Infinite Latent Attribute Model

Proposal: allow an explicit representation of the partitioning of each general feature into subclusters.

ILA model: Links are generated as follows:

- every entity is assigned a binary vector $z_i$ indicating which features it has active. Draw the $N \times M$ latent feature matrix $Z$ from the Indian Buffet Process.
- all the members of the $m$th feature, are assigned to $K(m)$ subclusters, with each entity belonging to a single subcluster in that feature. $c(m)$ is a vector of length $N$ and $c_{im}$ denotes the subcluster of the $i$th entity belongs to in the $m$th feature.
- $c(m) | y \sim CRP(\gamma)$
- Each feature $m$ has a real-valued $K(m) \times K(m)$ weight matrix $W(m)$. $w_{kk}^{(m)} \equiv W^{(k)}(k, k')$ is the weight that affects the probability of there being a link from entity $i$ to entity $j$, given that entity $i$ belongs to subcluster $k$ and entity $j$ belongs to subcluster $k'$ of the $m$th feature.
- to generate a link $Y(i, j)$ $\in \{0, 1\}$, from entity $i$ to entity $j$, draw

$$Y_{ij} \mid Z, C, W \sim Bernoulli(\sigma)$$

where $\sigma$ is a bias parameter. Only classes that are on for both entities influence the probability of a link between them.

Inference

Sampling $Z$:

- for $m \leq M$: Bernoulli sample. Integrate over $c(m)$, including case of new subcluster. $P(w)$ non-conjugate to likelihood so use auxiliary variable approach [5] (Algorithm 8)
- for $m > M$: sample the number of new features and the associated weights. Due to non-conjugacy, use Metropolis-Hastings.

Sampling $W$:

- Non-conjugate so use Metropolis-Hastings or slice sampling.

Latent Class & Latent Feature Models

Latent Class models assume a number of clusters $K$ and each entity belongs to a single cluster. The link probability between two entities depends only on their cluster assignments. The Infinite Relational Model (IRM) [3] belongs to this category. In the IRM:

- The cluster assignments $c_i = k, k \in \{1, 2, \ldots, K\}$ are drawn from the Chinese Restaurant Process (CRP).
- A $K \times K$ weight matrix $W$ contains the link probability between each pair of clusters.
- To generate a link $Y_{ij} \in \{0, 1\}$, draw $Y_{ij} \sim Bernoulli(W(c_i, c_j))$

Latent Feature models relate each entity with a vector of $M$ features and determine the link probability based on feature interactions. Such a model is the Nonparametric Latent Feature Relational Model (NLFRM) [4]:

- each entity $i$ is assigned a binary feature vector $z_i$. The $N \times M$ latent feature matrix, $Z$ is drawn from the Indian Buffet Process.
- a $M \times M$ weight matrix $W$ contains the real valued weights between each pair of features.
- to generate a link $Y_{ij} \in \{0, 1\}$, draw $Y_{ij} \sim Bernoulli(\sigma z_i^T W z_j)$

Motivation

An example: a friendship network at a collegiate University. A person might belong to more than one cluster e.g. a college, a department and a sport team. A latent class model would need a new cluster for each combination of the types of cluster, e.g. ‘Gryffindor College, Department of Mathematics, Football’. A latent feature model uses the feature vector representation to implicitly account for the possible combination of clusters.

Motivation: existing models only account for a flat clustering. The ‘college’ feature might be divided into subclusters, e.g. ‘Slytherin College’, ‘Gryffindor College’ etc. A latent feature model must represent each cluster with a new feature.

Results

Synthetic data

NIPS Coauthorship network

We used the NIPS 1-17 coauthorship dataset [1]. We kept only the 234 most connected authors, ran 10 repeats, holding out 20% of the data.

Gene Interactions network

We used a subset of the interaction data by [2]. We used 156 genes.


c| True links | IRM | LFRM | ILA | ILA

Test error (0-1 loss) | 0.0140 ± 0.0014 | 0.0129 ± 0.0011 | 0.0141 ± 0.0012 | 0.0116 ± 0.0007

Test log likelihood | -0.0593 ± 0.0043 | -0.0547 ± 0.0079 | -0.0322 ± 0.0058 | -0.0318 ± 0.0094

AUC | 0.9565 ± 0.0037 | 0.9631 ± 0.0150 | 0.9908 ± 0.0048 | 0.9910 ± 0.0056

Inference

- ILA is able to capture the complex nature of real world networks, with corresponding gains in empirical performance.
- ILA could be made even more flexible by allowing multiple membership of subclusters within a feature, corresponding to a nested IBP.

References