Optimal Dynamic Graphs for Video Content Analysis

Tao Xiang and Shaogang Gong
Department of Computer Science
Queen Mary, University of London, London E1 4NS, UK
{txiang,sgg}@dcs.qmul.ac.uk

Abstract

This study addresses the problem of learning the optimal structure of a dynamic graphical model for video content analysis given sparse data. We propose a Completed Likelihood AIC (CL-AIC) scoring function that differs from existing ones by optimising *explicitly* both the explanation and prediction capabilities of a model simultaneously. We demonstrate that CL-AIC is superior to existing scoring functions including BIC, AIC and ICL in building dynamic graph models for video content analysis.

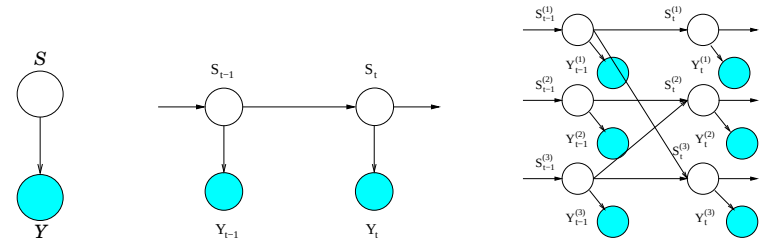
1 Introduction

Dynamic graph models, and in particular Dynamic Bayesian Networks (DBNs) including Hidden Markov Models (HMMs) and their variants, have become increasingly popular for modelling and analysing space-time visual data [8, 13, 4, 7, 12, 9]. By using a DBN, we assume that dynamic visual data are generated sequentially by some hidden states of a dynamic scene evolving over time. Since the hidden states cannot be observed directly, they can only be inferred from the observed visual data given a learned DBN. Learning a DBN involves estimating both its structure and parameters from data. The structure of a DBN refers primarily to (1) the number of hidden states of each hidden variable in a model and (2) the conditional dependence among hidden states of all the hidden variables of a model, i.e, factorisation of the model state space for determining the topology of a graph network. There have been extensive studies in the machine learning community on efficient parameter learning when the structure of a model is known *a priori* (i.e. assumed) [11]. However, much less efforts have been made to tackle the more challenging problem of learning the optimal structure of an unknown DBN [2, 10, 6]. As a consequence, most previous DBNs-based visual data modelling approaches avoid the structure learning problem by setting the structure manually [13, 4, 9]. However, it has been shown that a learned structure can be advantageous over those that are manually set [12].

Previous automatic structure learning techniques have adopted a search-and-score paradigm [10] [1], within which one first defines a scoring function consisting of a maximum likelihood term and a penalty term to penalise complex model structures whilst optimising data fitting. The model structure space is then searched to find the optimal model structure with the highest score. The most commonly used scoring functions include Bayesian Information Criterion (BIC) [19] , Minimum Description Length (MDL) [18], BDe [10], and Akaike's Information Criterion (AIC) [1]. The selected models are 'optimal' in a sense that they can either best explain the existing data (BIC, MDL), or best predict unseen data (AIC). It has been demonstrated both theoretically and experimentally

[1] Alternatives include the Bayesian approach to model selection [2] and context-specific independence [5].

BMVC 2006 doi:10.5244/C.20.19



(a) GMM (b) HMM (c) DML-HMM

Figure 1: Three different types of graphical models with hidden nodes, among which HMM and DML-HMM are DBNs. Observation nodes are shown as shaded circles and hidden nodes as clear circles.

in the case of static models that explanation oriented scoring functions suffer from model under-fitting while prediction oriented ones suffer from model over-fitting [14, 3, 21]. In the case of dynamic models, this is also true (see experimental results presented later in this paper). To address the problems associated with existing scoring functions, especially in determining the structure of dynamic models, we extend in this work Completed Likelihood AIC (CL-AIC) for learning an optimal dynamic Bayesian graph model and demonstrate its effectiveness in video content analysis when only sparse and noisy visual data are available. CL-AIC was first introduced in the case of a Gaussian Mixture Model (GMM) which can be represented as a static graphical model [21] (see Figure 1(a)). In this paper, we show that CL-AIC can be derived for any graphical models with hidden variables, with GMMs and DBNs as special cases. In particular, CL-AIC is formulated for determining the number of hidden states of a HMM and for learning the topology of a Dynamically Multi-Linked HMM (DML-HMM) (see Figure 1(b)&(c)). The effectiveness of CL-AIC on DBNs structure learning is demonstrated through comparative experiments against BIC, AIC and Integrated Completed Likelihood (ICL) [3].

2 CL-AIC for Graphical Models with Hidden Variables

We extend the formulation of Completed Likelihood AIC (CL-AIC) from GMMs to the more general case of graph models with hidden variables. Consider an observed data set $\mathcal{Y}$ modelled by a graphical model $\mathcal{M}_K$ with hidden variables. $\mathcal{M}_K$ can be used to perform three tasks: (1) estimating the unknown distribution that most likely generates $\mathcal{Y}$, (2) inferring the values of hidden variable in $\mathcal{M}_K$ from $\mathcal{Y}$, and (3) predicting unseen data. Computing (1) and (2) emphasises data explanation while solving (3) concerns with data prediction and synthesising. In this context, scoring functions based on approximating the Bayesian Model Selection Criterion [17] such as BIC choose a model that maximises $p(\mathcal{Y}|\mathcal{M}_K)$, the probability of observing $\mathcal{Y}$ given $\mathcal{M}_K$. They thus enforce mainly task (1). AIC, on the other hand, chooses the model that best predicts unseen data, therefore optimising (3). CL-AIC utilises Completed Likelihood (CL) in order to makes explicit the task (2) while following a similar derivation procedure as AIC.

Completed Likelihood (CL) was originally derived for mixture models [3]. We wish to extend the definition of Completed Likelihood (CL) to a general dynamic graphical model with hidden variables. The complete data, denoted as $\bar{\mathcal{Y}}$, for such a model is a combination of the observed data ($\mathcal{Y}$) and the values of the hidden variables ($\mathcal{Z}$): $\bar{\mathcal{Y}} = \{\mathcal{Y}, \mathcal{Z}\}$, where

$\mathcal{Z}$ is unknown, and must be inferred from $\mathcal{Y}$. The completed log-likelihood of $\bar{\mathcal{Y}}$ is:

$$\mathrm{CL}(K) = \log p(\mathcal{Y}|\mathcal{M}_K, \boldsymbol{\theta}_{\mathcal{M}_K}) + \log p(\mathcal{Z}|\mathcal{Y}, \mathcal{M}_K, \boldsymbol{\theta}_{\mathcal{M}_K})$$

where $\boldsymbol{\theta}_{\mathcal{M}_K}$ are the true model parameters and K is the index of the candidate models. In practice, $\boldsymbol{\theta}_{\mathcal{M}_K}$ are replaced using the ML estimate $\hat{\boldsymbol{\theta}}_{\mathcal{M}_K}$ and the unknown values of the hidden variables $\mathcal{Z}$ is replaced by $\hat{\mathcal{Z}}$, the values inferred from the observed data $\bar{\mathcal{Y}}$. The completed log-likelihood is thus rewritten as:

$$\mathrm{CL}(K) = \log p(\mathcal{Y}|\mathcal{M}_K, \hat{\boldsymbol{\theta}}_{\mathcal{M}_K}) + \log p(\hat{\mathcal{Z}}|\mathcal{Y}, \mathcal{M}_K, \hat{\boldsymbol{\theta}}_{\mathcal{M}_K}) \tag{1}$$

CL-AIC aims to choose a model that best explains the the observed data and has the minimal divergence to the true model, which thus best predicts unseen data. The divergence between a candidate model and the true model is measured using the Kullback-Leibler information [15]. CL-AIC for dynamic graphical models with hidden variables is formulated as:

$$\mathrm{CL\text{–}AIC(K)} = -\log p(\mathcal{Y}|\mathcal{M}_K, \hat{\boldsymbol{\theta}}_{\mathcal{M}_K}) - \log p(\hat{\mathcal{Z}}|\mathcal{Y}, \mathcal{M}_K, \hat{\boldsymbol{\theta}}_{\mathcal{M}_K}) + C_K \tag{2}$$

where C_K is the dimensionality of the parameter space. The derivation of CL-AIC follows a similar procedure as that of AIC [1].

Unlike previous scoring functions, CL-AIC attempts to optimise *explicitly* the explanation and prediction capabilities of a model. This makes CL-AIC theoretically attractive. The effectiveness of CL-AIC in practice is demonstrated through experiments in the following sections. It has been shown that Completed Likelihood can be combined with BIC which leads to an Integrated Completed Likelihood (ICL) criterion [3]. However, the experiments reported in [3] indicated that in the case of mixture models, ICL performs poorly when data belonging to different mixture components are severely overlapped. This is caused by ICL being a combination of two explanation oriented criteria without considering any prediction capability of a model. Since CL-AIC integrates an explanation criterion with a prediction criterion, it is theoretically better justified than ICL. Our experiments in the following reinforces this observation.

Let us now consider a specific problem of learning the structure of a Hidden Markov Model (HMM). A HMM can be represented by one hidden variable and one observation variable at each time instance t (see Figure 1(b)). The hidden variable is discrete in most applications. The structure learning problem for a HMM thus refers to how to determine the number of hidden states that the hidden variable can assume. Assuming that at each time instance t, the discrete hidden variable S_t can assume K different values (states), the complete data for the model is $\bar{\mathcal{Y}} = \{\mathcal{Y}, \mathcal{Z}\}$ where $\mathcal{Z}$ is the true hidden variable values (i.e. the true hidden state sequence). The completed log-likelihood of $\bar{\mathcal{Y}}$ is computed as:

$$\mathrm{CL}(K) = \log\left(\sum_S p(\mathcal{Y}|S, \hat{\boldsymbol{\theta}}_K) p(S|\hat{\boldsymbol{\theta}}_K)\right) + \log p(S = \hat{\mathcal{Z}}|\mathcal{Y}, \hat{\boldsymbol{\theta}}_K). \tag{3}$$

where $S = \{S_1, ..., S_T\}$ represents all the possible hidden state sequences, T is the length of the sequence, $\hat{\boldsymbol{\theta}}_K$ are the ML (maximum likelihood) estimate of the model parameters of a HMM with K hidden states, $\hat{\mathcal{Z}}$ is the most probable state sequence (i.e. the hidden state sequence among S that best explains the observation sequence) given $\hat{\boldsymbol{\theta}}_K$ and $\mathcal{Y}$. $\hat{\boldsymbol{\theta}}_K$

can be computed using the Baum-Welch method and $\hat{\mathcal{Z}}$ can be obtained using the Viterbi algorithm (see [16] for details). We thus have:

$$\text{CL–AIC}(K) = -\log\left(\sum_{S} p(\mathcal{Y}|S,\hat{\boldsymbol{\theta}}_K)p(S|\hat{\boldsymbol{\theta}}_K)\right) - \log p(S=\hat{\mathcal{Z}}|\mathcal{Y},\hat{\boldsymbol{\theta}}_K) + C_K. \quad (4)$$

We now consider the problem of determining the unknown topology of a Dynamically Multi-Linked HMM (DML-HMM) [12] from data using CL-AIC as the scoring function. Instead of being fully connected as in the case of a Coupled HMM (CHMM) [8], a DML-HMM aims to *only* connect a subset of relevant hidden state variables across multiple temporal processes. Given a data set, we assume that at each time instance the temporal process responsible for each data sample is known and the number of hidden states for each hidden variable is also known. The unknown structure to be learned is the topology of the graph, i.e. the links among different hidden nodes within two consecutive time instances. CL-AIC can be computed using Eqn. (4) where the subscript K becomes the index of different topologies. The total number of candidate topologies K_{max} is exponential in the number of temporal processes N_t. Each candidate topology can be represented using a $N_t \times N_t$ inter-connection matrix whose elements have value 1 if there is a directed link between the corresponding two hidden nodes within two consecutive time instances and 0 otherwise.

3 Experiments

3.1 Synthetic Experiments on Learning HMM Structure

Synthetic experiments were conducted to compare the effectiveness of CL-AIC with that of BIC, AIC and ICL on determining the number of hidden states of a HMM given data of different sample sizes. One-dimensional data were first generated from a 3-state HMM (i.e. the hidden variable at each time instance can assume 3 states) whose parameters are:

$$\mathbf{A} = \begin{bmatrix} 1/3 & 1/6 & 1/2 \\ 0 & 1/3 & 2/3 \\ 1/2 & 1/2 & 0 \end{bmatrix}, \pi = \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix}, \mathbf{B} = \left\{ \begin{array}{l} \mu_1 = 1, \sigma_1^2 = 0.5 \\ \mu_2 = 3, \sigma_2^2 = 0.5 \\ \mu_3 = 5, \sigma_3^2 = 0.5 \end{array} \right\}, \quad (5)$$

where $\mathbf{A}$ is the transition matrix, π is the initial state probability and $\mathbf{B}$ contains the parameters of the emission density (Gaussians with the indicated means and variances). The total number of model parameters is 14 for this HMM. The data were then perturbed by uniformly distributed random noise with a range of $[-0.5\ \ 0.5]$. HMMs with the number of hidden states K varying from 1 to 10 were evaluated. Four different scoring functions were tested on the data sets with the sample size T varying from 25 to 4000. The results are shown in Figure 2 using the mean and ± 1 standard deviation of the selected number of hidden states over 50 trials, with each trial having a different random number seeds.

Figure 2(b) shows the mean of the number of states estimated by different scoring functions over 50 trials in a single plot. It can be seen that when the sample sizes were small, all four scoring functions tends to favour under-fitted models, with AIC and CL-AIC clearly outperforming BIC and ICL. As the sample sizes increased, the number of hidden states determined using all scoring functions converged to the true number 3. Given densely sampled data sets ($T > 400$), our results show that both AIC and BIC tended to slightly over-fit while ICL and CL-AIC yielded accurate estimation of the

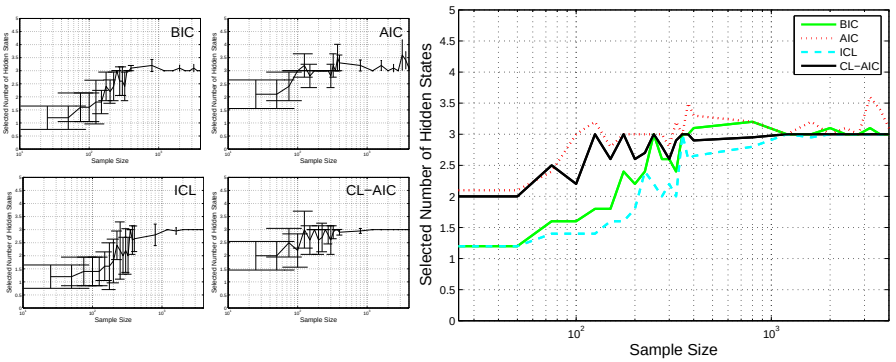


Figure 2: Synthetic data experiment results for determining the number of hidden states of a HMM using different scoring functions. (a) Selected number of hidden states (mean and ± 1 standard deviation over 50 trials); (b) Mean of selected number of hidden states (The true number of hidden states is 3).

number of hidden states. Figure 2(a) shows variations in the structure learning results across different trials, and in particular, that AIC exhibited large variations in the estimated number of states no matter what the sample size was, whilst other scoring functions had smaller variations given larger sample sizes.

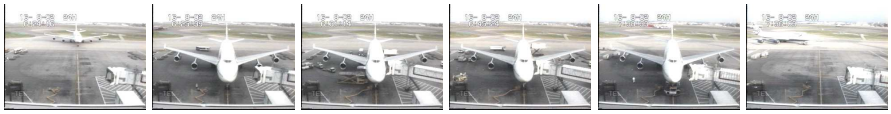
The experimental results show that the performance of CL-AIC on determining the number of hidden states for a HMM is superior to that of existing popular alternatives especially when the given dataset is sparse. Similar results were reported in the case of GMMs in [21]. However, there is a difference in the definition of 'data sparseness' for dynamic graph models and for static models such as GMMs. The sparseness of a dataset is normally measured according to the number of free parameters of a model. The experiments reported in [21] show that a sample size smaller than 5 times of the parameter number should be considered as sparse while our experiments on HMMs here show that any sample size smaller than 20 times of the true number of parameters would qualify for being sparse (see Figure 2).

3.2 Surveillance Video Segmentation

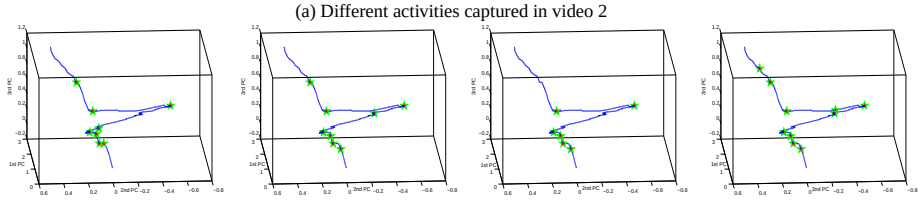
To segment a continuous surveillance video based on activities captured in the video, a L-dimensional feature vector is first extracted from each image frame. The video content is thus represented a video trajectory in this L-dimensional feature space. This feature vector is then represented as the observational variable of a HMM at each time instance. The conditional probability distributions (CPDs) of each observation variable are Gaussian for each of the K states of its parent hidden variable. The video content is then monitored using the discrete hidden variables in the model. The changes of video content can thus be detected as the changes of hidden states which correspond to breakpoints on a video trajectory (N detected change points/breakpoints result in N+1 video segments for a continuous video). Using a left-to-right HMM model, the number of hidden states would correspond to the number of video segments.

Our experiments were conducted on CCTV surveillance videos monitoring an aircraft ramp area (see Figure 3(a)). A fixed CCTV analogue camera took continuous recordings.

After digitisation, the final video sequences have a frame rate of 2Hz. Each image frame has a size of 320×240 pixels. Our database for the experiments consists of 7 sequences of aircraft docking lasting from 6470 to 17262 frames per sequence (around 50 to 140 minutes of recording), giving in total 72776 frames (10 hours) of video data. They are referred as video 1 to video 7 respectively. The 7 videos were first manually segmented into activities to give the ground truth of the breakpoints for segmentation, resulting in a total of 64 breakpoints and 71 segments. The lengths of these video segments were within the range of 127 to 3210 frames. In our experiment, a scene event based method proposed in [20] was adopted for feature extraction, which resulted in each image frames being represented as a 8 dimensional feature vector. The problem to be solved here is to automatically determine K, the number of hidden states which corresponds to the number of video segments.



(a) Different activities captured in video 2



(b) Ground truth (8 breakpoints) (c) CL-AIC (8 breakpoints) (d) BIC & ICL (6 breakpoints) (e) AIC (10 breakpoints)

Figure 3: Determining the number of video segments for an aircraft docking video (video 2 of the 7 videos) using a HMM with different score functions. (a) Representative frames of different activities captured video 2. These activities were (from left to right): 'aircraft arrival', 'airbridge connected', 'frontal cargo service', 'catering service', 'airbridge disconnected', and 'aircraft departure'. (b) Ground truth obtained by manually segmenting the video. (c)-(e) segmentation results using different scoring functions with the detected breakpoints shown on the video trajectory. Note that in (b)-(e) the video trajectories are shown in a 3D PCA space of the original 8D video content feature space just for the illustration purpose.

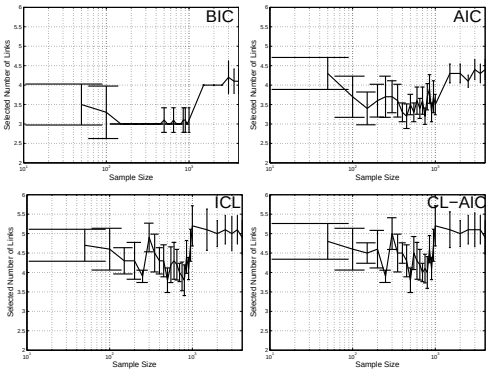
	BIC	AIC	ICL	CL-AIC
# Det. B. points	49	73	45	**62**
# True Positives	39	52	37	**54**
# False Positives	10	21	8	**8**

Table 1: Comparing scoring functions for video segmentation. True breakpoints was 64.

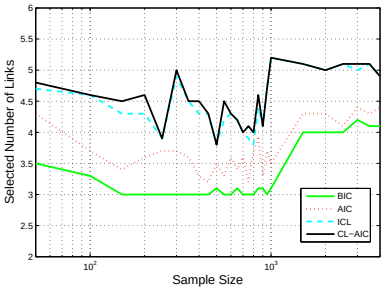
The performance of different score functions are compared by looking at the number of detected breakpoints, the number of true positives and the number of false positives. The results are shown in Table 1 and Figure 3[2]. Given the true number of breakpoints 64, it can be seen from Table 1 that both BIC and ICL underestimated the number of segments

[2]due to space limitation, only results on one of the 7 videos are shown in Figure 3

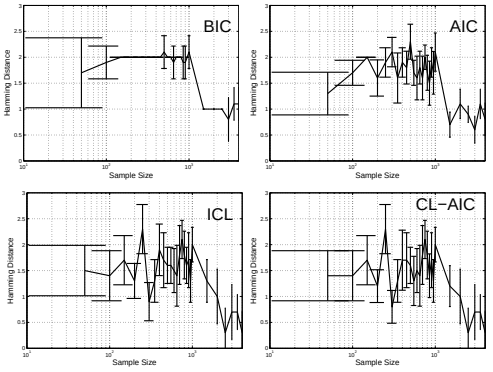
while AIC overestimated the segment number. In the meantime, the number of segments estimated using CL-AIC was the closest to the true number. On the accuracy of breakpoint detection, Table 1 shows that CL-AIC yielded the highest true positive number and lowest false positive number. In the meantime, both BIC and ICL gave low false positive number but low true positive number as well. As for AIC, high true positive number was obtained at the price of high false positive number.



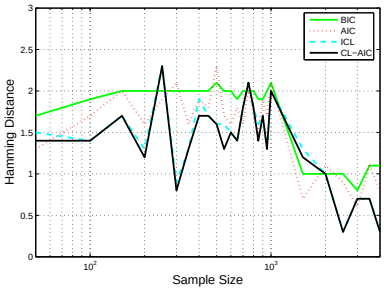
(a) Selected number of links (mean and ± 1 standard deviation over 50 trials)



(b) Mean of selected number of links



(c) Hamming distance (mean and ± 1 standard deviation over 50 trials)



(d) Mean of hamming distance

Figure 4: Synthetic data experiment results for determining the topology of a DML-HMM using different scoring functions. The true number of links is 5.

3.3 Synthetic Experiments on Learning DML-HMM Structure

Training datasets were generated using a DML-HMM with three temporal processes whose topology is shown in Figure 1(c). Both the observation and hidden variables are discrete with three possible values. One fourth of the observational data were replaced by random numbers to synthesise noise contained in the observation. The model parameters are not presented here due to the space limitation. DML-HMMs with $K_{max} = 64$ different topologies were evaluated by four different scoring functions using data sets with sample size T varying from 25 to 4000. The performance of different scoring functions was measured by looking at both the number of links connecting hidden nodes within two consecutive time instances (the true number is 5) and the hamming distance between the estimated inter-connection matrices and the true one (the distance is zero if the structure

is estimated correctly). The former measures complexity of the selected models while the latter measures the accuracy of the learned structures. The experimental results, shown in Figure 4, were obtained over 50 trials.

Figure 4 shows that given sparse data, the optimal models selected using all four different scoring functions tended to underfit with ICL and CL-AIC outperforming the other two. As the sample sizes increased, the optimal number of links among hidden nodes selected by CL-AIC and ICL converged towards the true number 5, while those selected by BIC and AIC converged to 4, (i.e. underfitting). In the meantime, the hamming distance obtained using different scoring functions decreased, with that obtained using CL-AIC being the smallest.

3.4 Discovering Causal Relationships among Visual Events

A group activity involves multiple objects co-existing and interacting in a shared common space. Examples of group activities include 'people playing football' and 'shoppers checking out at a supermarket'. Group Activity modelling is concerned with not only modelling actions executed by different objects in isolation, but also the interactions and causal/temporal relationships among these actions. Adopting a DML-HMM based activity modelling approach [12], we consider that a group activity is composed of different classes of dynamically linked visual events representing significant changes in the image over time caused by different objects in the scene. An event is represented by a multi-dimensional feature vector and automatically detected and classified into different event classes (see [12] for details). The detected events are then taken as the observational input to a DML-HMM so that learning causal and temporal relationships among different classes of events can be achieved by learning the optimal structure of the DML-HMM for modelling the dynamics of the detected events and the interactions among them. More specifically, each temporal process of the DML-HMM is used to model the dynamics of one class of events and those links among different processes capture the causal/temporal relationships of different classes of events.

A simulated 'shopping scenario' was captured on a 20 minutes video. Some typical scenes can be seen in Figure 5(a). The scene consists of a shopkeeper sitting behind a table on the right side of the view. Drink cans were laid out on a display table. Shoppers entered from the left and either browsed without paying or took a can and paid for it. The data used for this experiment were sampled at 5 frames per second with total number of 5699 frames of images sized 320×240 pixels. In the 20 minutes video, a total of 4634 events were automatically detected and classified into 5 event classes, which corresponded rather well to 5 known key constituents of the shopping activity. They were labelled as `canTaken`, `entering/leaving`, `shopkeeper`, `browsing` and `paying` respectively (see Figure 5(a)). It was noted that different classes of events occurred simultaneously. It is also true that our event recognition model made errors. Some of the errors were caused by the occlusion, closeness and visual similarity among different events. Some others were due to the factor that the causal/temporal relationships among events were not considered at the level of event detection. For example, when a shopper stands in front of the shopkeeper, it is impossible to tell whether he/she is going to pay unless one takes into consideration whether any drink can was taken a moment ago. The event classifier is therefore expected to make such errors without taking into account the temporal and causal correlations among different classes of events. Such causal/temporal relationships are modelled using a DML-HMM.

There are 5 temporal processes in this DML-HMM, each corresponding to one class

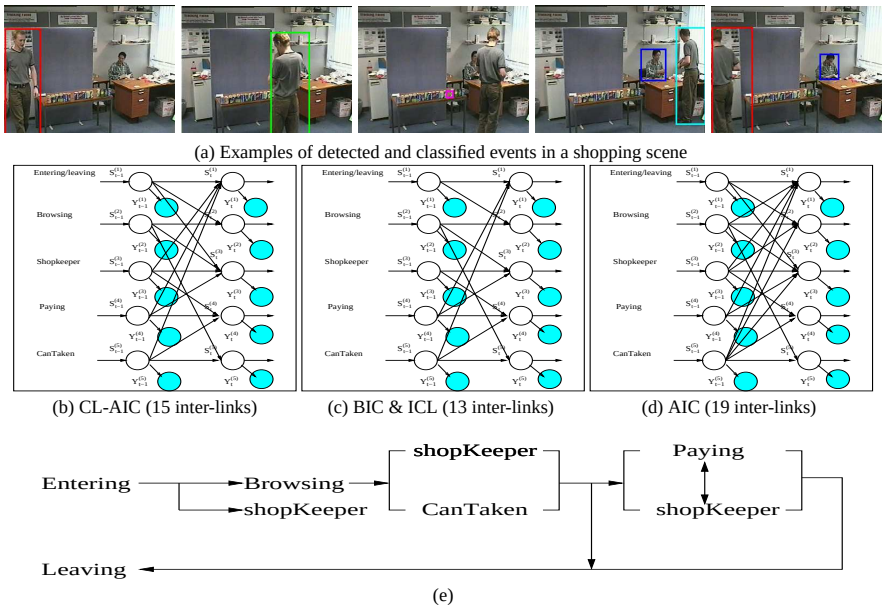


Figure 5: Discovering causal relationships among visual events in a shopping scene. In (a), events belonging to 5 event classes `canTaken`, `entering/leaving`, `shopkeeper`, `browsing` and `paying` are indicated with bounding boxes in magenta, red, blue, green and cyan respectively. (b)-(d): topologies of DML-HMMs learned using different scoring functions. (e): The expected causal and temporal structure of the shopping activity.

of events. We also consider two states for each hidden state variable, i.e. a binary variable switching between the status of `True` and `False`, corresponding to whether or not event of a certain class is truly present in each frame. Each observation variable is continuous and given by a 7-D feature vector representing a event [12]. Their distributions are mixtures of Gaussian with respect to the states of their discrete parent nodes. For model learning, the distributions of the detected events are used to initialise the distributions of the observation vectors. The priors and transition matrices of states are initialised randomly. The number of candidate topologies for a 5-temporal-process DML-HMM is too large to be searched exhaustively. The Structural EM algorithm [10] was thus adopted to search for the optimal structure more efficiently using different scoring functions.

The discovered causal/temporal relationships among different classes of events are embodied in the learned topologies of the DML-HMMs. For instance, a link pointing from the `canTaken` process towards the `paying` process indicates the causality between these two classes of events. Compared with the expected structure of the shopping activity as shown in Figure 5(e), it can be seen that the causal relationships among different classes of events and the temporal structure of the activity were discovered correctly by CL-AIC (Figure 5(b)). In comparison, a over-complicated DML-HMM topology was selected using AIC ((Figure 5(d))) while both BIC and ICL underestimated the number of inter-links among different temporal processes, resulting in over-simplified causal relationships (Figure 5(c)).

4 Conclusion

We proposed a novel scoring function (CL-AIC) for selecting the optimal structure of dynamic graph models, especially DBNs. The effectiveness of CL-AIC was demonstrated on solving challenging video content analysis problems.

References

[1] H. Akaike. Information theory and an extension of the maximum likelihood principle. In *2nd International Symposium on Information Theory*, pages 267–281, 1973.

[2] M. Beal and Z. Ghahramani. The variational bayesian em algorithm for incomplete data: with application to scoring graphical model structures. *Bayesian Statistics*, 7, 2003.

[3] C. Biernacki, G. Celeux, and G. Govaert. Assessing a mixture model for clustering with the integrated completed likelihood. *PAMI*, 22(7):719–725, 2000.

[4] J. Boreczky and L. Wilcox. A hidden markov model framework for video segmentation using audio and image features. In *International Conference on Acoustics, Speech, and Signal Processing*, pages 3741–3744, 1998.

[5] C. Boutilier, N. Friedman, M. Goldszmidt, and D. Koller. Context-specific independence in bayesian networks. In *Uncertainty in AI*, pages 115–123, 1996.

[6] M. Brand. Structure discovery in conditional probability models via an entropic prior and parameter extinction. *Neural Computation*, 11(5):1155–1182, 1999.

[7] M. Brand and V. Kettnaker. Discovery and segmentation of activities in video. *PAMI*, 22(8):844–851, August 2000.

[8] M. Brand, N. Oliver, and A. Pentland. Coupled hidden markov models for complex action recognition. In *CVPR*, pages 994–999, Puerto Rico, 1996.

[9] T. Duong, H. Bui, D. Phung, and S. Venkatesh. Activity recognition and abnormality detection with the switching hidden semi-markov model. In *CVPR*, pages 838–845, 2005.

[10] N. Friedman, K. Murphy, and S. Russell. Learning the structure of dynamic probabilistic networks. In *Uncertainty in AI*, pages 139–147, 1998.

[11] Z. Ghahramani. Learning dynamic bayesian networks. In *Adaptive Processing of Sequences and Data Structures. Lecture Notes in AI*, pages 168–197, 1998.

[12] S. Gong and T. Xiang. Recognition of group activities using dynamic probabilistic networks. In *ICCV*, pages 742–749, 2003.

[13] N. Johnson, A. Galata, and D. Hogg. The acquisition and use of interaction behaviour models. In *CVPR*, pages 866–871, 1998.

[14] R. Kass and A. Raftery. Bayes factors. *Journal of the American Statistical Association*, 90:377–395, 1995.

[15] S. Kullback. *Information theory and statistics*. Dover: New York, 1968.

[16] L.R.Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

[17] A. Raftery. Bayes model selection in social research. *Sociological Methodology*, 90:181–196, 1995.

[18] J. Rissanen. *Stochastic Complexity in Statistical Inquiry*. World Scentific, 1989.

[19] G. Schwarz. Estimating the dimension of a model. *Annals of Statistics*, 6:461–464, 1978.

[20] T. Xiang and S. Gong. Activity based video content trajectory representation and segmentation. In *BMVC*, pages 177–186, 2004.

[21] T. Xiang and S. Gong. Visual learning given sparse data of unknown complexity. In *ICCV*, pages 701–708, 2005.