Latent Feature Models for Link Prediction

Kurt T. Miller
EECS
University of California
Berkeley, CA 94720
tadayuki@cs.berkeley.edu

Thomas L. Griffiths
Psychology
University of California
Berkeley, CA 94720
tom_griffiths@berkeley.edu

Michael I. Jordan
EECS
University of California
Berkeley, CA 94720
jordan@cs.berkeley.edu

Abstract

In recent years, several Bayesian models have been developed for link prediction in relational data. Models such as the Infinite Relational Model (IRM) [3] and the Mixed-Membership Stochastic Blockmodel (MMSB) [1] assume that there exists a set of latent classes that each object we observe can belong to and that each object either belongs to a single class or has a distribution over the classes. Conditioned on the latent class membership of each node, all relations are assumed to be generated independently.

We propose a model in which instead of there being a latent set of classes that objects can belong to, there are a set of latent features that each object can have. For example, in social networks, these features might be clubs, activities, or even geographic attributes that affect the probability of two people knowing each other. In the IRM, each person can only belong to a single class, so in order to fully explain the data, the IRM would need to create a different class for many of the possible combinations of features. In the MMSB, since people now have distributions over classes, there is more flexibility in how it defines the classes, but the more a person belongs to one class, the less they can belong to others. With our latent feature representation, each pair of features contributes a certain probability of there being a link. The more features possessed by two objects that indicate the presence of a link between then, the higher the probability of there being a link. This is an attribute the neither the IRM nor the MMSB possesses.

In our model, we assume that we observe a set of binary relations X_{ij} where $X_{ij} = 1$ indicates there is a link from i to j and $X_{ij} = 0$ if there is no link. We also assume that each object i has a set of K binary latent features stored in a vector z_i. For each pair of features k and k', there is a weight $w_{kk'}$ that affects the likelihood of there being a link between two objects when they have features k and k', respectively. These weights are stored in a matrix W. Given z_i, z_j, and W, the probability of there being a link from object i to object j is

$$p(X_{ij} = 1 | z_i, z_j, W) \quad = \quad \sigma(z_i^\top W z_j)$$

where $\sigma(\cdot)$ maps real numbers to the interval $[0, 1]$ such as the sigmoid or probit function. If $\sigma(\cdot)$ is the sigmoid function, this is similar in flavor to the exponential random graph model [4].

We place a nonparametric prior known as the Indian Buffet Process (IBP) [2] on the latent feature matrix Z and a Gaussian prior on W. By using the IBP as a prior on Z, we are able to simultaneously infer both the number of features K as well as which objects have them. We perform inference via Markov Chain Monte Carlo and demonstrate our method on synthetic data as well as the Alyawarra kinship data used in [3] which itself is believed to come from a factorial representation.

Topic: graphical models

Preference: oral/poster

References

[1] E. M. Airodi, D. M. Blei, E. P. Xing, and S. E. Fienberg. Mixed membership stochastic block models for relational data, with applications to protein-protein interactions. In *Proceedings of International Biometric Society-ENAR Annual Meetings*, 2006.

[2] T. L. Griffiths and Z. Ghahramani. Infinite latent feature models and the Indian Buffet Process. In *Advances in Neural Information Processing Systems (NIPS)*, 2006.

[3] C. Kemp, J. B. Tenenbaum, T. L. Griffiths, T. Yamada, and N. Ueda. Learning systems of concepts with an infinite relational model, 2006.

[4] G. L. Robins, T. A. B. Snijders, P. Wang, M. Handcock, and P. Pattison. Recent developments in exponential random graph (p*) models for social networks. *Social Networks*, 29:192–215, 2007.