The Nonparametric Metadata Dependent Relational Model

Dae Il Kim; Michael C.Hughes; Erik B.Sudderth

Brown University, Department of Computer Science

February 1, 2013
Presented by Zhengming Xing
Outline

- Introduction
- Nonparametric Relational Modeling
  - Model Formulation
  - Generating Process
- Learning via MCMC
- Experiment Results
  - Synthetic Data
  - Sampson Monastery
  - Otago Harbour Food Web
  - Lazega Lawyers Network
Introduction

Main idea: Bayesian nonparametric block model for network data.

\( N \) nodes, a feature vector (metadata) is available for each node. To learn \( K \) latent communities, which are block of nodes with similar behavior (e.g. predator with common prey, individual with common interests), and missing edges.
Extend the previous Block model in two ways:

- Each node has mixed membership in an unbounded set of communities.
- Node-specific metadata directly influences the generation of latent community distribution.
Nonparametric Relation Modeling

- For $N$ nodes, let $\phi_i \in \mathbb{R}^F$ denotes a feature vector and captures the metadata associated with node $i$ and $\phi \in \mathbb{R}^{F \times N}$ a matrix of corpus metadata.
- For every community $k$, we let $\eta_{fk} \in \mathbb{R}$ denotes an associated significance weight for feature $f$ in community $k$, and $\eta_k \in \mathbb{R}^F$ a vector of these weights.
  \[
  \eta_k \sim \mathcal{N}(\mu, \Lambda^{-1})
  \]
- Given $\eta$ and $\phi$, the node-specific score for community $k$ is sampled as $\nu_{ki} \sim \mathcal{N}(\eta_k^T \phi_i, \lambda_V^{-1})$
Logistic stick breaking process is employed to model the community distribution. Let $\pi_{ki}$ denote the probability that node $i$ chooses community $k$, where $\sum_{k=1}^{\infty} \pi_{ki} = 1$.

$$\pi_{ki} = \psi(v_{ki}) \prod_{l=1}^{k-1} \psi(-v_{li})$$

$$\psi(v_{ki}) = \frac{1}{1+\exp(-v_{ki})}$$

where $\psi(v_{ki})$ is the logistic function.
Nonparametric Relation Modeling

- For node $i$ and $j$, given $\pi_i$ and $\pi_j$, we sample a pair of community indicator variables $s_{ij} \sim \text{mult}(\pi_i)$ and $r_{ij} \sim \text{mult}(\pi_j)$ for each directed interaction $y_{ij}$.
- Binary edge $y_{ij} \sim \text{Ber}(s_{ij}^T W r_{ij})$, $W_{kl} \sim \text{Beta}(\gamma_a, \gamma_b)$.
- Above model formulation can be extend to $M$ relations.
Generating Process

1. Sample global parameters:
   (a) Draw $\lambda_S, \lambda_F, \lambda_V$ from gamma priors
   (b) Draw $\mu \sim N(0, \lambda_S^{-1} I_F)$
   (c) For each community $k$, $\eta_{ik} \sim N(\mu, \lambda_F^{-1} I_F)$

2. For each node $i = 1, 2, \ldots, N$:
   (a) Draw $v_{ki} \sim N(\eta^T \phi_{ki}, \lambda_V^{-1} I)$
   (b) Let $\pi_{ki} = \psi(v_{ki}) \prod_{\ell=1}^{k-1} \psi(-v_{\ell i})$

3. For each relation $m = 1, 2, \ldots, M$:
   (a) For each community pair $k, \ell$:
      i. Draw $W_{k\ell m} \sim Beta(\gamma_a, \gamma_b)$
   (a) For each potential edge $\{i, j\} \in N \times N$:
      i. Draw $s_{ijm} \sim Mult(\pi_{i})$
      ii. Draw $r_{ijm} \sim Mult(\pi_{j})$
      iii. Draw $y_{ijm} \sim Ber(s_{ijm}^T W_{::m} r_{ijm})$
Learning via MCMC

- Implement a dynamic truncation technique based on retrospective sampling of latent community assignments $r$ and $s$.
- Metropolis-Hastings sampling is employed to sample $v_{i}$ which is non-conjugate because of logistic transformation.
Experiment Results

Synthetic Data: Two toy datasets and each with 80 nodes. First one has 5 blocks and the second noiser one has 4 blocks.
Experiment Results

Sampson Monastery Data: 12 months period, 18 novice monks and 8 relations about their peer were recorded. Sampson’s observation (considered as ground truth): Four groups which are "Young Turks" (Gregory and John Bosco), "Loyal Opposition" (Peter), "Outcast" and "Interstitials". Incorporate metadata which includes each monk’s arrival and departure time, each monk’s rank, binary judgement on sociability and maturity.
Experiment Results

Otago Harbour dataset: Single "who eats whom" binary relationship for 123 organisms. Metadata classifies each node as one of 21 possible organism types (e.g. annelids, birds), and assigns one of three mobility rating (low, intermediate, high).
Experiment Results

Lazega lawyers dataset: Social network between partners and associates of several law firms. Contains three directed binary relations encoding friendship, coworker and advisory among 71 lawyers. Metadata includes status, gender, office location, years employed, age, practice and law school.

Variational distance $D_{ij} = 1/2 \sum_{K=1}^{K} |\pi_{ki} - \pi_{kj}|$ between learned community membership distributions to measure node similarity.
Experiment Results

(c) AUC single

(d) AUC multiple