Joint Bayesian Modelling of Internal Dependencies and Relevant Multimorbidities of a Heterogeneous Disease

Peter Marx

September 6, 2016

Budapest University of Technology and Economics,
Deparment of Measurement and Information Systems

Table of contents

Goals

- Examine relevant factors in a multitarget setup
  - Large sample size
  - Heterogeneous targets (depression)

Unitarget (UT): Choosing one representative target

Individual targets (IT): Multiple analyses using individual targets

Multitarget (MT): Multitarget representation of the disease group

Introduction

- Large sample size 117.392 (UK Biobank)
- 109 variables/factors (disorders)
- Heterogeneous target variable (depression)
- Prevailing approaches:
    - Standard (pairwise) statistics
    - Associative connections

A PGM VIEW

- Bayesian Multilevel Analysis [1]
- Global features [2, 3, 4]:
  - Markov Blanket Membership (MBM)
  - Markov Blanket Set (MBS)
  - Markov Blanket Graph (MBG)

Effect of single target

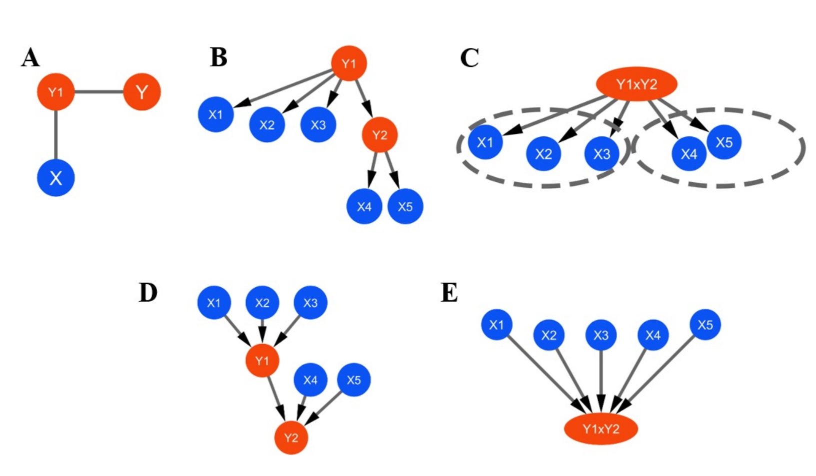

Figure 1: Effects of single target in relevance analysis.

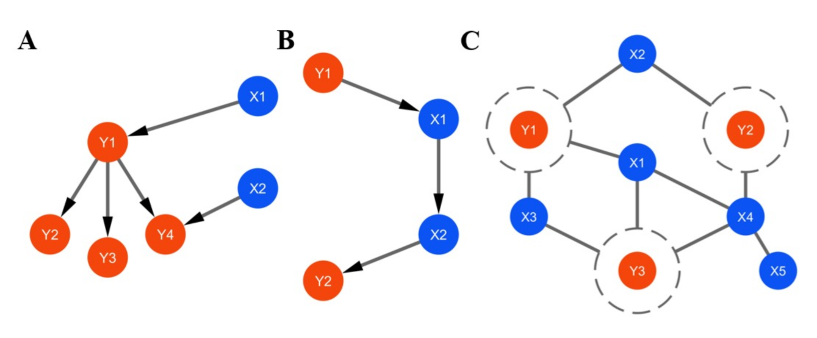

Figure 2: Effects of multiple separated analyses in relevance analysis.

Approach

Definition: k-subMBS, k-supMBS

For a distribution $p(\boldsymbol{V})$ with Markov Boundary set mbs, a set of variables $\boldsymbol{s}$ is called sub-relevant if it is a k-ary Markov Boundary subset (k-subMBS), i.e. $|\boldsymbol{s}| = k$ and $\boldsymbol{s} \subseteq$ mbs. A set of variables $\boldsymbol{s}$ is called sup-relevant if it is a k-ary Markov Boundary superset (k-supMBS), i.e. $|\boldsymbol{s}| = k$ and mbs $\subseteq \boldsymbol{s}$.

$$\underline{p}(s|D_N) = p(\mathrm{MBS}(Y,G) = s|D_N) + \sum_{s':s\subset s'} p(\mathrm{MBS}(Y,G) = s'|D_N)$$

Approximations I

Table 1: Approximations for the 3 scenarios

	MT MBS	MT MBM
MT	MBM	
IT	MBS	MBM
UT	MBM	MBM

Approximations II

MBM approximation:

$$P(\mathrm{MBM}(X_i, \mathbf{Y}) | D_N) \approx 1 - \prod_j \left(1 - P(\mathrm{MBM}(X_i, Y_j) | D_N)\right) \tag{1}$$

MBS approximation by utilizing single target MBM:

$$P(\mathrm{MBS}_i(\mathbf{Y})) \approx \prod_{X_i \in \mathrm{MBS}_i} P(\mathrm{MBM}(X_i, \mathbf{Y})) * \prod_{X_i \notin \mathrm{MBS}_i} (1 - P(\mathrm{MBM}(X_i, \mathbf{Y}))) \tag{2}$$

MBS approximation using single target MBS:

$$P(\mathrm{MBS}_k(\mathbf{Y})) \approx \sum_{\{I_i\}_{i=1}^n : \cup \mathrm{MBS}_{I_i}(Y_i) = \mathrm{MBS}_k} \prod_{i=1}^n P(\mathrm{MBS}_{I_i}(Y_i)) \tag{3}$$

Interaction redundancy score

Definition: Interaction redundancy score (IRS)

$$IRS(X_1; X_2) = \log \frac{p(\{X_1, X_2\} \subseteq \mathrm{MBS}(\mathbf{Y}))}{p(\mathrm{MBM}(\mathbf{Y}, X_1, G))p(\mathrm{MBM}(\mathbf{Y}, X_2, G))}$$

Results

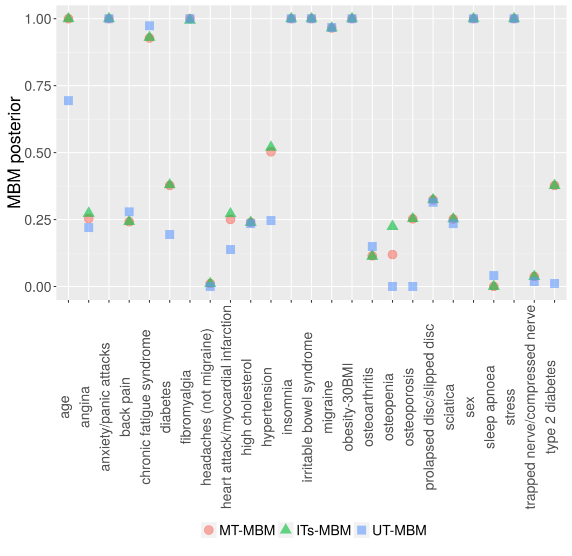

Figure 3: MBM approximations

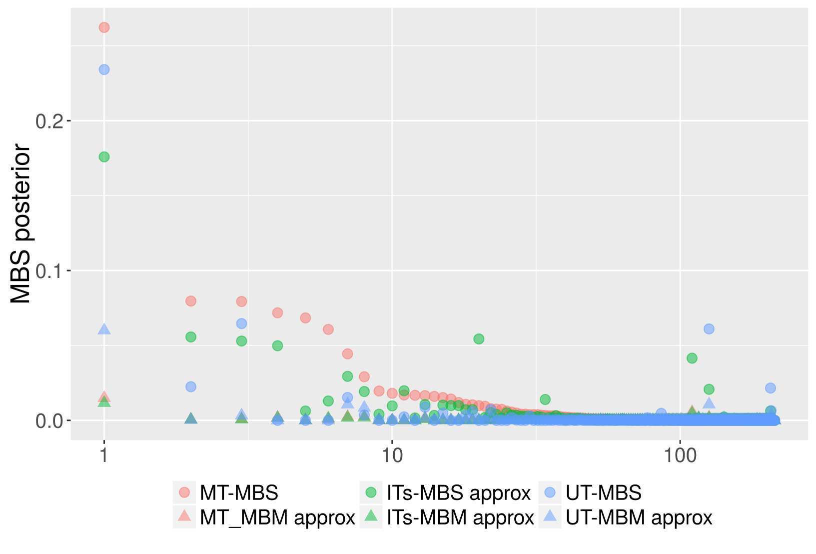

Figure 4: MBS approximations

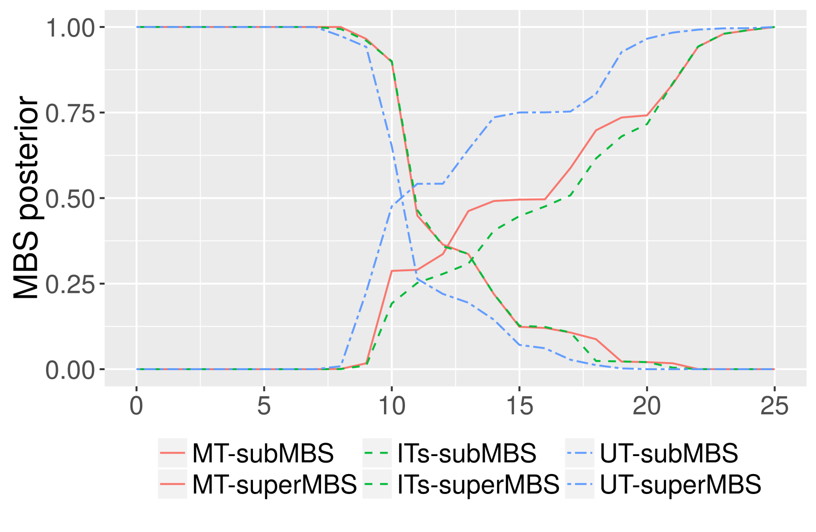

Figure 5: Sub- and supMBS curves for the 3 scenarios.

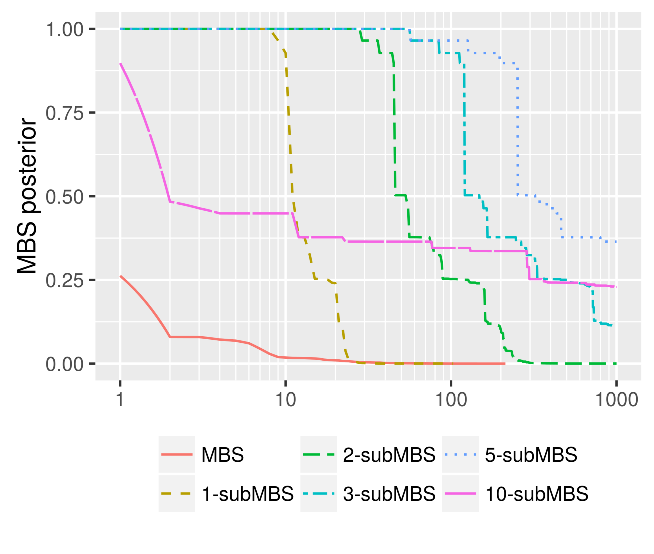

Figure 6: Posterior probability of the k-subMBS sets.

Interactions I

Joint relevant comorbidities for depression	MT P	MT $\hat{P}$	MT IRS	ITs P	ITs $\hat{P}$	ITs IRS	UT P	UT $\hat{P}$	UT IRS	$\frac{IRS_{MT}}{IRS_{ITs}}$	$\frac{IRS_{MT}}{IRS_{UT}}$	$\frac{IRS_{ITs}}{IRS_{UT}}$
osteoporosis osteoarthritis	0.091	0.029	3.151	0.024	0.028	0.844	0.001	0.006	0.234	3.732	13.477	3.611
heart attack osteoarthritis	0.113	0.029	3.927	0.044	0.031	1.455	0.023	0.006	3.684	2.698	1.066	0.395
angina osteoarthritis	0.114	0.029	3.905	0.045	0.031	1.452	0.023	0.006	3.693	2.691	1.057	0.393
diabetes osteopenia	0.119	0.045	2.648	0.085	0.085	0.997	0.000	0.000	0.185	2.656	14.332	5.397
type 2 diabetes osteopenia	0.119	0.045	2.646	0.085	0.085	0.999	0.000	0.000	0.363	2.647	7.287	2.752
osteoarthritis high cholesterol	0.114	0.028	4.133	0.042	0.027	1.576	0.023	0.006	3.905	2.622	1.058	0.404
type 2 diabetes osteoarthritis	0.091	0.043	2.106	0.035	0.042	0.830	0.001	0.009	0.155	2.538	13.577	5.351
diabetes osteoarthritis	0.091	0.043	2.108	0.036	0.043	0.831	0.001	0.009	0.158	2.536	13.359	5.268

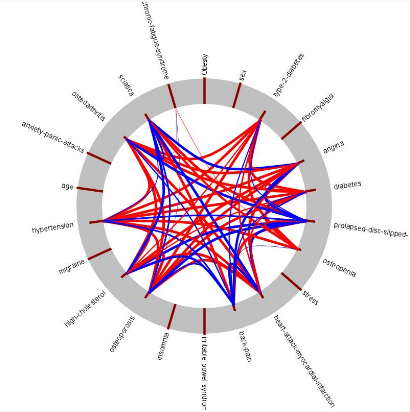

Figure 7: Interactions among the disorders using the MT approach.

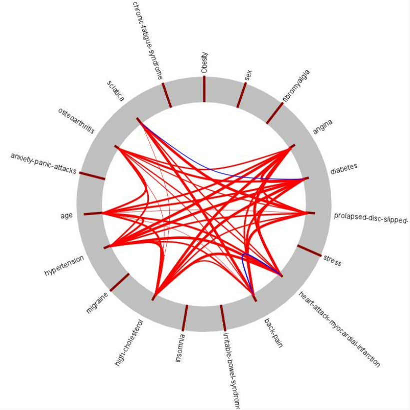

Figure 8: Interactions among the disorders using the IT approach.

Conclusion

- Flat posterior landscape
- subMBS concept
- Approximation of higher level global features
- Multitarget interactions

Collaborators:

Péter Antal (BME-MIT)

András Millinghoffer (BME-MIT)

Gabriella Juhász (MTA-SE)

Questions?

P. Antal, A. Millinghoffer, G. Hullám, C. Szalai, and A. Falus.
A Bayesian view of challenges in feature selection: Feature aggregation, multiple targets, redundancy and interaction.
Journal of Machine Learning Research: Workshop and Conference Proceedings, 4:74–89, 2008.

J. Pearl.
Probabilistic Reasoning in Intelligent Systems.
Morgan Kaufmann, San Francisco, CA, 1988.

J. Pearl.
Causality: Models, Reasoning, and Inference.
Cambridge University Press, 2000.

I. Tsamardinos and C. Aliferis.
Towards principled feature selection: Relevancy, filters, and wrappers.
In *Proc. of the Artificial Intelligence and Statistics*, pages 334–342, 2003.