

1 Introduction

In the urban space, explain the data variety, volume, and velocity. They pose challenges and opportunities in addressing real urban computing problem.

Challenges:

- Data sparse and noisy
- Heterogeneous data type
- Data is huge

Problems with existing methods:

- model of interaction is ad-hoc
- the spatial non-stationarity

My research focus is to model complicated interactions in the urban space. It could be interactions of regions, people, and features. And I want to propose a unified graphical model to do this.

1.1 Research questions include:

- How does the crime (or any other feature) in one region correlate and influence that variable of other regions.
- What influence crime in focal region, the crime in its neighbor, or the demographics in its neighbor.
- Measure the similarity of two person, given their observations.
- How would the traffic change, if build a new transportation center at location X.

Understanding the interactions can help us solve three fundamental problems in urban space.

- Inference (single source from others) problem. [Prediction]
- Structural learning. Does two variable correlated? How?
- Partition regions (address misalignment issue). community detection based on high-dimension similarity.

1.2 Why use graphical model?

A graphical model support three kinds of learning problem:

1. Infer unobserved values.
2. Learn parameters. Given graph, learn dependency.
3. Structural learning. Given data, the graph is too complicated to build. Learn the graph.

Therefore, a graphical model is a unified approach.

2 Inference Problem

2.1 Take crime inference as example

Spatial autoregressive model

Enhance with 1) newer type of data, and 2) Negative binomial regression.

Issues with current model

- Spatial non-stationarity \rightarrow adaptive model
- Simplified assumption \rightarrow graphical model

In the literature, GWR solves the non-stationarity problem.

What is the weakness of GWR.

There are multiple network adjacency structure. How to learn the importance of each network? Namely, spatial nearby neighbors show higher influence to its neighbors than that of taxi flow connected neighbors.

3 Structural Learning Problem

The graphical model on crime network.

Model the assumptions of crime inference **What influence the crime count in focal area? Crime in neighbors or demographics in neighbors.**