Bayesian nonparametric comorbidity analysis of psychiatric disorders

Francisco J. R. Ruiz[(1)], Isabel Valera[(1)],
Carlos Blanco[(2)], Fernando Perez-Cruz[(1,3)]

[(1)]Department of Signal Theory and Communications, University Carlos III in Madrid
[(2)]Department of Psychiatry, New York State Psychiatric Institute Columbia University
[(3)]Machine Learning Scientist at Amazon

Outline

Outline

Introduction

- Individuals with multiple coexisting diseases.
- 80% of US Medicare spending devoted to patients with 4+ chronic conditions.
- Impact of comorbidity: mortality, quality of life, quality of health care, . . .
- Psychiatry: etiological and treatment implications.

Introduction

Goal

Find out the latent relationship among psychiatric disorders.

Database

NESARC database (*National Epidemiologic Survey on Alcohol and Related Conditions*):

- Samples the U.S. population.
- Around 3K questions and 43K subjects.
- Mainly yes-or-no questions, and some multiple-choice and questions with ordinal answers.

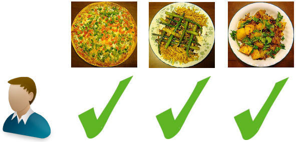
Our approach

Latent feature modeling and Indian buffet process (IBP).

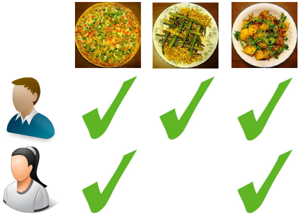
Outline

Indian Buffet Process

Indian Buffet Process



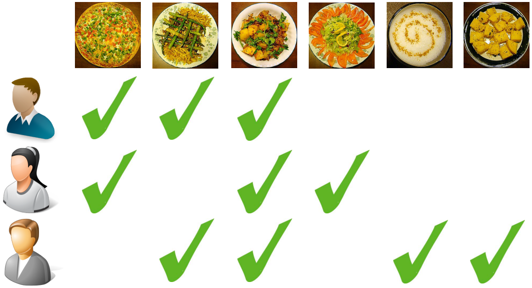
Indian Buffet Process



Indian Buffet Process

Indian Buffet Process

Indian Buffet Process



Indian Buffet Process

Indian Buffet Process

. . .

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

⋮

Indian Buffet Process

. . .

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

⋮

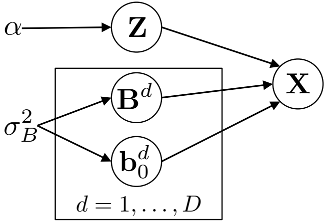
Indian Buffet Process

- Prior distribution over binary matrices.
- Number of columns (features) $K \to \infty$.
- Matrix $\mathbf{Z}_{N \times K} \sim \mathrm{IBP}(\alpha)$.
- Finite N implies finite number of non-zero columns K_+.

Outline

Observation Model

- Categorical observations: $x_{nd} \in \{1, \ldots, R\}$ ('Yes', 'No', 'Unkown', 'Blank', . . .).
- How to link latent features $\mathbf{z}_{n\cdot}$ and observations?



Observation Model

- Multiple-logistic (softmax) function:

$$p(x_{nd} = \text{`yes'} | \mathbf{z}_{n\cdot}, \mathbf{b}_0^d, \mathbf{B}^d) \propto \exp\left(b_{0\text{yes}}^d + \mathbf{z}_{n\cdot}\mathbf{b}_{\cdot\text{yes}}^d\right),$$
$$p(x_{nd} = \text{`no'} | \mathbf{z}_{n\cdot}, \mathbf{b}_0^d, \mathbf{B}^d) \propto \exp\left(b_{0\text{no}}^d + \mathbf{z}_{n\cdot}\mathbf{b}_{\cdot\text{no}}^d\right),$$
$$\cdots$$

$$p(x_{nd} = r | \mathbf{z}_{n\cdot}, \mathbf{b}_0^d, \mathbf{B}^d) = \frac{\exp\left(b_{0r}^d + \mathbf{z}_{n\cdot}\mathbf{b}_{\cdot r}^d\right)}{\sum_{r'=1}^{R} \exp\left(b_{0r'}^d + \mathbf{z}_{n\cdot}\mathbf{b}_{\cdot r'}^d\right)}, \quad r = 1, \ldots, R.$$

- The weighting factors are placed Gaussian priors.

Observation Model

Given the latent feature matrix $\mathbf{Z}$ and the weighting factors $\mathbf{B}^d$ and $\mathbf{b}_0^d$, the elements in $\mathbf{X}$ are independent:

$$p(\mathbf{X}|\mathbf{Z}, \mathbf{B}^1, \ldots, \mathbf{B}^D, \mathbf{b}_0^1, \ldots, \mathbf{b}_0^D) = \prod_{n=1}^{N} \prod_{d=1}^{D} p(x_{nd}|\mathbf{z}_{n\cdot}, \mathbf{B}^d, \mathbf{b}_0^d).$$

Outline

Inference

- The model is not conditionally conjugate.
- Posterior on the weighting factors:

$$\overbrace{p(\mathbf{B}^d, \mathbf{b}_0^d | \mathbf{X}, \mathbf{Z})}^{\text{Non Gauss}} = \frac{\overbrace{p(\mathbf{x}_{\cdot d} | \mathbf{B}^d, \mathbf{b}_0^d, \mathbf{Z})}^{\text{Non Gauss}} \overbrace{p(\mathbf{B}^d) p(\mathbf{b}_0^d)}^{\text{Gauss}}}{p(\mathbf{x}_{\cdot d} | \mathbf{Z})}.$$

- The integral in the denominator is intractable:

$$p(\mathbf{x}_{\cdot d} | \mathbf{Z}) = \int p(\mathbf{x}_{\cdot d} | \mathbf{B}^d, \mathbf{b}_0^d, \mathbf{Z}) p(\mathbf{B}^d) p(\mathbf{b}_0^d) d\mathbf{B}^d d\mathbf{b}_0^d.$$

Gibbs sampling

Gibbs sampling

- Iteratively sample each element of the IBP matrix.
- Integrate out all the weighting factors $\Rightarrow$ Intractable.
- Gaussian approximation of the posterior:
  1. Laplace approximation.
  2. Expectation propagation.
     - Multinomial probit likelihood instead of softmax.

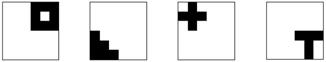
Variational inference

Variational inference

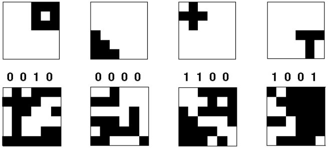
- Computationally less expensive than MCMC.
- Solve a non-convex optimization problem.
- Coordinate ascent.
- Has to deal with non-conjugacy arising from:
  - Multiple-logistic likelihood.
    - Bound the ELBO through a first order Taylor series expansion.
  - Stick-breaking construction of the IBP.
    - Bound the ELBO applying Jensen's inequality.

Outline

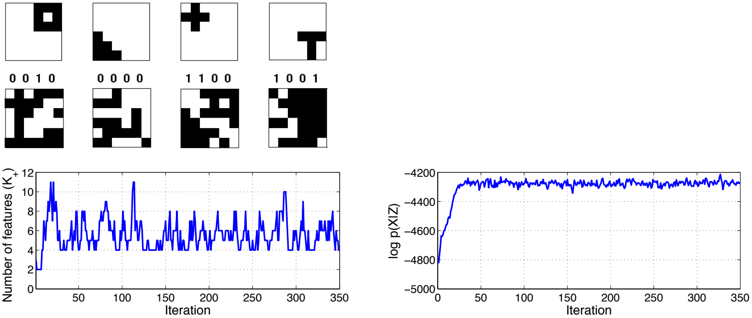
Toy Example



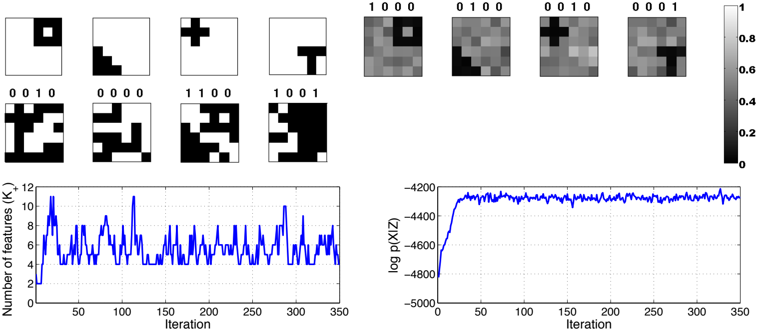
Toy Example



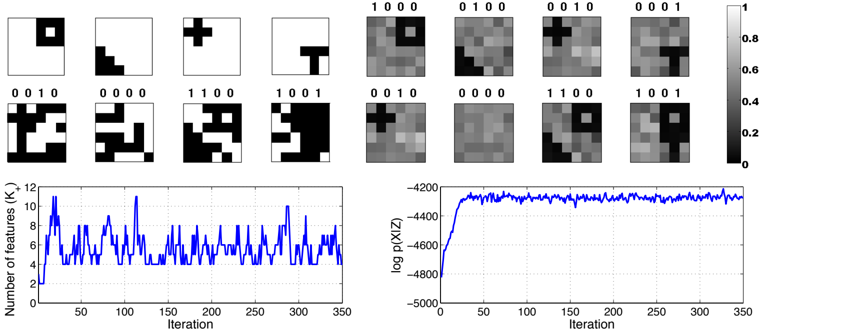
Toy Example


0 0 1 0
0 0 0 0
1 1 0 0
1 0 0 1
Number of features (K_+)
Iteration
log p(X|Z)
Iteration

Toy Example


1 0 0 0
0 1 0 0
0 0 1 0
0 0 0 1
0 0 1 0
0 0 0 0
1 1 0 0
1 0 0 1
Number of features (K_+)
Iteration
log p(X|Z)
Iteration

Toy Example



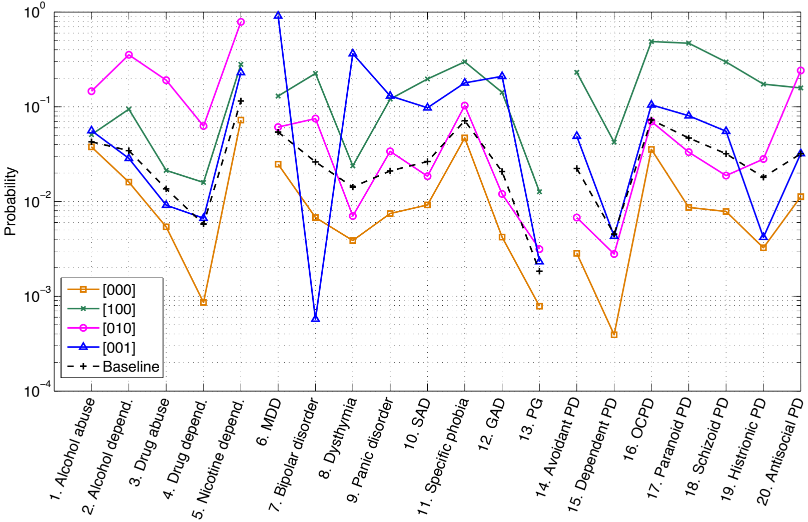
1 0 0 0
0 1 0 0
0 0 1 0
0 0 0 1
0 0 1 0
0 0 0 0
1 1 0 0
1 0 0 1
Number of features (K_+)
Iteration
log p(X|Z)

Experiments on NESARC

Real data

- Inputs: Diagnoses of 20 common psychiatric disorders.
- Previous studies: Factor analysis.
  - Specify the number of factors.
  - Assume Gaussian observations.
  - 3 latent factors seem enough.

Experiments on NESARC



Outline

Conclusions

Summary

- Likelihood model for the IBP with categorical observations.

Inference Algorithm	Likelihood
Gibbs sampling + Laplace approximation	Softmax
Gibbs sampling + Nested EP	Multinomial probit
Variational inference	Softmax

- Results are in agreement with previous work and also provide new information.

Ongoing work

- Individual-specific severity terms:

$$p(x_{nd} = r | \mathbf{w}_{n\cdot}, \mathbf{b}_0^d, \mathbf{B}^d) \propto \exp\left(b_{0r}^d + \mathbf{w}_{n\cdot}\mathbf{b}_{\cdot r}^d\right), \quad r = 1, \ldots, R.$$

- Instead of on/off features, each term in $[0, 1]$.

Outline

References

C. Blanco, R. F. Krueger, D. S. Hasin, S. M. Liu, S. Wang, B. T. Kerridge, T. Saha, and M. Olfson.
Mapping common psychiatric disorders: Structure and predictive validity in the National Epidemiologic Survey on Alcohol and Related Conditions.
Journal of the American Medical Association Psychiatry, 70(2):199–208, 2013.

D. M. Blei and J. D. Lafferty.
A correlated topic model of Science.
Annals of Applied Statistics, 1(1):17–35, August 2007.

F. Doshi-Velez, K. T. Miller, J. Van Gael, and Y. W. Teh.
Variational inference for the Indian buffet process, 2009.

T. L. Griffiths and Z. Ghahramani.
The Indian buffet process: an introduction and review.
Journal of Machine Learning Research, 12:1185–1224, 2011.

J. Riihimäki, P. Jylänki, and A. Vehtari.
Nested expectation propagation for Gaussian process classification with a multinomial probit likelihood.
Journal of Machine Learning Research, 14:75–109, 2013.

F. J. R. Ruiz, I. Valera, C. Blanco, and F. Perez-Cruz.
Bayesian nonparametric modeling of suicide attempts.
Advances in Neural Information Processing Systems (NIPS), 25:1862–1870, 2012.

F. J. R. Ruiz, I. Valera, C. Blanco, and F. Perez-Cruz.
Bayesian nonparametric comorbidity analysis of psychiatric disorders.
Journal of Machine Learning Research (Submitted), 2013.

C. K. I. Williams and D. Barber.
Bayesian classification with Gaussian Processes.
IEEE Transactions on Pattern Analysis and Machine Intelligence, 20:1342–1351, 1998.