2006 年情報論的学習理論ワークショップ
2006 Workshop on Information-Based Induction Sciences (IBIS2006)
Osaka, Japan, October 31 - November 2, 2006.

A Frequency-based Stochastic Blockmodel

Kenichi Kurihara [*] Yoshitaka Kameya [†] Taisuke Sato [‡]

Abstract: We propose a frequency-based infinite relational model (FIRM), which takes into account the frequency of relation whereas stochastic blockmodels ignore frequency. We also derive a variational inference method for the FIRM to apply to a large dataset. Experimental results show that the FIRM gives better clustering results than a stochastic blockmodel on a dataset which has the frequency of relation.

key words: stochastic blockmodels, Dirichlet process, variational inference

1 Introduction

Recently in machine learning, relational learning has received a great deal of attention, for example, to find social roles in social network. While the stochastic blockmodel has been popular for such relational learning in sociology, it is now widely used for various applications including clustering proteins, discovering concepts, etc [9, 13].

Infinite relational models (IRMs) [9, 13] are stochastic blockmodels exploiting the Dirichlet process [6, 2] so that we do not need to determine the number of clusters a priori. Stochastic blockmodels of mixed membership (SBMM) are also stochastic blockmodels that model multiple observation of tables [1].

In this paper, we propose a frequency-based infinite relational model (FIRM). The FIRM takes into account the frequency of relation, which is statistically informative but ignored by stochastic blockmodels. Our model generalizes the IRM and can also observe multiple tables as the SBMM. To apply the FIRM to a large dataset, we also derived a variational inference algorithm for the FIRM. Experimental results show the FIRM gives better clustering results than the IRM on a dataset which has the frequency of relation.

2 Stochastic Blockmodels

Stochastic blockmodels are proposed for social network data in sociology [8]. Fig.1 shows an example of stochastic blockmodels. Let's say we have type 1 entities consisting of people 1–10, type 2 entities consisting of animals a–j and a relation $likes(\cdot,\cdot)$. The left hand side of Fig.1 shows which animal each person likes by black dots. Stochastic blockmodels find clusters of each type as the right hand side of Fig.1 shows.

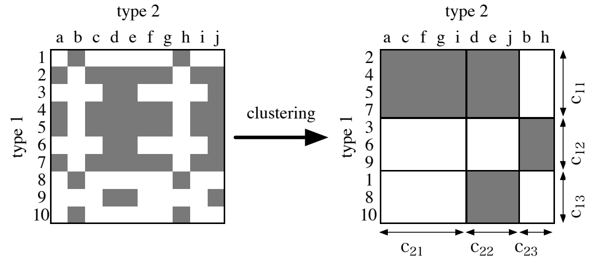

Figure 1: A toy example of stochastic blockmodels. Type 1 entities consist of people 1–10, and type 2 entities consist of animals a–j. Black dot indicates that a person likes an animal. Type 1 and 2 are partitioned into clusters $c_{11} - c_{13}$ and into $c_{21} - c_{23}$, respectively.

More formally, the stochastic blockmodel for one binary relation between two types is defined as,

$$p(R, Z^1, Z^2|\eta, \pi^1, \pi^2) = p(R|Z^1, Z^2, \eta)p(Z^1, Z^2|\pi^1, \pi^2)$$

$$p(R|Z^1, Z^2, \eta) = \prod_{e^1=1}^{N^1}\prod_{e^2=1}^{N^2} \text{Bernoulli}(R(e^1, e^2); \eta(z^1_{e^1}, z^2_{e^2})) \quad (1)$$

$$p(Z^1, Z^2|\pi^1, \pi^2) = \text{Multinomial}(Z^1; \pi^1)\text{Multinomial}(Z^2; \pi^2) \quad (2)$$

where for $i = 1, 2$, $Z^i = (z^i_1, ..., z^i_{N^i})$, $\pi^i = (\pi^i_1, ...\pi^i_{K^i})$, $z^i_n \in \{1, ..., K^i\}$ and $\sum_c \pi^i_c = 1$. Note that we can extend this model to relations involving an arbitrary

[*] Department of Computer Science, Tokyo Institute of Technology, 152-8552, 2-12-1 Ookayama, Meguro-ku Tokyo, Japan, tel. +81-3-5734-2186, e-mail kurihara@mi.cs.titech.ac.jp
[†] e-mail kameya@mi.cs.titech.ac.jp
[‡] e-mail sato@mi.cs.titech.ac.jp

number of entities like $R(\cdot, ..., \cdot)$.

The main task for this model is to infer Z^1 and Z^2, i.e. the assignments of entities to clusters. Since stochastic blockmodels are quite general approach to relational learning, its extensions have been proposed in machine learning.

In the following section, we review the recent variants of stochastic blockmodels.

3 Variants of the Stochastic Blockmodel

3.1 Infinite Relational Model

Xu et al and Kemp et al. proposed the infinite relational model (IRM) independently in [13] and [9]. Although their inference methods are different, their models are basically the same. The difference from the traditional blockmodel is the use of the Dirichlet process (DP) [6, 2]. The DP is a non-parametric Bayesian method which infers the number of clusters for clustering. The DP overcomes a weak point that the traditional blockmodel requires the number of clusters a priori. Their model is identical to the stochastic blockmodel except that Z^1 and Z^2 are drawn from the DP instead of Eq.2,

$$Z^1|\gamma \sim \mathrm{DP}(Z^1;\gamma), \quad Z^2|\gamma \sim \mathrm{DP}(Z^2;\gamma).$$

To infer Z^i for $i = 1, 2$, Xu et al. derived a deterministic maximum likelihood method using the mean field approximation, and Kemp et al. used a simpler hill-climbing method. Note that the exact inference of Z^1 and Z^2 can be carried out using Markov chain Monte Carlo (MCMC) methods although it may take long time until convergence.

3.2 Stochastic Blockmodels of Mixed Membership

Airoldi et al. proposed stochastic blockmodels of mixed membership (SBMM) in [1]. Although their model is a relational version of a mixed membership model [5], it can also be seen as a blockmodel that accepts multiple observations. Remembering Fig.1, stochastic blockmodels partition just one observation of a table of a relation. The SBMM also partitions types, but handle the multiple observations of a table. For example, chemical reactions of proteins can vary for each observation. Stochastic blockmodels handle only a single trial of the experiment, whereas the SBMM can model an arbitrary number of trials.

However, it only accepts multiple "tables", i.e. each relation must be observed the same number of times. Let's say that we have type $1 = \{1, 2, ..., \}$ and type $2 = \{a, b, ...\}$. If we observe $R(1, a)$ three times, other relations must be observed three times as well in the SBMM. We also note that the SBMM only model data consisting of a single type, for example, R : protein × protein → $\{1, 0\}$.

The condition of observing each relation the same times is too restrictive in some cases, for example, modification relation in natural language and causality in medical data, etc. We propose a more flexible model that accepts unevenly distributed observation in the following section.

4 Frequency-based Infinite Relational Model

4.1 General Definition

We propose a frequency-based infinite relational model (FIRM), which is a generative model of relation. Relations sampled from the FIRM are i.i.d., which allows us to observe relations different times for each relation. We introduce discrete distributions to draw entities,

$$e^1|\boldsymbol{u}^1 \sim \mathrm{Multinomial}(e^1;\boldsymbol{u}^1), \; e^2|\boldsymbol{u}^2 \sim \mathrm{Multinomial}(e^2;\boldsymbol{u}^2).$$

Let D be a dataset, $D = \{(e^1_i, e^2_i)|i = 1, ..., m\}$. The probability of D given other variables is,

$$p(D|\boldsymbol{u}^1, \boldsymbol{u}^2, Z^1, Z^2, \eta) = \prod_{i=1}^{m} \boldsymbol{u}^1_{e^1_i} \boldsymbol{u}^2_{e^2_i} \eta(z^1_{e^1_i}, z^2_{e^2_i})^{R(e^1_i, e^2_i)} \\ (1 - \eta(z^1_{e^1_i}, z^2_{e^2_i}))^{1-R(e^1_i, e^2_i)}.$$

where m is the number of observations, and Z^1 and Z^2 are drawn from the DP as the IRM. Note that the FIRM is identical to the IRM when $D = \{R(e^1, e^2)|e^1 = 1, ..., N^1, e^2 = 1, ..., N^2\}$. Therefore, the FIRM generalizes the IRM. Moreover, the FIRM can also model multiple tables as the SBMM.

We are interested in inferring Z^1 and Z^2. This can be done by Markov chain Monte Carlo (MCMC) methods. However, it is well known that DP inference by MCMC is too slow to apply to a large dataset. Therefore, we utilize a variational inference for the FIRM. A variational inference for the Dirichlet process (VDP) has been proposed in [4]. Although the FIRM is not a simple DP mixture, we can derive a variational inference for the FIRM. We will briefly review VDP first.

4.2 Variational Dirichlet Process

Variational inference methods are alternatives to sampling methods for Bayesian learning [3, 7] especially in the context of large-scale problems. Blei and Jordan have applied a variational inference for the DP [4] in the stick-breaking (SB) representation [12]. The SB representation introduces new random parameters $\boldsymbol{v} = (v_1, ...)$. The DP in the SB representation is represented as,

$$\begin{array}{lcll} v_t|\alpha & \sim & \mathrm{Beta}(v_t; 1, \alpha), & \text{for } t = 1, ... \\ \eta_t|G_0 & \sim & G_0, & \text{for } t = 1, ... \\ z_i|\boldsymbol{v} & \sim & \mathrm{Multinomial}(z_i; \pi(\boldsymbol{v})), & \text{for } i = 1, ..., m \\ x_i|z_i, \eta & \sim & p(x_i|\eta_{z_i}), & \text{for } i = 1, ..., m \end{array}$$

where η_t is the parameter of the tth component, $\pi_t(\boldsymbol{v}) = v_t \prod_{s=1}^{t-1}(1 - v_s)$, and z_i and x_i are the ith assignment and observation, respectively.

VDP infers $q(Z, \eta, \boldsymbol{v})$ as an approximate posterior, $p(Z, \eta, \boldsymbol{v}|X)$, assuming the following factorization,

$$q(Z, \eta, \boldsymbol{v}) = \prod_{t=1}^{T-1} q(v_t) \prod_{t=1}^{T} q(\eta_t) \prod_{i=1}^{m} q(z_i), \tag{3}$$

where T is a truncation level. At truncation level T, we assume $p(v_T = 1) = 1$ and $q(v_T) = 1$. This assumption leads to $\pi_t = 0$ for all $t > T$. Therefore, the infinite mixture boils down to a finite mixture. Note that if we set T large enough, the approximation is quite good in practice.

Using Jensen's inequality, we find a bound of $p(X)$,

$$p(X) \geq E\left[\log \frac{p(X, Z, \eta, \boldsymbol{v})}{q(Z, \eta, \boldsymbol{v})}\right]_{q(Z,\eta,\boldsymbol{v})} \tag{4}$$

where $E[f(x)]_{g(x)} = \int dx\ g(x) f(x)$. The approximate posterior, q, is derived by taking variation of Eq.4 and setting to zero.

4.3 Variational Inference for FIRM

Let $W = (Z^1, Z^2, v^1, v^2, \eta, \boldsymbol{u}^1, \boldsymbol{u}^2)$, which is a set of hidden variables. We are interested in inferring the posterior distribution, $p(W|D)$. Using a variational inference, we approximate the posterior as $q(W)$. First, we make the following bound of $p(D)$,

$$p(D) \geq E\left[\log \frac{p(D, W)}{q(W)}\right]_{q(W)}. \tag{5}$$

To make the approximate posterior, q, tractable, we assume the following factorization,

$$q(W) = q(Z^1) q(Z^2) q(\boldsymbol{v}^1) q(\boldsymbol{v}^2) q(\eta) q(\boldsymbol{u}^1) q(\boldsymbol{u}^2). \tag{6}$$

The joint distribution of the FIRM is factorized as follows,

$$\begin{aligned} p(D, W) =& p(D|Z^1, Z^2, \eta) p(Z^1|\boldsymbol{v}^1) p(Z^2|\boldsymbol{v}^2) \\ & \times p(\boldsymbol{v}^1) p(\boldsymbol{v}^2) p(\eta) p(\boldsymbol{u}^1) p(\boldsymbol{u}^2). \end{aligned} \tag{7}$$

We put the following priors into Eq.7.

$$\begin{aligned} & p(\eta(t^1, t^2)) = \mathrm{Beta}(\eta(t^1, t^2); \beta, \beta) \\ & p(Z^1|\boldsymbol{v}) = \prod_{e=1}^{N^1} \pi^1_{z^1_e}, \qquad p(Z^2|\boldsymbol{v}) = \prod_{e=1}^{N^2} \pi^2_{z^2_e} \\ & p(v^1_{t^1}) = \mathrm{Beta}(v^1_{t^1}; 1, \gamma), \qquad p(v^2_{t^2}) = \mathrm{Beta}(v^2_{t^2}; 1, \gamma) \\ & p(\boldsymbol{u}^1) = \mathrm{Dirichlet}(\boldsymbol{u}^1; \alpha^1_0), \quad p(\boldsymbol{u}^2) = \mathrm{Dirichlet}(\boldsymbol{u}^2; \alpha^2_0) \end{aligned}$$

Taking the variation of Eq.5, we will find the approximate posterior, q[1],

$$\begin{aligned} & q(v^1_t) = \mathrm{Beta}(v^1_t; \gamma^1_{1,t}, \gamma^1_{2,t}) \\ & q(v^2_t) = \mathrm{Beta}(v^2_t; \gamma^2_{1,t}, \gamma^2_{2,t}) \\ & q(\eta(t^1, t^2)) = \mathrm{Beta}(\eta(t^1, t^2); \tau_{1,t^1,t^2}, \tau_{2,t^1,t^2}) \\ & q(z^1_e) \propto \exp E[\log p(R, Z^1, Z^2, \eta, \boldsymbol{v}^1, \boldsymbol{v}^2)]_{q(Z^1_{\neg e}, Z^2, \eta, \boldsymbol{v}^1, \boldsymbol{v}^2)} \\ & q(z^2_e) \propto \exp E[\log p(R, Z^1, Z^2, \eta, \boldsymbol{v}^1, \boldsymbol{v}^2)]_{q(Z^1, Z^2_{\neg e}, \eta, \boldsymbol{v}^1, \boldsymbol{v}^2)}, \end{aligned}$$

where

$$\begin{aligned} & \gamma^1_{1,t^1} = 1 + m^1_{t^1}, \qquad \gamma^1_{2,t^1} = \gamma + \sum_{j=1}^{T^1} m^1_j \\ & \gamma^2_{1,t^2} = 1 + m^2_{t^2}, \qquad \gamma^2_{2,t^2} = \gamma + \sum_{j=1}^{T^2} m^2_j \\ & m^1_{t^1} = \sum_{e=1}^{N^1} q(z^1_e = t^1), \quad m^2_{t^2} = \sum_{e=1}^{N^2} q(z^2_e = t^2) \\ & \tau_{1,t^1,t^2} = \beta + \sum_{i=1}^{m} q(z^1_{e^1_i} = t^1) q(z^2_{e^2_i} = t^2) I(R(e^1_i, e^2_i) = 1) \\ & \tau_{2,t^1,t^2} = \beta + \sum_{i=1}^{m} q(z^1_{e^1_i} = t^1) q(z^2_{e^2_i} = t^2) I(R(e^1_i, e^2_i) = 0). \end{aligned}$$

[1] We notice that $q(\boldsymbol{u}^1)$ and $q(\boldsymbol{u}^2)$ are equal to $p(\boldsymbol{u}^1|D)$ and $p(\boldsymbol{u}^2|D)$. Therefore, it is easy to derive them. We omit them here to save space.

Note that we set the truncation level of type 1 clusters and type 2 clusters to T^1 and T^2, respectively, and that $I(\cdot)$ is the indicator function.

We have derived a variational inference algorithm for the FIRM. We also apply the variational inference for the IRM in experiments while it is quite similar to derive the variational IRM.

5 Experimental Results

In this section, we experimentally show that the FIRM works better than the IRM on a dataset including multiple observations. We use a word co-occurrence dataset extracted from Mainichi newspaper 1993–2002 by CaboCha, a Japanese dependency structure analyzer. The dataset has more than one million adjective–noun pairs consisting of 210,605 distinct pairs, 1,291 adjectives and 3,705 nouns. Note that we cannot apply stochastic blockmodels of mixed membership to this dataset due to the violation of its strict assumption on frequency. We apply the FIRM and the IRM to this dataset to make clusters of adjectives and nouns[2]. Since the dataset does not have labels, we apply another word co-occurrence clustering called semantic aggregate model (SAM) [10]. Because it is verified by psychological experiments that clusters discovered by the SAM are consistent with humans' intuitions [11], we evaluate the FIRM and the IRM using the SAM clusters as correct labels,

We call adjective type 1 and noun type 2, and distinguish them in superscript like N^1 and N^2. For both of the IRM and the FIRM, we set truncation levels T^1 and T^2 to 80 and 120, and set α, β, γ and δ to 1.0, 0.1, 1.0 and 1.0, respectively. This experiment was repeated 30 times. Each run of the experiment took less than 5 minutes. The best results of 30 trials in terms of the free energy are depicted in Fig.2 and Fig.3 with the most likely Z^1 and Z^2, which maximize $q(Z^1)$ and $q(Z^2)$. Each row corresponds to one adjective, and each column corresponds to one noun. Clusters are ordered in descending order of the size, i.e. the top-most and left-most clusters are the largest clusters. Each black dot represents the existence of an adjective-noun co-occurrence. Therefore, dense cells signify the strength of the relations between adjective clusters and noun clusters.

Table 1: Clusters that are not discovered by the SAM but discovered by relational models.

adjective cluster	new, good, different,...
noun cluster	thing, object, place,...

We compare the FIRM with the IRM using two criteria, *coverage* and *purity*. We regard clusters discovered by the SAM as correct labels. Since the SAM found that the number of clusters is 50, relational models should also discover the same 50 labels. The *coverage*[3] shows the percentage of labels rediscovered by the relational models. Even when relational models achieve high *coverage*, the extent to which each relational cluster contains co-occurrence from one label might be low. *purity*[4] measures this extent.

We show the distribution of covered SAM clusters in Fig.4. The FIRM discovered clusters which cover all of the SAM clusters, i.e. *coverage* = 100%. On the other hand, the *coverage* by the IRM is 86%.

We plot the *purity* in Fig.5 varying a hyperparameter, $\beta = 1, 0.1$ and 0.01. For every β and i, the FIRM achieved higher *purity* $S_i(Z^1, Z^2)$ than the IRM.

One may suspect that the better *coverage* and *purity* is simply because the FIRM discovered more clusters

[2] Co-occurrence data has only positive data, i.e. $R(e^1, e^2) = 1$ because we can not observe pairs of words which do not make co-occurrences. In other words, we can observe only co-occurrences existing in a dataset. Let D be a dataset consisting of positive data. We use the following dataset D',

$$D' = D \cup \{R(e^1, e^2) = 0 | R(e^1, e^2) = 1 \notin D\}. \qquad (8)$$

[3] The *coverage* shows how many SAM labels are discovered, whose definition is (#covered SAM clusters) / (#SAM clusters). Let's say we are looking at a cell specified by adjective cluster t^1 and noun cluster t^2, cell(t^1, t^2). cell(t^1, t^2) contains adjective–noun pairs, $\{(a, n) | a \in t^1 \text{ and } n \in t^2\}$. We can predict the most likely SAM cluster for each pair by $p(c|a, n)$. Let $d(c, t^1, t^2)$ be the number of pairs in cell(t^1, t^2) whose most likely SAM cluster is c. A SAM cluster c is covered if and only if there exist t^1 and t^2 such that $c = \arg\max_j d(j, t^1, t^2)$.

[4] The *purity* is defined for each cell. Cells which have high *purity* consist of pairs that have the same most likely SAM cluster. For example, when a cell has *purity* 0.9, 90% of pairs in the call has the same most likely SAM cluster (see [14] for more general definition of the *purity*). More formally, *purity* $S_i(t^1, t^2)$ is defined as

$$S_i(t^1, t^2) = \frac{\sum_{j=1}^{i} \tilde{d}(j, t^1, t^2)}{\sum_j d(j, t^1, t^2)} \quad \text{for } i = 1, ..., K. \qquad (9)$$

where K is the number of clusters of the SAM and $\tilde{d}$ is sorted d in descending order. We also define the *purity* of assignments,

$$S_i(Z^1, Z^2) = \sum_{t^1=1}^{T^1} \sum_{t^2=1}^{T^2} \frac{\left\{\sum_j d(j, t^1, t^2)\right\} S_i(t^1, t^2)}{N^1 N^2} \qquad (10)$$

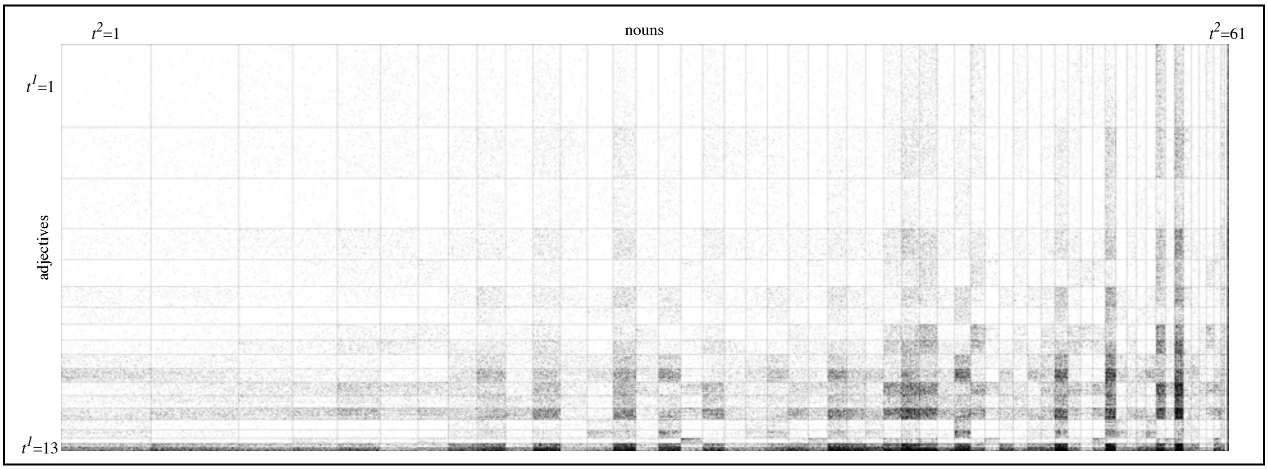

Figure 2: Clustering results by the IRM. 13 adjective clusters and 61 noun clusters were discovered.

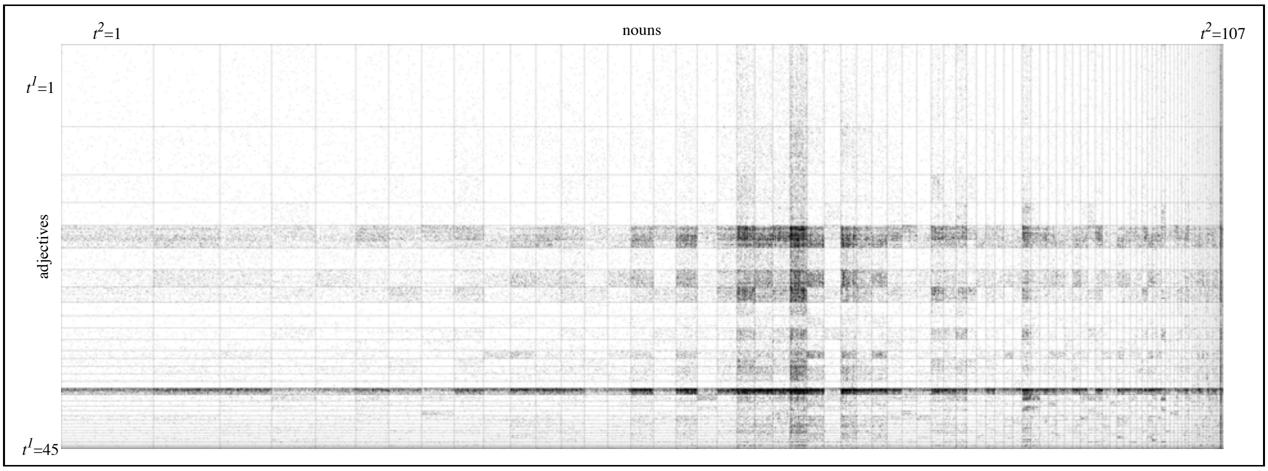

Figure 3: Clustering results by the FIRM. 45 adjective clusters and 107 noun clusters were discovered.

than the IRM. It may be the reason, but we emphasize that the number of clusters was automatically detected by inference. In other words, taking into account frequency enabled us to discover fine-grained clusters.

For both of the IRM and the FIRM, the *purity* of whole assignments was not very high because *purity* of some cells are quite small. The reason is that relational models also find clusters that are not discovered by the SAM, like Table 1. We can see that these words make co-occurrences with many words. Therefore, these clusters are uninformative for the SAM, and hence the SAM does not discover them.

6 Conclusion and Future Work

We proposed a frequency-based infinite relational model (FIRM). The FIRM takes into account the frequency of relation, which is statistically informative whereas traditional stochastic blockmodels ignore it. The FIRM closely relates to recent work, infinite relational models (IRM) by Xu et al. [13] and Kemp et al. [9], and stochastic blockmodels of mixed membership (SBMM) by Airoldi [1]. Our model generalizes the IRM and can also observe multiple tables as the SBMM. To apply the FIRM to a large dataset, we derived a variational inference algorithm for the FIRM.

We experimentally showed that the FIRM achieved better results than the IRM on a dataset which has the frequency of co-occurrence relation. We also found that the FIRM discovered clusters that are consistent with the ones discovered by the SAM, which was verified by psychological experiments.

The IRM is initially proposed for relational cluster-

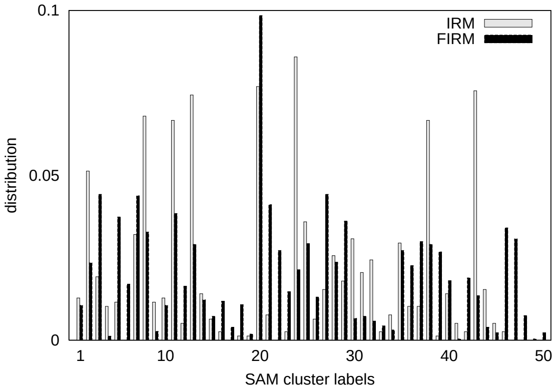

Figure 4: Distributions of cells over the SAM clusters. Y axis shows the ratio of the number of cells that are covered by each SAM cluster. The IRM and the FIRM covered 43 SAM clusters and 50 SAM clusters, respectively. The *coverage* of the IRM and the FIRM is 86% (= 43/50) and 100% (= 50/50).

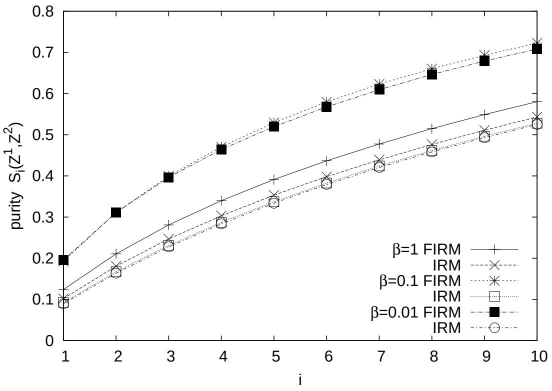

Figure 5: *Purity* $S_i(Z^1, Z^2)$ varying β.

ing with an arbitrary number of types. Therefore, it is straight forward to apply the FIRM to feature-rich datasets like co-occurrences of subject–verb, verb–objective and adjective–noun. It should be interesting to see clustering results of such a dataset.

Acknowledgment

This research was funded by the 21st Century COE Program "Framework for Systematization and Application of Large-scale Knowledge Resources."

References

[1] Edoardo M. Airoldi, David M. Blei, Stephen E. Fienberg, and Eric P. Xing. Stochastic block models of mixed-membership. In *Workshop on Statistical Network Analysis at ICML*, 2006.

[2] C.E. Antoniak. Mixtures of Dirichlet processes with applications to bayesian nonparametric problems. *The Annals of Statistics*, 2:1152–1174, 1974.

[3] Hagai Attias. A variational Bayesian framework for graphical models. In *Advances in Neural Information Processing Systems*, volume 12, 2000.

[4] David M. Blei and Michael I. Jordan. Variational inference for dirichlet process mixtures. *Bayesian Analysis*, 1(1):121–144, 2006.

[5] Elena A. Erosheva and Stephen E. Fienberg. Bayesian mixed membership models for soft classification. In *Classification - the Ubiquitous Challenge: Proceedings of the 28th Annual Conference of the Gesellschaft Fnr Klassifikation*, 2004.

[6] T.S. Ferguson. A bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1:209–230, 1973.

[7] Zoubin Ghahramani and Matthew J. Beal. Variational inference for Bayesian mixtures of factor analysers. In *Advances in Neural Information Processing Systems*, volume 12, 2000.

[8] Paul W. Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. Stochastic blockmodels: First steps. *Social Networks*, 5(2):109–137, 1983.

[9] Charles Kemp, Joshua B. Tenenbaum, Thomas L. Griffiths, Takeshi Yamada, and Naonori Ueda. Learning systems of concepts with an infinite relational model. In *AAAI*, 2006.

[10] Daichi Mochihashi and Yuji Matsumoto. Probabilistic representatin of meaning. In *IPSJ-NL*, volume 4, pages 77–84, 2002.

[11] Masanori Nakagawa, Asuka Terai, and Taisuke Sato. A computational model of metaphor understanding using a statitical analysis of japanese corpora based on soft clustering – toward a metaphorical search engine –. In *Framework for Systematization and Application of Large-scale Knowledge Resources*, 2006.

[12] Jayaram Sethuraman. A constructive definition of dirichlet priors. *Statistica Sinica*, 4:639–650, 1994.

[13] Zhao Xu, Kai Yu Volker Tresp, and Hans-Peter Kriegel. Infinite hidden relational models. In *Proceedings of the 22nd International Conference on Uncertainty in Artificial Intelligence*, 2006.

[14] Ying Zhao and George Karypis. Criterion functions for document clustering: Experiments and analysis. Technical Report Technical Report TR #01–40, Department of Computer Science, University of Minnesota, Minneapolis, 2001.