

Decision support for medical diagnosis

Bert Kappen and Wim Wiegerinck, Nijmegen University
Edith ter Braak, University Medical Centre Utrecht,

December 7, 2000

1 Introduction

“Do we need computerised Diagnostic Decision Support Systems in medical practice?”

Problems in modern medicine are often very complex, and evidence for the best choice to be made is often lacking. Decisions made by physicians are arbitrary and highly variable (within one physician and between physicians) and lacking explanation or “rationalisation” [1, 2]. Clinical examples of this phenomenon in diagnosis making are abundant.

The body of potentially useful knowledge that is relevant to even a relatively narrow diagnostic area may be too large to make the optimal (diagnostic) decision on the spot. Ironically, modern information technology (especially through the Internet) further increases the amount of available knowledge, potentially even further complicating this situation. Moreover, individual patients need “individualised” decisions, because their characteristics differ from the “average” and because of their individual wishes [3]. Apparently, individualising the general results of research may be cumbersome and time consuming, while on the other hand, modern medical practice demands for efficiency, cost-effectiveness and high technical quality.

The derivation of diagnostic protocols is a main problem in health care. In some environments diagnostic support as proposed in this proposal is not likely to influence physician’s decisions, e.g. on a neurological intensive care unit, since the diagnosis is often obvious [4]. In contrast, general internal medicine covers an enormous range of, sometimes relatively rare, diagnostic categories. Hence the tendency of medicine to be divided in superspecialties. A diagnostic decision support system covering general internal medicine may be appreciated by both generalists and super-specialists alike: by the generalist because this field of work typically covers a very broad range of diagnoses, by the super-specialist because he/she may not feel completely at ease outside his/her specific field of expertise.

It is readily understandable that the above comprises an enormous task and challenge for modern medicine in general and individual doctors in particular,

illustrating the need for decision support techniques. Obviously, computerised decision aids may be very promising from a theoretical point of view. However, the currently available systems have not yet been very successful and certainly their use is still not widespread and not established in daily routine.

A variety of reasons may be responsible for this:

- **Lack of accuracy:**
Those current systems that intend to cover a broad diagnostic domain of medicine [5, 6] generally lack diagnostic accuracy. This is mainly due to the levels of detail [7] (e.g. diagnostic categories at the level of ICD-10 [8]) and completeness in the knowledge base. In contrast, systems that are based on detailed modelling of knowledge, resulting in good performance, are restricted to a relatively narrow field [9, 10].
- **Lack of transparency:**
In the era of evidence based medicine the advice of “a machine”, functioning as a black box is unacceptable: an advice must be accounted for on the basis of research published in the peer reviewed literature. The majority of conventional protocols and consensus guidelines also often fail to refer explicitly to the literature. Therefore, (diagnostic) advices suggested by a computerised tool should come with the appropriate references from the literature.
- **Users attitude:**
In a subset of (potential) users there may be a misunderstanding about what computers can and cannot do for them. Generally, decision support systems need intelligent and responsible users, who are able to interpret the advice given and estimate its merit [11]. This, however is not exclusively a matter of users attitude. Producers of decision support tools should take this issue into account as well, especially when designing the user interface and deciding which facilities are needed.
- **Lack of integration of information:**
Patient oriented decision support needs data from several sources. A decision support system will generate new information (e.g. a diagnostic advice) through inference, using patient-specific information. Integration of information, multiple usability of patient data, integration of databases and knowledge bases are common problems when using a heterogeneous Hospital Information System (HIS). In practice, the completeness of patient information, and the accuracy and level of detail of diagnoses stored in the HIS is often very poor [12].
- **Lack of a controlled terminology:**
This is a problem that even might not be solved completely in the near future. Most standard classification systems are at a general level [8, 13], thus lacking the required detail, or specialised [14] and therefore too limited to meet the needs for a broad decision support system. Furthermore,

there is not always a standard classification available, for instance for specific terminology used in text books.

- Careful introduction:
Introduction of a decision support system should be done as careful and thorough as is done for drugs that are new on the market. Oddly enough this tradition of careful introduction (and marketing !) is common in the field of therapeutics, but not quite as established for support tools in general and for diagnostics in particular. After introduction, the decision support system will need constant monitoring of users needs and maintenance to keep up with the latest results of medical research.
- The need of an integrated clinical workstation:
The appropriate infrastructure and workstations are not yet available in all hospitals. Physicians will need on-line support during the implementation of the various functionalities of a reliable clinical workstation, that integrates all the required information.

In conclusion, modern medicine is in need of computerised decision aids both to meet its own high standards and to keep pace with the stage of development in other domains such as manufacturing or the services industry. Although decision support appears to be exceptionally suitable for the medical domain, computer aided decision making in medicine is still in its infancy. The development, implementation, assessment and further improvement of decision support systems in medicine still need a lot of research.

2 Our approach

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 [5, 6]. The main reasons are that the modeling of the relations between diseases and findings is at a very coarse level. Therefore, the diagnoses suggested by these systems are too superficial for clinical use. Secondly, the diagnostic process requires reasoning from causes to effects (diseases \rightarrow finding) and vice 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. [15, 16, 17]) 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 [9, 10]. 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 [18, 19, 20]. 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 [12]. 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.

3 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 [21].

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.

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. Such simplified conditional probability tables require only the specification of order k parameters, where k is the number of parents.

Even though the conditional probability tables are modeled in a compact way, inference is still intractable.

4 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 : $Q(S) = \prod_{\gamma} Q(S_{\gamma}|\pi_{\gamma})$, [22, 23, 20]. The next step is to optimize the parameters of Q such that the Kullback-Leibler (KL) divergence between Q and P_E ,

$$D(Q, P_E) = \sum_{\{S\}} Q(S) \log \frac{Q(S)}{P_E(S)} \quad (1)$$

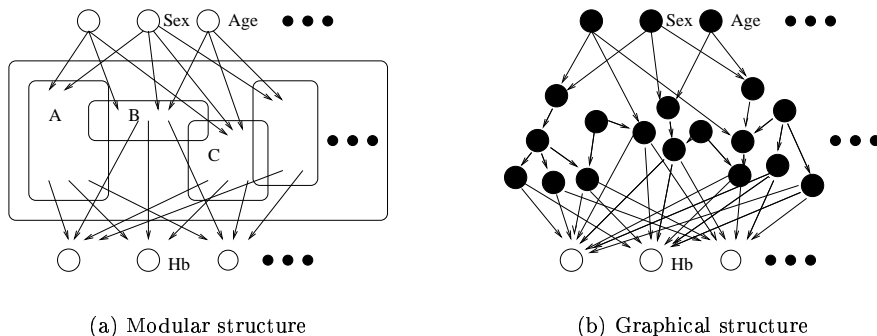


Figure 1: Modular and graphical network structure. Left: modular structure of the network. A, B, C ... represent (overlapping) sub-domains. Each sub-domain is modeled by a number of nodes (cf. right figure) representing variables that are relevant in that domain. The upper nodes, e.g. 'sex' and 'age'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

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

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

(see [24]).

$D(Q, P_E)$ 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 a coupled set of non-linear equations that can be solved numerically.

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. $Q(S) = \prod_i Q(S_i)$. The other extreme is to factorize Q according to a triangulated graph [15, 17] of P . In this case, one obtains the exact 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.

In figure 2 we plotted the maximal error $\max_d |Q(s_d) - P(s_d|s_f)|$ as a function of the network size for an artificially generated network. 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. For more information on the variational approach see [25].

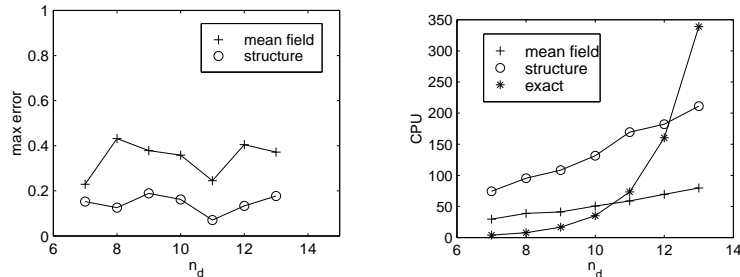


Figure 2: Left: The maximal error as a function of the network size. Right: CPU-time in Matlab seconds for exact and approximate inference as a function of the network size

5 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.

To develop Promedas, we use our internally developed software environment, called BayesBuilder. 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 (fig. 3). 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

¹A demonstration version of Promedas is available on CD-ROM. See www.mbfys.kun.nl/snn/Research/promedas

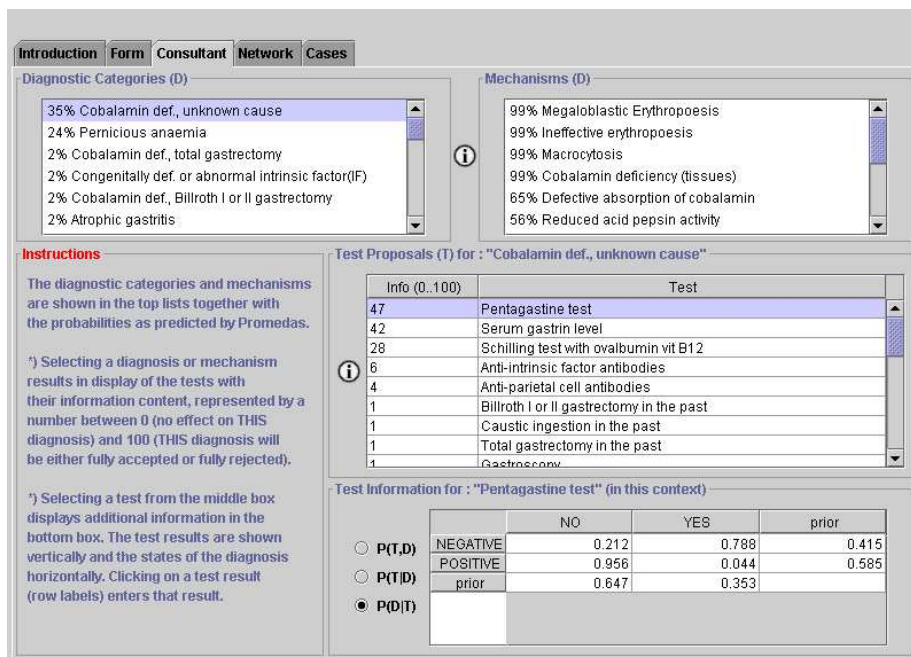


Figure 3: The Promedas diagnostic decision support system.

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. In addition, Promedas provides help information, medical background information and pointers to the literature.

6 Utilisation

Taking into account the need for decision support system in general and diagnostic decision support in particular, we strongly believe that a diagnostic decision support system is viable and, eventually, marketable. However, even a pilot regarding the implementation and assessment a diagnostic decision support system that covers only a relatively narrow diagnostic field (i.e. anaemia) will need careful, preferably step-wise, introduction to its target users, followed by continuing support and mutual feed-back. It is expected that this will result in growing acceptance and enthusiasm by its target users and finally leading to

wide-spread use.

The physicians that participate within this project work in daily hospital practice as well. They have good contacts in the medical community in many fields of medicine (general internal medicine, oncology, endocrinology, haematology) within Utrecht University Hospital, affiliated (regional) hospitals and various professional circles. In addition, the user group includes specialists in internal medicine from other academic hospitals. Therefore, we feel that we are able to “market” and to follow up a diagnostic decision support system at least for research purposes (assessment) in The Netherlands. Wide spread acceptance of computer aided diagnostic decision support tools will probably need some “trend setters”, who are most likely to be found in the academic circle.

7 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 500-1000 variables in order to be able to address this issue properly.

Further reading on graphical models can be found in [16, 26]. Further reading on diagnostic decision support systems can be found in [4, 2].

The Promedas project is supported by the Technology Foundation STW.

References

- [1] D. M. Eddy. The challenge. *Journal of American Medical Association*, 263:287–290, 1990.
- [2] M. Berg. *Rationalizing Medical Work, Decision-support Techniques and Medical Practices*. MIT press, 1997.
- [3] R.J. Lilford, S.G. Pauker, D.A. Braunholz, and J. Chard. Decision analysis and the implementation of research findings. *BMJ*, 317:405–409, 1998.
- [4] 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.
- [5] E.S. Berner, G.D. Webster, A.A. Shugerman, J.R. Jackson, J. Algina, A.L. Baker, E.V. Ball, C.G. Cobbs, V.W. Dennis, E.P. Frenkel, L.D. Hudson, E.L. Mancall,

- C.E. Racle, and O.D. Taunton. Performance of four computer-based diagnostic systems. *N-Engl-J-Med.*, 330(25):1792–6, 1994.
- [6] 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.
- [7] M.A Shwe, B. Middleton, D.E. Heckerman, M. Henrion, Horvitz E.J., H.P. Lehman, and G.F. Cooper. Probabilistic Diagnosis Using a Reformulation of the Internist-1/ QMR Knowledge Base. *Methods of Information in Medicine*, 30:241–55, 1991.
- [8] WHO. *International Classification of Diseases, 10th revision, Clinical Modification ICD-10*, 1992.
- [9] 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.
- [10] 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.
- [11] J.J. Dijkstra. *On the use of computerised decision aids*. PhD thesis, State University of Groningen, 1998.
- [12] 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.
- [13] J. Boersma, R.S. Gebel, and H. Lamberts. *International Classification of Primary Care, Short titles (vertaald in het Nederlands) en Nederlandse subtitels*. Nederlands Huisartsen Genootschap, 1993.
- [14] WHO. *International Classification of Diseases for Oncology*.
- [15] S.L. Lauritzen and D.J. Spiegelhalter. Local computations with probabilities on graphical structures and their application to expert systems. *J. Royal Statistical society B*, 50:154–227, 1988.
- [16] J. Pearl. *Probabilistic Reasoning in Intelligent systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers, Inc., 1988.
- [17] E. Castillo, J. M. Gutierrez, and A. S. Hadi. *Expert Systems and Probabilistic Network Models*. Springer, 1997.
- [18] 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.
- [19] 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.
- [20] 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 11*. MIT Press, 1999.
- [21] W.J. ter Burg *et al.* A diagnostic advice system based on pathophysiological models of diseases. In *Medical Informatics Europe*, Ljubljana, Slovenia, 1999. Accepted.

- [22] W. Wiegierinck 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*. Springer, 1998.
- [23] W. Wiegierinck 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)*, Amsterdam, 1998. CWI.
- [24] J. Whittaker. *Graphical models in applied multivariate statistics*. Wiley, Chichester, 1990.
- [25] H.J. Kappen and W. Wiegierinck. Mean field theory for graphical models. In D. Saad and M. Opper, editors, *Advanced mean field theory*. MIT Press, 2001.
- [26] F.V. Jensen. *An introduction to Bayesian networks*. UCL Press, 1996.