TrajectoryNet: A Dynamic Optimal Transport Network for Modeling Cellular Dynamics

Alexander Tong, Jessie Huang, Guy Wolf, David van Dijk, Smita Krishnaswamy

July 2020
Motivation

Longitudinal inference from cross sectional snapshot measurements
Motivation

Longitudinal inference from cross sectional measurements

Tasks:
• Predict trajectory of a point
Motivation

Longitudinal inference from cross sectional measurements

Tasks:
• Predict trajectory of a point
• Predict distribution at test timepoint
Normalizing Flows (NFs)

- Begin with a simple distribution
  \[ p_{t_0}(x) \sim \mathcal{N}(0, 1) \]
- Apply an invertible transformation(s)
  \[ x_{t_1} = f(x_{t_0}) \]
- Use change of variables to calculate probability
  \[ \log p_{t_1}(x_{t_1}) = \log p_{t_0}(x_{t_0}) - \log \det \left| \frac{\partial f}{\partial x_{t_0}} \right| \]
Deep Normalizing Flows (NFs)

• Apply a series of transformations

$$x_{t_1} = f(x_{t_0}) \quad \Rightarrow \quad x_{t_N} = f_N \circ f_{N-1} \circ \cdots \circ f_1(x_{t_0})$$

• Use change of variables to calculate probability

$$\log p_{t_1}(x_{t_1}) = \log p_{t_0}(x_{t_0}) - \log \det \left| \frac{\partial f}{\partial x_{t_0}} \right| \quad \Rightarrow \quad \log p_{t_N}(x_{t_N}) = \log p_{t_0}(x_{t_0}) - \sum_{n=1}^{N} \log \det \left| \frac{\partial f_n}{\partial x_{t_{n-1}}} \right|$$
Continuous Normalizing Flows

\[ x_{t_N} = f_N \circ f_{N-1} \circ \cdots \circ f_1(x_{t_0}) \quad \Rightarrow \quad x_{t_1} = F(x_{t_0}) = \int_{t_0}^{t_1} f(x(t), t) \, dt \]

\[ \log p_{t_N}(x_{t_N}) = \log p_{t_0}(x_{t_0}) - \sum_{n=1}^{N} \log \det \left| \frac{\partial f_n}{\partial x_{t_n-1}} \right| \quad \Rightarrow \quad \log p_{t_1}(x_{t_1}) = \log p_{t_0}(x_{t_0}) - \int_{t_0}^{t_1} \text{Tr} \left( \frac{\partial f}{\partial x(t)} \right) \, dt \]

Cannot model:

\[ F(x) = -x \]

[Chen et al. 2018]
CNFs create continuous paths

Creates continuous paths, but they may not be biologically plausible — no restriction on circuitous paths!

[Chen et al. 2018]
Obtaining straight paths via regularization

Penalize path energy: the squared L2-norm of the derivatives
Regularized CNF approximates dynamic optimal transport

Dynamic OT:

\[ W(\mu, \nu)_2^2 = \inf_{p,f} \int_{\mathbb{R}^d} \int_0^1 p(x, t) \| f(x, t) \| \, dt \, dx = \inf_{p,f} \mathbb{E} \int_0^1 \| f(x(t), t) \| \, dt \]

Subject to:

\[ \partial_t p + \nabla \cdot (pf) = 0 \]

Mass is preserved

\[ p(\cdot, 0) = \mu, \quad p(\cdot, 1) = \nu \]

Benamou and Brenier 2000
Regularized CNF approximates dynamic optimal transport

Dynamic OT: regularized CNF

\[ W(\mu, \nu)^2_2 = \inf_{p,f} \int_{\mathbb{R}^d} \int_0^1 p(x, t) \| f(x, t) \| \, dt \, dx = \inf_{p,f} \mathbb{E} \int_0^1 \| f(x(t), t) \| \, dt \]

Subject to:

\[ \partial_t p + \nabla \cdot (pf) = 0 \]

Mass is preserved

\[ p(\cdot, 0) = \mu \]

\[ p(\cdot, 1) = \nu \]
CNFs model Dynamic Optimal Transport

a) Source

b) Target

c) Dynamic OT
d) CNF
e) CNF + Energy reg.
Dynamic OT via TrajectoryNet

• Dynamic OT via TrajectoryNet can be utilized to infer continuous trajectories of any populations adhering to energy or transport constraints
  • Population migration
  • Disease spread

• However, cellular systems are more constrained, and other domain specific priors apply
Single Cell Embryonic Stem Cell Data

27 day timecourse collected at 5 timepoints, measurements destroy cells at each timepoint (same cell cannot be measured at more than one timepoint)
Inferring Continuous Flow in Static Snapshots
Additional Properties of cells

1. Cells are not simply transported from one timepoint to another, *they cells divide and die.*
2. Cells cannot travel in straight paths through Euclidean space in terms of measured dimensions, *cells only travel along a cellular manifold.*
3. Though cells are destroyed when measured, we can estimate their direction of transition-based *RNA velocity*
Cell Death and Growth

• Allowing unbalanced transport can let cells “die” instead of moving them to implausible locations

• Unbalanced transport hard to achieve dynamically

• We use discrete optimal transport to assign growth and death rates

Liero et al. 2018
Cells have to transition through allowable parts of the state space.

Enforce this with a density penalty. Based on a knn density estimate:

\[ L_{\text{density}}(x, t) = \sum_k \max(0, \min-k(\{\|x(t) - z\| : z \in \mathcal{X}\}) - h) \]
Velocity Regularization

RNA Velocity, estimate of direction of change
[La Manno et al. 2018 Velocyto; Volker et al. ScVelo]

\[(x, t, \overrightarrow{dx/dt}) = \frac{f(x, t) \cdot \overrightarrow{dx/dt}}{\|f(x, t)\| \|\overrightarrow{dx/dt}\|}\]
Toy Example
Continuous Trajectories in Single Cell Data

Density

Single Cell Trajectories
Results — Embryoid body dataset

<table>
<thead>
<tr>
<th></th>
<th>t=1</th>
<th>t=2</th>
<th>t=3</th>
<th>mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base</td>
<td>0.764</td>
<td>0.811</td>
<td>0.863</td>
<td>0.813</td>
</tr>
<tr>
<td>Base + D</td>
<td>0.759</td>
<td><strong>0.783</strong></td>
<td>0.811</td>
<td><strong>0.784</strong></td>
</tr>
<tr>
<td>Base + V</td>
<td>0.816</td>
<td>0.839</td>
<td>0.865</td>
<td>0.840</td>
</tr>
<tr>
<td>Base + D + V</td>
<td>0.930</td>
<td>0.806</td>
<td><strong>0.810</strong></td>
<td>0.848</td>
</tr>
<tr>
<td>Base + E</td>
<td>0.737</td>
<td>0.896</td>
<td>0.842</td>
<td>0.825</td>
</tr>
<tr>
<td>Base + G</td>
<td><strong>0.700</strong></td>
<td>0.913</td>
<td>0.829</td>
<td>0.814</td>
</tr>
</tbody>
</table>

• Wasserstein distance between predicted and true distributions for different left out timepoints
• Different regularizations have different assumptions and tradeoffs
Tracing Ancestry
Summary

- Energy regularized CNF performs dynamic optimal transport to find flows between cross-sectional populations.

- TrajectoryNet includes additional regularizations that allow for optimal transport on a manifold, with growth and death of individuals over time, and respecting individual velocity data.

- Trajectories of individual cells, and gene expression activity can be inferred.
Thanks!

Code: https://github.com/krishnaswamylab/TrajectoryNet
Lab Website: https://www.krishnaswamylab.org