Gaussian Process Models for Colored Graphs

Zhao Xu, Kristian Kersting
Fraunhofer IAIS
1stname.2ndname@iais.fraunhofer.de

Volker Tresp
Siemens Corporate Technology
volker.tresp@siemens.com

Abstract

Many real-world domains can naturally be represented as complex graphs, i.e., in terms of entities (nodes) and relations (edges) among them. In domains with multiple relations, represented as colored graphs, we may improve the quality of a model by exploiting the correlations among relations of different types. To this end, we develop a multi-relational Gaussian process (MRGP) model. The MRGP model introduces multiple GPs for each type of entities. Each random variable drawn from a GP represents profile/preference of an entity in some aspect, which is the function value of entity features at the aspect. These GPs are then coupled together via relations between entities. The MRGP model can be used for relation prediction and (semi-) supervised learning. We give an analysis of the MRGP model for bipartite, directed and undirected univariate relations.

1 Introduction

The analysis of complex, relational data is of growing interest within the machine learning community. Example domains are social networks, the world wide web, transportation networks, and gene interaction networks, among others. Gaussian processes (GP) [6] can be used for modeling and analyzing such structured data. Assume that we are interested in classifying persons based on their friendship relations (i.e. a social network). To do so, we introduce each person a latent variable drawn from a GP, which covariance matrix encodes the correlations among persons, and can be computed on the friendship graph using e.g. spectral graph methods [13]. Then we predict class labels of persons conditioned on their latent variables in a GP classification framework.

A common assumption of existing relational GP models, however, is that entities and hence relations are of uniform type, i.e. colorless graphs. In multi-relational domains, represented as colored graphs, one could indeed fit each type of relations separately; however, this approach would not take advantage of any correlations between relations. A simple example on social recommendation system is shown in Figure 1. There are two types of entities (persons, movies) and two types of relations (friend: person $\times$ person, likes: person $\times$ movie). We would like to exploit the additional correlations revealed by multi-relational knowledge to improve our recommendation quality, e.g. if friends of a person tend to rate dramas higher than comedies, then so does he with high probability. Furthermore, in the above person classification example, relations are not directly modeled in the probabilistic model, but rather play a role as 'prior' information reflected in the covariance matrix, thus it is difficult – if not impossible – to estimate relations in such a model, e.g. whether two persons are friends, or whether a person likes a movie.

To overcome these limitations, we extend relational/colorless GP models to multi-relational and hence colored domains. The basic idea of the resulting nonparametric Bayesian framework, which we called *multi-relational GPs* (MRGPs), is as follows. We introduce each entity several latent variables, each of which is drawn from a distinct GP. These latent variables represent the function values of entity features at different aspects, and can be intuitively viewed as the preference of the entity on the aspects. For example, in the social recommendation system, there are three latent variables associated with each person: $f_i^p \sim GP(0, \Sigma^p)$ representing person basis pro-

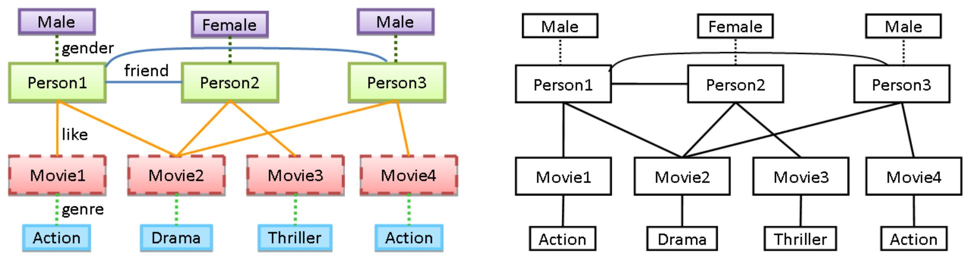

Figure 1: An example of social movie recommendation. Left: Colors resp. lines styles encode the different types of entities and relations. Right: Colorless variant.

file and being the function value of person attributes, $g_i^{friend,p} \sim GP(0, \Sigma^{friend,p})$ representing person preference on friendships and being the function value of the person's friendships, and $h_i^{like,p} \sim GP(0, \Sigma^{like,p})$ representing person preference on movies and being the function value of person's favorite movies. Similarly, there are two latent variables associated with each movie: $f_k^m \sim GP(0, \Sigma^m)$, and $h_k^{like,m} \sim GP(0, \Sigma^{like,m})$. These GPs are coupled together with relations by linear model. For example, a relation between person i and movie k is modeled as $P(R_{i,k}^{like}|\xi_{i,k}^{like}) = \Phi(\xi_{i,k}^{like})$, where $\Phi(\cdot)$ can be any sigmoid function, and $\xi_{i,k}^{like}$ is linear combination of related latent variables $\xi_{i,k}^{like} = (1, f_i^p, f_k^m, h_i^{like,p}, h_k^{like,m}) \cdot \omega^{like} + \epsilon$. ω^{like} is a weight vector and ϵ is a Gaussian noise. Note that with the latent variable f_i^p representing basis profiles of entities, relations of different types are coupled together. Similarly, the class label y_i^p of a person i can be model as $P(y_i^p|\xi_i^p) = \Phi(\xi_i^p)$, where $\xi_i^p = (1, f_i^p, h_i^{like,p}, g_i^{friend,p}) \cdot \omega^p + \epsilon$. It corresponds to classify persons with all (heterogeneous) information about the persons.

The basic idea underlying the MRGP model can also be formulated in a function-space view. For example, assume that the class label y_i^p of a person i is conditioned on a function value ξ_i^p of person attributes $\mathbf{x}_i$, friendships $\mathbf{u}_i$ and favorite movies $\mathbf{v}_i$:

$$\xi_i^p = \phi(\mathbf{x}_i, \mathbf{u}_i, \mathbf{v}_i)^T \omega, \quad \omega \sim \mathcal{N}(\mathbf{0}, \Omega), \tag{1}$$

where $\phi(\cdot)$ is a set of basis functions mapping the input vector into a higher dimensional feature space where the input vectors are (obviously) distinct. Since $\mathbf{x}_i$, $\mathbf{u}_i$ and $\mathbf{v}_i$ come from different information sources, it is natural that in different, higher dimensional feature spaces (i.e., mapping functions), the heterogeneous inputs are distinguished respectively. Thus, we introduce for each information source a distinct set of basis functions, and get a multiple GP model.

$$\xi_i^p = (1, \phi_1(\mathbf{x}_i)^T \omega_1, \phi_2(\mathbf{u}_i)^T \omega_2, \phi_3(\mathbf{v}_i)^T \omega_3) \cdot \omega^p + \epsilon. \tag{2}$$

Note, that ω_1, ω_2 and ω_3 are three different dimensional vectors, which dimensionality is related to the spaces the input vectors ($\mathbf{x}_i$, $\mathbf{u}_i$, and $\mathbf{v}_i$) are mapped. ω^p is a four dimensional weight vector, which combines the latent function values together.

The paper is organized as follows. We start off by describing MRGPs. Then, we will describe an EP approximation for inference and a EM method for parameter learning in Sec. 3. In Sec. 4, we will present our initial experimental evaluation. Before concluding, we will touch upon related work.

2 Multi-relational Gaussian Process Model

In this section, we will first provide a more detailed introduction of the multi-relational GP (MRGP) model, as well as an analysis of it for bipartite, directed, and undirected univariate relations. Finally, we will discuss the covariance matrices involved in the GPs.

The MRGP model can be formulated as:

- For each entity type c, there are multiple Gaussian process associated: $GP(0, \Sigma^c)$ for entity attributes, and $GP(0, \Sigma^{b,c})$ for each type b of relations the entities participate.
- For each entity i of type c:

- There is a latent variable f_i^c drawn from $GP(0, \Sigma^c)$, which represents the essential property of the entity. The latent value is function value of entity attributes.
- There is a latent variable $g_i^{b,c}$ drawn from $GP(0, \Sigma^{b,c})$, if the entity is involved in the relations of type b. It represents the preference of or hidden cause for the entity on such type of relations, and is function value of all such relations the entity participates.

- Each relation $R_{i,j}^b$ of type b between entities i and j has two states, +1 if the relation exists, -1 otherwise. The likelihood is $P(R_{i,j}^b = +1|\xi_{i,j}^b) = \Phi(\xi_{i,j}^b)$, where $\Phi(\cdot)$ can be any sigmoid function. $\xi_{i,j}^b$ is latent variable for the relation, which is a function of related latent variables. For simplification, we employ a linear combination

$$\xi_{i,j}^b = \omega_0^b + \omega_1^b f_i^{c_i} + \omega_2^b f_j^{c_j} + \omega_3^b g_i^{b,c_i} + \omega_4^b g_j^{b,c_j} + \epsilon. \tag{3}$$

c_i and c_j are the types of the entities i and j. $\omega^b = (\omega_0^b, \omega_1^b, \omega_2^b, \omega_3^b, \omega_4^b)$ is a vector of weights. ϵ is a Gaussian noise with zero mean and variance σ^2. Actually the weights of latent variables can be integrated into the hyperparameters of GP (mean and parameters in kernel functions) in inference, we can reduce the Equation (3) as:

$$\xi_{i,j}^b = f_i^{c_i} + f_j^{c_j} + g_i^{b,c_i} + g_j^{b,c_j}. \tag{4}$$

In the MRGP model, the latent variable f_i^c plays an important role. It is involved in all relations the entity i participates, by which different types of relations are coupled.

We notice that latent variables of relations of same type follow a Gaussian process with mean function $\mu = \mathbb{E}(\xi_{i,j}^b) = 0$, and covariance function $k(\xi_{i,j}^b, \xi_{i',j'}^b)$. Note, that the covariance function is different for bipartite, directed and undirected univariate relations.

Bipartite relations Such relations involve entities of different types, e.g. Like: person $\times$ movie. The latent variable of a relation between a person i and a movie k is written as

$$\xi_{i,k}^{like} = f_i^p + f_k^m + g_i^{like,p} + g_k^{like,m}. \tag{5}$$

Intuitively, f_i^p and f_k^m respectively represent the essential profiles of the person i and the movie k, $g_i^{like,p}$ and $g_k^{like,m}$ respectively represent the preference of the person i on movies and the movie k on persons. Equation (5) reveals that whether a person likes a movie is dependent on the person's profile, the movie's profile, and the person's preference on movies, as well as the movie's preference on persons. The covariance function of latent variables of bipartite relations is

$$k(\xi_{i,j}^{like}, \xi_{i',j'}^{like}) = \Sigma_{i,i'}^p + \Sigma_{j,j'}^m + \Sigma_{i,i'}^{like,p} + \Sigma_{j,j'}^{like,m}. \tag{6}$$

Directed univariate relations In these relations, the pairs of entities are of the same type, and ordered, e.g. Cite: paper $\times$ paper. If we describe directed univariate relations with a matrix, then the square matrix is not necessarily symmetric. Generally the roles the entities play in such relations are distinguished, e.g. the citing and cited papers in a citation relation. The latent variable of a relation from a paper i to a paper j is written as

$$\xi_{i,j}^{cite} = f_i^p + f_j^p + g_i^{citing} + g_j^{cited}. \tag{7}$$

Note, that f_i^p and f_j^p are drawn from the same GP, since f_i^p and f_j^p are functions of entity attributes and the entities involved in univariate relations are of the same type. However g_i^{citing} and g_j^{cited} are drawn from different GPs, since they have distinct roles in the relations. The covariance function of latent variables of directed univariate relations is

$$k(\xi_{i,j}^{cite}, \xi_{i',j'}^{cite}) = \Sigma_{i,i'}^p + \Sigma_{i,j'}^p + \Sigma_{j,i'}^p + \Sigma_{j,j'}^p + \Sigma_{i,i'}^{citing} + \Sigma_{j,j'}^{cited}. \tag{8}$$

Undirected univariate relations There is no orientation in such relations, i.e. a relation from i to j equals to a relation from j to i, e.g. Friend: person $\times$ person. The latent variable of a relation from a person i to a person j is written as

$$\xi_{i,j}^{friend} = f_i^p + f_j^p + g_i^{friend,p} + g_j^{friend,p}. \tag{9}$$

Note, that in an undirected univariate relation, the two involved entities play the same role, and we do not need to distinguish them, thus $g_i^{friend,p}$ and $g_j^{friend,p}$ are drawn from the same GP. The covariance function of latent variables of undirected univariate relations is

$$k(\xi_{i,j}^{friend}, \xi_{i',j'}^{friend}) = \Sigma_{i,i'}^{p} + \Sigma_{i,j'}^{p} + \Sigma_{j,i'}^{p} + \Sigma_{j,j'}^{p} + \Sigma_{i,i'}^{friend,p} + \Sigma_{i,j'}^{friend,p} + \Sigma_{j,i'}^{friend,p} + \Sigma_{j,j'}^{friend,p}. \quad (10)$$

Note, that the covariance function exactly satisfies the constraint on symmetric relations:

$$k(\xi_{i,j}^{friend}, \xi_{i',j'}^{friend}) \equiv k(\xi_{i,j}^{friend}, \xi_{j',i'}^{friend}) \equiv k(\xi_{j,i}^{friend}, \xi_{i',j'}^{friend}) \equiv k(\xi_{j,i}^{friend}, \xi_{j',i'}^{friend}) \quad (11)$$

To wrap up the introduction of MRGPs, let us discuss the covariance functions involved in the GPs. There are two types of covariance matrices: *one for entity attributes and one for relations*.

The latent variable $f_i^c \sim GP(0, \Sigma^c)$ represents the essential profile of an entity. The covariance matrix of the real process can be derived from entity attributes $\mathbf{x}_i^c$. The kernel function $k(\mathbf{x}_i^c, \mathbf{x}_j^c)$ is assumed as a function of $|\mathbf{x}_i^c - \mathbf{x}_j^c|$, e.g., squared exponential covariance function.

The other type of Gaussian processes involve relations. For each relation type b between entity types c_i and c_j, there are two GPs associated: $GP(0, \Sigma^{b,c_i})$ for c_i and $GP(0, \Sigma^{b,c_j})$ for c_j. There are generally two strategies to define their covariance matrices.

- The known relations of entity i can be represented as a vector. The covariance matrix Σ^{b,c_i} can be computed like the covariance matrix of entity attributes.
- Alternatively, the graph-based kernels are employed (e.g., [13, 8]). However, there is one difficulty in computing the graph-based kernels for multi-relational learning. If the relations are bipartite, i.e., $c_i \neq c_j$, then the graph kernels for univariate relations are not applicable. One solution is projecting the bipartite relations to univariate ones. In particular, we add a relation between entities i and j, iff both the entities link to the same (heterogeneous) entity. All entities linking to the same (heterogeneous) entity form a clique. Then we can compute the graph kernels on the projected graphs.

If attributes or relations are not informative, missing or unavailable, the covariance matrices computed with the information will not be reliable. In this case, we can learn the covariance matrices in the training process. We assume the covariance matrix has an inverse Wishart distribution $\Sigma \sim W^{-1}(\Sigma_0, \beta)$.

3 Inference and Parameter Learning

We have given a detailed description of the MRGP model. In this section, we will discuss the inference and hyperparameter estimation of MRGP.

Inference We use the movie recommendation example to illustrate the inference method. Assume there are N persons, M movies and ratings. The latent variables are $\mathbf{f}^p = \{f_1^p, f_1^p, \ldots, f_N^p\}$, $\mathbf{f}^m = \{f_1^m, f_1^m, \ldots, f_M^m\}$ and $\mathbf{g}^{like,p} = \{g_1^{like,p}, g_1^{like,p}, \ldots, g_N^{like,p}\}$, $\mathbf{g}^{like,m} = \{g_1^{like,m}, g_1^{like,m}, \ldots, g_M^{like,m}\}$. The key inference problem is computing the posterior of latent variables given the ratings $\mathbf{R}$:

$$P(\mathbf{f}^p, \mathbf{f}^m, \mathbf{g}^{like,p}, \mathbf{g}^{like,m}|\mathbf{R}) \propto P(\mathbf{f}^p)P(\mathbf{f}^m)P(\mathbf{g}^{like,p})P(\mathbf{g}^{like,m})P(\mathbf{R}|\boldsymbol{\xi}), \quad (12)$$

where $\boldsymbol{\xi}$ is a linear combination of $\mathbf{f}^p$, $\mathbf{f}^m$, $\mathbf{g}^{like,p}$, $\mathbf{g}^{like,m}$ (see Equation 5). We assume a rating from a person i to a movie k has the likelihood

$$P(R_{i,k}|\xi_{i,k}) = \frac{1}{1 + \exp(-R_{i,k}\xi_{i,k})}. \quad (13)$$

Unfortunately, computing this posterior is intractable because $P(\mathbf{R}|\boldsymbol{\xi})$ is non-Gaussian, and because $\boldsymbol{\xi}$ is sum of multiple latent variables, which couples the GPs together. To overcome the difficulty, we alternatively compute the posterior of $\boldsymbol{\xi}$ and use the linear relationship between $\boldsymbol{\xi}$ and $\mathbf{f}^p$, $\mathbf{f}^m$, $\mathbf{g}^{like,p}$ and $\mathbf{g}^{like,m}$. More precisely,

$$P(\boldsymbol{\xi}|\mathbf{R}) \propto P(\boldsymbol{\xi})P(\mathbf{R}|\boldsymbol{\xi}), \tag{14}$$

where $P(\boldsymbol{\xi})$ is the prior of $\boldsymbol{\xi}$, which is a GP, which mean and covariance matrix has been discussed in last section. Then we use the expectation propagation (EP) algorithm [5] to approximate the posterior, i.e., we use unnormalized Gaussian distributions $t_{i,k}(\xi_{i,k}|\tilde{\mu}_{i,k}, \tilde{\sigma}^2_{i,k}, \tilde{Z}_{i,k})$ to approximate Equation 13. In the inference procedure, we update the approximations for each latent variable $\xi_{i,k}$ sequentially until convergence. Note, however, that *if there are N persons and M movies, then the number of $\xi_{i,k}$ is NM whereas the number of unobservable variables ($\mathbf{f}^p, \mathbf{f}^m, \mathbf{g}^{like,p}, \mathbf{g}^{like,m}$) is only $2N + 2M \ll NM$*. In other words, there is much room for improvement of our inference approach to scale it to large data sets.

Hyperparameter Estimation We learn the hyperparameters ς under the empirical Bayesian framework. The hyperparameters include the parameters of kernel functions and the means of the GPs. We have

$$\begin{aligned}
\varsigma* &= \arg\max_{\varsigma} \ \log P(\mathbf{R}|\varsigma) \\
&= \arg\max_{\varsigma} \ \log \int P(\boldsymbol{\xi}|\varsigma)P(\mathbf{R}|\boldsymbol{\xi}) \ d\boldsymbol{\xi},
\end{aligned} \tag{15}$$

where the prior $P(\boldsymbol{\xi}|\varsigma)$ is Gaussian process, but the likelihood $P(\mathbf{R}|\boldsymbol{\xi})$ is not Gaussian, thus the integral is analytically intractable. To solve the problem, we approximate the likelihood with Gaussian distributions:

$$\begin{aligned}
\log P(\mathbf{R}|\varsigma) &\approx \log \int P(\boldsymbol{\xi}|\varsigma) \prod_{i,k} t_{i,k}(\xi_{i,k}|\tilde{\mu}_{i,k}, \tilde{\sigma}^2_{i,k}, \tilde{Z}_{i,k}) \ d\xi_{i,k}. \\
&= -\frac{1}{2}\log|\mathbf{K} + \tilde{\boldsymbol{\Sigma}}|^{-\frac{1}{2}} - \frac{1}{2}((\tilde{\boldsymbol{\mu}} - \boldsymbol{\mu}))^T(\mathbf{K} + \tilde{\boldsymbol{\Sigma}})^{-1}((\tilde{\boldsymbol{\mu}} - \boldsymbol{\mu})) + C,
\end{aligned} \tag{16}$$

where C can be viewed as a constant, which do not influence the optimization of hyperparameters. $\tilde{\boldsymbol{\Sigma}}$ is a diagonal matrix, which entries are $\tilde{\sigma}^2_{i,k}$. We propose an Expectation Maximum solution to learn the hyperparameters. In the E step, the EP parameters ($\tilde{\mu}_{i,k}$, $\tilde{\sigma}^2_{i,k}$, $\tilde{Z}_{i,k}$) are optimized to approximate the posterior of latent variables with the current values of hyperparameters. The approximation procedure has been discussed above. In the M step, the hyperparameters are optimized to maximize the log marginal likelihood. Referring to [6], we employ a conjugate gradient method to solve the optimization problem.

4 Experimental Evidence

We evaluated the MRGP model on the MovieLens data [7]. We randomly selected a subset of 150 users, 100 movies and 2763 relations for the experiment. The task is predicting person preference on movies given person/movie attributes and previous ratings. The persons have attributes Age, Gender, Occupation, and the movies have attributes Published-year, Genres and so on. The like relations have two states, where $R = +1$ indicates that the person likes the movie and -1 otherwise. Note that the ratings in MovieLens are originally based on a five-star scale, so we transfer each rating to binary value with $R = +1$ if the rating is higher than the person's average rating, vice versa. For testing, the relation is predicted to exist (i.e., $R = +1$) if the predictive probability is larger than a threshold $\varepsilon = 0.5$.

We compared the MRGP model with the infinite hidden relational model (IHRM) [10], which introduces nonparametric hierarchical Bayesian modeling into relational learning. The predictive accuracy of the IHRM was 61.67%; slightly lower than the predictive accuracy of the MRGP, which was 63.33%.

5 Discussion and Related Work

Multi-relational learning has a long tradition in the machine learning community [3, 2]. So far, however, Gaussian process models have been mainly developed for the single-relational, i.e., col-

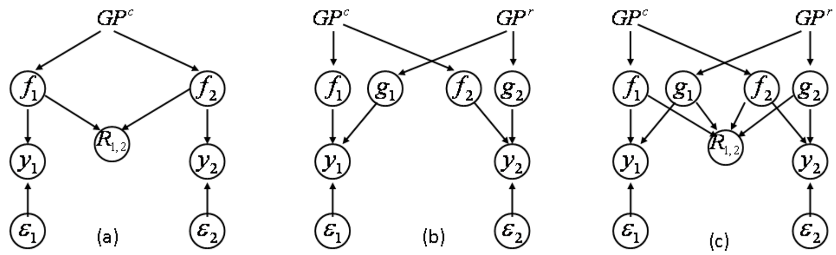

Figure 2: The graphical representation of (a) RGPs [1], (b) XGPs [8], and (c) MRGPs for *one univariate relation*. y_i and $R_{1,2}$ respectively denote entity class labels and the univariate relation. ϵ_i is a Gaussian noise. f_1 and f_2 are latent variables sharing a common GP, which can be viewed as function values of entity attributes. g_1 and g_2 are latent variables sharing another GP, which can be viewed as function values of relations. MRGPs can simply be viewed as the combination of RGPs and XGPs, but are especially applicable for multi-relational learning.

orless graph case. In general, to develop relational GP models, there are two major challenges to be addressed. One is at which level to define the GPs, and the other is how to incorporate the relation knowledge. One line of research [1, 12, 8, 11] introduced GP models where the same type of entities share a common GP. For each entity i of type c, there are d latent variables associated ($d = 1$ in [1, 8]), all of which are generated from GP^c. In the MRGP, there are additional GPs introduced for each type of relations. Similar to MRGPs, Chu *et al.* [8] use an extra GP related to relations but only one. For modeling relations, there are generally two strategies in existing work. One is encoding relations in the *covariance matrix*, e.g., [13, 8]. The other is encoding relations as *random variables* conditioned on the latent variables of entities involved in relations [1, 12, 11]. In the MRGP, we essentially combine the two strategies together. An important property of relational system is that future relations depend on previous relations. For example, prescribed procedures $\mathbf{u}$ of a patient will influence his future procedure u_* with likelihood $P(u_*|f(\mathbf{u}))$, where $f(\cdot)$ is a nonparameterized function of the prescribed procedures. Thus it is natural to encode relations both in covariance matrix and as random variables. Figure 2 illustrates the differences between these models. The linear combination of random variables assumed in MRGPs is similar to Singh and Gordon's [9] generalization of linear models to multi-relational domains.

6 Conclusion

We developed a nonparametric Bayesian framework for multi-relational data, i.e., colored graphs based on Gaussian processes. The proposed multi-relations Gaussian process (MRGP) model combines the *covariance* and the *random variables* approaches to model relations within Gaussian processes and hence provides a flexible tool for many relational learning tasks. The initial experimental evidence is promising and suggests that mixing information from multiple relations indeed leads to better performance. We will perform further experiments to evaluate the MRGP model.

MRGPs suggest several interesting directions for future research such as scaling MRGPs to large data sets, developing multi-relational variants of dimensionality reduction techniques along the lines of [4], and applying MRGPs in spatial-relational domains and robotics.

Acknowledgments Zhao Xu and Kristian Kersting are supported by the Fraunhofer ATTRACT Fellowship "Statistical Relational Activity Mining" (STREAM).

References

[1] W. Chu, V. Sindhwani, Z. Ghahramani, and S. Keerthi. Relational learning with gaussian processes. In *Neural Information Processing Systems*, 2006.

[2] L. De Raedt. *Logical and Relational Learning*. Springer, 2008.

[3] L. Getoor and B. Taskar, editors. *Introduction to Statistical Relational Learning*. The MIT Press, 2007.

[4] N. Lawrence. Probabilistic non-linear principal component analysis with gaussian process latent variable models. *Journal of Machine Learning Research (JMLR)*, 6:1783–1816, 2005.

[5] T. Minka. *A family of algorithms for approximate Bayesian inference*. PhD thesis, Massachusetts Institute of Technology, 2001.

[6] C. E. Rasmussen and C. K. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, 2006.

[7] B. M. Sarwar, G. Karypis, J. A. Konstan, and J. Riedl. Analysis of recommender algorithms for e-commerce. In *Proc. ACM E-Commerce Conference*, pages 158–167. ACM, 2000.

[8] R. Silva, W. Chu, and Z. Ghahramani. Hidden common cause relations in relational learning. In *Neural Information Processing Systems*, 2007.

[9] A. Singh and G. Gordon. Relational learning via collective matrix factorization. In *In Proc. 14th Intl. Conf. on Knowledge Discovery and Data Mining (KDD)*, 2008.

[10] Z. Xu, V. Tresp, K. Yu, and H.-P. Kriegel. Infinite hidden relational models. In *Proc. 22nd UAI*, 2006.

[11] K. Yu and W. Chu. Gaussian process models for link analysis and transfer learning. In *Neural Information Processing Systems*, 2007.

[12] K. Yu, W. Chu, S. Yu, V. Tresp, and Z. Xu. Stochastic relational models for discriminative link prediction. In *Neural Information Processing Systems*, 2006.

[13] X. Zhu, J. Kandola, J. Lafferty, and Z. Ghahramani. Graph kernels by spectral transforms. In O. Chapelle, B. Schoelkopf, and A. Zien, editors, *Semi-Supervised Learning*. MIT Press, 2005.