Approximate inference for medical diagnosis

WAJJ Wiegerinck[1], HJ Kappen[1], EWMT ter Braak [2], WJPP ter Burg [3],
MJ Nijman [1], YL O [2], and JP Neijt [2]

(1) Foundation for Neural Networks, University of Nijmegen, Nijmegen, The Netherlands
(2) University Medical Centre, University of Utrecht, Utrecht, the Netherlands
(3) Academic Medical Centre, University of Amsterdam, Amsterdam, The Netherlands

Abstract

Computer-based diagnostic decision support systems (DSS) will play an increasingly impor-
tant role in health care. Due to the inherent probabilistic nature of medical diagnosis, a DSS
should preferably be based on a probabilistic model. In particular Bayesian networks provide
a powerful and conceptually transparent formalism for probabilistic modeling. A drawback is
that Bayesian networks become intractable for exact computation if a large medical domain
would be modeled in detail. This has obstructed the development of a useful system for in-
ternal medicine. Advances in approximation techniques, e.g. using variational methods with
tractable structures, have opened new possibilities to deal with the computational problem.
However, the only way to assess the usefulness of these methods for a DSS in practice is by
actually building such a system and evaluating it by users. In the coming years, we aim to
build a DSS for anaemia based on a detailed probabilistic model, and equipped with approxi-
mate methods to study the practical feasibility and the usefulness of this approach in medical
practice.

In this paper, we will sketch how variational techniques with tractable structures can be
used in a typical model for medical diagnosis. We provide numerical results on artificial
problems. In addition, we describe our approach to develop the Bayesian network for the
DSS, and show some preliminary results.

Keywords: medical decision support, Bayesian belief networks, variational approximations.

1 Introduction

Health care will change profoundly with the introduction of clinical diagnostic decision support
systems (DSSs), preferably integrated with electronic patient data and together with on-line com-
puter communication [1]. However, such systems have not yet entered daily clinical practice for
a variety of reasons [2]. Some of the reasons are the lack of adequate computer infrastructure in
hospitals, the lack of standardization of terminology and the poor performance of current medical
diagnostic systems. In this paper we focus on this last issue.

Diagnostic reasoning in the medical domain is a typical example of reasoning with uncertainty.
This uncertainty has different sources: missing patient information, uncertainty in medical tests
results or observations, and the uncertainty about the physiological processes involved. A model
on which a DSS is based should be able to deal with these uncertainties. The different systems
that have been developed so far use a variety of modeling approaches which can be roughly divided
into two categories: The large systems, that attempt to cover the whole of internal medicine use
a rule-based approach with some rather heuristic method to quantify uncertainty. These methods
perform poorly in practice [3, 4]. The main reasons are that the modeling of the relations between
diseases and findings is at a very course level. Therefore, the diagnoses suggested by these systems
are too superficial for clinical use. Secondly, the diagnostic process requires reasoning from causes

1

to effects (diseases → finding) and vise versa at the same time. The rule based approach, together with the heuristics for uncertainty, is not well suited for such bidirectional reasoning.

For smaller systems, the probabilistic approach is typically used. The probabilistic approach has the important advantage of mathematical consistency and correctness. In particular Bayesian networks (see e.g. [5, 6, 7]) provide a powerful and conceptual transparent formalism for probabilistic modeling. In addition, they allow for easy integration of domain knowledge and learning from data. Systems that are based on detailed modeling have been restricted to a relatively small domain [8, 9]. The reason for this restriction is that Bayesian networks will become intractable for exact computation if a large medical domain would be modeled in detail.

To proceed one has to rely on *approximate* computations. Recently, variational methods for approximation are becoming increasingly popular [10, 11, 12]. An advantage of variational methods techniques is that they provide bounds on the quantity of interest in contrast to stochastic sampling methods which may yield unreliable results due to finite sampling times. Until now, variational approximations have been less widely applied than Monte Carlo methods, arguably since their use is not so straightforward. We argue that variational methods are indeed applicable to large, detailed Bayesian networks for medical diagnosis constructed by human experts.

Although the formalism of Bayesian networks is very powerful, the construction of networks for medical diagnosis is not straightforward. A learning approach depends crucially on the availability of high quality patient data. In particular, rare disorders should be well covered. In general, unfortunately, this is rather exception than rule [13]. Therefore, to reach a successful diagnostic DSS requires explicit modeling effort by human experts. The existing medical literature is not sufficient to define the probabilistic model. Not all probabilistic relations between variables have been documented. But it provides a useful starting point for model design. Once a minimal performance is thus obtained, the model can be improved by learning from patient data.

The paper is organised as follows. In section 2 we discuss the modular structure of a broad and detailed probabilistic model for medical diagnosis constructed by human experts. In section 3 we show how variational methods can provide approximations for the conditional probabilities in these networks. This is illustrated by results on the Lauritzen's chest clinic model and a random intractable network. In section 4 we describe our approach to model medical diagnosis, and we give a short description of our system for megaloblastic anaemia, called Promedas. We conclude with a discussion in section 5.

2 Probabilistic modeling in the medical domain

We here outline what the structure of a broad and detailed Bayesian network will typically look like. This is based on an extrapolation of our current modeling experiences. Details of the medical domain are beyond the scope of this paper and are discussed elsewhere [14].

The variables to consider in the network are of different types. There are diseases variables, which are typically of the binary type, signalling whether a disease is present or not. The findings encode the results from laboratory measurements, physical examination etc. As a simplification, these variables are discretized, with medically relevant cut-off points. In practice, such discretisation does not lead to significant loss of information. In addition, there are prior variables that describe the patient, such as sex and age.

In constructing the graph for the Bayesian network, human experts mostly use "causal" relationships between variables as a guideline (the arrows in fig. 1). Often, the expert can relate (large numbers of) variables via additional hidden variables. These hidden variables may represent pathophysiological variables that are known to have certain relations to the observable variables, but are themselves not accessible during clinical investigation. Often, the use of hidden variables results in a simplified and more transparent network.

The majority of probabilistic relations between the variables involve only a small number of parents. Consequently, modeling using explicit probability tables is feasible. These are estimated on the basis of data in the literature or on "educated guesses" based on local statistics/experience if no data from the literature are available.

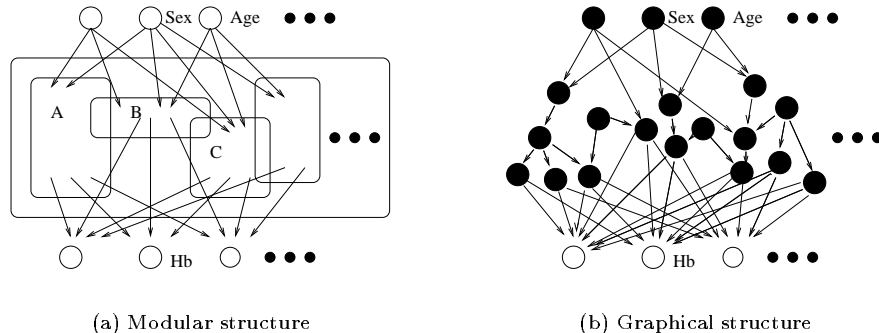|	
(a) Modular structure	(b) Graphical structure

Figure 1: Modular and graphical network structure. Left: modular structure of the network. A, B, C . . . represent (overlapping) sub-domains. Each sub-domain is modeled by a number of nodes (cf. right figure) representing variables that are relevant in that domain. The upper nodes, e.g. 'sex' and 'age' represent common ancestors of nodes in several sub-domains. The lower nodes, e.g. 'Hb' represent common children of nodes in several sub-domains (e.g. related to anaemia). Right: underlying bipartite graphical structure of same network. Filled circles: nodes in sub-domains and their common ancestors. Open circles: common children

Medical experts tend to divide knowledge concerning a medical domain into sub-domains with a relatively small overlap. Therefore, the network will typically have a modular structure (cf. fig. 1). Each module represents a disease with its relevant findings. In practice, the modules are rather small, containing between 20-50 variables. Different modules are connected via shared variables (e.g. pathophysiological variables that are relevant in different modules), common prior nodes, and/ or common findings nodes. The computational complexity of the network $\mathcal{N}_1$ consisting of the modules and their parents (black nodes in fig. 1) can be assumed to be tractable.

The probabilistic relations for the findings require somewhat more care. For example, 'hemoglobin level' (Hb) is a variable whose value is affected by many diseases. Such nodes may have parents in many sub-domains. This makes the use of a conditional probability table not feasible, as the size of the table grows exponentially in the number of parents. Fortunately, this is neither necessary, since medical experts are likely to agree with a 'sum of univariate relations' between this finding and its parents. This can be modeled by a noisy-OR gate [6]. Assuming binary variables for convenience, the noisy-OR gate for finding $f = 0, 1$ given its parents $S_k = 0, 1, k = 1, \ldots, n_f$ is

$$P(f = 0|S_1, \ldots, S_{n_f}) \;=\; \exp(-z_f) \tag{1}$$

where $z_f \equiv \sum_k^{n_f} \theta_{fk} S_k + \theta_{f0}$, with parameters satisfying $0 < \theta_{fk} < \infty$. The total model on all variables is given as $P(S_1, \ldots, S_n) = \prod_k^n P(S_k|\pi_k)$, where the conditional probability distributions are either tables or noisy-ORs. π_k denotes the state of the parent nodes of node k. Thus, if node k has n_k parents, π_k can take on 2^{n_k} values.

Even though the findings are modeled in a compact way, instantiated findings are likely to cause computational problems in inference. Let $\{f_{+,-}\}$ denote the subset of instantiated positive and negative findings, respectively. The marginal probability of the findings is

$$\sum_{\{S_H\}} P(S_H, S_{\{f_+\}} = 1, S_{\{f_-\}} = 0) = \sum_{\{S_H\}} \prod_k P(S_k|\pi_k) \prod_{f_+}(1 - \exp z_{f_+}) \prod_{f_-} \exp z_{f_-} \tag{2}$$

where S_H are the states of the remaining nodes in $\mathcal{N}_1$. As a result of the linearity of z_f in S_k, $\exp z_{f_-}$ factorizes as a product over parent states. Thus, the negative findings can be absorbed in $\mathcal{N}_1$ via the transformation $P \to R$, where $R(S_k|\pi_k) = \exp(-\theta_{fk} S_k) P(S_k|\pi_k)$ $(k \in \mathcal{N}_1)$. These terms can be summed efficiently in linear time. This is not possible with positive findings. Therefore the computational costs will grow exponentially in the number of positive findings f_+ [11, 15], and the total network will be intractable.

3 Variational Approximations

In general, the problem of inference is to find the conditional probability distribution $P(S_i|E)$ of each of the nodes i given the evidence E. If P is intractable, one has to approximate these conditional probabilities. In the variational method, the intractable probability distribution $P(S|E) = P_E(S)$ is approximated by a tractable distribution $Q(S)$ (on the non-evidential nodes). Then Q is used to compute the node probabilities $Q(S_i)$. To construct Q, one first has to define a tractable graphical structure for Q: $Q(S) = \prod_\gamma Q(S_\gamma|\pi_\gamma)$, [16, 17, 12]. The next step is to optimize the parameters of Q such that the Kullback-Leibler (KL) divergence between Q and P_E,

$$D(Q, P_E) = \sum_{\{S\}} Q(S) \log \frac{Q(S)}{P_E(S)} \tag{3}$$

is minimized. The KL-divergence is related to the difference of the marginals of Q and P_E,

$$\max_i |P(S_i|E) - Q(S_i)| \le \sqrt{\frac{1}{2} D(Q, P_E)} \tag{4}$$

(see [18]). The KL-divergence satisfies $D(Q, P_E) \ge 0$, and $D(Q, P_E) = 0 \Leftrightarrow Q = P_E$. Using $P(S|E) = P(S, E)/P(E)$ and substituting the graphical structures for P and Q, we can rewrite D as

$$D(Q, P_E) = \sum_{\{S\}} Q(S) \left(\sum_\gamma \log Q(S_\gamma|\pi_\gamma) - \sum_i \log P(S_i|\pi_i) \right) + \text{constant.} \tag{5}$$

Parent sets π_γ and π_i are understood with respect to the probability distribution in which they appear and are in principle different. As a consequence of the factorization of $P(S, E)$ into conditionals, the average $\langle \log P(S, E) \rangle_Q$ reduces to the sum of local averages $\sum_i \langle \log(P(S_i|\pi_i)) \rangle_Q$, which facilitates the tractability of D.

$D(Q, P_E)$ depends on the numerical values of the conditional probability tables $Q(S_i|\pi_i)$. Setting the gradient of D with respect to these parameters equal to zero, yields the equations

$$
\begin{aligned}
Q(S_i|\pi_i) &= \frac{1}{Z_i} \exp \langle \log P(S, E) - \log Q(S) \rangle_{Q(S|S_i, \pi_i)} \\
&= \frac{1}{Z_i} \exp \left\langle \sum_k \log P(S_k|\pi_k) - \sum_\gamma \log Q(S_\gamma|\pi_\gamma) \right\rangle_{Q(S|S_i, \pi_i)}.
\end{aligned} \tag{6}
$$

$\langle \cdot \rangle_{Q(S|S_i, \pi_i)}$ denotes the average with respect to Q with node i and its parents clamped to S_i and π_i, respectively. Z_i is a normalisation factor. Eqs. 6 are a coupled set of non-linear equations that must be solved for $Q(S_i|\pi_i)$. For each i, the right hand side of Eqs. 6 does not depend on the parameters $Q(s_i|\pi_i)$. This means that asynchronous iteration of Eqs. 6 is guaranteed to converge to a local minimum of the KL-divergence.

The quality of the approximation depends strongly on the structure of Q. The simplest approach is the so called mean-field approach, in which the graph of Q is completely disconnected, i.e. $Q(S) = \prod_i Q(S_i)$. Then Eqs. 6 reduces to the standard mean field equations

$$Q(S_i) = \frac{1}{Z_i} \exp \left\langle \sum_k \log P(S_k|\pi_k) \right\rangle_{Q(S|S_i)}$$

The other extreme is to factorize Q according to a triangulated graph [5, 7] of P. In this case, iteration of Eqs. 6 leads to the solution $Q = P_E$ and $D = 0$. This solution is only theoretically of interest, since its computational complexity is equal to the original inference problem. However, it indicates that the variational approach using structure interpolates between the standard mean field theory and the exact solution. In general one must choose a structure for Q that is a good compromise between approximation error and complexity.

4

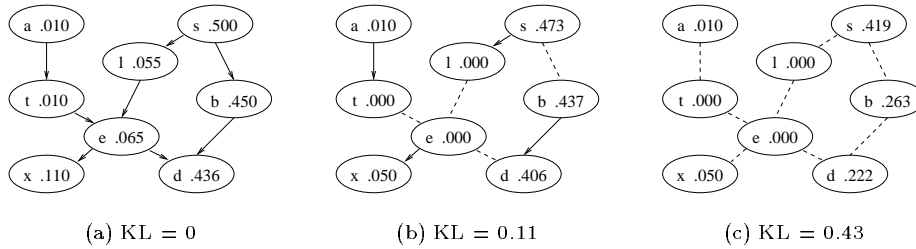(a) KL = 0   (b) KL = 0.11   (c) KL = 0.43

Figure 2: Chest clinic model. Arrows: links in the exact model (a): exact model with marginal probabilities. (b-c): approximating models with approximated marginal probabilities. Arrows indicate the graphical structure of the exact and approximating models. Dashed lines indicate the underlying links in the exact model. KL is the KL-divergence $D(Q, P)$ between the approximating model Q and the true model P.

The complexity of the variational method is at least exponential in the parent size of the original model P, since it requires the computation of averages of the form $\langle \log P(S_k|\pi_k)\rangle_Q$. This is also true for noisy-OR gates. This means that computational advantage is obtained if the parent size is much smaller than the clique [5, 7] size of P. For large parent sets, one can use additional approximations. For instance, for the noisy-OR gates with positive findings, we can use the approximation proposed in [15]

$$\langle \log(1 - \exp(-z_f))\rangle_Q \approx -\sum_{\kappa=0}^{K} \log(\langle 1 + \exp(-2^\kappa z_f)\rangle_Q) \tag{7}$$

which is tractable if Q is tractable.

3.1 Simulation results

We illustrate the theory by two toy problems. The first one is inference in the chest clinic model (see [5] for more details). The model is defined on binary variables $\{a, t, s, l, b, e, x, d\}$, with conditional probability tables given in table 1.

We compared exact marginals with approximate marginals using the approximating models in figure 2. ¿From the results, we can conclude that adding structure to the approximating network decreases the error in the approximation. However, we also can see from the simulation results that even the fully disconnected mean field approximation is qualitatively correct (maximum error between marginals $P(S_i)$ and $Q(S_i)$ is about 0.2).

In the second toy problem we simulated inference in networks with a structure that is similar to the structure we encounter in medical diagnosis, as discussed in section 2. We generated models with graphical structure as in figure 3. The upper node is a mixture node with m mixture components. The next layer consists of $n_d + 1$ binary nodes. These two layers represent prior probabilities and hidden pathophysiological mechanisms. The third layer of n_d binary nodes s_d represent the diseases. Each of these nodes has two parents in the preceding layer. Up to this

$p(a)$	=	.01	$p(l\|s)$	=	.10	$p(e\|l,t)$	=	1	$p(d\|e,b)$	=	0.90
			$p(l\|\bar{s})$	=	.01	$p(e\|l,\bar{t})$	=	1	$p(d\|e,\bar{b})$	=	0.70
$p(t\|a)$	=	.05				$p(e\|\bar{l},t)$	=	1	$p(d\|\bar{e},b)$	=	0.80
$p(t\|\bar{a})$	=	.01	$p(b\|s)$	=	.60	$p(e\|\bar{l},\bar{t})$	=	0	$p(d\|\bar{e},\bar{b})$	=	0.10
			$p(b\|\bar{s})$	=	.30						
$p(s)$	=	.50				$p(x\|e)$	=	.98			
						$p(x\|\bar{e})$	=	.05			

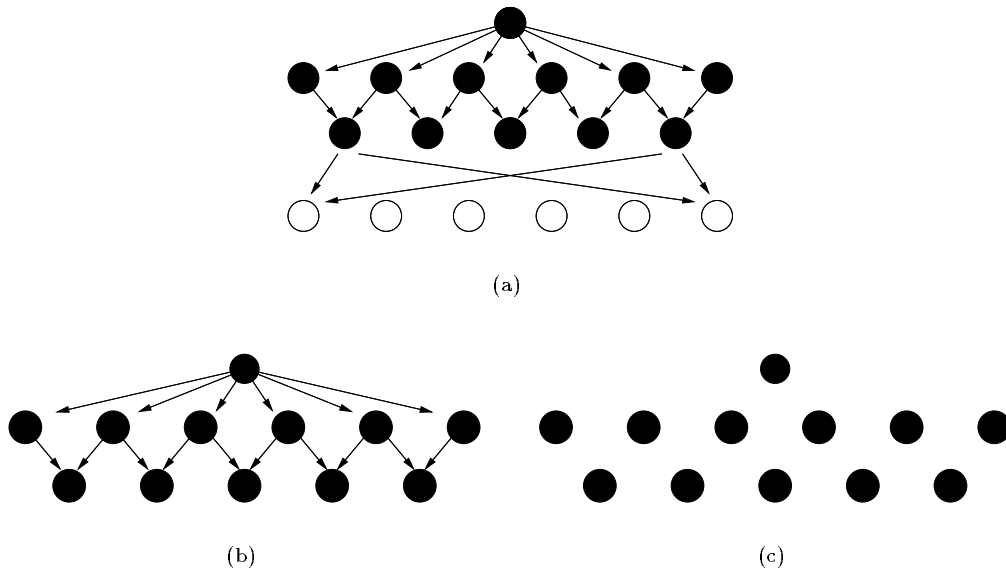Table 1: Conditional probability table for chest clinic model (Asia problem).

Figure 3: (a): Graphical structure of artificially generated probability distribution P. Non-evidential nodes are black. White nodes are positive findings. (b) and (c): Graphical structure on the non-evidential nodes of the approximating distributions Q.

layer the network is tractable. We refer to this part of the network as $\mathcal{N}_1$. Finally, there is a layer of n_f findings s_f. These are modeled as noisy-OR gates, receiving inputs from all the nodes of the preceding disease layer. In the simulations, the findings are all clamped on positive values $s_f = 1$. This makes the network intractable for large n_d and n_f. We choose $m = 10$, $n_f = n_d$ and varied $n_d = 7, \ldots, 13$. Networks of this size are still tractable for exact computation. The values in the probability tables of $\mathcal{N}_1$ are drawn uniformly. The parameters θ_{fd} in the noisy-OR gates are drawn from the exponential distribution and are normalised by a factor $\sqrt{n_d}$. We computed exact and approximated marginals for the diseases s_d. As approximating models we used the model with structure $\mathcal{N}_1$ and a factorized model (fig 3).

In figure 4 we plotted the maximal error $\max_d |Q(s_d) - P(s_d|s_f)|$ as a function of the network size. We also plotted the required computer time for exact and approximate inference as a function of the network size.

We conclude that variational methods using structure significantly improves the quality of approximation, within feasible computer time. In a network with tractable substructures, as can be expected in medical diagnosis, these substructures provide a useful starting point for the approximating model.

4 Promedas, a demonstration DSS

Promedas (PRObabilistic MEdical Diagnostic Advisory System)[1] is a DSS that we are developing for the problem of anaemia. The aim is to use Promedas to assess the usefulness of approximate methods for a DSS in practice. The problem domain anaemia is chosen because we expect that the computational problems described in the previous sections will be encountered in this domain. For instance, anaemia can be subdivided in a large number of sub-domains, each of which share a large number of findings. Furthermore, anaemia is a common medical problem. This facilitates evaluation in practice. To cover the domain completely, we expect that approximately 1000 nodes are needed.

---

[1] A demonstration version of Promedas is available on CD-ROM. See www.mbfys.kun.nl/snn/Research/promedas
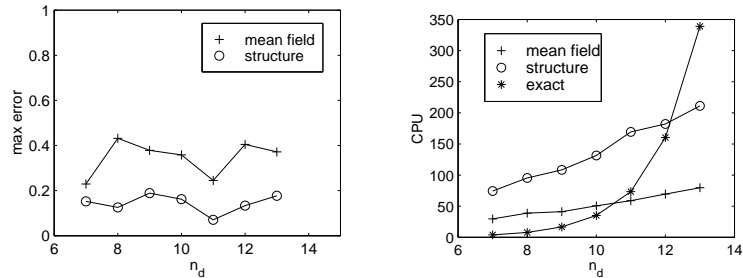
Figure 4: Left: The maximal error as a function of the network size. Right: CPU-time in Matlab seconds for exact and approximate inference as a function of the network size

To develop Promedas, we use our internally developed software environment, called Bayes-Builder. BayesBuilder has graphical tools for network construction, evaluation, and maintenance. So far, Promedas covers megaloblastic anaemia. It is currently based on a network of 91 variables, and is still tractable for exact algorithms.

Promedas consists of a graphical user interface (GUI) to enter patient data and for diagnostic consultation. It provides a differential diagnosis, i.e. the probabilities of potentially relevant diagnoses and the probabilities of potentially involved underlying mechanisms (e.g. pathophysiology) as percentages (ranked in descending order). These probabilities are computed on the basis of the available findings entered in the system. In addition, Promedas computes which additional tests it expects to be most informative to decide about a diagnosis, specified by the user. This information is computed given the values of the variables previously entered and is defined as $I(D,T) = \sum_{D,T} P(D,T) \ln(P(D,T)/P(D)P(T))$ with $P(D,T)$ the joint probability of diagnosis and test result, and $P(D), P(T)$ the marginal probabilities of diagnoses and tests, respectively. These probabilities are computed by marginalizing over all the missing variables in the network. The information is normalised between 0 and 100, and displayed in descending order (see Table 2). In addition, Promedas provides help information, medical background information and pointers to the literature.

In Table 2, we illustrate the capability of the system to guide the medical decision-making process.

5   Discussion

The development of a DSS for comprehensive medical diagnosis in internal medicine represents a great challenge for AI. A broad and detailed probabilistic network is intractable for exact inference in this context. It is currently unknown, whether variational or other approximate methods are sufficiently powerful to provide a practical solution. The "quality of approximation" is to a large extent a user defined (medical) issue, since (1) comparison with exact inference is not possible due to the size of the networks and (2) errors in the approximation will be judged as acceptable not just on their numerical values but more importantly on their medical implications. The only way to assess the usefulness of approximate methods for modeling medical domains is by actually building such a system and evaluating it by users. The Promedas model must be extended to several 100 variables in order to be able to address this issue properly.

Our results on the Asia problem with evidence (not reported here), as well as the results reported here, show that the factorized variational approximation is qualitatively correct in the sense that it correctly estimates whether probabilities are high or low. However, the numerical errors can be rather large. The results of variational approximations using structure gives a significant improvement. Our results seem to indicate that this improvment is independent of the problem size. The choice of optimal structure for Q is a topic for further investigation.

Visit 1

		Diagnosis
36	%	Cobalamin deficiency, no cause established
25	%	Pernicious Anemia
2	%	Cobalamin def. due to total gastrectomy
2	%	Cobalamin def. or abnormal intrinsic factor (IF)
2	%	Cobalamin def. due to Bilroth I or II gastrectomy
2	%	Cobalamin def. due to caustic ingestion

	Test
30	Anti-parietal cell antibodies
26	Anti-intrinsic factor antibodies
17	Pentagastine test
14	Serum gastrin level
10	Schilling test with ovalbumin vit B12
1	Caustic ingestion in the past
1	Total gastrectomy in the past

Selected test: Anti-parietal cell antibodies   *Result:* Yes

Visit 2

		Diagnosis
60	%	Pernicious Anemia
19	%	Cobalamin deficiency, no cause established
1	%	Cobalamin def. due to Bilroth I or II gastrectomy
1	%	Cobalamin def. due to total gastrectomy
1	%	Cobalamin def. due to caustic ingestion
1	%	Cobalamin def. due or abnormal intrinsic factor (IF)

	Test
26	Anti-intrinsic factor antibodies
24	Pentagastine test
19	Serum gastrin level
15	Schilling test with ovalbumin vit B12
1	Caustic ingestion in the past
1	Total gastrectomy in the past
1	Bilroth I or II gastrectomy in the past

Selected test: Anti-intrinsic factor antibodies   *Result:* Yes

Visit 3

		Diagnosis
97	%	Pernicious Anemia
2	%	Cobalamin deficiency, no cause established

	Test
3	Pentagastine test
2	Serum gastrin level
1	Schilling test with ovalbumin vit B12

Selected test: Pentagastine test   *Result:* Yes

Visit 4

		Diagnosis
98	%	Pernicious Anemia
1	%	Cobalamin deficiency, no cause established

Test
all tests < 1

Table 2: Results of the decision support of Promedas. At visit 1, a small number of patient findings are entered in the system (not shown). Promedas displays the most likely diagnoses. After selecting Pernicious Anemia, Promedas computes the expected information for all unknown tests. We select the test 'Anti-parietal cel antibodies', whose subsequent measurement yields 'Yes'. Ranking of hidden mechanisms are not shown. For subsequent visits of the patient we see how the additional test results change the likelihood of the diagnoses.

Acknowledgements

References

[1] Y. Reisman. Computer-based clinical decision aids. a review of methods and assessment of systems. *Med. Inform.*, 21(3):179–197, 1996.

[2] B. Brigl, P. Ringleb, T. Steiner, P. Knaup, W. Hacke, and R. Haux. An integrated approach for a knowledge-based clinical workstation: Architecture and experience. *Methods of Information in Medicine*, 37:16–25, 1998.

[3] AS Elstein, CP Friedman, FM Wolf, G Murphy, J Miller, P Fine, P Heckerling, T Miller, J Sisson, S Barlas, K Biolsi, M Ng, X Mei, and T Franz. Performance of four computer-based diagnostic systems. *N-Engl-J-Med.*, 330(25):1792–6, 1994.

[4] ES Berner, JR Jackson, and J Algina. Relationships among performance scores of four diagnostic decision support systems. *J-Am-Med-Inform-Assoc.*, 3(3):208–15, 1996.

[5] S.L. Lauritzen and D.J. Spiegelhalter. Local computations with probabilties on graphical structures and their application to expert systems. *J. Royal Statistical society B*, 50:154–227, 1988.

[6] J. Pearl. *Probabilistic Reasoning in Intelligent systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers, Inc., 1988.

[7] E. Castillo, J. M. Gutierrez, and A. S. Hadi. *Expert Systems and Probabilistic Network Models*. Springer, 1997.

[8] D.E. Heckerman, E.J. Horvitz, and B.N. Nathwani. Towards normative expert systems: part I, the Pathfinder project. *Methods of Information in medicine*, 31:90–105, 1992.

[9] D.E. Heckerman and B.N. Nathwani. Towards normative expert systems: part II, probability-based representations for efficient knowledge acquisition and inference. *Methods of Information in medicine*, 31:106–116, 1992.

[10] L.K. Saul, T. Jaakkola, and M.I. Jordan. Mean field theory for sigmoid belief networks. *Journal of Artificial Intelligence Research*, 4:61–76, 1996.

[11] T.S. Jaakkola and M.I. Jordan. Variational methods and the QMR-DT database. MIT Computational Cognitive Science Technical Report 9701, Massachusetts Institute of Technology, 1997.

[12] D. Barber and W. Wiegerinck. Tractable variational structures for approximating graphical models. In M.S. Kearns, S.A. Solla, and D.A. Cohn, editors, *Advances in Neural Information Processing Systems*, volume 11. MIT Press, 1999.

[13] W. Wiegerinck, W. ter Burg, E. ter Braak, P. van Dam, Y. O, Neijt J., and H Kappen. Inference and advisory system for medical diagnosis. Technical report, SNN, 1997.

[14] W.J. ter Burg *et al.* A diagnostic advice system based on pathophysiological models of diseases. In *Medical Informatics Europe*, Ljubljana, Slovenia, 1999. Accepted.

[15] T. Jaakkola. *Variational Methods for Inference and Estimation in Graphical Models*. PhD thesis, Massachusetts Institute of Technology, 1997.

[16] W. Wiegerinck and D. Barber. Mean field theory based on belief networks for approximate inference. In L. Niklasson, M. Bodén, and T. Ziemke, editors, *ICANN'98: Proceedings of the 8th International Conference on Artificial Neural Networks, Skövde, Sweden, 2-4 September 1998*, volume 2, pages 499–504, London, 1998. Springer.

[17] W. Wiegerinck and D. Barber. Variational belief networks for approximate inference. In H. La Poutre and J. van den Herik, editors, *Proceedings of the Tenth Netherlands/Belgium Conference on Artificial Intelligence (NAIC'98)*, pages 177–83, Amsterdam, 1998. CWI.

[18] J. Whittaker. *Graphical models in applied multivariate statistics*. Wiley, Chichester, 1990.