ICDE
40th IEEE International Conference on
Data Engineering
Utrecht, Netherlands | 13th - 17th May
2024

Representation Learning of Tangled Key-Value Sequence Data for Early Classification

Tao Duan, Junzhou Zhao*, Shuo Zhang, Jing Tao, Pinghui Wang

Xi'an Jiaotong University

西安交通大學
XI'AN JIAOTONG UNIVERSITY

Outline

☞ ☐ **Background and Problem Formulation**

☐ Motivation

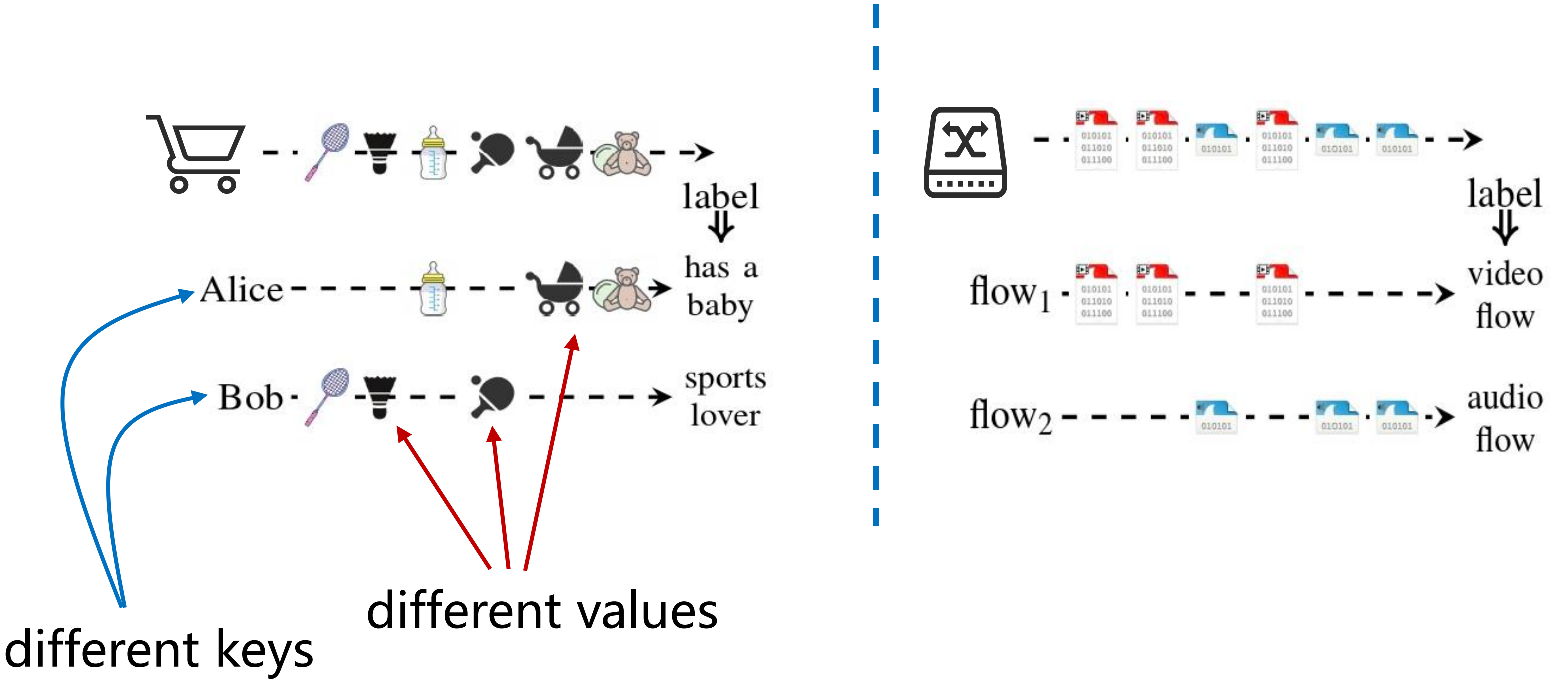
☐ Framework Design

☐ Evaluation

☐ Conclusion

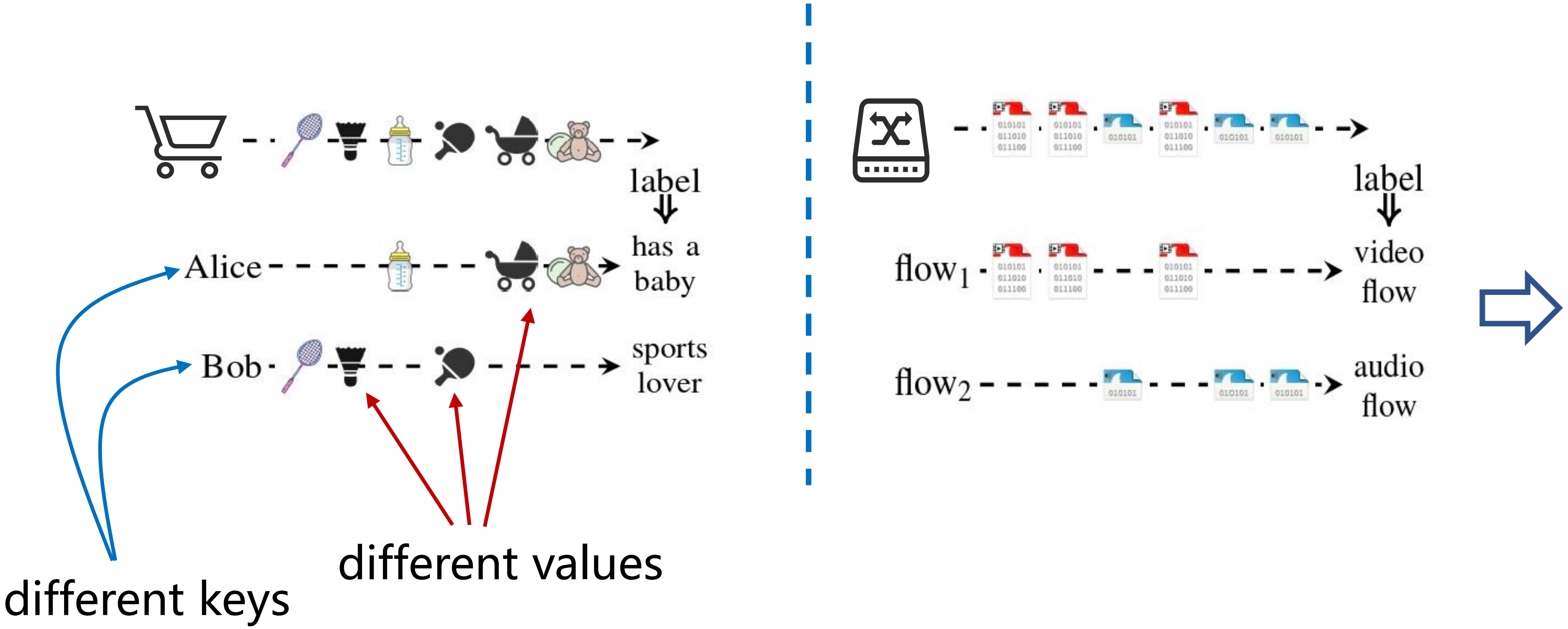
Background

□ Tangled Key-Value Sequence



Background

◻ Tangled Key-Value Sequence



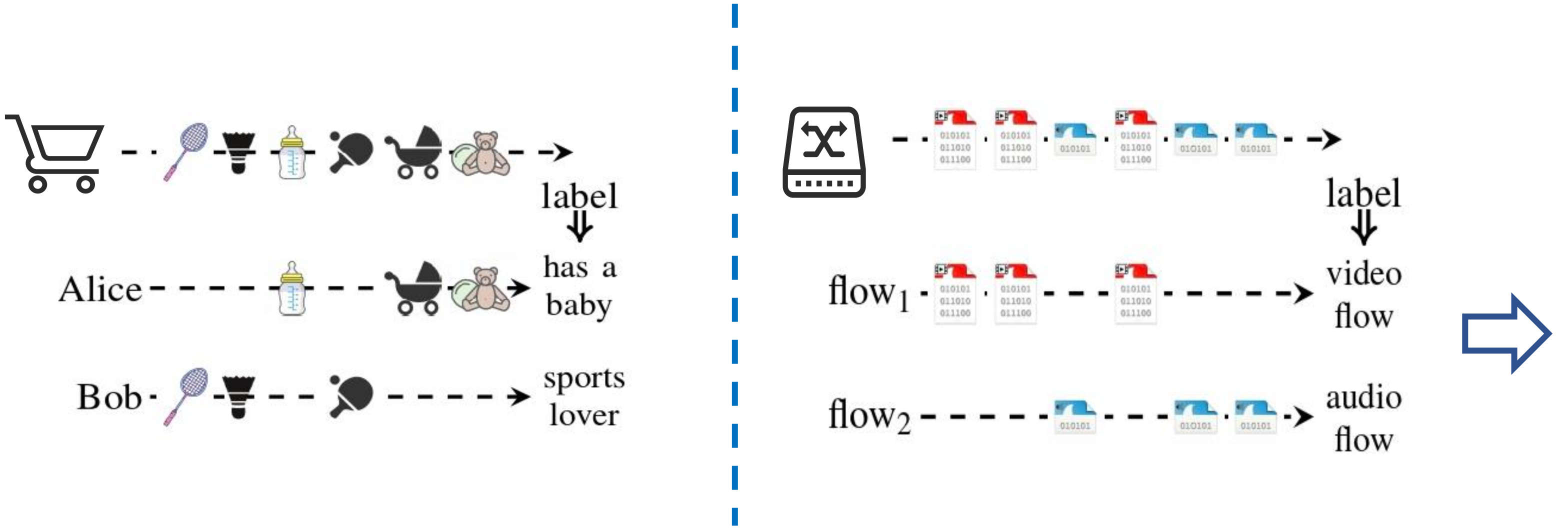
Application:

- ✓ Product Recommendation
- ✓ Networking QoS Improvement
- ✓ Malicious Intrusion Detection

…

Background

- Tangled Key-Value Sequence

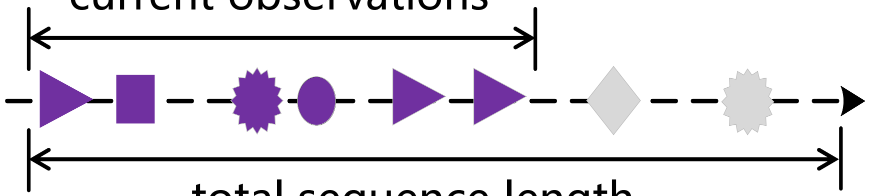


Application:

- ✓ Product Recommendation
- ✓ Networking QoS Improvement
- ✓ Malicious Intrusion Detection

…

- Two core performances for sequence data classification:
  - Earliness
  - Accuracy

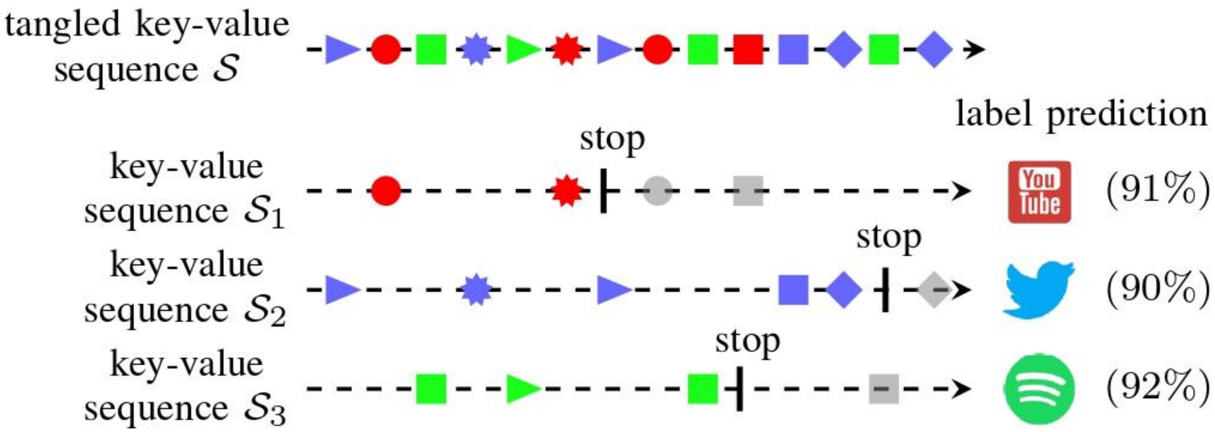


Earliness = observations/sequence length

Problem Formulation

◻ Tangled Key-Value Sequence Early Classification Problem:

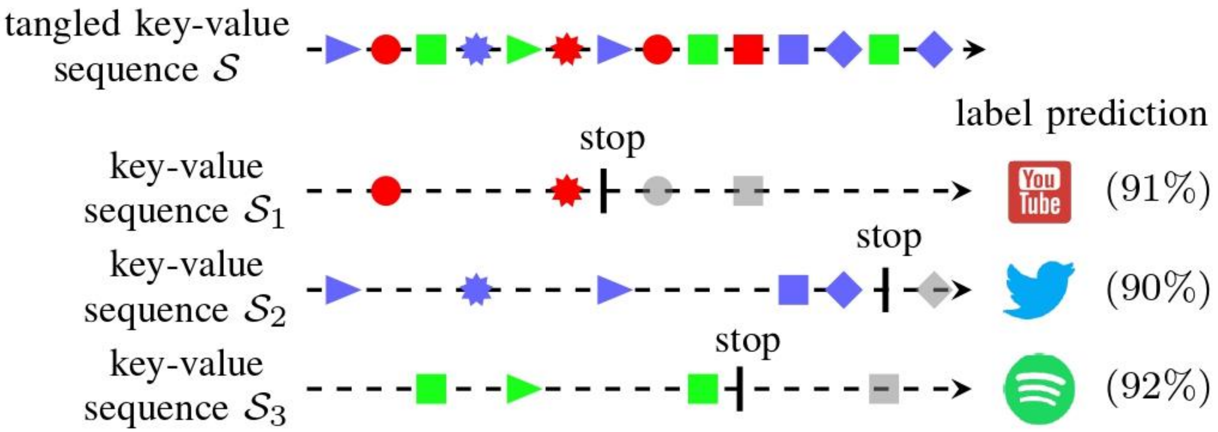
- Given a **tangled key-value sequence**: $\mathcal{S} \triangleq (\langle k, \mathbf{v}\rangle : k \in \mathcal{K}, \mathbf{v} \in \mathcal{V}_1 \times \cdots \times \mathcal{V}_l)$
- Each **key-value sequence sharing a same key**: $\mathcal{S}_k \triangleq (\langle k, \mathbf{v}\rangle : \langle k, \mathbf{v}\rangle \in \mathcal{S})$



Our Target: classify each $\mathcal{S}_k$ within $\mathcal{S}$ both **early** and **accurately.**

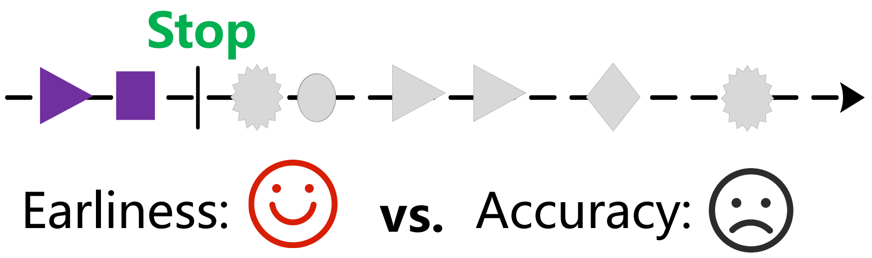
Problem Formulation

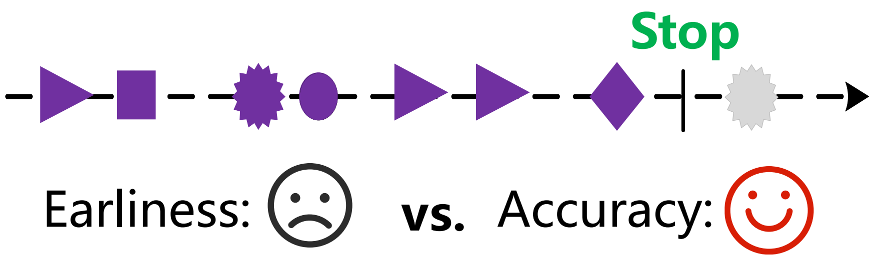
- Tangled Key-Value Sequence Early Classification Problem:
  - Given a **tangled key-value sequence**: $\mathcal{S} \triangleq (\langle k, \mathbf{v}\rangle : k \in \mathcal{K}, \mathbf{v} \in \mathcal{V}_1 \times \cdots \times \mathcal{V}_l)$
  - Each **key-value sequence sharing a same key**: $\mathcal{S}_k \triangleq (\langle k, \mathbf{v}\rangle : \langle k, \mathbf{v}\rangle \in \mathcal{S})$



Our Target: classify each $\mathcal{S}_k$ within $\mathcal{S}$ both **early** and **accurately.**

- Challenges: Earliness and Accuracy are two conflicting goals.



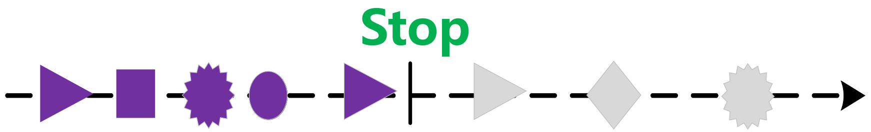


Outline

- ☐ Background and Problem Formulation
- ☞ ☐ **Motivation**
- ☐ Framework Design
- ☐ Evaluation
- ☐ Conclusion

Motivation

- The multi-objective optimization of "earliness-accuracy"



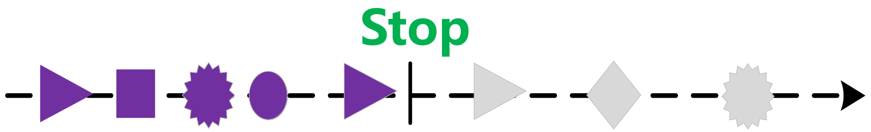
When?

Earliness:  & Accuracy: 

- Decompose it to two targets:
  - how to learn an informative representation from partial observations?
  - how to adaptively determine the number of observations for each S_k ?

Motivation

- ☐ The multi-objective optimization of "earliness-accuracy"



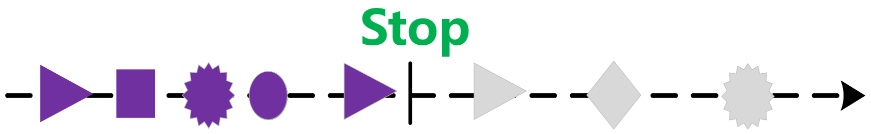
Earliness: & Accuracy:

When?

- ☐ Decompose it to two targets:
  - how to learn an informative representation from partial observations ?
    - ✓ <u>exploit rich inner- and inter-sequence correlations in $\mathcal{S}$.</u>

Motivation

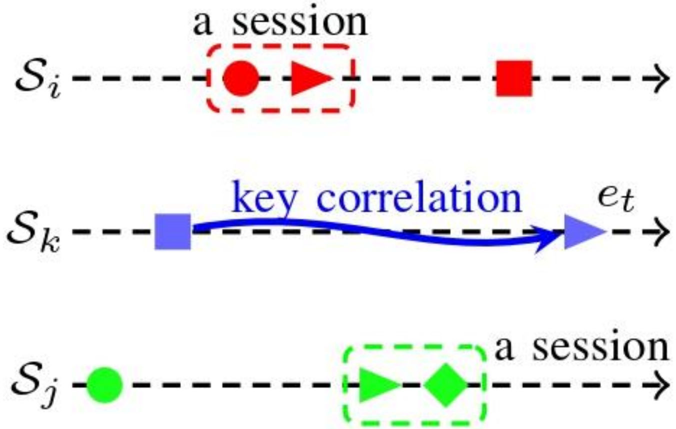
- The multi-objective optimization of "earliness-accuracy"



Earliness: ☺ & Accuracy: ☺

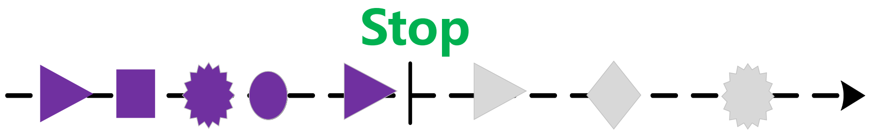
When?

- Decompose it to two targets:
  - how to learn an informative representation from partial observations ?
    - ✓ <u>exploit rich inner- and inter-sequence correlations in $\mathcal{S}$.</u>



Motivation

- The multi-objective optimization of "earliness-accuracy"



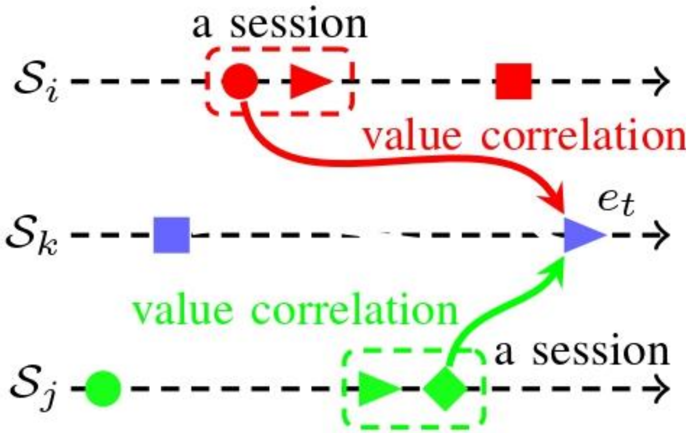
When?

Earliness: & Accuracy:

- Decompose it to two targets:

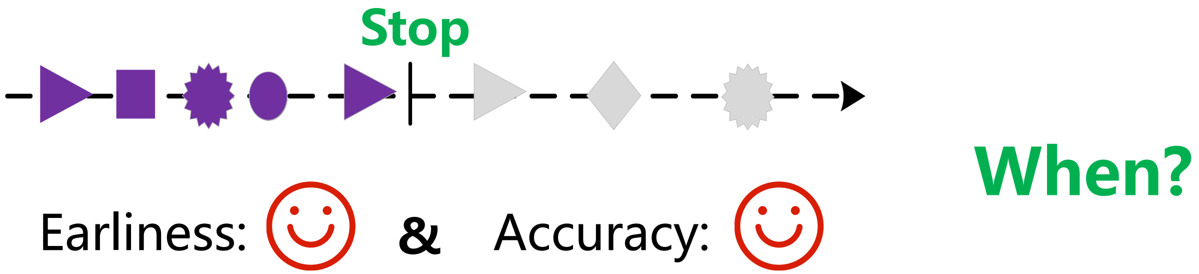
  - how to learn an informative representation from partial observations ?

    ✓ <u>exploit rich inner- and inter-sequence correlations in $\mathcal{S}$.</u>

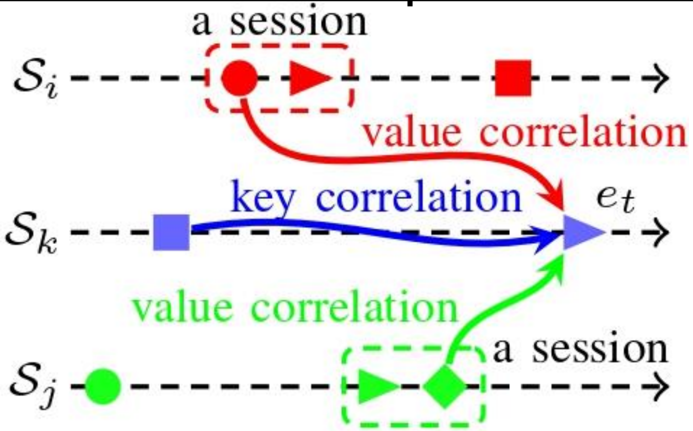


Motivation

- The multi-objective optimization of "earliness-accuracy"

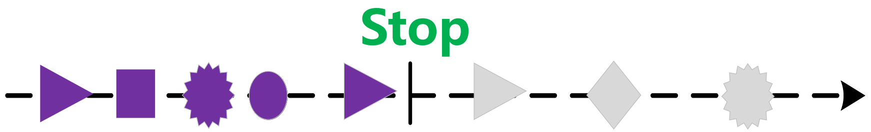


- Decompose it to two targets:
  - how to learn an informative representation from partial observations ?
    - ✓ <u>exploit rich inner- and inter-sequence correlations in $\mathcal{S}$.</u>



Motivation

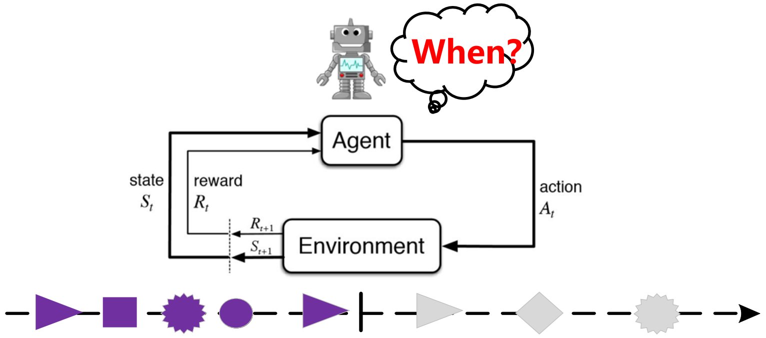
- The multi-objective optimization of "earliness-accuracy"



Earliness: & Accuracy:

When?

- Decompose it to two targets:
  - how to learn an informative representation from partial observations?
  - how to adaptively determine the number of observations for each S_k ?
    - ✓ <u>formulate it as the Partially Observable Markov Decision Process (POMDP), and solve it through a halting policy.</u>

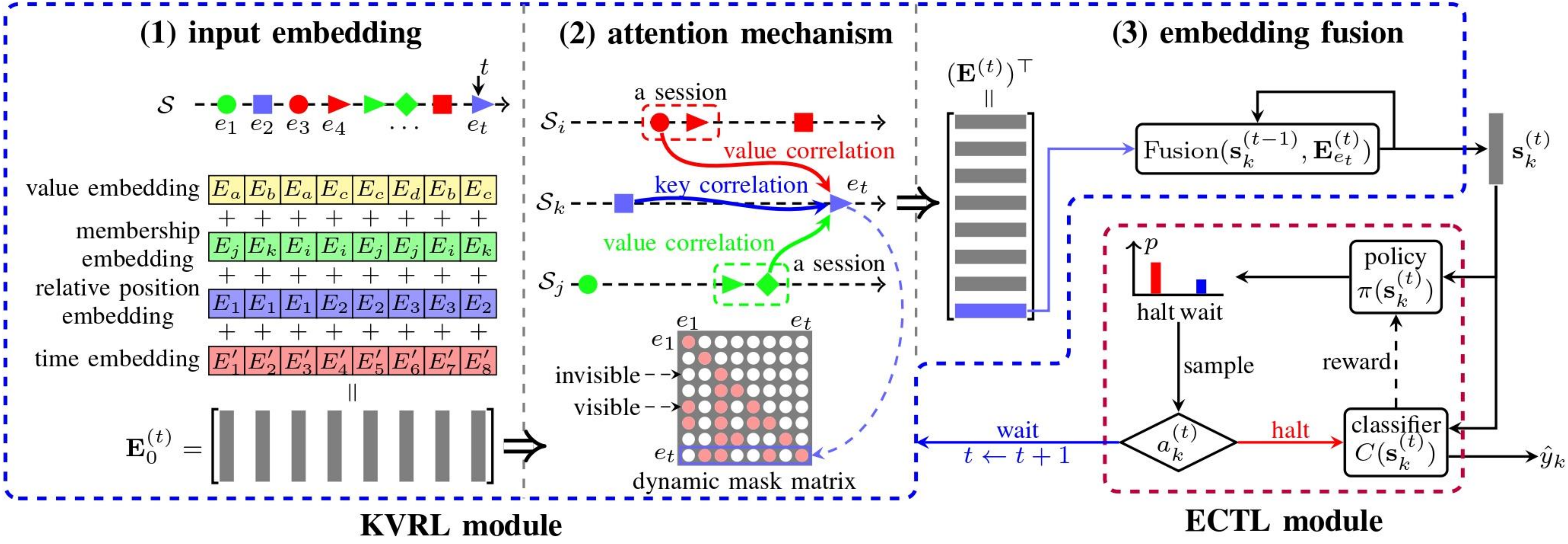


Outline

- ☐ Background and Problem Formulation
- ☐ Motivation
- ☞ ☐ **Framework Design**
- ☐ Evaluation
- ☐ Conclusion

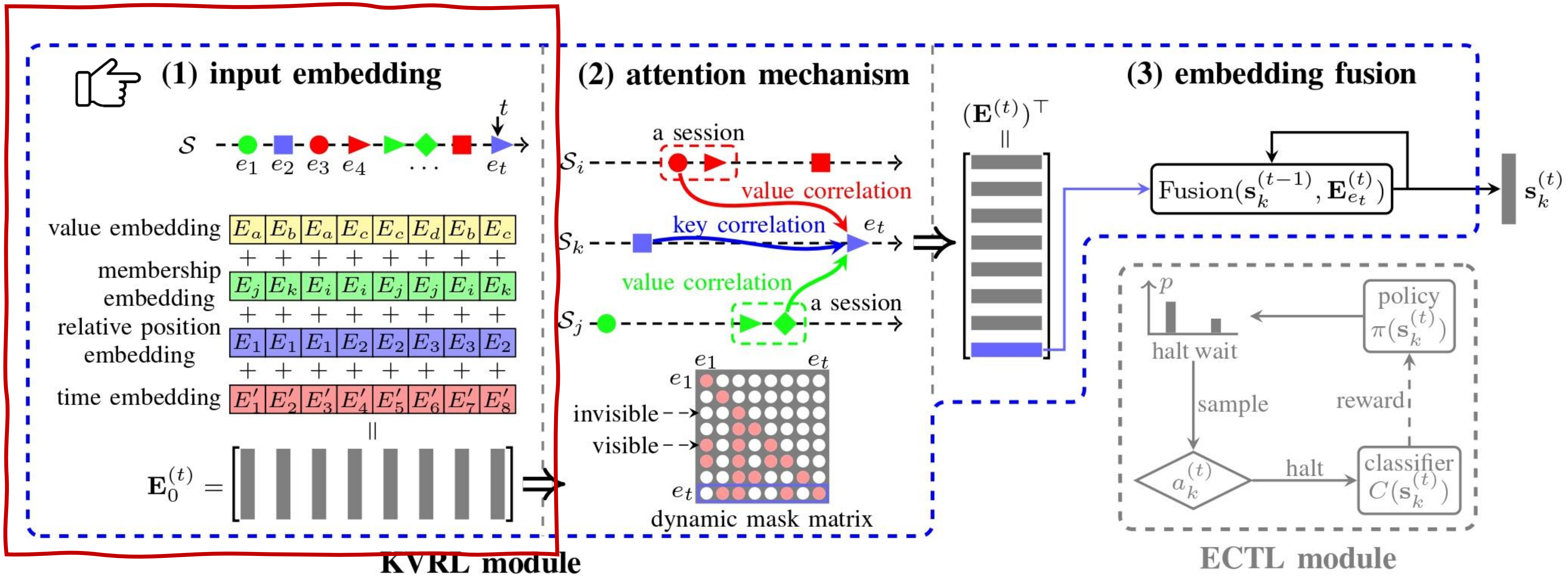
Overview

Key-Value sequence Early Co-classification (KVEC) framework



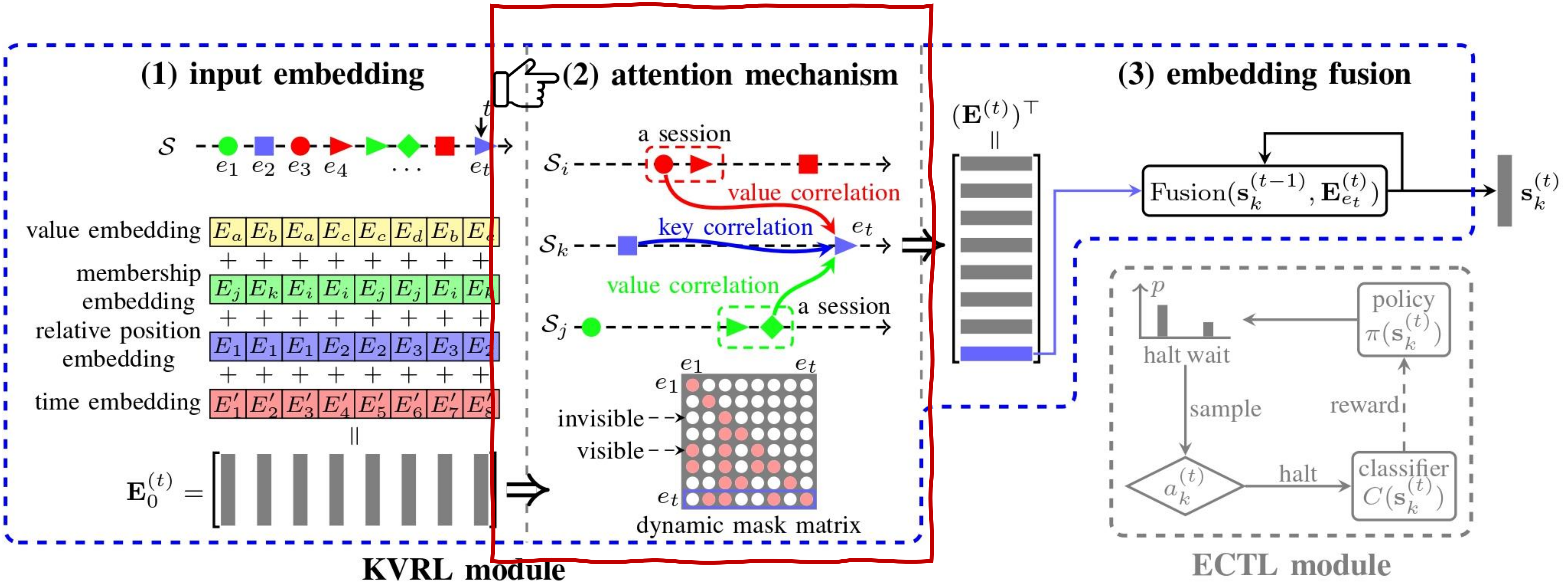
Model Detail

◻ Key-Value Sequence Representation Learning (KVRL) module



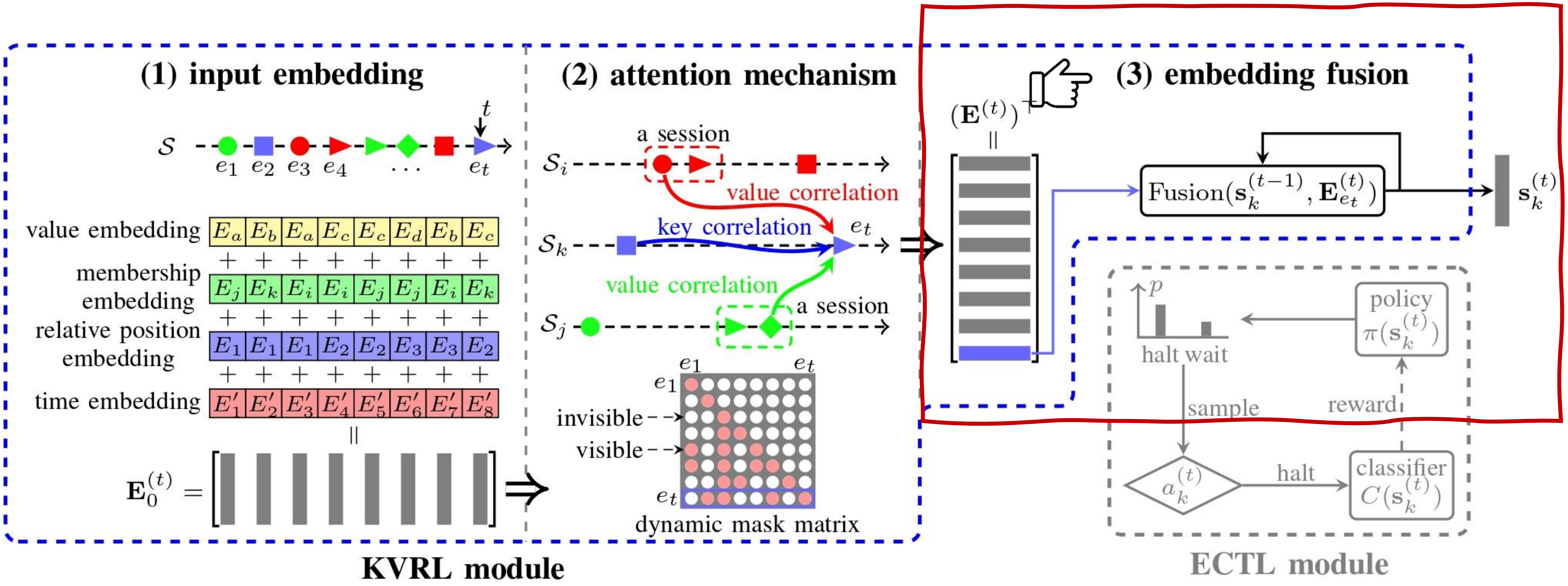
Model Detail

- Key-Value Sequence Representation Learning (KVRL) module



Model Detail

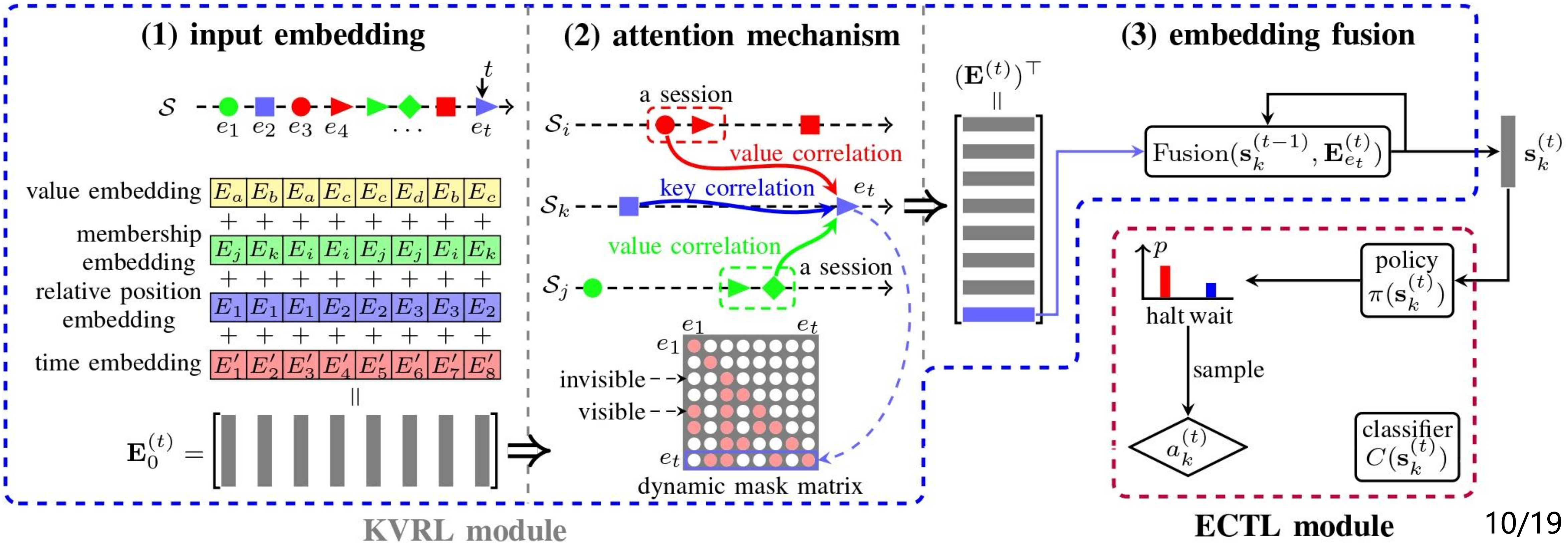
- Key-Value Sequence Representation Learning (KVRL) module



Model Detail

- Early Co-classification Timing Learning (ECTL) module

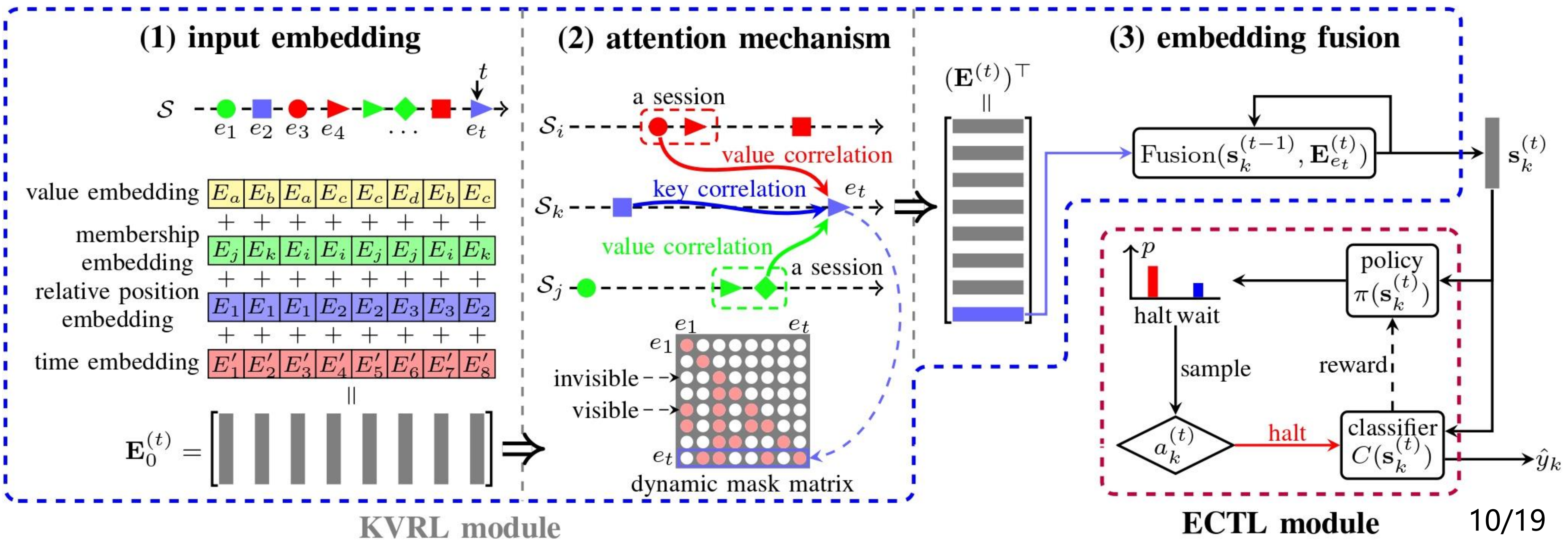
sample action from halting policy
- halt:
- wait:



Model Detail

- Early Co-classification Timing Learning (ECTL) module

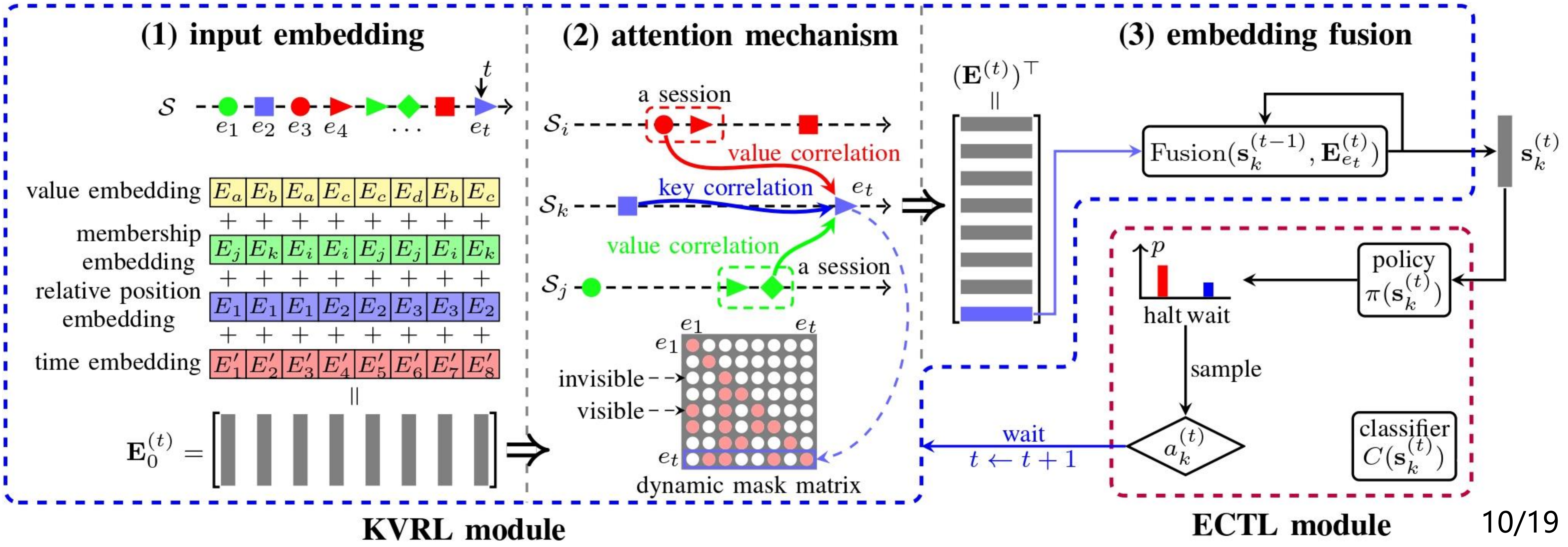
sample action from halting policy
- halt: enough items have been observed
- wait: more items in need to be observed



Model Detail

- Early Co-classification Timing Learning (ECTL) module

sample action from halting policy $\begin{cases} \text{halt: enough items have been observed} \\ \text{wait: more items in need to be observed} \end{cases}$



Model Detail

Multi-objective Optimization

- minimize the prediction error of the classification network

$$l_1(\boldsymbol{\theta}_1) \triangleq -\sum_{k=1}^{K}\sum_{c=1}^{C} \mathbf{1}(y_k = c)\log \mathbf{p}_{k,c}(\boldsymbol{\theta}_1)$$

- maximize the accumulate reward gained by the policy network

$$l_2(\boldsymbol{\theta}_2) \triangleq -\sum_{k=1}^{K}\sum_{i=1}^{n_k} \left(R_k^{(i)} - b_k^{(i)}(\boldsymbol{\theta}_b)\right)\log P(a_k^{(i)}|\mathbf{s}_k^{(i)};\boldsymbol{\theta}_\pi)$$

- encourage early prediction

$$l_3(\boldsymbol{\theta}_3) \triangleq -\sum_{k=1}^{K}\sum_{i=1}^{n_k} \log P(a_k^{(i)} = \mathrm{Halt}|\mathbf{s}_k^{(i)};\boldsymbol{\theta}_\pi)$$

Total Training Loss: $l(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2,\boldsymbol{\theta}_3) \triangleq l_1(\boldsymbol{\theta}_1) + \alpha l_2(\boldsymbol{\theta}_2) + \beta l_3(\boldsymbol{\theta}_3)$

Outline

- ☐ Background and Problem Formulation
- ☐ Motivation
- ☐ Framework Design
- ☞ ☐ **Evaluation**
- ☐ Conclusion

Evaluation

- Datasets:
  - USTC-TFC2016: malware and intrusion network traffic dataset
  - MovieLens-1M: movie rating dataset
  - Traffic-FG / Traffic-App: service-level / App-level traffic dataset
  - Synthetic-Traffic: synthetic network traffic dataset

dataset	#keys	avg $\|\mathcal{S}_k\|$	avg session length	#classes
USTC-TFC2016	$3,200$	31.2	8.3	9
MovieLens-1M	$6,040$	163.5	1.7	2
Traffic-FG	$60,000$	50.7	2.4	12
Traffic-App	$50,000$	57.5	2.7	10
Synthetic-Traffic	$10,000$	100.0	2.1	2

Evaluation

- Metrics:
  - Earliness
  - Accuracy, Precision, Recall, F1-score
  - HM: harmonic mean of Accuracy and Earliness, measure the multi-objective balancing ability of different methods.

$$\mathrm{HM} \triangleq \frac{2 \times (1 - \mathrm{Earliness}) \times \mathrm{Accuracy}}{1 - \mathrm{Earliness} + \mathrm{Accuracy}}$$

Evaluation

- Results: KVEC achieves 4.7−17.5% accuracy improvement, 3.7−14.0% HM improvement.

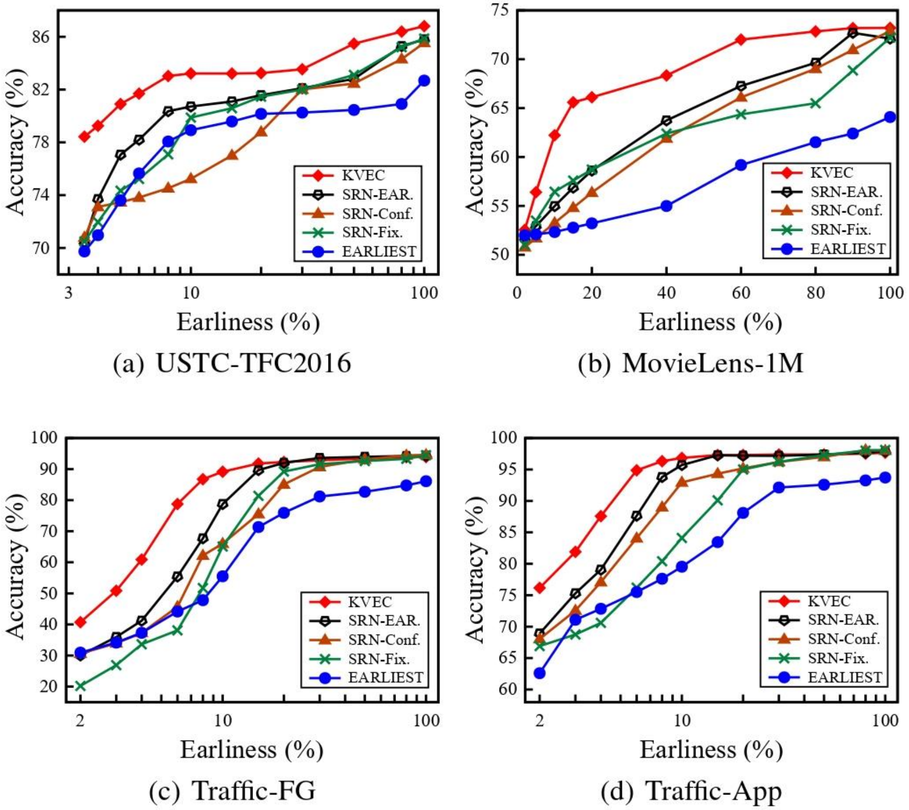


Figure 3. Accuracy comparison

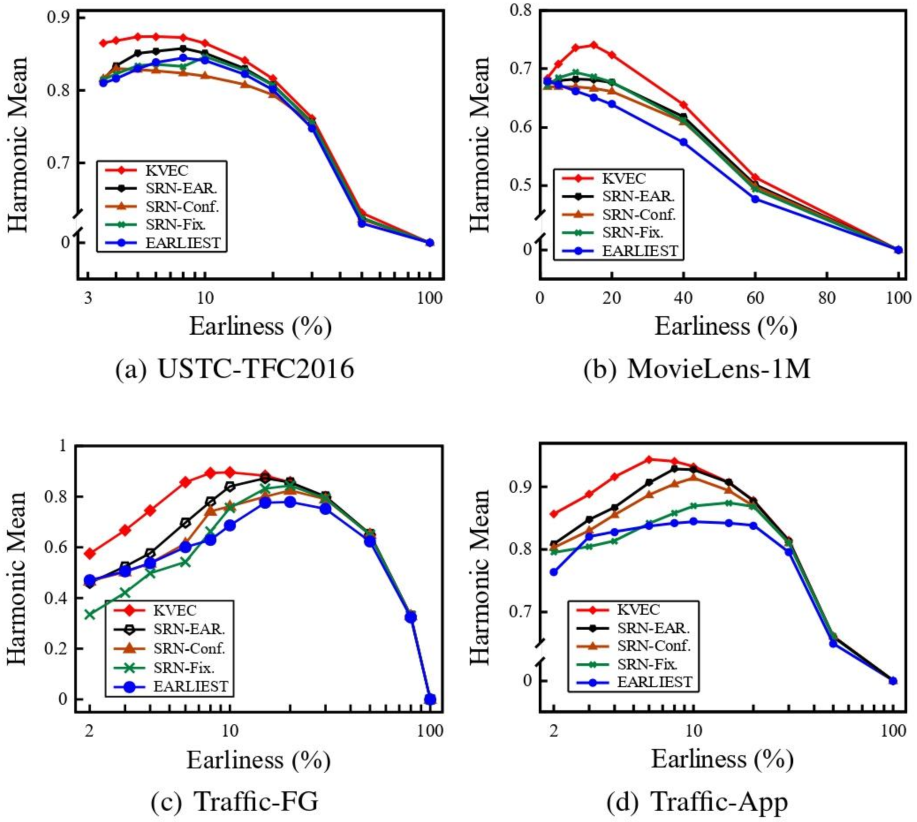
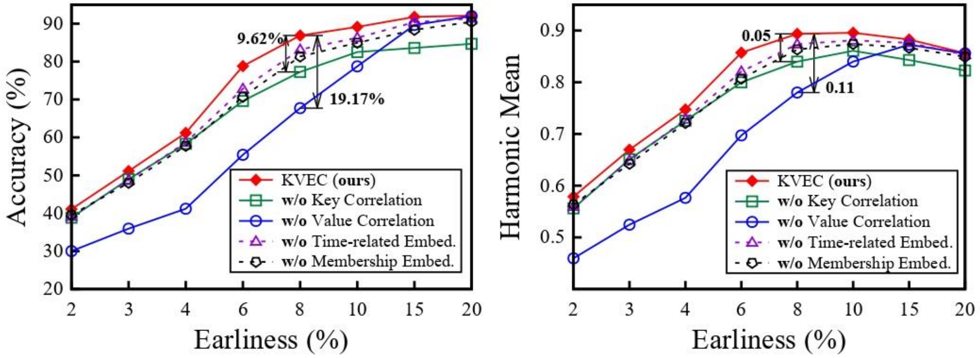


Figure 7. HM comparison

- KVEC has a better early classification performance and a better multi-objective balancing ability.

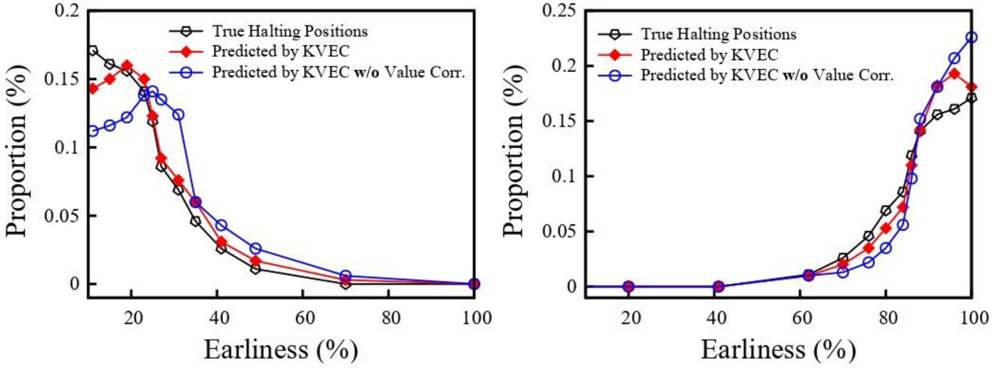
Evaluation

◻ More Discussions



(a) Effect on Accuracy (b) Effect on Harmonic Mean

Figure 9. Ablation study



(a) Early-stop subdataset (b) Late-stop subdataset

Figure 11. Distribution of halting positions predicted by different methods.

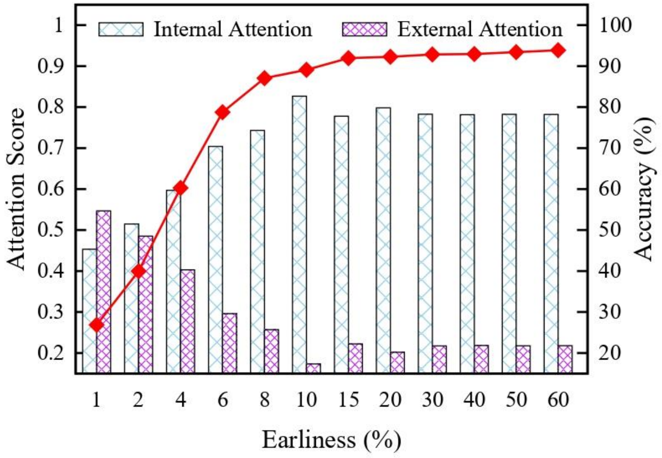


Figure 10. Distribution of attention score in various halting positions.

◆ <u>The advantages of KVEC primarily comes from the proposed representation learning.</u>

Outline

- Background and Problem Formulation
- Motivation
- Framework Design
- Evaluation
- **Conclusion**

Conclusion

- Formulate the tangled key-value sequence early classification problem.

- We propose the KVEC framework to solve this problem which mainly consists of two core modules, i.e., KVRL and ECTL.

- Extensive experiments conducted on both real-world and synthetic datasets demonstrate that KVEC outperforms all alternative methods.

Thanks for listening!

Paper Link

duantao@stu.xjtu.edu.cn