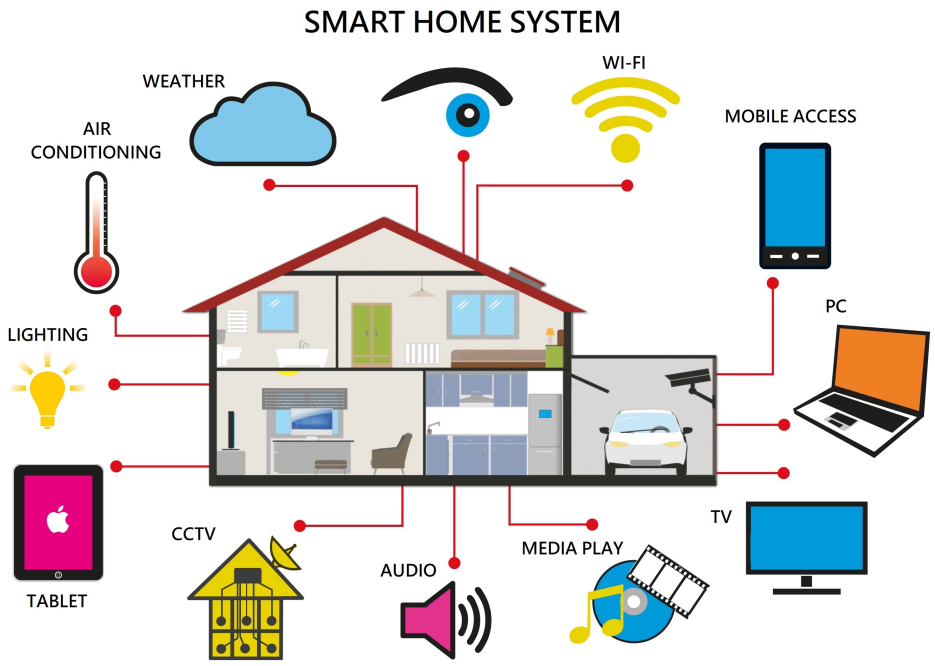
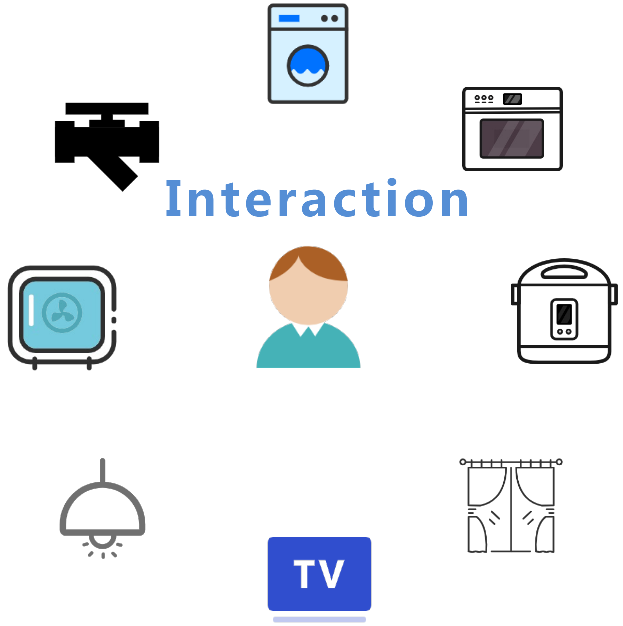
User Device Interaction Prediction via Relational Gated Graph Attention Network and Intent-aware Encoder

Jingyu Xiao*, Qingsong Zou*, Qing Li ‡, Dan Zhao, Kang Li,
Wenxin Tang, Runjie Zhou, Yong Jiang

1 Background

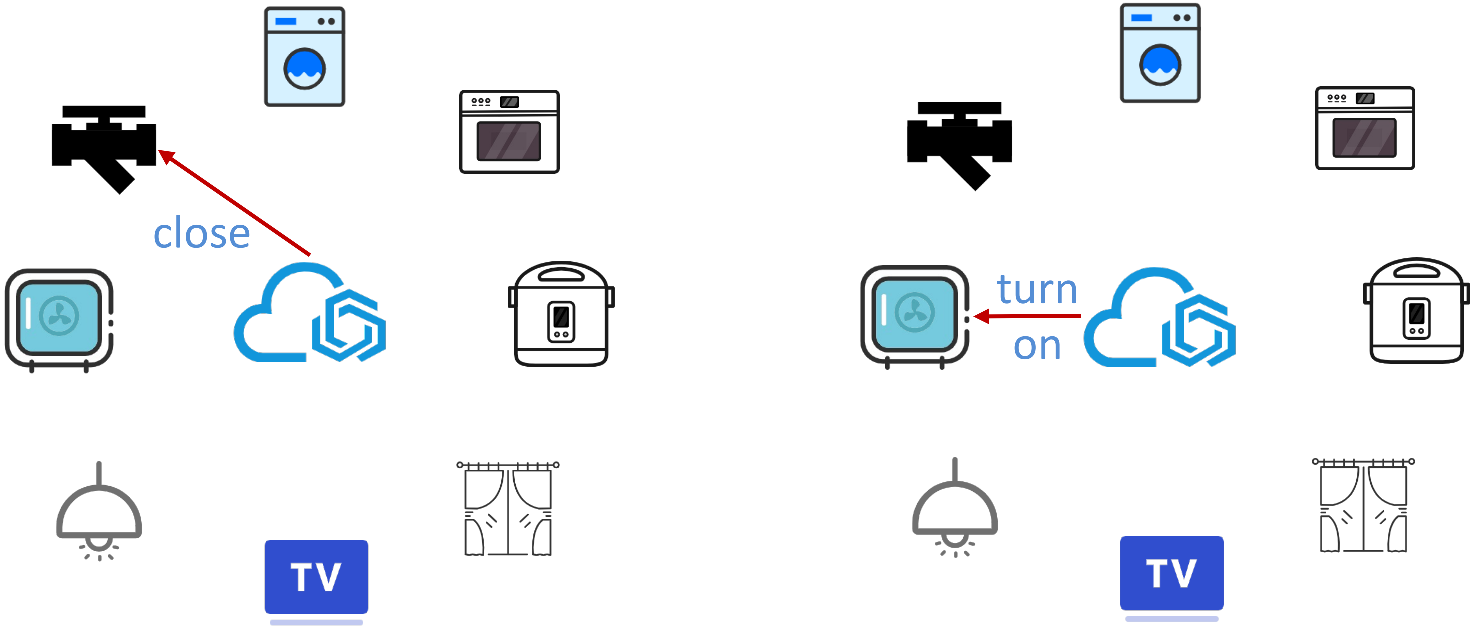
- **Smart Home**: Internet of Things (IoT) devices have been increasingly involved in home life.
- **User Device Interaction (UDI):** interaction between user and device reflects the habits of users using IoT devices.





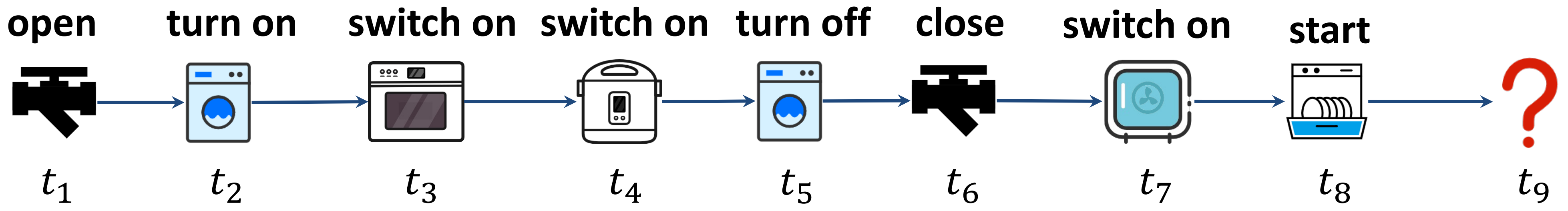
1 Background

- **User Device Interaction (UDI) prediction is necessary for smart homes**
  - It can help **remind users of key operations** they have forgotten, e.g., “close the water valve”
  - It can help **automate device control** and realize fully automatic whole-house intelligence, e.g., turn on the dryer automatically after washing the clothes.



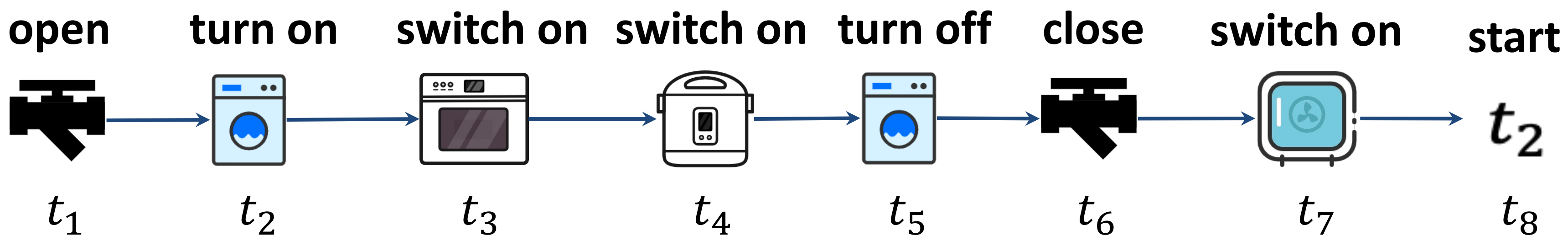
2 Problem Definition

- **User Device Interaction (UDI) Prediction in Smart Home:**
  - 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}.



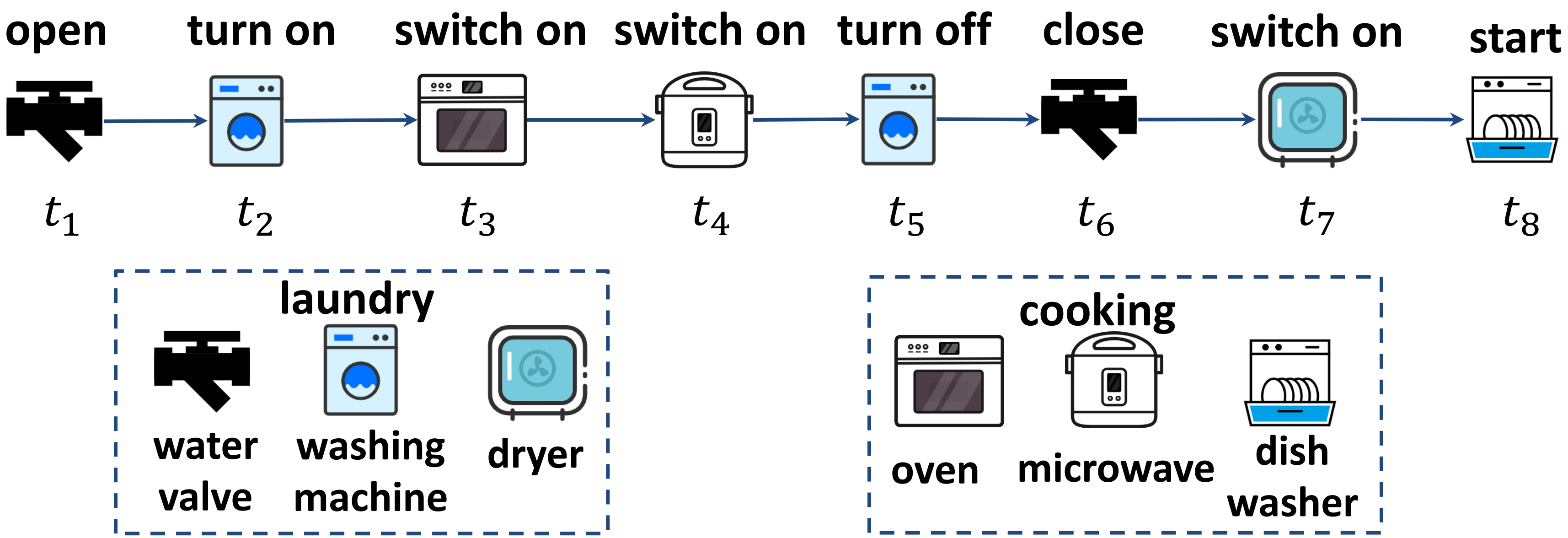
3 Challenge #1

- **Complex and Heterogeneous Transitions:**
  - **Complex:** there are transitions not only between consecutive devices, but also in broader contexts.
  - **Heterogeneous:** transitions caused by different device control.



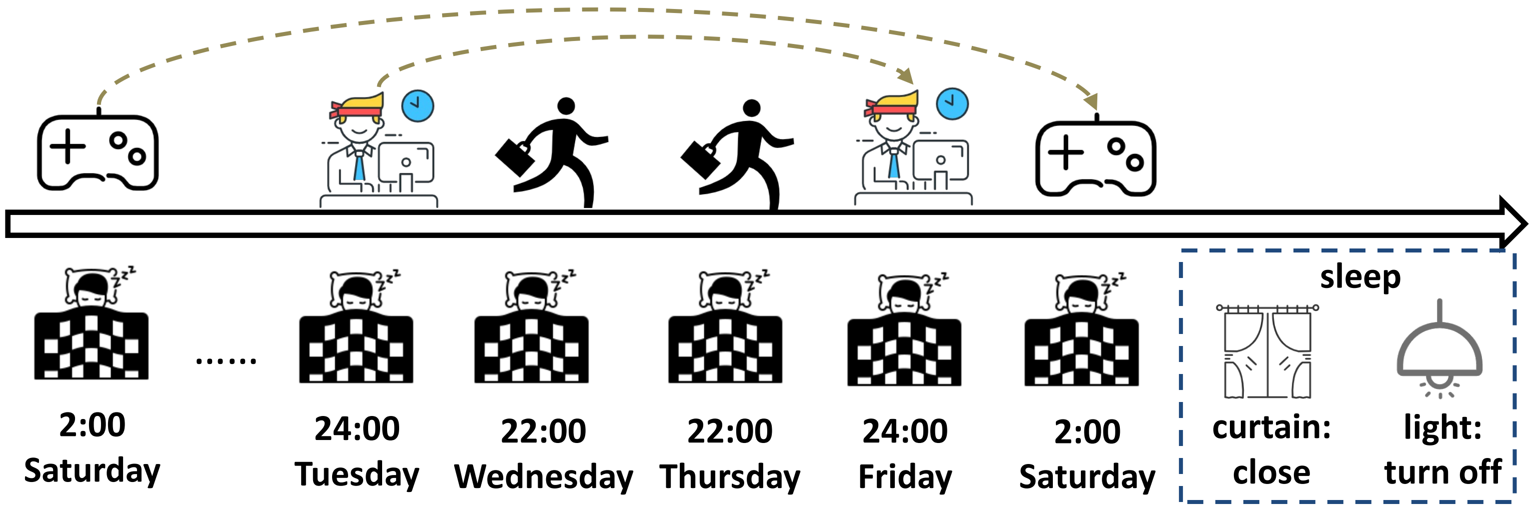
3 Challenge #2

- **Multiple Intents:**
  - **Laundry-related:** water valve, washing machine and dryer.
  - **Cooking-related:** oven, microwave and dish washer.



3 Challenge #3

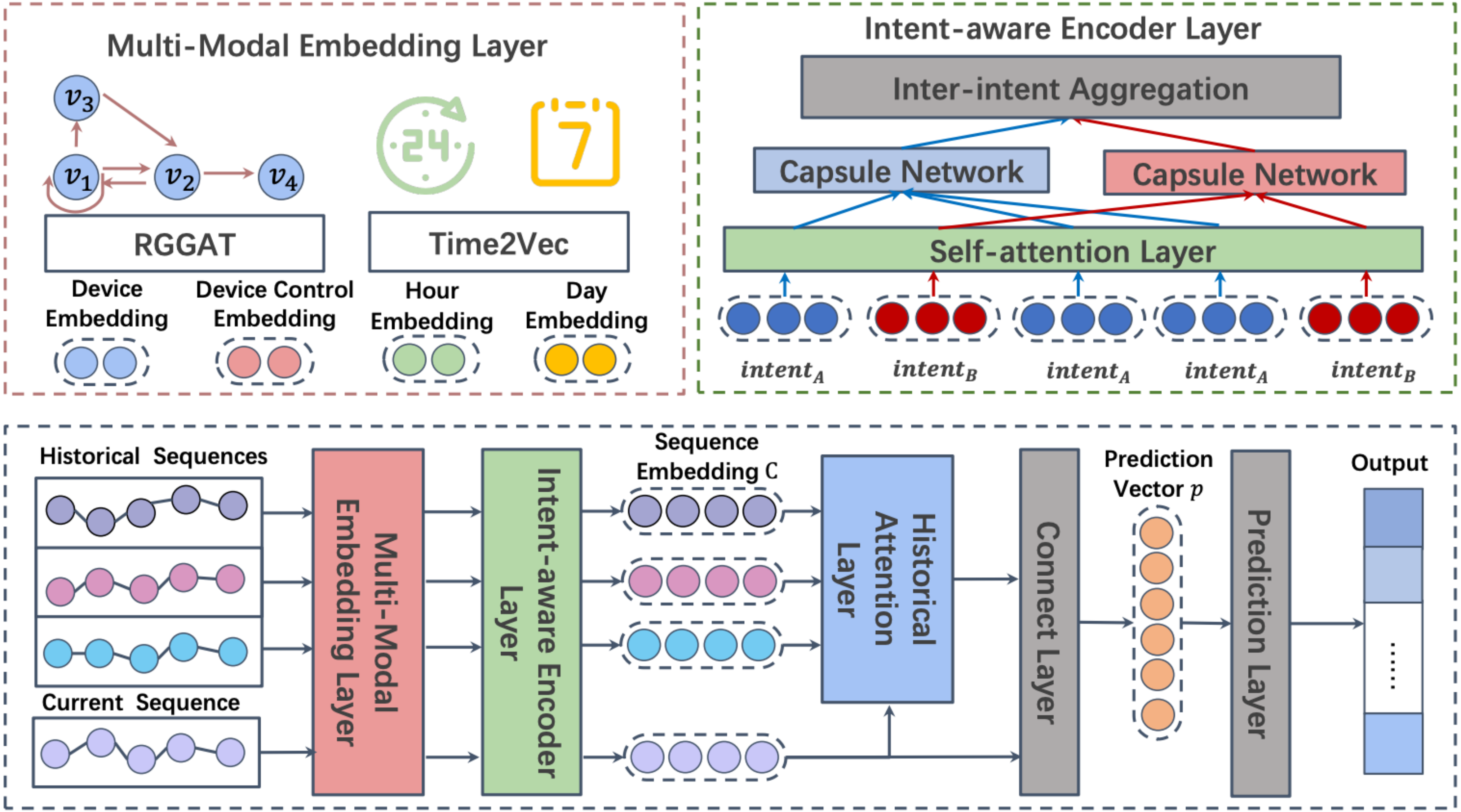
- **Multi-level Periodicity:**
  - **Day-level Periodicity:** user leaves work on time on Wednesday and Thursday, but work overtime once **every 3 days**, on Tuesday and Friday.
  - **Week-level Periodicity:** user stays up late for games **every Saturday** night.



4 Overview

- **We propose DeepUDI:**
  - A novel approach for accurate UDI prediction.
- **Idea #1: Relational Gated Graph Attention Network**
  - To construct graph from behavior sequence and leverage relational gated GNN to capture the transitions between different device.
- **Idea #2: Intent-aware Encoder**
  - To view intents as capsules and capture multiple intents by capsule network.
- **Idea #3: Historical Attention Mechanism**
  - To model correlation between current sequence and historical sequences by attention mechanism.

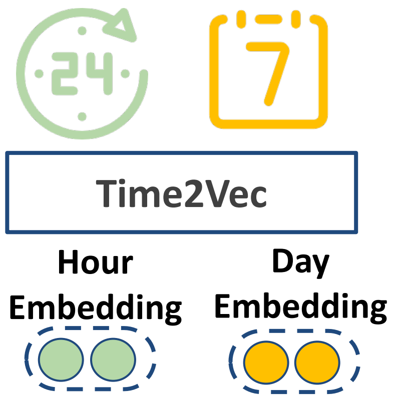
4 Overview



Multi-Modal Embedding Layer
v_3
v_1
v_2
v_4
RGGAT
Time2Vec
Device Embedding
Device Control Embedding
Hour Embedding
Day Embedding
Intent-aware Encoder Layer
Inter-intent Aggregation
Capsule Network
Capsule Network
Self-attention Layer
$intent_A$
$intent_B$
$intent_A$
$intent_A$
$intent_B$
Historical Sequences
Current Sequence
Multi-Modal Embedding Layer
Intent-aware Encoder Layer
Sequence Embedding C
Historical Attention Layer
Connect Layer
Prediction Vector p
Prediction Layer
Output

5 Multi-Modal Embedding Layer

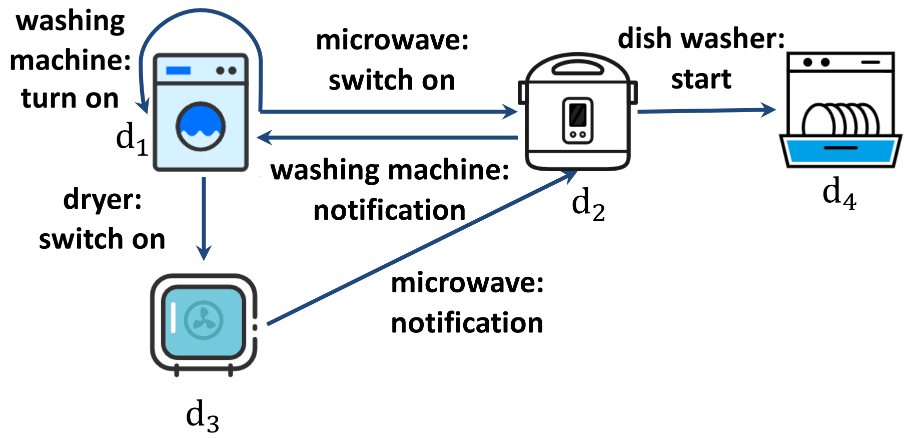
- **Time Embedding**
  - Time2vec can capture both periodic and non-periodic patterns of time.
  - We denotes time as day of week and hour of day.
  - For time τ, $t2v(\tau)[i]$ denotes the i-th element of time embedding, ω_i and φ_i are learnable parameters, $\mathcal{F}$ is a periodic activation function e.g., a sine function.



$$\mathbf{t2v}(\tau)[i] = \begin{cases} \omega_i \tau + \varphi_i, & \text{if } i = 0 \\ \mathcal{F}(\omega_i \tau + \varphi_i), & \text{if } 1 \le i \le L - 1 \end{cases}$$

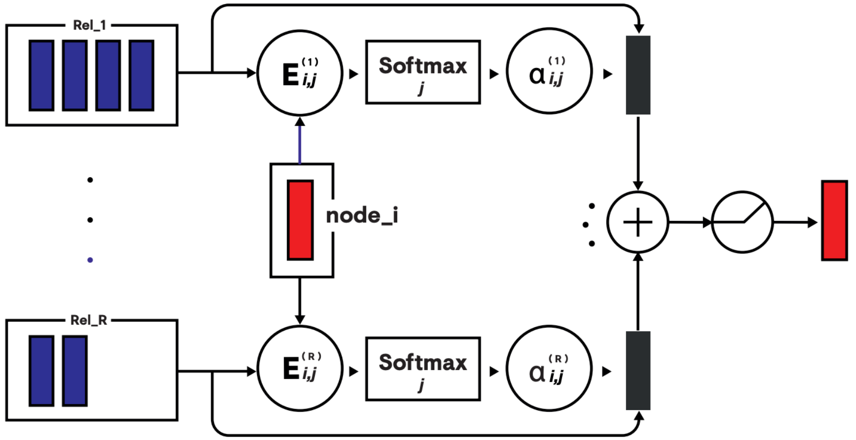
5 Multi-Modal Embedding Layer

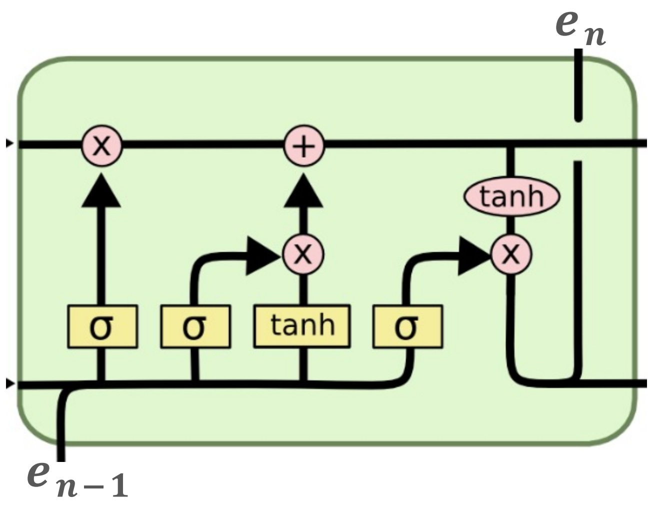
- **Device and Device Control Embedding**
  - **Relational Sequence Graph(RS-Graph):**
    - each node in the graph represents a device $\boldsymbol{d}$, each edge $(\boldsymbol{d}_{n-1}, \boldsymbol{d}_n)$ indicates that the user accesses device $\boldsymbol{d}_n$ after accessing device $\boldsymbol{d}_{n-1}$
    - each device control represents a relation r



5 Multi-Modal Embedding Layer

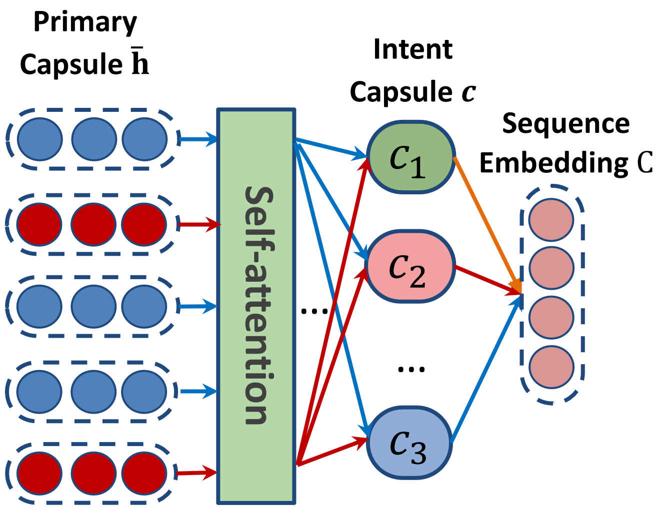
- **Device and Device Control Embedding**
  - **Relational Gated Graph Attention Network:**
    - message passing based on relation.
    - feature update by GRU.





6 Intent-aware Encoder

- **Self-attention and capsule network are employed as Intent-aware Encoder**
  - Self-attention Layer: mine global semantic information of behavior.
  - Capsule Network: we treat behavior as primary capsules and use dynamic routing to learn the probability of high-level intent capsule.
  - Finally, we aggregate intent vector c as sequence embedding C.
  - W_C is the learnable weight of different intents.



Algorithm 1: DeepUDI Dynamic Routing.

Input: primary capsules $\bar{\mathbf{h}}_i$, iteration times T, number of intent capsules K
Output: intent capules $\{\mathbf{c}_j, j = 1, \ldots, K\}$

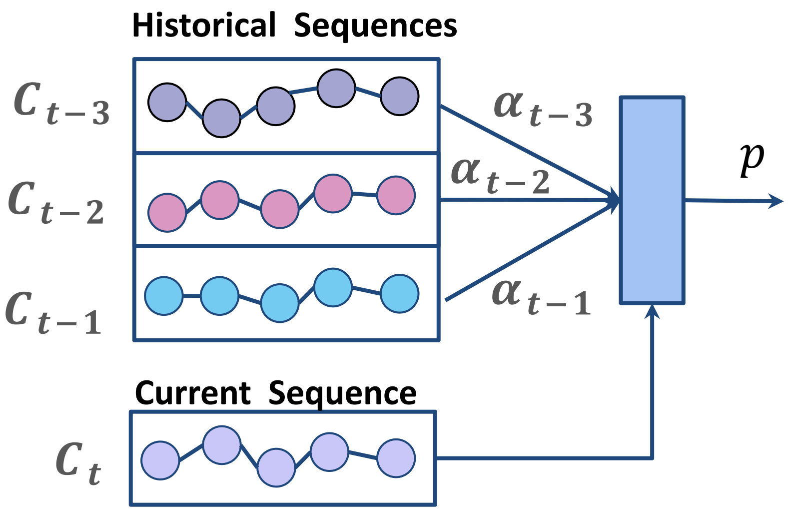
```
for each primary capsule i and corresponding intent capsule j:
initialize b_ij = 0
for iter = 1, ...T do
    for each primary capsule i: w_i = softmax(b_i).
    for each intent capsule j: s_j = Σ_i w_ij W_ij h_i.
    for each intent capsule j: c_j = squash(s_j).
    for each primary capsule i and intent capsule j:
    b_ij = b_ij + c_j^T W_ij h_i.
end for
return {c_j, j = 1, ..., K}
```

$$\mathbf{C} = \mathbf{W}_C[\mathbf{c}_1, \ldots, \mathbf{c}_K]$$

7 Historical Attention Mechanism

AAMAS 2023 LONDON

- **Summarize the history sequence vector $[C_1, C_2, C_3, \ldots C_{t-1}]$ into p to capture the multi-level periodicity**
  - α_i and β_i are normalized and unnormalized scores of C_i for C_t, respectively, W_H is the learnable weight.



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

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

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

$$\hat{\mathrm{y}} = \text{softmax}(\mathbf{W}_p p)$$

8 Datasets

- **Datasets: we use four real-world datasets to evaluate DeepUDI**
  - Three datasets **(US/SP/FR)** from public dataset
  - One 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 DeepUDI with 7 competitors
  - Traditional Models: **HMM** and **FPMC**
  - RNN-based Models: **LSTM**, **CA-RNN** and **DeepMove**
  - GNN-based Models: **SR-GNN**
  - Transformer-based Models: **SmartSense**

- **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

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

10 Questions

- We answer the following research questions:
  - RQ1 (Performance). Compared with other methods, does DeepUDI have higher prediction performance of user device interaction?
  - RQ2 (Ablation study). How do main components of DeepUDI affect the performance of UDI prediction?
  - RQ3 (Parameter study). How key parameters affect the performance of DeepUDI?
  - RQ4 (Interpretability study). Can DeepUDI give a reasonable explanation for the prediction results?

Experiment Results

- **RQ1:** Compared with other methods, does DeepUDI have higher prediction performance of user device interaction?
- **A1:** DeepUDI outperforms all competitors in most cases.

Dataset	Metric	HMM	FPMC	LSTM	CA-RNN	DeepMove	SRGNN	SmartSense	DeepUDI(Ours)
AN	Acc@1	0.6099	0.6557	0.7062	0.7026	0.7116	0.9245	0.9407	**0.9784**
	Acc@3	0.7501	0.7959	0.7843	0.8302	0.9272	0.9864	0.9731	**0.9865**
	Acc@5	0.7714	0.7902	0.8328	0.9003	0.9542	0.9872	0.9838	**0.9892**
	Macro-F1	0.2439	0.2845	0.3759	0.4159	0.5027	0.7368	0.7519	**0.7997**
FR	Acc@1	0.6536	0.6814	0.6962	0.7893	0.7762	0.7819	0.7923	**0.8144**
	Acc@3	0.7813	0.8271	0.8011	0.9148	0.9221	0.9197	**0.9371**	0.9238
	Acc@5	0.8242	0.8508	0.8565	0.9425	0.9446	0.9435	**0.9628**	0.9512
	Macro-F1	0.1127	0.1279	0.1302	0.2102	0.2288	0.2482	0.2603	**0.3425**
SP	Acc@1	0.6315	0.6964	0.7517	0.7853	0.7756	0.7815	0.7921	**0.7923**
	Acc@3	0.7863	0.7916	0.8864	0.8915	0.9125	0.9303	0.9342	**0.9375**
	Acc@5	0.8361	0.8605	0.9346	0.9117	0.9521	0.9603	0.9511	**0.9642**
	Macro-F1	0.1382	0.1586	0.1756	0.1745	0.2159	0.2239	0.2244	**0.3112**
US	Acc@1	0.3327	0.3543	0.4286	0.5212	0.5527	0.5784	0.5935	**0.6056**
	Acc@3	0.6881	0.6992	0.8209	0.8577	0.8844	0.8955	0.9056	**0.9123**
	Acc@5	0.7258	0.7712	0.8929	0.9135	0.9418	0.9463	0.9489	**0.9521**
	Macro-F1	0.1069	0.1123	0.1265	0.1396	0.2388	0.2431	0.2451	**0.3538**

Experiment Results

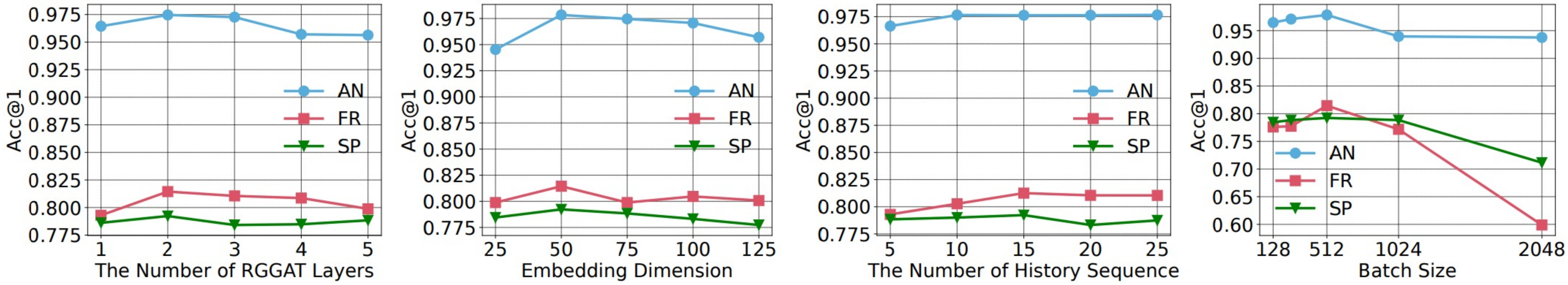
- **RQ2:** How do main components of DeepUDI affect the performance of UDI prediction?
- **A2:** All three components (**MME: Multi-Modal Embedding Layer, IAE: Intent-aware encoder** and **HAM: Historical Attention Mechanism**) of DeepUDI are contributive for UDI prediction.

Model	AN		FR	
	Acc@1	Macro-F1	Acc@1	Macro-F1
DeepUDI-MME	0.9487	0.7364	0.7851	0.2805
DeepUDI-IAE	0.9595	0.7585	0.7853	0.3103
DeepUDI-HAM	0.9676	0.7812	0.7988	0.3234
DeepUDI-ALL	0.9137	0.6934	0.7578	0.2511
DeepUDI	**0.9784**	**0.7997**	**0.8144**	**0.3425**

AAMAS 2023
LONDON

11 Experiment Results

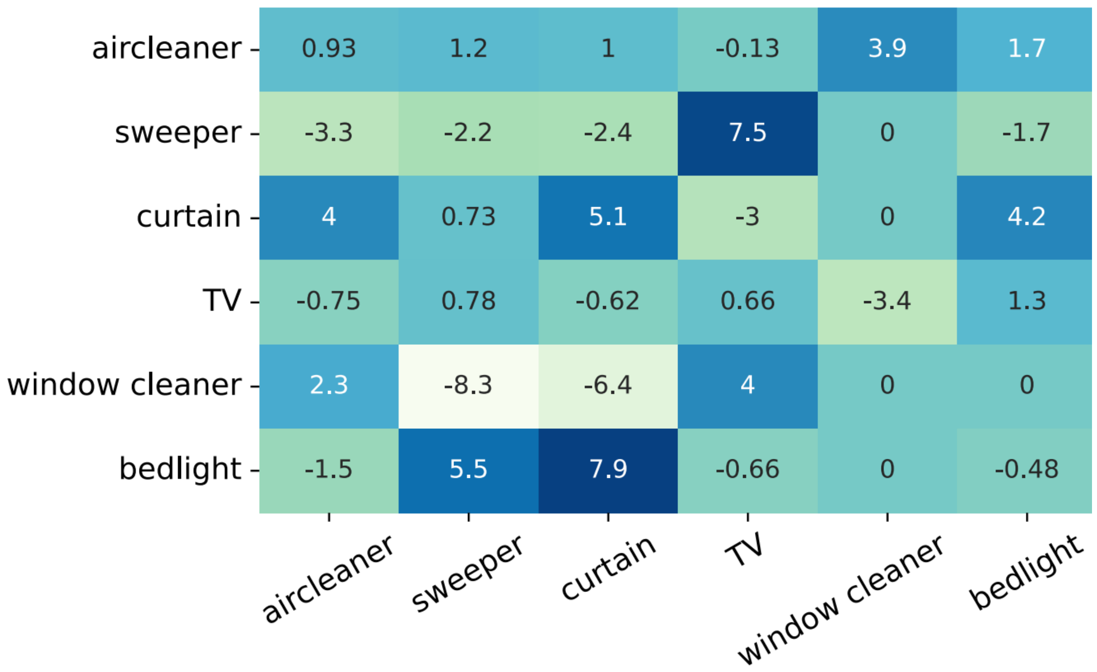
- **RQ3:** How key parameters affect the performance of DeepUDI?
- **A3:** The best parameter combination: #of layers of RGGAT=2, Embedding Dimension=50, # of History Sequence=15, Batch Size=512.



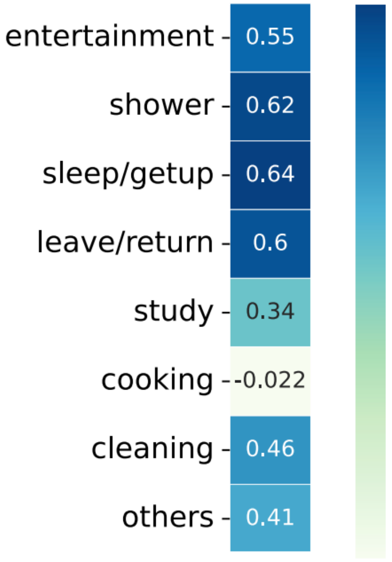
11 Experiment Results

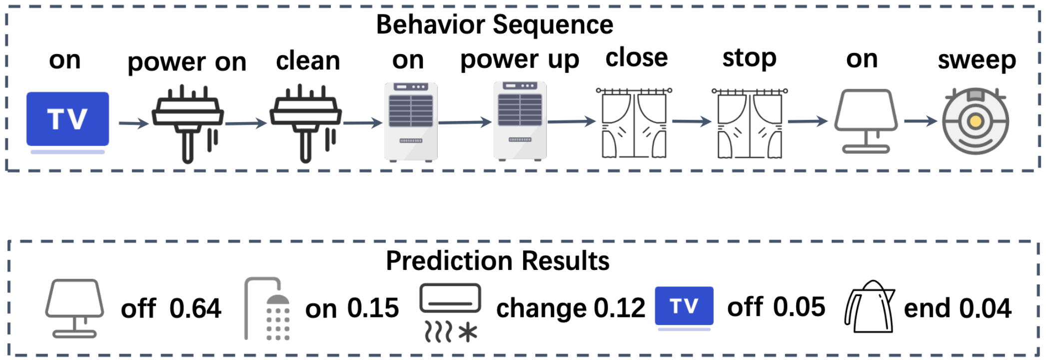
- **RQ4:** Can DeepUDI give a reasonable explanation for the prediction results?
- **A4:** DeepUDI successfully learns the correlation between devices by RGGAT and intents of user by capsule network.

RGGAT Attention



Capsule Weight





12

Conclusions

- We propose DeepUDI for accurate UDI prediction.
- Our main contributions are summarized as follows:
  - Idea #1: Relational Gated Graph Attention Network
  - Idea #2: Intent-aware Encoder
  - Idea #3: Historical Attention Mechanism
- DeepUDI 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