Statistical Models for Discovering Knowledge from Relational Data

Iku Ohama

Panasonic

JAFOE18@Tsukuba International Congress Center, June 18-20, 2018

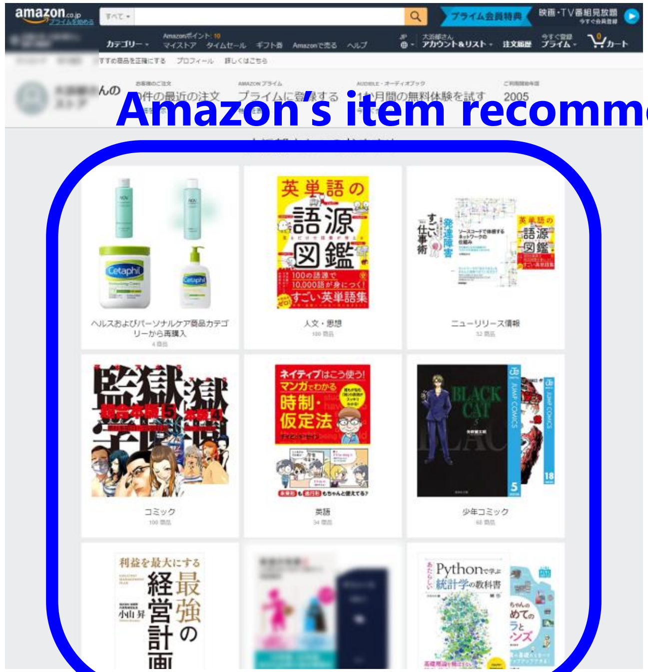
Background

Nowadays, data analysis is a very important technology for business to provide better UX.

Recommended TV programs for me

dimora.jp

Amazon's item recommendation



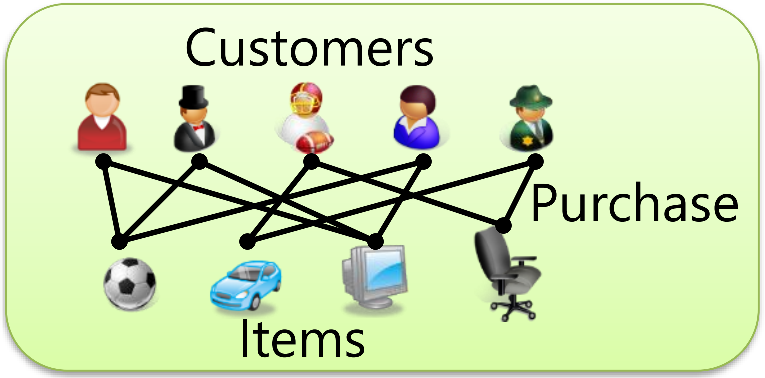
www.amazon.co.jp

A key technology underlying such intelligent systems is Relational Data Analysis.

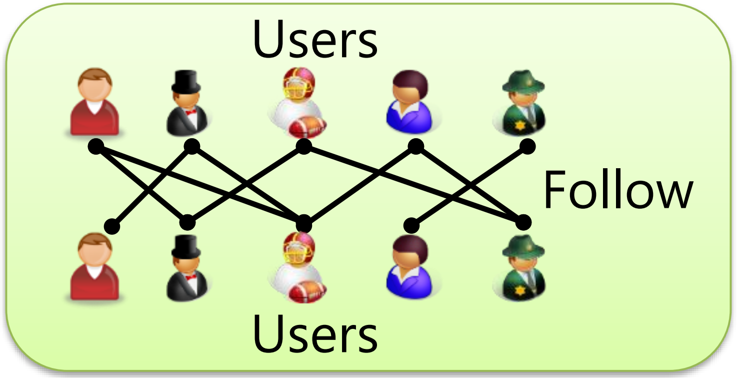
What is Relational Data?

Relational Data encoding pairwise relationships appears in many business fields.

Point-of-sale (POS) data



Social Relationships



1	0	1	0
1	0	1	0
0	1	0	1
1	0	1	0
0	1	0	1

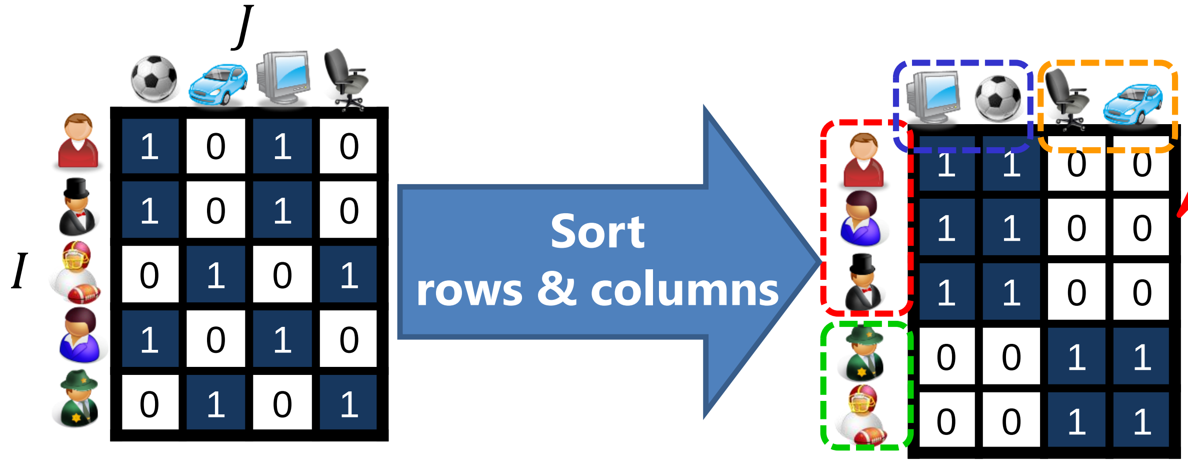
0	1	1	0	0
1	0	1	0	0
0	1	0	0	1
0	0	1	0	1
0	0	0	1	0

Relational Data is equivalently represented by a binary matrix

Relational Data Analysis

A major motivation behind analyzing relational data is to discover significant interaction patterns.

Simple case of Relational Data Analysis



$I \times J$ Relational Data $\boldsymbol{x}$

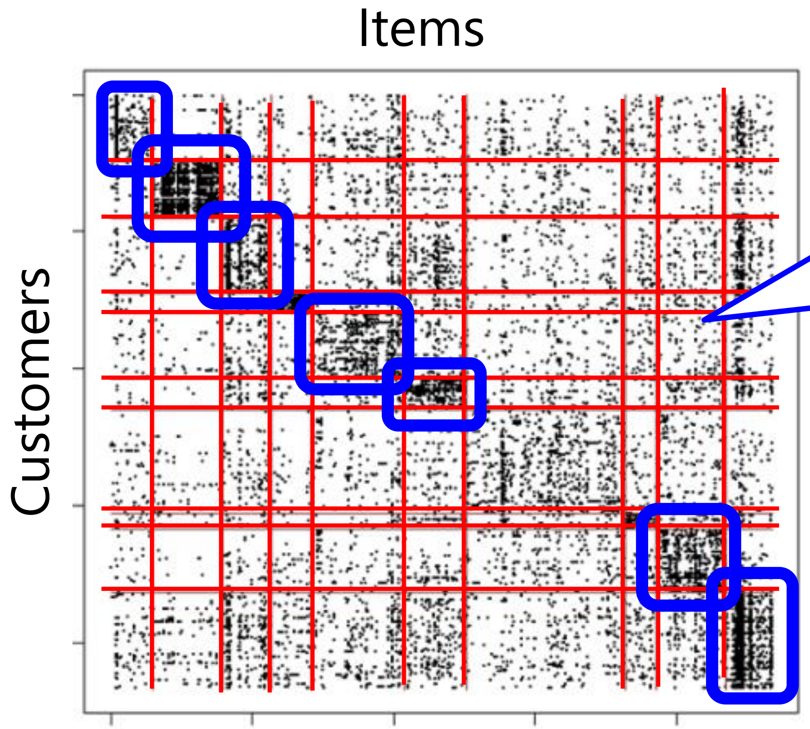
Result

Block structure indicating preference of customers

Applications:
- **Recommendation**
- **Ad Targeting**
- **Market Analysis**

Example

Recommender system for an E-commerce business.



A result obtained from purchase records on EC site

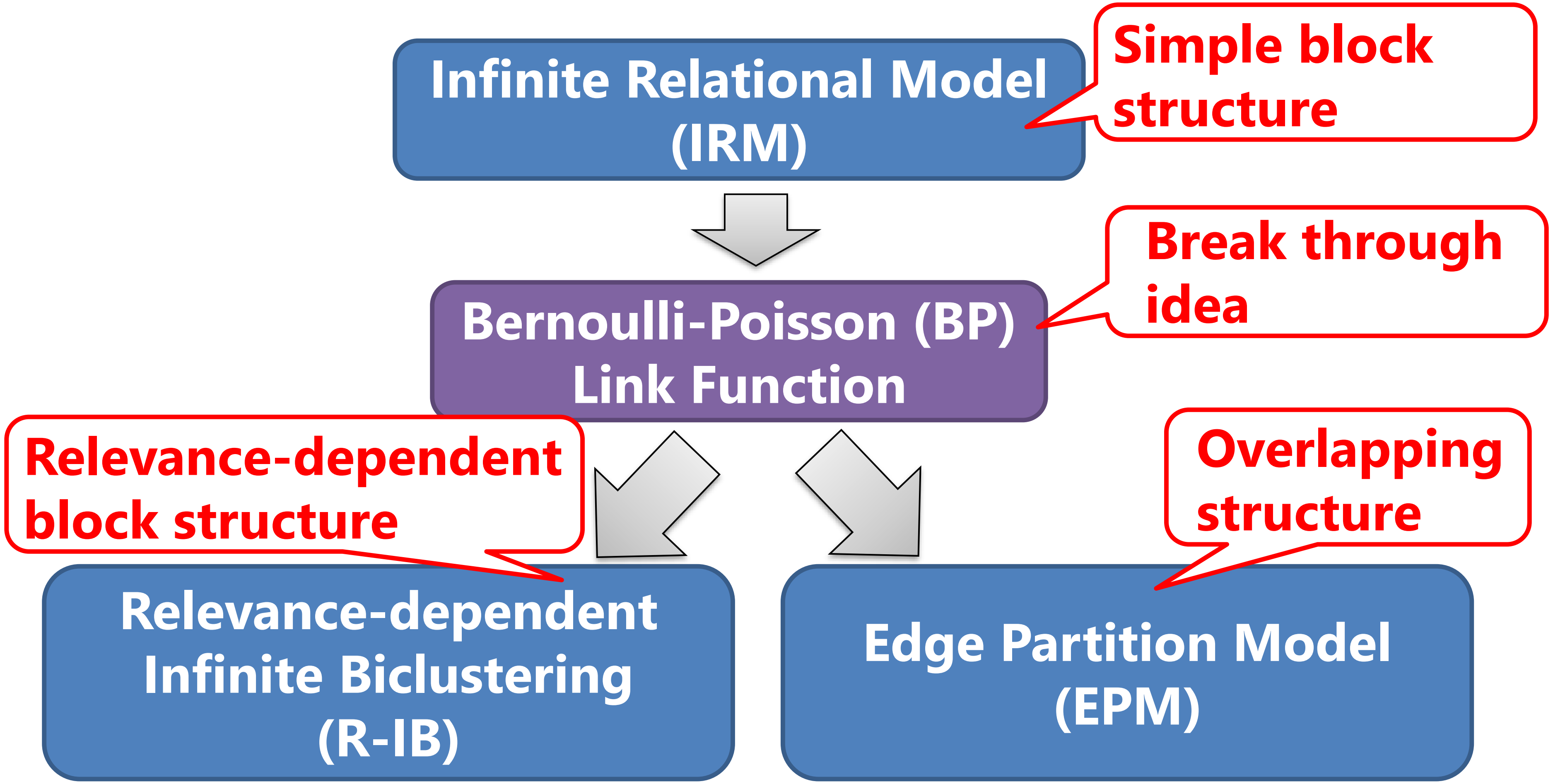
You can find pairs of customer group & item group that sell well.

We can increase total sales by highlighting the items those are strongly related to each target customer.

Relational Data Analysis is a key technology for industries in AI era.

Overview of my Talk

Statistical machine learning models for discovering advanced structure from relational data.



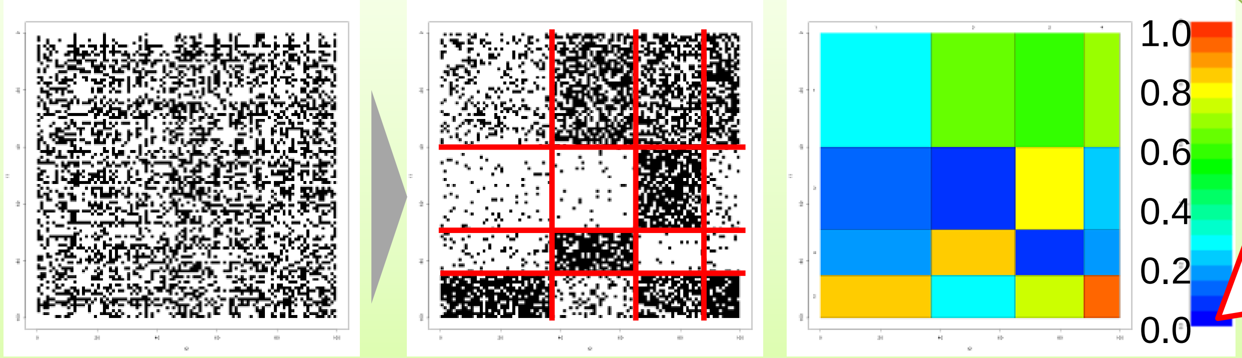
The Infinite Relational Model (IRM) [Kemp+, AAAI06]

A well-known statistical model that extract hidden block structure from noisy real-world relational data.

Model description of the IRM

$$P(x_{i,j} = 1, | z_{1,i} = k, z_{2,j} = l) = p_{k,l}$$

- $x_{i,j}$: Observed Entry between row i and column j
- $z_{1,i}$: Group assignment for i-th row
- $z_{2,j}$: Group assignment for j-th column
- $p_{k,l}$: Link Probability for block (k,l)



Input relational data → IRM solution → Link probabilities

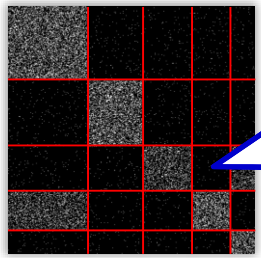
By maximizing $P(z_1, z_2, p|x)$, we can obtain blocks, where each block has uniform density.

Drawbacks of the IRM

The IRM is not acceptable in real-world data analysis.

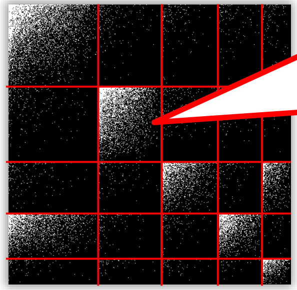
Observation model of the IRM

$$P(x_{i,j} = 1, \mid z_{1,i}, z_{j,2}) = p_{k,l}$$



The IRM assumes each block has a uniform density.

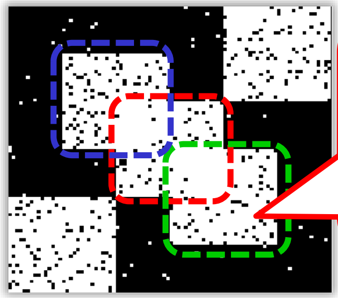
1. Relevance-dependent Blocks



Both active and passive users exist in same community.

(e.g., mail transactions within a SNS community)

2. Overlapping Structure



Some users might belong to multiple groups.

Efficient models for capturing such advanced structure have not been developed for a long time.

Bernoulli-Poisson Link Function

[Zhou+, AISTATS15]

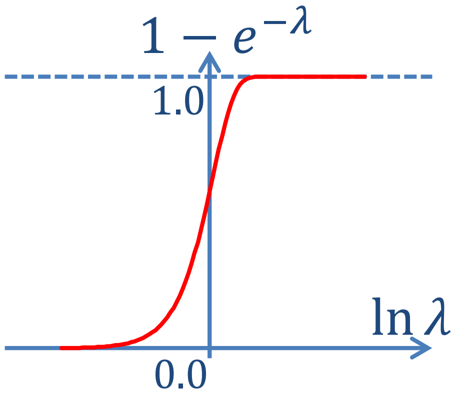
Key Idea: Bernoulli-Poisson Link Function

Modeling link probability using
Bernoulli-Poisson (BP) link function.

BP link function [Zhou, AISTATS15]

$$f(\lambda) = 1 - e^{-\lambda}.$$

λ :non-negative variable in $[0, \infty)$.



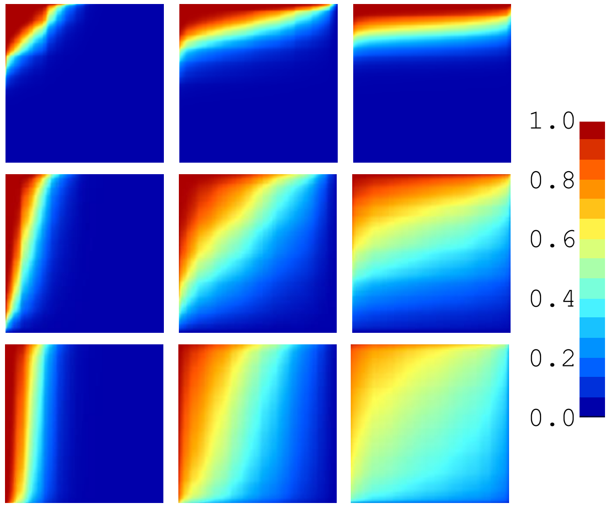
BP link function enable us to define link probabilities using non-negative parameters.

Multiplicative Property of BP Link Function

The parameter λ can be straightforwardly extended to product of multiple parameters.

$$f(\lambda_1 \lambda_2) = 1 - e^{-\lambda_1 \times \lambda_2}.$$

Factor 1 Factor 2



Examples of prob. density

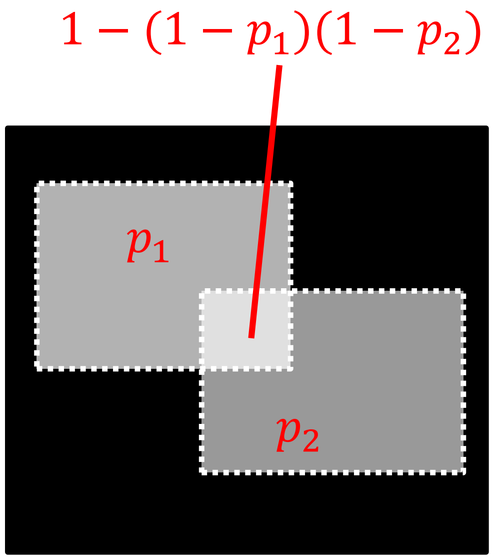
We can easily model the effect of multiple factors for a link probability.

Additive Property of BP Link Function

Sum of multiple parameters also have a useful meaning.

$$
\begin{aligned}
f(\lambda_1 + \lambda_2) &= 1 - e^{-(\lambda_1+\lambda_2)} \\
&= 1 - e^{-\lambda_1} \times e^{-\lambda_2} \\
&= 1 - (1 - f(\lambda_1))(1 - f(\lambda_2)).
\end{aligned}
$$

$P(x = 0 \mid f(\lambda_1))$ $P(x = 0 \mid f(\lambda_2))$



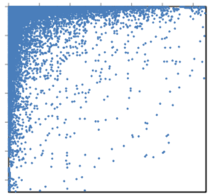
We can easily model the interaction of multiple factors with probabilistic OR manner.

Relevance-dependent Infinite Biclustering (R-IB)

[Ohama+, IJCAI17]

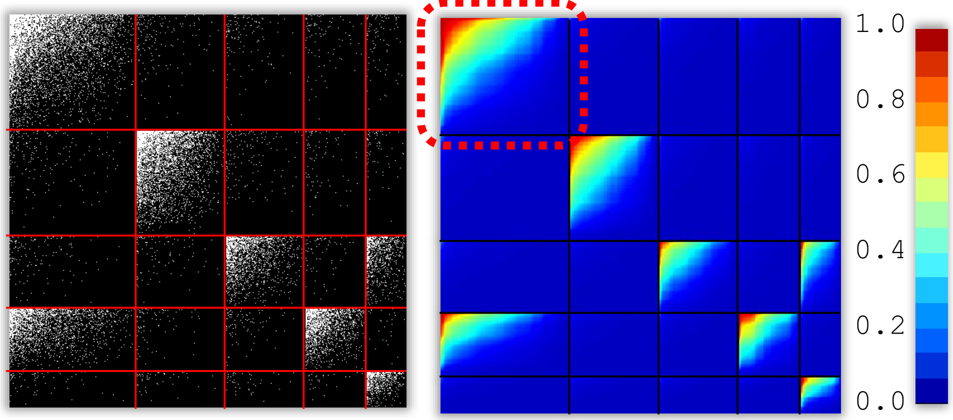
Motivation

Real-world relational data often have relevance-dependent block structure.



[Araujo, ECML14]

relevance-dependent density



Clustering result Density

We can naturally extend the IRM so that it can capture the relevance-dependent block structure.

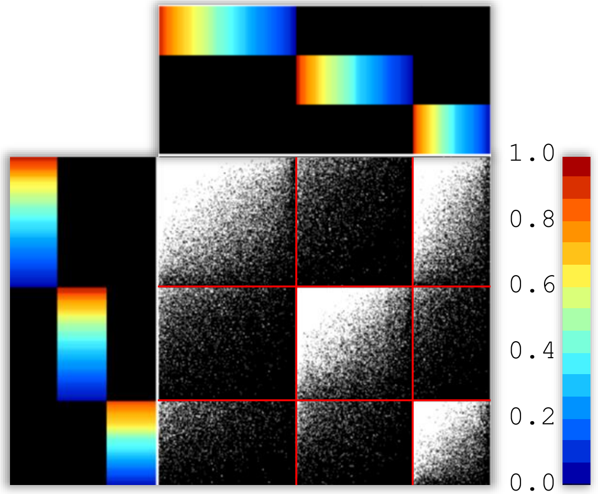
Relevance-dependent Infinite Biclustering (R-IB)

The key strategy is to define a link probability by BP link function with product of three nn-parameters.

R-IB: $P(x_{i,j} = 1, | \boldsymbol{z}) = 1 - e^{-\theta_{1,i}\theta_{2,j}\lambda_{k,l}}$

- $\theta_{1,i}$: Relevance (i-th row)
- $\theta_{2,j}$: Relevance (j-th column)
- $\lambda_{k,l}$: Typical link strength (block (k,l))

In R-IB model, link probability for an entry is modulated by related relevance parameters.



Example of R-IB solution
(colors indicate the relevance)

R-IB can simultaneously estimate the block structure and relevance values for each row and column object.

Remarkable Property of R-IB Inference

Most of parameters of the R-IB can be integrated out.

$$x_{i,j} \sim \mathrm{Bernoulli}\left(1 - e^{-\theta_{1,i}\theta_{2,j}\lambda_{k,l}}\right).$$

Equivalently rewritten by truncating Poisson random count! [Zhou, 2015]

$$x^*_{i,j} \sim \mathrm{Poisson}(\theta_{1,i}\theta_{2,j}\lambda_{k,l}),$$

$$x_{i,j} = \begin{cases} 0 & \text{if } x^*_{i,j} = 0, \\ 1 & \text{if } x^*_{i,j} \geq 1. \end{cases}$$

Remarkable Property of R-IB Inference

Most of parameters of the R-IB can be integrated out.

$$\boldsymbol{\theta}_1/I \sim \text{Dirichlet}(\overbrace{c_1, \ldots, c_1}^{I}),$$

$$\boldsymbol{\theta}_2/J \sim \text{Dirichlet}(\overbrace{c_2, \ldots, c_2}^{J})$$

$$\lambda \sim \text{Gamma}(a, b)$$

Prior

Prior

$$x^*_{i,j} \sim \text{Poisson}(\theta_{1,i}\theta_{2,j}\lambda_{k,l}),$$

$$x_{i,j} = \begin{cases} 0 & \text{if } x^*_{i,j} = 0, \\ 1 & \text{if } x^*_{i,j} \geq 1. \end{cases}$$

We can derive analytical marginal likelihood function☺

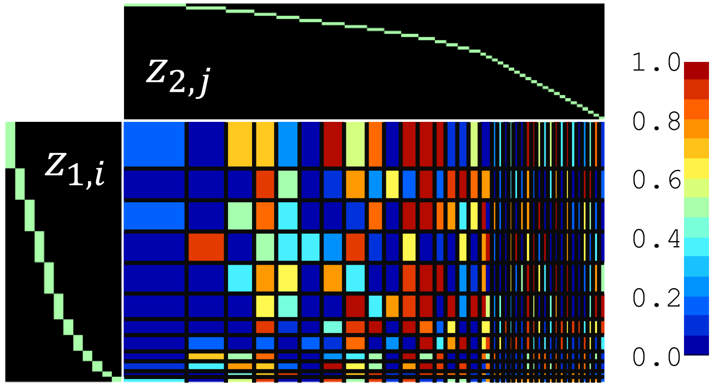
$\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\lambda}$ can be integrated out

$$P(\boldsymbol{x}^* \mid a, b, c_1, c_2) = \frac{1}{\prod_{i,j} x^*_{i,j}!} \times \prod_i \frac{\Gamma(c_1 + M_{i,\cdot})}{\Gamma(c_1)} \times \prod_j \frac{\Gamma(c_2 + M_{\cdot,j})}{\Gamma(c_2)} \times \frac{I^M \Gamma(Ic_1)}{\Gamma(Ic_1 + M)} \times \frac{J^M \Gamma(Jc_2)}{\Gamma(Jc_2 + M)} \times \frac{G(a + M, b + IJ)}{G(a, b)},$$

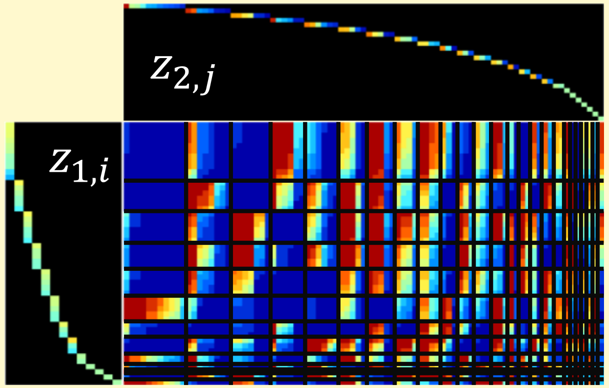
Inference for R-IB can be efficiently performed by optimizing only remaining parameters z_1 & z_2 ☺

Experimental Result using Animal Dataset

Relationships between 50 mammals & 85 attributes



IRM solution



(Proposed) R-IB solution

Row Groups

Aquatic

Name	Relevance
H. Whale	**(1.175)**
Seal	(1.151)
Walrus	(1.017)
Dolphin	(0.837)
B. whale	(0.820)

Big

Name	Relevance
Elephant	**(1.303)**
Rhinoceros	(0.922)
Hippopotamus	(0.775)

Carnivorous

Name	Relevance
Wolf	**(1.379)**
Leopard	(1.172)
⋮	⋮
Grizzly bear	(0.830)
G. shepherd	(0.666)

Apes

Name	Relevance
Chimpanzee	**(1.566)**
Gorilla	(0.804)
S. monkey	(0.630)

Column Groups

Aquatic

Name	Relevance
(Eat) fish	**(6.318)**
Water	(1.111)
Swims	(1.023)
⋮	⋮
Blue	(0.188)
Plankton	(0.156)
Skimmer	(0.079)

Carnivorous

Name	Relevance
Meatteeth	**(2.036)**
Fierce	(1.967)
⋮	⋮
Cave	(0.102)
Stalker	(0.088)

Herbivorous

Name	Relevance
Vegetation	**(4.492)**
Fields	(0.585)
⋮	⋮
Longneck	(0.049)
Horns	(0.045)

Top-ranked objects facilitate understanding what each group means ☺

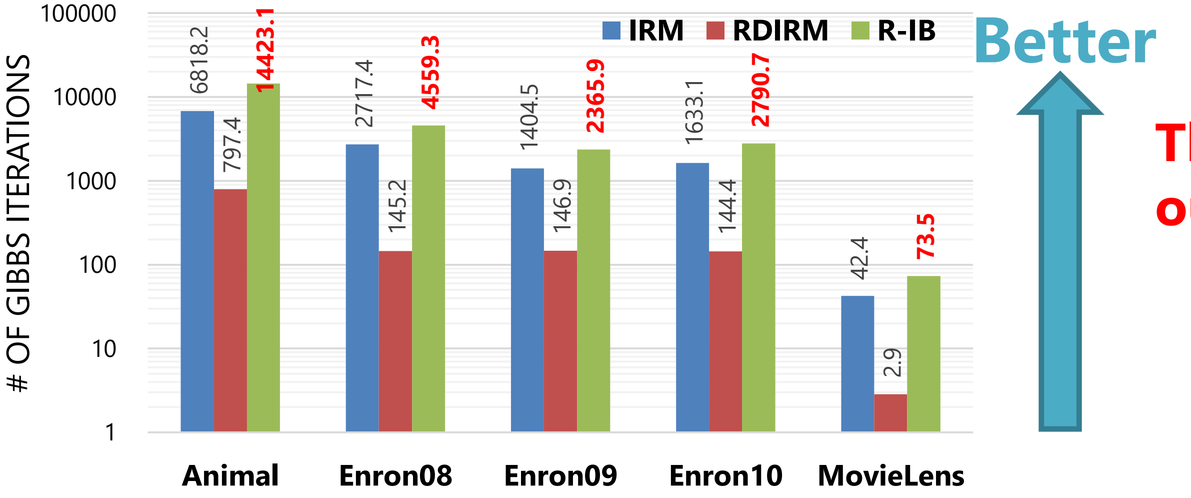
Quantitative Results

◆Link prediction accuracy (AUC of precision-recall curve)

Dataset	IRM	RDIRM	R-IB
Animal	**0.8106**±0.0356	0.7515±0.0525	0.8023±0.0255
Enron08	0.2736±0.0693	0.2044±0.0482	**0.2885**±0.0741
Enron09	0.2710±0.0488	0.2129±0.0447	**0.2964**±0.0546
Enron10	0.3519±0.0398	0.3100±0.0327	**0.3810**±0.0421
MovieLens	0.4098±0.0062	0.4130±0.0061	**0.4473**±0.0058

The R-IB indicated better link prediction accuracy especially for larger datasets.

◆Computational Efficiency (# of Learning Epochs / 5 min.)



The R-IB significantly outperformed the IRM.

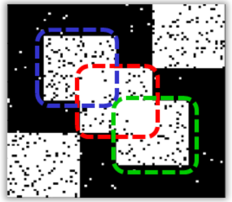
The R-IB can extract more essential structure with better computational efficiency ☺

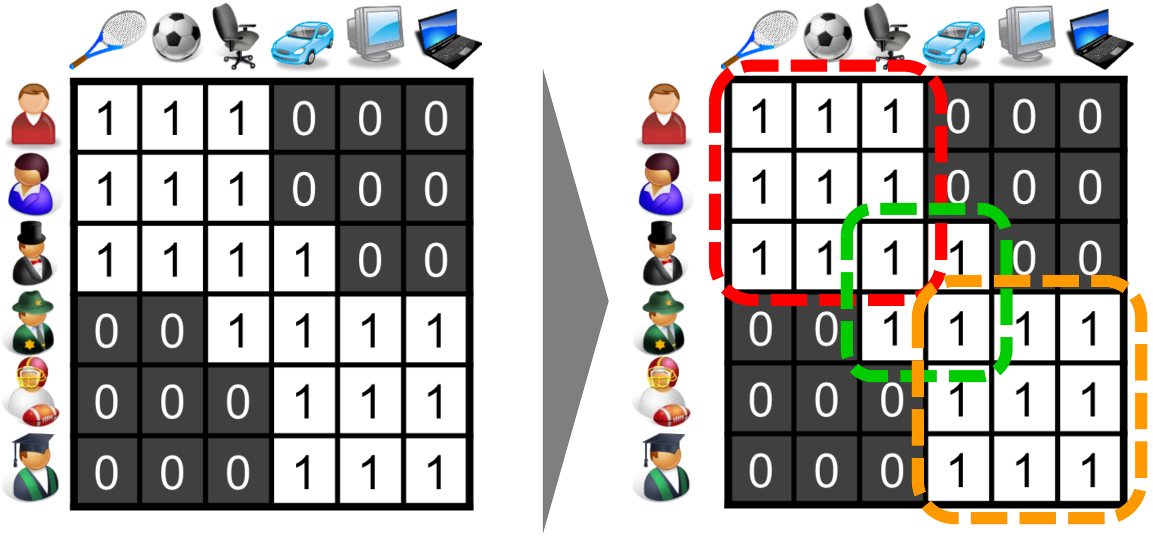
Gamma Process
Edge Partition Model (EPM)

[Zhou, AISTATS15] [Ohama+, NIPS17]

Motivation

Real-world relational data often have overlapping block structure.





$I \times J$ Relational Data $\boldsymbol{x}$ Overlapping Structure

BP link function is also useful for developing relational model for overlapping structure.

Gamma Process Edge Partition Model (EPM)

The key strategy is to utilize additive property of BP link function.

EPM model

Relevance for i-th row **to k-th block**

Relevance for j-th column **to k-th block**

$$P(x_{i,j} \mid \boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{\lambda}) = 1 - e^{-\sum_{k=1}^{K} \phi_{i,k}\psi_{j,k}\lambda_k}$$

Typical link strength for k-th block.

$$= 1 - \prod_{k=1}^{K} e^{\phi_{i,k}\psi_{j,k}\lambda_k}$$

Probability that at least one of K blocks generate a link.

Probability that k-th block DOES NOT generate a link.

EPM can capture overlapping block structure with probabilistic OR manner.

Efficient Inference for the EPM

Similar to the R-IB, model parameters of the EPM can be integrated out.

Marginal likelihood function for the EPM

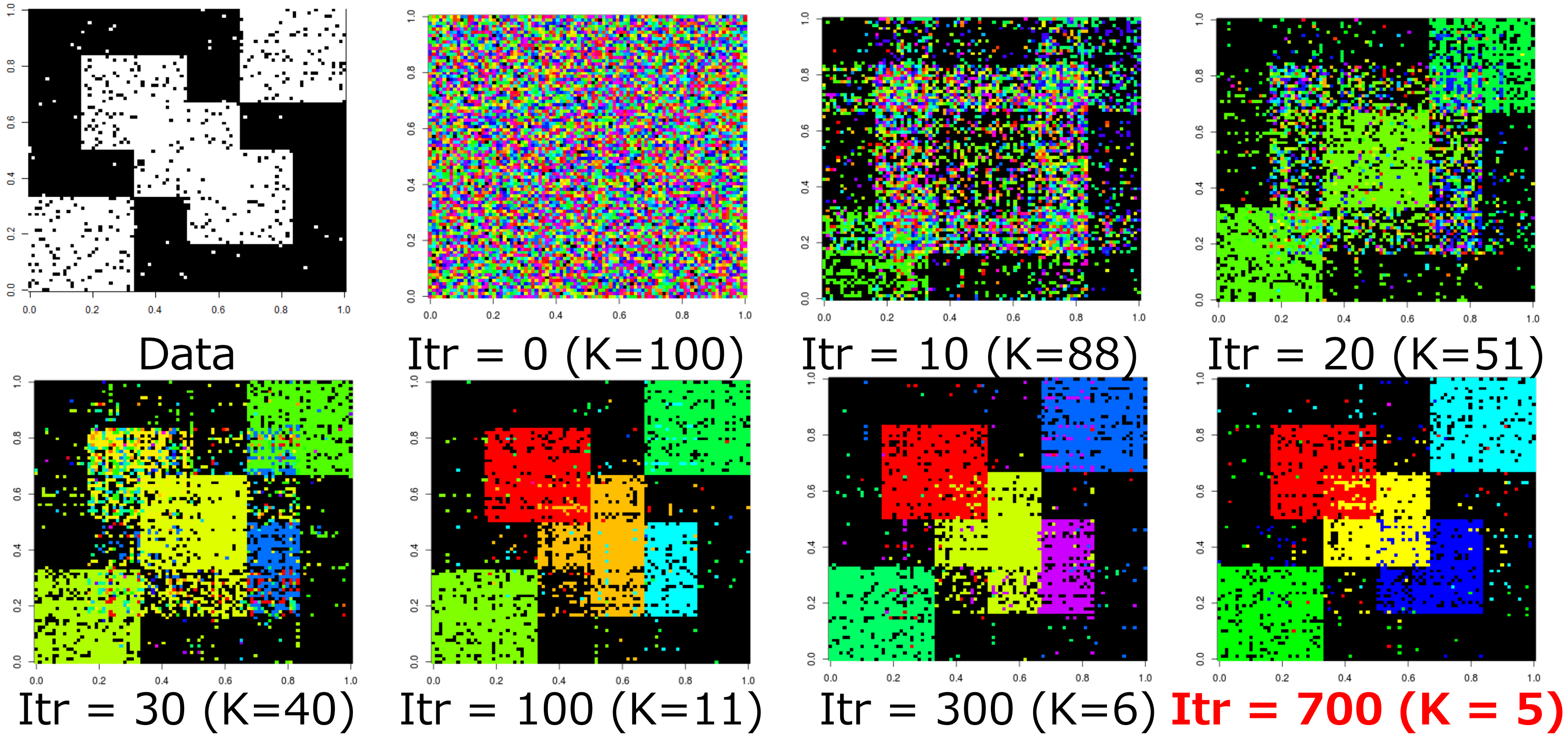
$$P(\boldsymbol{m},[\boldsymbol{z}])_\infty = \prod_{i=1}^{I}\prod_{j=1}^{J}\frac{1}{m_{i,j,\cdot}!}\times\prod_{k=1}^{K_+}\frac{\Gamma(I\alpha_1)}{\Gamma(I\alpha_1+m_{\cdot,\cdot,k})}\prod_{i=1}^{I}\frac{\Gamma(\alpha_1+m_{i,\cdot,k})}{\Gamma(\alpha_1)}$$

$$\times\prod_{k=1}^{K_+}\frac{\Gamma(J\alpha_2)}{\Gamma(J\alpha_2+m_{\cdot,\cdot,k})}\prod_{j=1}^{J}\frac{\Gamma(\alpha_2+m_{\cdot,j,k})}{\Gamma(\alpha_2)}\times\gamma_0^{K_+}\left(\frac{c_0}{c_0+1}\right)^{\gamma_0}\prod_{k=1}^{K_+}\frac{\Gamma(m_{\cdot,\cdot,k})}{(c_0+1)^{m_{\cdot,\cdot,k}}},$$

We no longer have to explicitly estimate the parameters for the EPM!

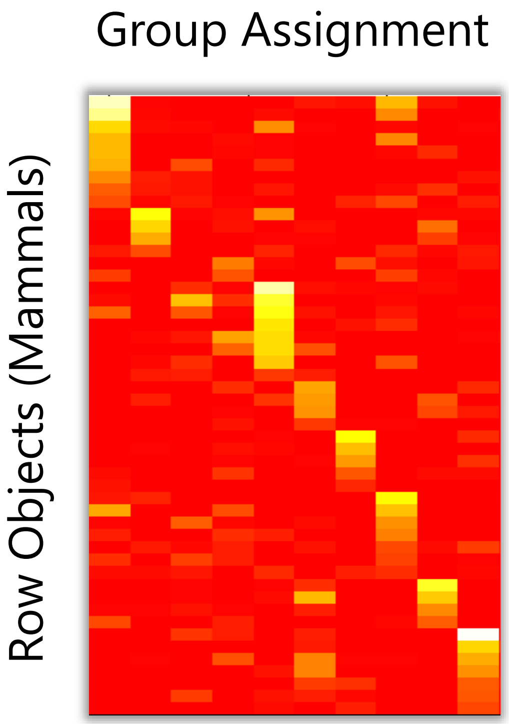
EPM can be inferred efficiently by optimizing its marginal likelihood function.

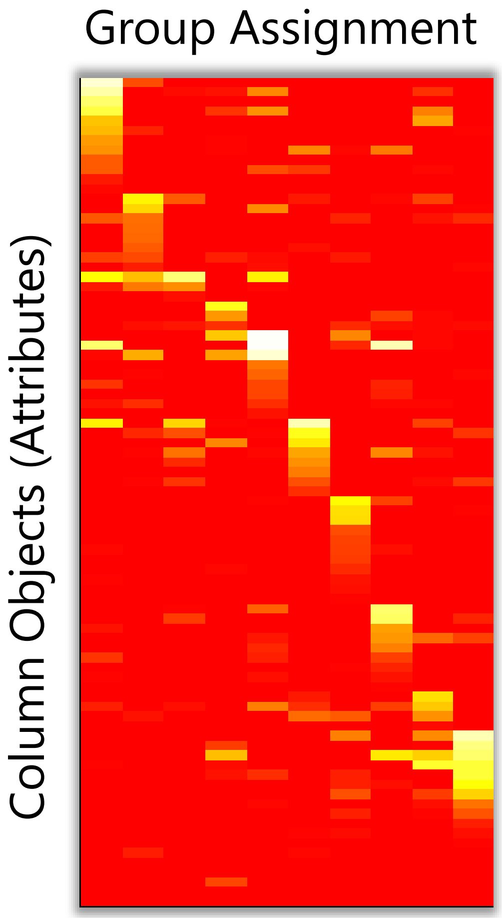
Result



The EPM accurately discover overlapping block structure underlying the data.

Result on Animal Dataset





The EPM is more helpful for finding deeper insights from noisy real-world data.

Conclusions

- We addressed that discovering hidden structure from relational data is an important technical problem for many industries.
- We introduced Bernoulli-Poisson (BP) link function, which is a great idea to capture advanced structure underlying the data.
- We presented two novel relational model:
  - R-IB for relevance-dependent block structure
  - EPM for overlapping structure
- Using the BP modeling to bridge the statistical modeling and deep neural networks is a promising aspect for future work.

Acknowledgement

- Prof. Hiroki Arimura (Hokkaido Univ.)
- Assoc. Prof. Takuya Kida (Hokkaido Univ.)
- Lecturer Issei Sato (Univ. of Tokyo)

Thank you for your kind attention