Zhang, K., Wang, M., 2022. A survey on accuracy-oriented neural recommendation: From collaborative filtering to information-rich recommendation. IEEE Transactions on Knowledge and Data Engineering 35, 4425–4445.

Wu, Y., Li, K., Zhao, G., Qian, X., 2020. Personalized long-and short-term preference learning for next poi recommendation. IEEE Transactions on Knowledge and Data Engineering 34, 1944–1957.

Xing, Q., Chen, Z., Zhang, Z., Wang, R., Zhang, T., 2021. Modelling driving and charging behaviours of electric vehicles using a data-driven approach combined with behavioural economics theory. Journal of Cleaner Production 324, 129243.

Yu, J., Jiang, J., 2016. Learning sentence embeddings with auxiliary tasks for cross-domain sentiment classification, in: Proceedings of the 2016 conference on empirical methods in natural language processing, pp. 236–246.

Zhai, X., Oliver, A., Kolesnikov, A., Beyer, L., 2019. S4l: Self-supervised semi-supervised learning, in: Proceedings of the IEEE/CVF international conference on computer vision, pp. 1476–1485.

Zhang, Y., Liu, X., 2021. Learning attention embeddings based on memory networks for neural collaborative recommendation. Expert Systems with Applications 183, 115439.

Zhao, H., Stretcu, O., Smola, A.J., Gordon, G.J., 2020. Efficient multitask feature and relationship learning, in: Uncertainty in Artificial Intelligence, PMLR. pp. 777–787.

Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A., 2016. Learning deep features for discriminative localization, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2921–2929.

Zuo, Y., Fang, Y., Yang, Y., Shang, X., Wang, B., 2019. Residual dense network for intensity-guided depth map enhancement. Information Sciences 495, 52–64.