Neural Relational Inference

Learning latent structure of interacting systems

Steven Morse
William & Mary, Data Science
DATA 691, Graph Learning

March 18, 2025

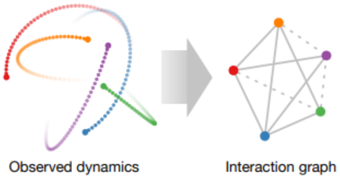
Overview

Idea

Can we learn the relational (graph) structure of a dynamical system and use that structure to predict future dynamics?

"Neural Relational Inference," Thomas Kipf and Ethan Fetaya, Kuan-Chieh Wang, Max Welling, Richard Zemel (Univ Amsterdam, Univ Toronto). ICML 2018. [Kipf et al., 2018]

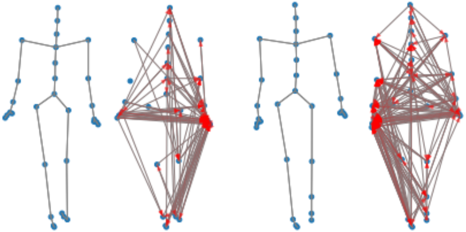
$F_{ij} = -k_{ij}(r_i - r_j) \longrightarrow$



$\longrightarrow e_{ij} \approx F_{ij}$

Introduction and Related work

1. How can we model and predict behavior in a dynamical system?
2. With structure (F_{ij}), we can use physics simulations.
3. Without structure we can regress directly on trajectories (e.g. LR, LSTM, ...)
4. Can we learn the latent structure to inform our predictions? Yes!

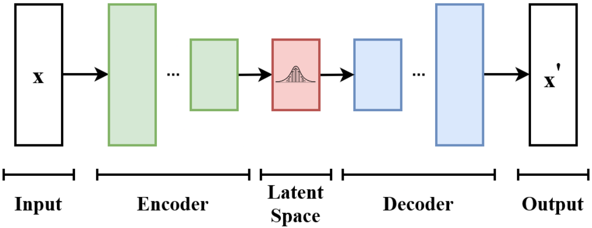


(a) Right hand focus (b) Left hand focus

What is an auto-encoder?

Idea

An encoder neural net compresses the input data into a lower dimensional latent space. The decoder uses this latent space as input to generate the full-dimensional output. This framework allows the model to learn hidden structure.



Uses: image denoising, machine translation, dimension reduction, generative models

Variational auto-encoder (VAE)

Idea

Apply a hierarchical Bayesian framework to autoencoders! Assume $p(x) = \int_z p(x, z)\, dz$. Then instead of $p(x) = \int_z p(x|z)p(z)\, dz$ we sample $\hat{z} \sim p(z|x)$ (encoder) and compute $p(x|z = \hat{z})$ (decoder).

Example

Gaussian mixture models are a shallow version of this. $z \sim$ Categorical, $p(x|z = k) = \mathcal{N}(\mu_k, \Sigma_k)$, then $p(x) = \sum_k z_k \mathcal{N}(\mu_k, \Sigma_k)$

- Two problems we need to solve:
  - How to approximate and learn $p(z|x) = p(x, z)/p(x)$? $\rightarrow$ ELBO
  - How to do backprop? $\rightarrow$ reparameterization trick

VAEs: the ELBO

- Finding the encoder $p(z|x)$ is hard. Instead, let's find a $q(z|x) \approx p(z|x)$.
- We notice that

$$\log p(x) = \log \sum_z q(z) \frac{p(x,z)}{q(z)} = \log \mathbb{E}_{z\sim q}\left[\frac{p(x,z)}{q(z)}\right] \geq \mathbb{E}_{z\sim q}\left[\log \frac{p(x,z)}{q(z)}\right] \stackrel{\text{def}}{=} \mathcal{L} \tag{1}$$

  where $\mathcal{L}$ is the evidence lower bound (ELBO). You can also show that

$$\log p(x) = \mathcal{L} + D_{\text{KL}}\big(q(z|x) \,||\, p(z|x)\big) \tag{2}$$

  where D_{KL} is the KL-divergence, a measure of the "distance" between distributions q and p.
- Since the KL-divergence is nonnegative, and $p(x)$ is fixed, this shows that maximizing the ELBO minimizes the KL divergence between our approximation $q(z|x)$ and the true $p(z|x)$.

VAEs: more ELBO

It also just so happens that we can decompose the ELBO into two terms:

Evidence Lower Bound (ELBO)

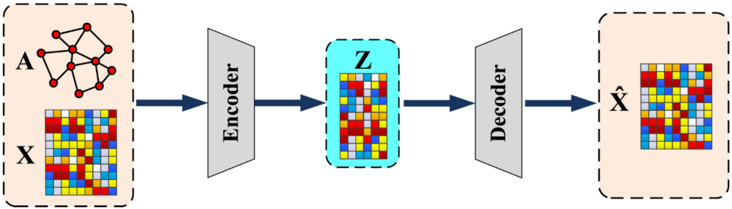
$$\mathcal{L}_{\theta,\phi}(x) \stackrel{\text{def}}{=} \mathbb{E}_{z\sim q_\phi}\left[\ln \frac{p_\theta(x,z)}{q_\phi(z|x)}\right] = \underbrace{\mathbb{E}_{z\sim q_\phi}\big[\ln p_\theta(x|z)\big]}_{\text{reconstruction loss}} - \underbrace{D_{\mathrm{KL}}(q_\phi(z|x) \parallel p_\theta(z))}_{\text{regularizer}}$$

(Later, we will see what these terms simplify to in the NRI model.)

(Variational) Graph auto-encoder (VGAE)

Idea

GAE is an autoencoder where the encoder and decoder are GNNs.



Neural Relational Inference

- Denote with $\mathbf{x}_i^t$ the feature vector of object v_i at time t (e.g. location and velocity).
- Assume we can capture dynamics with an unknown graph $\mathbf{z}$ with $\mathbf{z}_{ij}$ representing the discrete edge type between objects v_i and v_j.

Encoder

$$q_\phi(\mathbf{z}_{ij}|x) = \text{softmax}(f_{\text{enc},\phi}(\mathbf{x})_{ij})$$

Decoder

$$p(\mathbf{x}_j^{t+1}|\mathbf{x}^t, \mathbf{z}) = \mathcal{N}(f_{\text{dec}}(\mathbf{x}_j^t, \mathbf{z}), \sigma^2\mathbf{I})$$

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

$$\begin{aligned}
\tilde{\mathbf{h}}_{(ij)}^t &= \sum_k z_{ij,k} \tilde{f}_e^k([\mathbf{x}_i^t, \mathbf{x}_j^t]) \\
\boldsymbol{\mu}_j^{t+1} &= f_{\text{dec}}(\mathbf{x}_j^t, \mathbf{z}) \\
p(\mathbf{x}_j^{t+1}|\mathbf{x}^t, \mathbf{z}) &= \mathcal{N}(\boldsymbol{\mu}_j^{t+1}, \sigma^2\mathbf{I})
\end{aligned}$$

NRI: Training

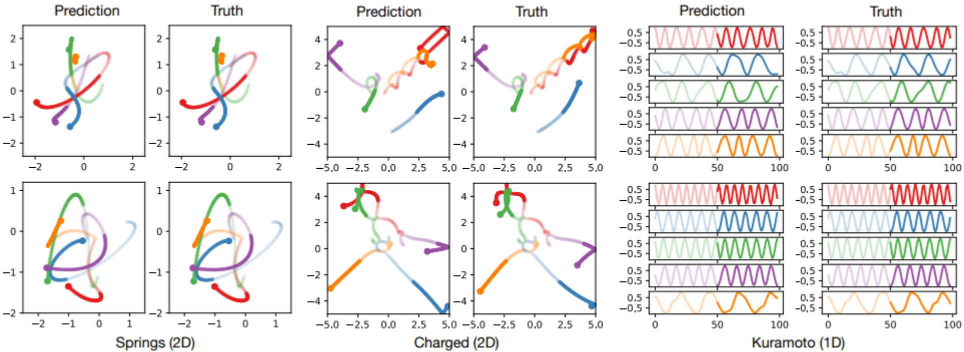
The ELBO in this specific model can be estimated by

$$\text{ELBO} = \underbrace{-\sum_j \sum_t \frac{||\mathbf{x}_j^t - \boldsymbol{\mu}_j^t||}{2\sigma^2}}_{\text{reconstruction loss}} + \underbrace{\sum_{i \neq j} H(q_\phi(\mathbf{z}_{ij}|\mathbf{x}))}_{\text{regularizer}} + \text{const} \tag{3}$$

and we do reparameterization by using a continuous approximation of the discrete $\mathbf{z}$:

$$\mathbf{z}_{ij} = \text{softmax}((\mathbf{h}^2_{(ij)} + \mathbf{g})/\tau) \tag{4}$$

with $\mathbf{g} \sim \text{Gumbel}(0, 1)$ (double-exponential)

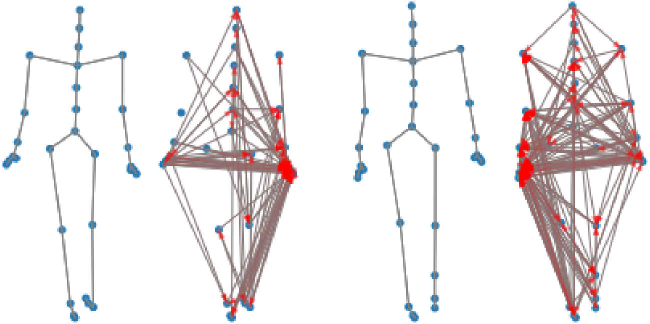


Springs
$F_{ij} = -k(r_i - r_j)$

Charged particles
$F_{ij} = C \cdot \text{sign}(q_i \cdot q_j) \frac{r_i - r_j}{||r_i - r_j||^3}$

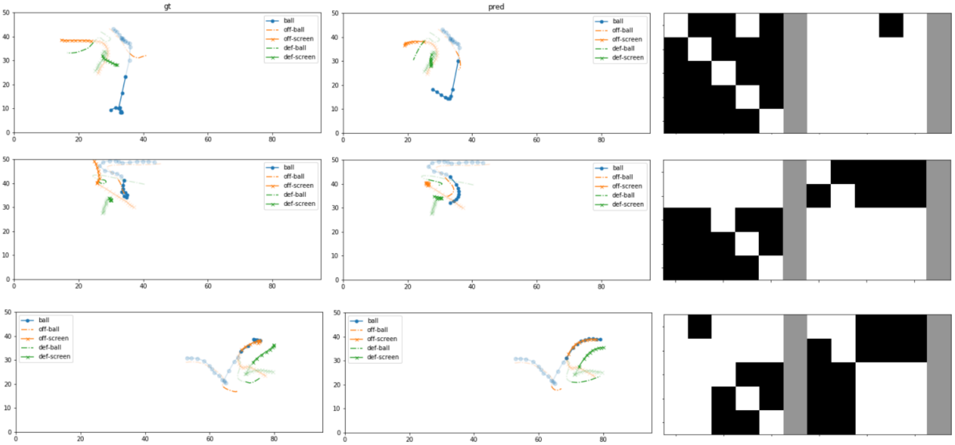
Kuramoto oscillators
$\frac{d\phi_i}{dt} = \omega_i + \sum_{j \neq i} k_{ij} \sin(\phi_i - \phi_j)$

Experiments - Motion capture

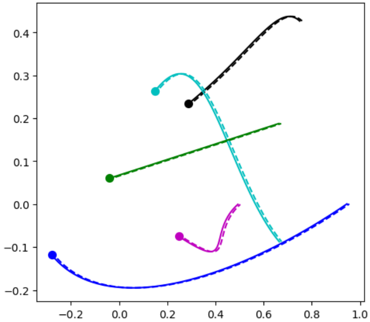


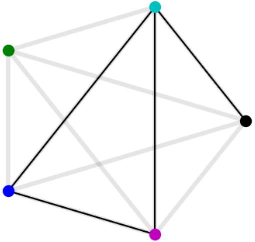
(a) Right hand focus (b) Left hand focus

Experiments - Basketball Pick and Roll



Reimplementation

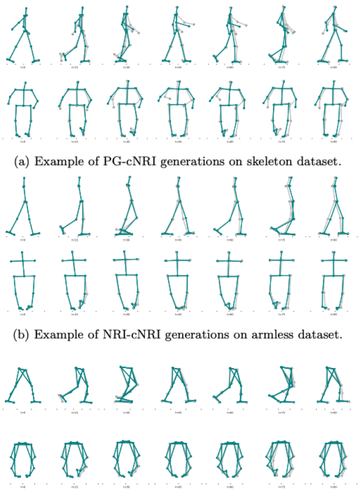




https://github.com/ethanfetaya/NRI/

Subsequent work

- Dynamic NRI ([Graber and Schwing, 2020]) — allow edge types to change over time
- Conditional NRI ([Candido Ramos et al., 2021]) — model multiple systems (repo)
- *Actional-Structural GCN ([Li et al., 2019]) — extend NRI with higher order relationships as structural links
- NRI for allosteric interactions in proteins ([Zhu et al., 2022]) — learn long-range molecular dynamics
- NRI with Node-specific Information ([Banijamali, 2021]) — incorporate additional node info as new nodes



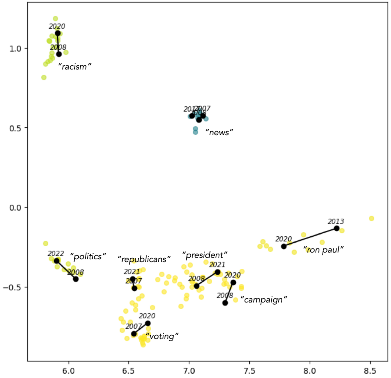
(a) Example of PG-cNRI generations on skeleton dataset.

(b) Example of NRI-cNRI generations on armless dataset.

(c) Example of NRI-cNRI generations on lower body dataset.

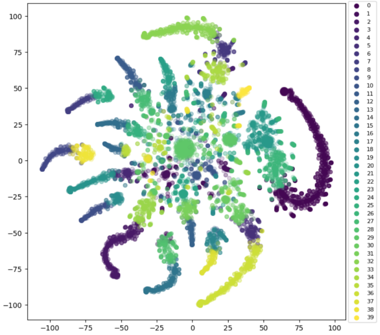
Application to topic dynamics?

1. Sentence embedding captures contextualized semantic similarity in a metric space
2. Clustering → topics, over time → persistent topics
3. (Ongoing work) Many topics exhibit significant drift. The embedding carries semantic meaning — so does the drift.
4. Can we effectively model topics as a dynamical system and predict their trajectories?



Application to social networks?

1. Consider the embedded output of a community forum, $(\mathbf{e}_i, u_i)$, topic clusters $\{\boldsymbol{\mu}_k\}$, and the mixture vector $\mathbf{x}_i^t$ of each user i over those topics (at time t)
2. Using $\mathbf{x}_i^t$ as a feature vector of (node) i, can we learn a latent edge structure $\mathbf{z}_{ij}$ that is predictive of $\mathbf{x}_i^{t+1}$?
3. What about with fixed t?



References I

Banijamali, E. (2021).
Neural relational inference with node-specific information.
In *International Conference on Learning Representations*.

Candido Ramos, J. A., Blondé, L., Armand, S., and Kalousis, A. (2021).
Conditional neural relational inference for interacting systems.
In *Machine Learning and Knowledge Discovery in Databases. Applied Data Science Track: European Conference, ECML PKDD 2021, Bilbao, Spain, September 13–17, 2021, Proceedings, Part V 21*, pages 182–197. Springer.

Graber, C. and Schwing, A. G. (2020).
Dynamic neural relational inference.
In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8513–8522.

References II

Kipf, T., Fetaya, E., Wang, K.-C., Welling, M., and Zemel, R. (2018).
Neural relational inference for interacting systems.
In *International conference on machine learning*, pages 2688–2697. Pmlr.

Li, M., Chen, S., Chen, X., Zhang, Y., Wang, Y., and Tian, Q. (2019).
Actional-structural graph convolutional networks for skeleton-based action recognition.
In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3595–3603.

Zhu, J., Wang, J., Han, W., and Xu, D. (2022).
Neural relational inference to learn long-range allosteric interactions in proteins from molecular dynamics simulations.
Nature communications, 13(1):1661.

Questions?