BELIEF NETWORKS FOR CYTOGENETIC IMAGE CATEGORIZATION

Boaz Lerner and Roy Malka

Pattern Analysis and Machine Learning Laboratory
Department of Electrical & Computer Engineering
Ben-Gurion University, Beer-Sheva, Israel
{boaz,malkah}@ee.bgu.ac.il

ABSTRACT

The structure and parameters of a belief network are learned in order to categorize cytogenetic images enabling the detection of genetic syndromes. We compare a structure learned from the data to another obtained utilizing expert knowledge and to the naive Bayesian classifier. We also study feature quantization needed for parameter learning in comparison to density estimation. Both networks achieve comparable accuracy for the cytogenetic database with a slight advantage to that based on expert knowledge.

1. INTRODUCTION

Fluorescence in-situ hybridization (FISH) signal categorization is essential for clinical diagnosis of genetic abnormalities. One of the most important applications of this categorization in 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 automation of dot counting is needed.

We extend previous work accomplishing dot counting using a neural network (NN) classifier [1] by applying belief networks (BNs) to discriminate valid FISH signals from artifacts. Belief networks besides providing a well-defined formalism for probabilistic inference extend model interpretability and exhibit causal relationships within the domain which are missing in the NN. This is accomplished by manifesting dependencies and independencies between graph nodes representing the domain variables in edges and lack of edges, respectively. The theory of BNs is well established [2] and a several applications of BNs have already been suggested. For example, face images were discriminated using a BN maximizing the Kullback-Leibler divergence between the corresponding joint distributions [3], and decisions of a several classifiers were fused utilizing a BN learned using the K2 algorithm (Section 2) [4].

We address here both learning network structure and parameters (conditional probabilities), and compare structures learned from the data to others obtained utilizing expert knowledge. The obtained BN classifiers are then evaluated in discriminating valid and artifact signals representing two genetic syndromes in order to assess the accuracy of BN FISH signal classification.

Section 2 presents the BN model and structure learning strategies based on expert knowledge or the data, as well as parameter learning and inference. Section 3 demonstrates the capabilities of the strategies in learning a BN for FISH signal classification, while Section 4 summarizes the work.

2. LEARNING IN A BN

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

The second component of a BN is a set of parameters, $\boldsymbol{\theta}$, that specify the conditional probability distributions (or densities) that quantify graph edges. The probability distribution of each $X_i \in \mathbf{U}$ conditioned on its parents in the graph $\mathbf{Pa}_i \in \mathbf{U}$ is $P(X_i = x_i|\mathbf{Pa}_i) \in \boldsymbol{\theta}$.

The joint probability distribution over $\mathbf{U}$ given a structure $\mathcal{G}$ assumed to encode this probability is given by [5, 2]

$$P(\mathbf{X} = \mathbf{x}|\mathcal{G}) = \prod_{i=1}^{n} P(X_i = x_i|\mathbf{Pa}_i, \mathcal{G}) \qquad (1)$$

where $\mathbf{x}$ is the assignment of states to each of the n variables in $\mathbf{U}$, x_i is the value taken by X_i, 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 based on expert knowledge or learned from the data.

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 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.

One common approach to learning a structure from data is a score-based [6, 5]. This approach comprises of a search procedure after a network structure and a score employed to evaluate each structure found. An exhaustive search in which every possible DAG being scored is limited to structures having small numbers of nodes. A sub-optimal hill-climbing method is the heuristic K2 algorithm [6]. The algorithm begins using node ordering and assuming a node has no parents. It adds incrementally that parent from the ordering whose addition increases the probability (score) of the resulting structure the most. We stop adding parents to a node when the score no longer increases.

A score commonly employed is the Bayesian score,

$$P(\mathcal{G}|D) = \frac{P(D|\mathcal{G})P(\mathcal{G})}{P(D)} = \frac{P(\mathcal{G}, D)}{P(D)} \qquad (2)$$

of a structure $\mathcal{G}$ given a data set $D = \{d_1, d_2, \ldots, d_N\}$ which is a random sample from the joint probability distribution of $\mathbf{U}$. Since $P(D)$ does not depend on the structure, we may compute $P(\mathcal{G}|D)$ or $P(\mathcal{G}, D)$ as the score. Moreover, if the graphs are equally probable than the likelihood $P(D|\mathcal{G})$ is also an equivalent score . For the K2 algorithm this score is defined as [6]:

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

where $P(\mathcal{G})$ is the structure prior probability, n is the number of nodes in the graph, and q_i and r_i are the numbers of parent configurations and mutual exclusive states of the ith node, respectively. N_{ijk} is the number of instances of the ith node being 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.

Learning the BN parameters

Equation 1 has 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 are continuous, hence we have either to estimate densities [7] or quantize the variables and estimate the distribution using relative frequencies. Taking the second approach and following quantization, we model local probabilities using the unrestricted multinomial distribution, where the distribution parameters are obtained using the maximum-likelihood solution [6, 5].

Inference

In inference we calculate the conditional probability distribution of a subset of the nodes in the graph ('hidden' nodes) given another subset of the nodes ('observed' nodes). We used here a reputable method for exact probabilistic inference known as the junction tree [8].

3. EXPERIMENTATION AND RESULTS

Valid signals and artifacts of red and green fluorophores demonstrating two genetic syndromes are analyzed in FISH images and classified as Real-Red, Artifact-Red, Real-Green and Artifact-Green. Data preparation and image acquisition is described thoroughly in [1] and avoided here. Twelve features are measured for the signals including Area, Eccentricity and a number of spectral features. We compute 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 compute four HSI (hue-saturation-intensity) 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 features are represented by the observable variables (nodes) of the BN, and the (hidden) class variable takes up four states associated with the four possible classes.

First, we learn BN structures in comparison to the naive Bayesian classifier (NBC) having the class node the only parent of all other nodes. The first learning method uses expert knowledge to improve the NBC by adding necessary, and removing unnecessary, edges. The expert-based structure is presented in Figure 1. Added edges from the Max-Hue (6) and Average-Hue (7) nodes to the Delta-Hue (9) node, and from the Area (1) and Total (3) and Average (4) Channel Intensity nodes to the Average Gray Inten-

sity (12) node manifest dependency relations between parent and child nodes as reflected from expert knowledge. In the second method we start with an initial topological ordering coinciding with that inferred from the expert structure and apply the K2 algorithm to search for the structure maximizing the likelihood for the data. The resulting structure is presented in Figure 2. Second, in order to learn the BN probabilities we evaluate the required number of quantization levels, q, using the classification accuracy and a CV-10 experiment. Figure 3 shows the four-class classification accuracy on the test set for the expert-based BN for increasing sizes of q. It reveals that using a few quantization levels ($q \leq 6$) smooths the densities impeding sufficient accuracy. Figure 3 also demonstrates the 'curse of dimensionality' phenomenon in which the accuracy is reduced for a fixed data set and increasing dimensionality. 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 1 the accuracy of classifying FISH signals into the four classes by employing BN structures derived from either expert knowledge (expert-based BN) or the K2 algorithm (K2-based BN) in comparison to that of the NBC either utilizing kernel density estimation (Continuous NBC) [7] or density quantization (NBC). The comparison reveals only slight differences between the models all achieving around 80% of test accuracy with some advantage to the expert-based structure. We explain the inferiority of the K2-based structure compared to the expert-based structure by the minimization of a score other than the miss-classification error, and the inferiority of the NBC to the Continuous NBC by the not necessarily accurate representation of all variables by the same number of quantization levels.

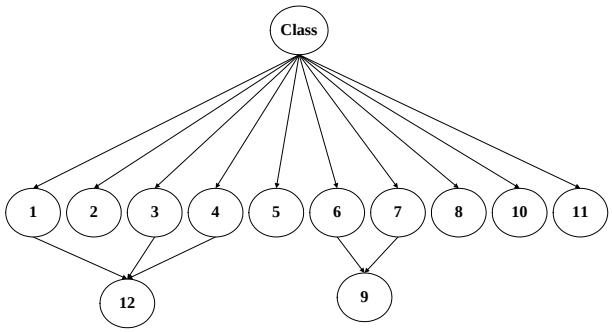

Fig. 1. A BN structure modifying the naive Bayesian classifier to incorporate expert knowledge. Numbers represent features by their order of appearance in Section 3.

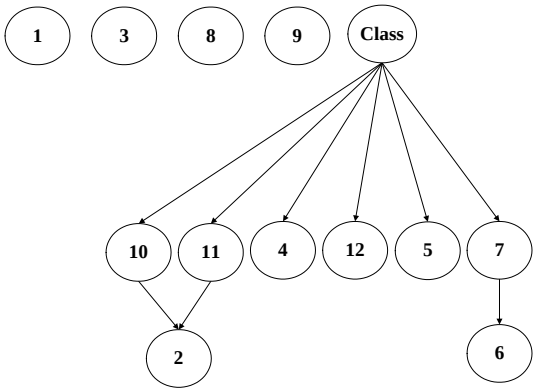

Fig. 2. A BN structure obtained using the K2 algorithm and a topological ordering coinciding with that of the expert-based structure of Figure 1. Numbers represent features by their order of appearance in Section 3.

4. DISCUSSION

BN structures based on expert knowledge and the K2 algorithm in comparison to the NBC interpreted and represented the FISH domain differently. For each structure, we learned model parameters by quantizing variable values and studied the classifier sensitivity to the number of quantization levels. Maintaining or improving classifier accuracy by raising the number of quantization levels was subject to the curse-of-dimensionality.

Classification using each of the structures has led to around 80% accuracy with slight superiority of the expert structure. Each of the evaluated structures suffers from a particular shortcoming. The NBC endures from the unjustified to the FISH domain independent assumption [7]. The expert-based structure is restricted to the expert knowledge which is not guaranteed to reflect all dependencies between the variables. The relative inferiority of the K2-based structure is attributed to the particular expert-based initial topological ordering and the greedy rather than exhaustive search procedure that is employed.

Both the expert-based and K2-based structures derived from expert knowledge and achieved accuracy inferior to that of the neural network [1]. Thus, we believe that further exploration into the FISH domain yielding structures manifesting variable relations more accurately will lead to performance improvement of the BN.

Acknowledgment: We implemented the BNs using the Bayes Net Toolbox [9], and thank K. Murphy for making it publicly available. This work was supported in part by the Paul Ivanier Center for Robotics and Production Management, Ben-Gurion University, Beer-Sheva, Israel.

5. REFERENCES

[1] B. Lerner, W. F. Clocksin, S. Dhanjal, M. A. Hultén, and C. M. Bishop, "Feature representation and signal classification in fluorescence in-situ hybridization image analysis," *IEEE Transactions on Systems, Man, and Cybernetics A*, vol. 31, pp. 655–665, 2001.

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

[3] T.V. Pham, M. Worring, and A.W.M. Smeulders, "Face detection by aggregated Bayesian network classifiers," *Pattern Recognition Letters*, vol. 23, pp. 451–461, 2002.

[4] I. Inza, P. Larranaga, B. Sierra, R. Etxeberria, J.A. Lozano, and J.M. Pena, "Representing the behaviour of supervised classification learning algorithms by Bayesian networks," *Pattern Recognition Letters*, vol. 20, pp. 1201–1209, 1999.

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

[6] G. F. Cooper and E. Herskovits, "A Bayesian method for the induction of probabilistic networks from data," *Machine Learning Journal*, vol. 9, pp. 309–347, 1992.

[7] B. Lerner, "Bayesian fluorescence in-situ hybridization signal classification," *Artificial Intelligence in Medicine*, vol. 30, pp. 301–316, 2004, Special issue on Bayesian models in medicine.

[8] S. L. Lauritzen and D. J. Spiegelhalter, "Local computation with probabilities on graphical structures and their application to expert systems," *Journal of Royal Statistics B*, vol. 50, pp. 157–224, 1988.

[9] K. Murphy, "The Bayes Net Toolbox for Matlab," *Computing Science and Statistics*, vol. 33, 2001.

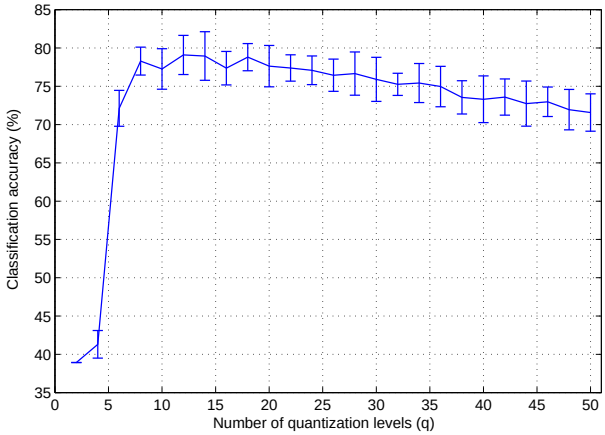

Fig. 3. Classification accuracy of the expert-based BN of Figure 1 for increasing numbers of quantization levels employed in estimating probability densities.

Table 1. Classification accuracy of the expert and K2-based BNs in comparison to the NBC.

Model	Classification Accuracy (Mean (Std) in %)
Expert-based BN	79.13 (2.18)
K2-based BN	78.02 (1.62)
NBC	77.06 (2.55)
Continuous NBC	78.66 (1.90)