note that according to this analysis, inadequate diagnosis due to shortcomings of MUNIN is not on the list of reasons.

4 Conclusion

From the evaluation of the current prototype of MUNIN we have seen that the diagnostic capabilities are quite good in most cases. The evaluation also showed that some improvements must be expected if MUNIN could be allowed to use other clinical findings, apart from the neurophysiological findings. This also includes knowledge of whether the examination is performed to provide a diagnosis or to describe the extent of a known lesion.

The major limitation of MUNIN remains its limited anatomy, but we are confident that the system can be expanded to cover most relevant neuromuscular diseases encountered in an EMG lab. The major limitation wrt. clinical testing of MUNIN has been the small number of cases (30). We have also used the same cases for both learning and testing purposes. A more rigorous evaluation of MUNIN will have to await an expansion of the anatomy of MUNIN, which in turn will make it possible to test MUNIN on a much larger case database. The modifications of MUNIN presented here is one of the last steps before this can be achieved.

References

1. Andreassen S, Jensen FV, Andersen SK, Falck B, Kjærulff U, Woldbye M, Sørensen AR, Rosenfalck A and Jensen F 1989. MUNIN - an expert EMG assistant. In: Computer-aided electromyography and expert systems, ed. JE Desmedt. Elsevier, Amsterdam, pp. 255-277.
2. Andreassen S, Jensen FV and Olesen KG 1991. Medical expert systems based on causal probabilistic networks. *International Journal of Biomedical Computing* 28: 1-30.
3. Andreassen S, Falck B and Olesen KG 1992. Diagnostic function of the microhuman prototype of the expert system - MUNIN. *Electroencephalography and Clinical Neurophysiology* 85: 143-157.
4. Andreassen S, Rosenfalck A, Falck B, Olesen KG and Andersen SK 1996. Evaluation of the diagnostic performance of the expert EMG assistant MUNIN. *Electroencephalography and Clinical Neurophysiology* 101: 129-144.
5. Jensen FV 1996. An introduction to Bayesian networks. UCL Press.
6. Olesen KG, Kjærulff U, Jensen F, Jensen FV, Falck B, Andreassen S and Andersen SK 1989. A MUNIN network for the median nerve - a case study on loops. *Applied Artificial Intelligence* 3: 385-403.
7. Olesen KG and Andreassen S 1996. A heuristic method for diagnosing multiple diseases in complex medical domains modelled by causal probabilistic networks. In: Medical Informatics Europe '96, eds. J Brender et al. IOS Press, pp. 614-618.
8. Pearl J 1988. Probabilistic reasoning in intelligent systems. Morgan Kaufmann.
9. Suojanen M, Olesen KG and Andreassen S 1997. A method for diagnosing in large medical expert systems based on causal probabilistic networks. In: Lecture Notes in Artificial Intelligence 1211, Proceedings of the 6th Conference on Artificial Intelligence in Medicine Europe (AIME'97), eds. E Keravnou et al. Springer, pp. 285-295.
10. Suojanen M, Andreassen S and Olesen KG 1999. The EMG diagnosis - an interpretation based on partial information. *Medical Engineering & Physics* 21: 517-523.
11. Suojanen M, Andreassen S and Olesen KG 2001. A Method for Diagnosing Multiple Diseases in MUNIN. *IEEE Transactions on Biomedical Engineering*. In press.

01_60

Extended Bayesian Regression Models: A Symbiotic Application of Belief Networks and Multilayer Perceptrons for the Classification of Ovarian Tumors

Peter Antal[1], Geert Fannes[1], Bart De Moor[1], Joos Vandewalle[1], Y. Moreau[1], and Dirk Timmerman[2]

[1] Electrical Eng. Dept. ESAT/SISTA, Katholieke Universiteit Leuven, Kasteelpark Arenberg 10, B-3001 Heverlee (Leuven), Belgium
[2] Department of Obstetrics and Gynecology, University Hospitals Leuven, Herestraat 49, B-3000 Leuven, Belgium

Abstract. We describe a methodology based on a dual Belief Network-Multilayer Perceptron representation to build Bayesian classifiers. This methodology combines efficiently the prior domain knowledge and statistical data. We overview how this Bayesian methodology is able (1) to define constructively a valuable "informative" prior for black-box models, (2) to provide uncertainty information with predictions and (3) to handle missing values based on an auxiliary domain model. We assume that the prior domain model is formalized as a Belief Network (since this representation is a practical solution to acquiring prior domain knowledge) while we use black-box models (such as Multilayer Perceptrons) for learning to utilize the statistical data. In a medical task of predicting the malignancy of ovarian masses we demonstrate these two symbiotic applications of Belief Network models and summarize the practical advantages of the Bayesian approach.

1 Introduction

The Bayesian approach is becoming more attractive for the machine learning community because it can cope with the valuable subjective prior information in a principled way and it provides more detailed information for decision support. These properties are particularly attractive in medical applications, since detailed uncertainty information can be vital in a medical decision and frequently abundant prior domain knowledge is available beside the statistical data. The appearance of powerful stochastic algorithms further helped the wider application of the Bayesian approach, since more complex models became manageable in a Bayesian way. However, the Bayesian application of such efficient black-box models is hindered by the fact that the domain knowledge cannot be formulated directly in terms of such models (i.e., as a prior probability distribution over its parametrization). To solve this problem, we used Belief Networks to formalize the prior knowledge in a Bayesian way [3] and suggested an algorithm based on

S. Quaglini, P. Barahona, and S. Andreassen (Eds.): AIME 2001, LNAI 2101, pp. 177–187, 2001.

Belief Networks to derive *informative* priors for black-box regression models [2]. Since the constructed Belief Network is a complete probabilistic domain model, it can be attached to regression models as an auxiliary model to handle missing values frequently occurring in a medical context [1]. The derived *informative* prior and the probabilistic auxiliary domain model can be important elements for the successful application of black-box regression models (for a related overview on knowledge-based neurocomputing, see [5]. In the paper we summarize these methods retrospectively from a unified perspective and compare the performance of these extended black-box regression models with Belief Network models in a real-life medical problem.

The paper is organized as follows: Section 2 reviews the Bayesian approach in classification problems. Section 3 recapitulates the medical problem which will serve as a test case, introduces the data and defines relevant performance measures. In Section 4 we discuss the applied Bayesian models, particularly the derived *informative* prior distributions and the missing value management. In Section 5 we describe shortly the algorithms used to approximate the Bayesian performance measures. Section 6 presents the performance of the models and the comparison of the models in the Bayesian framework. In Section 7 we summarize the advantages of Bayesianism in this classification problem and the proposed *informative* priors.

2 Classification in the Bayesian Context

Starting with a prior distribution expressing the initial beliefs concerning the parameter values of the model, we can use the observations to transform this into the posterior distribution for the model parameters expressing the beliefs after observing the data. Using this posterior distribution over the model parameters, useful random variables can be defined for functions depending on the model parameters, like predictions and error measures.

In a binary classification task this rationale means the following. We are primarily interested in the correct classification of an observation $\boldsymbol{x} \in \mathbb{R}^l$. This can be achieved by constructing a *binary decision function* $g(\boldsymbol{x}, \omega) \in \{0,1\}$ where $\omega \in \Theta$ are the model parameters. A more informative predictive model provides not only a class label, but also the *class probabilities*, though it is a more complex task both from a statistical and computational point of view. As a further step in improving the decision support, *uncertainty information* can be provided for the class probabilities, for example the posterior distribution of class probabilities in the Bayesian framework. In this paper we follow the Bayesian approach to solve the classification problem for two main reasons: to incorporate prior background information in a general and principled way and to provide detailed information with clear semantics for decision support. Therefore we use a *probabilistic regression* model $P(T=1|\boldsymbol{x},\omega) = f(\boldsymbol{x},\omega) \in [0,1]$ with model parameters $\omega \in \Theta$ and a prior distribution $p_\Omega(.)$ over Θ. We assume a supervised learning scheme, that is the existence of a labeled training set $\underline{\boldsymbol{d}} = \{\boldsymbol{x}_k, t_k\}_{k=1}^n$, $(\boldsymbol{x}_k, t_k) \in \mathbb{R}^l \times \{0,1\}$, where $\boldsymbol{x}$ is a real valued l-dimensional input vector and t is the corresponding class label. In the paper we use capitals for random variables, bold indicates a vector and an underline indicates a matrix.

Using the observed data $\underline{\boldsymbol{d}}$ and applying Bayes' rule, the prior distribution can be transformed to the posterior distribution

$$p_\Omega(\omega|\underline{\boldsymbol{d}}) = \frac{p_T(t_1,\ldots,t_n|\omega,\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n)\,p_\Omega(\omega|\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n)}{p_{\underline{D}}(\underline{\boldsymbol{d}})} \propto L(\omega|\underline{\boldsymbol{d}})\,p_\Omega(\omega)$$

where $L(\omega|\underline{\boldsymbol{d}})$ denotes the probability of the data given the parameters.

Once we have this posterior distribution for the model parameters, we can define random variables related to predictions, performance, etc. In classification problems for example, we are interested, for a given $\boldsymbol{x}$, in the random variable $f(\boldsymbol{x},\boldsymbol{\Omega})$ where $\boldsymbol{\Omega}$ is a random parameter vector. In this way we have uncertainty information about the predicted class probability.

We can simplify this result to scalar values for the class probabilities $P(T=1|\boldsymbol{x},\underline{\boldsymbol{d}})$. The optimal step back depends on the cost function attached to the reported scalar value. Assuming the L_2 loss function, the optimal strategy is to report the expectation of the class probability in the posterior parameter probability space $f(\boldsymbol{x}) := E_{\Omega|\underline{d}}[f(\boldsymbol{x},\omega)]$. A further simplification is to discretize this scalar value using some user specified threshold λ, deriving a binary decision function

$$g_\lambda(\boldsymbol{x}) := \begin{cases} 1 \text{ if } E_{\Omega|\underline{d}}[f(\boldsymbol{x},\omega)] \geq \lambda \\ 0 \text{ else.} \end{cases}$$

These three distinct levels ($f(\boldsymbol{x},\boldsymbol{\Omega})$, $f(\boldsymbol{x})$ and $g_\lambda(\boldsymbol{x})$) provide diminishing possibilities for decision support. This is illustrated by the increasing information provided by the class labels, class probabilities and distributions of class probabilities, though the burdening statistical and computational complexity should be considered too.

3 Classification of Ovarian Masses

Ovarian malignancies represent the greatest challenge among gynaecologic cancers. A reliable preoperative prediction in terms of benign and malignant ovarian tumors would be of considerable help to clinicians selecting an appropriate treatment. There are two sources of information to construct such predictive models: prior knowledge and data.

The available relevant medical literature and expert knowledge is abundant and very diverse (for an overview, see [7]. In addition to the prior background information, data were collected prospectively from 300 consecutive patients who were referred to a single institution (University Hospitals Leuven, Belgium) from August 1994 until June 1997. The data collection protocol ensure that the patients had an apparent persistent extrauterine pelvic mass and excludes other causes that may have similar symptoms such as infection or pregnancy, so the primary aim is differentiation between benign and malignant masses (for a detailed description, see [7]. Univariate statistics of data set are presented in Table 1.

Table 1. Univariate statistics for the ovarian cancer data set.

	Age	Parity	CA 125	Color score	RI
$E[.\|Benign]$	47.77	1.50	110.34	1.98	0.12
$E[.\|Malignant]$	58.62	1.57	1222.299	3.20	0.41
$Std[.\|Benign]$	15.60	1.40	976.56	0.84	0.77
$Std[.\|Malignant]$	15.18	1.73	3779.64	0.95	0.46

Standard statistical studies indicate that a multi-modal approach – the combination of various types of variables – is necessary for the discrimination between benign and malignant tumors. Therefore Logistic Regression models, Multilayer Perceptrons and Belief Networks were previously applied [7, 3]. These models predicted the scalar class probabilities and they were developed and tested in the classical statistical framework.

A natural step to provide more detailed information for medical decision support is to apply the Bayesian approach to provide the distribution of class probabilities. We can use the classical statistical performance measures for the evaluation of the models in the Bayesian framework, since any performance measure is a function of the model parameters (for fixed observations/test data). These performance measures then become random variables which provide more information than a point estimate.

Because in the medical literature the Receiver Operator Characteristics (ROC) curve was advocated to assess and compare the performance of probabilistic classifiers, we use three performance measures: the misclassification rate $MR(\Omega, \underline{d})$, the ROC curve $ROC(., \Omega, \underline{d})$ and the area under the ROC curve $AUC(\Omega, \underline{d})$ (for the definition and interpretations of the ROC curve, see [8]. Finally, we computed the Bayes factors [9] to compare the models in a Bayesian way.

4 Applied Bayesian Models

In the paper a Multilayer Perceptron (MLP) and a Belief Network (BN) model are discussed. For each model we define and investigate two types of prior distribution: *informative* and *non-informative* priors, depending on the amount of incorporated prior background knowledge. Additionally, an auxiliary probabilistic domain model is introduced for the MLP models, which are not capable of coping with incomplete input samples. This auxiliary probabilistic domain model is not updated by the observed samples and is only used for coping with missing input values in the original data set[1].

4.1 Belief Network Models

A Belief Network model defines a joint probability distribution over the domain variables. It is a probabilistic *white-box* model, since it consists of a graph model about the conditional independencies of the domain variables and a quantitative part specifying the conditional probabilities for the domain variables. The prior knowledge available from experts and the literature can be directly represented in an *informative* prior distribution for this model using the following technique: Assuming that the parameter independence holds, we use the Dirichlet family to represent the prior distribution [4]. Using a Dirichlet distribution, an expert can express his belief in parametrizations, instead of giving only a point estimation for the parameters. The *non-informative* distribution for Belief Networks is a uniform distribution corresponding to the Dirichlet distribution with all hyperparameters equal to one.

We tried to build Belief Networks to formalize the available prior knowledge from expert and literature in three different ways [3]. In the first phase we experimented with "biological" models in which various causal models of the disease are incorporated. The specification of the structure was relatively easy, but the quantification was not possible from the literature, nor from the expert and we had a too small data set to quantify additionally introduced hidden variables. In the second phase we built "expert" models that reflect the expert's experience. The qualitative dependence-independency structure specification was again relatively easy. However the results were too biased because the medical expert participating in the project previously worked with the same collected data, so his estimates were largely based on the data set. In the third phase we built "heterogeneous" models containing biological models of the underlying mechanism quantifiable by the literature, parts quantified by a medical expert and parts quantified by previously published studies. The graphical structure of the Belief Network model is shown on the left side of Fig. 1.

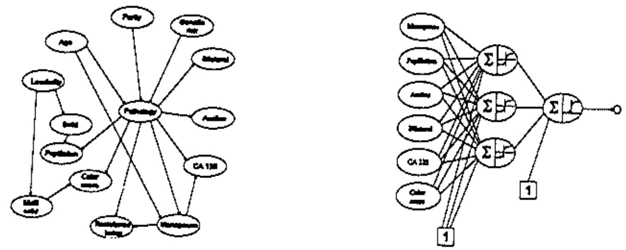


Fig. 1. The BN and MLP model structures.

[1] Note that the BN missing value mechanism is fundamentally different from this in two important respects: first, the mechanism is immediately provided since a BN model defines a joint probability distribution, second the prior distributions for Belief Network models (and therefore the underlying mechanism to cope with the missing values) is updated by the data.

It is worth to note from a computational point of view that all the variables are discrete in this model class. Because of the extensive and complex usage of the prior knowledge we used a strict documentation method to track the route of the prior information from studies into the model. Conversion formulas were constructed to compile the raw prior knowledge to be compatible with the conditions of the model and the format of the Bayesian network.

Since a Bayesian network model is inherently capable of performing inference for incomplete cases, there is no additional technique for handling missing values. Consequently, we evaluate two Belief Network models: the same structure with *informative* and *non-informative* prior distribution.

4.2 Multilayer Perceptron Models

These probabilistic regression models are defined over a continuous input space specifying an input-output mapping which can be interpreted as a conditional probability $P(T=1|\boldsymbol{x},\omega)=f(\boldsymbol{x},\omega)$.

The MLP model structure used in the paper is shown on the right of Fig. 1 (for data preprocessing details, input and MLP structure selection, see [7, 2].

Since the Multilayer Perceptron (MLP) model is a *black-box* model, it is not possible to specify directly a prior distribution incorporating the available prior knowledge in a general way. To solve this problem we use the Belief Network models as a tool to acquire and represent the prior domain knowledge. The *informative* prior distribution for the BN is then transformed into a so called *informative* prior distribution for the MLP parameters.

Omitting the technical details, it is possible to define a mapping $\mathcal{T}:\Theta^{BN}\rightarrow\Theta$ that transforms a prior distribution $p_{\Omega^{BN}}(.)$ over the Belief Network parameter space to a prior probability distribution $p_{\Omega}(.)$ over the black-box model parameter space[2]: a black-box regression model $f(\boldsymbol{x},\omega)$ is used for approximating the conditional distribution of the output class $P(T=1|\boldsymbol{x},\omega^{BN})$ conditioned on the input $\boldsymbol{x}$, which is defined by the Belief Network. Thus we can define a mapping from every Belief Network parametrization $\omega^{BN}\in\Theta^{BN}$ to the "best" approximating regression model parametrization $\omega\in\Theta$.

The main steps for the application of this technique in the case of Multilayer Perceptron are the following (see [2] for details):

1a. Generate Belief Network parametrizations $\{\omega_1^{BN},\ldots,\omega_l^{BN}\}$ from the Dirichlet distribution by standard methods.
1b. Generate block of prior samples $\{\underline{d}_1^p,\ldots,\underline{d}_l^p\}$ from each parametrization. It is advantageous to use the prior probabilistic domain model to compute $P(T=1|\boldsymbol{x},\omega^{BN})$ for each sample instead of Bernouilli generated class labels according to this probability.
2. Train a Multilayer Preceptron for each block of samples resulting in a block of MLP parametrizations $\{\omega_1,\ldots,\omega_l\}$ (we applied the scaled conjugate gradient algorithm [11].

[2] For simplicity, only the Belief Network notations are differentiated with a *BN* superscript and we use the intact notation for the black-box MLP models.

3. Estimate the transformed distribution $p_{\Omega}(.)$ from the generated MLP parametrization $\{\omega_1,\ldots,\omega_l\}$ with a mixture of Gaussians.

The *non-informative* prior for the MLP model is a non-restrictive, complexity based prior $\mathcal{N}(\boldsymbol{0},\sigma^2\boldsymbol{I})$ (cfr. weight decay for neural networks in the classical framework).

As mentioned previously, an auxiliary probabilistic domain model for the MLP models is introduced, which is represented by a Belief Network[3]. We use the prior Belief Network described earlier with its most probable a priori parametrization and assume a fixed data collection procedure with the *ignorable* condition [10]. In this case $P(\boldsymbol{X}^{miss}|\boldsymbol{X}^{obs},\ T,\ \boldsymbol{Z},\ \mathrm{BN})$ gives the distribution of the missing values (for details see [1]) where $\boldsymbol{Z}$ are those variables that occur in the data set and in the attached auxiliary domain model, but not in the MLP model. It means that the auxiliary Belief Network model provides a theoretically optimal solution to use incomplete samples for learning and inference. Using these assumptions it can be shown [1] that the posterior and inference for such an extended regression model can be written as follows with $\mathcal{I}$ a binary random vector, where $\mathcal{I}_i$ denotes the observed or missing status of the i^{th} variable.

$$p_{\Omega}(\omega|\boldsymbol{t},\underline{\boldsymbol{x}}^{obs},\underline{\boldsymbol{z}},\mathcal{I},\mathrm{BN})=\int\ldots\int p_{\Omega}(\omega|\boldsymbol{t},\underline{\boldsymbol{x}}^{obs},\underline{\boldsymbol{x}}^{miss})\prod_{i=1}^{n}dP(\boldsymbol{x}_i^{miss}|t_i,\boldsymbol{x}_i^{obs},z_i,\mathrm{BN})$$

$$P(T=1|\boldsymbol{x}^{obs},\ z,\ \mathcal{I},\ \omega)=\int f(\boldsymbol{x},\omega)dP(\boldsymbol{x}^{miss}|\boldsymbol{x}^{obs},\ z)$$

5 Inference Algorithms

The target random variables to be estimated are hierarchical: the inference $P(T=1|\boldsymbol{\Omega},\boldsymbol{x}^{obs},z,\underline{\boldsymbol{d}})$ and the performance related $AUC(\boldsymbol{\Omega},\underline{\boldsymbol{d}}),MR(\boldsymbol{\Omega},\underline{\boldsymbol{d}})$.

In the case of Belief Network prediction we sample the posterior distribution $p_{\Omega}(\omega|\underline{\boldsymbol{d}})$ by direct sampling from the updated Dirichlet (see [6]) and compute the conditional probabilities of malignancy for the drawn parametrizations by an exact inference algorithm using a join tree (see [4]). Based on these predictions the corresponding AUC and MR values can be computed.

In the case of MLP prediction, at first we sample the unknown input variables conditioned on all the known variables using the auxiliary probabilistic BN model (i.e., according to $P(\boldsymbol{X}^{miss}|\boldsymbol{X}^{obs},\ T,\ Z,\ \mathrm{BN})$). Then we sample from the posterior distribution based on the completed data set $p_{\Omega}(\omega|\boldsymbol{t},\underline{\boldsymbol{x}}^{obs},\underline{\boldsymbol{x}}^{miss})$ by the Hybrid Monte Carlo method [13, 12]. Finally we compute the predictions and the corresponding AUC and MR values for the drawn parametrization on the test set by averaging over the missing input variables sampling $P(\boldsymbol{X}^{miss}|\boldsymbol{X}^{obs},\ Z,\ \mathrm{BN})$.

[3] Note that the applied Belief Network is discrete valued, so we use additional probabilistic models for transformation between discrete and continuous values.

6 Results

We report results for the four models described in Section 4.1 using the algorithms summarized in Section 5. We partitioned the data set described in Section 3 randomly to a test (30%) and training (70%) set, the reported results are based on the test set. This was repeated 50 times to eliminate dependency on separation. The model classes are denoted as follows: Multilayer Perceptron model with *informative* prior (MLP-I) and with *non-informative* (MLP-N) and Belief Network model with *informative* prior (BN-I) and with *non-informative* prior (BN-N). The performance of the models are shown in Table 2. Note that most of the outliers for models with *non-informative* prior are omitted to focus on the most interesting region. Figure 2 shows the more detailed effect of the prior for varying sample size.

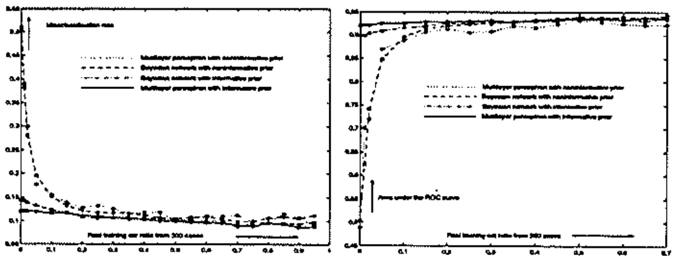


Fig. 2. Learning curves of MLP-N, MLP-I, BN-N, and BN-I models.

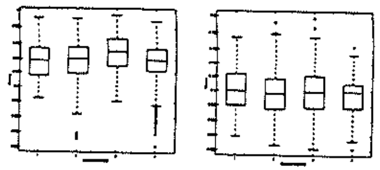
The Bayes-factors against the less probable model class (BN-N) are 1.29 (MLP-N), 1.45 (MLP-I) and 1.19 (BN-I) assuming uniform prior probabilities for the model classes[4].

7 Discussion

As can be seen from Table 2 the Belief Network models have the best performance, which is the consequence of the larger number of input variables (a more extensive comparison including Logistic Regression models can be found in [2]).Using the same input set as for the MLP models, the performance lags behind the performance of the MLP models, as Figure 2 shows (for details please see [2]).

[4] To eliminate the effect of the sample size we take the n^{th} root of the likelihoods where $n = 300$ is the size of the sample (i.e., we estimate $\int_{\Theta} \sqrt[n]{\frac{P(d|\omega, M=.)}{P(d|BN-N)}} dP_{\Omega}(\omega)$ by the Monte Carlo method).

Table 2. The distributions of the area under the ROC curve, of the misclassification rates and the corresponding means and variances.



E[.]/Var[.]	MR	ROC
MLP-N	0.1019/0.0007	0.9361/0.0006
MLP-I	0.0964/0.0010	0.9362/0.0010
BN-N	0.0980/0.0008	0.9485/0.0006
BN-I	0.0944/0.0006	0.9371/0.0006

Figure 2 illustrates clearly the subtle advantages of the Bayesian incorporation of prior domain knowledge through *informative* priors: it has a large advantageous effect in the small sample region ($[0 - 0.4]$) and it has no restrictive effect in the large sample range.

The Bayes factors indicate that the posterior probability of MLP model with informative priors is the highest, but the Bayesian model comparison needs further investigation, such as the selection of appropriate priors for the models.

Beside numerical comparison, we performed a manual evaluation of the predictions. The application of the Bayesian approach, particularly for the MLP models, shows that when the model misclassifies a sample, the distribution of the predicted class probability is widespread. Therefore the introduction of an "uncertain" label can have important consequences on performances and thus on medical decision support. A further interesting medical result that all the models are regularly "uncertain" for a well-determined subset of patients, which suggests that the currently used input features are not enough to achieve reliable classification for these patients.

8 Conclusions

In the paper the applicability of complex regression models, Multilayer Perceptron models, in the Bayesian context was examined. We summarized a technique that derives informative prior distributions for such black-box models and overviewed the application of Belief Networks as auxiliary probabilistic models to handle incomplete cases for probabilistic regression models. We performed a retrospective analysis to predict malignancy in ovarian masses in which various Bayesian models were investigated: Multilayer Perceptron and Belief Network models. We reported standard performance measures in the Bayesian framework, investigated the usability of the more detailed predictions of Bayesian models and compared the models in the Bayesian framework.

Currently we are investigating a hybrid Belief Network model composed of a prior Belief Network and a probabilistic regression model, where we treat the prior Belief Network (including the auxiliary model) and a probabilistic regression model together in the Bayesian framework.

Acknowledgements. Joos Vandewalle is Full Professor at the K.U.Leuven. Bart De Moor is Full Professor at the K.U.Leuven. Sabine Van Huffel is a senior research associate with the F.W.O. (Fund for Scientific Research-Flanders). Peter Antal is a Research Assistant with the K.U.Leuven. Geert Fannes is a Research Assistant with the F.W.O. Vlaanderen. This work was carried out at the ESAT laboratory and supported by grants and projects from the Flemish Government: Concerted Research Action GOA-MEFISTO-666 (Mathematical Engineering for Information and Communication Systems Technology) and F.W.O. project G.0262.97: Learning and Optimization: an Interdisciplinary Approach and the F.W.O. Research Communities: ICCoS (Identification and Control of Complex Systems) and Advanced Numerical Methods for Mathematical Modelling and Bilaterale Wetenschappelijke en Technologische Samenwerking Flanders-Hungary, BIL2000/19; from National Fund for Scientific Research (OTKA) under contract number T030586; from National Fund for Scientific Research (OTKA) under contract number F-030763; from the Belgian State, Prime Minister's Office-Federal Office for Sci., Tech. and Cult. Affairs-Interuniversity Poles of Attraction Programme (IUAP P4-02 (1997-2001): Modeling, Identification and Control of Complex Systems. The scientific responsibility is assumed by its authors; from the European Community TMR Programme, Networks, project CHRX-CT97-0160; from the Brite Euram Programme, Thematic Network BRRT-CT97-5040 'Niconet'; from the IDO/99/03 project (K.U.Leuven) "Predictive computer models for medical classification problems using patient data and expert knowledge", of the FWO grants G.0326.98 and G.0360.98, the FWO project G.0200.00.

References

1. P. Antal, G. Fannes, S. Van Huffel, B. De Moor, J. Vandewalle, and Dirk Timmerman, *Bayesian predictive models for ovarian cancer classification: evaluation of logistic regresseion, multi-layer perceptron and belief network models in the bayesian context,* Proceedings of the Tenth Belgian-Dutch Conference on Machine Learning, BENELEARN 2000, 2000, pp. 125–132.
2. P. Antal, G. Fannes, H. Verrelst, B. De Moor, and J. Vandewalle,*Incorporation of prior knowledge in black-box models: Comparison of transformation methods from bayesian network to multilayer perceptrons,* Workshop on Fusion of Domain Knowledge with Data for Decision Support, 16th UAI Conf., 2000, pp. 42–48.
3. P. Antal, H. Verrelst, D. Timmerman, Y. Moreau, S. Van Huffel, B. De Moor, and I. Vergote, *Bayesian networks in ovarian cancer diagnosis: Potential and limitations,* Prpc. of the 13th IEEE Symp. on Comp.-Based Med.Sys., 2000, Houston, pp. 103–109.
4. E. Castillo, J.M. Guttiérrez, and A.S. Hadi, *Expert systems and probabilistic network models,* Springer, 1997.
5. J. Cloete and J.M. Zurada, *Knowledge-based neurocomputing,* MIT Press, 2000.
6. D.J. Spiegelhalter et al., *Bayesian analysis in expert systems,* Statistical Science **8** (1993), no. 3, 219–283.
7. D. Timmerman et al., *Artificial neural network models for the pre-operative discrimination between malignant and benign adnexal masses,*Ultrasound Obstet. Gynecol. **13** (1999), 17–25.

8. J.A. Hanley et al., *The meaning and use of the area under receiver operating characteristic (roc) curve,* Radiology **143** (1982), 29–36.
9. J.M. Bernardo et al., *Bayesian theory,* Wiley & Sons, 1995.
10. A. Gelman, J.B. Carlin, H.S. Stern., and D.B. Rubin, *Bayesian data analysis,* Chapman & Hall, 1995.
11. M.F. Moller, *A scaled conjugate gradient algorithm for fast supervised learning,* Neural Networks **6** (1993), 525–533.
12. P. Müller and R.D. Insua, *Issues in bayesian analysis of neural network models,* Neural Computation **10** (1998), 571–592.
13. R.M. Neal, *Bayesian learning for neural networks,* Springer, 1996.