Bayesian Analysis, Pattern Analysis and Data Mining in Health Care*

Peter Lucas
Institute for Computing and Information Sciences
Radboud University Nijmegen
Toernooiveld 1, 6525 ED Nijmegen, The Netherlands
Tel: +31 24 365 2611/3456; Fax: +31 24 3653366
E-mail: peterl@cs.kun.nl

Abstract

Purpose of review: To discuss the current role of data mining and Bayesian methods in biomedicine and heath care, in particular critical care.
Recent findings: Bayesian networks and other probabilistic graphical models are beginning to emerge as methods for discovering patterns in biomedical data and also as a basis for the representation of the uncertainties underlying clinical decision making. At the same time, techniques from machine learning are being used to solve biomedical and health-care problems.
Summary: With the increasing availability of biomedical and health-care data with a wide range of characteristics there is an increasing need to use methods which allow modeling the uncertainties that come with the problem, are capable of dealing with missing data, allow integrating data from various sources, explicitly indicate statistical dependence and independence, and allow integrating biomedical and clinical background knowledge. These requirements have given rise to an influx of new methods into the field of data analysis in health care, in particular from the fields of machine learning and probabilistic graphical models.

1 Introduction

The process of data analysis in health care is becoming more and more complicated for a number of reasons: (1) new techniques, such as microarrays, have given rise to the generation of data with unusual characteristics, such regarding few patients described by a large number of variables; (2) the integrated analysis of data from different sources concerning the same health-care topic, including the wish to incorporate background knowledge in the analysis process, is becoming more and more relevant; (3) with the wide availability of sophisticated and cheap computing equipment the exploitation of models to support clinical decision making has become a practical option. As traditional statistical methods have been unable to meet all of these requirements, there has been an influx of new methods from a number of fields, in particular from machine learning. This broader field, which includes methods for data preprocessing and visualization, non-statistical methods from machine learning, but also new methods from probability theory and statistic, is called *data mining*.

*Published in: Current Opinion in Critical Care 2004, 10:399–403

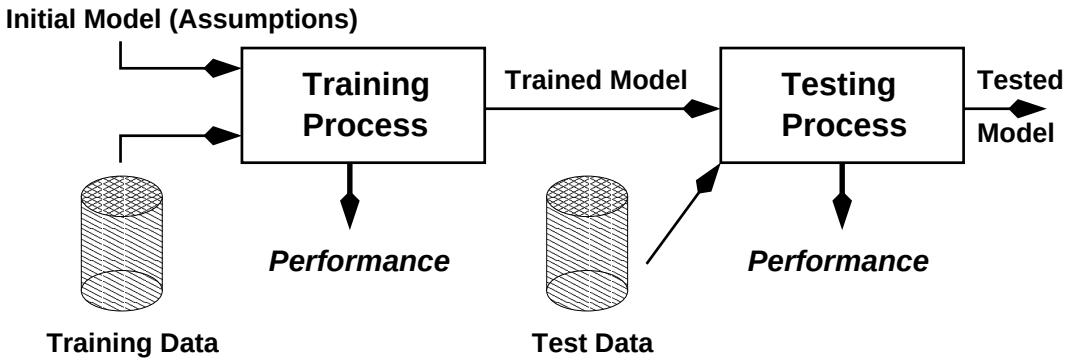


Figure 1: Model construction and evaluation in data mining.

Another gradually increasing trend in the health-care field is the use Bayesian statistical methods in data analysis, i.e. methods which allow taking into account prior knowledge when analyzing data. Data analysis is then viewed as the process of updating prior knowledge based on available biomedical and health-care evidence in the form of data. Clearly, the assumptions that are made in formulating this prior knowledge are crucial, and is a topic of much debate [3, 6].

Finally, statistical models which allow making part of the model assumptions about statistical dependence and independence explicit, called *probabilistic graphical models*, are gaining ground [17]. In particular, Bayesian networks, which are a type of graphical network model with directed links, offer a general and versatile approach to capturing and reasoning with uncertainty in medicine and health care [7, 17].

Below we will outline general trends in data mining in health care, which will be illustrated by recent results from health care and, when possible, from critical care in particular.

2 Finding and representing logical patterns

In this section, an overview is given of the model-construction process in data mining, and a number of recent modeling techniques are briefly reviewed.

2.1 Model construction

Unlike traditional statistical methods where the (mathematical) structure of models is given and probabilistic information is learnt from data, a characteristic of some machine-learning approaches is that the structure of models can be learnt automatically. An outline of the process of model construction in machine learning is shown in Figure 1; it is essentially identical to the process underlying statistical model development. In the case of limited availability of data, *cross validation*, where learning is done on subsequent parts of the data, and *bootstrapping*, where repeated random samples are taken from a dataset, are being used instead of the scheme shown in Figure 1. For both cross validation and bootstrapping, evaluation of learnt models is done on data not used in the training process. Both in traditional statistics and data-mining, there are always some initial assumptions about the model that have to be made; in Bayesian statistics these include assumptions about the prior probability distribution and the way in which prior information is updated based on observed evidence.

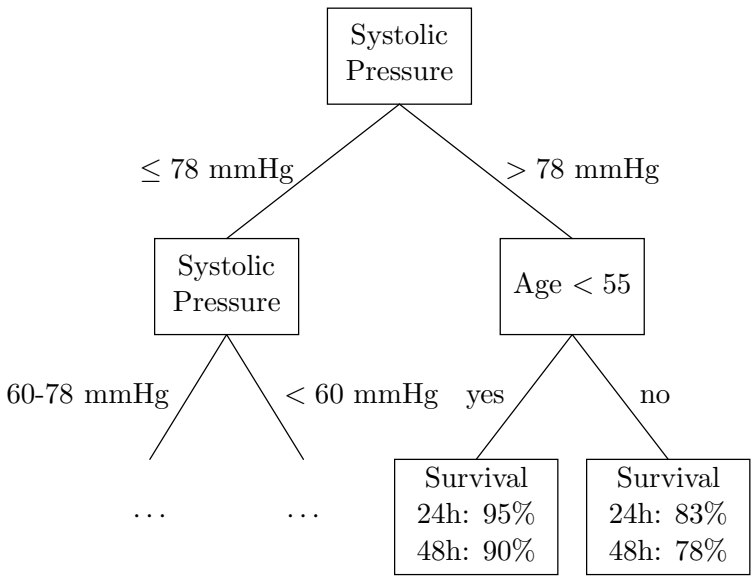


Figure 2: A fictitious machine-learning decision tree for predicting survival based on minimum systolic pressure and age.

A typical distinction made in the machine-learning and data-mining community is between supervised and unsupervised learning. In *supervised learning*, outcome or output variables are known in the available health-care data, and the aim of model construction is to predict these outcome variables as well as possible based on a subset of the other variables. In *unsupervised learning* such outcome variables are not available; a typical application of this in health care is discovering (diagnostic or prognostic) groups in data.

Many problems in machine learning are cast as *classification problems*, i.e. in terms of a limited set of possible outcomes, such as 'death' and 'alive' or a diagnostic category. There exists a great deal of experience in developing classifier models, based on a wide range of principles and techniques. The book by Hastie et al. is of particular interest to readers with a wish to understand the principles underlying data mining and statistical learning [14].

2.2 Classification trees and rules

A simple, yet powerful way to extract classification patterns from data is by means of a *decision tree*; this is a popular methods from machine learning that, however, is only used infrequently in health-care settings. To illustrate, the prognostic decision tree in Figure 2 predicts patient survival based on minimum systolic pressure and age in the intensive care. The rightmost path in the tree can also be represented as a *classification rule*:

> IF (Systolic Pressure $>$ 78 mmHg) **and** $\neg(age < 55)$ THEN
> (Survival_24h = 83%) **and** (Survival_48h = 78%)

which offer an alternative, and also easy to understand representation of clinical knowledge. It is also possible to learn classification rules from data without this intermediate step of decision-tree construction

2.3 Regression trees

Methods discussed so far were focused on producing output of a qualitative nature, such as whether a disorder is present or absent. In some problems it is necessary to produce numerical output in terms of input variables. This is traditionally done in statistics using regression. However, tree-based methods can be used for that purpose as well, giving rise to the same rectangular splitting up of the data space as in decision trees. *Regression trees* allow splitting up the input space in such way that the resulting output provides a suitable piecewise approximation of the problem. Regression trees are often mentioned together with classification trees because of the CART software, which combines the two approaches [4].

2.4 Neural networks and support vector machines

A *neural network* is a computational representation that takes as input a sequence of numbers, say encoded patient features, and outputs another sequence of numbers that is interpreted as, say, survival probability of that patient. Neural networks are very powerful at learning complicated, non-linear patterns in data, but have a tendency of adapting themselves too much to the data, a phenomenon known as *overfitting* [14]. This may give rise to the discovery of spurious patterns. They are an examples of back-box methods, i.e. the relationship between a neural network and the problem it represents is not easy to understand.

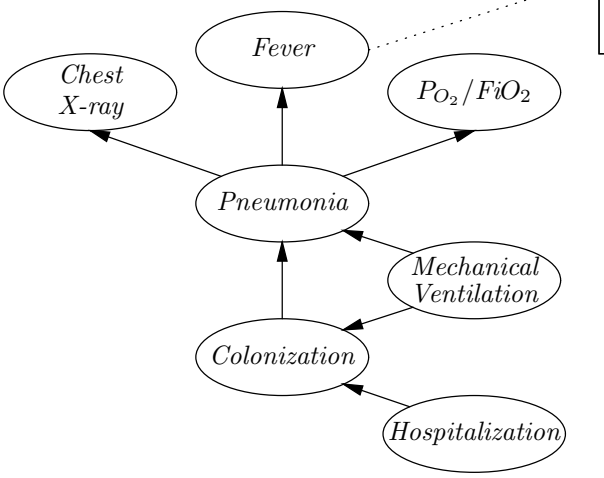
As neural networks, *support vector machines* are powerful pattern recognizers, capable of optimizing their content based on classification performance [14]. This is done by adjusting structure of the space of features. They have the same disadvantage as neural networks, viz. that they yield black-box models.

2.5 Bayesian networks

A *Bayesian network*, or probabilistic network, is a model of a joint, or multivariate, probability distribution over a set of random variables; it consists of a network structure and an associated probability distribution Pr. Bayesian networks can be seen as an alternative to logistic regression, where, in contrast to logistic regression, statistical dependence and independence are explicitly represented, and not hidden in approximating weights as in logistic regression. Figure 3 shows an example Bayesian network; the notations v and $\neg v$ are used as abbreviations for the variable V taking the value 'yes' and 'no', respectively. The network expresses the occurrence of *Pneumonia* to be dependent of *Colonization* by certain bacteria in the ICU and duration of *Mechanical Ventilation*. In turn, bacterial colonization of the tracheobronchial tree of the patient is assumed to be dependent of the duration of stay in the ICU (*Hospitalization*), as shown by the arrow from *Hospitalization* to *Colonization*. Pneumonia gives rise to signs and symptoms; mentioned in the network are abnormalities on the *Chest X-ray*, *Fever* and low P_{O_2}/FiO_2 ratio. These are all random variables. On these variable is defined a joint probability distribution in terms of 7 local probability tables, one for every variable in the network, of which is an example is shown to the right of the figure. The conditional probability distribution associated with the variable *Fever* in the figure,

$$\Pr(\mathit{Fever} \mid \mathit{Pneumonia})$$

expresses, for example, that *Fever* is conditionally independent of all the other variables, including *Colonization* and *Mechanical Ventilation*, given *Pneumonia*; in others words: as soon as we know an ICU patient has pneumonia it does for example no longer matter whether



$\Pr(fever \mid pneumonia) =$ e.g. 0.95	$\Pr(fever \mid \neg pneumonia) =$ e.g. 0.01
$\Pr(\neg fever \mid pneumonia) =$ e.g. 0.05	$\Pr(\neg fever \mid \neg pneumonia) =$ e.g. 0.99

Figure 3: An example Bayesian network. A Bayesian network consists of a network structure, shown to the left, which encodes dependence and independence information among the random variables shown, and a set of local probability distribution tables, of which one is shown to the right.

we know that a patient has been colonized by particular bacteria for our knowledge about the likelihood of fever. This is an assumption that probably makes sense clinically. However, the network also expresses that once we know a patient has pneumonia, learning that the P_{O_2}/FiO_2 is low does not tell us anything about the likelihood of the co-occurrence of fever, which is an simplifying assumption which may not be true. Thus, many simplifying assumptions can be read off from the network, and even if they are not true, they are at least explicitly visualized by means of the network structure. Ref. [18] discusses the development of a more extensive version of a Bayesian network for computer-based support of the diagnosis and treatment of ventilator-associated pneumonia in the ICU, which is of special interest to those involved in the construction of diagnostic and therapeutic models of ventilator-associated pneumonia.

For handling classification problems a specific class of networks of restricted structure have become popular [9, 11]. In these networks, a distinction is made between a single class variable C, e.g. 'disorder', and one or more feature variables, in clinical medicine usually signs and symptoms; the latter variables serve to describe the characteristics of patients to be classified. In a *naïve Bayesian network*, for example, no links are allowed between the feature variables. In a *tree-augmented Bayesian network*, or TAN for short, links are allowed between the feature variables. The general structure of a naïve Bayesian network and of a TAN network are shown in Figure 4. Structures in between these two extremes can also be usefully exploited [16].

Many of the Bayesian networks developed to date for real-life applications in biomedicine and health-care have been constructed by hand, i.e. they are based on medical background knowledge. 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. Integrating background knowledge and evidence derived from data is also supported by Bayesian networks, and missing data can be handled both in the construction process and in using a Bayesian network model.

Researchers with a particular interest in health-care applications of Bayesian networks are referred to Ref. [17].

3 Solving health-care problems

So far we have summarized general principles of data mining; in the next sections recent advances in using the methods mentioned above are described.

3.1 Support of clinical management

Classification trees and rules are normally used to obtain simple models that can be employed in the clinic without any need for computer assistance. Recent examples of such models include the classification of cancer using gene-expression data [23], the classification of tumors of the tongue [21], and the deferral of rectal examination in trauma patients [12].

Neural networks have repeatedly been shown to be suitable for the analysis of complicated patterns in health-care data. A well-known example is the PAPNET system that is used for the screening of cervical smears. However, despite the fact that it has been in use since the beginning of the 1990s, there is still controversy about its clinical value [13]. The observation that neural network technology can be usefully exploited for the construction of classifier systems in medicine is also supported by the recent successful construction of a neural network that is able to classify hematological malignancies based on hematological laboratory data [24]. As mentioned above, the black-box nature of the resulting models is seen as a significant drawback; it limits the clinical usefulness of neural network technology, and it may be one of the reasons behind the ongoing controversy about PAPNET.

Already in the 1970s several systems were developed according to principles closely related to that used in naïve Bayesian networks with varying amounts of success. Well-known early work is that by De Dombal et al. which dealt with the diagnosis of causes of acute abdominal pain (e.g. Ref. [8]). Nowadays, due to the significant progress that has been made in the theory of probabilistic graphical models, the early limitations no longer exist. Furthermore, computers are now widely available and used. As a consequence, systems that are capable of assisting clinicians in the management of patients are now being developed in a more principled way, and, given the high level of computerization of hospitals, slowly emerging in the clinic. Ironically, the last few years it has become apparent that the early naïve approaches,

(a) (b)

Figure 4: A naïve Bayesian network, (a), and a tree-augmented Bayesian network, (b); the nodes F_j indicate the feature variables and C is the class variable.

where conditional independence of signs, symptoms and test results given the disorders was assumed, thus giving rise to an inappropriate representation of the uncertainties involved in a problem, are often sufficient for the purpose of classification of patients, e.g. in diagnostic or prognostic categories. For example, the decision-support system for the classification of preinvasive cervical squamous lesions developed by Price et al. is a naïve Bayesian network, and therefore similar to the early approaches. However, it was developed according to modern principles and appeared to perform well [19]. Acid et al. have studied the usefulness of Bayesian networks in capturing the knowledge involved in the management of a medical emergency service [1]. This is one of the few example of work which goes beyond immediate clinical use of decision aids. Often it is difficult to obtain sufficient (clinical) data that can be used as a basis for the construction of machine-learning models; Antal et al. have shown that textual documents, in this case papers concerning ovarian tumors, can also be used as a basis for prior information in learning Bayesian networks [2].

3.2 Discovering regulatory processes

The insights 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 structure 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 unraveling of molecular mechanisms at the cellular level. For example, finding interactions between genes and protein-protein interactions based on experimentally obtained expression data in microarrays is currently a significant research topic. Segal et al. have shown that by learning Bayesian networks from data it is possible to obtain insight into the way genes are regulated [22]. The results suggested regulatory roles for previously uncharacterized proteins. Bulashevka et al. have used Bayesian-network learning to uncover the mechanisms underlying the progression in genetic changes in the development of urothelial cancers of the bladder [5]. Using probabilistic reasoning, taking into account the statistical dependences and independences encoded in the Bayesian network, the authors were able to propose hypotheses regarding the primary and secondary events in tumor pathogenesis.

The use of Bayesian networks in discovering protein-protein interaction has been studied by Jansen et al. [15], using a naïve Bayesian network, as shown in Figure 4(a), and a fully connected network, i.e. where each pair of variables in the network is connected by a link. The first structure was used to encode data where information about interactions between proteins were not available; the fully connected network captures all possible interactions among proteins.

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 [20]. An overview of recent successes in discovering cellular processes by means of Bayesian networks is given in Ref. [10], which is of particular interest to clinical researchers engaged in unraveling the molecular mechanisms underlying disease.

4 Conclusions

With the current rapid increase in the amount of biomedical data being collected electronically in critical care and the wide-spread availability of cheap and reliable computing equipment,

many researchers have already started, or are eager to start, exploring these data. In the end this will, without doubt, significantly enhance our insights into the various facets, both patient specific and patient aspecific, of the care process, and may result in new diagnostic, therapeutic and prognostic methods.

The complicated nature of real-world biomedical data has made it necessary to look beyond traditional biostatistics. New techniques were brought into the field from related areas, such as machine learning; the realization that modern data analysis draws up from a whole range of related disciplines has give rise to the establishment of the broader field of data mining. The results obtained by data mining, in particular from the subfield of machine learning, may not only be exploited to improve the quality of care by implementing particular changes to care policies but can also be used as a basis for the construction of computer-based decision-support systems. At the moment, it is unclear whether this will be done in the future on a more widely scale than has been done so far. The easy with which such systems can be embedded in clinical information systems, which are now starting to emerge in most critical care units, renders this at least an attractive possibility.

References

[1] Acid S, De Campos LM, Ferández-Luna JM, Rodríguise S, Rodríguise JM, Salcedo JL. A comparison of learning algorithms for Bayesian networks: a case study based on data from an emergency medical service. Artificial Intelligence in Medicine 30; 215–232, 2004.

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

[3] Beaumont MA, Rannala B. The Bayesian revolution in genetics. Nature Reviews Genetics 5; 251–261, 2004.

[4] Breiman L, Friedman J, Olshen R, Stone C. Classification and Regression Trees. Wadsworth, 1984.

[5] Bulashevka S, Skakacs O, Brors B, Eils R, Kovacs G. Pathways of urothelial cancer progression suggested by Bayesian network analysis of allelotyping data. Int J Cancer 110; 850–856, 2004.

[6] Couzin J. The new maths of clinical trials. Science 303; 784–786, 2004.

[7] * Cowell RG, Dawid AP, Lauritzen SL, Spiegelhalter DJ. Probabilistic Networks and Expert Systems. New York: Springer, 1999.
Technical book about the principles of Bayesian networks, illustrated by clinical applications.

[8] De Dombal FT, Leaper DJ, Staniland JR, McAnn AP, Horrocks JC. Computer-aided diagnosis of acute abdominal pain. British Medical Journal ii; 9–13, 1972.

[9] Domingos P, Pazzani M. On the optimality of the simple Bayesian classifier under zero–one loss. Machine Learning 29; 103–130, 1997.

[10] ** Friedman N. Inferring cellular networks using probabilistic graphical models. Science 3003; 799–805, 2004.
Survey paper about recent advances in uncovering cellular processes using Bayesian networks. Finding the appropriate structure of a Bayesian network is of the key questions. The relationship between structure and represented cellular mechanisms is reviewed in this paper.

[11] Friedman NIR, Geiger D, Pazzani M. On the optimality of the simple Bayesian network classifier. Machine Learning 29; 131–163, 1997.

[12] Guldner G, Babbitt J, Boulton M, O'Callaghan T, Feleke R, Hargrove J. Deferral of the rectal examination in blunt trauma patients: a clinical decision rule. Acad Emerg Med 11(6); 635–41, 2004.

[13] Hartmann KE, Nanda K, Hall S, Myers E. Technologic advances for evaluation of cervical cytology: is newer better? Obstet Gynecol 56(12): 765–74, 2001.

[14] ** Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. New York: Springer, 2001.
Excellent book about the mathematical principles of data mining and statistical learning. It discusses the bias-variance decomposition underlying model development, and various supervised and supervised learning methods. However, Bayesian networks are not discussed.

[15] Jansen R, Yu H, Greenbaum D, Kluger Y, Krogan NJ, Chung S, Emili A, Snyder M, Greenblatt JF, Gerstein M. A Bayesian network approach for predicting protein-protein interaction from genomic data. Science 302; 449–253, 2003.

[16] Lucas PJF. Restricted Bayesian network structure learning. In: Gâmez JA, Moral S, and Salmeron A (Eds.). Advances in Bayesian Networks, Studies in Fuzziness and Soft Computing, volume 146, Springer-Verlag, Berlin, 2004, pp. 217–232.

[17] ** Lucas PJF, Gaag van der LC, Abu-Hanna A. Bayesian networks in biomedicine and health-care. Artificial Intelligence in Medicine 30; 201–214, 2004.
Survey paper on recent advances in Bayesian networks in health care.

[18] Lucas PJF, De Bruijn NC, Schurink K, Hoepelman IM. A Probabilistic and decision-theoretic approach to the management of infectious disease at the ICU. *Artificial Intelligence in Medicine* 2000; **19**(3): 251–279.

[19] Price GJ, McCluggage WG, Morrison ML, McClean G, Venkatraman L, Diamond J, Bharucha H, Montironi R, Bartels PH, Thompson D, Hamilton PW. Computerised diagnostic decision support system for the classification of preinvasive cervical squamous lesions. Human Pathology 34(11); 1193–1203, 2003.

[20] Ramoni M, Sebastiani P, Cohen P. Bayesian clustering by dynamics. Machine Learning 2002; 47: 91–121.

[21] Schwarzer G, Nagata T, Mattern D, Schmelzeisen R, Schumacher M. Comparison of fuzzy inference, logistic regression, and classification trees (CART). Prediction of cervical lymph node metastasis in carcinoma of the tongue. Methods Inf Med 42(5): 572-7, 2003.

[22] Segal E, Shapira M, Regev A, Pe'er D, Botstein D, Koller D, Friedman N. Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. Nat Genet 34(2); 166–76, 2003.

[23] Tan AC, Gilbert D. Ensemble machine learning on gene expression data for cancer classification. Appl Bioinformatics 2(3 Suppl); S75–83, 2003.

[24] Zini G, d'Onofrio G. Neural network in hematopoietic malignancies. Clin Chim Acta 333(2); 195–201, 2003.