

UNIVERSITÄT DES SAARLANDES

CISPA HELMHOLTZ CENTER FOR INFORMATION SECURITY

NEURAL INFORMATION PROCESSING SYSTEMS

CPEC CENTER FOR PERSPICUOUS COMPUTING

NEURO EXPLICIT MODELS

ZUSE SCHOOL ELIZA

What Comes Next and Why: Interpretable Next-Event Prediction with Neuro-Symbolic Rules

Presenter: Tim Nico Bauerschmidt

Tim Nico Bauerschmidt



Isabel Valera



Jilles Vreeken

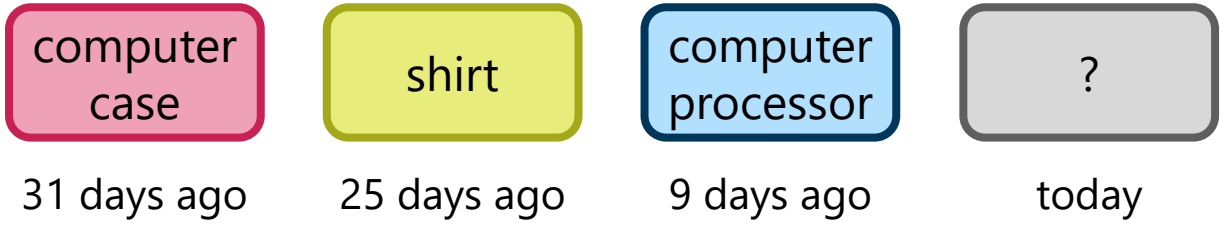


Motivation

Next-event prediction

- **What** categorical event **comes next** in a timestamped sequence?

What will the customer buy next?



Neural Networks

✓ predictive model
✓ accurate
✗ interpretable

Sequential Rule Miners

✗ predictive model
? accurate
✓ interpretable

Goal

Address the accuracy–interpretability tension

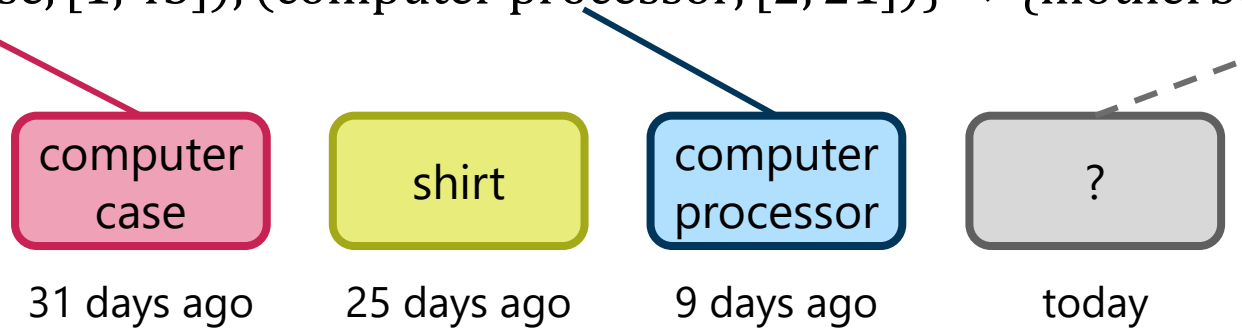
Sequential Rules

Sequential rules $A \Rightarrow B$ symbolically describe **why** we predict an event

- Antecedent A: When does the rule activate?
- Consequent B: What do we predict if A holds?

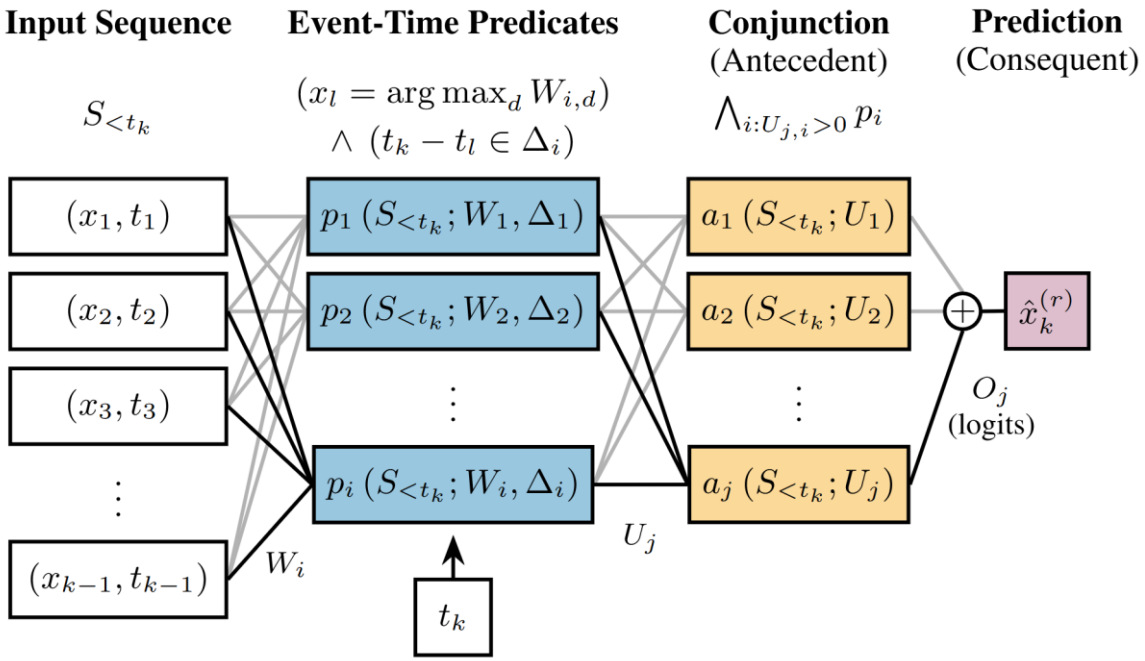
Our Rules:

$$\{(\text{computer case}, [1, 43]), (\text{computer processor}, [2, 21])\} \Rightarrow \{\text{motherboard}, \text{power supply unit}\}$$



The next event is likely "motherboard" or "power supply unit".

RULES: Learning Rules in a Neural Network



First Layer: Which events are relevant at which offset? (Event-Time Predicates)

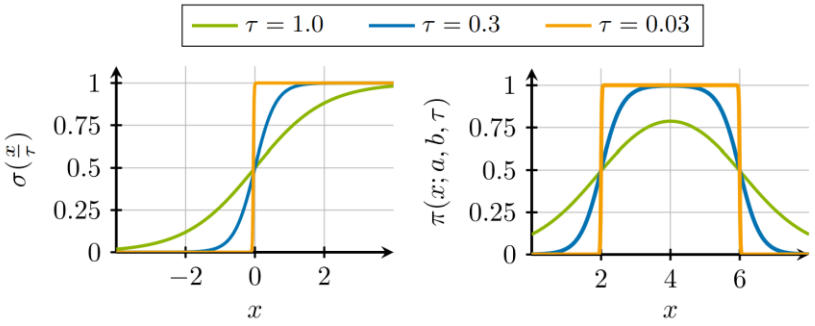
Second Layer: Which predicates from the first layer should form an antecedent?

Output Layer: Which events likely follow the antecedent? (Consequent)

Training

- Rule discovery is a combinatorial problem
  - **Temperature annealing** to enable gradient-based learning
  - Next-event prediction loss (cross-entropy)

Temperature annealing: $\tau \to 0$

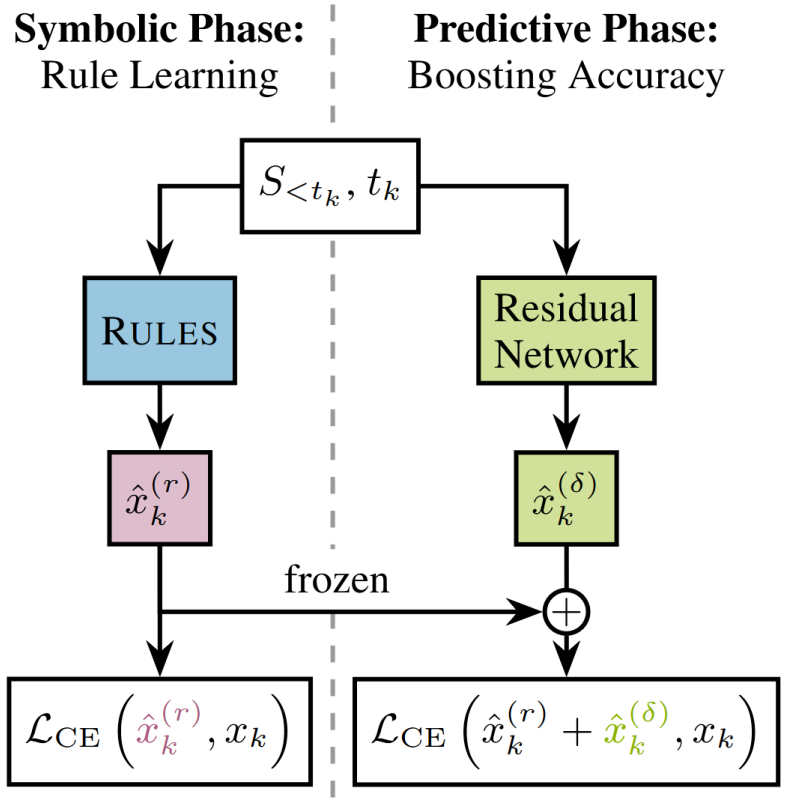


RUNES: Capturing Structure Beyond Rules

Not all data strictly follows symbolic rules.

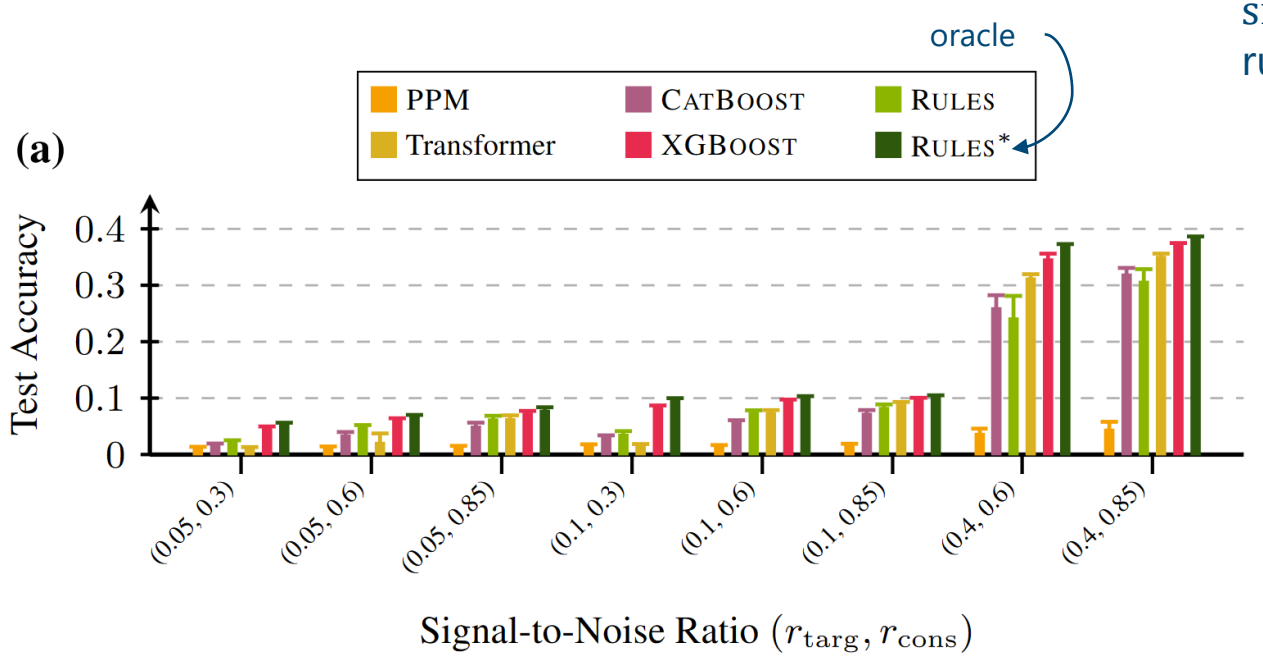
→ Combine RULES with a **residual network**

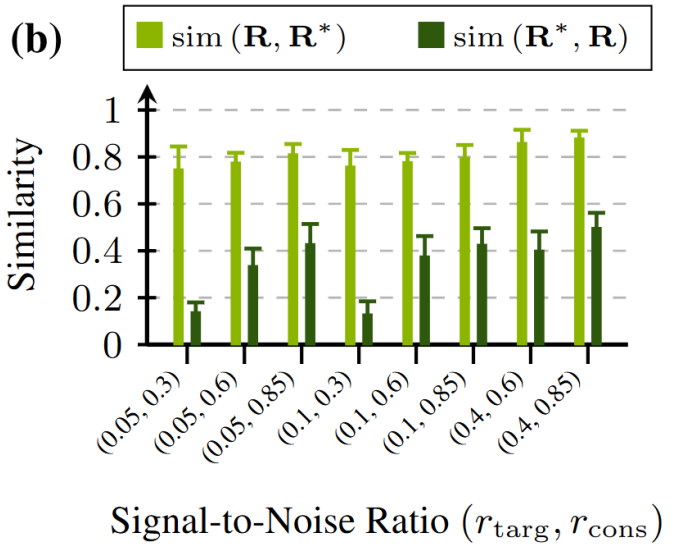
- Transparent combination via additive logits
  - Traceability: *Does the residual change the prediction of RULES? Which rules agree with the full model? …*
- Expose the **accuracy–interpretability trade-off**
  - User can decide whether to prioritize interpretability or accuracy.



Experiments: Synthetic Data

$\text{sim}(\mathbf{R}, \mathbf{R}^*)$ = How well does the discovered ruleset $\mathbf{R}$ map onto the ground truth $\mathbf{R}^*$?





RULES faithfully **recovers ground-truth** generative structure from synthetic data.

Experiments: Real Data

UNIVERSITÄT DES SAARLANDES

Table 1: *Test accuracy (Acc.) in % and number of rules ($|\mathbf{R}|$) on real-world datasets.* Best accuracy results are in **bold**, second best are underlined. RUNES performs close to the black-box transformer.

	RULES		RUNES	PPM		CATBOOST		XGBOOST		Transformer
Datasets	$\|\mathbf{R}\|$	Acc.	Acc.	$\|\mathbf{R}\|$	Acc.	$\|\mathbf{R}\|$	Acc.	$\|\mathbf{R}\|$	Acc.	Acc.
Holidays	26	97.57	**100.00**	14	**100.00**	1129	**100.00**	1455	**100.00**	**100.00**
Chorales	33	21.00	37.00	180	31.40	11596	37.10	34462	37.72	**39.35**
Commits	51	19.60	18.31	80	18.11	380	15.55	83965	17.03	**20.10**
Rolling Mill	77	34.13	70.44	9583	67.64	13809	65.64	640723	68.84	**71.24**
Amazon	42	6.53	17.55	993	15.16	13123	11.79	5149683	15.37	**17.61**
Avg. Rank		5.2	2.4		3.6		4.0		2.8	**1.0**

RULES produces compact, interpretable rule sets and works well when symbolic structure suffices.

RUNES closes the accuracy gap.

Experiments: Real Data

Across real-world datasets RUNES corrects errors (✗)
but rarely flips correct predictions (✓) of RULES.

RUNES	RULES			
	Rule Active		No Rule Active	
	✓	✗	✓	✗
✓	0.40	0.17	0.19	0.36
✗	0.03	0.40	0.03	0.42
Data Frac.	0.64		0.36	

Decision Rule: Trust the rules when the rule and residual networks agree and incorporate residual logits otherwise.

Unsupervised metrics also provide practical signals for distinguishing corrective from spurious residual changes.

Implementation:
https://github.com/IamginE/runes-rules

Thank you!

