JMLR: Workshop and Conference Proceedings 8: 30-43 Machine Learning in Systems Biology

Evaluation of a Bayesian model-based approach in GA studies

Gábor Hullám GABOR.HULLAM@MIT.BME.HU
Department of Measurement and Information Systems
University of Technology and Economics
Budapest, H-1117, Hungary

Péter Antal ANTAL@MIT.BME.HU
Department of Measurement and Information Systems
University of Technology and Economics
Budapest, H-1117, Hungary

Csaba Szalai SZALAI@HEIMPALKORHAZ.HU
Inflammation Biology and Immunogenomics Research Group
Semmelweis University
Budapest, H-1445, Hungary

András Falus FALAND@DGCI.SOTE.HU
Department of Genetics, Cell- and Immunobiology
Semmelweis University
Budapest, H-1445, Hungary

Editor: Sašo Džeroski, Pierre Geurts, and Juho Rousu

Abstract

In a typical Genetic Association Study (GAS) several hundreds to millions of genomic variables are measured and tested for association with a given set of a phenotypic variables (e.g., a given disease state or a complete expression profile), with the aim of identifying the genetic background of complex, multifactorial diseases. These highly varying requirements resulted in a number of different statistical tools applying different approaches either bayesian or non-bayesian, model-based or conditional. In this paper we evaluate dedicated GAS tools and general purpose feature subset selection (FSS) tools including a Bayesian model-based tool *BMLA* in a GAS context. In the evaluation we used an artificial data set generated from a reference model with 113 genotypic variables that was based on a real-world genotype data.

Keywords: Bayesian Networks, Feature Subset Selection, Genetic Association Studies

1. Introduction

The research on genomic variability received much attention in the past years as one of the most promising areas of genetics research, and several tools were created to aid GAS analysis, particularly the discovery of gene-gene and gene-environment interactions (for a review on detecting gene-gene interactions see Cordell (2009), for an overview on the Bayesian approach see Stephens and Balding (2009)).

Earlier multivariate methods designed to detect associations between genotypic variables and the target variable in GAS include *MDR* (Multifactor Dimensionality Reduction, Moore et al. (2006)), a nonparametric model-free data mining method, which can also be used in conjunction with several filters such as *ReliefF* (Robnik-Sikonja and Kononenko, 2003; Moore and White, 2007). Other GAS methods such as *BEAM* (Bayesian Epistasis Association Mapping, Zhang and Liu (2007)) are based on computing posterior probabilities (i.e., the probability whether each marker set is associated with the disease) via a Markov chain Monte Carlo method. Another approach is to compute metrics that indicate associations, e.g. Bayes factors such as in the case of *BIMBAM* (Bayesian IMputation-Based Association Mapping, Servin and Stephens (2007)).

In this paper we compare the performance of these methods and our previously introduced Bayesian network based method in a typical GAS context assuming that the primary goal is (1) the analysis of the relevance of input variables (e.g. SNPs) w.r.t. the target variable (e.g. an indicator of a certain disease); and (2) the exploration of the interdependencies of relevant variables. Note that there are other applicable methods such as *PIA* (Polymorphism Interaction Analysis, Mechanic et al. (2008)), and an interaction search method based on external, a priori networks (Emily et al., 2009) that were not included in this comparative study.

Earlier, we presented the methodology of the Bayesian Multilevel Analysis (BMLA) of the relevance of input variables in Antal et al. (2006); Millinghoffer et al. (2007). BMLA enables the analysis of relevance at different abstraction levels: model-based pairwise relevance, relevance of variable sets, and interaction models of relevant variables. In the Bayesian model averaging framework each of these levels corresponds to a structural property of Bayesian networks (i.e. Markov Blanket Memberships, Markov Blanket sets, and Markov Blanket graphs respectively). The essence of BMLA is the estimation of posteriors of these structural properties by using a Markov chain Monte Carlo method (for details see Antal et al. (2006)). Based on the resulting estimated posteriors the relevance of subsets of input variables can be assessed. Furthermore, Markov blanket graph posteriors provide principled confidence measures for multivariate variable selection and facilitates the identification of interaction models of relevant variables (for an extension with scalable structural properties see Antal et al. (2008)).

Due to its model-based semantics, the BMLA method indicates "direct" associations, and omits the transitive dependencies, which results in more peaked posteriors than in conditional models working with associations. Furthermore, its model-based nature facilitates the application of more powerful imputation methods and prior incorporation. In contrast, conditional models are typically simple such as logistic regression or black boxes, and the implementation of imputation methods and prior incorporation requires additional efforts. Finally, as a Bayesian approach, it has an in- built automated correction for the multiple testing problem (i.e., the posterior is less peaked with increasing model complexity and decreasing sample size) contrary to the hypothesis testing framework.

From another point of view, the Bayesian statistical framework is ideal for trading sample complexity for computational complexity (i.e., applying computation intensive model-averaging to quantify the sufficiency of the data). Bayesian conditional methods e.g. using logistic regression or multilayer perceptrons, are widely used in biomedicine and in GASs (e.g., see Antal et al. (2003); Kooperberg and Ruczinski (2005); Balding (2006); Province

and Borecki (2008); Park and Hastie (2007); Stephens and Balding (2009)). Although the conditional approach is capable of multivariate analysis and also copes with conditional relevance and interactions, the model-based approach offers many advantages such as listed below.

1. *Strong relevance.* Clear semantics for the explicit, faithful representation of strongly relevant (e.g. non-transitive) relations (cf. associations)

2. *Structure posterior.* In case of complete data the parameters can be analytically marginalized.

3. *Independence map and causal structure.* It offers a graphical representation for the dependence-independence structure, (e.g. about interactions and conditional relevance) and optionally for the causal relations (Pearl, 2000; Glymour and Cooper, 1999).

4. *Multiple-targets.* It is applicable for multiple target variables (Antal et al., 2008).

5. *Incomplete data.* It is applicable for incomplete data sets.

We investigated several probabilistic domain models with promising results (Antal et al., 2006, 2008), relying on these properties of our model-based framework.

2. Probabilistic concepts for GAS

Despite the centrality of "associations" in GASs the refinements of this concept are hardly gaining acceptance in biomedicine, such as strong and weak relevance (cf. non-transitivity and redundancy), conditional relevance (cf. pure interaction), contextual relevance, multivariate relevance (cf. epistasis, complete interaction model, haplotype-level association) or causal relevance. In the following paragraphs, we provide a partial overview on these probabilistic concepts, for a detailed description see Antal et al. (2008).

We start with a pure probabilistic definition of relevance, which is defined in a model-free, method-free, cost-free and data-free way (Kohavi and John, 1997).

Definition 1 (Relevance) *A feature (stochastic variable) X_i is strongly relevant to Y, if there exists some $X_i = x_i, Y = y$ 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.*

A probabilistic definition of relevance can also be given for a set of variables $\mathbf{X}'$ based on the concept of Markov blanket (Pearl, 1988).

Definition 2 (Markov boundary) *A set of variables $\mathbf{X}' \subseteq \boldsymbol{V}$ is called a* Markov blanket *set of X_i w.r.t. the distribution $p(\boldsymbol{V})$, if $(X_i \perp\!\!\!\perp V \setminus \mathbf{X}'|\mathbf{X}')_p$, where $\perp\!\!\!\perp$ denotes conditional independence. A minimal Markov blanket is called* $\mathrm{MBS}_p(X_i)$ Markov boundary.

For the representation of probabilistic relevance, Bayesian networks (BNs) are an adequate choice, since their structural properties are capable of serving such a purpose (Pearl, 1988). They even allow the unambiguous BN representation of relevant variables under a sufficient condition defined in Theorem 1.

Theorem 1 *For a distribution p defined by Bayesian network (G, θ) the variables* $\mathrm{bd}(Y, G)$ *form a Markov blanket of Y, where* $\mathrm{bd}(Y, G)$ *denotes the set of parents, children and the children's other parents for Y (Pearl, 1988). If the distribution p is stable w.r.t. the DAG G, then* $\mathrm{bd}(Y, G)$ *forms a unique and minimal Markov blanket of Y,* $\mathrm{MBS}_p(Y)$ *and* $X_i \in \mathrm{MBS}_p(Y)$ *iff X_i is strongly relevant (Tsamardinos and Aliferis, 2003).*

Note that in typical Bayesian scenarios (e.g., in case of Dirichlet distributions applied in the paper to specify $p(\theta|G)$), the graph-theoretic neighborhood $\mathrm{bd}(Y, G)$ is the unique Markov Boundary with probability 1, i.e. the parameterizations encoding independencies have the measure of 0 (Meek, 1995).

The induced (symmetric) pairwise relation $\mathrm{MBM}(Y, X_j, G)$ w.r.t. G between Y and X_j is called *Markov blanket membership*. $\mathrm{MBM}(Y, X_j, G)$ indicates whether X_j is in $\mathrm{bd}(Y, G)$ (i.e. X_j is an element of the Markov blanket set of Y).

To include interaction terms into the dependency model of a given variable we proposed the use of the Markov Blanket Graph (MBG) property, a.k.a. classification subgraph (Acid et al., 2005; Antal et al., 2006).

Definition 3 (Markov Blanket Graph) *A subgraph of Bayesian network structure G is called the Markov Blanket Graph or Mechanism Boundary Graph* $\mathrm{MBG}(Y, G)$ *of variable Y if it includes the nodes in the Markov blanket defined by* $\mathrm{bd}(Y, G)$ *and the incoming edges into Y and into its children.*

Finally, note that the definition of conditional relevance corresponds to the concept of pure interaction.

Definition 4 (Conditional Relevance) *Assume that $\boldsymbol{X} = \boldsymbol{X'} \cup \boldsymbol{C'}$ is relevant for $\boldsymbol{Y}$, that is $(\boldsymbol{Y} \not\perp\!\!\!\perp (\boldsymbol{X'} \cup \boldsymbol{C'}))$, and $(\boldsymbol{X'} \cap \boldsymbol{C'} = \emptyset)$. We say that $\boldsymbol{X'}$ is conditionally relevant if $(\boldsymbol{X'} \perp\!\!\!\perp \boldsymbol{Y})$, but $(\boldsymbol{X'} \not\perp\!\!\!\perp \boldsymbol{Y} | \boldsymbol{C'})$.*

3. GAS Tools

To demonstrate the performance of BMLA compared to other available GAS tools we present the results of a comparative study in Section 4. For this purpose we selected two groups of tools that are capable of analyzing case-control type GASs based on SNP measurements. The first group consists of dedicated GAS tools, designed specifically for GAS analysis, and the second group consists of general purpose feature subset selection methods that are applicable in this GAS context. In the following sections we give a short description for the tools dedicated for GAS analysis.

- *BEAM: Bayesian Epistasis Association Mapping 1.0* (Zhang and Liu, 2007): BEAM uses a Bayesian partitioning model to select SNPs associated with a disease (i.e. the

target variable) and their interactions, and computes the posterior probability that each set is associated with the disease via a Markov chain Monte Carlo method. *http://www.fas.harvard.edu/ junliu/BEAM*

- *BIMBAM: Bayesian IMputation-Based Association Mapping 0.99* (Servin and Stephens, 2007): BIMBAM computes Bayes Factors for each SNP, and multi Bayes factors for combinations of SNPs under a linear or logistic regression of target variable(s) on SNPs.
  http://stephenslab.uchicago.edu/software.html

- *Powermarker 3.25* (Liu and Muse, 2005): PowerMarker contains a set of statistical methods for SNP data analysis. It implements traditional statistical methods for population genetic analysis and also some newly developed methods.
  http://statgen.ncsu.edu/powermarker

- *SNPassoc 1.5.8* (Gonzlez et al., 2007): SNPassoc is an R package that provides tools for the analysis of whole genome association studies. It allows the identification of SNP-disease associations based on generalized linear models (depending on the selected genetic inheritance model) and the analysis of epistasis.
  http://www.creal.cat/jrgonzalez/software.htm

- *SNPMStat 3.1* (Lin et al., 2008): SNPMStat is an association analysis tool for case-control studies. The program performs a standard association analysis and provides estimated odds ratios, standard error estimates, and Armitage trend tests.
  http://www.bios.unc.edu/ lin/software/SNPMStat

Furthermore, we also investigated some general purpose feature subset selection tools, which are as follows:

- *Causal Explorer 1.4* (Aliferis et al., 2003): Causal Explorer is a library of causal discovery algorithms (such as HITON and IAMB) implemented in MatLab. The algorithms are based on Bayesian Network learning theory, and can also be used for variable selection for classification.
  http://discover1.mc.vanderbilt.edu/discover/public/causal_explorer

- *MDR: Multifactor Dimensionality Reduction 2.0.7* (Moore et al., 2006): MDR is a nonparametric and genetic model-free data mining method for detecting nonlinear interactions among discrete genetic and environmental variables. The MDR software also implements a couple of feature selection algorithms to aid the selection of relevant variables. *http://www.multifactordimensionalityreduction.org*

4. Results

We demonstrate the capabilities of BMLA and compare its performance with other GAS tools (presented previously) on an artificial data set, which consists of 5000 complete random samples generated from a reference model containing 113 SNPs (genomic variables) and a clinical variable *Asthma*. The reference model was learned from a real data set containing

1117 samples, and the 113 SNPs were selected from the asthma susceptibility region of chromosome 11q13 (Szalai, 2005).

The clinical variable *Asthma* served as the target variable, and the aim of the comparative study was to identify all the relevant variables w.r.t. this target variable. There are 11 SNPs in total that are relevant and therefore are part of the MBG of *Asthma* (see Fig. 1).

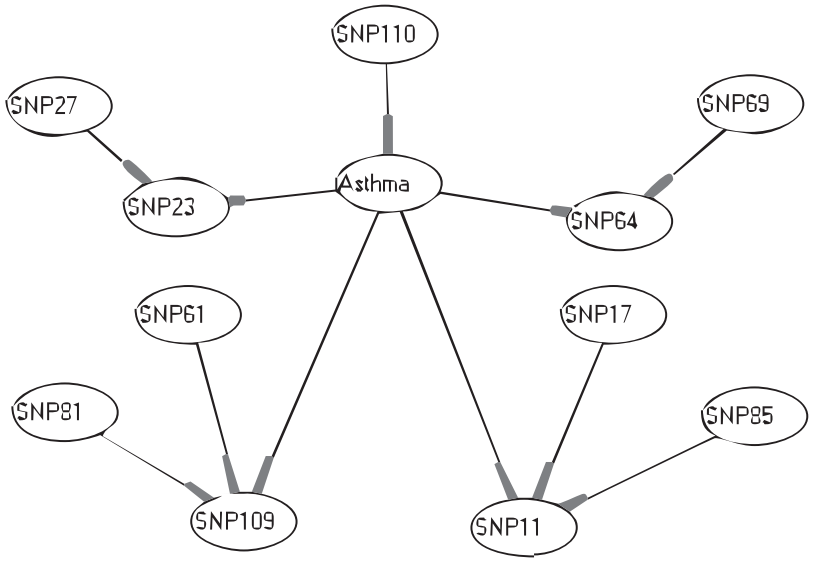

Figure 1: Markov blanket of the reference model containing all relevant SNPs

Out of the 11 relevant SNPs, 5 are in direct relationship with *Asthma* (i.e. 4 children: SNP11, SNP23, SNP64, SNP109 and 1 single parent: SNP110), and the remaining 6 are pure interaction terms (SNP17, SNP27, SNP61, SNP69, SNP81, SNP85). The performance of the tools was assessed by comparing their result set of relevant variables against the 11 relevant SNPs of the reference model. In order to measure the effect of varying sample size (i.e. the sufficiency of the data) the computations were run on data sets with sample sizes 500, 1000 and 5000, where the smaller data sets are subsets of larger ones.

Fig. 2 presents the sensitivity for selecting relevant variables for each of the tested dedicated GAS methods. Apart from the overall sensitivity, the sensitivity for identifying relevant variable subgroups (i.e. direct relationships and pure interactions) is also shown. Fig. 3 shows the accuracy of the GAS methods data set sizes 500,1000 and 5000.

The results confirm preliminary expectations, that is direct relationships are discovered by almost all of the methods, while pure interaction terms are ignored by most. Fig. 4 presents the sensitivity measures and Fig. 5 displays the accuracy of the tested general purpose feature subset selection (FSS) methods. The results indicate that the examined FSS methods identify pure interactions at a significantly higher rate than dedicated GAS tools.

Table 1 shows the sensitivity, specificity and accuracy of the five best performing methods using the complete data set of 5000 samples. Whereas there is only a slight difference in terms of specificity among the methods, the difference in sensitivity is much more significant.

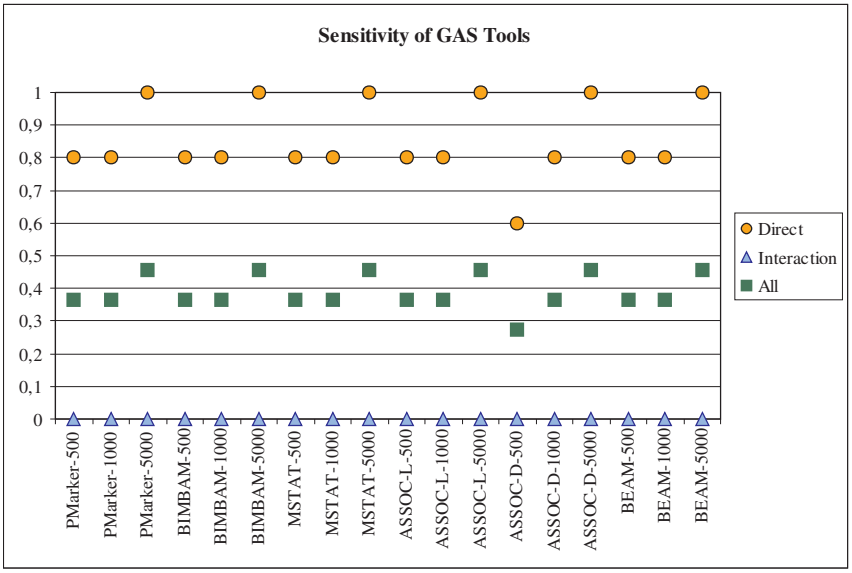

Figure 2: The performance of dedicated GAS tools: Sensitivity for selecting relevant variables. The figure indicates the sensitivity for identifying all associations, and the two main subtypes separately, i.e. direct relationships and interactions using data sets consisting of 500, 1000 and 5000 samples. Methods are denoted as follows: *PMarker* - PowerMarker, *MSTAT* - SNPMStat, *ASSOC-L* - SNPAssoc using a log-additive inheritance model, *ASSOC-D* - SNPAssoc using a dominant inheritance model.

Table 1: Sensitivity, specificity and accuracy of the five best performing methods with different parameter settings. The listed methods include BMLA, HITON-MB - with G^2 statistic and a varying local test set size k, MDR - with TurF and RelieF as two pre-filters, interIAMB - based on mutual information(MI), and the Koller-Sahami algorithm(KS).

Method	Sensitivity	Specificity	Accuracy
BMLA MBM	1	0.99	0.9912
HITON-MB(G2, k=1)	0.7692	0.98	0.9558
HITON-MB(G2, k=2)	0.7692	0.99	0.9646
HITON-MB(G2, k=3)	0.6923	0.99	0.9558
MDR-TurF	0.6154	0.97	0.9292
MDR-Relief	0.5385	0.96	0.9115
interIAMB(MI)	0.4615	0.96	0.9027
KS(k=3)	0.4615	0.97	0.9115

Although in the examined case MBM based results are sufficient to successfully identify the relevant interaction terms of the target variable, however, in several real-world domains a broader, multivariate interpretation is required (e.g. for the discovery of gene-gene in-

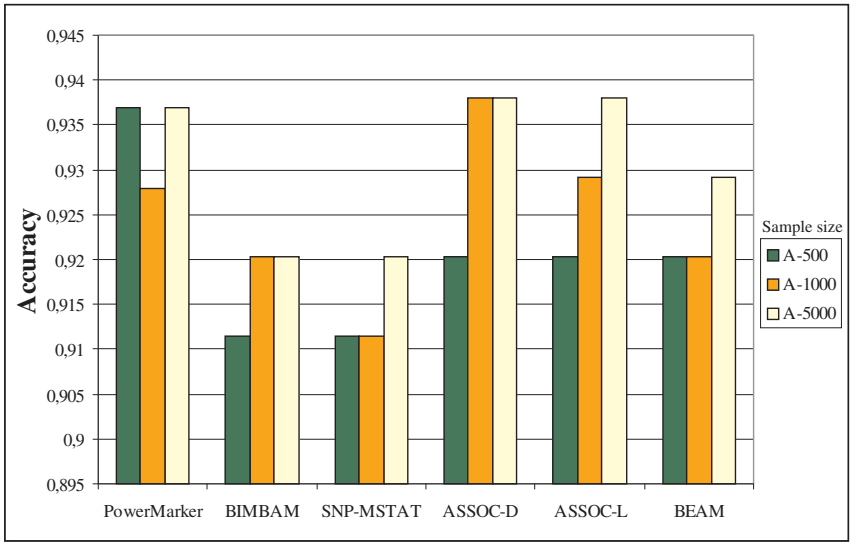

Figure 3: Accuracy of dedicated GAS tools. The figure indicates the accuracy of identifying all associations, using data sets consisting of 500, 1000 and 5000 samples. Methods are denoted as follows: *PMarker* - PowerMarker, *MSTAT* - SNPMStat, *ASSOC-L* - SNPAssoc using a log-additive inheritance model, *ASSOC-D* - SNPAssoc using a dominant inheritance model.

teractions) such as the investigation of MBS properties. Note that MBM probabilities are normatively model-based, despite being pairwise descriptors, since they are generated by Bayesian model averaging. Figures 6 and 7 display the specificity and the accuracy of the 10 most probable MBSs (estimated from the data) respectively for data set sizes 500,1000 and 5000. Note that even in the case of 500 samples the sensitivity of the top MBSs is close to 1 (for details see Table 2).

Table 2: Sensitivity, specificity and accuracy of the ten most probable MBSs.

	500 samples			1000 samples			5000 samples		
Top10 MBS	Sens.	Spec.	Acc.	Sens.	Spec.	Acc.	Sens.	Spec.	Acc
MBS-1	0.909	0.720	0.739	1.000	0.990	0.991	1.000	0.970	0.973
MBS-2	0.909	0.790	0.802	1.000	0.980	0.982	1.000	0.980	0.982
MBS-3	1.000	0.780	0.802	1.000	0.980	0.982	1.000	0.960	0.964
MBS-4	0.909	0.730	0.748	1.000	0.980	0.982	1.000	0.960	0.964
MBS-5	1.000	0.760	0.784	0.909	0.980	0.973	1.000	0.960	0.964
MBS-6	1.000	0.780	0.802	1.000	0.970	0.973	1.000	0.960	0.964
MBS-7	1.000	0.770	0.793	1.000	0.970	0.973	1.000	0.950	0.955
MBS-8	1.000	0.770	0.793	0.909	0.970	0.964	1.000	0.970	0.973
MBS-9	0.909	0.810	0.820	1.000	1.000	1.000	1.000	0.970	0.973
MBS-10	0.909	0.770	0.784	1.000	0.980	0.982	1.000	0.980	0.982

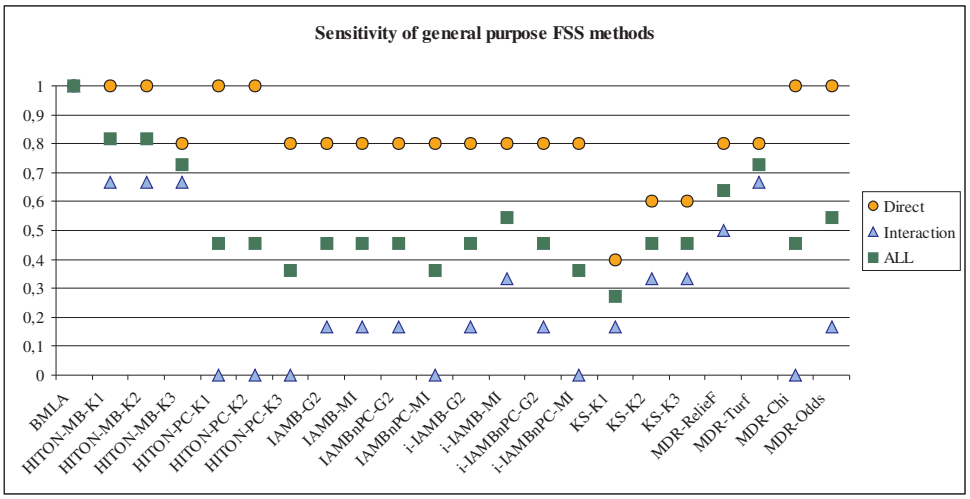

Figure 4: The performance of general purpose FSS tools: Sensitivity for selecting relevant variables. The suffixes for the methods are as follows: *G2* - based on G^2 statistic, *MI* - based on mutual information, Kn - uses a local test set size of n. Note that Turf, Relief, Chi and Odds denote filters used for variable selection with MDR.

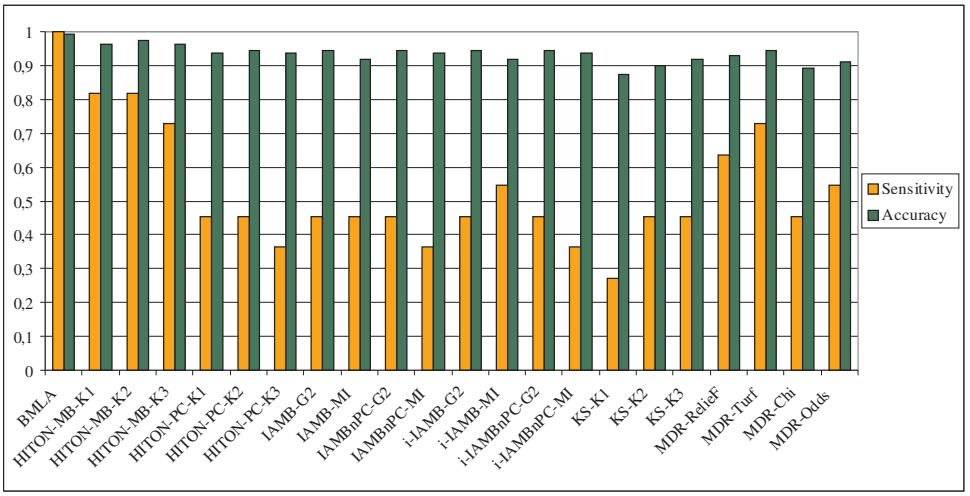

Figure 5: Accuracy of general purpose FSS tools. The suffixes for the methods are the same as in Fig. 4.

In terms of MBS posteriors, in the case of 500 samples, there are hundreds or thousands of MBSs with low posteriors, i.e. the posterior curve seen on Fig. 8 is flat. The posteriors for data sets with 1000 and 5000 samples are more peaked, but there are still numerous MBSs with relatively high posteriors. Although MBSs provide a perfect multivariate view of the examined domain, their cardinality is exponential, therefore their sufficient sample

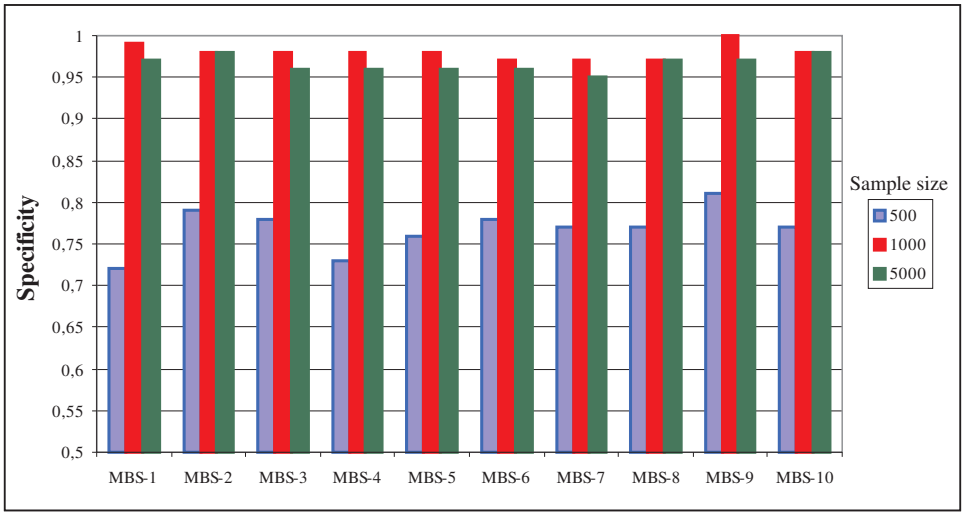

Figure 6: Specificity of the 10 most probable MBSs estimated from the data.

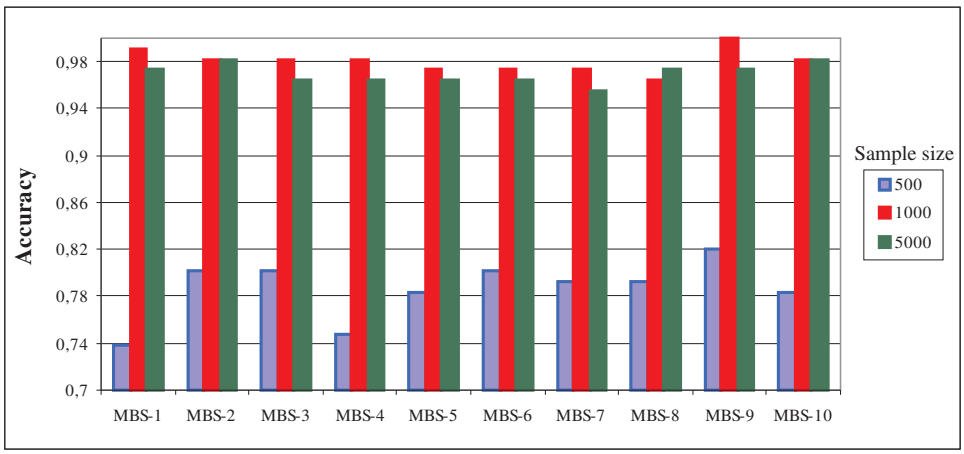

Figure 7: Accuracy of the 10 most probable MBSs estimated from the data.

size is larger than that of MBMs (which have linear cardinality). This is the cause of the phenomenon, that the posteriors for MBS properties are relatively flat compared to posteriors for MBM properties (average posterior for members: 0.9834 and non-members: 0.0109). Note that there is also an intermediate level between MBMs and MBSs, the so called k-MBS property with scalable polynomial cardinality, which we designed especially to be able to select a posterior with appropriate peakness (for details see Antal et al. (2008)).

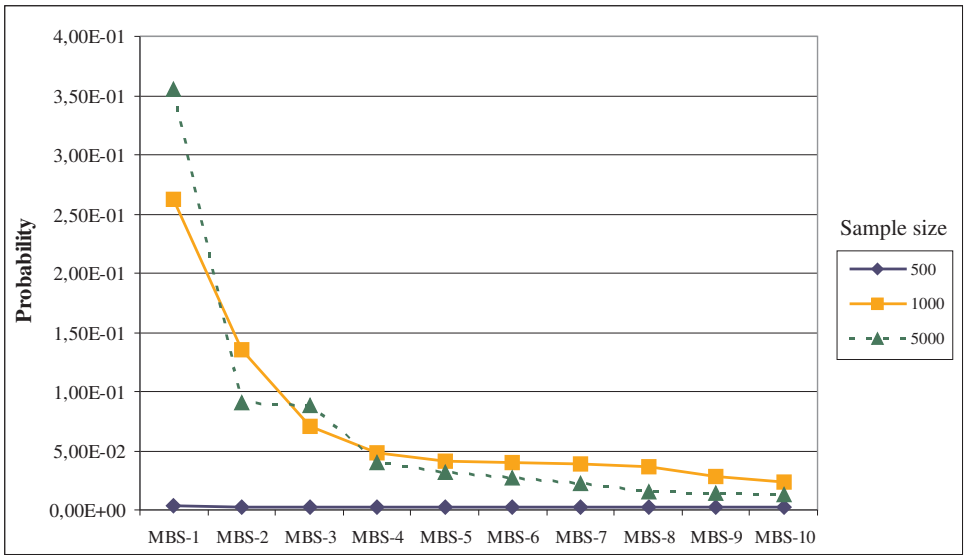

Figure 8: MBS posteriors for data sets of 500, 1000 and 5000 samples.

5. Discussion

The results indicate that the general purpose FSS tools significantly outperformed the tested GAS tools in terms of identifying interactions and conditional relevance in the examined domain.

Basically none of the GAS tools have identified any of the 6 interaction terms of the reference model successfully. Although *BEAM* produced a larger than zero posterior for 3 interaction terms (using 5 chains, 10^6 and $5 * 10^6$ steps for burn-in and length respectively), these were not larger than 0.3 (and thus they were ignored). On the other hand, direct associations were identified by most of the methods correctly. The variation of sensitivity seen on Fig. 2 is due to $SNP110$, which could only be identified from the data set of 5000 samples.

As it can be seen from Table 1, the methods producing the best results all belong to the FSS group. Among them, the best performance was achieved by BMLA. The second best method was HITON (with several different setups), and the third was MDR in conjunction with its filters ReliefF and TurF. Note that the performance of MDR highly depends on the used filter method, since the exhaustive evaluation of all variables is frequently not feasible. The other FSS methods identified only a portion of interactions, and missed even some of the direct associations.

However, note that this evaluation was undertaken in a partial genetic association study (<1000 variables) domain, which is the primary target of the Bayesian network-based BMLA method, and not in a genome-wide association study (10000< variables), which is the main target of dedicated GAS tools. This is also reflected in the computational complexity of these tools. On the other hand, the performance of BMLA comes at a high computational cost and therefore it is currently only applicable in partial genetic association studies. On an Intel® Core^{TM}2 CPU 6700 @ 2.66GHz the execution time of BMLA varied between 5.5

to 7 hours (depending on the used runtime parameters) in contrast to all other methods, amongst which the longest execution time was 39.8 minutes.

The high computational cost of BMLA is due to the estimation of non-analytic posteriors. Although Monte Carlo methods using direct sampling are insensitive to the curse of dimensionality, Markov chain Monte Carlo methods, which are used to estimate the posteriors, are not. They require longer runs in order to ensure convergence and a given confidence level (empirically it takes $\mathcal{O}(n^3)$ steps in this domain, where n denotes the number of variables).

Furthermore, note that the reference model used in this evaluation is a Bayesian network with general multinomial local models, which cannot represent contextual dependencies explicitly (Boutilier et al., 1996). On the other hand, MDR and other methods such as logistic regression are capable of the representation of contextual dependencies, thus in case of more specific reference models their relative performance can be better.

6. Conclusion

The presented comparative study has shown, that general purpose FSS tools can be successfully applied in partial genetic association studies and for the purpose of detecting interactions, particularly conditional relevance, among relevant variables, they perform better than dedicated GAS tools. The results also indicated, that BMLA is an adequate choice for evaluating GASs. Its Bayesian network based approach allowed an excellent reconstruction of the reference model, i.e. the identification of relevant variables either in a direct association or in an interaction with the target variable. Furthermore, being a model-based Bayesian method BMLA offers an automated approach to the correction of multiple testing problem (for the selection of proper priors for GAS in univariate case see Stephens and Balding (2009)).

Finally, note that the capabilities of BMLA can only be fully utilized when the analysis based on MBS and k-MBS structural properties is also carried out (Antal et al., 2008).

Acknowledgments

We would like to acknowledge support for this project from the Hungarian Fund for Scientific Research (OTKA grant 73496 and PD-76348), NKTH TECH 08-A1/2-2008-0120 (Genagrid), and the Janos Bolyai Research Scholarship of the Hungarian Academy of Sciences (P.Antal).

References

S. Acid, L. M. de Campos, and J. G. Castellano. Learning bayesian network classifiers: searching in a space of partially directed acyclic graphs. *Machine Learning*, 59:213–235, 2005.

C.F. Aliferis, I. Tsamardinos, and A. Statnikov. Causal explorer: A probabilistic network learning toolkit for biomedical discovery. In *International Conference on Mathematics*

and Engineering Techniques in Medicine and Biological Sciences (METMBS '03), pages 371–376, 2003.

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

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.

P. Antal, A. Millinghoffer, G. Hullám, Cs. Szalai, and A. Falus. A bayesian view of challenges in feature selection: Feature aggregation, multiple targets, redundancy and interaction. *JMLR Proceeding*, 4:74–89, 2008.

D. J. Balding. A tutorial on statistical methods for population association studies. *Nature*, 7:781–91, 2006.

C. Boutilier, N. Friedman, M. Goldszmidt, and D. Koller. Context-specific independence in bayesian networks. In Eric Horvitz and Finn V. Jensen, editors, *Proc. of the 20th Conf. on Uncertainty in Artificial Intelligence (UAI-1996)*, pages 115–123. Morgan Kaufmann, 1996.

H. J. Cordell. Detecting genegene interactions that underlie human diseases. *Nature Reviews: Genetics*, 10(1):392–404, 2009.

M. Emily, T. Mailund, J. Hein, L. Schauser, and M. H. Schierup. Using biological networks to search for interacting loci in genome-wide association studies. *European Journal of Human Genetics*, pages 1–10, 2009.

C. Glymour and G. F. Cooper. *Computation, Causation, and Discovery*. AAAI Press, 1999.

Juan R. Gonzlez, Llus Armengol, Xavier Sol, Elisabet Guin, Josep M. Mercader, Xavier Estivill, and Vctor Moreno. Snpassoc: an r package to perform whole genome association studies. *Bioinformatics*, 23(5):654–655, 2007.

R. Kohavi and G. H. John. Wrappers for feature subset selection. *Artificial Intelligence*, 97:273–324, 1997.

C. Kooperberg and I. Ruczinski. Identifying interacting snps using monte carlo logic regression. *Genet Epidemiol*, 28(2):157–170, 2005.

D.Y. Lin, Y. Hu, and B.E. Huang. Simple and efficient analysis of disease association with missing genotype data. *American Journal of Human Genetics*, 82(2):444–452, 2008.

K. Liu and S. V. Muse. Powermarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics*, 21(9):2128–2129, 2005.

L. E. Mechanic, B. T. Luke, J. E. Goodman, S. J. Chanock, and C. C. Harris. Polymorphism interaction analysis (pia): a method for investigating complex gene-gene interactions. *BMC Bioinformatics*, 9(1):146–158, 2008.

C. Meek. Causal inference and causal explanation with background knowledge. In Philippe Besnard, Steve Hanks, Philippe Besnard, and Steve Hanks, editors, *Proc. of the 11th Conf. on Uncertainty in Artificial Intelligence (UAI-1995)*, pages 403–410. Morgan Kaufmann, 1995.

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)*, pages 13–18, 2007.

J. H. Moore and B. C. White. Tuning relieff for genome-wide genetic analysis. In *Lecture Notes in Computer Science: Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics*, pages 166–175. Springer Berlin - Heidelberg, 2007.

J.H. Moore, J.C. Gilbert, C.T. Tsai, F.T. Chiang, T. Holden, N. Barney, and B.C. White . A flexible computational framework for detecting, characterizing, and interpreting statistical patterns of epistasis in genetic studies of human disease susceptibility. *Journal of Theoretical Biology*, 241(2):252–261, 2006.

M. Y. Park and T. Hastie. Penalized logistic regression for detecting gene interactions. *Biostatistics*, 9(1):30–50, 2007.

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

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

M. A. Province and I. B. Borecki. Gathering the gold dust: Methods for assessing the aggregate impact of small effect genes in genomic scans. In *Proc. of the Pacific Symposium on Biocomputing (PSB08)*, volume 13, pages 190–200, 2008.

M. Robnik-Sikonja and I. Kononenko. Theoretical and empirical analysis of relieff and rrelieff. *Machine Learning*, 53(1):2369, 2003.

B. Servin and M. Stephens. Imputation-based analysis of association studies: candidate genes and quantitative traits. *PLoS Genetics*, 3(7):e114, 2007.

M. Stephens and D.J. Balding. Bayesian statistical methods for genetic association studies. *Nature Reviews: Genetics*, 10(10):681–690, 2009.

C. Szalai. Genomic investigation of asthma in human and animal models. In A. Falus, editor, *Immunogenomics and Human Disease*, pages 419–441. Wiley, London, 2005.

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.

Y. Zhang and J. S Liu. Bayesian inference of epistatic interactions in case-control studies. *Nature Genetics*, 39(9):1167–1173, 2007.