Complex Activity Recognition using Granger Constrained Dynamic Bayesian Network

Eran Swears and Anthony Hoogs
Kitware Inc.
28 Corporate Drive, Clifton Park, New York 12065
[eran.swears|anthony.hoogs]@kitware.com

Many scenes in surveillance, sports, and other video domains involve complex multi-agent activities where the agents co-exist and are interacting in a time-varying manner. For example, in the surveillance domain one person may open a door of a vehicle so another person can load an object before they both enter the vehicle. Similarly, team sports involve multiple players acting in a coordinated manner. Our goal is to model and recognize such coordinated activities in video by capturing the most discriminative Granger causal relationships between pairs of time sequences extracted from event-clusters. An activity is represented as a collection of event-clusters that can be instantaneous or occur over a period of time. And, loosely speaking, Granger causality,[1], is an explicit measure of one temporal sequence's influence on another and is therefore ideal for explicitly capturing the causal relationships between agents.

The overall training approach is shown in Figure 1, where the feature data from the activity classes are automatically clustered using a hierarchical divisive clustering algorithm. Activity profiles are then extracted from each event-cluster by accumulating counts of moving-object-detections (MOD) from agents as they pass through each cluster. The pair-wise causal strength of the event-cluster activity profiles are then calculated using the continuous time version of the Granger Causality (GC) test, [2]. The GC statistic is then converted to a probability that represents the strength or weight of the GC links in order to obtain a measure that can be compared using distance metrics. The GC weights are used as features in an Adaboost temporal link selection algorithm while the activity profile is used in the Adaboost node selection algorithm. The resulting discriminative nodes and links completely define the DBN structure, where further refinement using a Structure Expectation Maximization (SEM) algorithm, [7], only decreases the classification performance. During testing unknown activities have features from their agents assigned to the event-clusters, which are then accumulated into activity profiles. Quantized versions of these activity profiles are used during inferencing as observations in the models, where the activity is classified with the label of the most likely model.

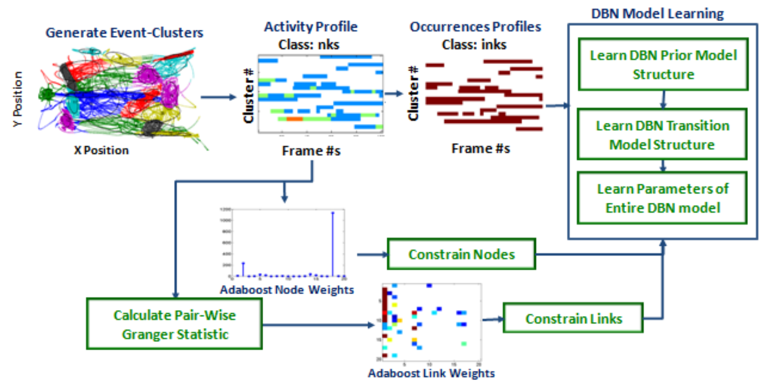

Figure 1, Overall approach flow chart for modeling complex activities using Granger Constraints

This work has several contributions that lead to more efficient and discriminative DBN models, resulting in improved classification performance. First, using the Granger statistics as a feature allows the temporal links in the DBN to be explicitly determined. Typically, score based [6,7,8,12] or constrained based [9,10,11] methods are used to determine the model structure; however, these either attempt to optimize the model fit to the training data, are independent of the classification performance, or require domain knowledge experts to define constraints. The approach used here is a constrained based approach that is automatic and unsupervised, where the constraints are learned from the data, not a domain expert.

The second contribution is that the Granger statistics are combined with an Adaboost feature selection algorithm in order to automatically define the number and type of the most discriminative nodes and temporal links in the GCDBN. This approach also removes any dependence on domain experts for applying constraints and speeds up the structure learning process by explicitly defining a smaller structure that is much closer to the desired structure.

The first well known graphical model for capturing the interactions of objects in activity recognition is the CHMM, [3]; which treats parallel and co-occurring agents as layers of Hidden Markov Networks (HMMs) that are fully coupled. The next evolutionary step is the Dynamic Multi-Linked HMM (DML-HMM), [6]. The DML-HMM is a data-driven approach for defining the temporal links between layers of HMMs using the SEM learning algorithm. This approach improves the computational burden by reducing the number of temporal links and allows the models to capture only the interactions that are represented by the current classes' data. But, as with the CHMM, it is computationally intractable for a larger number of events, forces all models to have the same number of events, and the links are independent of classification performance. On the other hand, the Time Delayed Probabilistic Graphical Model (TDPGM), [12], is the current state-of-the-art in this area and is a nice solution for automatically determining the spatial links in a graphical model based on the Time Delayed Mutual Information (TDMI) measure. The use of the TDMI in the TDPGM will detect the causal progression of a single agent between events or nodes, but does not capture the causal relationships with other co-occurring agents and is not designed to improve classification performance. The TDPGM also uses Prim's minimum spanning tree algorithm, [13], to determine the existence of a link, where the TDMI measures are the link weights. This automates the selection process; but also forces an initial tree structure onto the network while restricting the links based on its noncyclical nature.

Experiments are performed using our own synthetic data as well as real Handball data from the CVBASE06 dataset, [14]. The GCDBN is compared against a Time-Delayed DBN (TDDBN) that uses TDMI and Prim's algorithm from the TDPGM model to define the temporal links of a DBN. Results on the 5 classes in the synthetic data show that the GCDBN achieves an average probability of correct classification (Pcc), over 5-folds of a cross-validation analysis, of 73%, while the TDDBN achieves 56%. The GCDBN's higher performance can be contributed to the discriminative nature of the chosen temporal links, being able to establish causal relationships between all agents via the Granger Causality statistic, and the fact that there are no acyclic or tree requirements in the temporal link network. The 5 classes in the handball data are more separable than the classes in the synthetic data resulting in an average Pcc of 83% for the GCDBN.

References

[1] C. Granger. Investigating causal relations by econometric models and cross-spectral methods. Econometrica, 37(3):424–438, 1969.

[2] M.Ding, Y.Chen, S. Bressler, "Handbook of Time Series Analysis, Recent Theoretical Developments and Applications," chap 17. WILEY-VCH, 436:460, 2006

[3] Pradeep Natarajan, Ramakant Nevatia, "Coupled Hidden Semi Markov Models for Activity Recognition." IEEE Workshop on Motion and Video Computing, 2007.

[4] R. Li, R. Chellappa, S.K. Zhou, "Learning Multi-model Densities on Discriminative Temporal Interaction Manifold for Group Activity Recognition," Proceedings of CVPR, 2009.

[5] S. Intille, A. Bobick, "Recognizing Planned, Multiperson Action," Computer Vision and Image Understanding, 81, pp. 414-445, 2001.

[6] T. Xiang and S. Gong, "Beyond Modeling Activity and Understanding Behavior," IJCV 2006

[7] N. Friedman, "The Bayesian Structural EM algorithm," UAI, 1998

[8] G. Cooper and E. Heskovitz, "A Bayesian method for the induction of probabilistic networks from data," Machine Learning, pages 309-347, 1992.

[9] C. Campos, Z. Zheng, and Q. Ji, "Structure Learning of Bayesian Networks using Constraints," ICML 2009.

[10] Z. Zheng and Q. Ji, "Knowledge based Activity Recognition with Dynamic Bayesian Network," ECCV 2010.

[11] D. Dash and M. Druzdzel, "Robust Independence Testing for Constraint-Based Learning of Causal Structure," UAI, 2003.

[12] C.C. Loy, T. Xiang, and S. Gong, "Modeling Activity Global Temporal Dependencies using Time Delayed Probabilistic Graphical Model," ICCV, 2009

[13] R.C. Prim, "Shortest Connection Networks and Some Generalizations," Bell Sys. Tech. J., 36:1389-1401, 1957.

[14] J. Pers, M. Bon, G. Vuckovic, CVBASE 06 Dataset, available online: http://vision.fe.uni-lj.si/cvbase06/dataset.html

Topic: Visual Processing and Pattern Recognition
Preference: Poster
Presenter: Eran Swears