Available online at www.sciencedirect.com
SCIENCE DIRECT®
Pattern Recognition Letters 25 (2004) 1777–1785
Pattern Recognition Letters
www.elsevier.com/locate/patrec

Classification of fluorescence in situ hybridization images using belief networks

Roy Malka, Boaz Lerner *

Pattern Analysis and Machine Learning Lab, Department of Electrical and Computer Engineering, Ben-Gurion University, P.O. Box 653, Beer-Sheva 84105, Israel

Received 19 September 2003; received in revised form 3 June 2004
Available online 11 September 2004

Abstract

The structure and parameters of a belief network are learned in order to classify images enabling the detection of genetic abnormalities. We compare a structure learned from the data to another structure obtained utilizing expert knowledge and to the naive Bayesian classifier and study quantization in comparison to density estimation in parameter learning.

Keywords: Belief networks; Fluorescence in situ hybridization (FISH); Image classification; Naive Bayesian classifier; K2 algorithm

1. Introduction

Automatic fluorescence in situ hybridization (FISH) signal classification is essential for efficient clinical diagnosis of genetic abnormalities. FISH offers numerous advantages compared with conventional cytogenetic techniques since it allows numerical chromosome abnormalities to be detected during normal cell interphase. One of the most important applications of FISH for the detection of numerical aberrations such as Down syndrome is dot counting, i.e., the enumeration of signals (dots) within the nuclei, as the dots in the image represent the inspected DNA sequences. Manual dot counting is a time-consuming, laborious and tedious procedure hence the need in automation.

It has recently been proposed (Lerner et al., 2001) to base FISH dot counting on a classifier discriminating between valid signals and artifacts. The research indicated the accuracy of FISH signal classification for dot counting using a neural

* Corresponding author. Tel.: +972 8 6472 567; fax: +972 8 6472 949.
E-mail addresses: malkah@ee.bgu.ac.il (R. Malka), boaz@ee.bgu.ac.il (B. Lerner).

0167-8655/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2004.07.010

network (NN). Yet, in order to better understand the exact mechanisms underlying FISH signal classification and the relevance of each domain feature to the classification process and to each other, we are interested in this study in paradigms identifying and manifesting relationships among the features.

Probabilistic graphical models, and especially belief networks (BNs) also called Bayesian networks, exhibit dependencies and independencies inherent to the classification problem in edges and lack of edges between graph nodes representing the domain variables, respectively. Besides providing well-defined formalism for probabilistic inference, BNs extend model interpretability by representing causal relationships and dependencies among variables (features). They also allow the incorporation of prior (domain) knowledge into a data-driven model, thereby increasing its predictability especially in cases of scarce data. The theory of BNs is well established (Pearl, 1988; Heckerman, 1995; Jensen, 2001) and a several applications of BNs have already been suggested. For example, face images were discriminated from non-face images using a BN maximizing the Kullback–Leibler divergence between the corresponding joint distributions (Pham et al., 2002). In addition, a BN learned using the K2 algorithm (details in Section 3) fused decisions from a several machine learning classifiers also exploring their relative merit (Inza et al., 1999). In another study, a BN structure was learned successfully based on the naive Bayesian classifier using missing data to tackle a clustering problem (Pena et al., 1999).

Although originally proposed for knowledge representation and inference under uncertainty (Pearl, 1988), belief networks have lately been applied successfully also to classification (Langley et al., 1992; Friedman et al., 1997; Cheng and Greiner, 1999). We first identify a class node and observed nodes representing the features. The classifier is implemented by associating the computation of the conditional probability of a node given other nodes (inference in a BN) with the calculation of the posterior probability of the class node given the observed nodes. Then, according to Bayes' decision theory we assign a test pattern to the class corresponding to the maximum posterior probability, thereby minimizing the probability of misclassification.

A special case of a BN is the naive Bayesian classifier (NBC) that assumes variable independence conditioned on the class variable. When the dependency relationships among the variables are unknown (partially or completely), this simplification of the model is worthwhile leading to enhanced prediction accuracy using reduced burden of parameter estimation (Duda et al., 2001). Previous study have demonstrated the accuracy of the NBC in FISH signal classification (Lerner, 2004), however also its inferiority compared to the NN classifier (Lerner and Lawrence, 2001). In order to alleviate the independence restriction which may undermine performance, we allow in this study the BN to capture dependencies in the domain. Network structures expressing the true inter-connections between variables have higher chances to represent correctly the data yielding more accurate classification performance. We address here both learning network structure and parameters (the conditional probabilities quantifying the inter-connections), and compare structures learned from the data to others obtained utilizing expert knowledge and to the NBC. The obtained BN classifiers are then evaluated in discriminating valid and artifact FISH signals of two genetic syndromes in order to assess their benefit in automating FISH signal classification and dot counting.

Section 2 of the paper describes the analysis of FISH images and signals. Section 3 presents the BN model and discusses strategies of learning a BN structure based on expert knowledge or the data. The section also addresses BN parameter learning using variable quantization as well as inference. Section 4 demonstrates the capabilities of the strategies in FISH signal classification. Finally, Section 5 discusses the results of the study, while Section 6 concludes the research and presents topics of future research.

2. FISH image analysis and signal representation

FISH data preparation and acquisition have been described thoroughly and an example FISH

image has been given in (Lerner et al., 2001) and thus avoided here. A total of 400 images were collected from five slides demonstrating valid ('real') and artifact signals of red and green fluorophores (dyes) representing trisomy 21 and 13, respectively. Following segmentation, the system identified around 900 nuclei within these images and around 3000 signals within the nuclei. Based on labels provided by expert inspection, the system recognized a signal as 'real' or 'artifact' of the two fluorophores, thus accomplishing a four-class classification problem.

Following multi-spectral image analysis, twelve features were measured for the signals. These included area, eccentricity and a number of spectral features. We computed at the specific color plane three RGB (red–green–blue) intensity-based measurements: the total and average channel intensities and the channel intensity standard deviation. We also computed four hue-saturation-intensity (HSI) measurements: maximum hue, average hue, hue standard deviation and delta hue. Delta hue is the difference between the maximum and average hue normalized by the average hue. Two additional features of the set are the two coordinates of the eigenvector corresponding to the largest eigenvalue of the red and green intensity components of the signal. The last feature is the average gray intensity, i.e., average intensity over the three color channels. These twelve features have been described and motivated thoroughly as well as used successfully in previous study (Lerner et al., 2001), and they are listed and numbered in Table 1 to facilitate their identification in the rest of the paper. No feature selection has been applied to them.

3. Learning and inference in a BN

A belief network model $\mathscr{B}$ for a set of variables $\boldsymbol{U} = \{X_1, X_2, \ldots, X_n\}$ having each a finite set of mutually exclusive states consists of two main components, $\mathscr{B} = (\mathscr{G}, \boldsymbol{\theta})$. The first component $\mathscr{G} = (\boldsymbol{V}, \boldsymbol{E})$ is a model structure that is a directed acyclic graph (DAG) since it contains no directed cycles. $\boldsymbol{V}$ is a finite set of nodes of $\mathscr{G}$ corresponding to the variables of $\boldsymbol{U}$, and $\boldsymbol{E}$ is a finite set of directed edges of $\mathscr{G}$ where the edge $X_i \rightarrow X_j$ connects a parent node X_i to a child node X_j. $\boldsymbol{E}$ represents the dependence relations modeled by $\mathscr{G}$, where the presence or absence of an edge in $\boldsymbol{E}$ demonstrates dependence or independence, respectively, of two variables of $\boldsymbol{U}$ corresponding to the nodes being connected or not by that edge.

The second component of the BN is a set of parameters, $\boldsymbol{\theta}$, that specify all of the conditional probability distributions (or densities) that quantify graph edges. The probability distribution of each $X_i \in \boldsymbol{U}$ conditioned on its parents in the graph $\boldsymbol{Pa}_i \subset \boldsymbol{U}$ is $P(X_i = x_i | \boldsymbol{Pa}_i) \in \boldsymbol{\theta}$ when we use X_i and $\boldsymbol{Pa}_i$ to denote the ith variable and its parents, respectively, as well as the corresponding nodes.

The joint probability distribution over $\boldsymbol{U}$ given a structure $\mathscr{G}$ assumed to encode this probability is given by (Pearl, 1988; Heckerman, 1995; Jensen, 2001)

$$P(\boldsymbol{X} = \boldsymbol{x} \mid \mathscr{G}) = \prod_{i=1}^{n} P(X_i = x_i \mid \boldsymbol{Pa}_i, \mathscr{G}) \tag{1}$$

where $\boldsymbol{x}$ is the assignment of states to each of the n variables in $\boldsymbol{U}$, x_i is the value taken by X_i,

Table 1
The set of FISH signal features studied in the work

Number	Feature	Number	Feature
1	Area	7	Average hue
2	Eccentricity	8	Hue texture
3	Total channel intensity	9	Delta hue
4	Average channel intensity	10	Eig. 1
5	Texture	11	Eig. 2
6	Maximum hue	12	Average gray intensity

Numbers are used to identify the features. 'Texture' indicates standard deviation of intensity (5) or hue (8). Eig. 1, 2 are abbreviations for the two coordinates of the eigenvector corresponding to the largest eigenvalue of the red and green intensity components of the signal.

and the terms in the product compose the required set of local probability distributions $\boldsymbol{\theta}$ quantifying the dependence relations.

The computation of the joint probability distribution and any probability related to the joint is conditioned on the graph. Therefore, before computing probabilities of interest we must obtain a structure either from knowledge provided by an expert or learned from the data.

3.1. Learning the BN structure

Constructing the BN structure is commonly performed by employing expert knowledge to identify variables in, and extracting dependencies and independencies from, the problem domain. Due to its apparent simplicity this attitude to learning a BN structure seems very reasonable. However, the expert-based structure may be different from expert to expert, difficult to obtain in the absence of an expert and time-consuming to be manually constructed even when the domain is known. Thus, there are advantages in learning a BN structure directly from the data.

Most methods of learning a BN structure from the data are score-based (Cooper and Herskovits, 1992; Heckerman, 1995; Friedman et al., 1997). These methods comprise of a search procedure after a network structure and a score employed to evaluate each structure in the search space. An exhaustive search in which every possible DAG is scored is limited to structures with small numbers of nodes. Alternatively, a sub-optimal hill-climbing method is the heuristic K2 algorithm (Cooper and Herskovits, 1992). The algorithm begins using a total ordering on the nodes, i.e., a parent precedes its children, and the assumption that a node has no parents. It then adds incrementally that parent from the ordering whose addition increases the probability (score) of the resulting structure the most. The algorithm stops adding parents to a node when the addition of no single parent can increase the score.

A common score to evaluate structures is the Bayesian score,

$$P(\mathscr{G} \mid D) = \frac{P(D \mid \mathscr{G})P(\mathscr{G})}{P(D)} = \frac{P(\mathscr{G}, D)}{P(D)} \tag{2}$$

of a structure $\mathscr{G}$ given a data set $D = \{d_1, d_2, \ldots, d_N\}$ which is a random sample from the joint probability distribution of $\boldsymbol{U}$. Since $P(D)$ does not depend on the structure, we may employ for the score, the joint probability, $P(\mathscr{G}, D)$, the posterior probability, $P(\mathscr{G} \mid D)$, or the marginal likelihood, $P(D \mid \mathscr{G})$, assuming all structures are a priori equally probable. For the K2 algorithm this criterion is called the Bayesian Dirichlet metric, defined as (Cooper and Herskovits, 1992):

$$P(\mathscr{G}, D) = P(\mathscr{G}) \prod_{i=1}^{n} \prod_{j=1}^{q_i} \frac{(r_i - 1)!}{(N_{ij} + r_i - 1)!} \prod_{k=1}^{r_i} N_{ijk}! \tag{3}$$

where $P(\mathscr{G})$ is the (constant) structure prior probability, n is the number of nodes in the graph, q_i is the number of configurations (states) of the parents of the ith node, r_i is the number of mutual exclusive states of the ith node, N_{ijk} is the number of instances D in which the ith node is in the kth state when its parents are in their jth configuration and $N_{ij} = \sum_{k=1}^{r_i} N_{ijk}$. This score is scalable for large data sets and decomposable since it is assumed that the parameters associated with each variable are mutually independent (a priori and due to data completeness also a posteriori). Thus, the score can be written as a product of independent subscores one for each variable.

3.2. Learning the BN parameters

Eq. (1) summarized the joint probability over the graph as a product of local probability distributions one for each node conditioned on its parents. In this study, all except one of the variables (that representing the FISH signal area) are continuous, hence we have either to estimate densities (Lerner, 2004) or quantize the variables. We quantize the variables and study the dependence of classifier accuracy on the number of quantization levels. Following quantization, we model local probabilities using the unrestricted multinomial distribution. According to this model, each variable $X_i \in \boldsymbol{U}$ is discrete having r_i possible values $x_i^1, \ldots, x_i^{r_i}$, and each local probability of the decomposed distribution in (1) can be described as:

$$P(x_i^k \mid \boldsymbol{pa}_i^j, \boldsymbol{\theta}_i, \mathcal{G}) = \theta_{ijk} > 0,$$
$$\sum_{k=1}^{r_i} \theta_{ijk} = 1 \quad \forall i, j \tag{4}$$

where $\boldsymbol{pa}_i^j$ is the jth configuration of $\boldsymbol{Pa}_i$ and $\boldsymbol{\theta}_i = \{\theta_{ijk}\}_{jk}$ is the set of parameters of the ith node taking up every possible value x_i^k and configuration of parents $\boldsymbol{pa}_i^j$.

Learning probabilities in the Bayesian formalism is updating the prior probability for the parameters, $P(\boldsymbol{\theta}_s \mid \mathcal{G})$, using the training data set D to produce an improved a posteriori estimate $P(\boldsymbol{\theta}_s \mid \mathcal{G}, D)$, where $\boldsymbol{\theta}_s$ is the vector of parameters $(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_n)$. For the multinomial distribution, the Dirichlet distribution is a conjugate prior and thus the posterior is also Dirichlet. After seeing N cases of D (i.e., training) we summarize N_{ij} instances in which the parents of node i are in their jth configuration among them in N_{ijk} of the instances the node takes up its kth state. Thereby, we are able to compute the maximum a posteriori (MAP) estimation that in the case $X_{N+1}, x_i = x_i^k$ and $\boldsymbol{Pa}_i = \boldsymbol{pa}_i^j$ (i.e., test) as (Heckerman, 1995)

$$P(X_{N+1} \mid \mathcal{G}, D) = \prod_{i=1}^{n} \frac{\alpha_{ijk} + N_{ijk}}{\alpha_{ij} + N_{ij}}. \tag{5}$$

The hyper-parameter of the Dirichlet distribution $\alpha_{ijk} > 0$ corresponds to the a priori probability distribution of X_i taking up its kth state while its parents are in their jth configuration and $\alpha_{ij} = \sum_{k=1}^{r_i} \alpha_{ijk}$. Different estimates are obtained for different choices of hyper-parameters with most approaches making the simplifying assumption that $\alpha_{ijk} = \alpha$, $\forall i$ and $\alpha = 0$, 0.5 or 1 (Cooper and Herskovits, 1992; Heckerman, 1995). Degenerating the estimate by taking all hyper-parameters to be 0 (as in this study) we get the sample frequency $\theta_{ijk} = N_{ijk}/N_{ij}$ as the estimate for each local probability which is the maximum likelihood (ML) solution.

3.3. Inference

Probabilistic inference in BNs is the task of calculating the conditional probability distribution of a subset of the nodes in the graph (the 'hidden' nodes) given another subset of the nodes (the 'observed' nodes). In classification problems, the former subset includes only the class variable and the latter subset contains variables representing the features. The procedure used in this work for probabilistic inference is the junction tree algorithm (Lauritzen and Spiegelhalter, 1988; Huang and Darwiche, 1994), though other methods (Pearl, 1988) may do as well.

4. Experimentation and results

In order to perform the experiments, valid ('real') signals and artifacts of fluorophores demonstrating two genetic syndromes in the FISH images are represented using the features of Table 1 and classified as Real-Red, Artifact-Red, Real-Green and Artifact-Green. The features are represented by the observable variables (nodes), and the (hidden) class variable takes up four states associated with the four possible classes. Graph edges point from cause to effect, i.e., from the class node to the feature nodes responsible for signal representation.

In the first section of experimentation we compare different methods to learning a BN structure with that of the NBC shown in Fig. 1. The first method uses expert knowledge to improve the NBC by adding necessary, and removing unnecessary, edges. Added edges from the max-hue (6) and average-hue (7) nodes to the delta-hue (9) node, as well as from the area (1) and total (3) and average (4) channel intensity nodes to the average gray intensity (12) node manifest dependency relations

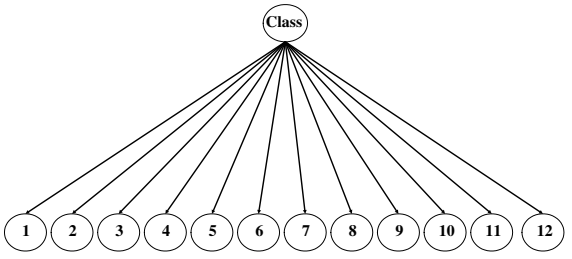

Fig. 1. The naive Bayesian classifier represented as a belief network applied to FISH signal classification. The class node is the root of the structure (cause) and the numbered nodes represent the features of Table 1 (effects).

between nodes as reflected from expert knowledge. The resulted expert-based structure is presented in Fig. 2. In the second method to learning a BN structure, we apply the K2 algorithm to search for the structure having the maximum marginal likelihood for classifying the FISH data when the initial topological ordering inferred from expert knowledge. The learned from data structure resulting from the K2 algorithm is presented in Fig. 3. The figure shows that the K2 algorithm performs a sort of feature selection by not connecting to the structure variables not contributing to the likelihood.

In the second section of experimentation aiming at learning the BN probabilities we study quantization of continuous variables composing the FISH domain letting states of nodes represent quantization levels. We evaluate the optimal number of quantization levels, q, using the classification accuracy and 10-fold cross-validation (CV) experiment. Fig. 4 shows the four-class classification accuracy on the test set of the expert-based BN of Fig. 2 for increasing sizes of q. It shows that using a few quantization levels ($q \leqslant 6$) smooths the densities and impends sufficient accuracy. Fig. 4 also demonstrates the 'curse of dimensionality' phenomenon in which the accuracy is reduced for a fixed data set and increasing dimensionality $|D| \propto q^n$, where $|D|$ is the required number of instances in the data set and n is the dimension of the feature space, i.e., the number of nodes in the BN. For the FISH data, accuracy deterioration begins beyond twelve quantization levels which is selected as the optimal number of levels for classification of this data.

Finally, we evaluate in Table 2 the accuracy of BNs classifying FISH signals into the four classes employing structures derived by either expert knowledge (expert-based BN) or the K2 algorithm (K2-based BN) in comparison to that of the NBC. The NBC either utilizes kernel density estimation (continuous NBC) (Lerner, 2004) or density quantization (NBC). The comparison reveals only slight differences between the models all achieving

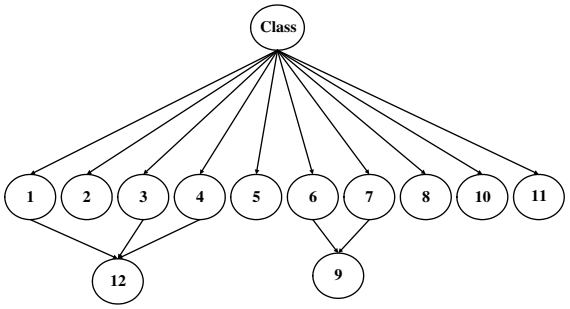

Fig. 2. A BN structure incorporating expert knowledge for FISH signal classification.

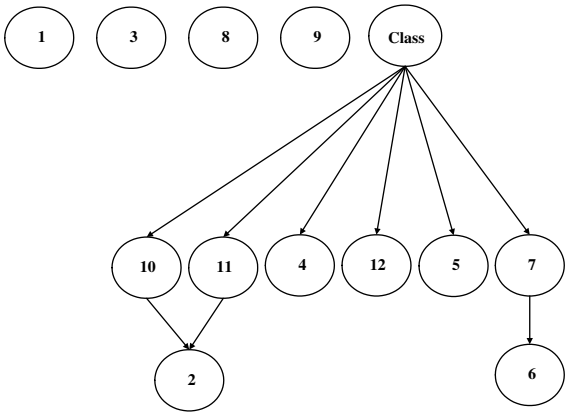

Fig. 3. A BN structure obtained using the K2 algorithm for FISH signal classification.

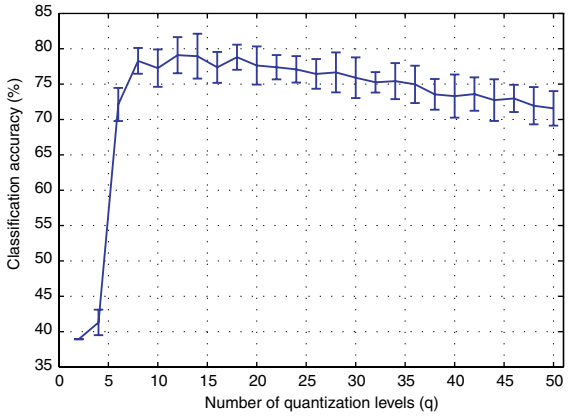

Fig. 4. The four-class classification accuracy (mean ± std) of the expert-based BN of Fig. 2 for increasing numbers of quantization levels employed in estimating probability densities.

Table 2
The accuracy of a BN classifying FISH signals for structures obtained using different methodologies

Model	Classification accuracy (mean (std) in %)
Continuous NBC	78.66 (1.90)
NBC	77.06 (2.55)
Expert-based BN	79.13 (2.18)
K2-based BN	78.02 (1.62)

almost 80% of test accuracy with a slight advantage to the expert-based structure.

5. Discussion

Automatic signal classification is a key task for efficient analysis of FISH images necessary in clinical diagnosis of genetic abnormalities. In this study, we suggested a classification methodology based on a belief network identifying and manifesting causation, dependencies and independencies among domain variables both qualitatively and quantitatively. We learned a BN structure capturing these relations and conditional probabilities quantifying them before predicting a new pattern. We studied structure learning based on expert knowledge and a search-and-score procedure called the K2 algorithm in comparison to the NBC, a restricted BN assuming variable independence given the class variable. For each of the learned structures, we learned model parameters by assuming variable multinomial distribution, quantizing continuous variable values and calculating the ML solution. For the NBC we both quantized and estimated densities.

When learning probabilities we experienced peaked accuracy of the BN classifier when changing the number of quantization levels. Dropping off the peak for increasing number of levels was a consequence of the curse-of-dimensionality. Indeed for the FISH database, variable states derived for too many quantization levels were estimated using very few examples thereby reducing classifier accuracy. Nevertheless, accuracies of the three BN schemes are attributed mainly to accuracies of their structures.

A comparison of classifier accuracy for the structures revealed no major differences where all structures led to almost 80% accuracy. The structure learned using expert knowledge outperformed slightly that of the (discrete) NBC, probably since the former structure was not affected by two features (9 and 12) not too relevant to the class posterior probability (Lerner, 2004). Also, the expert-based structure exceeded that derived from the K2 algorithm since the latter avoided relevant features (2 and 6) to the posterior probability. In addition, the K2 algorithm ignored a several of the features (1, 3, 8 and 9) since density estimation for them using uniform quantization was poor not contributing to the K2 score. To some extent the continuous NBC improved on the NBC quantizing probabilities due to quantization errors of the latter. Similarity of performances between the NBC and other BNs not assuming feature independence has also been reported elsewhere (Langley et al., 1992; Friedman et al., 1997). Nonetheless, the minor differences in our study between the structures deserve more consideration. Since a node in a structure is affected only by nodes in its Markov blanket (i.e., the node itself, its children, parents and children parents) (Pearl, 1988), classification using a BN is affected only by those variables within the class node Markov blanket. Inspection of the Markov blankets of the resulted structures of the expert (Fig. 2) and K2-based (Fig. 3) BNs reveals that both structures can be considered as variants of the NBC structure (Fig. 1). Although using different methods of learning, both methods end up having similar structures to the NBC leading to similar classification accuracies. Regarding the expert-based structure, the similarity to the NBC is the result of inducing the structure exploiting knowledge represented by the NBC (e.g., a sole root which is the class node). Concerning the K2 algorithm-based structure, the similarity is the result of positioning the class node first on the topological ordering rendering this node a graph root and a potential parent having no parents of itself. Then, the class node Markov blanket for both structures includes only the node itself, its children and possible co-parents for these children, all observed. Thus, most resulted structures are either variants of the NBC or of a tree augmented naive Bayes (TAN) (Friedman et al., 1997) allowing each observable

node to have an additional parent besides the class node. Interestingly enough, this result has not been explicitly emphasized in the Bayesian network classifier literature.

Since the same FISH database was previously classified by a neural network (NN) with 89% accuracy (Lerner et al., 2001), we attribute the around 10% gap between classifier accuracies to a several reasons. First, an NN utilizing the minimum-square-error criterion is a classification paradigm, whereas a BN utilizing the Bayesian Dirichlet metric (e.g., in the K2 algorithm) is a paradigm of data representation and probabilistic inference extended to classification but making use of a criterion not directly related to classification. Second, the NN handle continuous features naturally whereas the BN needs quantization leading to some error (e.g., 1.6% for the NBC). Third, the number of parameters to estimate for the BN is much larger (especially for a large number of quantization levels) than for the NN, causing for accuracy deterioration in the BN through the curse-of-dimensionality. Finally, the optimal structure is easily found for the NN, i.e., determining the number of hidden units using a validation set and an estimation method such as cross-validation. However, the number of possible BN structures grows exponentially with the number of nodes so the introduction of assumptions and conditions can make the search feasible (such as in the K2 algorithm) but cannot avoid the learned structure from being almost always non-optimal.

6. Conclusions

Classification of FISH images using BNs has been studied. In order to render the procedure attractive for clinical genetic diagnosis the proposed methodology should be extended in different ways, which are the topic of current research. First, more accurate structures should be sought as the inferiority of current structures in representing the problem is attributed to their inability to capture the exact dependencies within the domain. Unpublished analysis using other BN learning schemes reveals significant dependencies among FISH variables. Thus, regarding the expert-based method, structures not related to the NBC but reflecting the exact dependencies within the domain more closely should be employed. Concerning the K2 algorithm, topological orderings not restricting the class node to the beginning of the order should be evaluated. Moreover, ways making use of knowledge contained within the features in addition to that provided by the expert to better initialize the K2 algorithm will be studied. Inclusion of a feature selection phase in structure learning may also produce structures having less number of parameters to estimate thus more accurate. Also, ways incorporating hidden nodes into the BN in order to capture larger class node Markov blankets (considering the hidden node children) will be explored, all with the aim of improving structure performance. Indeed, a preliminary study of structures employing hidden nodes have shown improvement in accuracy performance. In addition, replacing the Bayesian Dirichlet metric of the K2 algorithm with a criterion directly associated with classification would also contribute to performance improvement and some empirical results (Friedman et al., 1997) support that. Regarding parameter learning, prior probabilities compensating for scarce data and methods optimizing the number of quantization levels separately for each variable or part of the distribution should both be considered.

Finally, although the BN paradigm has been found inferior to the NN in classifying FISH signals, we believe that implementing at least some of the above recommendations would provide graphs that besides explaining relations among variables qualitatively and quantitatively as required in this application and allowing the inclusion of a priori knowledge while learning will also provide better prediction accuracy.

Acknowledgment

The implementation of BNs in this study is based on the Bayes Net Toolbox for Matlab (Murphy, 2001), and the authors thank Kevin Murphy for making this toolbox freely available. This work was supported in part by the Paul Ivanier Center for Robotics and Production Management, Ben-Gurion University, Beer-Sheva, Israel.

References

Cheng, J., Greiner, R., 1999. Comparing Bayesian network classifiers. In: Laskey, K.B., Parde, H. (Eds.), Proc. of the Fifteenth Conf. on Uncertainty in Artificial Intelligence. Morgan Kaufmann Publishers, Stockholm, pp. 101–108.

Cooper, G.F., Herskovits, E., 1992. A Bayesian method for the induction of probabilistic networks from data. Mach. Learning J. 9, 309–347.

Duda, R.O., Hart, P.E., Stork, D.G., 2001. Pattern Classification. Wiley, NY.

Friedman, N., Geiger, D., Goldszmidt, M., 1997. Bayesian network classifiers. Mach. Learning 29, 131–163.

Heckerman, D., 1995. A tutorial on learning with Bayesian networks. Technical Report MSR-TR-95-06, Microsoft Research.

Huang, C., Darwiche, A., 1994. Inference in belief networks: A procedural guide. Int. J. Approx. Reasoning 15, 225–263.

Inza, L., Larranaga, P., Sierra, B., Etxeberria, R., Lozano, J., Pena, J., 1999. Representing the behaviour of supervised classification learning algorithms by Bayesian networks. Pattern Recogn. Lett. 20, 1201–1209.

Jensen, F.V., 2001. Bayesian Networks and Decision Graphs. Springer-Verlag, New York.

Langley, P., Iba, W., Thompson, K., 1992. An analysis of Bayesian classifiers. In: Besnard, P., Hanks, S. (Eds.), Proc. of the Tenth National Conf. on Artificial Intelligence. AAAI Press, San Jose, CA, pp. 223–228.

Lauritzen, S.L., Spiegelhalter, D.J., 1988. Local computation with probabilities on graphical structures and their application to expert systems. J. Royal Statist. B 50, 157–224.

Lerner, B., 2004. Bayesian fluorescence in-situ hybridization signal classification. Artificial Intell. Med. 30, 301–316, Special issue on Bayesian Models in Medicine.

Lerner, B., Clocksin, W.F., Dhanjal, S., Hultén, M.A., Bishop, C.M., 2001. Feature representation and signal classification in fluorescence in-situ hybridization image analysis. IEEE Trans. A Systems Man Cybernet. 31, 655–665.

Lerner, B., Lawrence, N.D., 2001. A comparison of state-of-the-art classification techniques with application to cytogenetics. Neural Comput. Applicat. 10, 39–47.

Murphy, K., 2001. The Bayes Net Toolbox for Matlab. Comput. Sci. Statist. 33.

Pearl, J., 1988. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann, San Francisco.

Pena, J., Lozano, J., Larranaga, P., 1999. Learning Bayesian networks for clustering by means of constructive induction. Pattern Recogn. Lett. 20, 1219–1230.

Pham, T., Worring, M., Smeulders, A., 2002. Face detection by aggregated Bayesian network classifiers. Pattern Recogn. Lett. 23, 451–461.