ISCAS 中国科学院软件研究所学术年会 暨重点实验室科技活动周 2025 第十届

学术论文

基于时态逻辑规则学习的可解释时间序列分类

Learning Reliable and Intuitive Temporal Logic Rules for Interpretable Time Series Classification

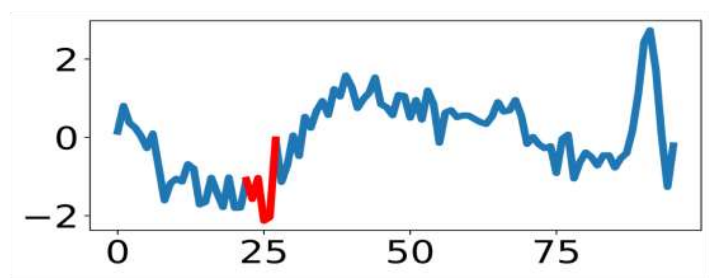
汪洋，朱嘉奇*，李妙妙，刘江，李依霖，杨翊，李佳璠，王宏安

To appear in *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD 2025)*, Research Track

联系人：朱嘉奇，zhujq@ios.ac.cn，13683257241

Introduction

- **Task:** Interpretable Time Series Classification in safety-critical scenarios
  - Such as medicine, finance and public security
- **Motivation:** requires not only accurate classification results, but also explicit and rigorous rationales -- Rule-based models
  - Reliable: consistent with the actual execution (decision process) of the neural network model
  - Intuitive: consistent with human understanding and the consecutive nature of temporal properties



Rationale:
T-wave declines until it inverts, indicating myocardial infarction

Represented as Logical Rule:

$$\phi := \psi_1 \mathbb{U}_{[22,27]} \psi_2$$
$$\psi_1 := (x^D < 0.16)$$
$$\psi_2 := (0.41 \leq x^D < 0.91)$$

Signal Temporal Logic (STL)

- **Preliminary**

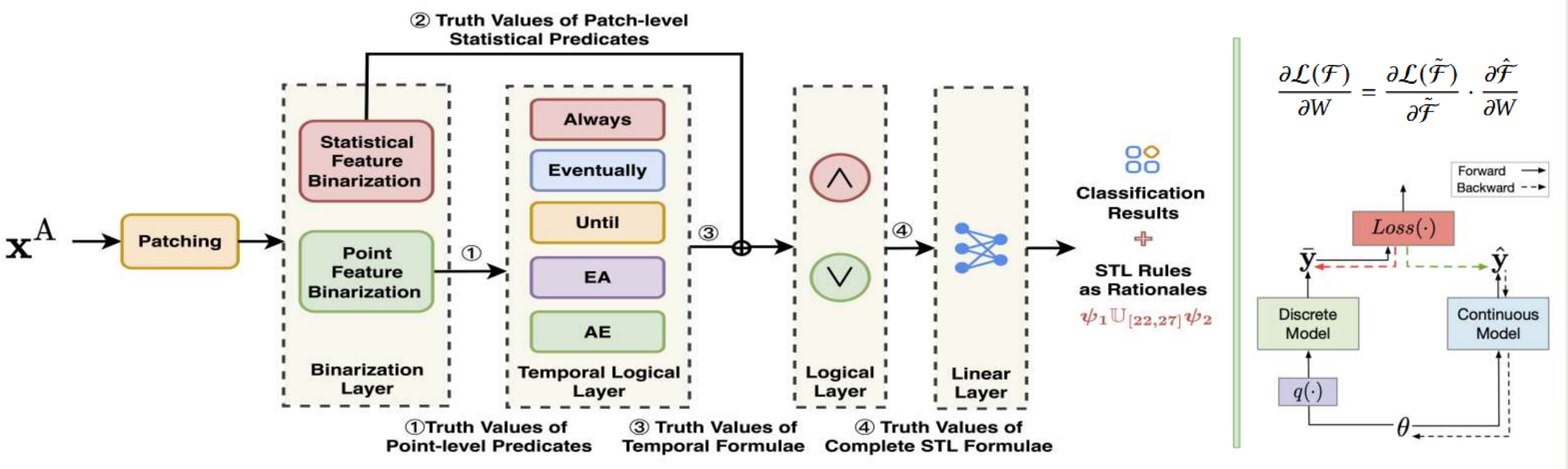
  Definition 3.1. STL formulae [22] can be recursively defined as:

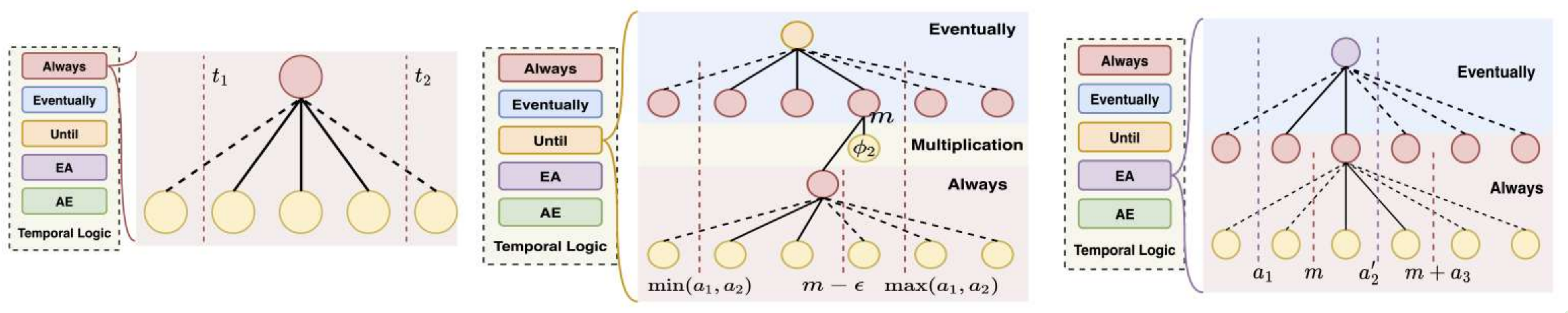
$$\phi := \top | \psi | \neg\phi | \phi_1 \wedge \phi_2 | \phi_1 \vee \phi_2 | \Box_I \phi | \Diamond_I \phi | \phi_1 \mathbb{U}_I \phi_2, \quad (1)$$

- **A subset for describing temporal properties of time series**
  - Atomic predicate (AP) $\psi := (f(x) \geq 0)$
  - Integer-bound time interval $I = [t_1, t_2]$
  - Always $\Box_{[t_1,t_2]}\phi$: each time point satisfies ϕ
  - Eventually $\Diamond_{[t_1,t_2]}\phi$: there exists a time point satisfying ϕ
  - Until $\phi_1 \mathbb{U}_{[t_1,t_2]} \phi_2$: there exists a time point satisfying ϕ_2, and each time point before it satisfies ϕ_1
  - EA $\Diamond_{[t_1,t_2]}\Box_{[0,t_3]}\phi$: Always ϕ on an unfixed interval
  - AE $\Box_{[t_1,t_2]}\Diamond_{[0,t_3]}\phi$: frequent occurrences of ϕ
  - Negation-free
  - Just allow temporal operators to act on atomic predicates

Model

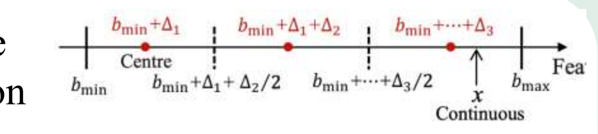
- **TemporalRule**: a novel neuro-symbolic model for interpretable time series classification
  - Automatically learns Signal Temporal Logic (STL) rules through four layers
  - Reliable: optimizes the model via gradient grafting (hard version + soft version)
  - Intuitive: simulates temporal operators through adaptively learning the time bounds





Binarization Layer

- Patch-wise bin division



- Probability $p_{t,k} = p_{(j,m),k} = \dfrac{\exp(-\tau \cdot D(x_{(j,m)}, c_{j,k}))}{\sum_{k'=1}^{K} \exp(-\tau \cdot D(x_{(j,m)}, c_{j,k'}))}$
- Truth value of AP $b_{t,k} = b_{(j,m),k} = \begin{cases} 1, & k = \operatorname{argmax}_{k'=1,\ldots,K} p_{(j,m),k'} \\ 0, & \text{otherwise} \end{cases}$

Temporal Logical Layer

- Always

$$w_{(j,m),k} = \mathrm{Sigmoid}((a_1 - m)(m - a_2))$$
$$r^{\mathsf{Always}}_{j,k} = \mathsf{Always}(\mathbf{b}_{j,k}, a_{j,k,1}, a_{j,k,2}) = Conj_+(\mathbf{b}_{j,k}, \mathbf{w}_{j,k}) = \mathbb{P}(\prod_{m=0}^{T^P-1} (F_c(b_{(j,m),k}, w_{(j,m),k}) + \epsilon))$$

- Until

$$r^{\mathsf{Until}}_{j,k_1,k_2} = \mathsf{Until}(\mathbf{b}_{j,k_1}, \mathbf{b}_{j,k_2}, a_{j,k_1,k_2,1}, a_{j,k_1,k_2,2}) = \mathsf{Eventually}(\mathbf{b}'_{j,k_1,k_2}, a_{j,k_1,k_2,1}, a_{j,k_1,k_2,2})$$
$$b'_{(j,m),k_1,k_2} = b_{(j,m),k_2} \times r^{\overline{\mathsf{Until}}}_{(j,m),k_1}$$
$$r^{\overline{\mathsf{Until}}}_{(j,m),k_1} = \mathsf{Always}(\mathbf{b}_{j,k_1}, \min(a_1, a_2), m - \epsilon)$$

- EA

$$r^{\mathsf{EA}}_{j,k} = \mathsf{Eventually}(\mathbf{r}^{\overline{\mathsf{EA}}}_{j,k}, a_{j,k,1}, a'_{j,k,2}) = 1 - Conj_+(1 - \mathbf{r}^{\overline{\mathsf{EA}}}_{j,k}, \mathbf{w}_{j,k})$$
$$r^{\overline{\mathsf{EA}}}_{(j,m),k} = \mathsf{Always}(\mathbf{b}_{j,k}, m, m + a_3)$$

Training Objective

- MSE loss for classification + clustering loss for binarization $\mathcal{L} = \mathcal{L}_{cla} + \lambda \mathcal{L}_{clu}$

Experiments

Overall Results and Ablation Study

- **Datesets:** 8 real-world datasets in 3 safety-critical domains
- **Baselines:** 6 rule-based models (2 dictionary-based, 2 shapelet-based and 2 neuro-symbolic)
- TemporalRule can actively guide the neural network to locate key intervals and generate advanced features, leading to accurate and stable classification results.

Method	Epilepsy2	NerveDamage	ECG200	ECG5000	EOGVerticalSignal	SharePriceIncrease	GunPoint	GunPointOVY
BOP	88.98	**100.00**	78.57	90.98	14.36	68.53	98.67	93.02
SAX-VSM	90.46	97.56	83.54	91.45	21.82	68.53	98.67	87.30
LS	76.38	46.34	87.14	94.02	17.96	68.43	**100.00**	99.68
ST	94.75	97.56	84.02	**94.34**	44.20	62.94	**100.00**	94.92
RRL	75.15	46.34	78.00	92.18	25.14	58.80	80.67	**100.00**
NSTSC	92.57	**100.00**	87.00	93.40	37.29	65.84	96.00	**100.00**
$-\mathbf{x}^P$	84.48	95.12	78.00	93.40	44.15	68.22	86.67	**100.00**
$-\mathbf{x}^D$	88.57	95.12	77.00	93.18	43.92	64.80	90.00	99.68
$-\mathbf{b}^S$	90.60	97.56	78.00	91.71	29.28	63.35	87.33	99.37
$-\mathbf{r}^A\mathbf{r}^E$	88.58	95.12	86.00	92.38	41.71	62.21	90.67	99.68
$-\mathbf{r}^{\mathsf{Until}}$	89.81	92.68	86.00	92.36	41.44	67.81	94.67	99.37
$-\mathbf{r}^{\mathsf{AE}}$	88.31	90.24	88.00	92.66	38.40	63.98	92.67	99.68
$-\mathbf{r}^{\mathsf{EA}}$	83.12	95.12	88.00	92.33	31.49	47.41	94.00	99.68
TemporalRule	**94.80**	**100.00**	**89.00**	**94.34**	**45.86**	**69.15**	**100.00**	**100.00**

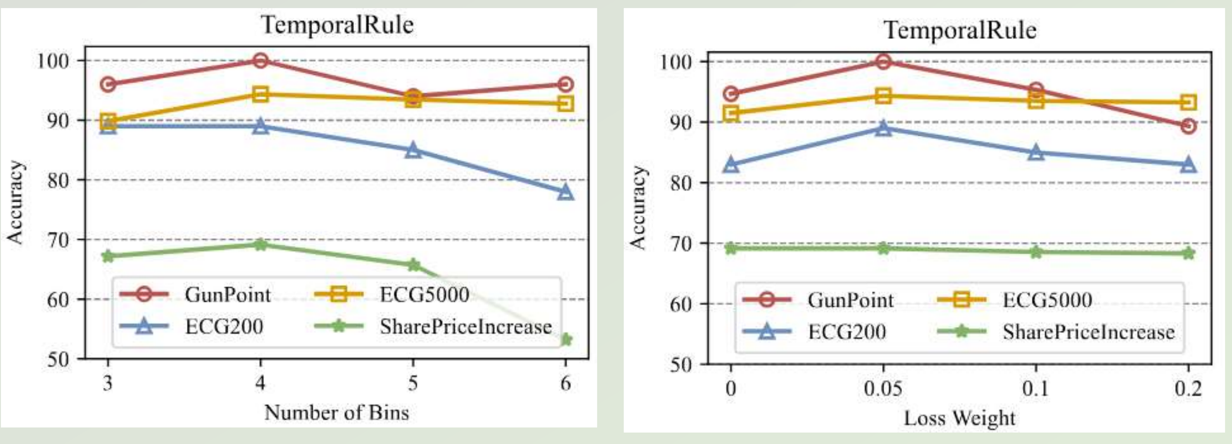
Category-level Interpretability Analysis

- TemporalRule directly learns holistic temporal features, focusing on condensed time intervals rather than isolated time points, so can produce more concise and intuitive rules consistent with human understanding of consecutive time series patterns.
- This arises from the adaptive learning of time bounds, instead of relying on point-wise weights.

Dataset	Model	Rule
ECG200	RRL	$\phi := (x_{29} > -1.835) \wedge (x_{82} > -0.455) \wedge (x_{83} > -0.305) \wedge (x_{87} > -0.485) \wedge (x_{94} > -0.250) \wedge (x_3 < 3.260) \wedge (x_{44} < 0.824) \wedge (x_{69} < 0.662) \wedge (x_{74} < 0.841) \wedge (x_{75} < 0.872) \wedge (x_{76} < 1.205) \wedge (x_{77} < 0.796)$
	NSTSC	$\phi := (x_{32} \leq -2.220) \wedge (x_{37} \leq -0.839) \wedge (x_{38} \leq -0.406) \wedge (x_{39} \leq -0.144)$
	TemporalRule	$\phi := \psi_1 \mathbb{U}_{[22,27]} \psi_2$, where $\psi_1 := (x^D < 0.16)$ and $\psi_2 := (0.41 \leq x^D < 0.91)$
GunPoint	RRL	$\phi := (x_{30} > -0.858) \wedge (x_{53} < 1.247) \wedge (x_{56} < 1.351) \wedge (x_{57} < 1.332) \wedge (x_{58} < 1.338) \wedge (x_{60} < 1.897) \wedge (x_{97} < 1.299) \wedge (x_{98} < 1.141)$
	NSTSC	$\phi := (x_{34} \leq -0.06) \vee (x_{35} \leq 0.04) \vee (x_{36} \leq 0.28) \vee (x_{109} \leq -0.48) \vee (x_{110} \leq -0.65)$
	TemporalRule	$\phi := \Diamond_{[96,113]}\psi$, where $\psi := (x^R < -0.604)$

Hyperparameter Analysis

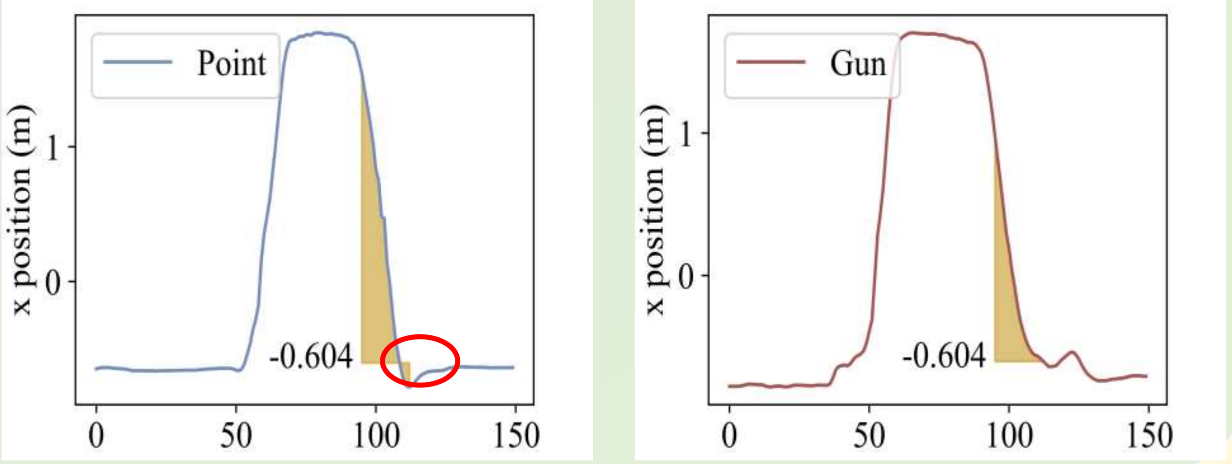
- The number of bin (K) can be consistently selected as a fixed value 4.
- The loss weight (λ) can be consistently selected as a fixed value 0.05.



Instance-level Interpretability Analysis

- We can specially examine the time interval from 96 to 113 in each time series instance to determine whether it should be classified as "Point" or "Gun".
- TemporalRule can provide precise and concise evidence within each time series for clear justification of classification decisions.

$$\phi := \Diamond_{[96,113]}\psi, \text{ where } \psi := (x^R < -0.604)$$



Future Work

- Future work will explore more expressive subsets of temporal logic and extend to multivariate time series classification tasks.