ELSEVIER

Artificial Intelligence in Medicine 30 (2004) 201–214

Artificial Intelligence in Medicine

www.intl.elsevierhealth.com/journals/aiim

Editorial

Bayesian networks in biomedicine and health-care☆

1. Introduction

Physiological mechanisms in human biology, the progress of disease in individual patients, hospital work-flow management: these are just a few of the many complicated processes studied by researchers in biomedicine and health-care. For controlling the ever increasing complexity of these fields, a proper understanding of their processes is important as is the ability to reason about them. The characteristics of the processes vary widely; however, typically only part of all the factors by which they are governed can be observed in practice. The processes, moreover, include the effects of individual as well as random variation. Essentially they are uncertain; the uncertainties involved render an overall understanding hard to achieve and reasoning a daunting task. Models capturing these processes and methods for using these models are thus called for to support decision-making in real-life practice.

Bayesian networks with their associated methods are especially suited for capturing and reasoning with uncertainty [33]. They have been around in biomedicine and health-care for more than a decade now and have become increasingly popular for handling the uncertain knowledge involved in establishing diagnoses of disease, in selecting optimal treatment alternatives, and predicting treatment outcome in various different areas. Bayesian networks are also increasingly developed in areas of health-care that are not directly related to the management of disease in individual patients. Examples include the use of Bayesian networks in clinical epidemiology for the construction of disease models and within bioinformatics for the interpretation of microarray gene expression data.

This special issue aims to convey an impression of the current state-of-the-art of the use of Bayesian networks in biomedicine and health-care. By devoting attention to new application areas, it complements what is known about the use of Bayesian networks in building decision-support systems for individual patient care. In this editorial, the various contributions are introduced. In addition, the scientific context of the contributions is sketched, to indicate their role and place in the broad field of Bayesian networks. In Section 2, the formalism of Bayesian networks is introduced and methods for their construction are reviewed. Section 3 introduces the biomedical problems involving uncertainty for which Bayesian networks are typically employed. The editorial concludes in Section 4 by introducing the five contributions to the issue.

☆ This special issue is a follow-up to the *Bayesian Models in Medicine* workshop, which was held on 1 July 2001 in Cascais, Portugal, during the AIME 2001 conference.

0933-3657/$ – see front matter
doi:10.1016/j.artmed.2003.11.001

2. Bayesian networks

In this section, the formalism of Bayesian networks and the basic methods for their development are reviewed. For a more thorough treatment of the topic, the reader is referred to Refs. [8,33].

2.1. The formalism

A Bayesian network, or probabilistic network, $\mathscr{B} = (\Pr, G)$ is a model of a joint, or multivariate, probability distribution over a set of random variables; it consists of a graphical structure G and an associated distribution Pr. The graphical structure takes the form of a directed acyclic graph, or DAG, $G = (V(G), A(G))$ with nodes $V(G) = \{V_1, \ldots, V_n\}$, $n \geq 1$, and arcs $A(G) \subseteq V(G) \times V(G)$. Each node V_i in G represents a random variable that takes one of a finite set of values. The arcs in the digraph model the probabilistic influences between the variables. Informally speaking, an arc $V_i \rightarrow V_j$ between two nodes V_i and V_j indicates that there is an influence between the associated variables V_i and V_j; absence of an arc between V_i and V_j means that the corresponding variables do not influence each other directly. More formally, a variable V_i is taken to be dependent of its parents and children in the digraph, but is conditionally independent of any of its non-descendants given its parents; this property is commonly known as the Markov condition [8,19].

Associated with the graphical structure of a Bayesian network is a joint probability distribution Pr that is represented in a factorised form. For each variable V_i in the digraph is specified a set of conditional probability distributions $\Pr(V_i|\pi(V_i))$; each of these distributions describes the joint effect of a specific combination of values for the parents $\pi(V_i)$ of V_i, on the probability distribution over the values of V_i. These sets of conditional probability distributions with each other define a unique joint probability distribution that factorises over the digraph's topology through

$$\Pr(V_1, \ldots, V_n) = \prod_{i=1}^{n} \Pr(V_i|\pi(V_i))$$

Fig. 1 shows an example Bayesian network; the notations v_i and $\neg v_i$ are used to indicate $V_i = \text{true}$ and $V_i = \text{false}$, respectively. The digraph of the network models cancer to be independent of heart disease given a value for their common parent smoking. The

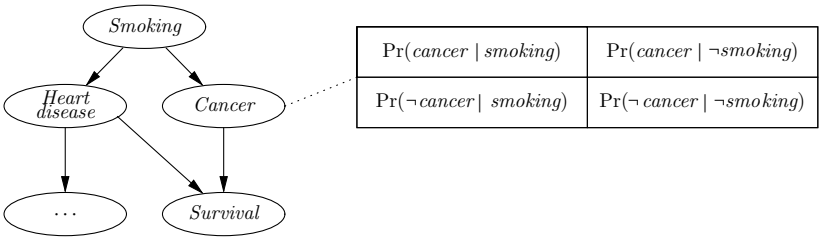

Fig. 1. An example Bayesian network.

conditional probability distributions associated with the variable cancer in the figure further demonstrate that the Markov condition provides for a localised representation of the joint probability distribution. The condition in fact serves to significantly reduce the amount of probabilistic information that has to be explicitly specified to uniquely describe the joint distribution. The condition also allows for the design of efficient algorithms for computing any probability of interest over a network's variables [28,33].

The digraph of a Bayesian network, apart from being acyclic, can have an arbitrarily complex topology to capture the intricacies of its application domain. For classification problems, however, a specific class of networks of limited topology have become popular [5,11,12]. In these networks, a distinction is made between a single class variable C and one or more feature variables; the latter variables serve to describe the characteristics of the instances to be classified. The class variable does not have any incoming arcs, but has arcs pointing to every feature variable. Between the feature variables, arcs are allowed under strict topological constraints. In a naive Bayesian network, for example, no arcs are allowed between the feature variables. In a tree-augmented Bayesian network (TAN), on the other hand, arcs are allowed between the feature variables as long as these constitute a tree. In a forest-augmented network (FAN), to conclude, the arcs should constitute a forest of trees [30]. The general structures of a naive Bayesian network and of a TAN network are shown in Fig. 2.

Although the variables in a Bayesian network are often assumed to be discrete, taking a value from a finite set of values, a network may also include continuous variables that adopt a value from a range of real values [27]. Generally Gaussian, or normal, distributions are assumed for the conditional probability distributions for such continuous variables. These distributions then are specified in terms of a limited number of parameters, such as their means and variance. Most Bayesian network tools nowadays allow for a mixture of discrete and continuous variables to be included in a network under some topological constraints.

2.2. Manual construction

Many of the Bayesian networks developed to date for real-life applications in biomedicine and health-care have been constructed by hand [2,3,17,21,22,31,32]. Manual construction of a network involves various development stages. For each of these stages, knowledge is acquired from experts in the domain of application, the relevant medical literature is studied, and available patient data are analysed. The following development stages are generally distinguished:

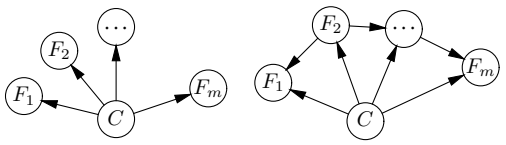

Fig. 2. (a) A naive Bayesian network and (b) a tree-augmented Bayesian network; the nodes F_j indicate the feature variables and C is the class variable.

(1) *Selection of relevant variables*: As a Bayesian network in essence is a graphical model of a joint probability distribution over a set of random variables, the first stage in its construction is the identification of the important variables to be captured, along with the values they may adopt. The selection of the relevant variables is generally based on interviews with experts, descriptions of the domain, and an extensive analysis of the purpose of the network under construction. Often, knowledge about the (patho)physiological processes concerned is used to guide the identification of the relevant variables [24,29].

(2) *Identification of the relationships among the variables*: Once the variables to be included in the network have been decided upon, the dependence and independence relationships between them have to be analysed and expressed in a graphical structure. For this purpose, generally the notion of causality is employed as a guiding principle: typical questions asked during the interviews with the domain experts are ''What could cause this effect?'' and ''What manifestations could this cause have?'' The elicited relationships are then expressed in graphical terms by taking the direction of causality for directing the arcs between the variables. The notion of causality often appears to match the experts' way of thinking about the (patho)physiological processes in their domain [14].

(3) *Identification of qualitative probabilistic and logical constraints*: Knowledge of qualitative probabilistic constraints and of logical constraints among the variables involved can help in the assessment and verification of the probabilities required for the network under construction. Qualitative probabilistic constraints are derived, for example, from properties of stochastic dominance of distributions. These constraints can be expressed as qualitative signs that can be used to study the reasoning behaviour of the projected network prior to its quantification [36]. Logical constraints are derived from functional relationships between the variables and can be used to significantly reduce the number of probabilities that have to be assessed for the network.

(4) *Assessment of probabilities*: In the next development stage, the local conditional probability distributions $\Pr(V_i|\pi(V_i))$ for each variable V_i are filled in. The required probabilities can be obtained from domain experts. Although the elicitation of judgmental probabilities is generally considered a daunting task, elicitation methods are available that are tailored to obtaining the large number of probabilities required in reasonable time [16,17,35]. Alternatively, the probabilities can be obtained from data. For a network with discrete variables, the conditional probability distributions are often computed as the weighted average of a probability estimate based on the available data and a prior Dirichlet distribution, that is, a multinomial distribution whose parameters can be interpreted as counts on a data set:

$$\Pr(V_i|\pi(V_i), D) = \frac{n}{n+n_0}\widehat{\Pr}_D(V_i|\pi(V_i)) + \frac{n_0}{n+n_0}\Theta(V_i|\pi(V_i))$$

where $\widehat{\Pr}_D$ is the probability distribution estimated from a given data set D, and Θ is the Dirichlet prior over the possible values of V_i; Θ is often taken to be uniform. The parameter n is the size of the data set D and n_0 is equal to an imaginary or real number of past cases on which the contribution of Θ is based. The resulting probability distribution Pr is again a Dirichlet distribution.

(5) *Sensitivity analysis and evaluation*: With the previous development stage, a fully specified Bayesian network is obtained. Before the network can be used in real-life practice, its quality and clinical value have to be established. One of the techniques for assessing a network's quality is to perform a sensitivity analysis with patient data. Such an analysis serves to provide insight in the robustness of the output of the network to possible inaccuracies in the underlying probability distribution [7,15]. Evaluation of a Bayesian network can be done in various different ways. Examples include measuring classification performance on a given set of real patient data and measuring similarity of structure or probability distribution to a gold-standard network or other probabilistic model.

As developing a Bayesian network is a creative process, the various stages are iterated in a cyclic fashion where each stage may, on each iteration, induce further refinement of the network under construction. An ontology may be developed to support the process [23].

2.3. Learning

In many fields of biomedicine and health-care, data have been collected and maintained, sometimes over numerous years. Such a data collection usually contains highly valuable information about the relationships between the variables discerned, be it implicitly. If a comprehensive data set is available, a Bayesian network can be learnt from the data, that is, it can be developed without explicit access to knowledge of human experts.

To be suitable for learning purposes, a data set has to satisfy various properties. First of all, the data comprised in the data set must have been collected very carefully. Biases that are introduced in the data set as a result of the data collection strategies used will have impact on the resulting Bayesian network, yet may not be desirable for the purpose for which the network is being developed. Also, the variables and associated values that occur in the data set should match the variables and values that are to be modelled in the network, or should at least admit easy translation. Moreover, the data set should comprise enough data to allow for reliable identification of probabilistic relationships among the variables discerned. In addition to these general prerequisites, a data set should satisfy several properties that are implicitly assumed by most learning algorithms. One of these is the assumption that each case in the data set specifies a value for every variable discerned, that is, there are no missing values. Unfortunately, for most real-life data sets this property does not hold. To use a data set with missing values for learning purposes, the missing values have to be filled in, or imputated, for example, based upon (roughly) estimated probabilities for these values or with the help of domain experts. Most learning algorithms further assume that the cases in the data set have been generated independently, that is, the values specified for the variables in a case are assumed not to be influenced in any way by the values in previously generated cases. Also, it is assumed that the process of data generation is not time-dependent.

Learning a Bayesian network from data involves the tasks of structure learning, that is, identifying the graphical structure of the network, and parameter learning, that is, estimating the conditional probability distributions to be associated with the network's

digraph. In many learning algorithms, the two tasks are performed simultaneously and, as a consequence, are not easily distinguished.

One of the early algorithms for learning a Bayesian network from data is the K2 algorithm [6]. Given a data set D, this algorithm searches, in a greedy heuristic way, for an acyclic digraph that, supplemented with maximum likelihood estimates for its probabilities, best explains the data at hand. More formally, it searches for a digraph G^* that maximises the joint probability $\Pr(G, D)$ over all possible digraphs G. Given a topological ordering on the random variables concerned, the algorithm constructs, for every subsequent variable V_i, an optimal set of parents. To this end, it starts by assuming the parental set to be empty and then adds, iteratively, the parent whose addition most increases the probability of the resulting structure and the data set; it stops adding variables to a parental set as soon as the addition of a single parent cannot increase the probability $\Pr(G, D)$. The K2 algorithm is an example of a search and scoring method. These methods search the space of all possible acyclic digraphs by generating various different graphs in a heuristic way and comparing these to their ability to explain the data at hand. Other search and scoring methods build, for example, upon the use of the minimum description length (MDL) principle [25] use a genetic algorithm for the search involved [26].

Another approach to learning a Bayesian network from data is to build upon the use of a dependence analysis [4]. A Bayesian network in essence models a collection of conditional dependence and independence statements, through its Markov condition. By studying the available data set, the dependences and independences between the various variables can be extracted, for example, by means of statistical tests, and subsequently captured in a graphical structure. The information-theoretical algorithm of Cheng et al. is an example of an algorithm taking this approach [4]. The algorithm has three subsequent phases termed drafting, thickening and thinning. In the drafting phase, the algorithm establishes, from the data, the mutual information for each pair of variables and constructs a draft digraph from this information. In the thickening phase, the algorithm adds arcs between pairs of nodes if the corresponding variables are not conditionally independent given a certain conditioning set of variables. In the thinning phase, to conclude, each arc of the graph obtained so far is examined using conditional independence tests, and is removed if the two variables connected by the arc prove to be conditionally independent.

Based upon the observation that independence tests quickly become unreliable for larger conditioning sets and the search space of all possible digraphs is infeasibly large, learning algorithms have been proposed that take a hybrid approach [10,38]. These algorithms are composed of two phases. In the first phase, a graph is constructed from the data, generally using lower-order dependence tests only. This graph is subsequently used to explicitly restrict the search space of graphical structures for the second phase in which a search algorithm is employed to find a digraph that best explains the data.

To conclude, there is also a great deal of interest in estimating probability distributions from data using maximum likelihood estimation [20]. The expectation maximisation (EM) algorithm is a two-step algorithm used by many researchers for this purpose [9]. It consists of a step of computing the expected value of the relevant parameter and a maximisation step, which are carried out in an interleaved fashion until convergence. In contrast with the learning algorithms reviewed above, the EM algorithm is able to deal with missing values.

2.4. Manual construction versus learning

Manual construction of a Bayesian network requires access to knowledge of human experts and, in practice, turns out to be quite time consuming. With the increasing availability of clinical and biological data, learning evidently is the more feasible alternative for developing a Bayesian network. Learning, as a consequence, is attracting considerable interest, both from developers and within the research community. Whether or not building a Bayesian network by hand would result in a network of higher quality when compared to learning it from data, is yet an open question. One would expect that, in many areas of biomedicine, human knowledge of the underlying (patho)physiological processes is more robust than the knowledge embedded in a data set of limited size. To date there is little evidence, however, to corroborate this expectation. It is an equally open question whether learning a Bayesian network of more complex topology pays off when compared to learning a simple Bayesian classifier. One would expect that the more faithful the digraph of a Bayesian network is in reflecting the dependences and independences embedded in the data, the better its performance. Research by Domingos and Pazzani has shown, however, that, when used for classification problems, naive Bayesian networks tend to outperform more sophisticated networks [11]. This finding has led to the suggestion that more complex network structures do not pay off. Friedman et al. [12], and Cheng and Greiner [5], on the other hand, have shown that tree-augmented networks, which in comparison to naive Bayesian networks, incorporate extra dependences among their feature variables, often outperform these naive Bayesian networks. Allowing for even more complex relationships between the feature variables, as in a forest-augmented network, moreover, has been shown to yield still better performance [30].

3. Problem solving in biomedicine and health-care

Bayesian networks are increasingly used in biomedicine and health-care to support different types of problem solving, four of which are briefly reviewed here.

3.1. Diagnostic reasoning

Establishing a diagnosis for an individual patient in essence amounts to constructing a hypothesis about the disease the patient is suffering from, based upon a set of indirect observations from diagnostic tests. Diagnostic tests, however, generally do not serve to unambiguously reveal the condition of a patient: the tests typically have true-positive rates and true-negative rates unequal to 100%. To avoid misdiagnosis, the uncertainty in the test results obtained for a patient should be taken into consideration upon constructing a diagnostic hypothesis. Bayesian networks offer a natural basis for this type of reasoning with uncertainty. A significant number of network-based systems for medical diagnosis have in fact been developed in the past and are currently being developed. Well-known early examples are the Pathfinder [21,22] and MUNIN [3] systems.

Formally, a diagnosis may be defined as a value assignment $\mathscr{D}^*$ to a subset of the random variables concerned, such that

$$\mathscr{D}^* = \underset{\mathscr{D}}{\operatorname{argmax}} \Pr(\mathscr{D}|\mathscr{E})$$

where $\mathscr{E}$ is the observed evidence, composed of symptoms, signs and test results. A diagnosis thus is a maximum a posteriori assignment (MPA) to a given subset of variables. Establishing a maximum a posteriori assignment from a Bayesian network, however, is extremely hard from a computational point of view. Since in addition combinations of disease do not occur very often, diagnostic reasoning is generally focused on single diseases. One approach is to assume that all diseases are mutually exclusive. The different possible diseases then are taken as the values of a single disease variable. Another approach is to capture each possible disease by a separate variable. Reasoning then amounts to computing the probability distribution for each such variable separately. The combination of the most likely values for these separate disease variables, however, need not be a maximum a posteriori assignment to these variables.

To assist physicians in the complex task of diagnostic reasoning, a Bayesian network is often equipped with a test-selection method that serves to indicate which tests had best been ordered to decrease the uncertainty about the disease present in a specific patient [1]. A test-selection method typically employs an information-theoretic measure for assessing diagnostic uncertainty. Such a measure is defined on a probability distribution over a disease variable and expresses the expected amount of information required to establish the value of this variable with certainty. An example measure often used for this purpose is the Shannon entropy. The measure can be extended to include information about the costs involved in performing a specific test and about the side effects it can have. Since it is computationally hard to look beyond the immediate next diagnostic test, test selection is generally carried out non-myopically, that is, in a sequential manner. The method then suggests a test to be performed and awaits the user's input; after taking the test's result into account, the method suggests a subsequent test, and so on.

3.2. Prognostic reasoning

Prognostic reasoning in biomedicine and health-care amounts to making a prediction about what will happen in the future. As knowledge of the future is inherently uncertain, in prognostic reasoning uncertainty is even more predominant than in diagnostic reasoning. Another prominent feature of prognostic reasoning when compared to diagnostic reasoning is the exploitation of knowledge about the evolution of processes over time. Even if temporal knowledge is not represented explicitly, prognostic Bayesian networks still have a clear general temporal structure, which is depicted schematically in Fig. 3. The outcome predicted for a specific patient is generally influenced by the particular sequence of treatment actions to be performed, which in turn may depend on the information that is available about the patient before the treatment is started. The outcome is often also influenced by progress of the underlying disease itself.

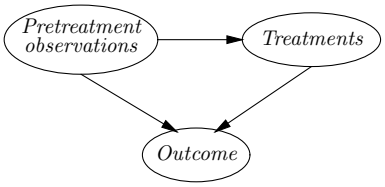

Fig. 3. General structure of a prognostic Bayesian network; each box denotes a part of the network.

Formally, a prognosis may be defined as a probability distribution

$$\Pr(\text{outcome}|\mathscr{E}, \mathscr{T})$$

where $\mathscr{E}$ again is the available patient data, including symptoms, signs and test results, and $\mathscr{T}$ denotes a selected sequence of treatment actions. The outcome of interest may be expressed by a single variable, e.g. modelling life expectancy. The outcome of interest, however, may be more complex, modelling not just length of life but also various aspects pertaining to quality of life. A subset of variables may then be used to express the outcome.

Prognostic Bayesian networks are a rather new development in medicine. Only recently have researchers started to develop such networks, for example, in the areas of oncology [18,31] and infectious disease [2,32]. There is little experience as yet with integrating ideas from, for example, traditional survival analysis into Bayesian networks. Given the importance of prognostication in health-care, it is to be expected, however, that more prognostic networks will be developed in the near future.

3.3. Treatment selection

The formalism of Bayesian networks provides only for capturing a set of random variables and a joint probability distribution over them. A Bayesian network therefore allows only for probabilistic reasoning, as in establishing a diagnosis for a specific patient and in making a prediction of the effects of treatment. For making decisions, as in deciding upon the most appropriate treatment alternative for a specific patient, the network formalism does not provide. Reasoning about treatment alternatives, however, involves reasoning about the effects to be expected from the different alternatives. It thus involves diagnostic reasoning and, even more prominently, prognostic reasoning. To provide for selecting an optimal treatment, a Bayesian network and its associated reasoning algorithms are therefore often embedded in a decision-support system that offers the necessary constructs from decision theory to select an optimal treatment given the predictions [2,31]. Alternatively, the Bayesian network formalism can be extended to include knowledge about decisions and preferences. An example of such an extended formalism is the influence diagram formalism [37]. Like a Bayesian network, an influence diagram includes an acyclic directed graph. In this graph, the set of nodes is partitioned into a set of probabilistic nodes modelling random variables, a set of decision nodes modelling the various different treatment alternatives, and a value node modelling the preferences involved. Influence diagrams for treatment selection once again have a clear general structure, which is depicted schematically in Fig. 4.

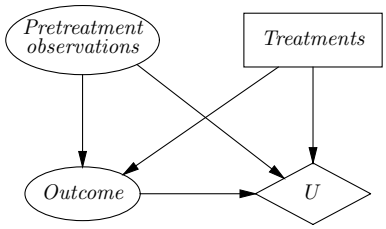

Fig. 4. General structure of an influence diagram, including a prognostic Bayesian network and a utility node U; each ellipse and box denotes a part of the diagram.

3.4. Discovering functional interactions

So far we have focused on the use of once constructed Bayesian networks for problem solving in biomedicine and health-care. However, the insight obtained by the construction process itself, in particular when done automatically by using one of the learning methods described above, may also be exploited to solve problems. As the topology of a Bayesian network can be interpreted as a representation of the uncertain interactions among variables, there is a growing interest in bioinformatics to use Bayesian network for the unravelling of molecular mechanisms at the cellular level. For example, finding interactions between genes based on experimentally obtained expression data in microarrays is currently a significant research topic [13]. Biological data are often collected over time; the analysis of the temporal patterns may reveal how the variables interact as a function of time. This is a typical task undertaken in molecular biology. Bayesian networks are now also being used for the analysis of such biological time series data [34].

4. Contents of the special issue

In the previous sections, we have sketched some of the developments in Bayesian networks research in biomedicine and health-care. We now introduce the papers that follow this editorial.

The paper by Silvia Acid and Luis de Campos, which is titled ‘‘A comparison of learning algorithms for Bayesian networks: a case study based on data of emergency medical services’’, is unusual as the area it focuses on is the management of health services instead of individual patient management. In the paper a number of structure-learning algorithms are explored and compared to one another using various different performance measures. The difficulties that must be overcome when using Bayesian networks in this domain are also described.

The next paper by Lise Getoor, Jeanne Rhee, Daphne Koller and Peter Small, which is titled ‘‘Understanding tuberculosis epidemiology using structured statistical models’’, addresses one of the limitations of the standard Bayesian network formalism as discussed in the previous sections. In the standard formalism, only fixed relationships

in a domain can be represented; general principles about similar objects, such as those expressed in object-oriented languages, cannot be represented explicitly. Statistical relational models are proposed as a means to increase the expressive power of Bayesian networks, and learning the structure and parameters of such models for the exploratory analysis of epidemiological data of patients with tuberculosis is investigated. The difference between learning statistical relational models and ordinary Bayesian networks is that in the former it is assumed that data are organised as a collection of tables (relations), so that learning takes place by inspecting tables in a relational data set that are explicitly linked to each other.

In the paper titled ''Using literature and data to learn Bayesian networks as clinical models of ovarian tumors'', Peter Antal, Geert Fannes, Dirk Timmerman, Yves Moreau and Bart De Moor explore the potential of the huge collection of information available on the World Wide Web as prior information for learning Bayesian networks. One of the problems that are often encountered upon learning Bayesian networks for clinical problems is that the available clinical data sets are too small to be exploited. As a consequence, it is usually necessary to extract information from various complementary sources. In this paper, techniques developed in the area of information retrieval are used as a basis for finding relationships among variables from the Web. The applicability of these techniques are studied with the construction of Bayesian networks for the classification of ovarian tumours in patients.

The discovery of latent, or hidden, variables in data for inclusion in Bayesian networks is the topic of the paper by Nevin Zhang, Thomas Nielsen and Finn Jensen, which is titled ''Latent variable discovery in classification models''. Much of the research in medicine is driven by the wish to extend what is known, and hence the question addressed in the paper, whether algorithms for latent variable discovery are able to find new phenomena, is a challenging one. The Bayesian networks studied in the paper are hierarchical naive Bayes models. These models have a central class variable as do naive Bayesian networks, but this class variable acts as the root of a tree in which latent variables are included as internal nodes and feature variables as leaves. Experimental evidence of the usefulness of this method is also provided in the paper.

The final paper included in this special issue is by Boaz Lerner, titled ''Bayesian fluorescence in situ hybridisation signal classification''. In this paper the usefulness of a variety of probability distribution estimation methods for naive Bayesian networks are discussed, and applied to the problem of classification of image features obtained by digital microscopy of in situ hybridisation. Previous research by the author has shown that a hierarchical neural network yields good performance in this task. The aim of the research presented in the paper was to see whether naive Bayesian networks were able to do a better job. Taking the naive Bayesian network as a base framework, three different estimation methods, single Gaussian estimation, kernel density estimation and a Gaussian mixture model, were developed and studied. It appears that none of these estimation methods is able to improve on the neural network, which is explained by the authors in terms of the restriction imposed by the assumption of conditional independence in the underlying naive Bayesian network. Note that this contradicts results obtained by naive Bayesian networks, TANs and FANs by other researchers, which may be attributed to the nature of the problem.

Acknowledgements

In addition to the editors, the following people were involved in both this workshop and the special issue, and in particular reviewed submitted papers: K.-P. Adlassnig, R. Bellazzi, C. Berzuini, G.F. Cooper, R.G. Cowell, F.J. Dìez, M.J. Druzdzel, P. Haddawy, D. Hand, I.S. Kohane, P. Larrañaga, A. Lawson, L. Leibovici, T.Y. Leong, S. Monti, L. Ohno-Machado, K.G. Olesen, M. Paul, M. Ramoni, A. Riva, P. Sebastiani, G. Tusch, J. Wyatt, and B. Zupan. We are thankful to all of them for their devotion to achieving success for both the workshop and this special issue.

References

[1] Andreassen S. Planning of therapy and tests in causal probabilistic networks. Artif Intell Med 1992;4: 227–41.
[2] Andreassen S, Riekehr C, Kristensen B, Schønheyder HC, Leibovici L. Using probabilistic and decision-theoretic methods in treatment and prognosis modeling. Artif Intell Med 1999;15:121–34.
[3] Andreassen S, Woldbye M, Falck B, Andersen SK. MUNIN—a causal probabilistic network for interpretation of electromyographic findings. In: McDermott J, editor. Proceedings of the 10th International Joint Conference on Artificial Intelligence. Los Altos, CA: Morgan Kaufmann, 1987. p. 366–72.
[4] Cheng J, Bell D, Liu W. Learning Bayesian networks from data: an efficient approach based on information theory. In: Proceeding of the Sixth ACM International Conference on Information and Knowledge Management, 1997. p. 325–31.
[5] Cheng J, Greiner R. Comparing Bayesian network classifiers. In: Proceedings of the UAI'99. San Francisco, CA: Morgan Kaufmann, 1999. p. 101–7.
[6] Cooper GF, Herskovitz E. A Bayesian method for the induction of probabilistic networks from data. Mach Learn 1992;9:309–47.
[7] Coupé VMH, van der Gaag LC. Sensitivity analysis: an aid for probability elicitation. Knowl Eng Rev 2000;15:215–32.
[8] Cowell RG, Dawid AP, Lauritzen SL, Spiegelhalter DJ. Probabilistic networks and expert systems. New York: Springer, 1999.
[9] Dempster A, Laird N, Rubin D. Maximisation likelihood from incomplete data via the EM algorithm. J R Stat Soc B 1977;39:1–38.
[10] Van Dijk S, Thierens D, van der Gaag LC. Building a GA from design principles for learning Bayesian networks. In: Cantú-Paz E, Foster JA, Deb K, et al., editors. Proceedings of the Genetic and Evolutionary Computation Conference. San Francisco, CA: Morgan Kaufmann, 2003. p. 886–97.
[11] Domingos P, Pazzani M. On the optimality of the simple Bayesian classifier under zero–one loss. Mach Learn 1997;29:103–30.
[12] Friedman NIR, Geiger D, Pazzani M. On the optimality of the simple Bayesian network classifier. Mach Learn 1997;29:131–63.
[13] Friedman NIR, Linial M, Nachman I, Pe'er D. Using Bayesian network to analyze expression data. J Comput Biol 2000;7:601–20.
[14] van der Gaag LC, Helsper EM. Experiences with modelling issues in building probabilistic networks. In: Gómez-Pérez A, Benjamins VR, editors. Proceedings of EKAW on Knowledge Engineering and Knowledge Management: Ontologies and the Semantic Web. Lecture Notes in Artificial Intelligence (LNAI) 2473. Berlin: Springer-Verlag, 2002. p. 21–6.
[15] van der Gaag LC, Renooij S. Analysing sensitivity data. In: Breese J, Koller D, editors. Proceedings of the 17th International Conference on Uncertainty in Artificial Intelligence. San Francisco, CA: Morgan Kaufmann, 2001. p. 530–7.
[16] van der Gaag LC, Renooij S, Witteman CLM, Aleman B, Taal BG. How to elicit many probabilities. In: Proceedings of the 15th International Conference on Uncertainty in Artificial Intelligence. San Francisco, CA: Morgan Kaufmann, 1999. p. 647–54.

[17] van der Gaag LC, Renooij S, Witteman CLM, Aleman BMP, Taal BG. Probabilities for a probabilistic network: a case study in oesophageal cancer. Artif Intell Med 2002;25:123–48.
[18] Galán SF, Aguado F, Dìez FJ, Mira J. NasoNet: joining Bayesian networks and time to model nasopharyngeal cancer spread. In: Proceedings of the Eighth International Conference on Artificial Intelligence in Medicine in Europe (AIME 2001), Lecture Notes in Artificial Intelligence (LNAI) 2101. Berlin: Springer-Verlag, 2001. p. 207–16.
[19] Glymour C, Cooper GF. Computation, causation & discovery. Menlo Park, CA: MIT Press, 1999.
[20] Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: data mining, inference, and prediction. New York: Springer, 2001.
[21] Heckerman DE, Horvitz EJ, Nathwani BN. Towards normative expert systems. I. The Pathfinder project. Meth Inform Med 1992;31:90–105.
[22] Heckerman DE, Nathwani BN. Towards normative expert systems. II. Probability-based representations for efficient knowledge acquisition and inference. Meth Inform Med 1992;31:106–16.
[23] Helsper EM, van der Gaag LC. Building Bayesian networks through ontologies. In: van Harmelen F, editor. Proceedings of ECAI2002. Amsterdam: IOI Press, 2002. p. 680–4.
[24] Korver M, Lucas PJF. Converting a rule-based expert system into a belief network. Med Inform 1993;18(3):219–41.
[25] Lam W, Bacchus F. Learning Bayesian belief networks: an approach based on the MDL principle. Comput Intell 1994;10:269–93.
[26] Larrañaga P, Poza M, Yurramendi Y, Murga R, Kuijpers C. Structure learning of Bayesian networks by genetic algorithms: a performance analysis of control parameters. IEEE Trans Pattern Anal Mach Intell 1996;18(9):912–26.
[27] Lauritzen SL. Propagation of probabilities, means and variances in mixed graphical models. J Am Stat Assoc 1992;87:1098–108.
[28] Lauritzen SL, Spiegelhalter DJ. Local computations with probabilities on graphical structures and their application to expert systems. J R Stat Soc B 1987;50:157–224.
[29] Lucas PJF. Knowledge acquisition for decision-theoretic expert systems. AISB Q 1996;94:23–33.
[30] Lucas PJF. Restricted Bayesian network structure learning. In: Gámez JA, Salmerón A, editors. Proceedings of the First European Workshop on Graphical Models (PGM'02). Cuenca, Spain, 2002. p. 117–26.
[31] Lucas PJF, Boot H, Taal BG. Computer-based decision-support in the management of primary gastric non-Hodgkin lymphoma. Meth Inform Med 1998;37:206–19.
[32] Lucas PJF, De Bruijn NC, Schurink K, Hoepelman IM. A probabilistic and decision-theoretic approach to the management of infectious disease at the ICU. Artif Intell Med 2000;19(3):251–79.
[33] Pearl J. Probabilistic reasoning in intelligent systems. San Mateo, CA: Morgan Kaufman, 1988.
[34] Ramoni M, Sebastiani P, Cohen P. Bayesian clustering by dynamics. Mach Learn 2002;47:91–121.
[35] Renooij S. Probability elicitation for belief networks: issues to consider. Knowl Eng Rev 2001;16(3): 255–69.
[36] Renooij S, van der Gaag LC. From qualitative to quantitative probabilistic networks. In: Darwiche A, Friedman N, editors. Proceedings of the 18th International Conference on Uncertainty in Artificial Intelligence. San Francisco, CA: Morgan Kaufmann, 2002. p. 422–9.
[37] Shachter RD. Evaluating influence diagrams. Oper Res 1986;34(6):871–82.
[38] Wong ML, Lee SY, Leung KS. A hybrid data mining approach to discover Bayesian networks using evolutionary programming. In: Langdon WB, et al., editors. Proceedings of the Genetic and Evolutionary Computation Conference. San Francisco, CA: Morgan Kaufmann, 2002. p. 214–22.

Peter J.F. Lucas*
Institute for Computing and Information Sciences, University of Nijmegen
Toernooiveld 1, ED-6525 Nijmegen
The Netherlands
Tel.: +31-24-365-2611/3456; fax: +31-24-365-3366
E-mail address: peterl@cs.kun.nl, lucas@cs.uu.nl (P.J.F. Lucas)

Linda C. van der Gaag
Institute of Information and Computing Sciences, Utrecht University
Utrecht
The Netherlands
E-mail address: @cs.uu.nl (L.C. van der Gaag)

Ameen Abu-Hanna
Department of Medical Informatics, Academic Medical Center (AMC)
University of Amsterdam, Amsterdam
The Netherlands
E-mail address: a.abu-hanna@amc.uva.nl (A. Abu-Hanna)