Proceedings of the Twenty-Third International Florida Artificial Intelligence Research Society Conference (FLAIRS 2010)

Towards a More Expressive Model for Dynamic Classification

Shengtong Zhong*
Norwegian University
of Science and Technology
N-7491 Trondheim (Norway)

Ana M. Martínez
University of
Castilla-La Mancha
02071 Albacete (Spain)

Thomas D. Nielsen
Aalborg University
Selma Lagerlöfs Vej 300
DK-9220 Aalborg (Denmark)

Helge Langseth
Norwegian University
of Science and Technology
N-7491 Trondheim (Norway)

1. Introduction

Monitoring a complex process often involves keeping an eye on hundreds or thousands of sensors to determine whether or not the process is in a normal state. We have been working with data from an oil production facility in the North Sea, where unstable situations should be identified as soon as possible. Currently, this data is analyzed in real time by experienced engineers, with the goal of avoiding a number of either dangerous or costly situations. The framework of dynamic Bayesian networks (Ghahramani 1998) is a viable framework for modeling the dynamic processes characterizing this domain. A simple instance of this framework is the Hidden Markov Model (HMM), which Smyth (1994) has previously considered for classification purposes; to this end the "hidden node" in the HMM is used as the classification node, and the attributes at time $t-1$ are assumed to be independent of those at time t given the class label at any of these time points. However, the independence assumptions underlying the HMM are often violated, and more expressive models may therefore be required in order to improve classification accuracy. In this paper we will take a first step in that direction, by examining the underlying assumption of some well-known probabilistic classifiers and their natural extensions to dynamic domains. We do so by carefully linking our analysis back to the Oil production dataset.

2. From static to dynamic classifiers

In this section we develop a general framework for performing dynamic classification in domains with only continuous attributes. The framework will be specified incrementally by evaluating its expressiveness relative to the Oil production data.

In the well-known HMM the underlying independence assumptions imply that temporal dynamics are described for the class variables only, and for any given time point the attributes are usually assumed to be conditional independent given the class variable at that time step. The latter implies that at any time t we have a NB model defined by a diagonal covariance matrix, thus requiring only parameters equal to the number of attributes at time t. Unfortunately, as for the (static) NB model, these independence assumptions are often violated in real-world settings and also in the Oil production data.

Modeling dependence using latent variables

There are several approaches to model attribute dependence. In this paper we introduce latent variables to encode the conditional dependencies among attributes. Specifically, for each time step t we have a collection $\{Z_1^t, \ldots, Z_k^t\}$ of latent variables that appear as children of the class variable and parents of all the attributes. The latent variable Z_i^t is assigned a Gaussian distribution conditional on the class variable and the attributes Y_j^t are assumed to be linear Gaussian distributions conditional on the latent variables, i.e.:

$$\mathbf{Z}^t \mid \{C^t = c\} \sim \mathcal{N}(\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c),$$
$$\mathbf{Y}^t \mid \{\mathbf{Z}^t = \mathbf{z}\} \sim \mathcal{N}(\mathbf{Lz}, \boldsymbol{\Theta}),$$

where $\boldsymbol{\Sigma}_c$ and $\boldsymbol{\Theta}$ are diagonal and $\mathbf{L}$ is the transition matrix.

In this model, the latent variables capture the dependencies between attributes. They are conditionally independent given the class but marginally dependent. Furthermore, the same mapping, $\mathbf{L}$, from the latent space to the attribute space is used for all classes, and hence, the relation between the class and the attributes is conveyed by the latent variables. More informally, at each time point the model can be seen as combining a naive Bayes model with a mixture of factor analyzers (FA), which has certain similarities to the model proposed by Langseth and Nielsen (2005).

This model assumes that the attributes in different time slices are independent given the class variable. This assumption implies that the temporal dynamics is captured at the class level. However, when the state specification of the class variable is coarse, then this assumption will rarely hold.[1] For example, if the class variable only contains the states *ok* and *not-ok*, then the model structure specifies that when the system state is e.g. *ok*, then the state specific dynamics of the system is stable. For the Oil production data, this assumption does not hold as we can see from Figure 1(a), which shows the correlation of one attribute in successive time slices.

*The first two authors have contributed equally to this paper.

Copyright © 2010, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

[1] Obviously, the finer the granularity of the state specification of the class variable, the more accurate this assumption will be.

We propose to address this apparent shortcoming by modeling the dynamics of the system at the level of the latent variables, which semantically can be seen as a compact representation of the "true" state of the system. Specifically, we encode the state specific dynamics by assuming that the latent variable Z_i^t follows a linear Gaussian distribution conditioned on Z_i^{t-1}:

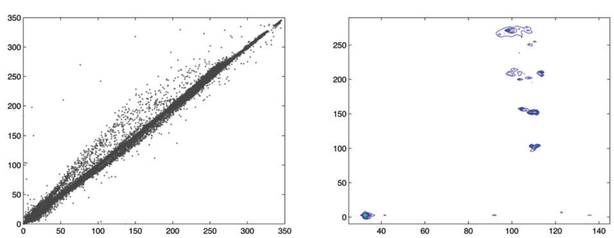
$$Z_i^t \mid \{C^t = c, Z_i^{t-1} = z\} \sim \mathcal{N}(w_{i,c}z, \sigma_c^2).$$

The latent variables of the FA are now used to capture the dynamics in the process as well as modeling dependences between attributes

Modeling non-linear systems

One of the main assumptions in the model above is that there is a linear mapping from the latent variables to the attribute space as well as a linear mapping from the latent variables in a given time slice to the latent variables in the succeeding time slice (i.e., that the state specific dynamics are linear). In combination, the two linearity assumptions imply that conditioned on all the class variables from time 1 to time t, the marginal distribution over the attributes at time t follows a multivariate Gaussian distribution.

Although this distribution assumption may be reasonable for some domains, clearly there are also domains where this assumption is violated. By investigating the class conditional empirical probability distribution of a pair of attributes, we obtain the empirical density shown in Figure 1(b). It is clear that the process that has generated the data does not follow a multivariate Gaussian distribution and hence, the previous linear approach does not provide a reasonable specification for the domain.



(a) Attribute correlation in successive time slices.

(b) Empirical distribution functions for two variables.

Figure 1: 1(a): Scatter plot of one attribute at time $t-1$ (x-axis) and at time t (y-axis) for the *ok* class. The conditional dynamic correlation of this attribute over time is evident. 1(b): Contour lines of the empirical distribution functions for two attributes in the Oil production data. It is evident that the process which generated the data does not follow a Gaussian distribution.

In order to extend the expressive power of the proposed classifier, we introduce a mixture variable M^t for each time slice to capture the non-linearity (Langseth and Nielsen 2005). Each time slice can now be seen as combining a naive Bayes model with a *mixture* of factor analyzers. In this case, the mixture variable follows a multinomial distribution conditioned on the class variable, and the attributes Y_j^t follow a Gaussian distribution conditioned on the latent variables and the discrete mixture variable, i.e.:

$$M^t \mid \{C^t = c\} \sim \text{Multinomial},$$
$$\mathbf{Y}^t \mid \{\mathbf{Z}^t = \mathbf{z}, M^t = m\} \sim \mathcal{N}(\mathbf{L}_m\mathbf{z}, \mathbf{\Theta}_m),$$

where $m \in sp(M)$ ($sp(M)$ indicates the space of M).

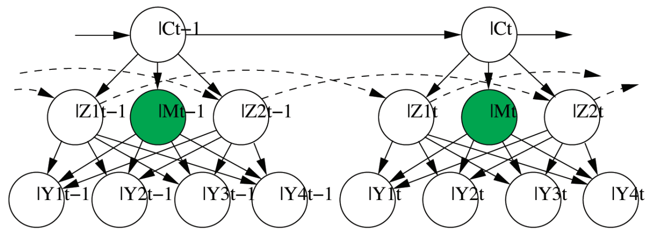


Figure 2: A mixture variable M^t is introduced at each time slice to extend the expressiveness of the model.

3. Conclusions and Future Work

In this paper we have proposed a generative model for performing dynamic classification of temporarily correlated data where latent variables are used to model class conditional correlations among the attributes as well as state specific dynamics. The work is part of an ongoing project involving data from an oil production facility in the North Sea. As part of our ongoing and future work is the implementation of the proposed model. The structure of the model also implies the study and posterior application of suitable approximate inference algorithms both for the learning phase and the inference/classification task.

Acknowledgments

This work has been partially supported by the Spanish Goverment under Project TIN2007-67418-C03-01 and the FPU grant with reference number AP2007-02736. We would also like to thank Verdande Technology for kindly providing us with access to the oil-production data.

References

Ghahramani, Z. 1998. Learning dynamic Bayesian networks. In *Adaptive Processing of Sequences and Data Structures, International Summer School on Neural Networks, "E.R. Caianiello"-Tutorial Lectures*, 168–197. London, UK: Springer-Verlag.

Langseth, H., and Nielsen, T. D. 2005. Latent classification models. *Machine Learning Journal – Special Issue on Probabilistic Graphical Models for Classification* 59(3):237–265.

Smyth, P. 1994. Hidden Markov models for fault detection in dynamic system. *Pattern Recognition* 27(1):149–164.