Learning Semantic Relationships of Geographical Areas Based on Trajectories

Presenter: Saim Mehmood
trajectories \((x, y, t)\)

(spatiotemporal information of moving objects)
Trajectory Data Mining
discovering patterns in trajectories to inform critical real-world applications
Trajectory Data Mining Tasks

- Trajectory similarity
- Trajectory clustering
- Trajectory anomaly detection
- Trajectory classification
- Trajectory prediction
- etc…
Trajectory Applications

- human mobility understanding
- healthcare (detecting change in gait pattern of seniors)

location-based services (e.g., recommendation of points-of-interest)
Research Questions
Research Question I

How people perceive different areas of their city?
Research Question II

To what extent people rely on geographical proximity of areas?
Research Question III

How the behavior of people compare in different geographical space?

New York

City of Porto
Overview

Method 1
Learning Semantic Relationships of Geographical Areas

Method 2
Statistical Method for Distinguishing Geographical Proximity to Semantic Proximity
Learning Semantic Relationships of Geographical Areas
How can we learn latent semantic relationships between geographical areas using trajectories?
Semantic Proximity

Geographical Proximity
Construction of a Uniform Grid
How I Convert Trajectory Into Grid Cells?

trajectory \((x_i, y_i, t_i)\)

trajectory \((x_j, y_j, t_j)\)

trajectory \((x_k, y_k, t_k)\)
Our Approach

learn relationships using
network representation learning (NRL)
Network Representation Learning (NRL)
Network Representation Learning (NRL)

several network structural properties can be learned/embedded
(nodes, edges, subgraphs, graphs, …)
Random Walk-based NRL

Input graph

Obtain a set of random walks

Learn a vector embedding for each node

Feed sentences to Skip-gram NN model

Treat the set of random walks as sentences

Obtain a set of random walks

Learn a vector embedding for each node

Feed sentences to Skip-gram NN model

Treat the set of random walks as sentences
NRL in our Approach
Construction of a lattice graph

Grid Cells

lattice graph
Trajectory as walks

lattice graph
Trajectory Permutations

nodes appearing in the same context window are more similar

for trajectories, every node should be in the context of every other node

Skip-gram (context window)

single walk

shuffling m-times

m-walks

feed walks to skip-gram NN model
Method 1 Overview
Statistical Method for Distinguishing Geographical Proximity to Semantic Proximity
Real vs Null Hypothesis
Real Model

Real model is based on real trajectory movements over lattice graph
Null Model

Null Model is based on random walks but satisfies the size constraint
Alternate Null (Intermediate) Model

Intermediate model is like Null model but satisfies the constraint for each walk starting from the same node \( u \) as Real model walks.
Model Comparative Analysis

how can we compare the real vs the null model?

metrics for both quantitative and visual comparison
Quantitative: Cosine Similarity

\[
\begin{align*}
\vec{v}_i (128D) & \quad \vec{v}_j (128D) & \quad \vec{v}_k (128D) \\
\end{align*}
\]

\[
\cos(\vec{v}_i, \vec{v}_j) \geq \lambda_a \quad \text{“similar”}
\]

\[
\cos(\vec{v}_i, \vec{v}_k) < \lambda_a \quad \text{“not similar”}
\]
Quantitative: Interesting Pairs of Nodes

Let’s say we have two models \((X \text{ and } Y)\)

\[
\cos_\theta_X (\vec{v}_i, \vec{v}_j) \quad \cos_\theta_Y (\vec{v}_i, \vec{v}_j)
\]

\[
d_{X,Y} (\vec{v}_i, \vec{v}_j) = |\cos_\theta_X (\vec{v}_i, \vec{v}_j) - \cos_\theta_Y (\vec{v}_i, \vec{v}_j)|
\]

\[
d_{X,Y} (\vec{v}_i, \vec{v}_j) \geq \lambda_b \quad \text{“similar”}
\]
Comparing Distributions of Models

Let’s say we have two Histograms ($H^A$ and $H^B$)

\[ \chi^2 = d(H^A, H^B) = \sum_{i=1}^{b} \frac{(H_i^A - H_i^B)^2}{H_i^A} \]

Where $b$ is the number of bins
Exploratory Analysis of Models

A many-to-many visualization

One-to-many visualization
Evaluation
Case Study I: New York City (NYC)
Exploratory Analysis: Many-to-Many

real

null

intermediate
Exploratory Analysis: One-to-Many
Quantitative: Cosine Similarity

The graph illustrates the cosine similarity as a function of rank for different categories:

- **real** (solid blue line)
- **intermediate** (dashed red line)
- **null** (dotted orange line)

The x-axis represents the rank (of a pair of nodes) multiplied by $10^5$. The y-axis shows the cosine similarity.
Quantitative: Interesting Pairs of Nodes

\[
\text{cosine similarity difference} = \begin{cases} 
0 & \text{for rank } 0 \\
\frac{1}{\text{rank}} & \text{for rank } 1, 2, 3, \ldots
\end{cases}
\]

- \text{real vs null}
- \text{real vs intermediate}
Distribution of Pair-wise Similarities

no of pairs of nodes

real

null

intermediate

cosine similarity
Case Study II: City of Porto
Exploratory Analysis: Many-to-Many

real

null

intermediate
Exploratory Analysis: One-to-Many
Quantitative: Cosine Similarity
Quantitative: Interesting Pairs of Nodes

![Graph showing cosine similarity difference versus rank (of a pair of nodes)]

- **Real vs Null**
- **Real vs Intermediate**
Distribution of Pair-wise Similarities

- **real**
- **null**
- **intermediate**

no of pairs of nodes vs. cosine similarity
Research Questions

How the behavior of people compare in different geographical space?
Chi-Square

\[ \chi^2 = d(H^A, H^B) = \sum_{i=1}^{b} \frac{(H_i^A - H_i^B)^2}{H_i^A} \]

City of New York

real distance from null: \( \chi^2 = 4.0854e + 05 \gg 0 \)
real distance from intermediate: \( \chi^2 = 3.0426e + 05 \gg 0 \)

City of Porto

real distance from null: \( \chi^2 = 6.1697e + 05 \gg 0 \)
real distance from intermediate: \( \chi^2 = 7.8492e + 05 \gg 0 \)
Summary
Summary of Contributions

learned nodes embeddings for real and null models

performed statistical analysis to distinguish geographical to semantic proximity

Learning Semantic Relationships of Geographical Areas based on Trajectories
Saim Mehmood and Manos Papagelis
IEEE Mobile Data Management 2020
References


Thank you!