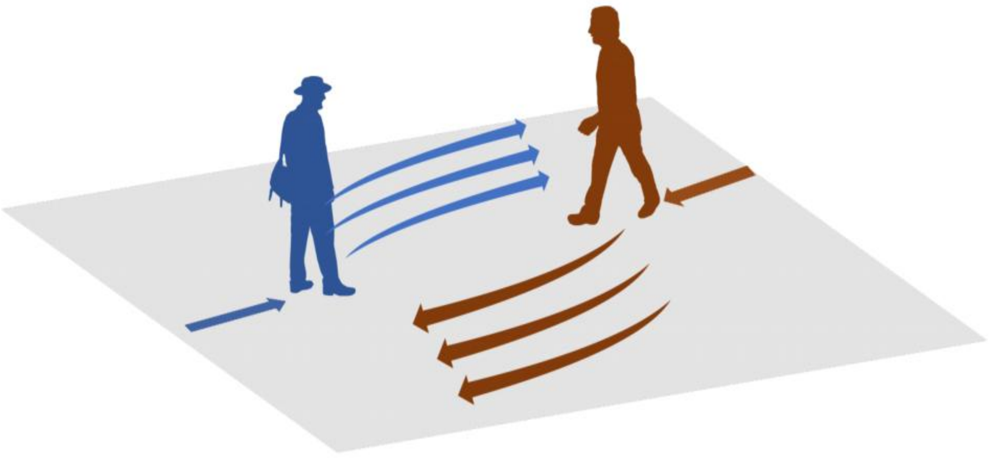
Discovering Intrinsic Spatial-Temporal Logic Rules to Explain Human Actions

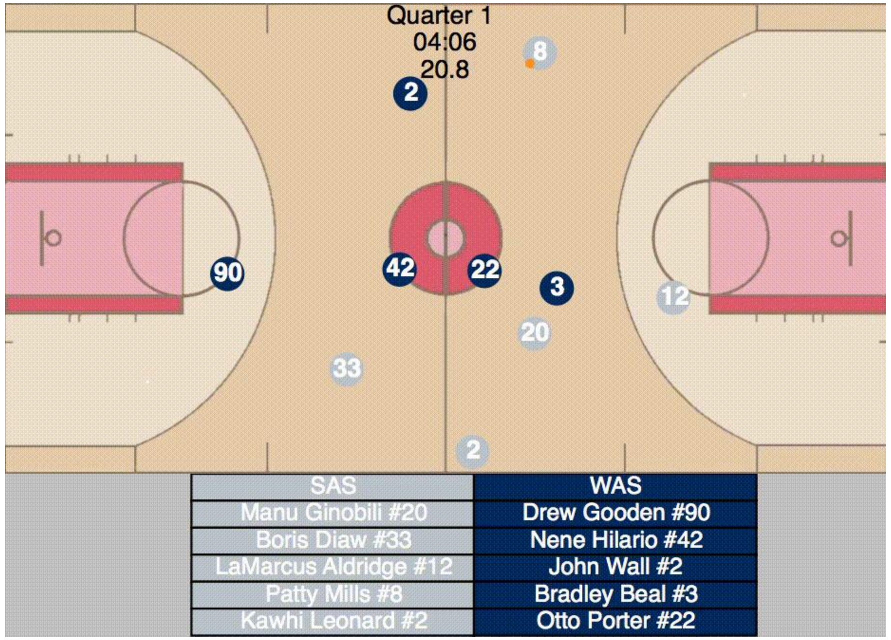
Chengzhi Cao, Chao Yang, Ruimao Zhang, Shuang Li

Introduction

There has been a great interest and business value in unveiling the human logic from the observational movements.



Pedestrian trajectory prediction



Offensive Strategies for Basketball Players

Introduction

Spatial-Temporal Predicate In our paper, we extend the above static predicates to spatial-temporal predicates, which include spatial-temporal *property* predicates and spatial-temporal *relation* predicates.

Specifically, the spatial-temporal *property* predicates are defined as

$$X(v) : \mathcal{C} \times \cdots \times \mathcal{C} \times \mathcal{T} \times \mathcal{S} \mapsto \{0, 1\}.$$

We will consider spatial-temporal logic rules where the body part contain spatial-temporal predicates as relation constraints. For example, a sensible rule will look like

$$f : Y_{\text{TurnAround}}(c, t, s) \leftarrow X_{\text{PickUpKey}}(c, t, s) \bigwedge R_{\text{InFront}}((c', t, s'), (c, t, s)) \bigwedge R_{\text{Behind}}((c'', t, s''), (c, t, s))$$

where $c \in \mathcal{C}_{\text{person}}$, $c' \in \mathcal{C}_{\text{block}}$, and $c'' \in \mathcal{C}_{\text{key}}$. In general, the *spatial-temporal logic rule* in our paper is defined as a logical connectives of predicates, including property predicates and spatial-temporal relation predicates,

$$f : Y(v) \leftarrow \bigwedge_{X_{\text{property}} \in \mathcal{X}_f} X_{\text{property}}(v') \bigwedge_{R_{\text{spatial-temporal}} \in \mathcal{R}_f} R_{\text{spatial-temporal}}(v'', v) \quad (1)$$

where $Y(v)$ is the *head predicate* evaluated at the entity-time-location triplet v, $\mathcal{X}_f$ is the set of property predicates defined in rule f, and $\mathcal{R}_f$ denotes the set of spatial-temporal relation predicates defined in rule f.

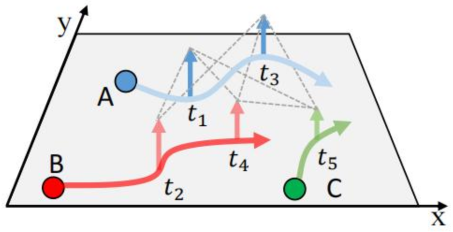


Figure 1: Illustration of feature construction using a simple logic formula with temporal relation predicate ($t_1 < t_2$), $f : Y \leftarrow A \wedge B \wedge C \wedge (A \text{ Before } B)$. The rule defines the template to gather combinations of the body predicate history events. Here predicate A has 2 events and predicate B has 1 event, the temporal relation constraint would lead to valid combinations (also called "paths"). This type of feature construction can be extended to spatial-temporal cases, where we count the valid paths as the feature.

Introduction

Given the rule set Fκ, we model the probability of the event κ as a log-linear function of the features, i.e.,

$$p(\kappa|v,\mathcal{H}_t) \propto \exp\left(\sum\nolimits_{f\in\mathcal{F}_\kappa} w_f \cdot \phi_f(\kappa|v,\mathcal{H}_t)\right), \quad (3)$$

where w = [wf]f∈F ≥ 0 are the learnable weight parameters associated with each rule. All the model parameters can be learned by maximizing the likelihood, which can be computed using the above Eq. (3)

$$\max_{\theta,w} \mathcal{O}(\theta,w) = \mathbb{E}_{(\kappa,v,\mathcal{H}_t)}[\log \mathbb{E}_{p_\theta}[p_w(\kappa|v,\mathcal{H}_t)]]. \quad (4)$$

Our goal is to maximize the likelihood of the observed human action events $\{\kappa^{(i)}\}_{i=1,\ldots,n}$. Using the chain rule, we have

$$\log p_w(\{\kappa^{(i)}\}_{i=1,\ldots,n}) = \sum_{i=1}^{n} \log p_w(\kappa^{(i)} \mid v^{(i)}, \mathcal{H}_{t(i-1)}). \quad (5)$$

We deploy Transformer-based framework to model the rule generator p_θ. We define the distribution of a set of rules as follows:

$$p_\theta(z \mid v,\mathcal{H}_t) = \Psi(z|N, \text{Trans}_\theta(v,\mathcal{H}_t)), \quad (8)$$

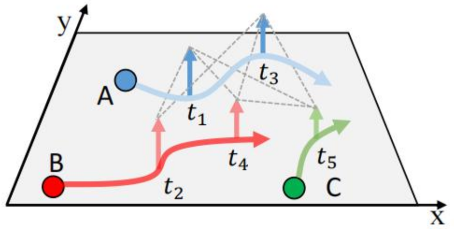
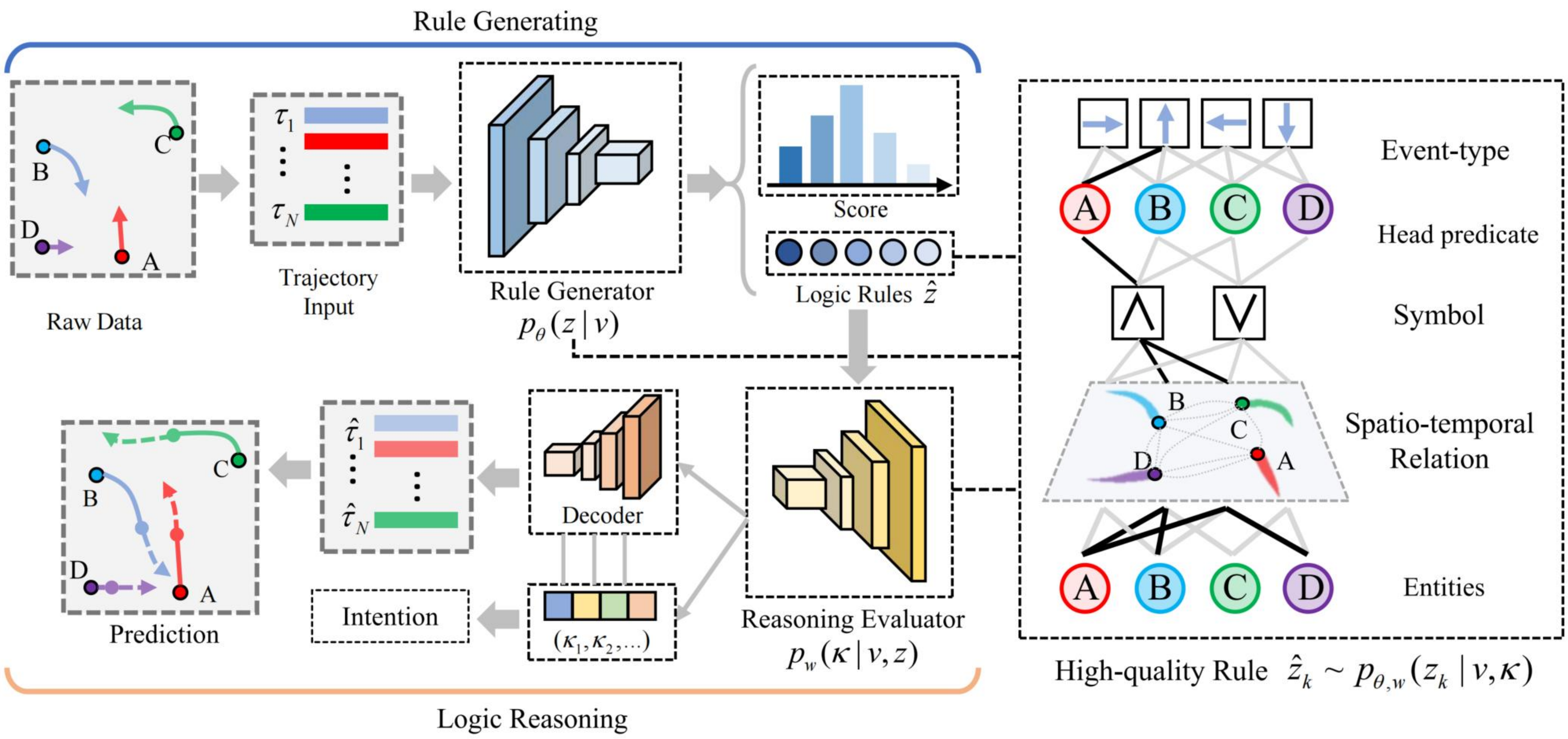


Figure 1: Illustration of feature construction using a simple logic formula with temporal relation predicate $(t_1 < t_2)$, $f: Y \leftarrow A \wedge B \wedge C \wedge (A \text{ Before } B)$. The rule defines the template to gather combinations of the body predicate history events. Here predicate A has 2 events and predicate B has 1 event, the temporal relation constraint would lead to valid combinations (also called "paths"). This type of feature construction can be extended to spatial-temporal cases, where we count the valid paths as the feature.

Framework



Rule Generating
Raw Data
Trajectory Input
Rule Generator
$p_\theta(z \mid v)$
Score
Logic Rules $\hat{z}$
Event-type
Head predicate
Symbol
Spatio-temporal Relation
Entities
High-quality Rule $\hat{z}_k \sim p_{\theta,w}(z_k \mid v, \kappa)$
Decoder
Intention
$(\kappa_1, \kappa_2, \dots)$
Reasoning Evaluator
$p_w(\kappa \mid v, z)$
Prediction
Logic Reasoning

Result

Table 2: Quantitative results (ADE_{20}/FDE_{20}, and accuracy) of trajectory prediction in NBA dataset. The bold/underlined font represent the best/second best result.

Times	Y-Net	MID	NSP-SFM	Social-SSL	Social Implicit	Social VAE	ABC+	Ours
1.0s	<u>0.38/0.48</u>	0.45/0.59	0.41/0.52	0.48/0.61	0.45/0.53	0.49/0.66	0.45/0.56	**0.30/0.40**
2.0s	<u>0.63/0.93</u>	0.76/1.06	0.67/0.94	0.76/1.08	0.72/0.96	0.77/1.11	0.75/0.98	**0.58/0.88**
3.0s	<u>0.94/1.34</u>	1.06/1.40	0.98/1.35	1.06/1.43	1.00/1.39	1.11/1.46	1.03/1.41	**0.87/1.31**
4.0s	<u>1.17/1.61</u>	1.32/1.74	1.18/1.63	1.35/1.78	1.19/1.66	1.37/1.79	1.23/1.67	**1.13/1.60**
Acc.	0.69	0.65	<u>0.70</u>	0.68	0.69	0.64	0.65	**0.73**

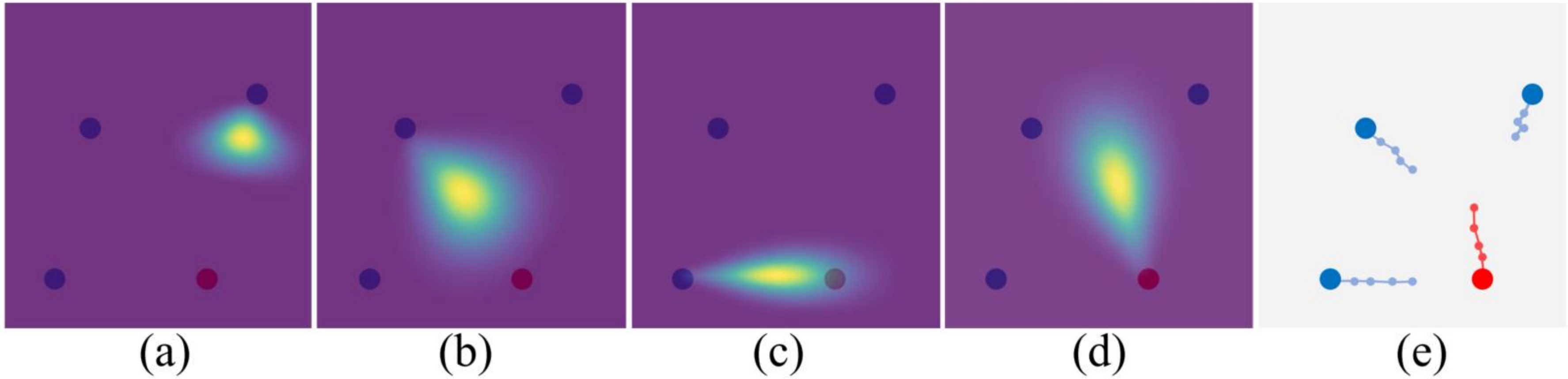
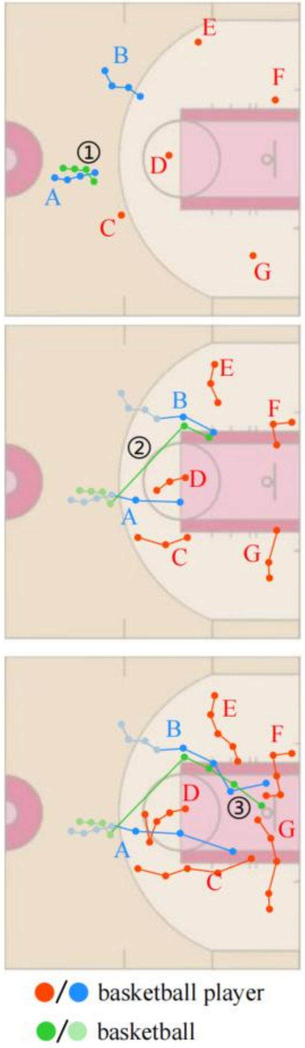


Figure 4: Estimated distributions for each player's intention in cartesian space.

Visualization



Player A Rule ①: Left(B,A)∧Right(C,A)∧Front(D,A)→Pass(A,B)
Explanation:

- Player A is carrying the ball
- Player A is defended by player C and player D
- The partner B is on the left of player A
- Player A passes the ball to player B

Player B Rule ②: Left(E,B)∧Right(D,B)→Go_Front(B)
Explanation:

- Player E and player F is moving to the player B
- Player B is defended by player E on left side
- Player B is defended by player D on right side
- Player B goes front and carry the ball

Player B Rule ③: Left(E,B)∧Right(G,B)∧Front(F,B)→Shoot(B)
Explanation:

- Player B is defended by player E on left side
- Player B is defended by player G on right side
- Player B is defended by player F on front side
- Player B shoots at the basket

Figure 5: Visualization and explanation of logic rules in NBA dataset.

Thanks for Listening!