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

University of Sheffield
July 9, 2014

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).

Introduction

Bayesian Nonparametrics

- Unbounded number of degrees of freedom in a model.
- E.g., clustering with unknown number of clusters.
- The posterior distribution chooses the complexity of the model to fit the data.

Introduction

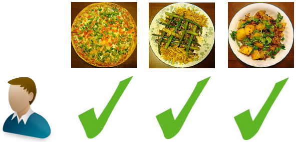
Bayesian Nonparametrics

- Unbounded number of degrees of freedom in a model.
- E.g., clustering with unknown number of clusters.
- The posterior distribution chooses the complexity of the model to fit the data.
- Applications (I've worked on / Would like to work on):
  - Psychiatry.
  - Power disaggregation.
  - Recommendation systems.
  - Multiuser MIMO channel estimation.
  - Channel coding.
  - Sports.
  - NYC marathon modeling.
  - Higgs boson challenge?

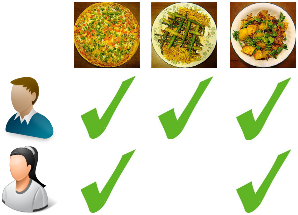
Outline

Indian Buffet Process

Indian Buffet Process



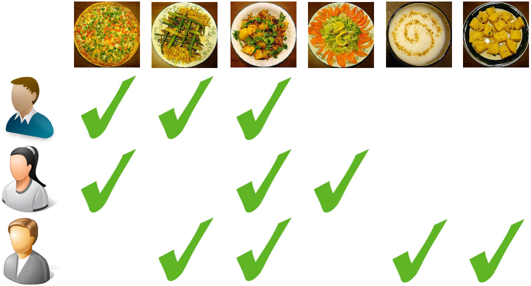
Indian Buffet Process



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

- Each subject characterized by a binary vector of latent features.

	$\mathbf{z}_n$	x_{nd}
	$[1 \quad 0 \quad 1 \quad 1 \quad \cdots]$	-1.05
	$[0 \quad 0 \quad 0 \quad 1 \quad \cdots]$	0.15
	$[1 \quad 0 \quad 0 \quad 1 \quad \cdots]$	1.35

Observation Model

- Each subject characterized by a binary vector of latent features.
- Under Gaussian observations, features are weighted and added.

	$\mathbf{z}_n$		$\mathbf{b}^d$		x_{nd}
	$[1 \quad 0 \quad 1 \quad 1 \quad \cdots]$				-1.05
	$[0 \quad 0 \quad 0 \quad 1 \quad \cdots]$	$\times$	$\begin{bmatrix} 1.2 \\ 3.2 \\ -2.4 \\ 0.15 \\ \vdots \end{bmatrix}$	$=$	0.15
	$[1 \quad 0 \quad 0 \quad 1 \quad \cdots]$				1.35

Observation Model

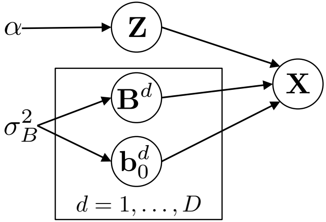
- Each subject characterized by a binary vector of latent features.
- Under Gaussian observations, features are weighted and added.

$$
\begin{array}{ccccc}
\mathbf{z}_n & & \mathbf{b}^d & & x_{nd} \\
[1 \quad 0 \quad 1 \quad 1 \quad \cdots] & & \left[\begin{array}{c} 1.2 \\ 3.2 \\ -2.4 \\ 0.15 \\ \vdots \end{array}\right] & & -1.05 \\
[0 \quad 0 \quad 0 \quad 1 \quad \cdots] & \times & & = & 0.15 \\
[1 \quad 0 \quad 0 \quad 1 \quad \cdots] & & & & 1.35
\end{array}
$$

$$\mathbf{ZB} + \text{noise} = \mathbf{X}$$

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:

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

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

- The weighting factors are placed Gaussian priors.

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.

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:

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

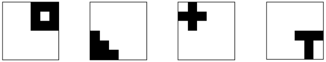
Gibbs sampling

Gibbs sampling

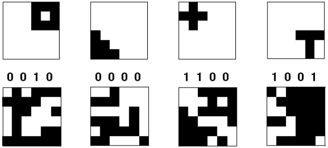
- Iteratively sample each element of the IBP matrix.
- Integrate out all the weighting factors $\Rightarrow$ Intractable.
- Gaussian approximation of the posterior:
  1. Laplace approximation.
  2. Expectation propagation.
     - Multinomial probit likelihood instead of softmax.

Outline

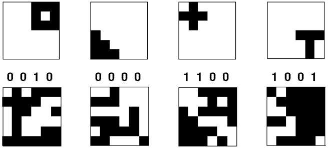
Toy Example

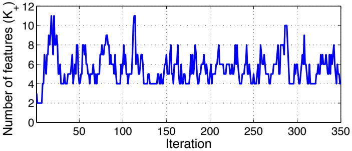


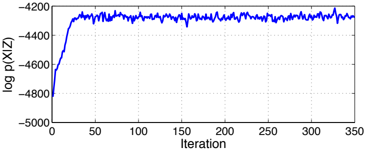
Toy Example



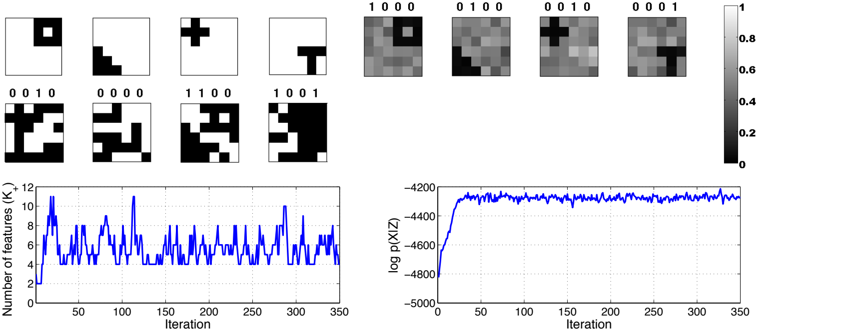
Toy Example


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

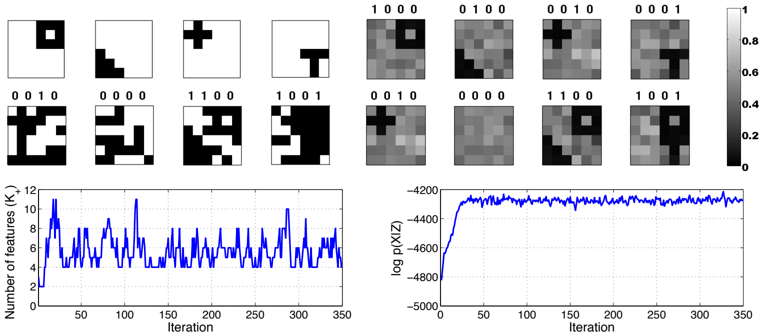

Number of features (K+)
Iteration
0
2
4
6
8
10
12
50
100
150
200
250
300
350


log p(X|Z)
Iteration
−4200
−4400
−4600
−4800
−5000
0
50
100
150
200
250
300
350

Toy Example


0 0 1 0
0 0 0 0
1 1 0 0
1 0 0 1
1 0 0 0
0 1 0 0
0 0 1 0
0 0 0 1
1
0.8
0.6
0.4
0.2
0
Number of features (K_+)
Iteration
log p(X|Z)
Iteration

Toy Example

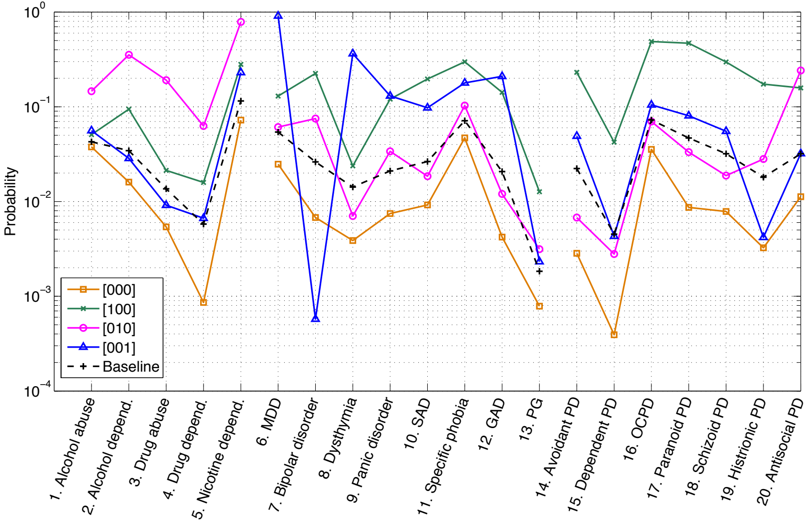

1 0 0 0
0 1 0 0
0 0 1 0
0 0 0 1
0 0 1 0
0 0 0 0
1 1 0 0
1 0 0 1
0 0 1 0
0 0 0 0
1 1 0 0
1 0 0 1
1
0.8
0.6
0.4
0.2
0
Number of features (K_+)
Iteration
log p(X|Z)
Iteration

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



Experiments on NESARC

Extension of the model

- 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 $[0, 1]$.

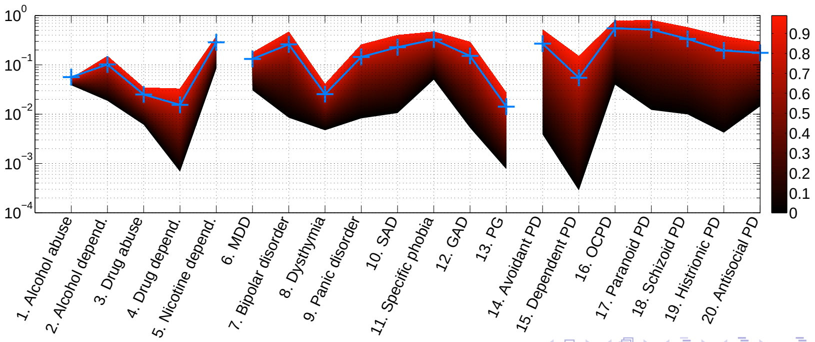
Experiments on NESARC

Extension of the model

- 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

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.

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.

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, 15:1215–1247, April 2014.

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