INTEGRATIVE ANALYSIS OF DATA, LITERATURE, AND EXPERT KNOWLEDGE BY BAYESIAN NETWORKS

Péter Antal

antal@mit.bme.hu

KATHOLIEKE UNIVERSITEIT LEUVEN

Department of Electrical Engineering

Promotor:

Prof. dr. ir. Bart De Moor

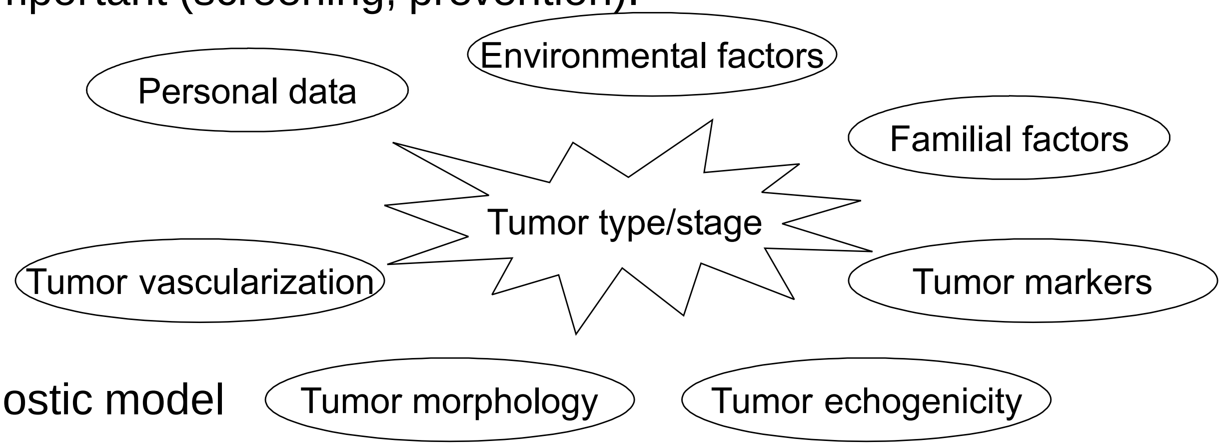
Prof. dr. Yves Moreau

Overview

1. The ovarian cancer problem, the IOTA project
2. Contributions
3. Priors for the ovarian cancer problem
4. Text-mining by Bayesian networks
5. Bayesian analysis of relevance using Bayesian networks
6. Informative priors for black-box conditional models
7. Applications
8. Challenges
9. Publications, CV

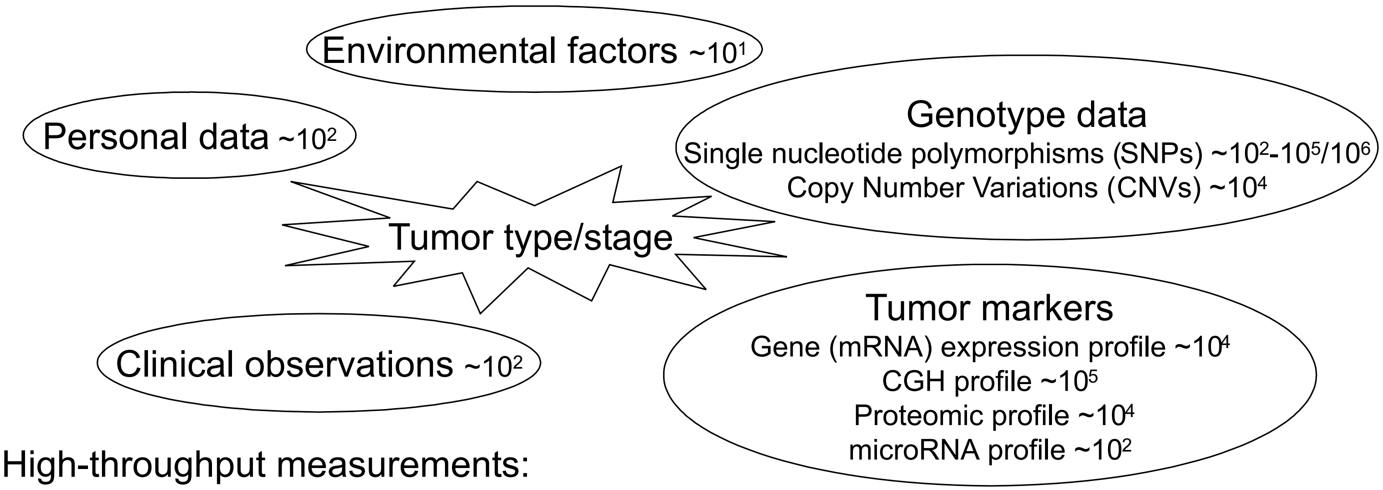
Introduction: the ovarian cancer problem

1. The first/second most commonly diagnosed gynecologic malignancy.
2. The leading cause of death from gynecological malignancy.
3. The fifth leading cause of cancer deaths in women.
4. Poor prognosis if diagnosed at an advanced stage.
5. $\Rightarrow$ Early diagnosis is important (screening, prevention).
6. Multifactorial disease.



1. Goal: learning a diagnostic model
   (a) Performance: Classification/prediction system
      i. Incorporation of prior expert knowledge
      ii. Handling incomplete data
      iii. Decision support, value of further information
      iv. Subclassification, suggestion of treatment
   (b) Understanding: Relevant variables/risk factors

The ovarian cancer problem in the high-throughput context



High-throughput measurements:
- higher number of variables with 2-4 orders

Goals
- Understanding
  - Relevance of variables (e.g. for diagnosis)
  - Mechanisms (e.g. for intervention)
  - Automated study design
    - Sensitive, but multiple-testing proof data analysis
    - Knowledge-rich data analysis
- Performance
  - Fusion of prior knowledge
    - Analysis and fusion of **electronic prior knowledge**
    - **Fusion of “classical” clinical expertise**
  - Value of further information analysis

} **Bayesian data analysis**

International Ovarian Tumor Analysis Consortium (IOTA)

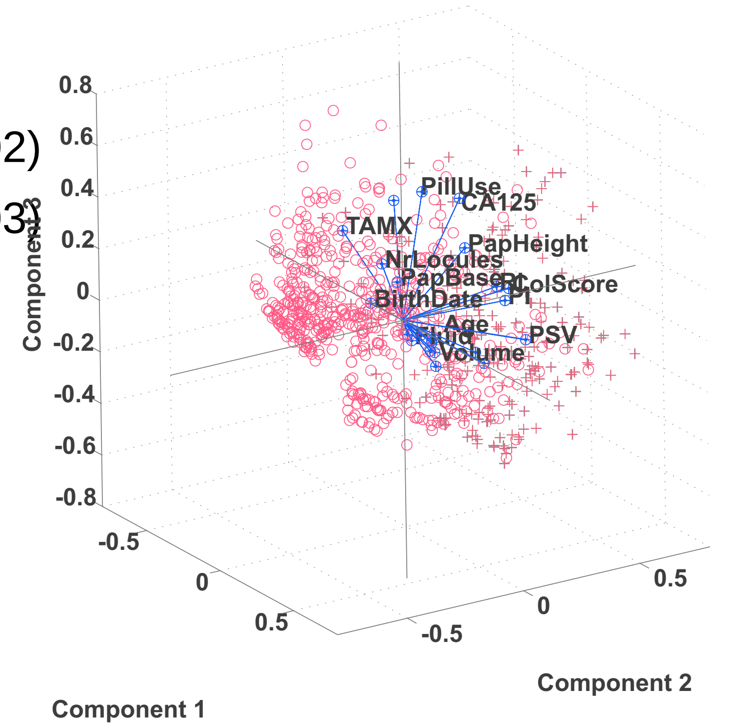
1. Data (9-19 international centers)
   (a) 68 parameters
   (b) 1,066 cases $IOTA-1$(2007)
2. Models
   (a) logistic regression (LR) (1997),
   (b) multilayer perceptrons (MLPs) (1998),
   (c) support vector machines (SVMs) (1998),
   (d) $\Rightarrow$Bayesian logistic regression/multilayer perceptrons (2001),
   (e) $\Rightarrow$Bayesian Belief networks (BNs) (2001),
   (f) least squares support vector machines (2002),
   (g) Bayesian kernel methods (2007)

Data

1. Parameters: 35 out of 68 are used in the thesis: *FamHistBrCa, FamHistOvCa, FamHist, PMenoAge, ReprYears, Meno, Age, PostMenoY, Hysterectomy, CycleDay, PillUse, Parity, HormTherapy, Pathology, PapFlow, PapSmooth, Papillation, Solid, WallRegularity, Septum, IncomplSeptum, Locularity, Echogenicity, Shadows, TAMX, PSV, PI, RI, ColScore, Volume, Ascites, Fluid, Bilateral, Pain, CA125*

2. Data sets in the thesis $\neq IOTA-1$(2007)
   (a) IDO, 11 variables, 300 cases (1997)
   (b) $IOTA-1.1$, 31 variables, 604 cases (2002)
   (c) $IOTA-1.2$, 35 variables, 782 cases (2003)

The biplot of the IOTA-1.2 data set: (variables:'**o**', cases:'+'/'o')



Models:

Probabilistic conditional and domain models

Logistic regression (LR): $P(y|\underline{x}) = \sigma[\sum_{i=0}^{n}(\beta_i x_i + \sum_{j=1}^{n}(\beta_{i,j}x_i x_j + \ldots)))]$,

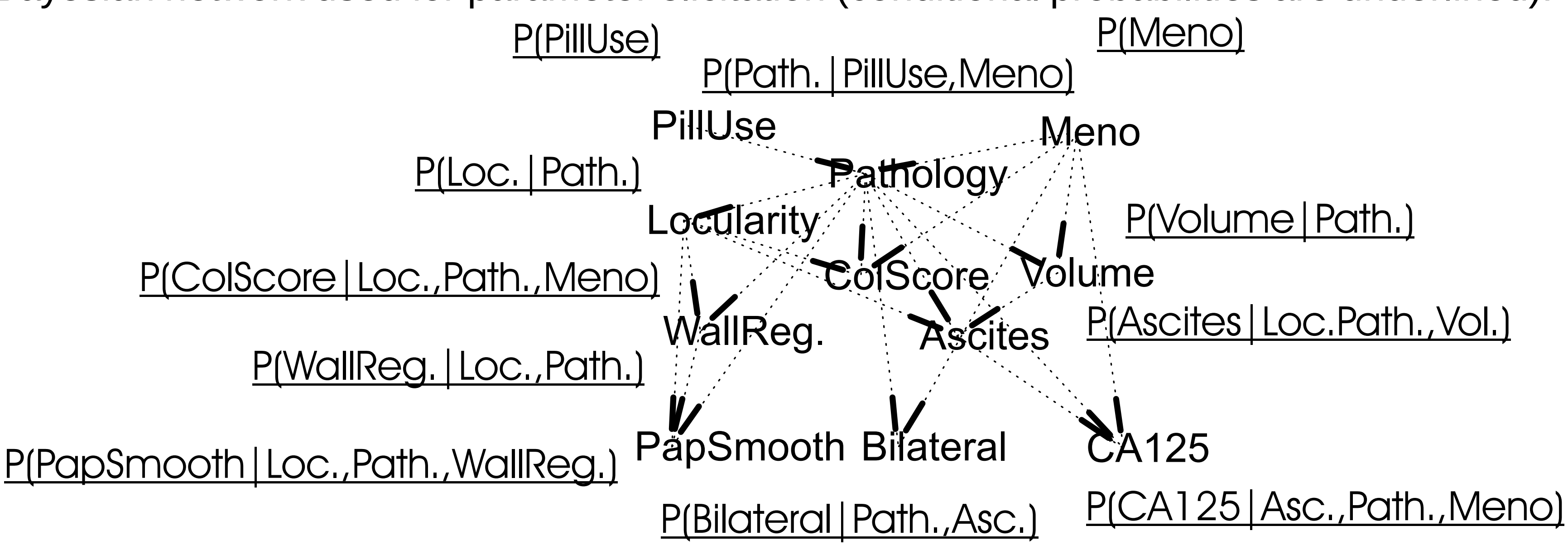
Multilayer perceptron (MLPs): $f(\underline{x},\underline{\omega}) = \sigma[\sum_{i=1}^{L}(\omega_i \ \tanh[\sum_{j=1}^{|\underline{X}|}(\omega_{ij}x_j + \omega_{i0})])]$,

Naive Bayesian networks (N-BNs): $p(y, x_1, \ldots, x_n|\underline{\theta}) = p(y)\prod_{i=1}^{n} p(x_i|y)$,

Bayesian networks (BNs): $p(x_1, \ldots, x_n|\underline{\theta}, G) = \prod_{i=1}^{n} p(x_i|\,\mathrm{pa}(X_i, G))$.

Model: structure and parameters (LR/MLP:$\underline{\theta}$, BN:$\underline{\omega}$)

Bayesian network used for parameter elicitation (conditional probabilities are underlined):



Methods

for Bayesian model averaging (BMA)

Frequentist statistics: optimization (w.r.t. likelihood)

Bayesian statistics: model averaging (w.r.t. posterior)

Predictive inference:

$$p(x|D) = \sum_k p(M_k|D) \int p(x|\theta_k) p(\theta_k|D, M_k) \, \mathrm{d}\theta_k. \tag{1}$$

"Parametric" inference (inferring about a structural model property):

$$p(F(G) = f|D_N) = \mathrm{E}_{p(G|D_N)}[F(G) = f] = \sum_G 1(F(G) = f) p(G|D_N). \tag{2}$$

Inference	Model	Target	Method
Predictive	LR/MLP	$p(y\|\underline{x}, S, D_N) = E_{p(\underline{\omega}\|D_N)}[f(\underline{x}, \underline{\omega}]$	hybrid-MCMC
Predictive	N-BN	$p(y\|\underline{x}, D_N) = E_{p(G\|D_N)}[p(y\|\underline{x}, G)]$	exact (by sum-prod flip)
Predictive	BN	$p(y\|\underline{x}, D_N) = E_{p(G\|D_N)}[p(y\|\underline{x}, G)]$	ordering MCMC
Parametric	BN	$p(F(G) = f\|D_N) = E_{p(\prec\|D_N)}[p(F(G) = f\| \prec)]$	ordering MCMC
Parametric	BN	K most probable model properties	integrated estimate&search

Contributions of the thesis

1. **Electronic prior knowledge**.
   (a) Statistical natural language processing
      i. Text-mining by Bayesian networks (BNs).
   (b) Bayesian logic.
      i. Fusion of factual and uncertain knowledge.
2. **Bayesian analysis of relevance**.
   (a) Generalizations of the feature subset selection (FSS) problem.
   (b) Ordering MCMC for Markov Boundary subGraphs (MBG).
   (c) Integrated estimate and search method for MBGs and Markov Boundary sets.
   (d) The Bayesian, four-level analysis of relevance (B4s).
3. Fusion of **prior expertise** in predictive systems.
   (a) Knowledge engineering textually enriched prior models (BNs).
   (b) Prequential analysis of the value of priors.
   (c) Structural priors for multilayer perceptrons (MLPs) using MBGs.
   (d) Parametric priors for MLPs by a distance minimization projection method.

Priors for the ovarian cancer

Document collections by querying Pubmed query with “ovarian cancer” ($35,562$)

- the *most relevant* papers ($2,256$) in the *most relevant* journals (2),
- the *highly relevant* papers ($3,301$) the *highly relevant* journals (3),
- the *moderately relevant* papers ($9,372$) in the *moderately relevant* journals (33),
- the *relevant* journals papers ($12,038$) in the *relevant* journals (93).

Domain vocabulary (phrases, synonyms).

Variables

- Discretization levels,
- “Text kernel”: name, synonyms, annotations, references.
- Hierarchical groupings of the variables.

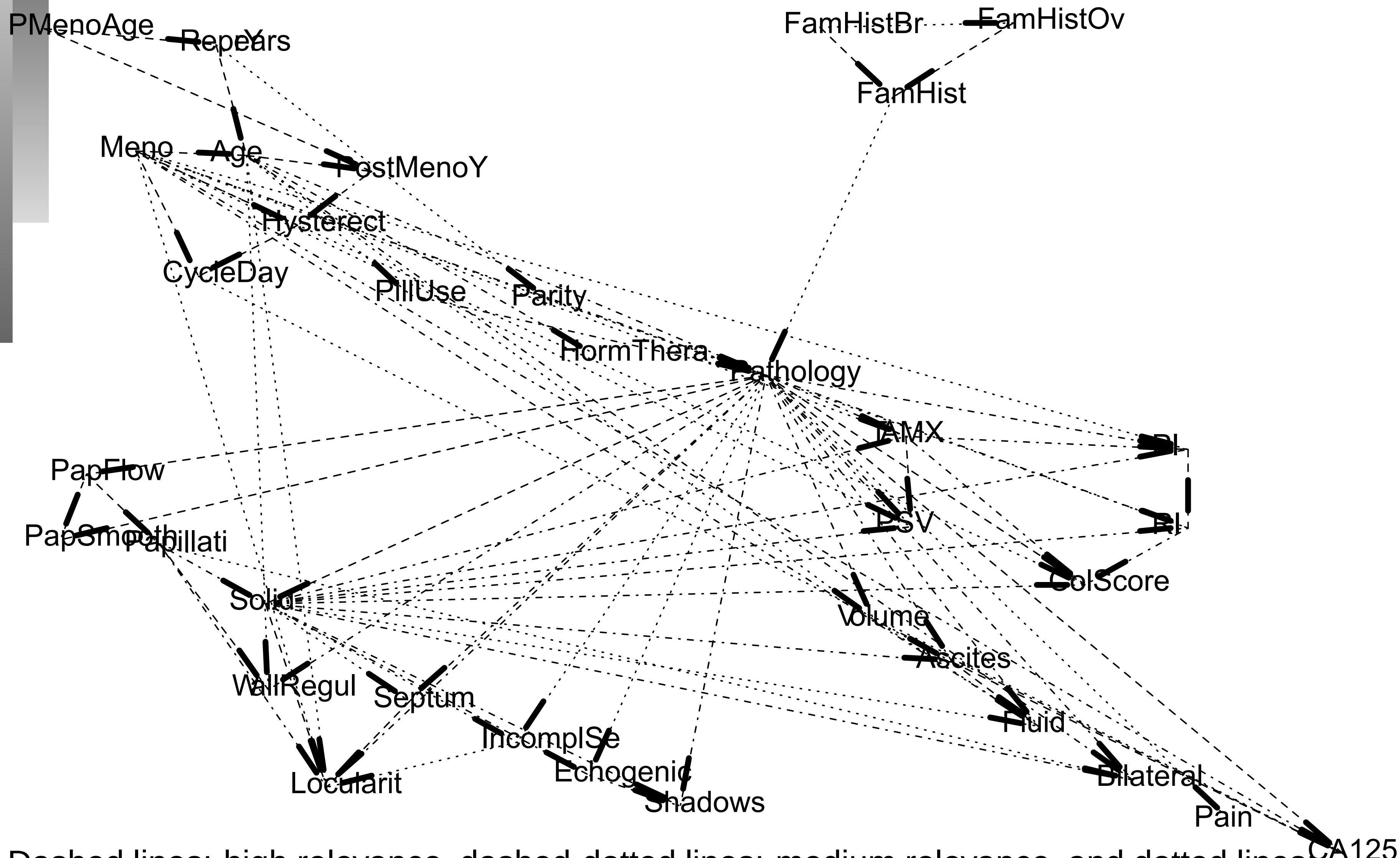
Pairwise relations: relevance (existential), sign, logical/causal,

Parameterized Bayesian network: 11 variables, 400 estimates,

Three embedded BN structures: 35 variables,

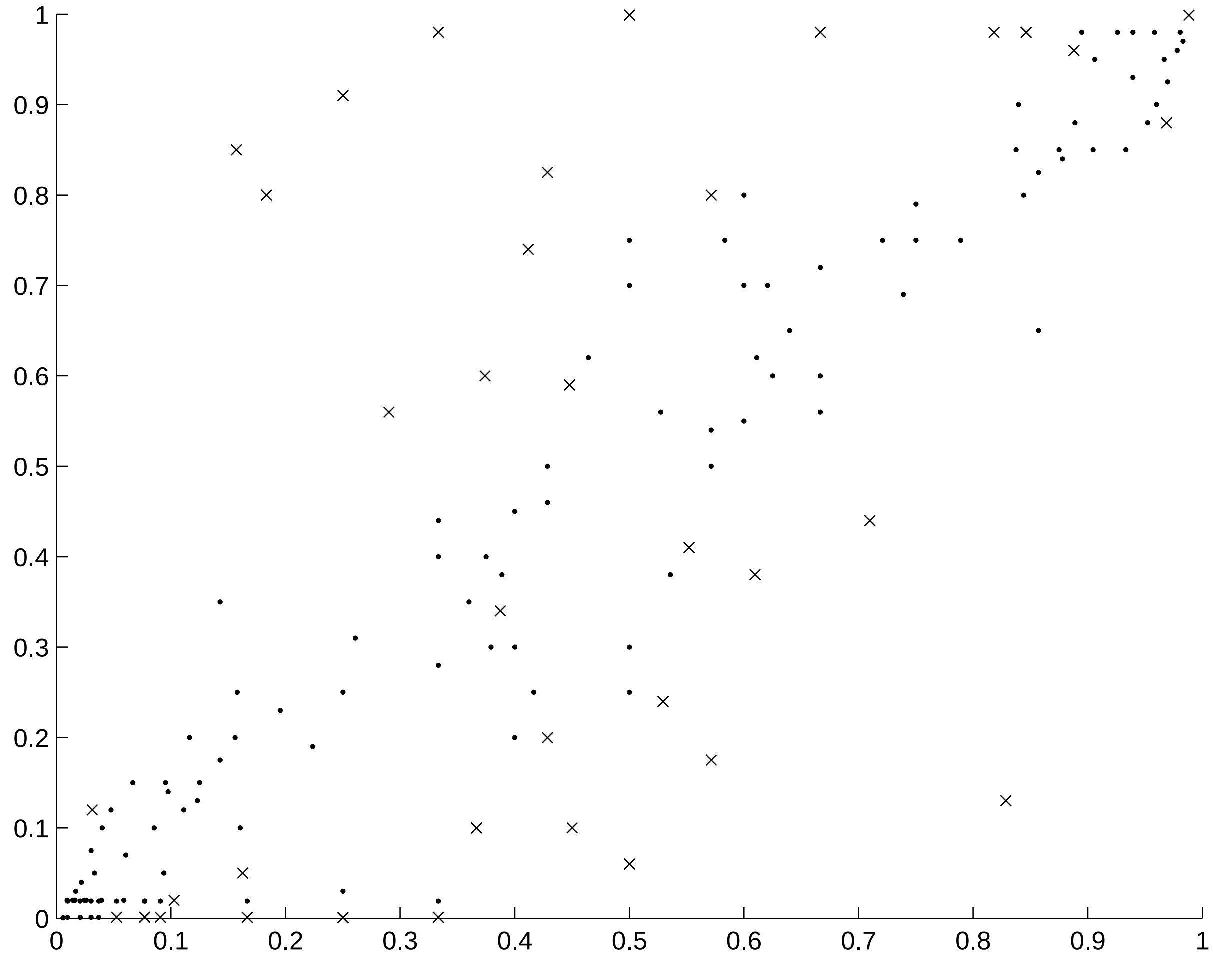
Partial and complete orderings of the variables.

Prior BN structures



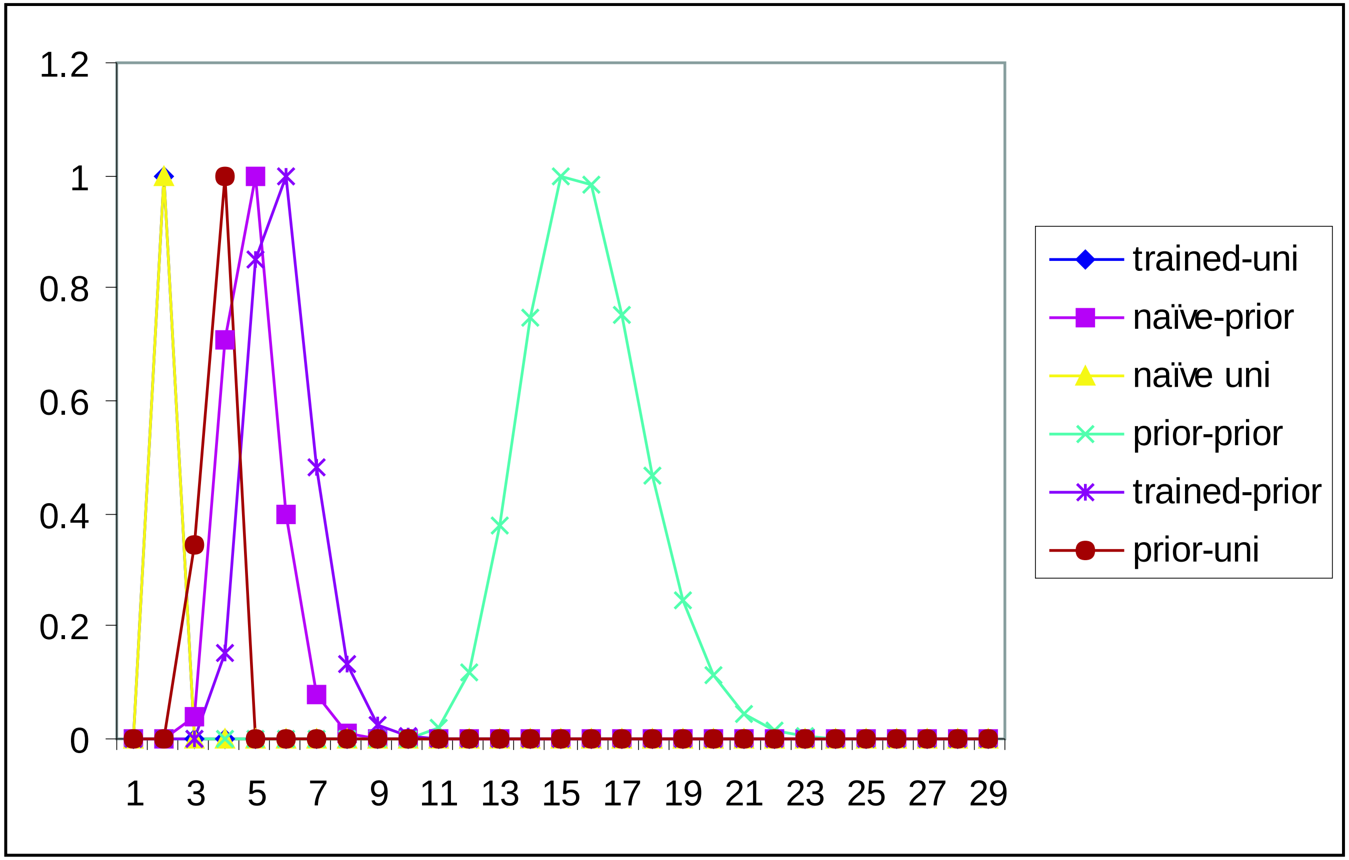
Dashed lines: high relevance, dashed-dotted lines: medium relevance, and dotted lines: relevance (structures are embedded).

Parametric priors



The scatterplot of the conditional probabilities for all the IOTA variables using the expert's estimates (vertical axis) and the posterior expectations with $\mathrm{BD_{eu}}$ priors (horizontal axis). The coordinates are labeled with being inside/outside (./x) a credible region(0.99).

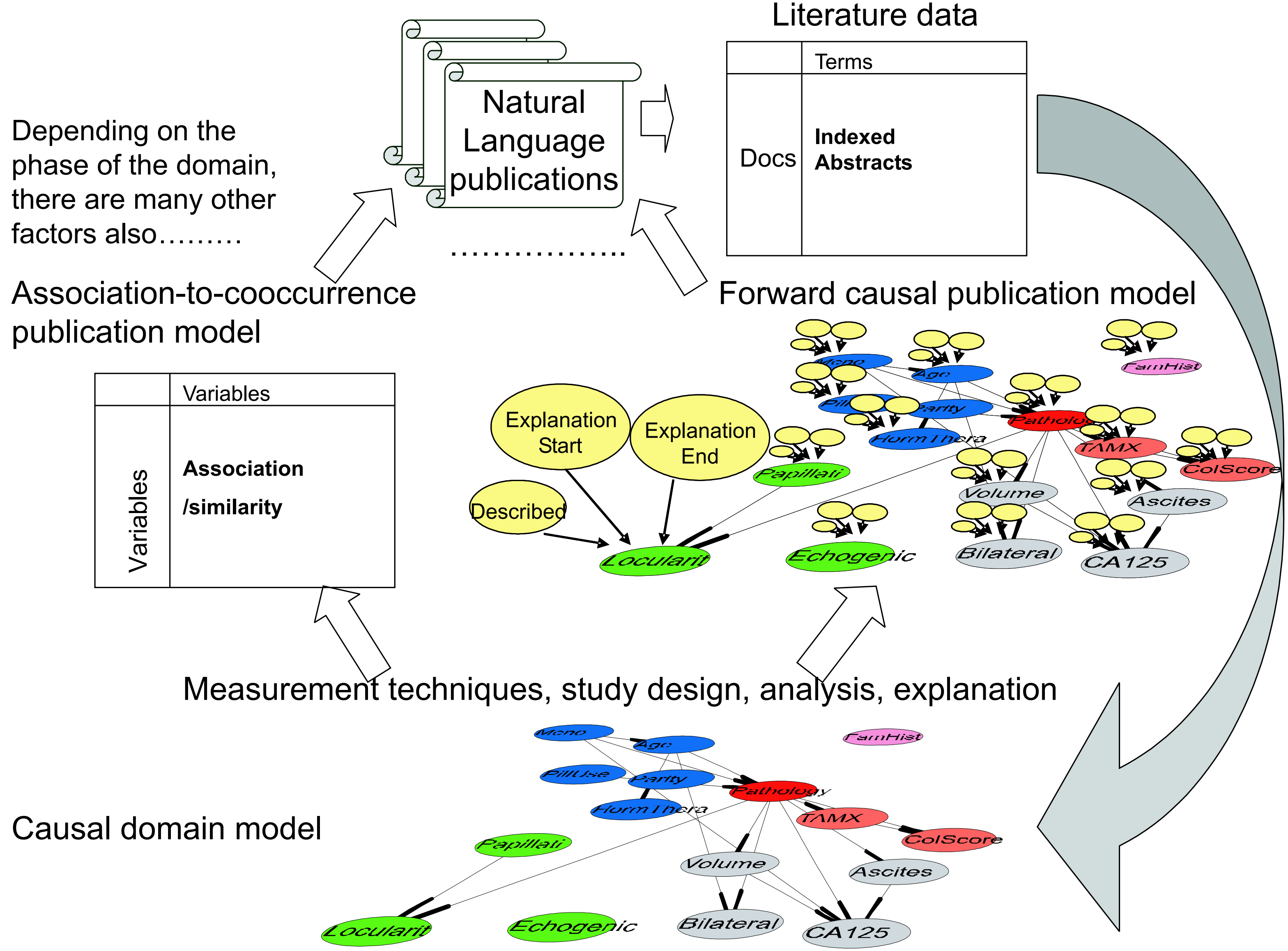
The value of parameter prior: the hyperposterior of the virtual sample size



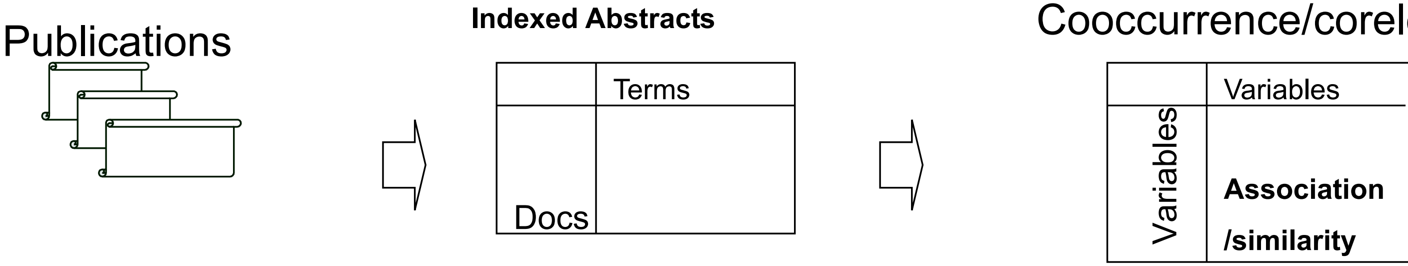
Observations:

1. *numeric estimates worth ≈150 complete cases,*
2. *in simpler model this drops to ≈50 cases,*
3. *even uniform pseudocounts are advantageous (corresponding to ≈10 cases).*

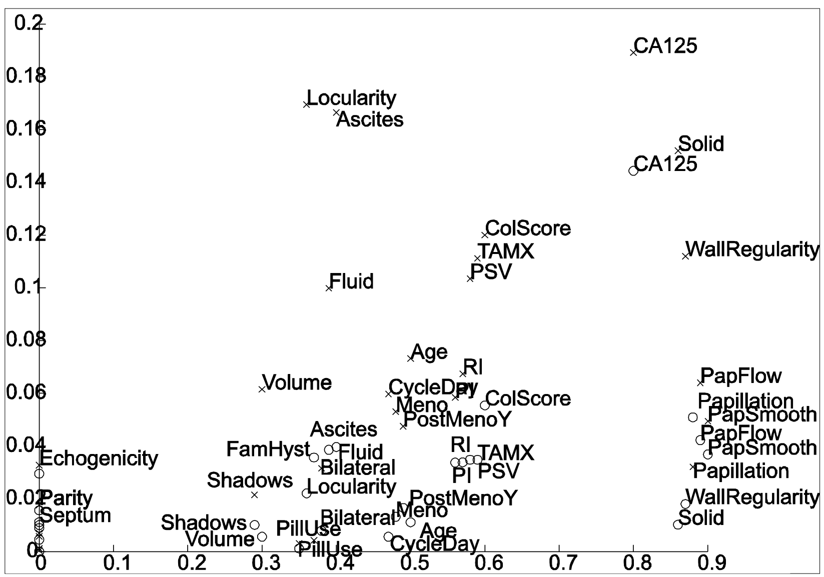
A multivariate, model-based, causal text-mining by Bayesian networks

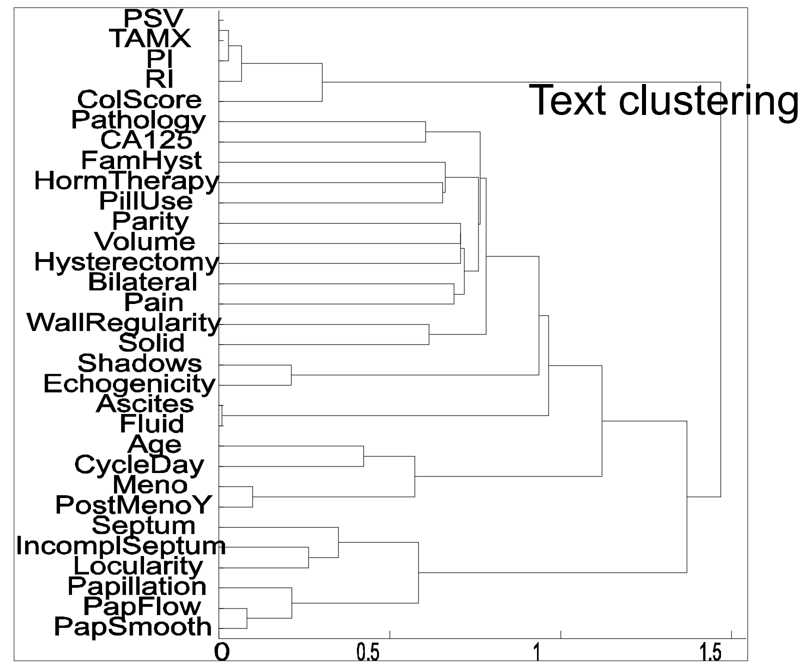


Analysis of the association-cooccurrence relation



Visualization of text- and data-based scores

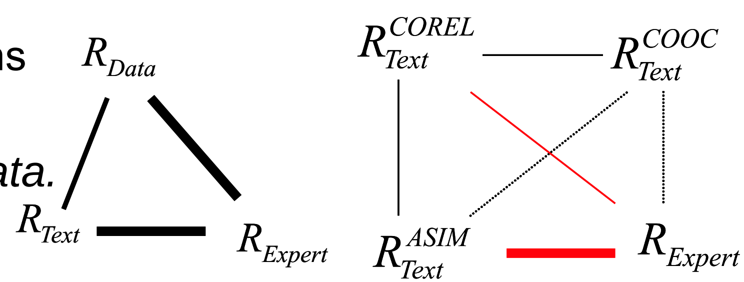




Correlations, rank correlations

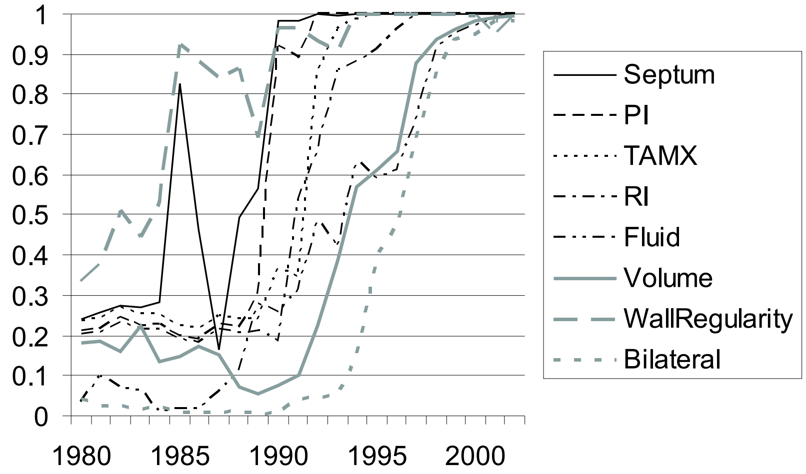
Observations:

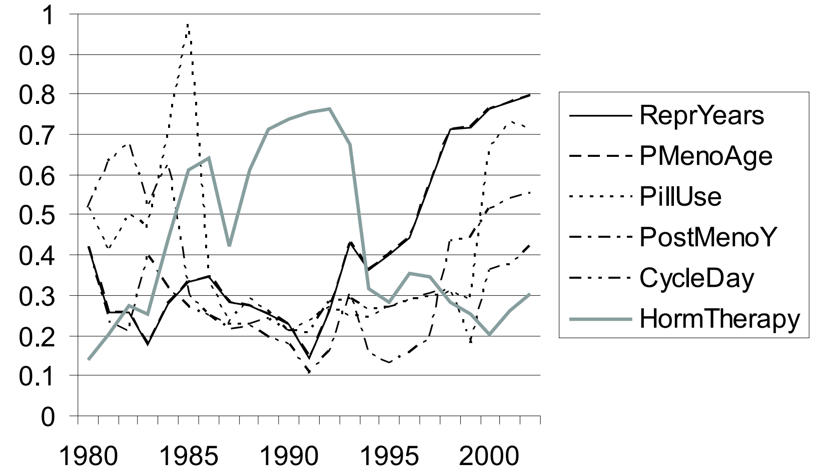
The expert is closer to the truth/data.



Analysis of the forward-causal assumption: Bayesian, sequential, pairwise approach

The temporal evolution of the collective belief — inferred from the literature — that a given variable is relevant for the preoperative diagnostics of ovarian cancer.





The figure shows the sequential posteriors of Bayesian network based relevances - the MBM(Pathology,X_i,G) relations - using the temporal sequence of publications between 1980 and 2005 in the large PubMed corpus.

Observations:

1. *convergence to 0/1 corresponds well to clinical irrelevance/relevance,*
2. *high level of uncertainty$\Rightarrow$MAP model-based evaluations are problematic!,*

Learning Bayesian network features from heterogeneous sources

From prior incorporation to BN features???

1. Scarcity of data $\Rightarrow$ Bayesian analysis of model properties (BN features).
2. 2000: Bayesian analysis of pairwise relevances [5, 6] $\Rightarrow$ fragmentary theory.
3. The questions are conditional ones (relevance of “attributes” for target).
4. Even the prior is conditionally biased.
5. Idea: Bayesian analysis of restricted, but “conditionally complete” feature.

Conditional relevance
and relevance by Bayesian networks

A probabilistic concept of relevance (the “fi lter” approach)

Definition 1. *A feature X_i is strongly relevant, if there exists some x_i, y and $s_i = x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n$ for which $p(x_i, s_i) > 0$ such that $p(y|x_i, s_i) \neq p(y|s_i)$. A feature X_i is weakly relevant, if it is not strongly relevant, and there exists a subset of features S_i' of S_i for which there exists some x_i, y and s_i' for which $p(x_i, s_i') > 0$ such that $p(y|x_i, s_i') \neq p(y|s_i')$. A feature is relevant, if it is either weakly or strongly relevant; otherwise it is irrelevant [7, 8].*

A graph-theoretic representation of relevance

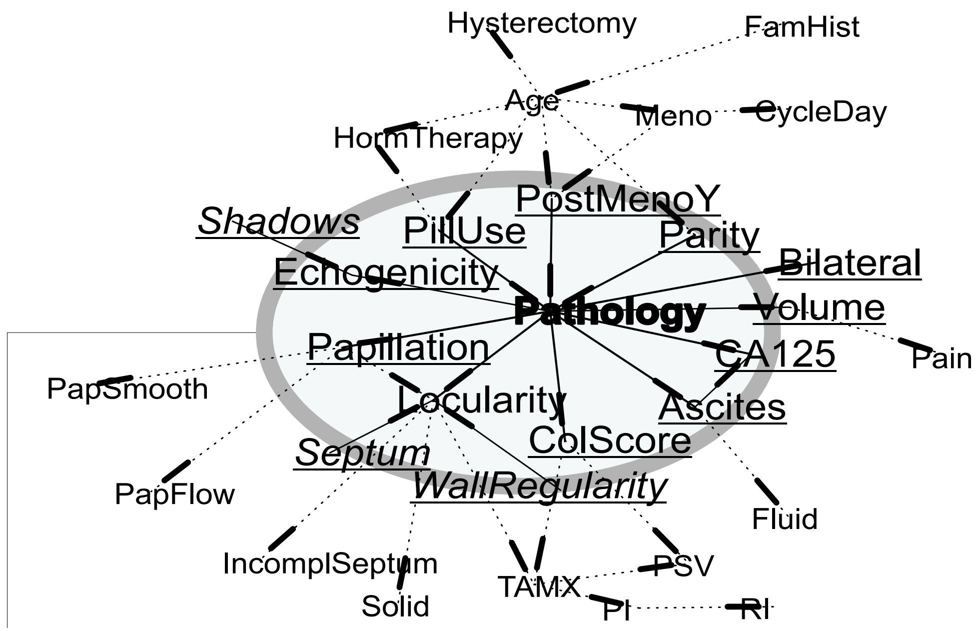
Theorem 1 ([16]). *If distribution P is stable w.r.t. the DAG G, then the variables corresponding to the nodes in the boundary of Y, $\mathrm{bd}(Y, G)$ (the parents and children of Y and other parents of its children) forms a unique and minimal Markov blanket of Y, $\mathrm{MB}_P(Y)$ (the Markov boundary). Furthermore, $X_i \in \mathrm{MB}_P(Y)$, if X_i is strongly relevant.*

Observation (filter vs. wrapper approach): this filter approach is

1. model-free: independent of the function class,
2. method-free: independent of the optimization method,
3. loss-free: independent of the performance measure,

Bayesian network features:

pairwise, set-based, subgraph-based



Multivariate with interactions (subgraph):Markov Blanket subgraph (MBG(Pathology,

Multivariate (set):Markov blanket [set] (MB(Pathology,G))

Univariate (pairwise): Markov Blanket Members MBM(Pathology,X,G))

Observation: these features form a hierarchy of decreasing complexity (cardinality):
BN→MBG→MB→MBM

The Markov Blanket (sub)Graph feature

Proposition 1. *Under standard assumptions, the Markov Blanket structural and parametric marginals define the conditional distribution of Y given other variables $V \setminus Y$:*

$$p(Y|V \setminus Y) = \sum_{\mathrm{MBG}(Y,G)=\mathrm{mbg}} p(\mathrm{mbg})p(Y|\,\mathrm{mbg}),$$

The space of DAGs can be collapsed to the MBG subspace in BMA ($\mathcal{O}(n!2^{n^2}) \rightarrow$).

Theorem 2 ([4]). *Under standard assumptions, the ordering-conditional posterior $p(\mathrm{MBG}(Y,G) = \mathrm{mbg} \,|\, \prec, D_N)$ can be computed in polynomial time.*

The MBG subspace can be reduced to the space of orderings in BMA ($\rightarrow \mathcal{O}(n!)$):

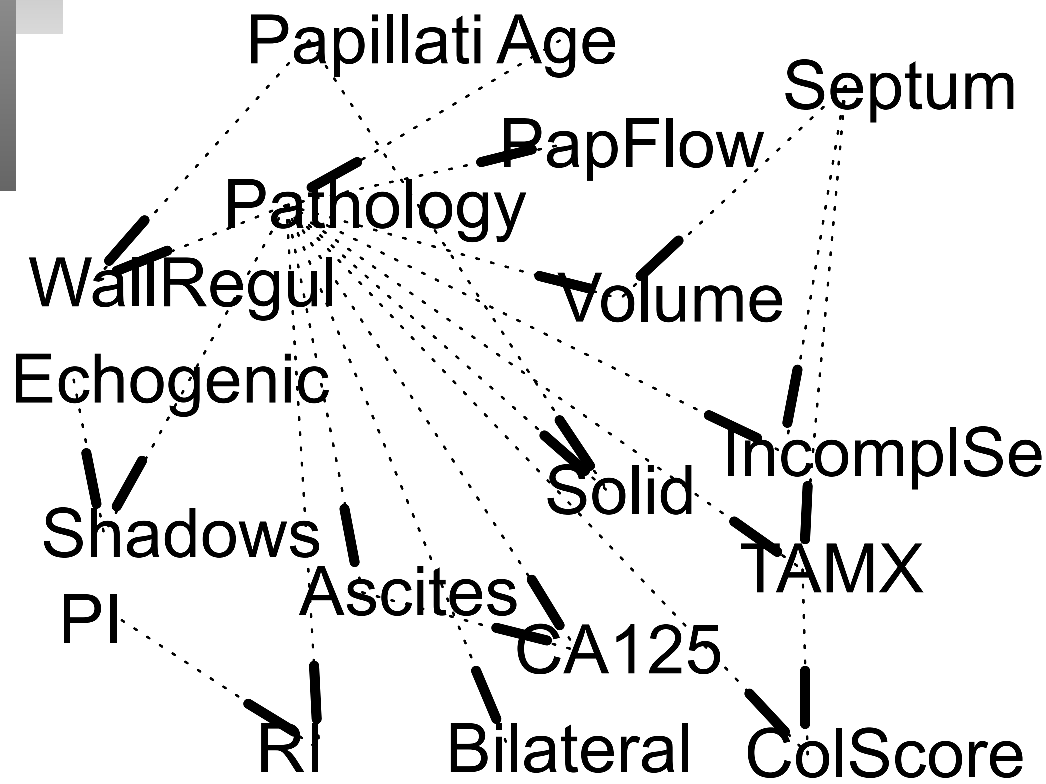
$$\mathrm{E}_{p(G|D_N)}[\mathrm{MBG}(Y,G) = \mathrm{mbg}] = \mathrm{E}_{p(\prec|D_N)}[p(\mathrm{MBG}(Y,G) = \mathrm{mbg} \,|\, \prec, D_N)]$$

Using dynamic programming with $\mathcal{O}(n2^n)$ space complexity [9]:

The space of orderings can be reduced to space of subsets in BMA ($\rightarrow \mathcal{O}(n2^n)$).

Generalizations of the feature subset selection (FSS) problem I.: learn subgraph

Definition 2. *In case of a stable distribution $p(Y, \underline{X})$, the feature (sub)graph selection problem (FGS) denotes the identifi cation of a Markov Blanket subgraph $MBG(Y, G)$, where DAG G denotes a perfect map of distribution p (i.e., Markov Blanket set + substructure).*



Generalizations of the FSS II.:

learn multiple, most probable features

Definition 3. *The* Most Probable Features *problem (MPFs(K)) consists of the selection of* K *most probable feature values* $f \in \mathcal{F}$.

The "Monte Carlo top K selection problem":

1. *estimation of the posteriors,*
2. *optimization (search).*

A statistical result using uniformly good estimates:

Theorem 3 ([10]). *Assuming* M *i.i.d. samples, the expected error in MPFs(K) is*

$$\mathrm{E}[\frac{1}{K}(\sum_{i=1}^{K}\hat{\mathcal{P}}_i - \sum_{i=1}^{K}\mathcal{P}_i^*)] \quad \leq \quad \sqrt{\frac{\log(2|\mathcal{F}|)+1}{M/2}}.$$

An algorithm for MPFs-MBG(K):
Integrated estimation and search using the MBG-ordering spaces [3].

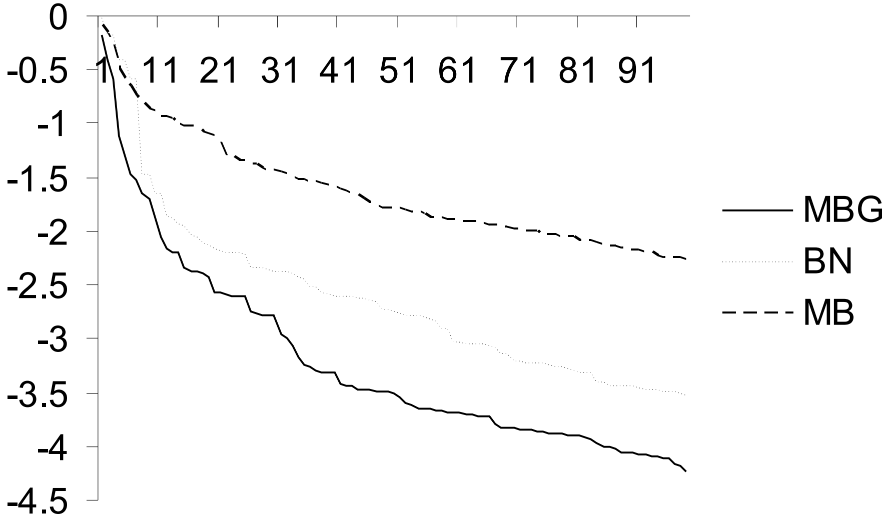
A method for the integrated estimation and search for MBGs (MBs)

Require: $p(\prec)$,p($pa(X_i)| \prec$),k,R,ρ,L^S,ρ^S,L^T,M;
Ensure: K MAP feature value with estimates

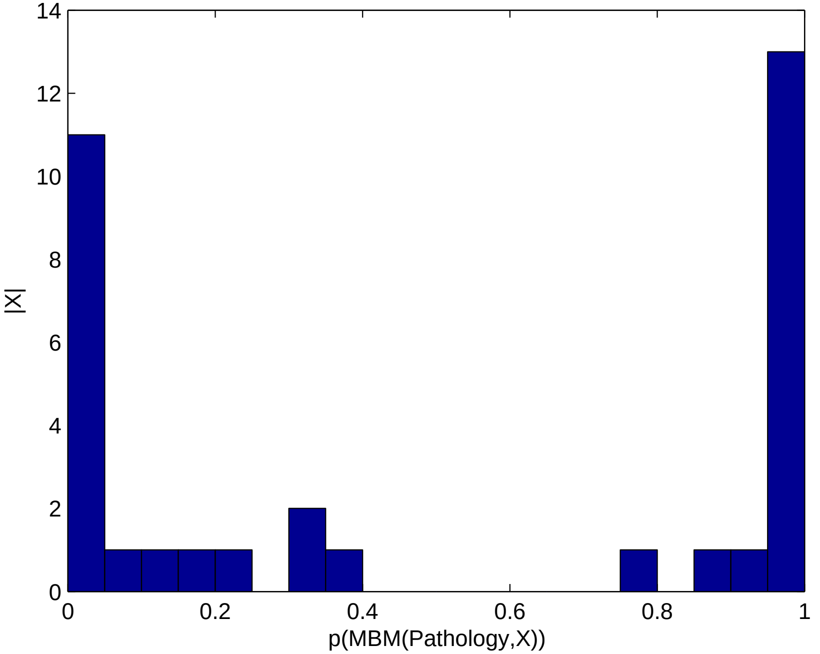
for $l = 0$ to M **do** {the sampling cycle}
  Draw next ordering;
  Compute $p(\prec_l |D_N)$;
  Construct *order conditional MBG-Subspace(*Π, Ψ, R, ρ*)*=Φ
  S^S=*UniformCostSearch(*Φ*,*L^S, ρ^S*)*;
  for all $mbg \in S^S$ **do**
    if $mbg \notin \mathcal{T}$ **then**
      Insert($\mathcal{T}, mbg$)
  if $L^T < |\mathcal{T}|$ **then**
    $\mathcal{T}$=PruneToHPD($\mathcal{T}$,L^T);
  for all $mbg \in \mathcal{T}$ **do**
    $\hat{p}(mbg|D_N) += p(mbg| \prec_l, D_N)$;

Why Bayesian? Uncertainty at all levels.

The relative log posteriors of ranked BNs, MBs, MBGs.



At the level of Markov blankets memberships.

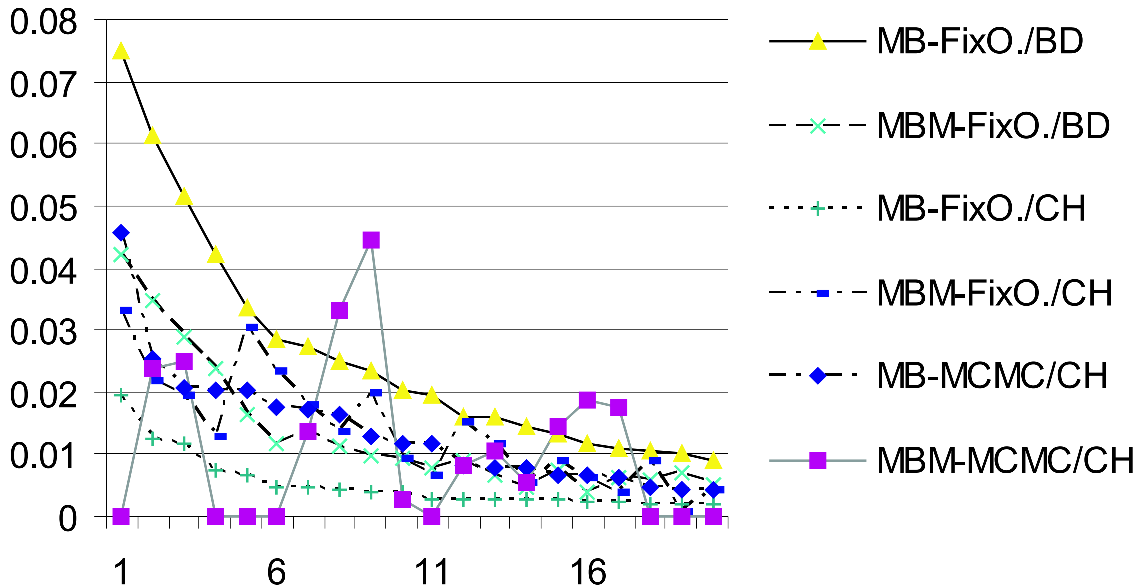


Why multilevel? Autonomy at all levels.

Idea: Low-order approximation of a distribution (at least for ranking).

The MB SET(!) posteriors can be approximated by the PAIRWISE(!) MBM posteriors:

$$p(\mathrm{MB}(Y) = \mathrm{mb}\,|D_N) \approx \prod_{X_i \in \mathrm{mb}} p(\mathrm{MBM}(Y, X_i)|D_N) \prod_{X_i \notin \mathrm{mb}} (1 - p(\mathrm{MBM}(Y, X_i)|D_N)). \quad (3)$$

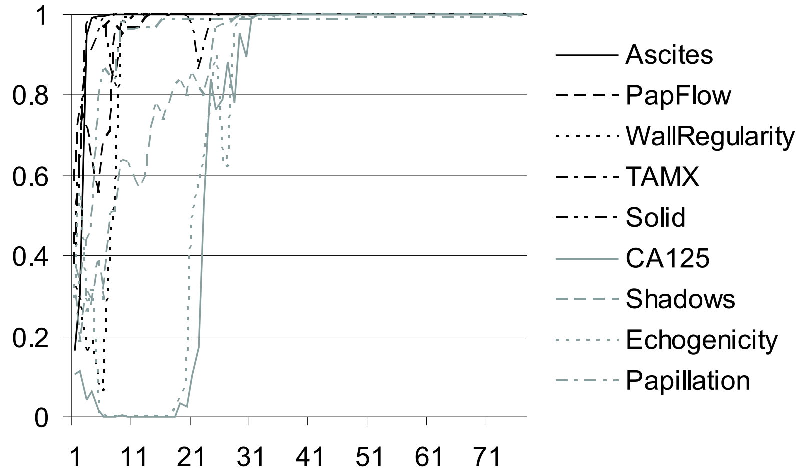


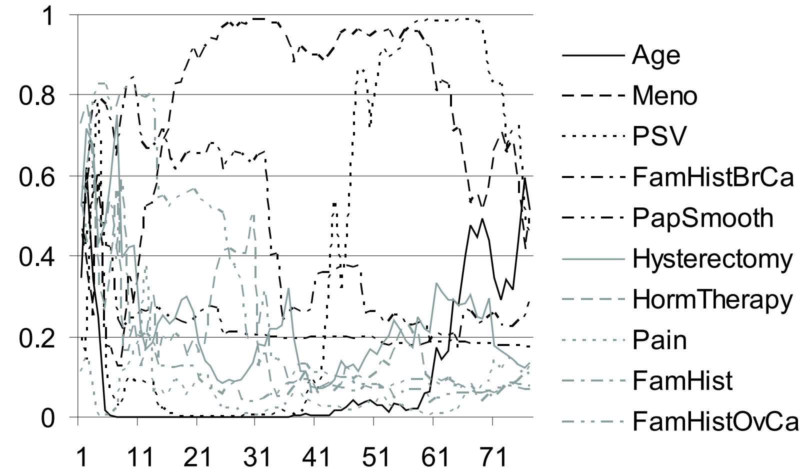
Observations:

1. *numeric approximation is tragic,*
2. *approximation of rankings is poor,*
3. *particularly in the unconstrained case,*
4. *In general: MAP-MBG≠MBG(MAP-BN), MAP-MB≠MB(MAP-MBG)≠MB(MAP-BN), MAP-MBM≠MBM(MAP-MB)≠...*

The sequential analysis of relevance: the pairwise (MBM) level

The temporal evolution of the belief — represented by the posterior of the MBM feature and inferred from growing amount of clinical data — that a given variable is relevant for the preoperative diagnostics of ovarian cancer (conditional on the expert's ordering).





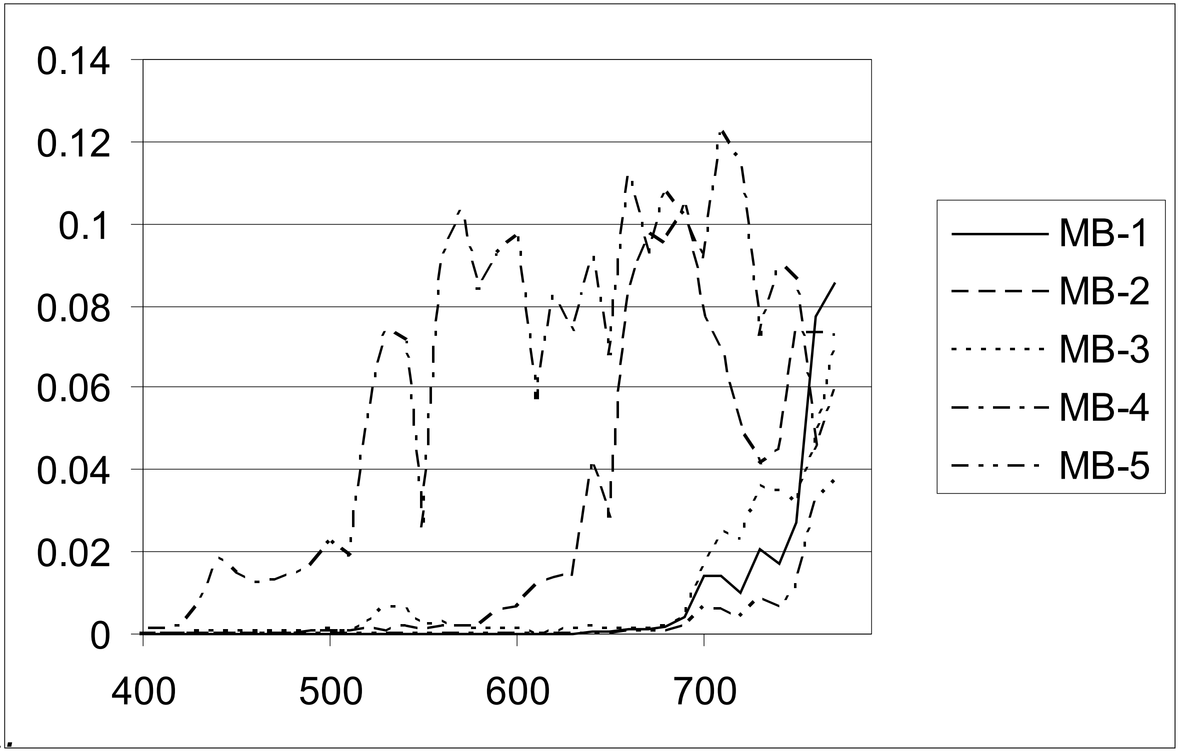
The horizontal axis is the sample size with step size 10.

Observations:

1. *non-monotonic behaviour (Age replacing Meno),*
2. *uncertainty even for all data,*
3. *trends after 500 cases,*
4. *corresponds well to expert's rating of relevance.*

The sequential analysis of relevance: the set (MB) level

The sequential posteriors of high-scoring $\mathrm{MB}(Pathology)$ feature values using the temporal sequence of the IOTA-1.2 data set and given the expert's total causal ordering.



Observations:

1. *The posteriors are less than 10^{-6} for sample size less than 400,*
2. *uncertainty even for all data, BUT POSTERIORS can be used to support the optimized design of subsequent studies.*

Informative priors for parametric black-box models

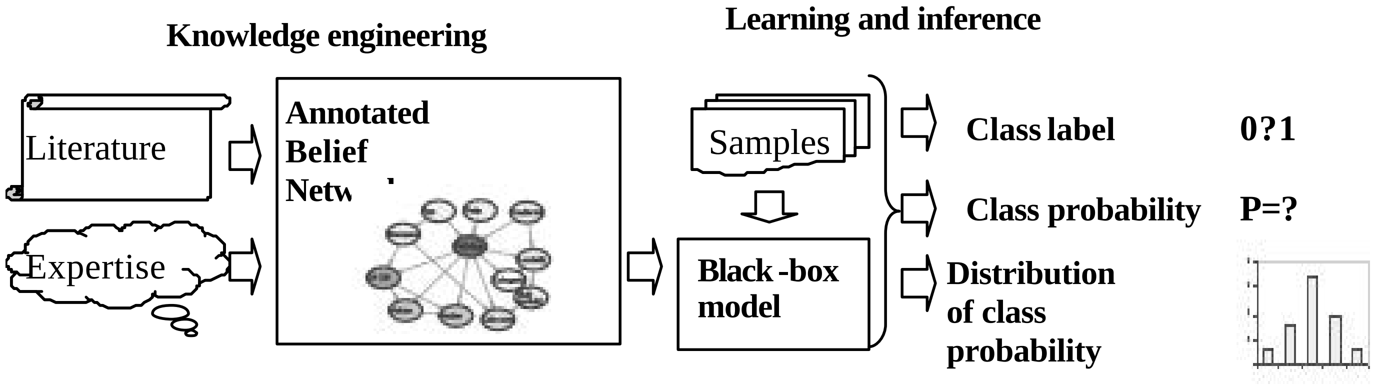
Goals:

1. improved Bayesian prediction,
2. fair Bayesian model comparison,
3. model interpretation,
4. bridging the gap between symbolic and sub-symbolic knowledge representation.

(Some) related works:

1. Knowledge-Based Artificial Neural Network [15],
2. Bayesian networks with neural local models [14],
3. Mapping of Bayesian networks onto stochastic neural networks [11],
4. Mixture of Experts [17],
5. Incorporating Prior Information by Creating Virtual Examples [13],
6. Donor-receiver link by imaginary samples [12].

The two-step, Bayesian methodology for the fusion of knowledge and data



First, we formalized the prior domain knowledge in a Bayesian network. Second, we induced informative structure and parameter priors for parametric conditional models to support various Bayesian inferences.

The prior sample/sample prior transformations

Definition 4.

$$P(\omega|D^+_{N'}, D_N, \xi^-) \propto P(D_N|\omega)P(\omega|D^+_{N'}, \xi^-) = P(D_N|\omega, \xi^-)P'(\omega|\xi^-). \quad (4)$$

Definition 5.

$$\begin{aligned}
p(\omega|\xi^+) &\triangleq \sum_{D^+_{N'}} p(\omega|D^+_{N'}, \xi^-)p(D^+_{N'}|\xi^+) && (5)\\
&= \sum_{D^+_{N'}} p(\omega|D^+_{N'}, \xi^-) \int p(D^+_{N'}|\theta)p(\theta|\xi^+)\,\mathrm{d}\theta && (6)\\
&\propto \sum_{D^+_{N'}} p(\omega|\xi^-)p(D^+_{N'}|\omega) \int p(D^+_{N'}|\theta)p(\theta|\xi^+)\,\mathrm{d}\theta. && (7)
\end{aligned}$$

The conditional distance minimization transformation

Definition 6 ([1, 2]). *Let θ and ω denote the parameters of a domain model and a conditional model. The direct transformation of an informative prior from a domain model into an informative prior over a parametric black box conditional model ($\mathcal{T} : \mathbf{\Theta} \rightarrow \mathbf{\Omega}$) is defi ned as*

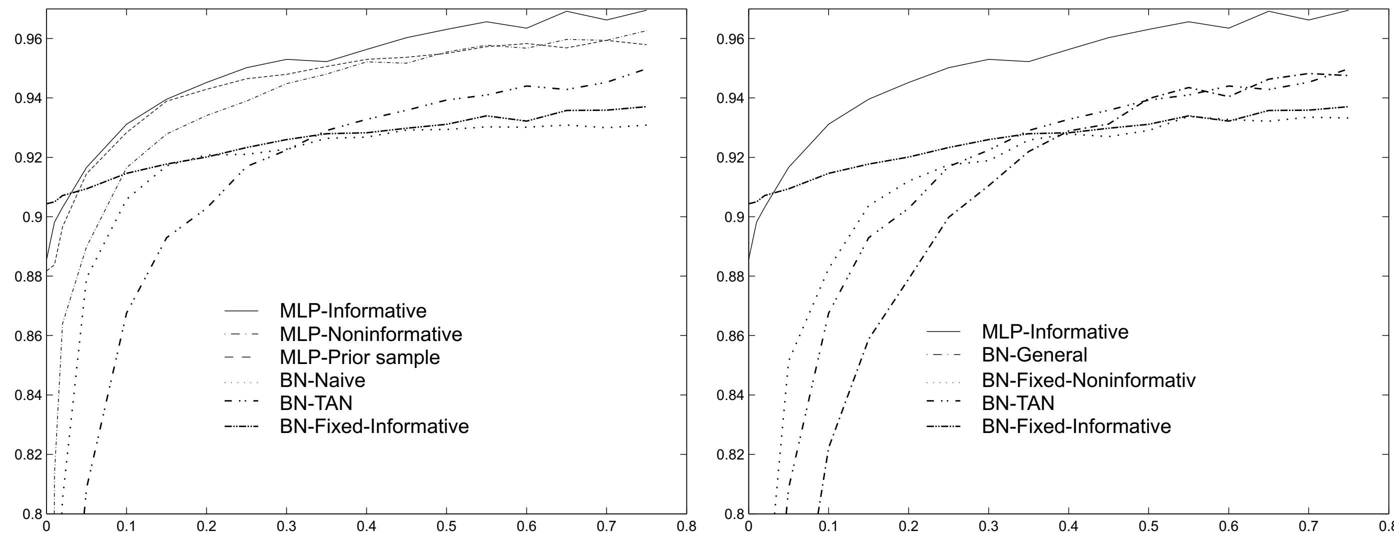
$$\mathcal{T}_{\mathrm{KL}}(\theta) = \arg\min_{\omega'} \mathrm{E}_{p(X|\theta)}[\mathrm{KL}(p(Y|X,\omega') \| p(Y|X,\theta))] + c(\omega) \quad (8)$$

$$\mathcal{T}_{L_2}(\theta) = \arg\min_{\omega'} \mathrm{E}_{p(X|\theta)}[L_2(p(Y|X,\omega'), p(Y|X,\theta))] + c(\omega). \quad (9)$$

1, Generate Bayesian network parameters $\{\theta_1, \ldots, \theta_l\}$
2, Generate block of prior samples from each parameter $\{D_1^p, \ldots, D_l^p\}$
3, Train a multilayer preceptron for each block of samples resulting in a block of perceptron parameters $\{\omega_1, \ldots, \omega_l\}$
3, Approximate the posterior with mixture of Gaussians (G.Fannes:symmetries in the MLP parameter space).

Performance of Bayesian classification with informative priors

The learning curves for the multilayer perceptron models using an informative prior (MLP-Informative), a noninformative prior (MLP-Noninformative) or prior samples (MLP-Prior sample).



1. the same performance as the prior Bayesian network,
2. better performance throughout,
3. not restrictive in the large sample region,
4. high computational complexity of deriving the informative prior,
5. lower complexity in the inference,
6. collapse of a complex general Bayesian model into a task specific, simpler conditional model.

Software platform

1. 10^5 lines of code written in C++ and MATLAB,
2. graphical user interface in the MS-Windows MFC environment,
3. command line version runs in a parallel computing grid environment,
4. Software Environment for Bayesian and Neural Networks (SEBANN),
5. LINUX version with continuous variables (Geert Fannes),
6. System for Probabilistic Annotated Networks (SPAN),
7. modules are migrated into the GEnomic Study Design and Analysis platform.

Applicability in the postgenomic era

1. **Statistical text-mining of multivariate relations**.
   (a) Text-mining by Bayesian networks.
   (b) Goals: minimal preparation, early applicability, model-based.
2. **Bayesian four-level analysis of relevance**.
   (a) Integrated estimate and search method for MBGs and Markov Boundary sets.
   (b) Goals: expression, representation, and communication (publishing) of uncertainty at multiple, linked levels.
3. Fusion of **electronic, factual knowledge** and probabilistic data analysis.
   (a) Bayesian logic by annotated Bayesian networks.
   (b) Goal: knowledge-rich data analysis using complex, semantical hypotheses.
4. Fusion of **prior expertise** in predictive systems.
   (a) Fusion using the conditional distance minimization transformation.
   (b) Goal:Fusing clinical diagnostic knowledge into models developed by high-throughput data.

Challenges

1. Text-mining by Bayesian networks.
   (a) Neutral omission,
   (b) Negation,
   (c) Temporality,
   (d) Utility models,
2. Bayesian analysis of relevance.
   (a) Multiple target variables,
   (b) Continuous variables,
   (c) Incomplete data,
   (d) Scaling up the number of variables from 100 to 1000: hierarchical-MCMC and coupled-MCMC.

Acknowledgment

Thank you for your attention!

Thank you for your support.

Publications about priors for ovarian cancer

[] P. Antal, H. Verrelst, D. Timmerman, S. Van Huffel, B. De Moor, and I. Vergote. How might we combine the information we know about a mass better? The use of mathematical models to handle medical data. 1st Monte Carlo Conf. on Updates in Gynaecology, 2000, Internal Report 00-145, ESAT-SISTA, K.U.Leuven (Leuven, Belgium), 2001.

[] P. Antal, H. Verrelst, D. Timmerman, Y. Moreau, S. Van Huffel, B. De Moor, and I. Vergote. Bayesian networks in ovarian cancer diagnosis: Potential and limitations. In *Proc. of the 13th IEEE Symp. on Comp.-Based Med. Sys. (CBMS-2000)*, pages 103–109, 2000.

[] S. Aerts, P. Antal, B. De Moor, and Y. Moreau. Web-based data collection for ovarian cancer: a case study. In *Proc. of the 15th IEEE Symp. on Computer-Based Medical Sys. (CBMS-2002)*, pages 282–287, 2002.

Publications about text-mining

[] P. Antal, G. Fannes, Y. Moreau, D. Timmerman, and B. De Moor. Using literature and data to learn Bayesian networks as clinical models of ovarian tumors. *Artificial Intelligence in Medicine*, vol. 30, pages 257–281, 2004.

[] P. Antal, P. Glenisson, G. Fannes, J. Mathijs, Y. Moreau, and B. De Moor. On the potential of domain literature for clustering and Bayesian network learning. In *Proc. of the 8th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (ACM-KDD-2002)*, pages 405–414, 2002.

[] P. Glenisson, P. Antal, J. Mathys, Y. Moreau, and B. De Moor. Evaluation of the vector space representation in text-based gene clustering. In *Proc. of the Pacific Symposium on Biocomputing (PSB03)*, pages 391–402, 2003.

[] P. Antal and A. Millinghoffer. Learning causal bayesian networks from literature data. *Proceedings of the 3rd International Conference on Global Research and Education, Inter-Academia'04*, pages 149–160, 2004.

[] P. Antal and A. Millinghoffer. Literature mining using bayesian networks. In *Proc. of third European Workshop on Probabilistic Graphical Models*, pages 17–24, 2006.

Publications about Bayesian based analysis of relevance

[] Y. Moreau, P. Antal, G. Fannes, and B. De Moor. Probabilistic graphical models for computational biomedicine. *Methods of Information in Medicine*, vol. 42(4), pages 161–168, 2002.

[] P. Antal, G. Fannes, Y. Moreau, D. Timmerman, and B. De Moor. Using literature and data to learn Bayesian networks as clinical models of ovarian tumors. *Artificial Intelligence in Medicine*, vol. 30, pages 257–281, 2004.

[3] P. Antal, G. Hullám, A. Gézsi, and A. Millinghoffer. Learning complex bayesian network features for classification. In *Proc. of third European Workshop on Probabilistic Graphical Models*, pages 9–16, 2006.

[10] A. Millinghoffer, G. Hullám, and P. Antal. On inferring the most probable sentences in bayesian logic. In *Workshop notes on Intelligent Data Analysis in bioMedicine And Pharmacology (IDAMAP-2007), 11th Conference on Artificial Intelligence in Medicine (AIME 07)*, pages 13–18, 2007.

Publications about informative priors for black-box conditional models

[2] P. Antal, G. Fannes, H. Verrelst, B. De Moor, and J. Vandewalle. Incorporation of prior knowledge in black-box models: Comparison of transformation methods from Bayesian network to multilayer perceptrons. In *Workshop on Fusion of Domain Knowledge with Data for Decision Support, 16th Uncertainty in Artificial Intelligence Conference*, pages 42–48, 2000.

[] P. Antal, G. Fannes, S. Van Huffel, B. De Moor, J. Vandewalle, and Dirk Timmerman. Bayesian predictive models for ovarian cancer classification: evaluation of logistic regression, multi-layer preceptron and belief network models in the Bayesian context. In *Proc. of the 10th Belgian-Dutch Conference on Machine Learning, BENELEARN 2000*, pages 125–132, 2000.

[] P. Antal, G. Fannes, Y. Moreau, and B. De Moor. Bayesian applications of belief networks and multilayer perceptrons for ovarian tumor classification with rejection. *Artificial Intelligence in Medicine*, vol. 29, pages 39–60, 2003.

[3] P. Antal, G. Hullám, A. Gézsi, and A. Millinghoffer. Learning complex bayesian network features for classification. In *Proc. of third European Workshop on Probabilistic Graphical Models*, pages 9–16, 2006.

Curriculum Vitae

1995: M.Sc. in Computer Science (Informatics Engineer), 1995, Faculty of Electrical Engineering and Informatics, Technical University of Budapest 1995-1998: PH.D. studies on the informatics Ph.D. programme of Faculty of Electrical Engineering and Informatics at the Department of Measurement and Information Systems, Technical University of Budapest.
1998 - 1999: International scholar, Department of Electrical Engineering, Katholieke Universiteit Leuven.
2000-2002: Studying for Ph.D. at the Department of Electrical Engineering, Katholieke Universiteit Leuven, Belgium. Research topic: Combination of prior domain knowledge and data in statistical learning methods.
2003-2004: Assistant lecturer at the Department of Measurement and Information Systems, Technical University of Budapest
2005-: Assistant professor at the Department of Measurement and Information Systems, Technical University of Budapest
2007-: founding member and CEO of GenoBys "GenomicsByBayes" (http://genobys.sote.hu)

References

[1] P. Antal. Applicability of prior domain knowledge formalised as Bayesian network in the process of construction of a classifi er. In *Proc. of the 24th Annual Conf. of the IEEE Industrial Electronic Society (IECON '98)*, pages 2527–2531, 1998.

[2] P. Antal, G. Fannes, H. Verrelst, B. De Moor, and J. Vandewalle. Incorporation of prior knowledge in black-box models: Comparison of transformation methods from Bayesian network to multilayer perceptrons. In *Workshop on Fusion of Domain Knowledge with Data for Decision Support, 16th Uncertainty in Artifi cial Intelligence Conference*, pages 42–48, 2000.

[3] P. Antal, G. Hullám, A. Gézsi, and A. Millinghoffer. Learning complex bayesian network features for classifi cation. In *Proc. of third European Workshop on Probabilistic Graphical Models*, pages 9–16, 2006.

[4] P. Antal and A. Millinghoffer. A probabilistic knowledge base using annotated bayesian network features. In *Proceedings of the 6th International Symposium of Hungarian Researchers on Computational Intelligence*, pages 1–12, 2005.

[5] N. Friedman and D. Koller. Being Bayesian about network structure. In *Proc. of the 16th Conf. on Uncertainty in Artifi cial Intelligence(UAI-2000)*, pages 201–211. Morgan Kaufmann, 2000.

[6] N. Friedman and D. Koller. Being Bayesian about network structure. *Machine Learning*, 50:95–125, 2003.

[7] G. H. John, R. Kohavi, and K. Pfleger. Irrelevant features and the feature subset selection problem. In *Proc. of the 11th International Conference on Machine Learning*, volume 97, pages 121–129. Morgan Kaufmann, 1994.

[8] R. Kohavi and G. H. John. Wrappers for feature subset selection. *Artifi cial Intelligence*, 97:273–324, 1997.

[9] M. Koivisto and K. Sood. Exact bayesian structure discovery in bayesian networks. *Journal of Machine Learning Research*, 5:549–573, 2004.

[10] A. Millinghoffer, G. Hullám, and P. Antal. On inferring the most probable sentences in bayesian logic. In *Workshop notes on Intelligent Data Analysis in bioMedicine And Pharmacology (IDAMAP-2007), 11th Conference on Artifi cial Intelligence in Medicine (AIME 07)*, pages 13–18, 2007.

[11] P. Myllymaki. *Mapping Bayesian Networks to Stochastic Neural Networks: A Foundation for Hybrid Bayesian-Neural systems*. Ph.D. dissertation, University of Helsinki, No. A-1995-1, 1995.

[12] R. M. Neal. Transferring prior information between models using imaginary data. Technical Report No. 0108, Dept. of Statistics, University of Toronto, 2001.

[13] P. Niyogi, T. Poggio, and F. Girosi. Incorporating prior information in machine learning by creating virtual examples. *Proceedings of the IEEE*, 86(11):2196–2209, 1998.

[14] R. Sowmya and R. J. Mooney. Theory refinement for Bayesian networks with hidden variables. In *Proc. 15th International Conf. on Machine Learning*, pages 454–462. Morgan Kaufmann, San Francisco, CA, 1998.

[15] G. Towell and J. Shavlik. Knowledge-based artificial neural networks. *Artificial Intelligence*, 70:119–165, 1994.

[16] 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.

[17] S. Waterhouse, D. MacKay, and T. Robinson. Bayesian methods for mixtures of experts. In *Neural Inf. Proc. Systems*, volume 8, pages 351–357, 1995.