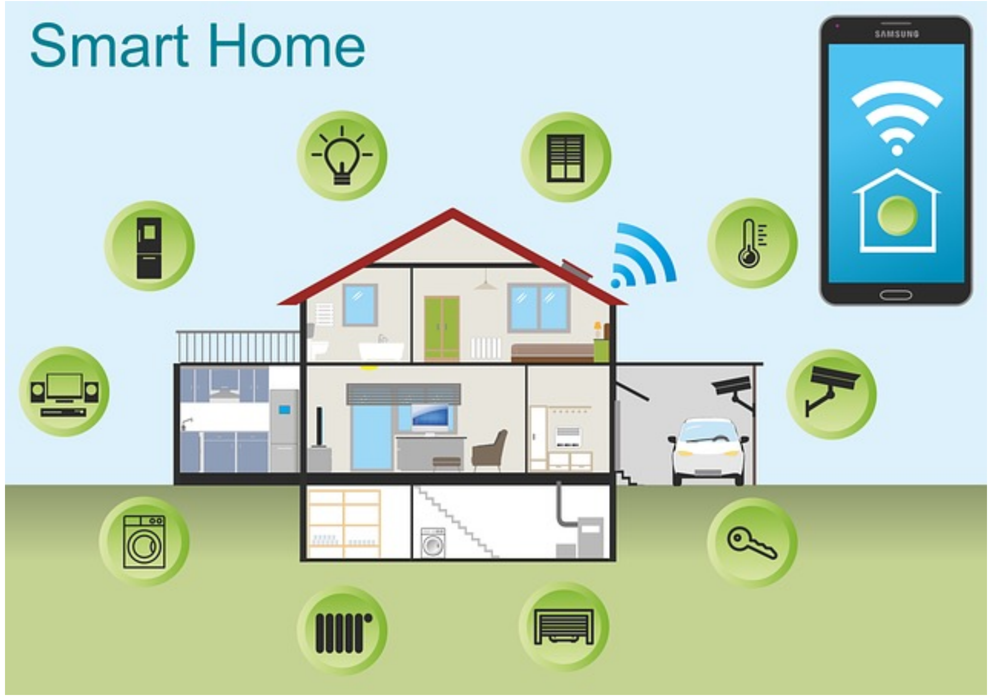
UBICOMP ISWC 2023 CANCUN, MEXICO

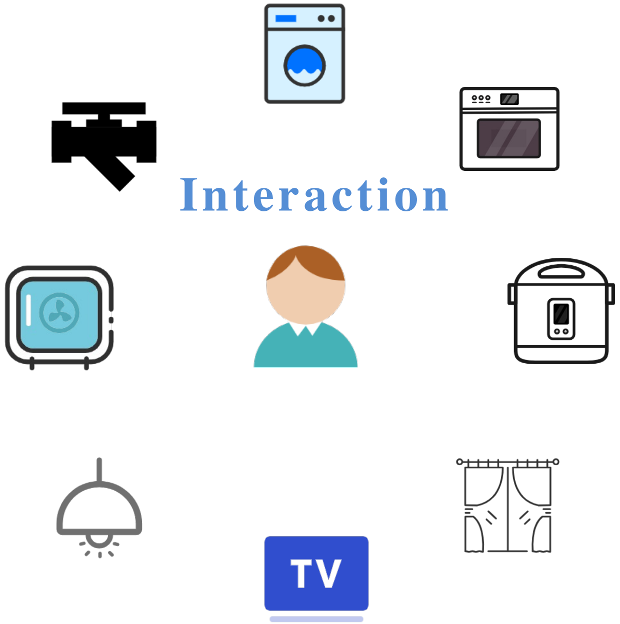
I Know Your Intent: Graph-Enhanced Intent-aware User Device Interaction Prediction via Contrastive Learning

Jingyu Xiao*, Qingsong Zou*, Qing Li ‡, Dan Zhao, Kang Li,
Zixuan Weng, Ruoyu Li, Yong Jiang

1 Background

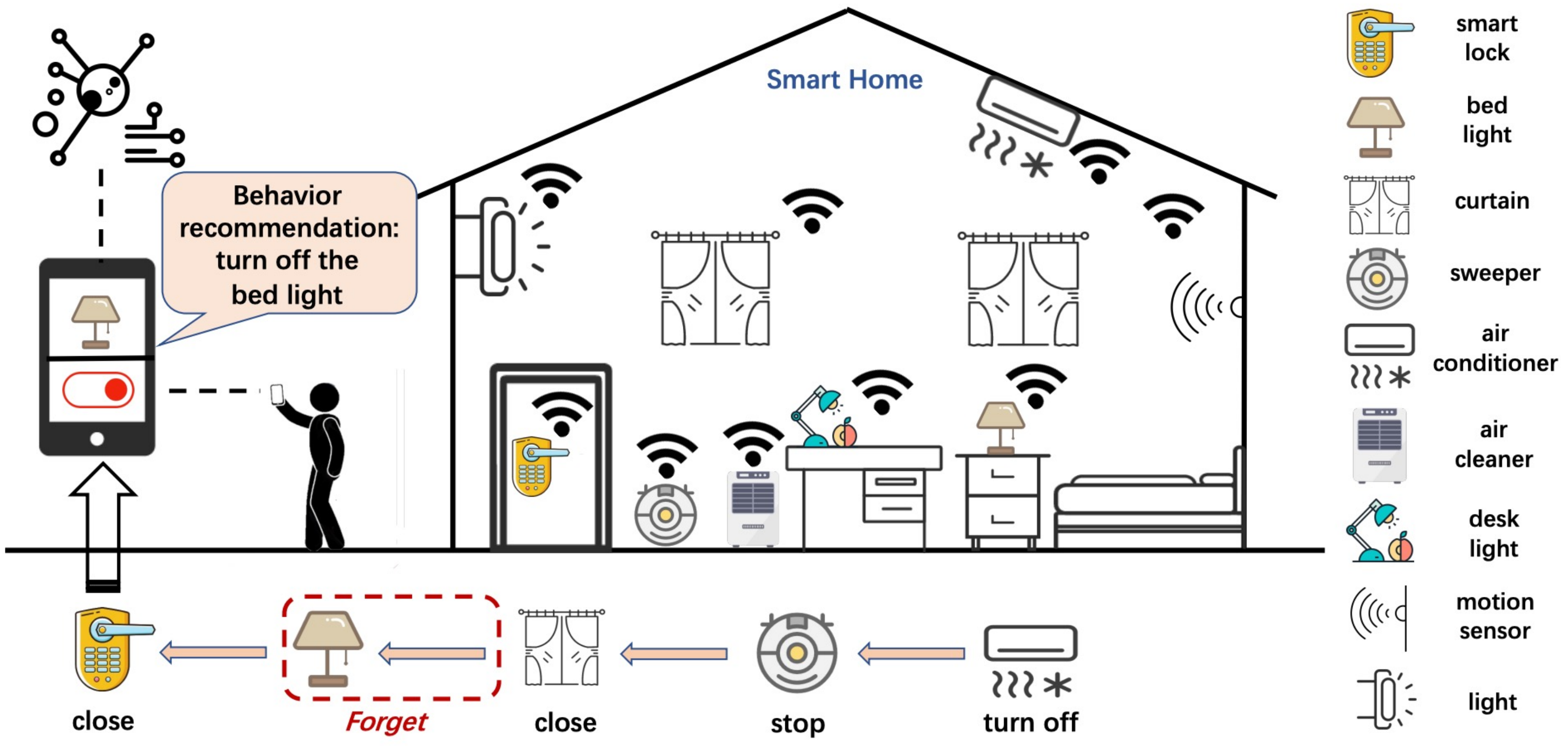
- Internet of Things (IoT) devices have been increasingly involved in smart home
- There are many User Device Interaction (UDI) sequences in people daily life.





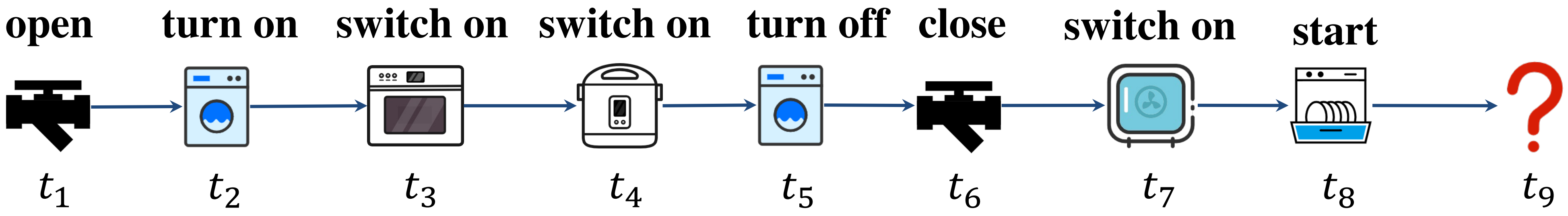
1 Background

- User Device Interaction (UDI) prediction is necessary for smart homes
  - Behavior Recommendation
  - Abnormal Behavior Identification



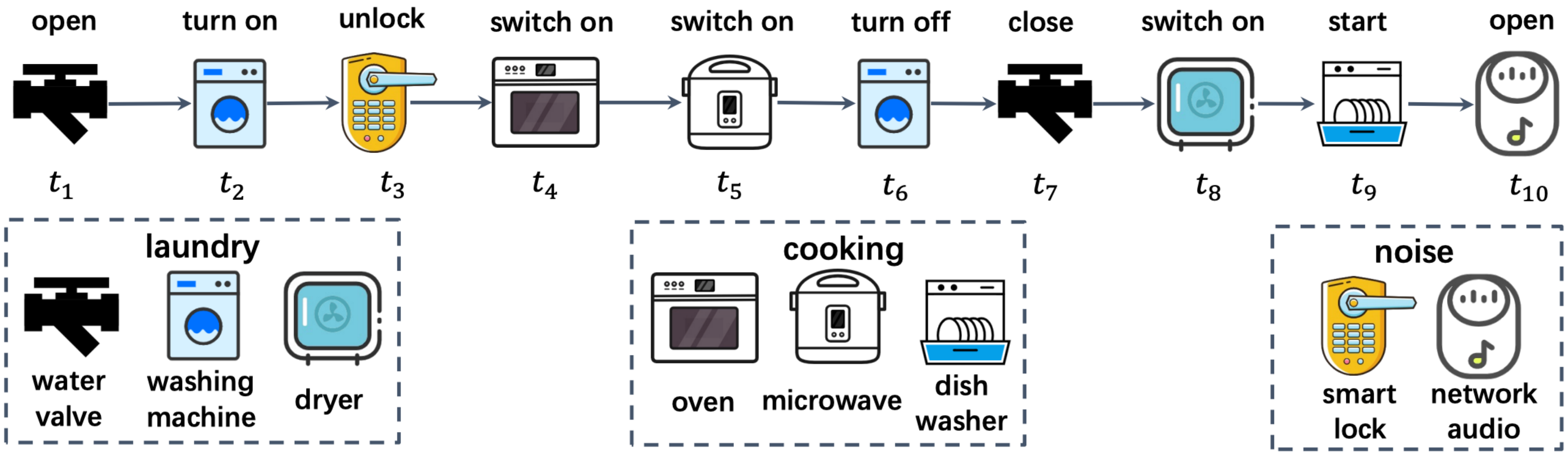
2 Problem Definition

- **User Device Interaction (UDI) prediction in smart homes**
  - **Given a behavior sequence s=[b_1, b_2,..., b_n.], where b=[t, d, c, i] consists of time t, device d, device control c and intent i. For example, b=[2022-10-15 11:30, oven, oven: switch, cooking] describes the behavior turn on the oven at 11:30 on 2022-10-15, with the intent of cooking.**
  - **The UDI prediction aims at predicting next behavior b_{n+1}.**



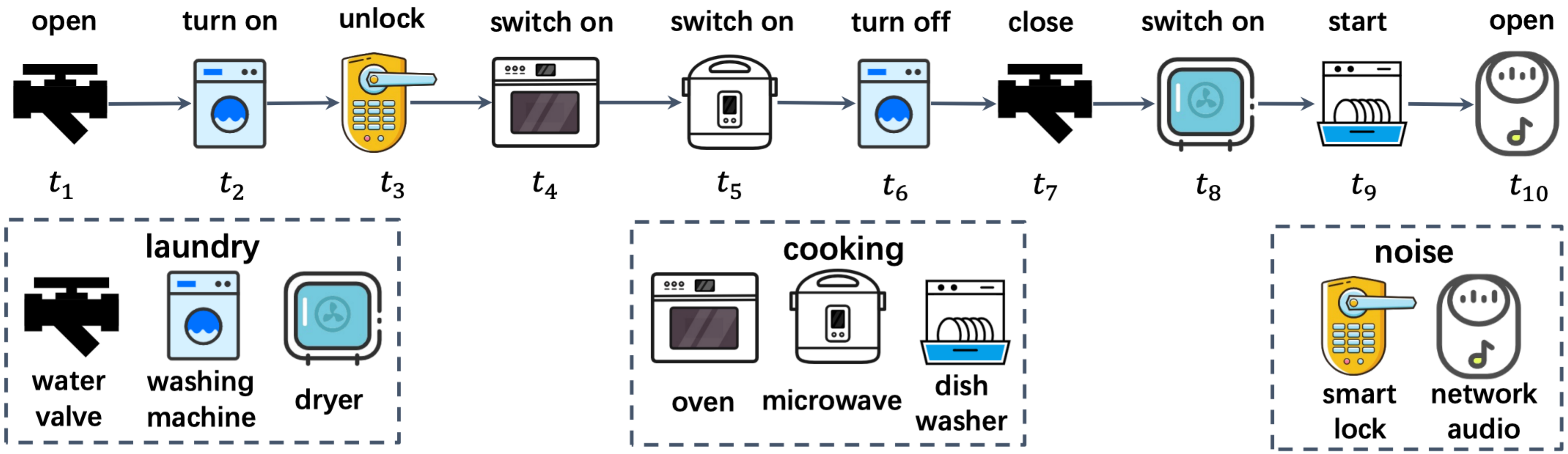
3 Key Factor #1

- **Routine** contains people's behavior correlations
  - Existence of noise behaviors between the routine behaviors causes the model to learn false correlations between noisy behaviors and routine behaviors which co-occur in the same sequence.



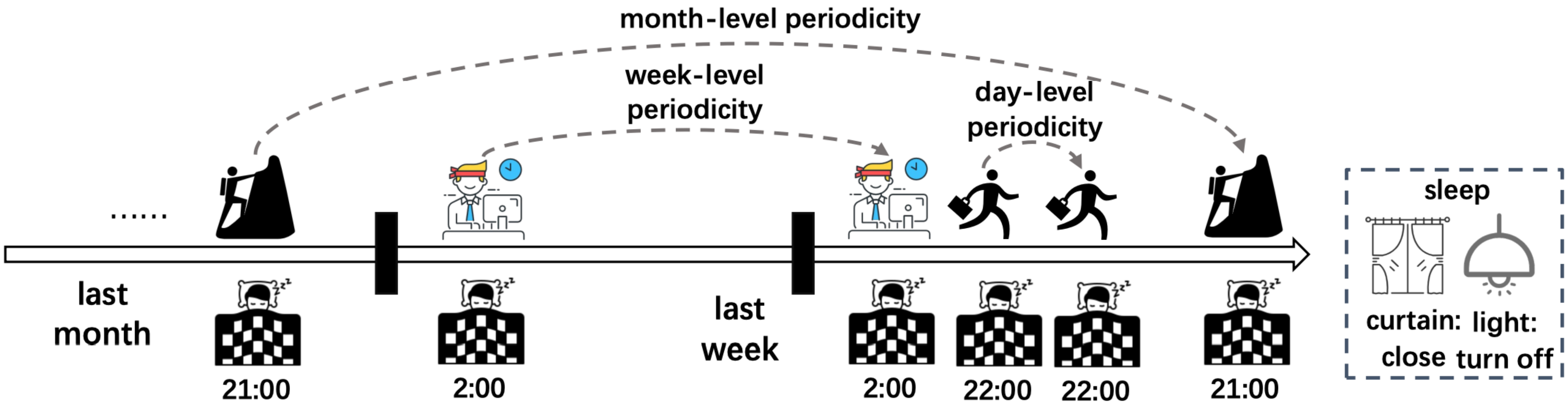
3 Key Factor #2

- **Intent determines the user behaviors:**
  - **There are multiple intents (e.g., laundry, cooking) in the behavior sequence.**
  - **There are complex transitions between different intents.**



3 Key Factor #3

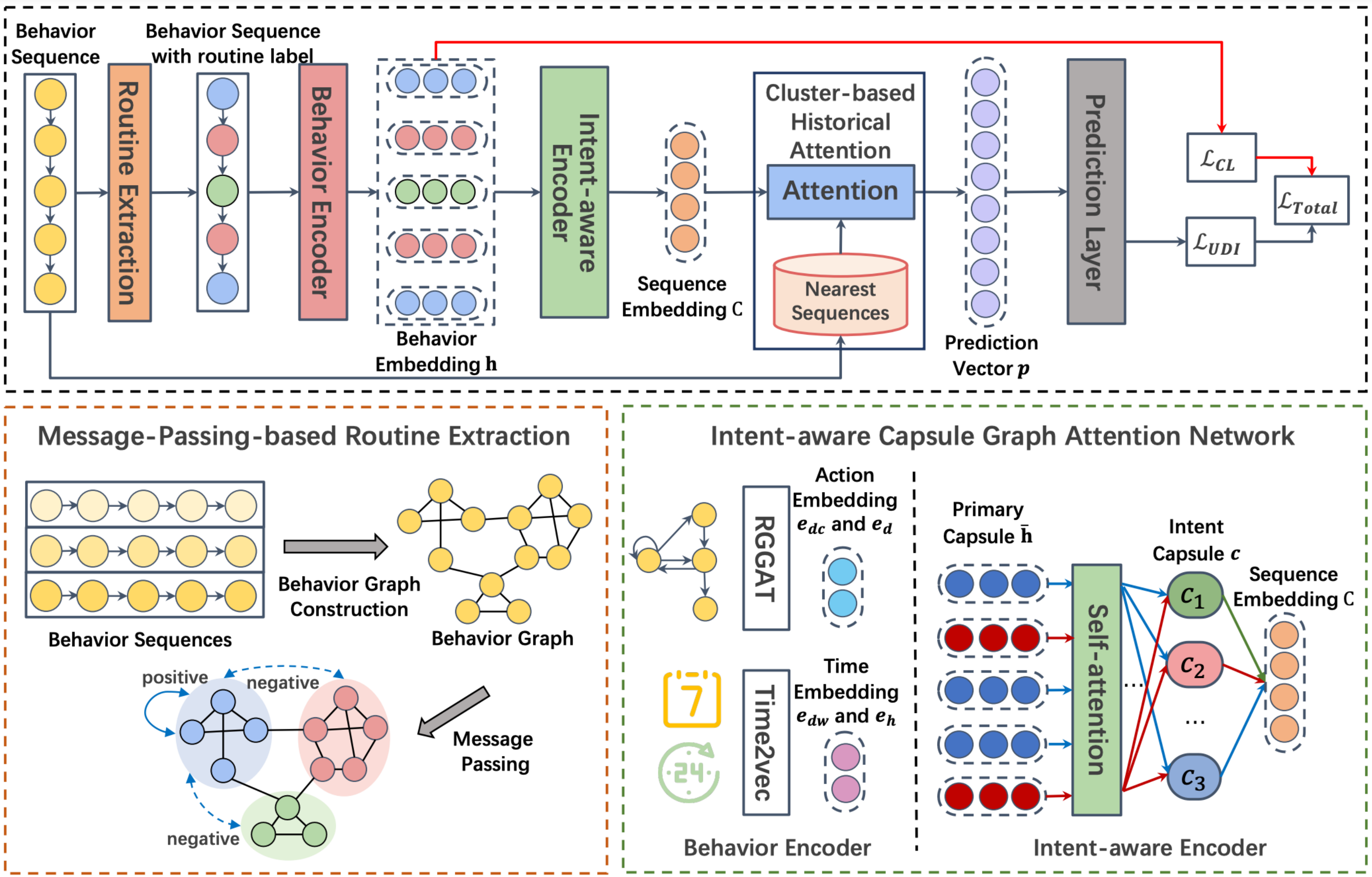
- **Multi-level Periodicity** reflects the behavior patterns:
  - There are different periodicities in the behavior sequences, such as month-level, week-level, day-level and so on.



4 Overview

- We propose **SmartUDI**:
  - A novel approach for accurate UDI prediction.
- Idea #1: **Message-Passing-based Routine Extraction**
  - Extract **Routine** via message passing and learn correct correlations via contrastive learning
- Idea #2: **Intent-aware Capsule Graph Attention Network**
  - View **Intents** as capsules and multiple intents by capsule network.
  - Leverage relational gated GAT to capture the transitions between different behaviors.
- Idea #3: **Cluster-based Historical Attention Mechanism**
  - Model correlation between current sequence and nearest historical sequences by attention mechanism to capture **Multi-level Periodicity**.

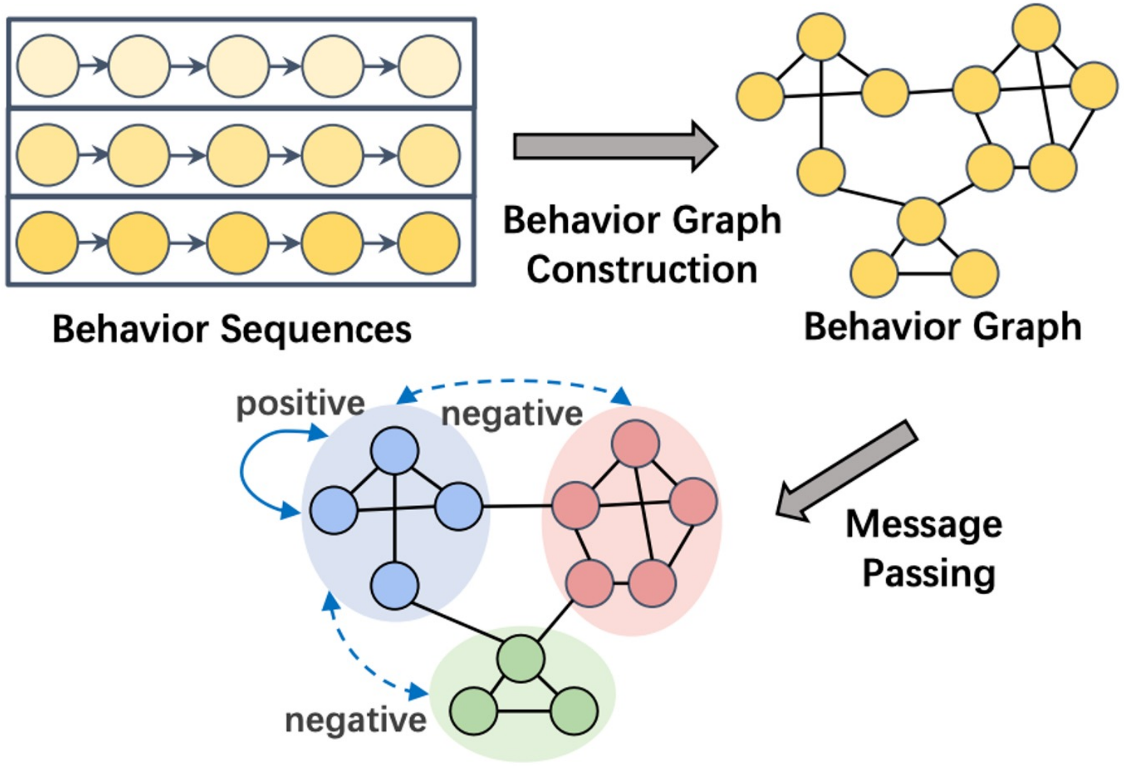
4 Overview



Behavior Sequence
Routine Extraction
Behavior Sequence with routine label
Behavior Encoder
Behavior Embedding h
Intent-aware Encoder
Sequence Embedding C
Cluster-based Historical Attention
Attention
Nearest Sequences
Prediction Vector p
Prediction Layer
$\mathcal{L}_{CL}$
$\mathcal{L}_{UDI}$
$\mathcal{L}_{Total}$
Message-Passing-based Routine Extraction
Behavior Sequences
Behavior Graph Construction
Behavior Graph
positive
negative
negative
Message Passing
Intent-aware Capsule Graph Attention Network
RGGAT
Action Embedding e_{dc} and e_d
Time2vec
Time Embedding e_{dw} and e_h
Behavior Encoder
Primary Capsule $\bar{h}$
Self-attention
Intent Capsule c
c_1
c_2
c_3
Sequence Embedding C
Intent-aware Encoder

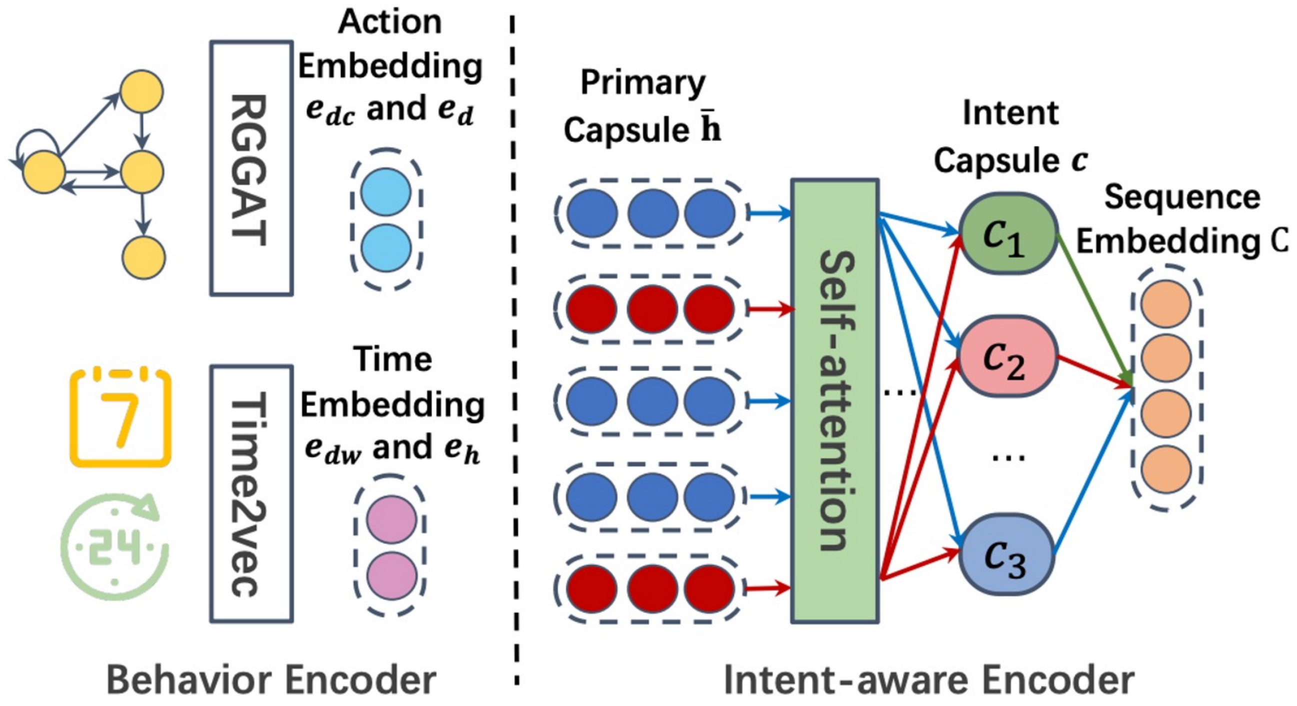
5 Message-Passing-based Routine Extraction

- **Construct the Behavior Graph: The weights in the graph represent the number of co-occurrences of behaviors.**
- **Initial the routine label based on K-clique algorithm.**
- **Update the routine label by message passing.**



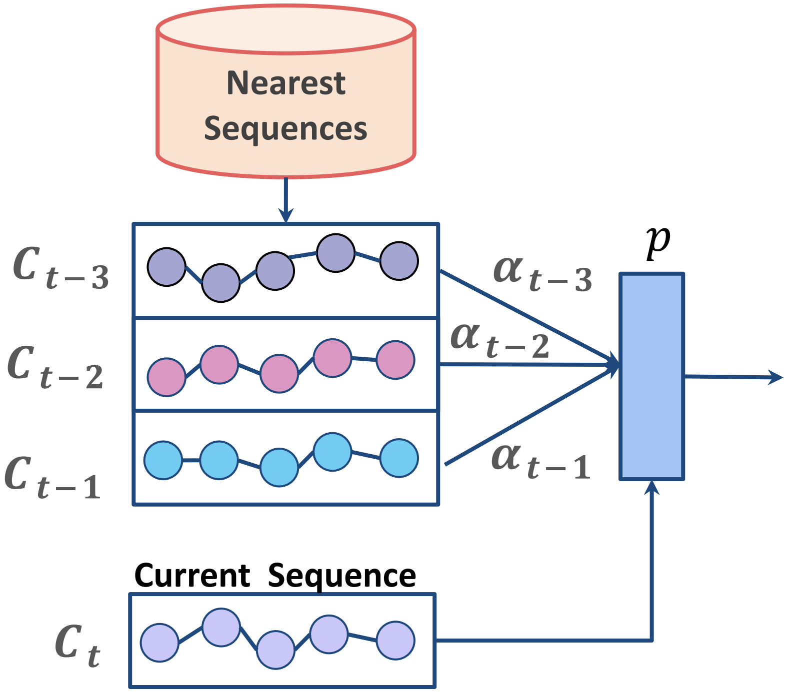
6 Intent-aware Graph Attention Network

- Time2Vec for Time Embedding.
- Relational Gated Graph Attention Network for Action Embedding.
- Self-attention and capsule network are employed to model multiple intents.



7 Cluster-based Historical Attention Mechanism

- **Cluster sequences by K-means.**
- **Summarize the nearest history sequence vector $[C_1, C_2, C_3, \ldots C_{t-1}]$ into p to capture the multi-level periodicity.**



$$\alpha_i = \frac{\exp(\beta_i)}{\sum_{j=1}^{t-1} \exp(\beta_j)}$$

$$\beta_i = \tanh(\mathbf{C}_t \mathbf{W}_H \mathbf{C}_i)$$

$$p = \mathrm{Concat}\left(\mathrm{C}_t, \sum_{i=1}^{t-1} \alpha_i \mathrm{C}_i\right)$$

8 Multi-task training

- **The similarity between two behavior representations**

$$\operatorname{sim}\left(\mathbf{h}_p, \mathbf{h}_q\right) = \mathbf{h}_p^{\top}\mathbf{h}_q / \left\|\mathbf{h}_p\right\| \left\|\mathbf{h}_q\right\|$$

- **Contrastive loss function**

$$\mathcal{L}\left(\mathbf{h}_i\right) = -\log \frac{\sum_{\mathbf{h}_j \in pos(\mathbf{h}_i), \mathbf{h}_j \neq \mathbf{h}_i} \exp\left(\operatorname{sim}\left(\mathbf{h}_i, \mathbf{h}_j\right)/\tau\right)}{\sum_{\mathbf{h}_k \in neg(\mathbf{h}_i), \mathbf{h}_k \neq \mathbf{h}_i} \exp\left(\operatorname{sim}\left(\mathbf{h}_i, \mathbf{h}_k\right)/\tau\right)}$$

$$\mathcal{L}_{CL} = \frac{1}{\|\mathcal{S}\|} \sum_{s \in \mathcal{S}} \sum_{\mathbf{h}_i \in s} \mathcal{L}\left(\mathbf{h}_i\right)$$

- **UDI prediction loss function**

$$\mathcal{L}_{UDI}(X, Y) = -\frac{1}{|\mathcal{S}|} \sum_{s \in \mathcal{S}} \mathbf{y}_s \log \hat{\mathbf{y}}_s$$

- **Total loss function**

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{UDI}} + \lambda \mathcal{L}_{\text{CL}}$$

9 Datasets

- We use four real-world datasets to evaluate SmartUDI
  - US/SP/FR from public dataset, AN collected by ourselves.
  - Datasets are split into training/validation/testing with a ratio of 7:1:2.
  - All sequence instances are of length 10, and we use the first 9 behaviors as input to predict the next behavior.
  - Eight intents: entertainment, shower, sleep/getup, leave/return, study, cooking, cleaning, others.

Name	Time period (Y-M-D)	Sizes	# Devices	# Device controls
US	2022-02-22~2022-03-21	67,882	40	268
SP	2022-02-28~2022-03-30	15,665	34	234
FR	2022-02-27~2022-03-25	4,423	33	222
AN	2022-07-31~2022-08-31	1,765	36	141

Baselines and Evaluation Metrics

- **Baselines: we compare SmartUDI with 11 competitors**
  - **Traditional Models: HMM and FPMC**
  - **RNN-based Models: LSTM, CA-RNN, SIAR and DeepMove**
  - **CNN-based Models: Casers**
  - **GNN-based Models: SR-GNN**
  - **Transformer-based Models: SASRec, SmartSense and DeepUDI**
- **Evaluation Metrics:**
  - **Acc@K: Top-K accuracy**

$$\text{Acc@K} = \frac{|\{s \in S : p(s) \in P_K(s)\}|}{|S|}$$

  - **Macro-F1: Macro averaging of F1 score**

$$\text{Macro-F1} = \frac{\sum_c \text{F1}_c}{|C|}$$

11 Questions

UBICOMP ISWC 2023 CANCUN, MEXICO

- **RQ1 (Performance).** Compared with other methods, can SmartUDI predict user device interaction more accurately?
- **RQ2 (Ablation study).** How does each main component of SmartUDI affects the performance of UDI prediction?
- **RQ3 (Parameter study).** How do key parameters affect the SmartUDI?
- **RQ4 (Interpretability study).** Can SmartUDI give a reasonable explanation?
- **RQ5 (Embedding space analysis).** Does SmartUDI successfully learn useful embedding vectors of behaviors and correct correlations between behaviors?

Experimental Results

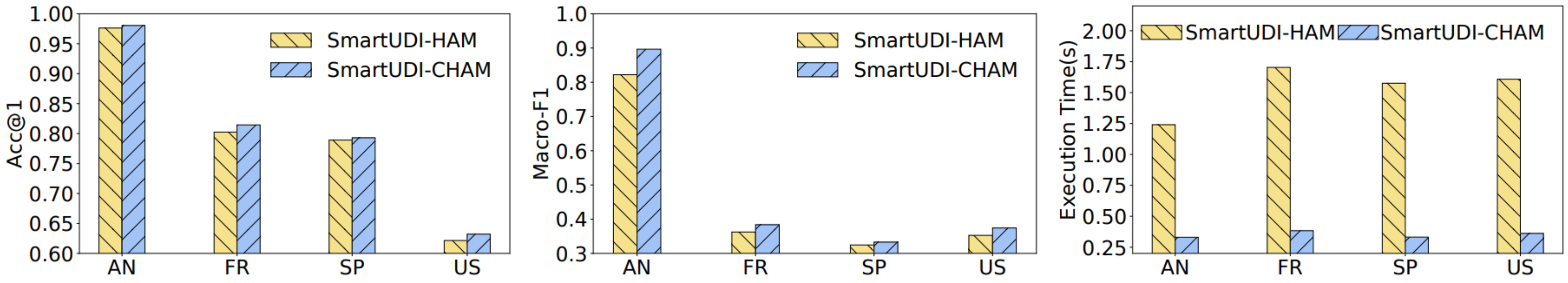
- **RQ1:** Compared with other methods, can SmartUDI predict user device interaction more accurately?
- **A1:** SmartUDI outperforms all competitors.

Dataset	Metric	HMM	FPMC	LSTM	CARNN	Caser	DeepMove	SIAR	SRGNN	SASRec	SmartSense	DeepUDI	SmartUDI
AN	Acc@1	0.6099	0.6557	0.7062	0.7026	0.7054	0.7116	0.7238	0.9245	0.9325	0.9407	0.9784	**0.9805**
	Acc@3	0.7501	0.7959	0.7843	0.8302	0.8569	0.9272	0.9354	0.9864	0.9665	0.9731	0.9865	**0.9874**
	Acc@5	0.7714	0.7902	0.8328	0.9003	0.9273	0.9542	0.9645	0.9872	0.9765	0.9838	0.9882	**0.9892**
	Macro-F1	0.2439	0.2845	0.3759	0.4159	0.4467	0.5027	0.5259	0.7368	0.7432	0.7519	0.7997	**0.8966**
FR	Acc@1	0.6536	0.6814	0.6962	0.7893	0.7742	0.7762	0.7796	0.7819	0.7821	0.7923	0.8144	**0.8145**
	Acc@3	0.7813	0.8271	0.8011	0.9148	0.9201	0.9221	0.9120	0.9197	0.9204	0.9232	0.9237	**0.9238**
	Acc@5	0.8242	0.8508	0.8565	0.9425	0.9414	0.9446	0.9420	0.9435	0.9362	0.9379	0.9511	**0.9512**
	Macro-F1	0.1127	0.1279	0.1302	0.2102	0.2158	0.2288	0.2312	0.2482	0.2473	0.2603	0.3425	**0.3837**
SP	Acc@1	0.6315	0.6964	0.7517	0.7853	0.7721	0.7756	0.7802	0.7815	0.7821	0.7921	0.7923	**0.7930**
	Acc@3	0.7863	0.7916	0.8864	0.8915	0.9045	0.9125	0.9217	0.9303	0.9321	0.9342	0.9375	**0.9427**
	Acc@5	0.8361	0.8605	0.9346	0.9117	0.9273	0.9521	0.9597	0.9603	0.9560	0.9511	0.9642	**0.9671**
	Macro-F1	0.1382	0.1586	0.1756	0.1745	0.1927	0.2159	0.2176	0.2239	0.2254	0.2244	0.3112	**0.3328**
US	Acc@1	0.3327	0.3543	0.4286	0.5212	0.5378	0.5527	0.5633	0.5784	0.5826	0.5935	0.6056	**0.6321**
	Acc@3	0.6881	0.6992	0.8209	0.8577	0.8632	0.8844	0.8902	0.8955	0.8972	0.9056	0.9123	**0.9058**
	Acc@5	0.7258	0.7712	0.8929	0.9135	0.9266	0.9418	0.9432	0.9463	0.9320	0.9489	0.9521	**0.9538**
	Macro-F1	0.1069	0.1123	0.1265	0.1396	0.1576	0.2388	0.2397	0.2431	0.2433	0.2451	0.3538	**0.3742**

Experimental Results

- **RQ2:** How does each main component of SmartUDI affects the performance?
- **A2:** All three components (MPRE, ICGAT and CHAM) of SmartUDI are contributive for UDI prediction. Cluster improve the performance and efficiency of historical attention mechanism.

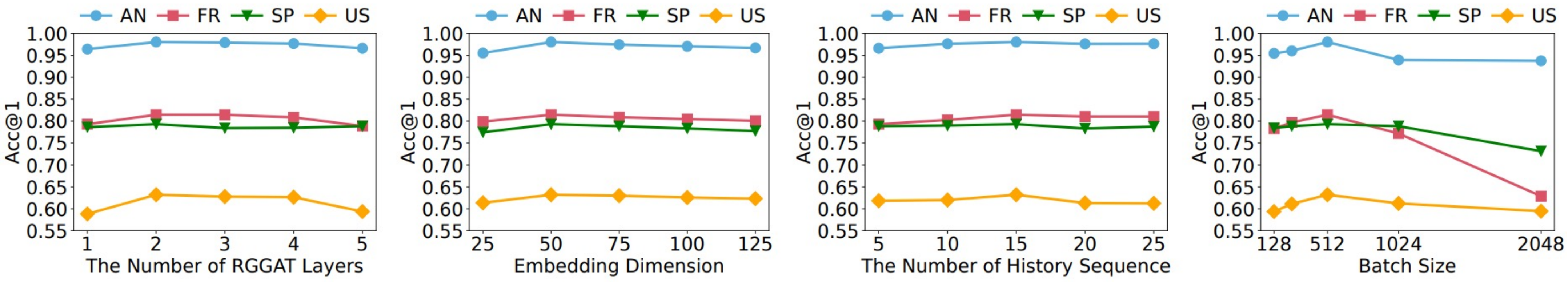
Model	AN		FR		SP		US	
	Acc@1	Macro-F1	Acc@1	Macro-F1	Acc@1	Macro-F1	Acc@1	Macro-F1
SmartUDI(w/o MPRE)	0.9622	0.7689	0.7871	0.3093	0.7832	0.3072	0.6028	0.3375
SmartUDI(w/o ICGAT)	0.9568	0.7354	0.7773	0.3199	0.7763	0.2846	0.5872	0.3252
SmartUDI(w/o CHAM)	0.9757	0.8018	0.7929	0.3458	0.7881	0.3235	0.6125	0.3486
SmartUDI(w/o ALL)	0.9137	0.6934	0.7578	0.2511	0.7685	0.2482	0.5736	0.2476
SmartUDI	**0.9805**	**0.8966**	**0.8145**	**0.3837**	**0.7930**	**0.3328**	**0.6321**	**0.3742**



(a) Acc@1. (b) Macro-F1. (c) Execution time per batch.

Experimental Results

- **RQ3:** **How do key parameters affect the SmartUDI?**
- **A3: The best parameter combination: #of layers of RGGAT=2, Embedding Dimension=50, # of History Sequence=15, Batch Size=512.**



(a) # of layers of RGGAT. (b) Embedding dimension. (c) # of history behavior sequences. (d) Batch size.

Experimental Results

- **RQ4:** Can SmartUDI give a reasonable explanation?
- **A4:** SmartUDI can interpret the results based on routine extraction results, intent capsule weight.

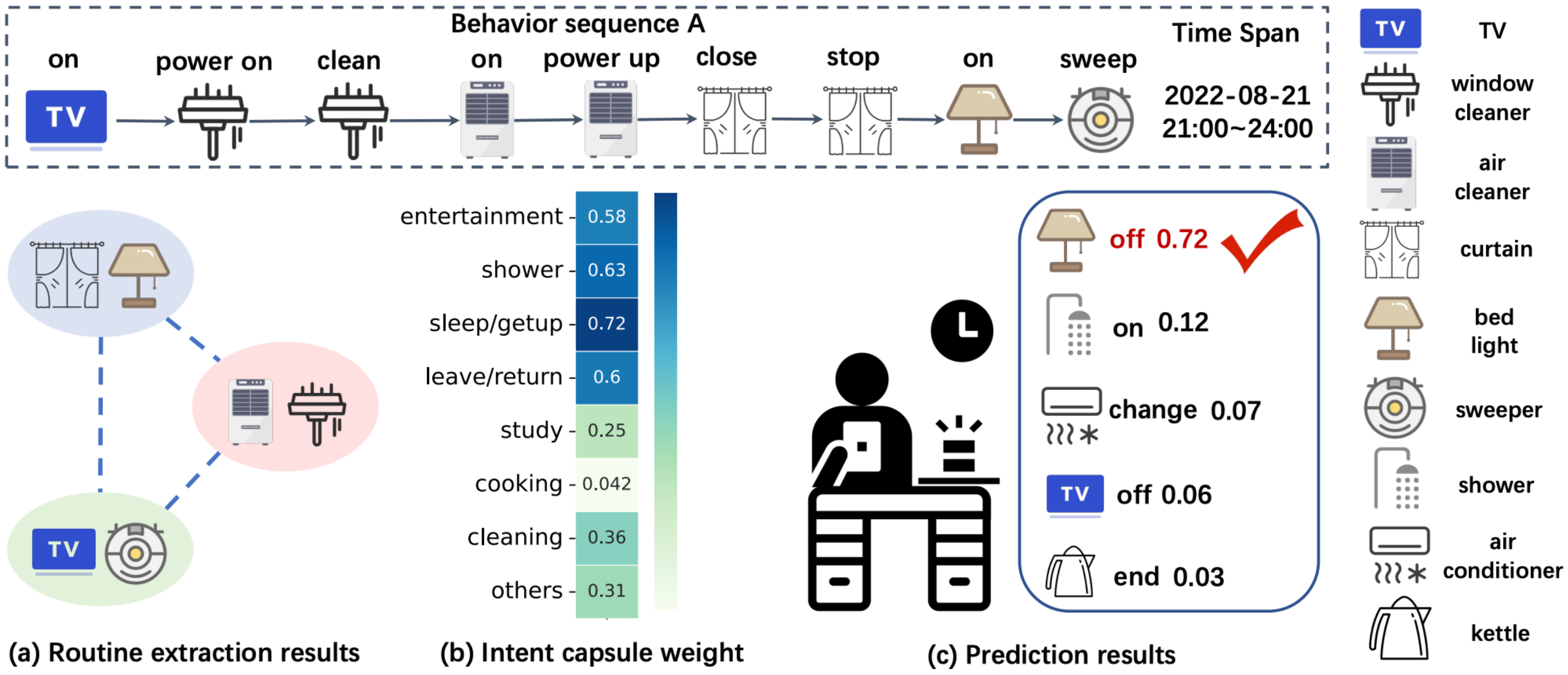


Fig. 8. (a) Routine extraction results, (b) capsule weight and (c) top 5 prediction results of the example.

Experimental Results

- **RQ4:** Can SmartUDI give a reasonable explanation?
- **A4:** SmartUDI can interpret the results based on historical attention score.

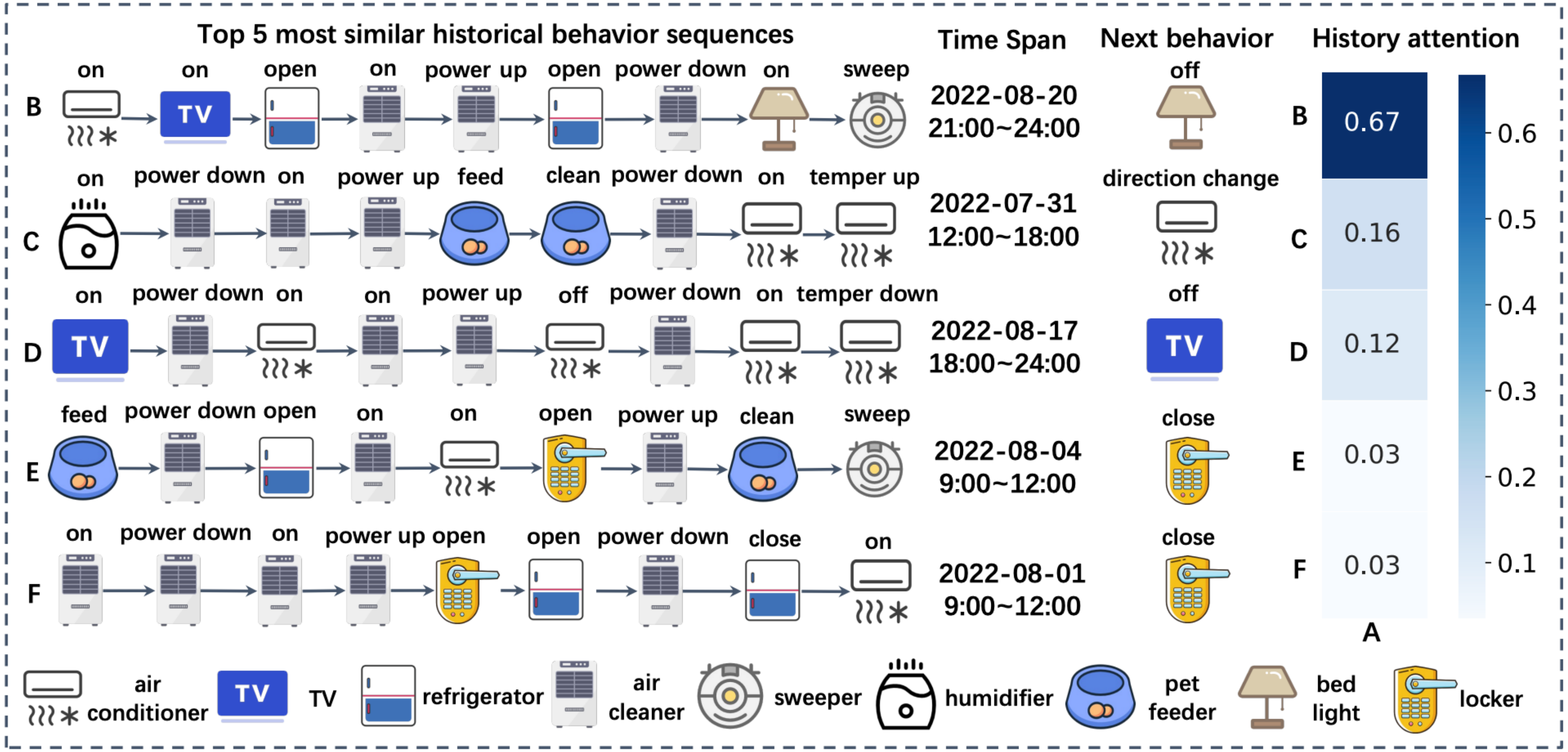
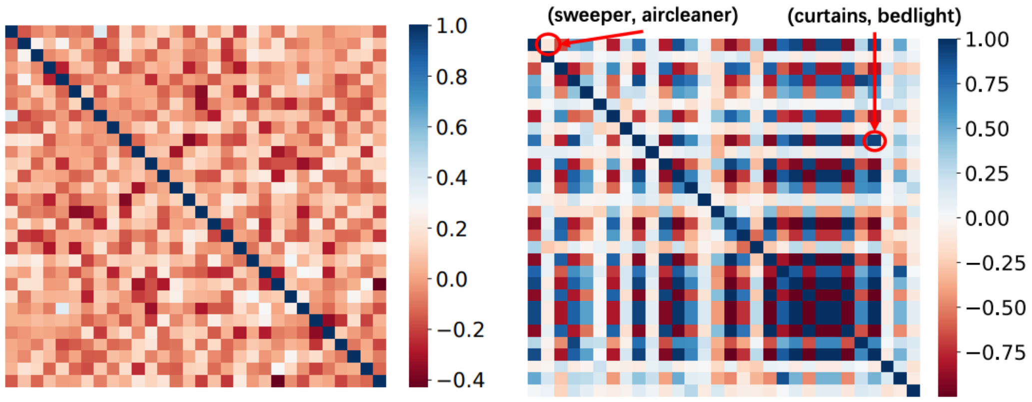


Fig. 9. Top 5 most similar historical sequences, time span, next behavior and historical attention score.

Experimental Results

- **RQ5:** **Does SmartUDI successfully learn useful embedding vectors of behaviors and correct correlations between behaviors?**
- **A5:** **After applying contrastive learning, SmartUDI can learn the correlations between device.**



(a) Device embedding w/o CL. (b) Device embedding with CL.

13 Conclusions

- **We propose SmartUDI for accurate UDI prediction.**
- **Our main contributions are summarized as follows:**
  - **Idea #1: Message-Passing-based Routine Extraction**
  - **Idea #2: Intent-aware Gated Graph Attention Network**
  - **Idea #3: Cluster-based Historical Attention Mechanism**
- **SmartUDI consistently outperforms state-of-the-art baselines and also offers highly interpretable results.**

Thank you!

- **Speaker: Jingyu Xiao**
- **Homepage**: https://whalexiao.github.io/
- **Email**: jy-xiao21@mails.tsinghua.edu.cn

UBICOMP ISWC
2023 CANCUN, MEXICO