Lifelong Sequential Modeling for User Response Prediction

Kan Ren, Jiarui Qin, Yuchen Fang, Weinan Zhang, Lei Zheng, Yong Yu

Weijie Bian, Guorui Zhou, Jian Xu, Xiaoqiang Zhu, Kun Gai

May 2019

上海交通大學
SHANGHAI JIAO TONG UNIVERSITY

上海交通大学
SHANGHAI JIAO TONG UNIVERSITY

User Response Prediction

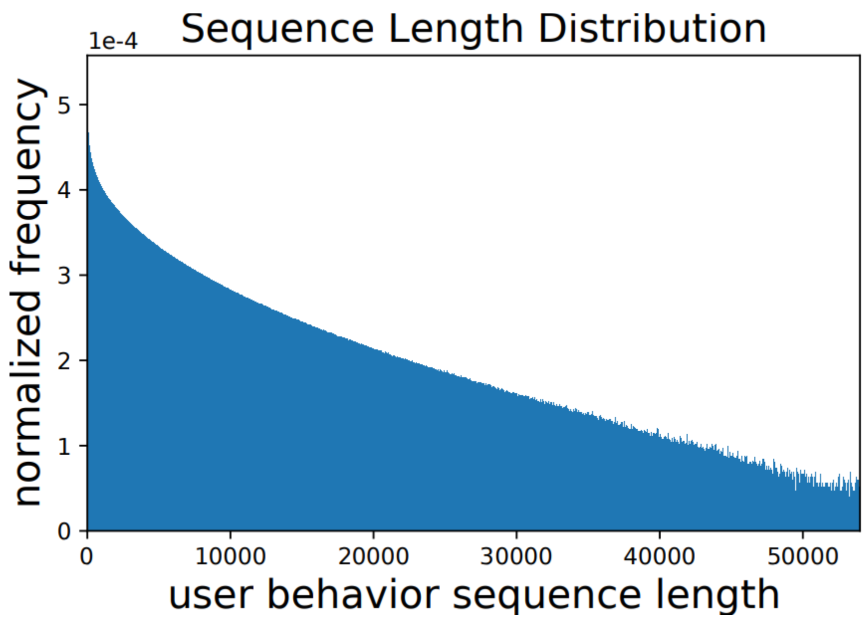
- Predict the probability of positive user response
  - Feature $\boldsymbol{x}$, including side-information and previous behaviors
  - Label y
  - Output $\Pr(y=1|\boldsymbol{x})$

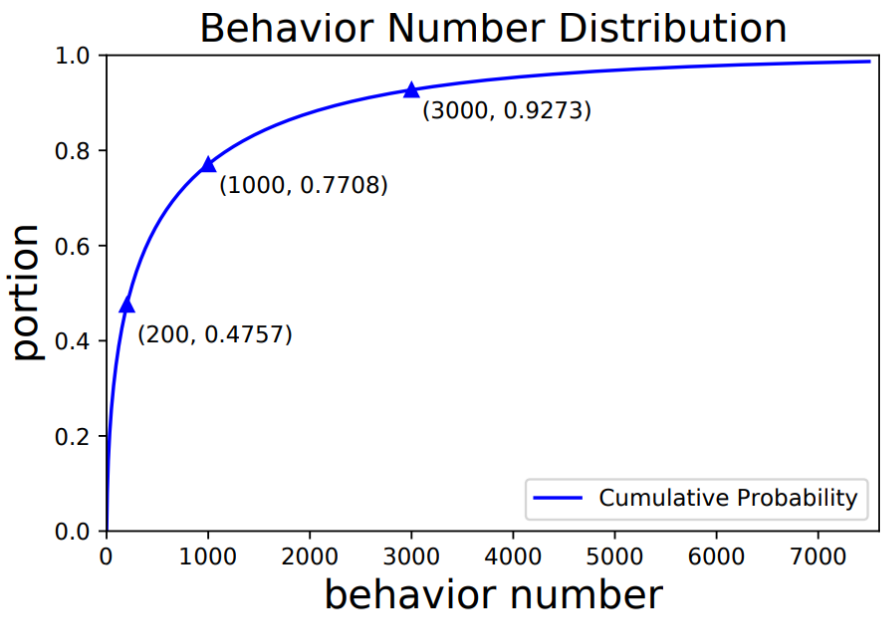
Response Type	Prediction Goal	Abbreviation
Click	Click-through Rate	CTR
Conversion	Conversion Rate	CVR

Sequential Modeling for User Behaviors

- Sequential user modeling
  - Conduct a comprehensive **user profiling** with the **historical user behaviors** and other side information and represent it in a unified framework.
- Usage
  - User targeting in online advertising
  - User behavior prediction
- Characteristics of user behaviors
  - Intrinsic and multi-facet user interests
  - Dynamic user interests and tastes
  - Multi-scale *sequential dependency* within behavior history

Analysis of User Behaviors (Alibaba)





Related Works

- Aggregation-base methods: w/o considering sequential dependencies
  - Matrix factorization (KDD’09)
  - SVD and other variants (KDD’09, KDD’13)
- State-based methods: simple state and transition assumption
  - Markov chain models (WWW’10, ICDM’16, RecSys’16)
- Deep learning methods: cannot handle long-term behavior sequences
  - Recurrent neural network models (ICLR’16, CIKM’18)
  - Convolutional neural network models (WSDM’18)

Lifelong Sequential Modeling

- Definition of Lifelong Sequential Modeling (LSM)
  - LSM is a process of continuous (online) user modeling with sequential pattern mining upon the lifelong user behavior history.

- Characteristics
  - supports **lifelong** memorization of user behavior patterns
  - conducts a **comprehensive** user modeling of intrinsic and dynamic user interests
  - continuous **adaptation** to the up-to-date user behaviors

Framework of LSM

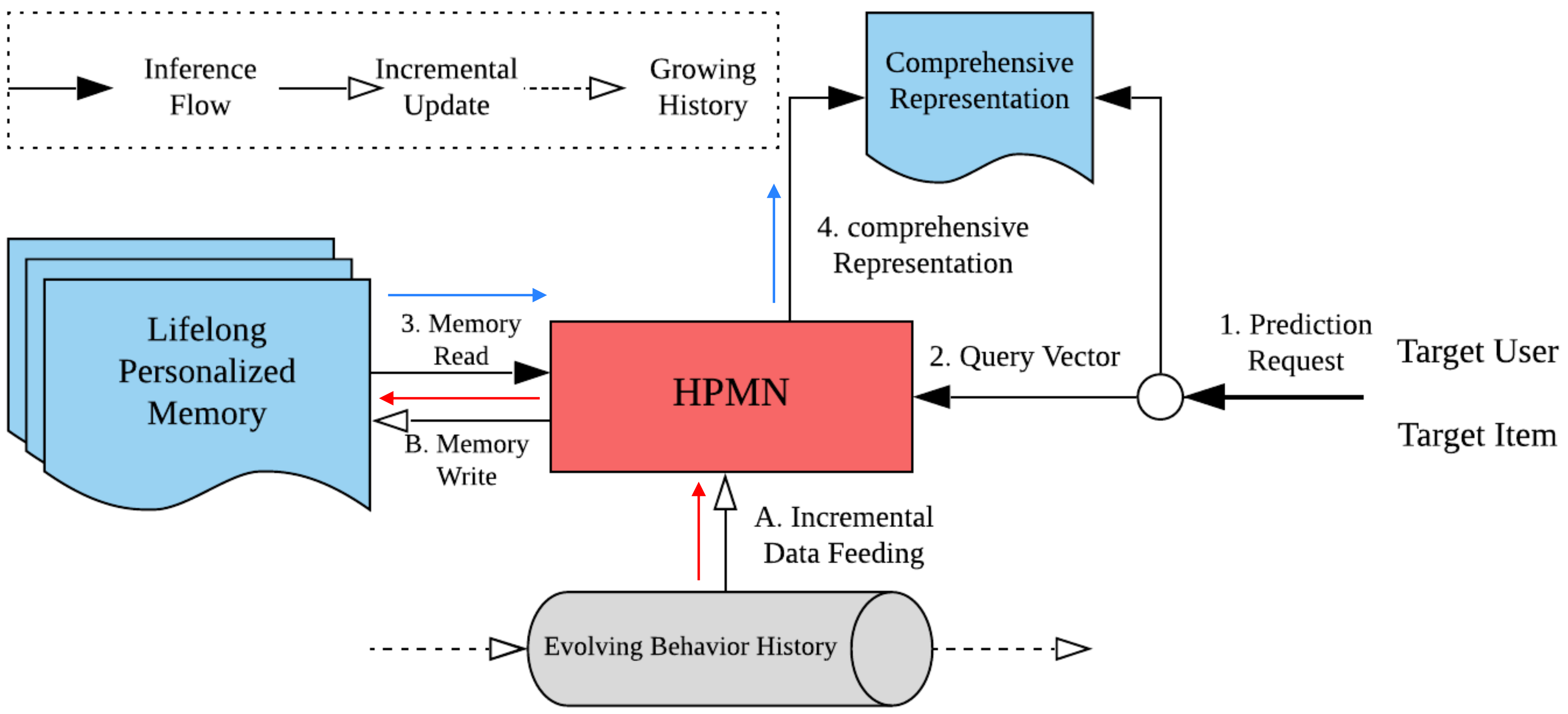


Figure 2: The LSM framework.

HPMN Model

- Hierarchical Periodical Memory Network, HPMN

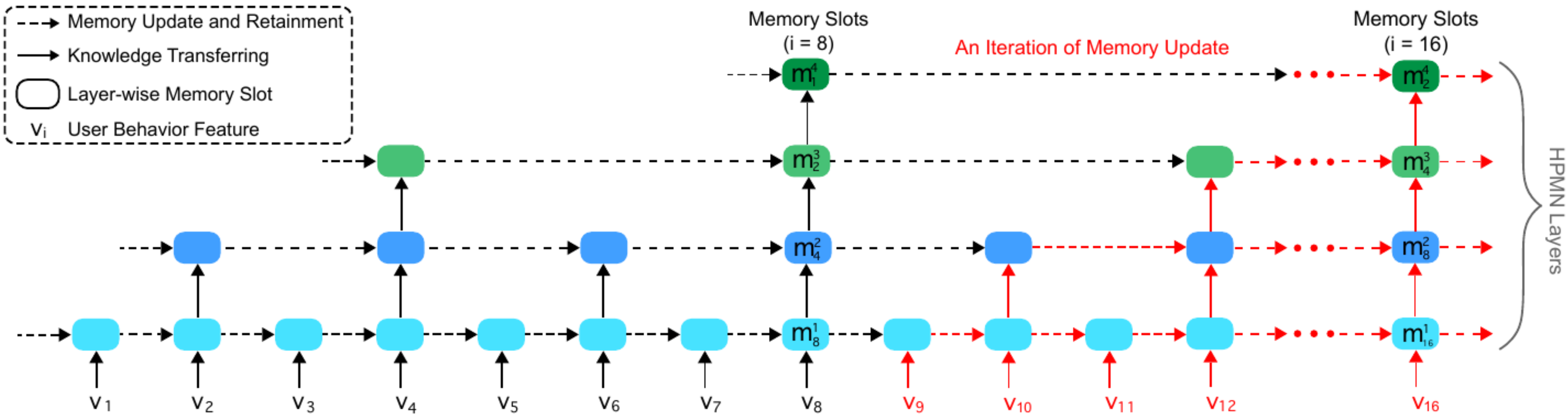


Figure 3: The framework of HPMN model with four layers maintaining user memory in four ($D = 4$) memory slots. The update period t^j of j-th layer follows an exponential sequence $\{2^{j-1}\}_{j=1}^{D}$ as an example. The red part means the incremental updating mechanism; the dotted line means the periodic memorization and forgetting.

User Response Prediction

- Real-time query only on the maintained user memory
  - w/o inference over the whole user behavior sequence online

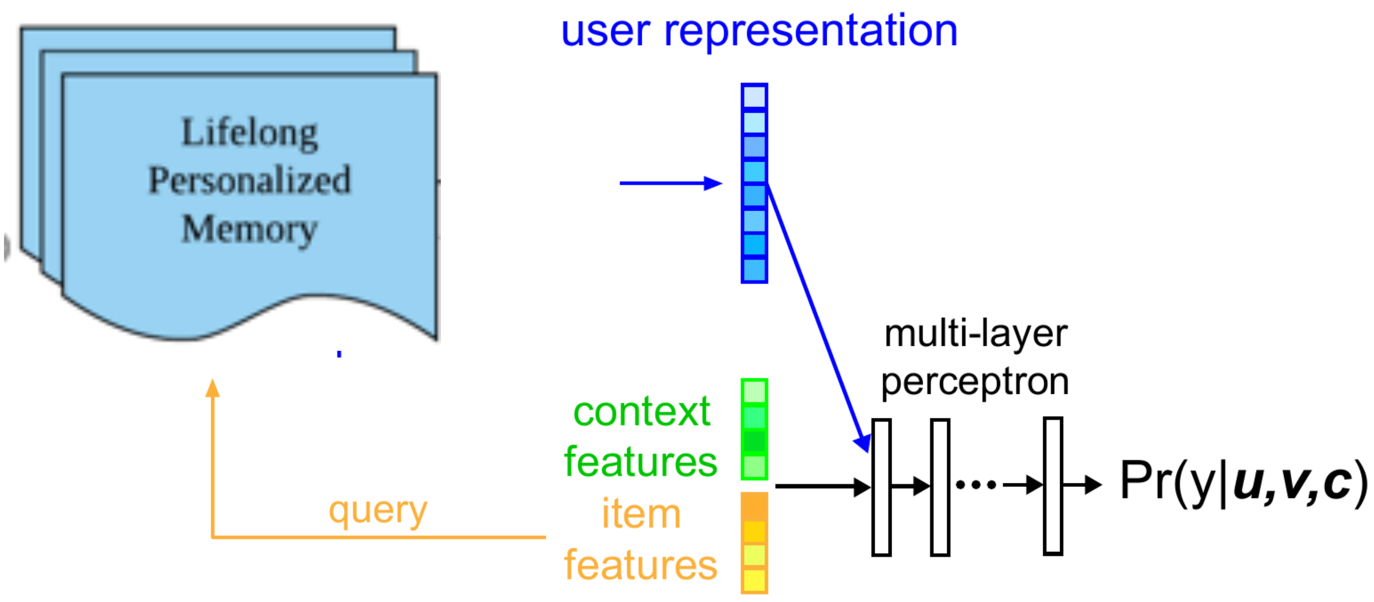


Figure 4: The overall user response prediction.

R/W Operations

- The content in the j-th memory slot at step i
  - $\{\boldsymbol{m}_i^j\}_{j=1}^D$
- Memory query and attentional reading
  - Given the query vector of the target item $\boldsymbol{v}$
  - Calculate the attention weight $w^j = E(\boldsymbol{m}^j, \boldsymbol{v})$ for each j-th memory slot
  - User representation $\boldsymbol{r} = \sum_j^D w^j \cdot \boldsymbol{m}^j$ at step i
- Periodical and gate-based (soft) writing

$$\boldsymbol{m}_i^j = \begin{cases} g^j\left(\boldsymbol{m}_i^{j-1}, \boldsymbol{m}_{i-1}^j\right) & \text{if } i \bmod t^j = 0\,, \\ \boldsymbol{m}_{i-1}^j & \text{otherwise}\,, \end{cases}$$

HPMN Model Training

- Offline model training
- Online memory maintaining

- Loss functions
  - Cross entropy loss
  - Memory covariance regularization
    - To enlarge covariance between each pair of memory slots
    - Help deal with multi-facet user interests
  - Parameter regularization

Experiment Setup

- Datasets

Table 2: The dataset statistics. T: length of the whole lifelong sequence (maximal length in the dataset). s: length of recent behavior sequence.

Dataset	Amazon	Taobao	XLong
User #	192,403	987,994	20,000
Item #	63,001	4,162,024	3,269,017
s	10	44	232
T	100	300	1,000

short → long
Sequence length

- Evaluation metrics
  - AUC
  - Log-loss

Compared Models

1. Aggregation-based methods
   1. DNN: utilizes sum-pooling for user behaviors
   2. SVD++: latent factor model
2. Short-term behavior modeling methods
   1. GRU4Rec: recurrent neural network model
   2. Caser: convolutional neural network model
   3. DIEN: dual RNN model w/ attention mechanism
   4. RUM: key-value memory network model
3. Long-term behavior modeling methods
   1. LSTM: long-short term memory model
   2. SHAN: hierarchical attention-based model
   3. HPMN: our model

Experiment Results

Table 4: Performance Comparison. (* indicates p-value < 10^{-6} in the significance test. ↑ and ↓ indicates the *performance* over lifelong sequences (with length T) is better or worse than the same model over short sequences (with length s). AUC: the higher, the better; Log-loss: the lower, the better. The second best performance of each metric is underlined.)

Model Group	Model	Len.	AUC Amazon	AUC Taobao	AUC XLong	Log-loss Amazon	Log-loss Taobao	Log-loss XLong
Group 2	GRU4Rec	s	0.7669	0.8431	0.8716	0.5650	0.4867	0.4583
	Caser	s	0.7509	0.8260	0.8467	0.5795	0.5094	0.4955
	DIEN	s	0.7725	0.8914	<u>0.8725</u>	0.5604	0.4184	<u>0.4515</u>
	RUM	s	0.7434	0.8327	0.8512	0.5819	0.5400	0.4931
Group 1	DNN	T	0.7546	0.7460	0.8152	0.6869	0.5681	0.5365
	SVD++	T	0.7155	0.8371	0.8008	0.6216	0.8371	1.7054
Group 2	GRU4Rec	T	0.7760 ↑	0.8471 ↑	0.8702 ↓	0.5569 ↑	0.4827 ↑	0.4630 ↓
	Caser	T	0.7582 ↑	0.8745 ↑	0.8390 ↓	0.5704 ↑	0.4550 ↑	0.5050 ↓
	DIEN	T	<u>0.7770</u> ↑	<u>0.8934</u> ↑	0.8716 ↓	<u>0.5564</u> ↑	<u>0.4155</u> ↑	0.4559 ↓
	RUM	T	0.7464 ↑	0.8370 ↑	0.8649 ↑	0.6301 ↓	0.4966 ↑	0.4620 ↑
Group 3	LSTM	T	0.7765	0.8681	0.8686	0.5612	0.4603	0.4570
	SHAN	T	0.7763	0.8828	0.8369	0.5595	0.4318	0.5000
	HPMN	T	**0.7809***	**0.9240***	**0.8929***	**0.5535***	**0.3487***	**0.4150***

Visualized Analysis

Conclusion

- First work proposes lifelong sequential modeling
- Construct hierarchical periodical memory network to model long-term sequential dependency
- Dynamic read-write operations
- Significantly improved the performance

- Acknowledgement
  - Alibaba Innovation Research (AIR)
  - National Natural Science Foundation of China

Alibaba Group