

Behavioral data assists decisions: exploring the mental representation of digital-self

Behavioral
data assists
decisions

185

Yixin Zhang

School of Software, Shandong University, Jinan, China

Lizhen Cui, Wei He and Xudong Lu

*School of Software, Shandong University, Jinan, China and
Joint SDU-NTU Centre for Artificial Intelligence Research (C-FAIR),
Jinan, China, and*

Shipeng Wang

School of Software, Shandong University, Jinan, China

Received 4 March 2021
Revised 15 April 2021
Accepted 19 April 2021

Abstract

Purpose – The behavioral decision-making of digital-self is one of the important research contents of the network of crowd intelligence. The factors and mechanisms that affect decision-making have attracted the attention of many researchers. Among the factors that influence decision-making, the mind of digital-self plays an important role. Exploring the influence mechanism of digital-selves' mind on decision-making is helpful to understand the behaviors of the crowd intelligence network and improve the transaction efficiency in the network of CrowdIntell.

Design/methodology/approach – In this paper, the authors use behavioral pattern perception layer, multi-aspect perception layer and memory network enhancement layer to adaptively explore the mind of a digital-self and generate the mental representation of a digital-self from three aspects including external behavior, multi-aspect factors of the mind and memory units. The authors use the mental representations to assist behavioral decision-making.

Findings – The evaluation in real-world open data sets shows that the proposed method can model the mind and verify the influence of the mind on the behavioral decisions, and its performance is better than the universal baseline methods for modeling user interest.

Originality/value – In general, the authors use the behaviors of the digital-self to mine and explore its mind, which is used to assist the digital-self to make decisions and promote the transaction in the network of CrowdIntell. This work is one of the early attempts, which uses neural networks to model the mental representation of digital-self.

Keywords Mental representation, Crowd behavior analysis, Behavioral decisions, CrowdIntell network

Paper type Research paper

© Yixin Zhang, Lizhen Cui, Wei He, Xudong Lu and Shipeng Wang. Published in *International Journal of Crowd Science*. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licenses/by/4.0/legalcode>

This work is partially supported by the National Key R&D Program (No. 2017YFB1400100), NFSC (NO. 61772316) and Ministry of Science and Technology (MOST) Research Project on Innovative Methodology (No. 2020IM20100).



1. Introduction

The network of CrowdIntell (Chai *et al.*, 2017; Wang *et al.*, 2019) refers to a complex self-organizing ecological network formed by multiple intelligent subjects in the physical space, as well as their behaviors, consciousness and information, which are mapped one by one to the digital-selves in the information space and interconnected. It is a three-dimensional fusion and deep superposition space of information space, physical space and psychological space.

The intelligent phenomena (Li *et al.*, 2017b) in the information space are not only large scale, but also deeply interconnected, widely interconnected and diversified in forms, which exist in the digital-selves. To realize the interaction, cooperation and development between numerous digital-selves, first of all, we need to fully, truthfully, correctly and synchronously project the intelligent subjects in the physical space and their mind in the psychological space into the information space. That is, we need to build a mental model and an interconnection model of digital-selves (Li *et al.*, 2018).

In recent years, with the development of machine learning and deep learning, deep representation learning has attracted more and more attention and provided new ideas, methods (Xudong *et al.*, 2019) and guidance for the development of CrowdIntell network. Through the powerful representation ability of neural network, a hidden vector representation of potential space is generated for each digital-self, so that each intelligent subject in the physical world corresponds to a dense vector and can express rich semantic information. A good representation can meet the needs of tasks and transactions in CrowdIntell network, such as a social network discovering new friends or a recommendation system recommending potential goods. So, one of the most critical questions in representation learning is: How can we learn representations that satisfy our needs?

The behavior of digital-self is influenced by its mind (Wang *et al.*, 2019). It is a reasonable way to explore the mind from the behavior of digital-self. Mind is a complex abstract concept, which includes the interests, preferences, intentions and other related factors of the intelligent subject, and may even be influenced by social opinions or friends. For example, for a Web celebrity cake, users may not like to eat cake, but still buy Web celebrity products. Then the user's behavior is implied by the influence of the underlying factors from a number of aspects, what called the mind.

We use the mind, which has been explored, to assist decision-making to facilitate the efficiency of CrowdIntell network. For example, when going shopping in the mall, everyone hopes to find the goods they need as soon as possible. In addition, when there is no clear demand, they will always be attracted by some goods that conform to their aesthetic taste and meet their potential intention (Yoshida *et al.*, 2020). Our goal is to explore the mind of intelligent subjects (users) in CrowdIntell network, retrieve the items most relevant to the mind and recommend them to the users, so as to facilitate transactions. At the same time, if the network can accurately provide optimal suggestions for users to make decisions, which will catalyze users' dependence on CrowdIntell network, and promote the development of CrowdIntell network and the intelligent degree of the crowd.

In this paper, we propose a novel method called adaptive multi-aspect mental exploration (*AM²E*). It uses behavioral patterns perceptual layer, multi-aspect perceptual layer and memory network enhancement layer to explore the mind of the digital-self and assist the digital-self to make decisions. Specifically, the authors use transformer encoders to encode behaviors, model dependencies and potential connections between behaviors. The potential representation of different aspects can be generated by multi-dimensional attention model modeling, which enables more differentiated attribute information reflecting the user's mind to have a greater impact on the corresponding potential representation. Then the complex and comprehensive mental representation is generated by the adaptive fusion module. In

addition, the ability of mental representation is enhanced with the assistance of memory network. The result is a representation of a digital-self, i.e. the mind of a digital-self.

Our contributions in this paper can be summarized as follows:

- We propose to explore the mind, to assist the digital-self to make decisions by mining the mental representation of the digital-self through its behaviors and finally to improve the efficiency of trading in the network of CrowdIntell.
- We explore the mind from the behavioral patterns and the multi-aspect factors behind the behaviors, and also enhance the mental representation through memory networks. Forming a mental representation framework for digital-selves.
- The experiments in real-world open data sets show that using mental representations to assist decision-making has a significant improvement compared with universal interest modeling baseline methods.

2. Related work

With the development of society, human beings have entered the era of CrowdIntell network (Li *et al.*, 2017b). We are in an environment where everything is interconnected. To realize the interaction, cooperation and evolution of various digital-selves in CrowdIntell network, the establishment of mental model is very important. The mental model is expected to lay a foundation for the theory of modeling and simulation in the research of crowd science and engineering. More specifically, mind is crucial in the decision-making (Kumar and Bishnu, 2019) process, and exploring mental representation can improve transaction efficiency.

2.1 Intent representation

Regarding the representation of user intent in behavioral items (Jingwei *et al.*, 2019), it has always been a research hotspot. DIN (Zhou *et al.*, 2018) model captures the user's interest points hidden in behavior items by introducing the attention mechanism. STAMP (Liu *et al.*, 2018) captures the user's long-term overall and current short-term interest preferences from behavior items. There are also efforts (Chen *et al.*, 2019a; Li *et al.*, 2019; Cen *et al.*, 2020) to capture multiple interest (or intent) representation in user behavior items. Although our work is deeper than modeling intent, which aims to explore the user's mind, there is still a lot to learn from and refer to.

2.2 Traditional method

Matrix factorization (MF) (Xiangnan *et al.*, 2017; Steffen *et al.*, 2012) is the most widely used method. This method can obtain the hidden factor vector of each user and each item to estimate the user's predicted score for a certain item through the inner product between the vectors. The implicit vector of the user can be simply understood as the user's mind. The principle behind this is to find the items related to the user's mind. BPR-MF (Steffen *et al.*, 2012) uses MF with the pairwise Bayesian personalized ranking (BPR) loss. NeuMF (Xiangnan *et al.*, 2017) uses both neural network architecture and MF to model linear and nonlinear user-item characteristics.

2.3 Sequential method

Different from the traditional method (Liu *et al.*, 2015, 2017; Peng *et al.*, 2018), the user's interactive behavior is strictly time sequential in the CrowdIntell network. Sequential methods use historical behavioral data arranged in chronological order to model mental representation for assisting decisions. Earlier applications in the sequential method (He and McAuley, 2016; Rendle *et al.*, 2010) are based on Markov chains (MCs), which were designed

to model sequential dependencies between user behaviors. A classic model FPMC (Rendle *et al.*, 2010) combines the two methods of MF and MC. With the wide application of deep learning technology, the recurrent neural network (RNN) shines brilliantly in the field of sequence problems. Two variants of RNN, LSTM (Hochreiter and Schmidhuber, 1997) and GRU (Kyunghyun *et al.*, 2014) are widely used. A large number of RNN-based works (Gharibshah *et al.*, 2020; Balázs and Alexandros, 2015; Li *et al.*, 2017a) has been explored as a decision assistant tool. Among them, GRU4Rec (Balázs and Alexandros, 2015) has attracted attention as a pioneering work. In addition, many new models have been proposed through variants of GRU. For example, add personalization (Quadrana *et al.*, 2017), contexture (Smirnova and Vasile, 2017) and attention mechanism (Li *et al.*, 2017a).

Simultaneously, the convolutional neural network-based sequential model has been explored. Caser (Tang and Wang, 2018) proposed to embed a sequence of recent items into the latent spaces and to learn sequential patterns using both horizontal and vertical convolutional filters. To solve the shortcoming that the perception of the model is limited by the size of the convolution kernel, NextItNet (Yuan *et al.*, 2019) has also been proposed inspired by temporal convolutional networks. MIAR (Zhang *et al.*, 2021) uses lightweight attention modules in convolution to extract fine-grained features of users' multi-interest representation.

Because of the great success of the BERT model and transformer in the field of NLP, the attention mechanism (Vaswani *et al.*, 2017) has been incorporated into the sequential filed recently. It is different from using the attention mechanism as an additional component of the original model [such as attention combined with attention (Li *et al.*, 2017a)]. Recently, pure attention-based models SASRec (Kang and McAuley, 2018) and BERT4Rec (Sun *et al.*, 2019) have been proposed. These models rely on self-attention mechanisms to model sequential patterns between sequences. Apart from this, HGN (Chen *et al.*, 2019a) adopts adaptive gating network to model sequential features.

3. Problem statement

CrowdIntell network projects the interactions between intelligent subjects (users) and intelligent objects (items or products) in the physical space onto the network environment. However, at the present stage, the development of CrowdIntell network is in its infancy. The interaction behavior in the network is extremely sparse, and the stability and robustness of the network are poor. For the healthy development of the network, we need to facilitate transactions in the network to obtain a more comprehensive representation of the digital-self and a more stable interconnection structure.

We aim to explore the mind from the behavior and the multi-aspect factors behind the behavior, and to use mental representation to assist the digital-self in making decisions. In other words, we will use the mind to conduct behavior retrieval and recommend the behaviors most relevant to the mental representation to users, which are used to assist decision-making and improve the trading efficiency of CrowdIntell network. We can define it as follows.

Definition 1: Given CrowdIntell network, we focus on one of the specific sub-domains (i.e. explore the mind to promote trading efficiency). In this paper, U , V and G represent a set of user, behavioral item and item attributes, respectively. An item $a \in V$ may have multiple attributes, which we denoted as $G_a \in G$.

Definition 2: For a specific user, $u \in U$ is associated with a sequence of historical behaviors, and we can sort the behavior records in chronological order $B^u = \{(X_1^u, A_1^u), (X_2^u, A_2^u), \dots, (X_{|B^u|}^u, A_{|B^u|}^u)\}$, where $X_i^u \in V, A_i^u = \cup_{a \in X_i^u} G_a$. The index i for X_i^u denotes the relative time index.

Definition 3: With the above notations, our goals are as follows. Given a user u and u 's historical behavior sequence B^u , our purpose is to infer u 's mind by considering comprehensively, and recommend a list of behavioral items that maximize the completion of the transaction.

We summarize the process by the following formula. The input data is processed through an encoder to explore mental representation. Then we can get the prediction score of candidate behavioral items through inner product, and sort the candidate behaviors by the score. The top-K items with the highest score will be recommended to the user. The prediction score can be simply expressed as follows:

$$\hat{r}_{u,a} = f_{enc}(B^u) \cdot e_a \quad (1)$$

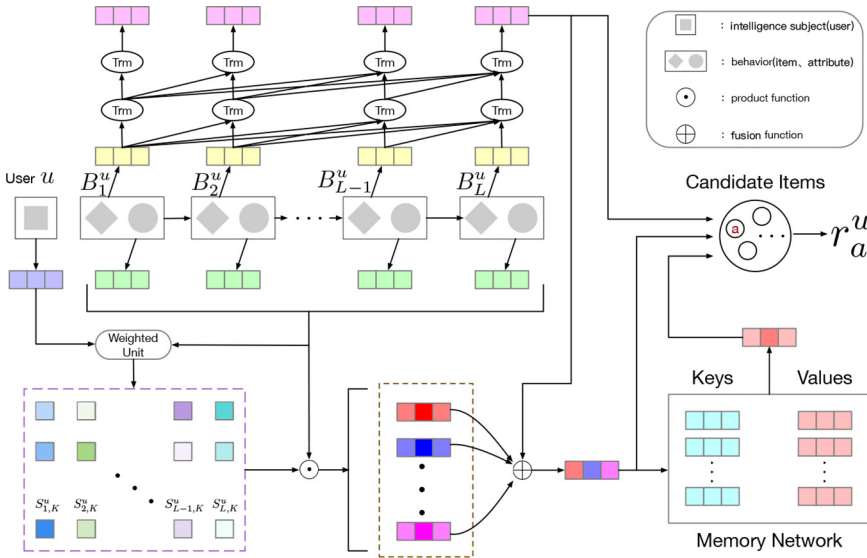
where f_{enc} represents encoder, e_a is the implicit vector of behavioral item a . Prediction score $\hat{r}_{u,a}$ is used to measure the user u interaction probability to item a .

4. Proposed approach

In this section, we introduce the proposed method AM^2E , which incorporates a multi-aspect perceptual module to learn multi-aspect faction representation and an adaptive mind fusion module to aggregation fine-grained mind representation. More importantly, the memory network and transformer encode are introduced, respectively, which significantly improves the ability to learn features. The overall architecture of AM^2E is shown in Figure 1.

4.1 Input layer

User ID and item ID are one-hot encoded, and item attributes data is multi-hot encoded, that is, an item may correspond to more than one attribute. The input of the model is the ID



Notes: With the user's previous behavioral items as input, the model obtains mental representation through three parts of behavioral patterns, multi-aspect perception and memory networks enhancement. And using the mental representation to assist digital-selves to make decisions

Figure 1.
The model architecture of AM^2E

coded representation after data preprocessing. It mainly contains the following data: the user input u , u 's historical behavior sequence B^u and auxiliary information related, such as the filling matrix and the actual length of the sequence.

4.2 Embedding layer

The original input of user and item IDs have very limited representation capacity, as higher characteristic dimension and extremely sparse. Through a special full-connection layer, the features obtained from the input layer are transformed into a dense low-dimensional vector representation. Embedding the user u 's ID characteristics as $u_{em} \in \mathbb{R}^d$, for user u at time step t , we retrieve the input embedding matrix $E^{(u,t)} \in \mathbb{R}^{L \times d}$ by looking up the previous L behavioral items $X_{t-L}^u, \dots, X_{t-1}^u$. Where d is the latent dimensionality. The experiment shows that the position information in the sequence also has an obvious promoting effect on the downstream tasks. Here, the sinusoidal position coding function PE is used to map the project position to position embedding. The embedded matrix using positional encoding is defined as follows:

$$E^{(u,t)} := E^{(u,t)} + PE(E^{(u,t)}) \quad (2)$$

At the same time, the interactive item attributes sequence is transformed into an embedded matrix with dimension $L * d$ by aggregating multiple attributes of each item, which is expressed as $C^{(u,t)} \in \mathbb{R}^{L * d}$. The common aggregation methods include maximum pooled aggregation and average pooled aggregation, and average aggregation is adopted here.

4.3 Behavioral pattern perception layer

By measuring the similarity of different behavioral items in the sequence, this layer models the potential relationship between behavioral items in the sequence, and then models behavioral patterns and the collaboration relationship between items. Transformer encoders are used as encoders here. Previous studies have shown that transformer encoders can effectively capture various types of sequence dependencies (such as point-level dependencies and group-level dependencies) and long-term dependencies.

When calculating the similarity of different items, various methods can be selected, such as inner product, multi-layer perceptron and addition and subtraction. Here, inner product is chosen to measure the similarity. The attention mechanism is calculated as follows:

$$Attention(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = softmax\left(\frac{\mathbf{Q} \cdot \mathbf{K}^T}{\sqrt{d}}\right) \mathbf{V} \quad (3)$$

where \mathbf{Q} represents the queries, \mathbf{K} represents the keys and \mathbf{V} represents the values. The factors \sqrt{d} play a regulatory role so that the inner product is not too large, otherwise $softmax$ may be invalid.

We use the transformer framework to model the interaction between the items in the sequence and their contextual neighbors. Given the item representation of the $b - 1$ sequence H^{b-1} , the output of the transformer encoder at layer b is defined as follows:

$$H^b = FFN(Concat(head_1, \dots, head_h)W^h) \quad (4)$$

$$head_i = Attention\left(H^{b-1}W_i^Q, H^{b-1}W_i^K, H^{b-1}W_i^V\right) \quad (5)$$

where the projection matrices $W_i^Q, W_i^K, W_i^V \in \mathbb{R}^{d \times d/h}$, $W^h \in \mathbb{R}^{d \times d}$. $FFN(s^2)$ represents feed-forward network, h represents the number of heads and b represents the number of

layers. Here, we omit the residual network, dropout and layer normalization strategies in the formula for convenience.

In our experiments, we can repeat the basic structure of the transformer encoder several times to obtain long-term and complex dependencies. The first self-attention block can consider similarities and potential connections between previous items. On this basis, we can model more complex relationships by stacking multiple attention blocks.

In general, the coding process can be summarized as:

$$H^{(u,t)} = f_{T-enc}(E^{(u,t)}) \quad (6)$$

where $f_{T-enc}(s^2)$ represents the abstract transformer encode function, and $H^{(u,t)} \in \mathbb{R}^{L*d}$ represents the result of the original sequence encoded by the encoder. After obtaining the output $H^{(u,t)}$ at the last layer, we obtain the representation of $H_{l-1}^u \in \mathbb{R}^d$ as the user's most recent interaction item X_{l-1}^u , which is denoted by h for simplicity. In addition, h also takes into account the influence of previous items and the synergistic effect of items in the sequence. The experiment also shows that the representation effect of using h as the sequence is better than average pooling or maximum pooling.

4.4 Multi-aspect perceptual layer

The multi-aspect perceptual layer is divided into two parts. The first part is multi-aspect perceptual module, which measures the different attributes of the behavioral items and abstracts the attributes into different aspects of the user's mental factors. The second part is the adaptively fusion module, which considers the recent item adaptively and fusion user's multiple factors to generate hybrid mind representation.

4.4.1 Multi-aspect perceptual module. The multi-aspect perceptual layer measures the different influences of different items in the sequence on the multi-aspect and combines the general preferences of users, through the multi-dimensional attention module assign weights to obtain the representation of users' minds in different factors.

This enables the more differentiated attribute information that reflects the user's mind to have a greater impact on the corresponding mental factors. At the same time, the general preferences of users will also have an impact on the behavior of users. For example, if a user likes goods with beautiful appearance, he will be inclined to choose goods with beautiful appearance no matter he buys daily necessities or electronic products. In the layer, the attribute characteristics of user's historical interaction behavior and user's general preference are integrated to explore user's mind. Formally, we can express this process as:

$$S^I = \text{softmax}(W_k \tanh(W_a \cdot C^{(u,t)} + (W_u \cdot u_{em}) \otimes I_\Phi)) \quad (7)$$

$$Z^{(u,t)} = \tanh(S^{IT} \cdot C^{(u,t)}) \quad (8)$$

where the model parameters $W_u \in \mathbb{R}^{d*d}$, $W_k \in \mathbb{R}^{d*k}$, $\tanh(\cdot)$ represent a nonlinear activation function, \otimes represents element-wise product, I_Φ represents a full ones matrix for dimensions such as $C^{(u,t)}$. One of the most important hyper-parameter k controls the number of aspects. $S^I \in \mathbb{R}^{L*k}$ represents a multi-dimensional attention score and $Z^{(u,t)} \in \mathbb{R}^{k*d}$ is a mental matrix representation of multiple factors, with each row in k rows representing a specific aspect.

4.4.2 Adaptively fusion module. For users, there are multiple factors affecting their behavioral decisions. However, items that are highly correlated with recent behaviors tend

to generate more feedback from users. Therefore, to maintain the diversity and novelty of recommendations, we need to give higher weight to the mental factors related to the user's recent interactions in the next recommendation, as well as other factors. The process of adaptive fusion is defined by the following formula:

$$S^Z = \text{softmax}(h \cdot \text{ReLU}(W_{mi} \cdot Z^{(u,t)})) \quad (9)$$

$$I^{(u,t)} = \tanh(S^{ZT} \cdot Z^{(u,t)}) \quad (10)$$

where the model parameters $W_{mi} \in \mathbb{R}^{d \times d}$, $\text{Relu}(\cdot)$ represent a nonlinear activation function. $S^Z \in \mathbb{R}^k$ represents the similarity score between the user's most recent interaction behavior and the user's mind factors representation and $I^{(u,t)} \in \mathbb{R}^d$ represents the mixed mind representation after adaptively fusion.

4.5 Memory network enhancement module

Adaptively fusion module uses the attributes of the items to model the user's multi-aspect mind factors in an implicit way (that is, it does not explicitly indicate what the mind factors are). An item has multiple attributes, indicating that there may be different users interacting with it for different purposes. While it is possible to model items through their attributes, it is not effective to model more complex underlying relationships and items associations. For example, the appearance of a commodity is beautiful, which is not well reflected in the attributes of the commodity. To model the deeper characteristics of items, we use a memory network to encode the complex underlying characteristics of items in the internal storage through read and write operations. By means of the associative addressing scheme in the memory network, the feature level fusion is realized by adaptively discovering the item features related to the user mind. Concrete, this article uses the key/value pair memory network, formally defined as $M_K \in \mathbb{R}^{|V| \times d_m}$, $M_V \in \mathbb{R}^{|V| \times d_m}$, respectively. Therefore, the enhanced mind of using memory network is defined as follows:

$$S^M = \text{softmax}\left(I^{(u,t)} \cdot \left(W_{mk} \cdot M_K^{(u,t)}\right)\right) \quad (11)$$

$$M^{(u,t)} = \tanh\left(W_{mv} \left(S^{MT} \cdot M_V^{(u,t)}\right)\right) \quad (12)$$

where d_m is the dimension of the memory unit. $M_K^{(u,t)}, M_V^{(u,t)} \in \mathbb{R}^{L \times d_m}$ represents the key/value of memory units corresponding to the given previous L behavioral items. Given the user's mixed mind as a query, using this query to find the appropriate combination of potential features of the items in the memory network, $S^M \in \mathbb{R}^L$ represents the score of the mind for the items, and $M^{(u,t)} \in \mathbb{R}^d$ represents the final representation of the enhanced mind by the memory network after passing a shared mapping matrix.

4.6 Prediction layer

As mentioned above, the user's mind needs to consider:

- user's behavioral patterns and collaborative relationships between behavioral items representation h ;
- user's mixed mental factors representation $I^{(u,t)}$; and
- the enhancement of the user's mixed mind representation $M^{(u,t)}$.

The predicted score for candidate behavioral item $a(a \in V)$ at t time step is calculated as follows:

$$r_a^{(u,t)} = (h + I^{(u,t)} + M^{(u,t)}) \cdot e_a^T \quad (13)$$

where $e_a \in \mathbb{R}^d$ shares the embedded item matrix parameter with the embedding layer.

4.7 Model training

In this paper, the truncation and padding strategy is adopted to convert each user behavior sequence (excluding the last operation) to a fixed-length sequence $B^{(u,t)}$, and the interaction item X_t^u corresponding to the time step t is taken as the prediction result. We convert the prediction score into probability and take the negative log likelihood function for model optimization. In other words, we use the cross-entropy loss as the objective function:

$$\mathcal{L} = - \sum_u \sum_t \left(\log \left(\sigma \left(r_{X_t^u}^{(u,t)} \right) \right) + \sum_{j \notin X_t^u} \log \left(1 - \sigma \left(r_j^{(u,t)} \right) \right) \right) + \lambda \|\Theta\| \quad (14)$$

where Θ is a set of the model trainable parameters and λ is the regularization parameter. $\sigma = 1/(1 + e^{-x})$ is sigmoid function. The network is optimized by the Adam optimizer (Diederik and Jimmy, 2014), which is a variant of stochastic gradient descent with adaptive moment estimation.

5. Experiments

We take e-commerce, online services and other typical crowd intelligence scenarios as examples to verify the effectiveness of the proposed method. On standard benchmark data sets from different domain, we compare the proposed method with the baseline method.

5.1 Data sets

This work conducts experiments on five common data sets collected from real-world platforms, which come from different domains and have different sparsity levels. To ensure that each user/item has enough interaction, we follow the preprocessing procedure in Zhou *et al.* (2020), which only keeps the “5-core” data sets. This means that users and items with fewer than five interaction records are deleted. The processed data statistics are summarized in the following Table 1.

Amazon [1] data set is widely used to evaluate the performance. According to the category of products on the Amazon platform, this work selects three subcategories, beauty, sports and toys, and uses the categories and the brands of the items as attributes. In the LastFM [2] data set, the artist tags given by the users are used as attributes. Yelp data set is collected by Yelp [3], which is the largest review site in the USA. We follow the preprocessing procedure in Zhou *et al.* (2020) and use the transaction records after January 1st, 2019. In addition, we treat the categories of businesses as attributes.

Data sets	# Users	# Items	# Attributes	# Interaction	Density (%)
Beauty	22,363	12,101	1,221	198,502	0.07
Sports	25,598	18,357	2,277	296,337	0.05
Toys	19,412	11,924	1,027	167,597	0.07
LastFM	1,090	3,646	388	52,551	1.32
Yelp	30,431	20,033	1,001	316,354	0.05

Table 1.
Statistics of the data
sets

5.2 Evaluation metrics

In our work, the strategy of “leave one out” was adopted to divide the data sets. For each user, we use the last behavioral item as test data and the item before the last item as validation data. The rest are used for training. To save computation resources and time, we randomly sampled 99 negative sample items according to the popularity of the items, and constituted a candidate set with ground-truth. We report the evaluated results by three popular top-K metrics, namely, hit ratio (HR@K), normalized discounted cumulative gain (NDCG@K) and mean reciprocal rank (MRR). Here, we empirically set K to 1, 5 and 10. We omit NDCG@1, because its result is equal to HR@1.

5.3 Baselines

We compare our method with the following baselines:

- NeuMF (Xiangnan *et al.*, 2017): This method combined with the traditional generalized MF and multilayer perceptron can capture both linear and nonlinear interaction features between users and items.
- BPRMF (Steffen *et al.*, 2012): This method is based on BPR, which uses pairwise coding method to sort all items for each user.
- FPMC (Rendle *et al.*, 2010): It combines MF and MC to fuse sequence and personalization information.
- GRU4Rec (Balázs and Alexandros, 2015): This method is a variant of RNN, which uses GRU to capture sequential dependencies and make recommendation.
- Caser (Tang and Wang, 2018): It uses vertical and horizontal convolution to learn users’ sequential patterns for sequential recommendation.
- STAMP (Liu *et al.*, 2018): It considers the impact of the user’s current actions on the next step and captures the user’s long-term overall and current short-term interest preferences.
- SASRec (Kang and McAuley, 2018): It uses self-attention mechanism to capture the user’s sequential pattern for sequential recommendation.
- BERT4Rec (Sun *et al.*, 2019): It uses the bi-direction self-attention mechanism to model the sequence of user behavior and constructs the bi-direction representation model by Cloze task learning.
- HGN (Chen *et al.*, 2019a): This method adopts adaptive hierarchical gating unit to model sequential features.

5.4 Experiment settings

For fair comparison, we collect open source code or source code provided by the corresponding authors. We implemented them by using PyTorch [4] based on those codes while keeping the data format and evaluation metrics consistent with our work. All hyper-parameters are tuned by grid search on the validation set.

For the proposed method AM^2E , we set batch size as 256, learning rate is 0.001 and the weight of the L2 regularization is set to 5×10^{-4} . The model latent dimension and memory unit dimension are both set to 64. The maximum sequence length L is set to 20, as the average length is low in most cases. As for the transformer encode, we set number of heads h and number of layers b as 2. Another important hyper-parameter, multi-aspect mind factors, is set to 5. All parameters are tuned by grid search. Our experiments are conducted with PyTorch running on GPU machines (Nvidia GeForce TITAN RTX).

5.5 Performance comparison

The performance comparison results are shown in Table 2. It is worth noting that improvement represents difference between best results and second-best results. From the results, we have the following observations.

Overall, the proposed model performs better than all baselines in the experiment. Among the baseline methods, the performance of the sequential methods (e.g. SASRec and GRU4Rec) is better than non-sequential methods (BPRMF and NeuMF). This shows that historical behavior records can be used to effectively model behavioral patterns and potential associations between items. In the sequential behavior baseline method, SASRec uses self-attention mechanism and outperforms the other baseline methods, indicating the effectiveness of attentional mechanism in modeling sequential behavior. In addition, the performance of HGN is comparable to SASRec. This indicates that the gating network can simulate the relationship between related items well.

We observed that all methods had generally low scores on the sports data set. When there are more attributes in the data set, it means that the distribution of items is more sparse. Therefore, the overall data distribution is sparse, which makes it more difficult for the methods to model the potential correlation.

Different from simple modeling of user interest or behavioral pattern, the proposed method comprehensively considers multi-aspect mental factors of users and incorporates behavioral pattern into them. Experiments show that the performance is outperforming all baseline methods. The significant improvement in the comparison results demonstrated the effectiveness of mental representations in decision-making.

5.6 Ablation study

There are three important components in our model. The first part represents behavioral patterns and potential relationship between behavioral items, which are hidden in historical behavior sequence. The second part represents adaptive fusion of multi-aspect mental factors. The third part represents hybrid mental representation enhanced by memory network. Then we verify the effectiveness of important structures through ablation experiments. For the validity of the data, we repeated the experiment several times and took the average to get the final result.

Figures 2 and 3 show the performance of our default method and its ablation variant on all data set. We introduce them respectively and analyze their influence.

We removed behavioral patterns, multi-aspect perception and memory networks, respectively, to test the validity of each component in mental representation. In addition, DEFAULT stands for the AM^2E method, and their variants are named W/O BP, W/O MA and W/O MN, respectively. At the same time, we chose SASRec and AM^2E as comparison. From our experiments, we can see that all the results outperform SASRec, and AM^2E is always at the optimal level, indicating that each part of the proposed method contributes to the modeling of mental representation.

The importance of the three components varies from one data set to another. As the experimental results show, memory network plays a more important role in Yelp data set, multi-aspect perception is more important in LastFM data set, while behavioral pattern is crucial in Amazon data sets. In general, behavioral patterns and multi-aspect perception play a more important role than mental representation enhancement. Mental representation enhancement is more like icing on the cake.

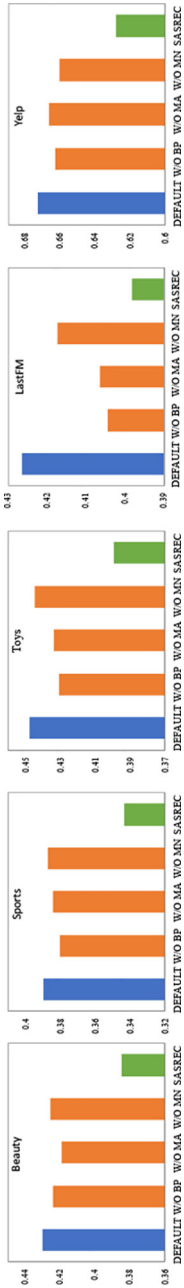
5.7 Influence of embedding dimension

We analyzed the key hyper-parameter, embedding dimension, to understand the impact of embedding dimension on the performance of the proposed model in this paper. The results

Table 2.
Performance
comparison on all data
sets

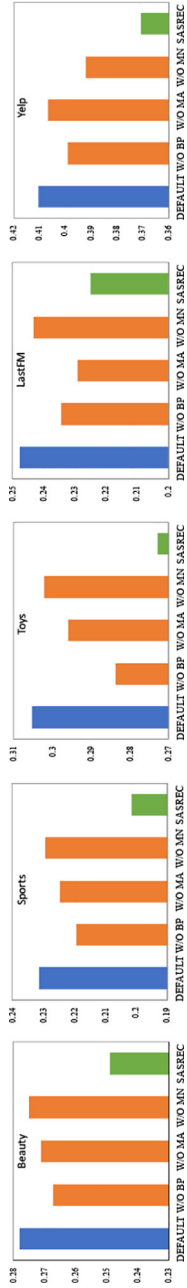
Data set	Metric	NeuMF	BPRMF	FPMC	GRU4Rec	Caser	STAMP	SASRec	BERT4Rec	HGN	AM ² E	Improve (%)
Beauty	HR@1	0.0603	0.0647	0.0909	0.1134	0.0787	0.1072	0.1381	0.0877	0.1139	0.1538	11.37
	HR@5	0.1746	0.1833	0.2273	0.2682	0.2133	0.2435	<u>0.2985</u>	0.2328	0.2849	0.3341	11.93
	NDCG@5	0.1182	0.1248	0.1610	0.1925	0.1473	0.1775	0.2211	0.1618	0.2014	0.2469	11.67
	HR@10	0.2685	0.2771	0.3136	0.3605	0.2983	0.3301	<u>0.3846</u>	0.3282	0.3953	0.4300	11.80
	NDCG@10	0.1483	0.1549	0.1889	0.2223	0.1747	0.2055	<u>0.2489</u>	0.1925	0.2370	0.2779	11.65
Sports	MRR	0.1343	0.1402	0.1706	0.1997	0.1576	0.1873	<u>0.2251</u>	0.1710	0.2081	0.2492	10.71
	HR@1	0.0373	0.0439	0.0572	0.0790	0.0491	0.0722	<u>0.0958</u>	0.0627	0.0845	0.1092	13.99
	HR@5	0.1289	0.1518	0.1561	0.2193	0.1530	0.1908	<u>0.2363</u>	0.1912	0.2439	0.2767	13.45
	NDCG@5	0.0826	0.0978	0.1070	0.1502	0.1011	0.1319	0.1671	0.1272	0.1651	0.1949	16.64
	HR@10	0.2208	0.2487	0.2396	0.3322	0.2471	0.2911	<u>0.3434</u>	0.2966	0.3642	0.3899	7.06
Toys	NDCG@10	0.1121	0.1290	0.1338	0.1865	0.1313	0.1641	0.2016	0.1610	0.2038	0.2313	13.49
	MRR	0.1032	0.1165	0.1241	0.1642	0.1196	0.1481	<u>0.1800</u>	0.1421	0.1773	0.2042	13.44
	HR@1	0.0575	0.0735	0.1189	0.1199	0.0774	0.1181	<u>0.1693</u>	0.1027	0.1281	0.1893	11.81
	HR@5	0.1431	0.1835	0.2386	0.2801	0.2103	0.2496	<u>0.3172</u>	0.2405	0.2853	0.3535	11.44
	NDCG@5	0.1009	0.1300	0.1808	0.2027	0.1457	0.1857	<u>0.2463</u>	0.1732	0.2095	0.2749	11.62
LastFM	HR@10	0.2179	0.2659	0.3137	0.3750	0.2992	0.3341	<u>0.3994</u>	0.3275	0.3807	0.4479	12.14
	NDCG@10	0.1250	0.1564	0.2051	0.2331	0.1743	0.2128	<u>0.2729</u>	0.2012	0.2401	0.3053	11.87
	MRR	0.1184	0.1451	0.1917	0.2095	0.1572	0.1957	<u>0.2520</u>	0.1827	0.2168	0.2793	10.83
	HR@1	0.0569	0.0862	0.0798	0.0716	0.0697	0.0468	<u>0.1000</u>	0.0780	0.1028	0.1138	10.79
	HR@5	0.2193	0.2596	0.2697	0.2321	0.1798	0.1633	0.2679	0.2450	0.2817	0.2890	2.59
Yelp	NDCG@5	0.1384	0.1755	0.1769	0.1523	0.1242	0.1050	0.1829	0.1625	0.1939	0.2033	4.85
	HR@10	0.3440	0.3862	0.4009	0.3651	0.2853	0.2587	0.3982	0.3651	0.4202	0.4266	1.52
	NDCG@10	0.1784	0.2167	0.2194	0.1949	0.1582	0.1361	0.2251	0.2012	0.2387	0.2477	3.77
	MRR	0.1524	0.1890	0.1870	0.1663	0.1438	0.1225	0.1956	0.1750	0.2051	0.2166	5.61
	HR@1	0.1476	0.1431	0.1320	0.1509	0.1179	0.1476	0.1685	0.1581	0.1923	0.2013	4.68
Yelp	HR@5	0.4005	0.3901	0.3771	0.4604	0.3729	0.4005	0.4574	0.4346	0.4797	0.5022	4.69
	NDCG@5	0.2751	0.2679	0.2561	0.3085	0.2465	0.2751	0.3153	0.2988	0.3395	0.3549	4.54
	HR@10	0.5795	0.5559	0.5418	0.6402	0.5476	0.5795	0.6284	0.6009	0.6466	0.6731	4.10
	NDCG@10	0.3329	0.3214	0.3093	0.3667	0.3029	0.3329	0.3707	0.3526	0.3935	0.4103	4.27
	MRR	0.2779	0.2695	0.2584	0.3004	0.2492	0.2779	0.3095	0.2947	0.3330	0.3459	3.87

Notes: The best results are highlighted in italic and the second-best results are highlighted with an underscore. All reported improvements over baseline methods are statistically significant at 0.05 level



Note: Using HR@10 to evaluate the various method on all datasets. “WO(without)” indicates that the corresponding component is removed in the method, while the rest objectives are kept

Figure 2. Ablation study of our approach



Note: Using NDCG@10 to evaluate the various methods on all data sets

Figure 3.
Ablation study of our
approach

of the different embedding dimensions on the beauty data set are shown in Figure 4 and compared with other representative approaches.

As shown in Figure 4, considering the different embedding dimensions, the AM^2E model outperforms the other approaches in most cases. This further demonstrates the effectiveness of the proposed model and we can see that SASRec performs better with lower embedding dimensions, possibly because transformer is expressive enough to capture sequence features, and the model presents an over-fitting situation as the dimensions get larger. However, AM^2E model proposed in this paper shows a trend of increasing first and then decreasing when the embedding dimension changes, which indicates that increasing the embedding dimension will improve the performance to a certain extent, but as the embedding dimension is too high, it will lead to the over-fitting of the model.

5.8 Influence of sequence length

We also analyzed another key hyper-parameter, sequence length, to understand the effect of sequence length on the performance of the proposed model in this paper. The results of different sequence lengths on the beauty data set are presented in Figure 5 and compared with other representative methods.

As shown in Figure 5, AM^2E is generally superior to other comparison methods under different sequence lengths. It is worth noting that the effect of HGN model decreases with the increase of sequence length, and the HGN model shows better performance when the sequence length is shorter, which indicates that HGN has modeled relatively shallow features, but has shortcomings when modeling longer dependencies. However, the proposed model in this paper,

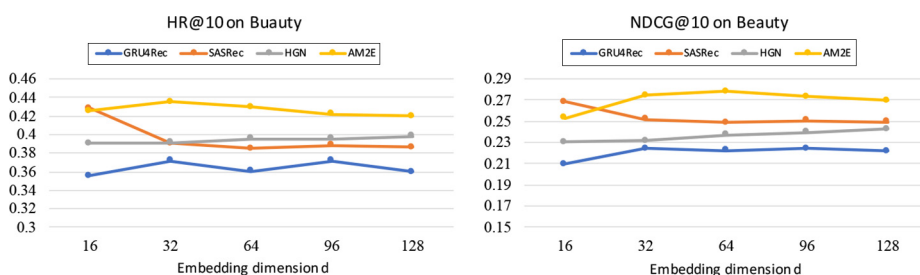


Figure 4.
Influence of
embedding
dimension

Note: Using HR@10 and NDCG@10 to evaluate the proposed method and other representative methods on beauty data set

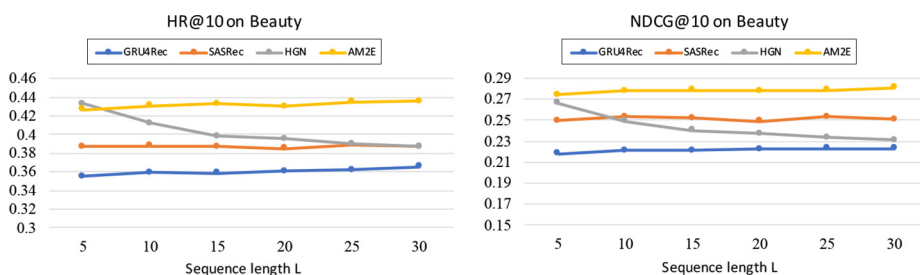
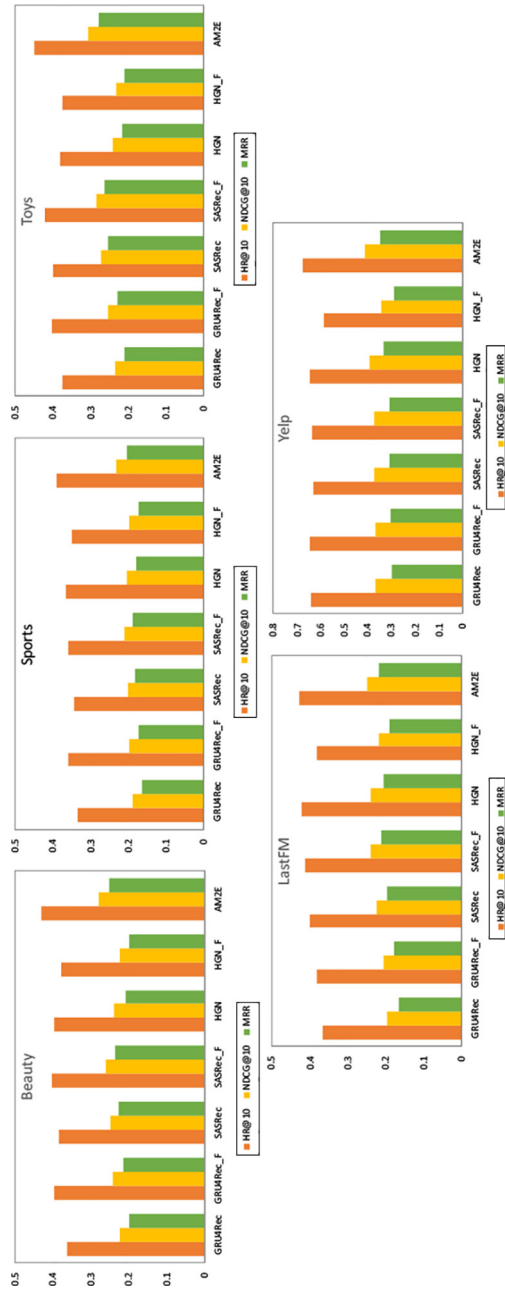


Figure 5.
Influence of sequence
length

Note: Using HR@10 and NDCG@10 to evaluate the proposed method and other representative methods on beauty data set



Note: Using HR@10, NDCG@10 and MRR to evaluate the proposed method and above methods on all data sets

Figure 6.
Influence of feature

AM^2E , shows a slight upward trend when the sequence length increases. Combining with the characteristics of which average sequence length is short on beauty data set, it indicates the stability and robustness of the model in the face of long sequences and complex dependencies.

5.9 Influence of feature

In the proposed method, we use attribute information of behavioral items. In some sequential methods, attribute information is often not included. Some works and experiments can prove that the introduction of attribute information can enhance the representational ability of the item and achieve better results. To demonstrate that our approach can model more complex underlying relationships, rather than relying on attribute information, we have implemented several methods, GRU4REC_F, SASREC_F and HGN_F, which are based on the combination of original sequential model and attribute information. Specifically, we merge attribute embedding with item embedding to replace the original item embedding.

Figure 6 shows the performance of the proposed method and the methods described above on all data sets. Through the experimental results, we can see that not all the sequential models can benefit from the introduction of attribute information, e.g. the performance of HGN is obviously degraded after the introduction of attribute information. In most cases, the introduction of attribute information has a significant impact on the performance of the model, which is in line with our expectations. Another obvious result is that although these sequential methods introduce attribute information, they do not outperform our method on all data sets.

6. Conclusion

In this paper, we propose a novel method called AM^2E . We designed three different structures to explore mental representations, incorporating behavioral patterns module, multi-aspect factors fusion module and memory network enhancement module in a comprehensive and integrated manner. Experiments on public data sets show that each structure plays an important role in mental representation. More importantly, the experiments show that the use of mental representations to assist decision-making has a significant improvement compared with traditional interest modeling method. CrowIntell network provides more precise recommendations for users' behavioral decisions, which can catalyze the dependence of digital-selves on the CrowIntell network and promote the improvement of transaction efficiency.

Notes

1. Available at: <http://jmcauley.ucsd.edu/data/amazon/>
2. Available at: <https://grouplens.org/datasets/hetrec-2011/>
3. Available at: www.yelp.com/dataset
4. Available at: <https://pytorch.org/>

References

- Balázs, H. and Alexandros, K. (2015), "Linus Baltrunas and Domonkos Tikk. Session-based recommendations with recurrent neural networks", arXiv preprint arXiv:1511.06939.
- Cen, Y., Zhang, J., Zou, X., Zhou, C., Yang, H. and Tang, J. (2020) "Controllable multi-interest framework for recommendation", in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 2942-2951.

-
- Chai, Y., Miao, C., Sun, B., Zheng, Y. and Li, Q. (2017), "Crowd science and engineering: concept and research framework", *International Journal of Crowd Science*, Vol. 1 No. 1.
- Chen, M., Kang, P. and Liu, X. (2019a), "Hierarchical gating networks for sequential recommendation", in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 825-833.
- Chen, Q., Zhao, H., Li, W., Huang, P. and Ou, W. (2019a) "Behavior sequence transformer for e-commerce recommendation in Alibaba", in *Proceedings of the 1st International Workshop on Deep Learning Practice for High-Dimensional Sparse Data*, pp. 1-4.
- Diederik, K. and Jimmy, B. (2014), *Adam: A Method for Stochastic Optimization*, *Computer Science*.
- Gharibshah, Z., Zhu, X., Hainline, A. and Conway, M. (2020), "Deep learning for user interest and response prediction in online display advertising", *Data Science and Engineering*, Vol. 5 No. 1, pp. 12-26.
- He, R. and McAuley, J. (2016), "Fusing similarity models with Markov chains for sparse sequential recommendation", in *2016 IEEE 16th International Conference on Data Mining (ICDM)*, pp. 191-200.
- Hochreiter, S. and Schmidhuber, J. (1997), "Long short-term memory", *Neural Computation*, Vol. 9 No. 8, pp. 1735-1780.
- Jingwei, M., Jiahui, W., Mingyang, Z., Weitong, C. and Xue, L. (2019), "MMM: multi-source multi-net micro-video recommendation with clustered hidden item representation learning", *Data Science and Engineering*, Vol. 4 No. 3, pp. 240-253.
- Kang, W.-C. and McAuley, J. (2018), "Self-attentive sequential recommendation", in *2018 IEEE International Conference on Data Mining (ICDM), IEEE*, pp. 197-206.
- Kumar, R. and Bishnu, P.S. (2019), "Identification of k-most promising features to set blue ocean strategy in decision making", *Data Science and Engineering*, Vol. 4 No. 4, pp. 367-384.
- Kyunghy0un, C., Bart, V.M., Caglar, G., Dzmitry, B., Fethi, B., Holger, S. and Yoshua, B. (2014), "Learning phrase representations using RNN encoder-decoder for statistical machine translation", *arXiv preprint arXiv:1406.1078*.
- Li, J., Ren, P., Chen, Z., Ren, Z. and Lian, T. and Ma, J. (2017a), "Neural attentive session-based recommendation", in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pp. 1419-1428.
- Li, P., Liu, L., Cui, L., Qingzhong, L., Zheng, Y. and Zhou, G. (2018), "Rim chain: bridge the provision and demand among the crowd", in *International Conference on Algorithms and Architectures for Parallel Processing*, Springer, pp. 18-31.
- Li, W., Wu, W.-J., Wang, H.-M., Cheng, X.-Q., Chen, H.-J., Zhou, Z.-H. and Ding, R. (2017b), "Crowd intelligence in AI 2.0 era", *Frontiers of Information Technology and Electronic Engineering*, Vol. 18 No. 1, pp. 15-43.
- Li, C., Liu, Z., Mengmeng, W., Yuchi, X., Zhao, H., Huang, P., Kang, G., Chen, Q., Wei, L. and Lee, D.L. (2019), "Multi-interest network with dynamic routing for recommendation at Tmall", in *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, pp. 2615-2623.
- Liu, Q., Zeng, Y., Mokhosi, R. and Zhang, H. (2018), "Stamp: short-term attention/memory priority model for session-based recommendation", in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1831-1839.
- Liu, Y., Zhao, P. and Sun, A. and Miao, C., (2015), "A boosting algorithm for item recommendation with implicit feedback", in *Twenty-Fourth International Joint Conference on Artificial Intelligence*.
- Liu, Y., Zhao, P., Liu, X., Min, W., Duan, L. and Li, X.-L. (2017), "Learning user dependencies for recommendation", in *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, pp. 2379-2385.
- Peng, Y., Peilin, Z., Yong, L. and Xin, G. (2018) "Robust cost-sensitive learning for recommendation with implicit feedback", in *Proceedings of the 2018 SIAM International Conference on Data Mining, SIAM*, pp. 621-629.

-
- Quadrana, M., Karatzoglou, A., Hidasi, B. and Cremonesi, P. (2017) "Personalizing session-based recommendations with hierarchical recurrent neural networks", *In Proceedings of the Eleventh ACM Conference on Recommender Systems*, pp. 130-137.
- Rendle, S., Freudenthaler, C. and Schmidt-Thieme, L. (2010), "Factorizing personalized Markov chains for next-basket recommendation", in *Proceedings of the 19th International Conference on World Wide Web, WWW '10, Association for Computing Machinery*, New York, NY, 811-820.
- Smirnova, E. and Vasile, F. (2017), "Contextual sequence modeling for recommendation with recurrent neural networks", *In Proceedings of the 2nd Workshop on Deep Learning for Recommender Systems*, pp. 2-9.
- Steffen, R., Christoph, F., Zeno, G. and Lars, S.-T. (2012), "BPR: Bayesian personalized ranking from implicit feedback", *arXiv preprint arXiv:1205.2618*.
- Sun, F., Liu, J., Wu, J., Pei, C., Lin, X., Ou, W. and Jiang, P. (2019), "BERT4rec: sequential recommendation with bidirectional encoder representations from transformer", *In Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, pp. 1441-1450.
- Tang, J. and Wang, K. (2018), "Personalized top-n sequential recommendation via convolutional sequence embedding", *In Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, pp. 565-573.
- Vaswani, A., Noam, S., Niki, P., Jakob, U., Llion, J., Gomez, A.N., Łukasz, K. and Illia, P. (2017), "Attention is all you need", *In Advances in Neural Information Processing Systems*, pp. 5998-6008.
- Wang, S., Cui, L., Liu, L., Lu, X. and Li, Q. (2019), "Projecting real world into crowdintell network: a methodology", *International Journal of Crowd Science*, Vol. 3 No. 2.
- Xiangnan, H., Liao, L., Zhang, H., Nie, L., Hu, X. and Chua, T.-S. (2017), "Neural collaborative filtering", *In Proceedings of the 26th International Conference on World Wide Web*, pp. 173-182.
- Xudong, L., Shipeng, W., Fengjian, K., Shijun, L., Li, H., Xu, X. and Lizhen, C. (2019), "An anomaly detection method to improve the intelligent level of smart articles based on multiple group correlation probability models", *International Journal of Crowd Science*.
- Yoshida, A., Higurashi, T., Maruishi, M., Tateiwa, N., Hata, N., Akira Tanaka, T., Wakamatsu, K., Nagamatsu, Akira Tajima, K. and Fujisawa, (2020), "New performance index 'attractiveness factor' for evaluating websites via obtaining transition of users' interests", *Data Science and Engineering*, Vol. 5 No. 1, pp. 48-64.
- Yuan, F., Karatzoglou, A., Arapakis, I., Jose, J.M. and He, X. (2019) "A simple convolutional generative network for next item recommendation", *In Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, pp. 582-590.
- Zhang, Y., Wei, H., Lizhen, C., Lei, L. and Zhongmin, Y. (2021), "Multi-interest aware recommendation in crowdintell network".
- Zhou, G., Zhu, X., Song, C., Fan, Y., Zhu, H., Ma, X., Yan, Y., Jin, J., Li, H. and Gai, K. (2018) "Deep interest network for click-through rate prediction", *In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1059-1068.
- Zhou, K., Wang, H., Zhao, W.X., Zhu, Y., Wang, S., Zhang, F., Wang, Z. and Wen, J.-R. (2020) "S3-rec: self-supervised learning for sequential recommendation with mutual information maximization", *In Proceedings of the 29th ACM International Conference on Information and Knowledge Management*, pp. 1893-1902.

Corresponding author

Lizhen Cui can be contacted at: clz@sdu.edu.cn

For instructions on how to order reprints of this article, please visit our website:

www.emeraldgrouppublishing.com/licensing/reprints.htm

Or contact us for further details: permissions@emeraldinsight.com