Vehicle Interaction Learning

Xiaosong Jia

08/26/2019

About covariance/correlation

$$Cov(X,Y) = E[(X-\mu_x)(Y-\mu_y)]$$

$$\rho = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

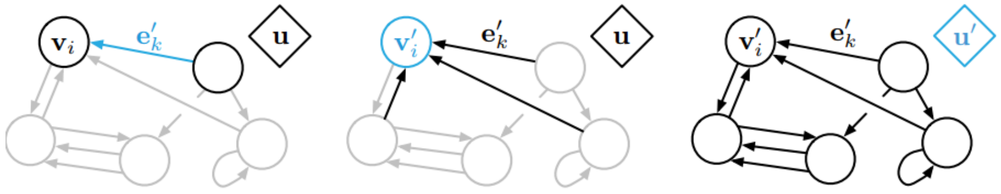
	Covariance (entire interval)	Correlation (entire interval)	Covariance (interaction interval)	Correlation (interaction interval)
HighD	0.018	0.519	0.001	0.321
NGSIM	0.68	0.401	0.033	0.092
FT	-0.018	0.084	-0.006	-0.096
SR	-0.004	0.023	-0.006	-0.086

Graph Neural Network

- Aggregation and Updating: [DeepMind, Google Brain, MIT 2018]

$$
\begin{aligned}
\mathbf{e}'_k &= \phi^e\left(\mathbf{e}_k, \mathbf{h}_{r_k}, \mathbf{h}_{s_k}, \mathbf{u}\right) & \bar{\mathbf{e}}'_i &= \rho^{e\rightarrow h}\left(E'_i\right) \\
\mathbf{h}'_i &= \phi^h\left(\bar{\mathbf{e}}'_i, \mathbf{h}_i, \mathbf{u}\right) & \bar{\mathbf{e}}' &= \rho^{e\rightarrow u}\left(E'\right) \\
\mathbf{u}' &= \phi^u\left(\bar{\mathbf{e}}', \bar{\mathbf{h}}', \mathbf{u}\right) & \bar{\mathbf{h}}' &= \rho^{h\rightarrow u}\left(H'\right)
\end{aligned}
\qquad (43)
$$

where $E'_i = \{(\mathbf{e}'_k, r_k, s_k)\}_{r_k=i,\ k=1:N^e}$, $H' = \{\mathbf{h}'_i\}_{i=1:N^v}$, and $E' = \bigcup_i E'_i = \{(\mathbf{e}'_k, r_k, s_k)\}_{k=1:N^e}$. The ρ functions must be invariant to permutations of their inputs and should take variable numbers of arguments.



(a) Edge update (b) Node update (c) Global update

- Simple Example: Neural FPs (NIPS 2015)

$$
\mathbf{x} = \mathbf{h}_v^{t-1} + \sum_{i=1}^{|\mathcal{N}_v|} \mathbf{h}_i^{t-1}
$$

$$
\mathbf{h}_v^t = \sigma\left(\mathbf{x}\mathbf{W}_t^{|\mathcal{N}_v|}\right)
$$

GNN Traffic [2019.5 Technische Universitat Munchen]

- Problem: Given 5s trajectories of several cars, predict the next 5s of their car.

- Motivation: whether representing interactions as graphs leads to better performance for prediction?

- Models: GCN (NIPS 2015) and GAT(ICLR 2018) with adaption to extract features for each node and use linear layers to predict

GNN Traffic [2019.5 Technische Universitat Munchen]

- Dataset:

NGSIM:

1. As Thiemann et al. [23] show, position, velocity, and acceleration data contain unrealistic values. We therefore smooth the positions using double-sided exponential smoothing with a span of 0.5s and compute velocities from these.
2. We subsample the trajectory data to 1 FPS.
3. The NGSIM dataset still contains many artifacts (errors in bounding boxes, undetected cars, complete non-overlap of bounding box and true vehicle)

HighD:

1. The dataset consists mainly of roads without on- or off-ramps and without traffic jams, interaction seems limited: Only about 5% of the cars experience a lane change.

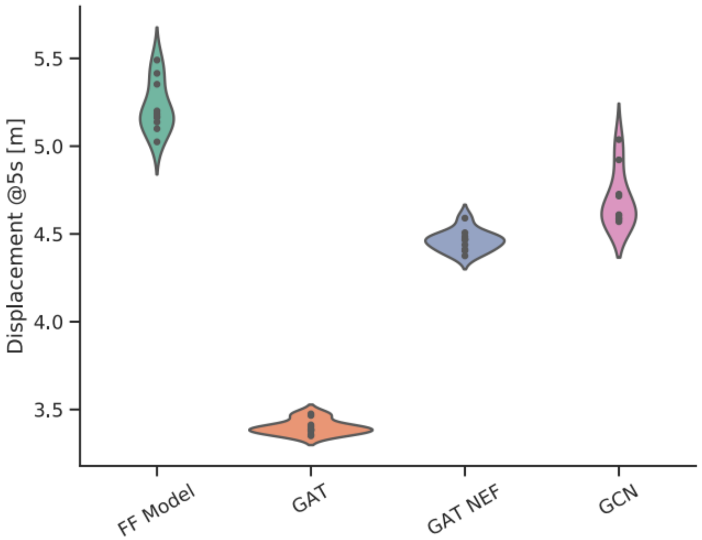
GNN Traffic [2019.5 Technische Universitat Munchen]

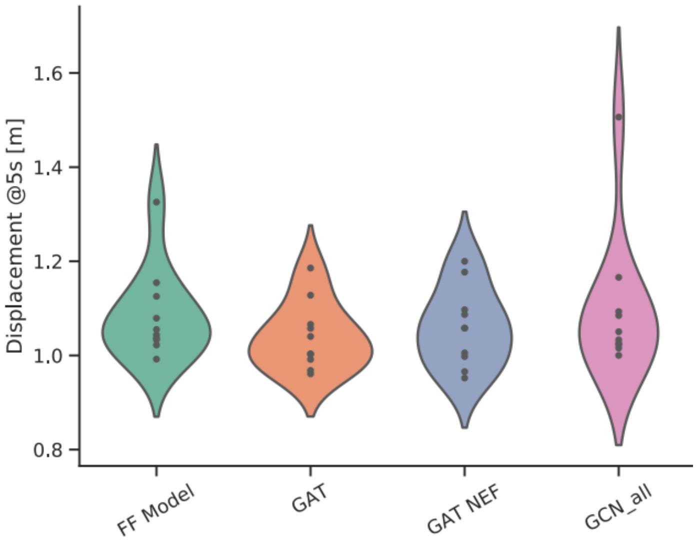
Graph Construction:

1. Only self connection or
2. All connection or
3. Preceding connection or
4. Close vehicles (at most 8 cars)
5. Features for nodes: all 5s trajectory
6. Features for edges: all 5s relative distance

GNN Traffic [2019.5 Technische Universitat Munchen]

- Experiments: baseline*: Constant Velocity Model, Intelligent Driver Model* [*Physical Review E 2000*], simple neural networks





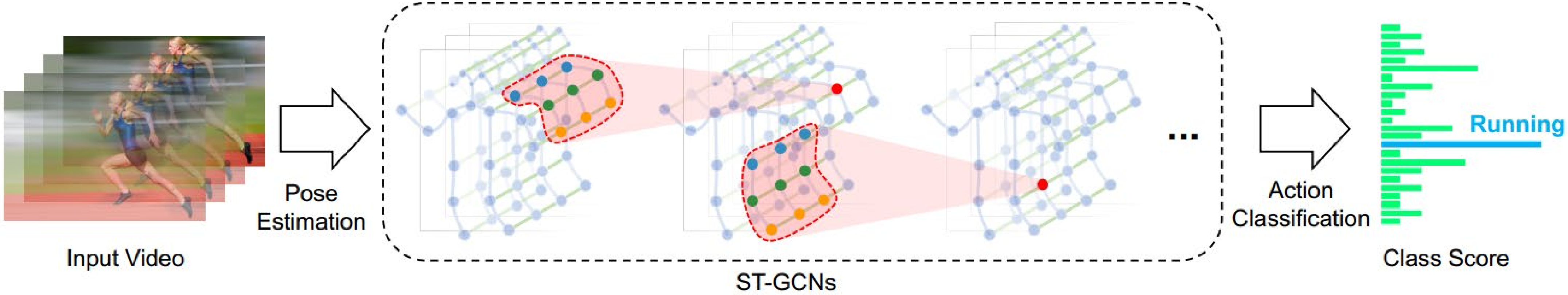
- About connection

Self-Connections	2.68 ± 0.05	5.08 ± 0.08
Preceding Connection	2.70 ± 0.04	5.11 ± 0.07
Neighbour Connection	1.93 ± 0.08	3.47 ± 0.13
All Connections ($\star$)	2.41 ± 0.02	4.42 ± 0.03

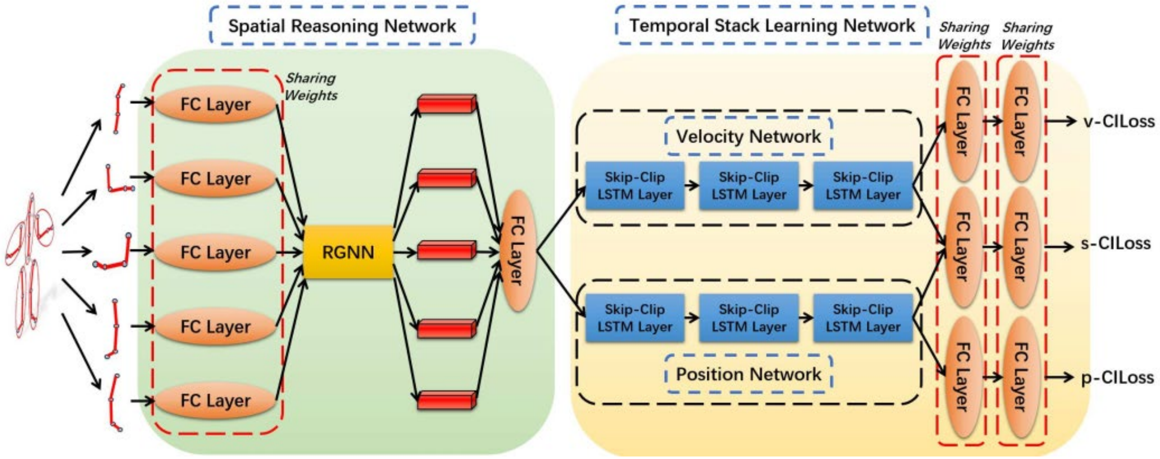
ST-GCN（AAAI 2018）

- What is a node's neighborhood? $\Delta t < T\ and\ Distrance < D$
- How to aggregate? (Kernel Design)

1. Uni-set
2. Distance-partition
3. Spatial-partition



Spatial Reasoning and Temporal Stack Learning [ECCV 2018]



Spatial Feature Extraction:

- e_k^t features of node k at time t
- r_k^t relation of node k at time t with other nodes
- m_k^t messages received by node k at time t
- q^t features of frame t

$$\boldsymbol{m}_k^t = \sum_{i \in \Omega_{v_k}} \boldsymbol{m}_{ik}^t = \sum_{i \in \Omega_{v_k}} \boldsymbol{W}_m \boldsymbol{s}_i^{t-1} + \boldsymbol{b}_m$$

$$\boldsymbol{s}_k^t = f_{lstm}\left(\boldsymbol{r}_k^{t-1}, \boldsymbol{m}_k^t, \boldsymbol{s}_k^{t-1}\right)$$

$$\boldsymbol{r}_k^t = \boldsymbol{r}_k^{t-1} + \boldsymbol{s}_k^t$$

$$\boldsymbol{r}^T = concat\left([\boldsymbol{r}_1^T, \boldsymbol{r}_2^T, ..., \boldsymbol{r}_k^T]\right), \forall k \in K$$

$$\boldsymbol{q} = f_r\left(\boldsymbol{r}^T\right)$$

Spatial Reasoning and Temporal Stack Learning [ECCV 2018]

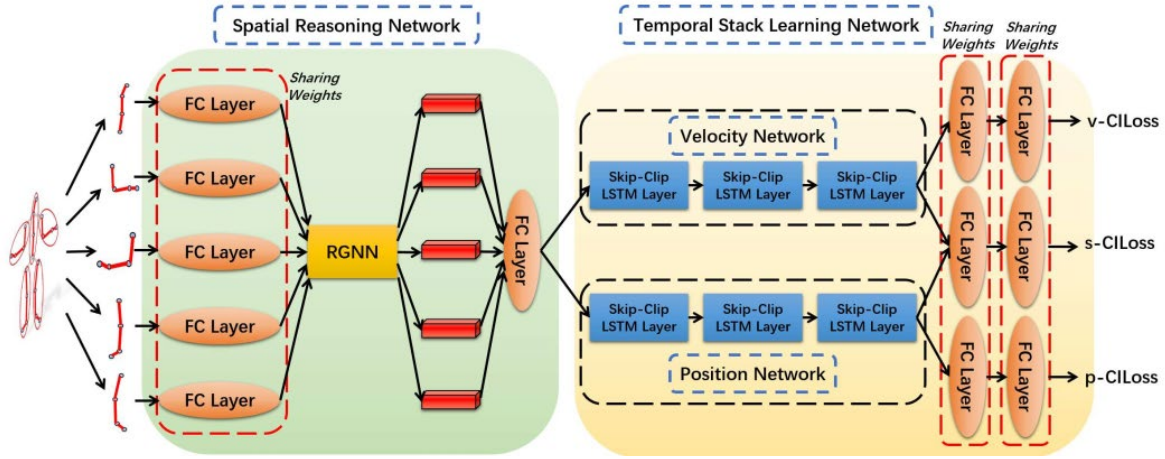
- The sequence is divided into M clips which includes d frames

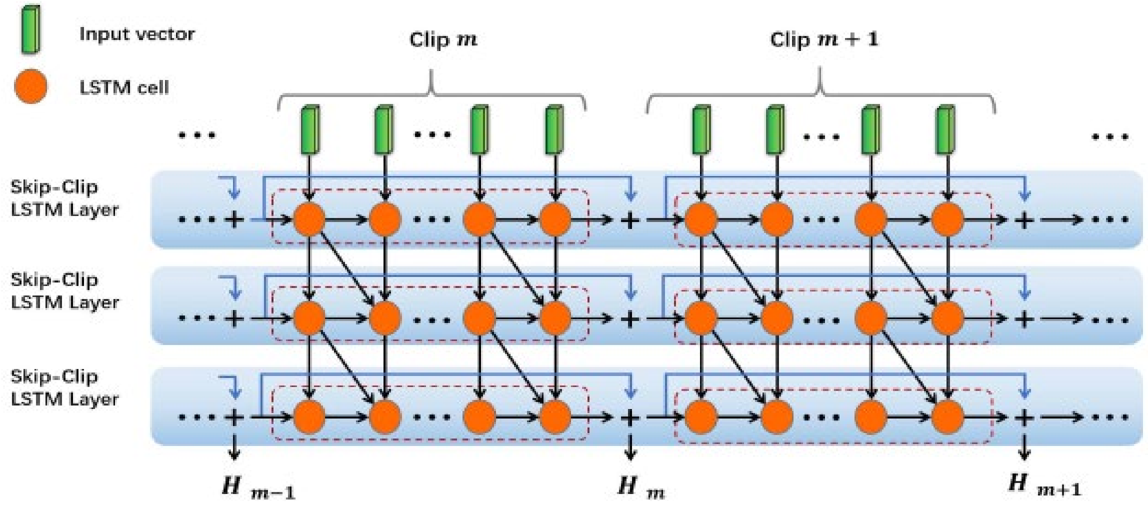
$\{Q_1, Q_2, \ldots, Q_M\}$ where $Q_i = \{q_{id+1}, q_{id+2}, \ldots, q_{(i+1)d}\}$

- Difference Vector $v_t = q_t - q_{t-1}$ and temporal difference features $V_j = \{q_{jd+1}, q_{jd+2}, \ldots, q_{(j+1)d}\}$

$$\boldsymbol{h}'_m = f_{LSTM}(Q_m) = f_{LSTM}(\{\boldsymbol{q}_{md+1}, \boldsymbol{q}_{md+2}, \ldots, \boldsymbol{q}_{(m+1)d}\})$$

$$\boldsymbol{H}_m = \boldsymbol{H}_{m-1} + \boldsymbol{h}'_m = \sum_{i=1}^{m} \boldsymbol{h}'_i$$

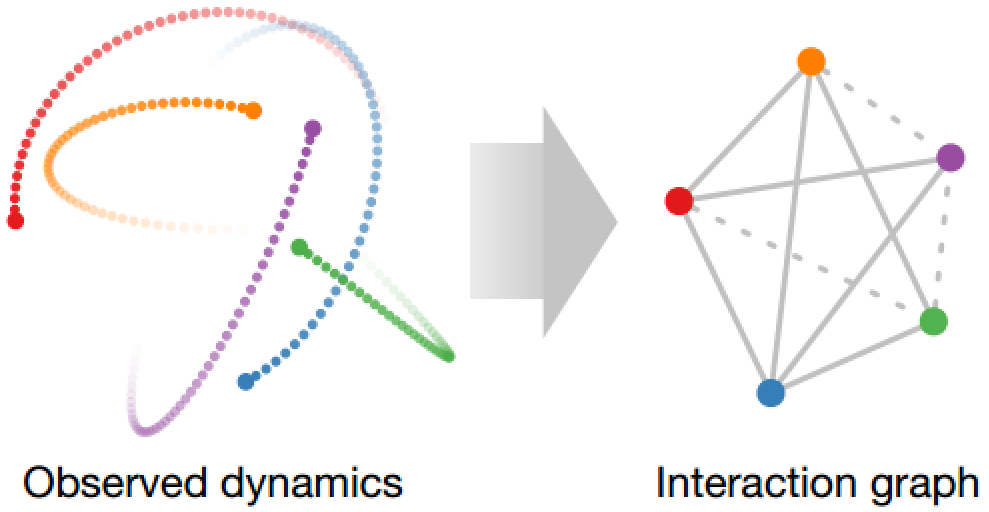




Neural Relational Inference for Interacting Systems [ICML 2018]

- an unsupervised model that learns to infer interactions while simultaneously learning the dynamics purely from observational data

- x_i^t feature vector of object i at time t (coordinates+velocity+⋯)
- N objects, T time steps



Neural Relational Inference for Interacting Systems [ICML 2018]

- Graph: node->object, edge->relationship (z_{ij} object I and object j's relation)

- Objective: $\mathcal{L} = \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}[\log p_\theta(\mathbf{x}|\mathbf{z})] - \mathrm{KL}[q_\phi(\mathbf{z}|\mathbf{x})||p_\theta(\mathbf{z})]$
- Decoder: Given the history trajectories of all objects and whether they have interactions, predict their future trajectories.

  $p_\theta(\mathbf{x}|\mathbf{z}) = \prod_{t=1}^{T} p_\theta(\mathbf{x}^{t+1}|\mathbf{x}^t, ..., \mathbf{x}^1, \mathbf{z})$

- Prior: constrains; To encourage a sparse graph, use a prior with higher probability on the non-edge label.

Neural Relational Inference for Interacting Systems [ICML 2018]

- Encoder: GNN on the fully connected graph

$$
\begin{aligned}
& \mathbf{h}_j^1 = f_{\text{emb}}(\mathbf{x}_j) \\
v{\to}e: \quad & \mathbf{h}_{(i,j)}^1 = f_e^1([\mathbf{h}_i^1, \mathbf{h}_j^1]) \\
e{\to}v: \quad & \mathbf{h}_j^2 = f_v^1(\textstyle\sum_{i \neq j} \mathbf{h}_{(i,j)}^1) \\
v{\to}e: \quad & \mathbf{h}_{(i,j)}^2 = f_e^2([\mathbf{h}_i^2, \mathbf{h}_j^2])
\end{aligned}
$$

$$\mathbf{z}_{ij} = \text{softmax}((\mathbf{h}_{(i,j)}^2 + \mathbf{g})/\tau) \qquad (9)$$

where $\mathbf{g} \in \mathbb{R}^K$ is a vector of i.i.d. samples drawn from a $\text{Gumbel}(0,1)$ distribution and τ (softmax temperature) is a parameter that controls the "smoothness" of the samples.

Neural Relational Inference for Interacting Systems [ICML 2018]

- Decoder:

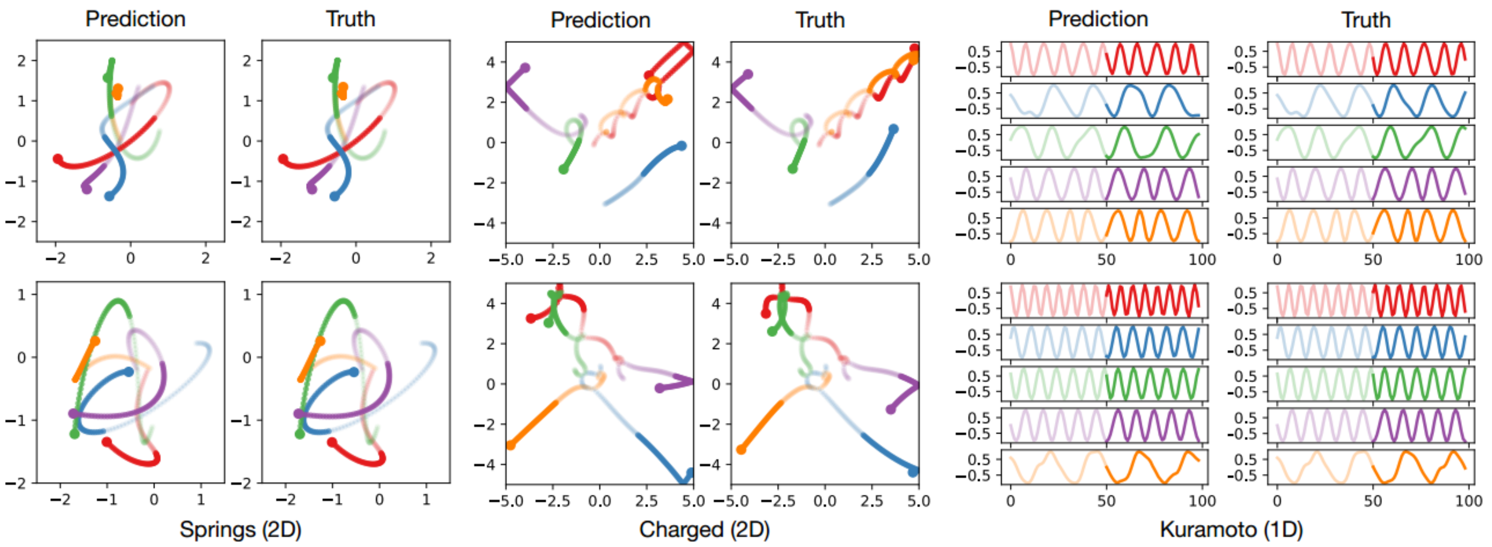
$$
\begin{aligned}
v{\to}e: \quad \tilde{\mathbf{h}}^t_{(i,j)} &= \sum_k z_{ij,k} \tilde{f}^k_e([\tilde{\mathbf{h}}^t_i, \tilde{\mathbf{h}}^t_j]) \\
e{\to}v: \quad \mathrm{MSG}^t_j &= \textstyle\sum_{i \neq j} \tilde{\mathbf{h}}^t_{(i,j)} \\
\tilde{\mathbf{h}}^{t+1}_j &= \mathrm{GRU}([\mathrm{MSG}^t_j, \mathbf{x}^t_j], \tilde{\mathbf{h}}^t_j) \\
\boldsymbol{\mu}^{t+1}_j &= \mathbf{x}^t_j + f_{\mathrm{out}}(\tilde{\mathbf{h}}^{t+1}_j) \\
p(\mathbf{x}^{t+1}|\mathbf{x}^t, \mathbf{z}) &= \mathcal{N}(\boldsymbol{\mu}^{t+1}, \sigma^2 \mathbf{I})
\end{aligned}
$$

- Training: reconstruction error and KL term for a uniform prior

$$
-\sum_j \sum_{t=2}^{T} \frac{||\mathbf{x}^t_j - \boldsymbol{\mu}^t_j||^2}{2\sigma^2} \qquad \sum_{i \neq j} H(q_\phi(\mathbf{z}_{ij}|\mathbf{x}))
$$

Neural Relational Inference for Interacting Systems [ICML 2018]

Experiments 1: Physics simulations: objects connected by springs (have/not), charged particles (attract/repel), phase-coupled oscillators (have/not)



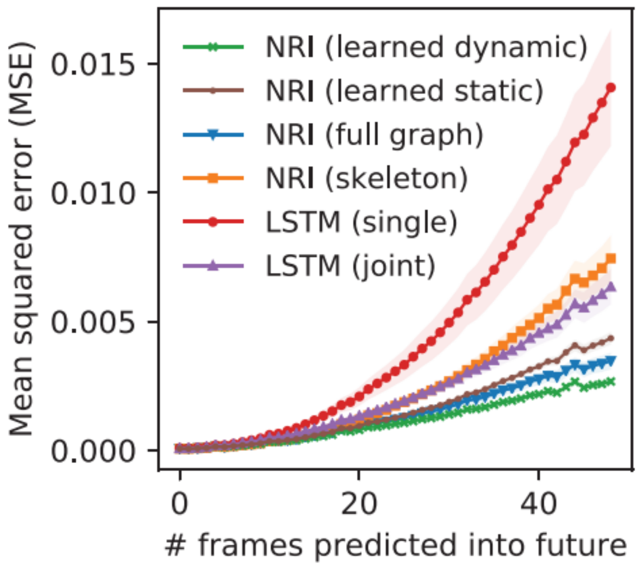
Model	Springs	Charged	Kuramoto
	5 objects		
Corr. (path)	52.4±0.0	55.8±0.0	62.8±0.0
Corr. (LSTM)	52.7±0.9	54.2±2.0	54.4±0.5
NRI (sim.)	**99.8**±0.0	59.6±0.8	–
NRI (learned)	**99.9**±0.0	**82.1**±0.6	**96.0**±0.1
Supervised	99.9±0.0	95.0±0.3	99.7±0.0
	10 objects		
Corr. (path)	50.4±0.0	51.4±0.0	59.3±0.0
Corr. (LSTM)	54.9±1.0	52.7±0.2	56.2±0.7
NRI (sim.)	**98.2**±0.0	53.7±0.8	–
NRI (learned)	**98.4**±0.0	**70.8**±0.4	**75.7**±0.3
Supervised	98.8±0.0	94.6±0.2	97.1±0.1

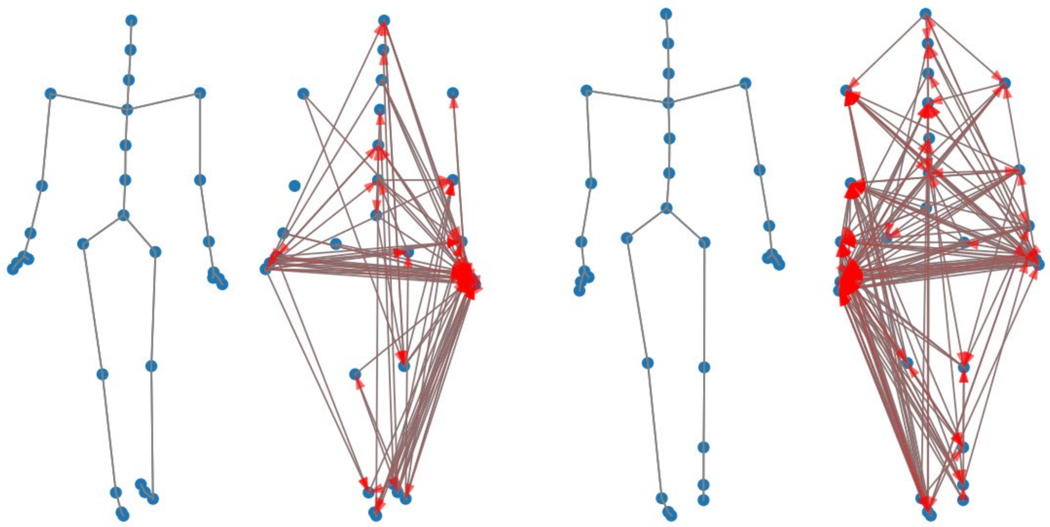
Neural Relational Inference for Interacting Systems [ICML 2018]

- Experiments 2: 3D trajectories of joints when walking

Observation: dynamic edge type helps

4-edge types, with non-edge prior=0.91





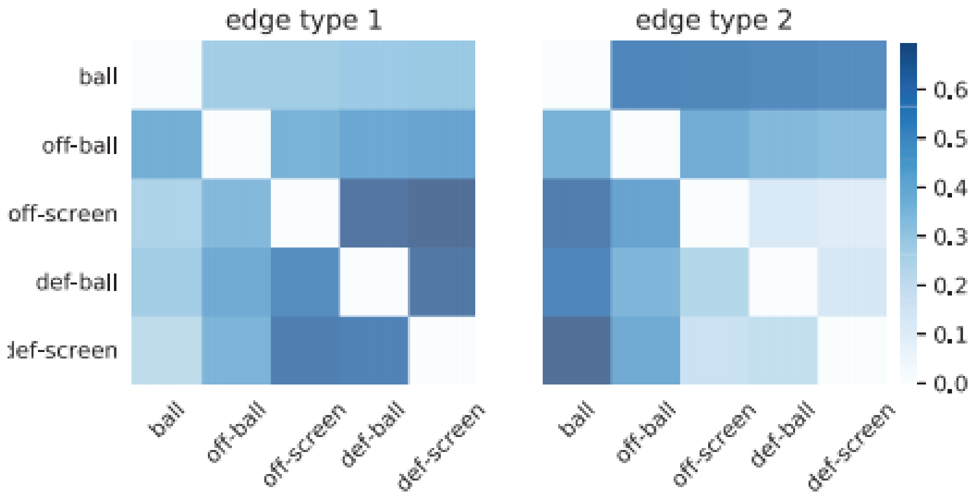
Neural Relational Inference for Interacting Systems [ICML 2018]

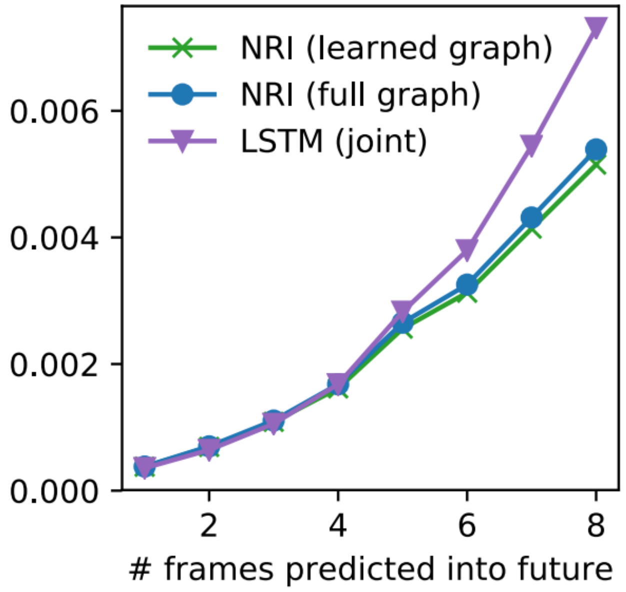
- Experiments

3. Pick and Roll NBA data

25 frames long (4s), five objects: ball, ball handler, screener, defensive matchup

2-edge types: ball, ball handler -> off-ball players;   among off-ball players





Factorised Neural Relational Inference for Multi-Interaction Systems [ICML 2019 Workshop]

- Multi-interaction at the same time -> only conflict interactions on the same layer

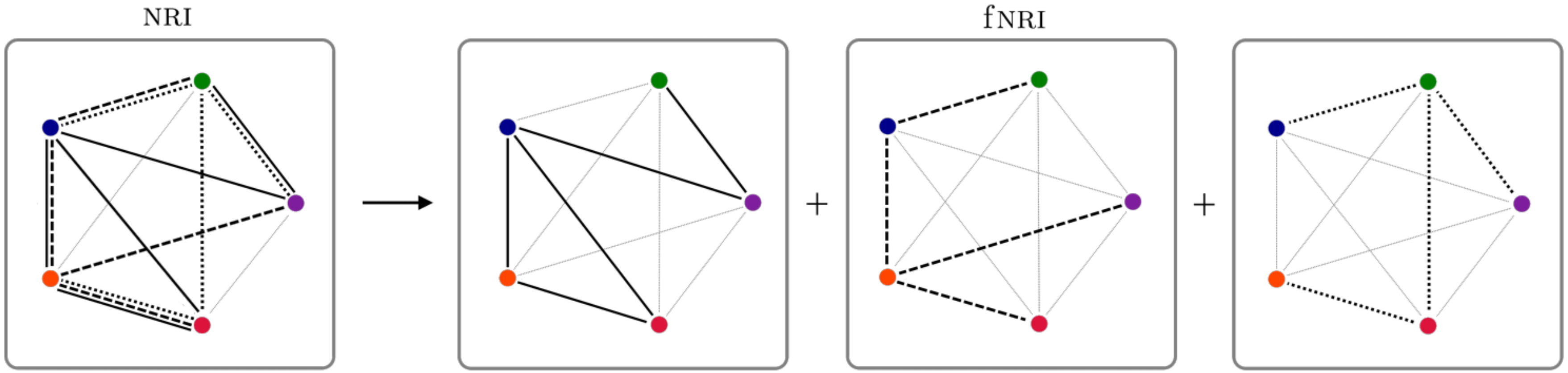
- Even only one output for each layer

Table 1. Accuracy (%) in recovering the ground truth interaction graph. Higher is better.

	I-Springs+Charges			I-Springs+Charges+F-springs			
Accuracy	**Combined**	I-Springs	Charges	**Combined**	I-Springs	Charges	F-Springs
Random	25.0	50.0	50.0	12.5	50.0	50.0	50.0
NRI (learned)	89.1 ± 0.4	97.9 ± 0.0	91.0 ± 0.4	57.9 ± 6.1	88.5 ± 0.9	87.3 ± 6.2	70.7 ± 2.3
fNRI (learned)	**94.0 ± 1.4**	98.0 ± 0.1	95.8 ± 1.3	**63.3 ± 6.5**	86.9 ± 2.7	97.7 ± 0.7	69.2 ± 5.5
sfNRI (learned)	88.8 ± 0.8	97.6 ± 0.1	91.1 ± 0.8	45.1 ± 5.1	90.0 ± 2.3	98.2 ± 0.8	52.4 ± 2.7
NRI (supervised)	**98.3 ± 0.0**	98.6 ± 0.0	99.7 ± 0.0	80.9 ± 0.7	92.4 ± 0.3	99.0 ± 0.1	84.4 ± 0.4
fNRI (supervised)	**98.3 ± 0.0**	98.8 ± 0.4	99.4 ± 0.4	**81.8 ± 0.1**	93.3 ± 0.1	99.3 ± 0.0	85.8 ± 0.1
sfNRI (supervised)	98.0 ± 0.0	98.3 ± 0.0	99.6 ± 0.0	81.0 ± 0.3	92.9 ± 0.1	99.2 ± 0.0	85.2 ± 0.2

Table 2. Mean squared error (MSE) / 10^{-5} in trajectory prediction. Lower is better.

	I-Springs+Charges			I-Springs+Charges+F-Springs		
Predictions Steps	1	10	20	1	10	20
Static	19.4	283	783	12.8	274	782
NRI (learned)	0.88 ± 0.06	4.05 ± 0.22	11.5 ± 0.5	0.95 ± 0.05	8.67 ± 0.45	29.1 ± 1.4
fNRI (learned)	**0.80 ± 0.04**	3.54 ± 0.09	9.93 ± 0.29	0.81 ± 0.05	7.78 ± 0.20	26.8 ± 0.8
sfNRI (learned)	1.03 ± 0.09	**3.32 ± 0.23**	**9.68 ± 0.74**	**0.77 ± 0.03**	**5.69 ± 0.21**	**19.3 ± 0.8**
NRI (true graph)	0.85 ± 0.04	1.59 ± 0.26	3.20 ± 0.15	0.75 ± 0.02	1.55 ± 0.07	3.43 ± 0.21
fNRI (true graph)	**0.70 ± 0.03**	**1.30 ± 0.06**	**2.52 ± 0.11**	**0.51 ± 0.05**	0.97 ± 0.08	2.44 ± 0.28
sfNRI (true graph)	0.86 ± 0.09	1.32 ± 0.06	2.77 ± 0.07	0.56 ± 0.04	**0.89 ± 0.06**	**2.28 ± 0.15**



About our model

- Supervised vs Semi-supervised vs Unsupervised
- Define a metric
- Temporal graph? Dynamic Edge Type?
- Define samples (which vehicles should be included)
- Features (Scale, Map information …)