A Survey Paper on Learning Pullback HMM Distance for Recognition of Action

Vanita Babane[1], Poonam Sangar[2]

[1] Professor, Department of Computer Engineering, RMD Sinhgad School of Engineering, University of Pune, India

[2] Department of Computer Engineering, RMD Sinhgad School of Engineering, University of Pune, India

Abstract: ***Recent work in action recognition has exposed the limitations of features extracted from spatiotemporal video volumes. Whereas, encoding the actions dynamics using generative dynamical models has a number of attractive features, in this respect Hidden Markov models (HMMs) is a popular choice. A general framework based on pullback metrics for learning distance functions of a given training set of labeled videos has been generated, The optimal distance function is selected among a family of pullback ones, which is generated by a parameterized automorphism of the space models. An experimental result shows that how pullback learning greatly improves action recognition performances with respect to base distances.***

Keywords: Distance learning, pullback metrics, hidden Markov models (HMM), action recognition

1. Introduction

Recognizing human activities from video is a natural application of computer vision. In Action recognition capturing image sequence of one or more people performing various actions and what categories these actions belong to. Motions, however, inherently possess an extremely high degree of variability and are subject to a large number of nuisance factors such as illumination, background, viewpoint and locality.

Generative dynamical models possess a number of desirable features for action recognition. Hidden Markov models (HMM) in particular have been widely used. They are typically classified by learning a new model for each test sequence, measuring its distance from the old models and attributing to it the label of the closest model. An interesting tool is provided by pullback metrics. If the model is belonging to a Riemannian manifold M, any diffeomorphism or automorphism of M onto itself induces such a metric on M. Pullback metrics adoption has been proposed by Lebanon for document retrieval.

In this paper a general framework for learning optimal pullback distances given a training set of generative dynamical models, identified from a collection of labeled observation sequences. This framework is then applied to hidden Markov models. The tests on the KTH and YouTube data sets are conducted which demonstrate the significant improvement in action classification rates delivered by pullback learning under challenging conditions.

2. Learning Pullback Distance For Dynamical Models

Suppose a training set of N image sequences is given and feature vector is extracted from each image also an algorithm able to identify the parameters of the dynamical model (of a chosen class) which best fits a given sequences of feature vectors .

2.1 Pullback Distance Learning Framework

The general framework for learning an optimal pullback distance from a training set of dynamical models is proposed:

1) Assume that a data set Y of N observation sequences {[ys(t),t=1,…., Ts], s=1,….., N} is available;
2) From each sequence a dynamical model ms of a certain class C (e.g., a HMM) is identified, yielding a data set of models D {m1…mN};
3) Such models belong to a metric space Mc endowed with a distance function dMc;
4) a family of automorphisms from Mc onto itself, parameterized by a vector (lemda) is designed;
5) Such family of automorphisms induces a family of distance functions on Mc;
6) Optimizing over this search space of pullback distances yields an optimal distance function, which can eventually be used to classify new “test” models.

3. Pullback HMM – Action Recognition

Here proposal is validated for learning optimal pullback distances between HMMs on challenging action testbeds.

3.1 Implementation Details

3.1.1 Feature Extraction

A “sliding window” approach can be applied in which for each time instant t features are extracted from the spatio-temporal sub-volume collecting the images from t to t+T and attributed to the state Xt of the Markov chain .

3.1.2 HMM Identification

For each feature sequence so obtained, the HMM parameters were identified via the EM algorithm. As EM suffers from local minima, this algorithm is applied 10 times for each sequence and selected the parameters yielding the highest likelihood.

3.1.3 HMM Space, Number of States and State Permutation

HMMs characterized by a different number n of states live in principle on different model spaces, due to the different spaces HA of transition matrices. Here number of states are set to n=3 to make the models comparable. Three state automata have been demonstrated to represent most simple actions effectively. Even then a Markov model is uniquely defined only up to a permutation of the states. Therefore, when measuring distances between two models consideration is given to the state permutation minimizing the Frobenius distance between the respective C matrices, so identifying states associated with the same "clusters" in the observation space.

3.1.4 Classification Protocol

For both benchmarks, an optimal pullback distance is looked by maximizing the classification performance on the training set via 5-fold cross validation. Then various parameters are measured: Accuracy (Acc), calculated as the #correctly classified testing clips /#total testing clips; Average Precision (AP), which considers the ordering in which the results are presented; For classification Nearest Neighbor (1-NN) is used ; each test sequence was attributed the class of the nearest model in the training set according to the considered distance. Having fixed a classification strategy, it is possible to fairly compare pullback and base distances.

4. Conclusion

The proposed distance learning framework for linear dynamical models, based on optimizing over a family of pullback metrics induced by automorphisms. The tests were conducted on KTH and YouTube data sets which demonstrate the significant improvement in action classification rates. Also the given method able to cope with the large variety of nuisance factors such as unconstrained camera motion of handheld cameras, and cluttered scenes. The linear automorphisms maps linear boundaries to different linear boundaries. Whereas nonlinear automorphisms have the capability to extend the search space towards the unknown 'linearising' automorphisms, delivering superior performances.

References

[1] S. Ali, A. Basharat and M. Shah "Chaotic Invariants for Human Action Recognition" 11th Proc. IEEE International Conference Comp. Vision (ICCV '07).
[2] M. Blank et al., "Actions as Space-Time Shapes" 10th Proc. IEEE International Conf. Comp. Vision (ICCV '05) 2005, pp. 1395-1402.
[3] M.E.H. Benbouzid, "Bibliography on Induction Motors Faults Detection and Diagnosis" IEEE Transaction Energy Conversion, vol. 14, no. 4, Dec. 1999, pp. 1065-1074.
[4] P. Burman, "A Comparative Study of Ordinary CrossValidation, V-Fold CrossValidation and the Repeated Learning Testing Method" Biometrika, 1989, vol. 76, no. 3, pp. 503-514.
[5] R. Chaudhry et al., "Histograms of Oriented Optical Flow and Binet- Cauchy Kernels on Nonlinear Dynamical Systems for the Recognition of Human Action" Proc. IEEE Conference CVPR '09, pp. 1932-1939.
[6] F. Cuzzolin, "Manifold Learning for Multi-Dimensional Auto-Regressive Dynamical Models" Machine Learning for Vision Based Motion Analysis Springer, 2011.
[7] M. Do, "Fast Approximation of Kullback-Leibler Distance for Dependence Trees and Hidden Markov Models" IEEE Signal Processing Letters, April 2003, vol. 10, no. 4.
[8] R. Elliot,J. Moore and L. Aggoun Hidden Markov Models: Estimation and Control Springer, 1995.
[9] P. Petersen, Riemannian Geometry Springer Verlag, 1992.
[10] Q. Shi, L. Wang, L. Cheng and A. Smola "Discriminative Human Action Segmentation and Recognition Using Semi Markov Model" Proc. IEEE Conference CVPR '08, pp. 1-8, 2008.
[11] G. Lebanon "Metric Learning for Text Documents" IEEE Transaction Pattern Analysis and Machine Intelligence (PAMI), vol. 28, no. 4, pp. 497-508, Apr. 2006.
[12] R. Li and R. Chellappa, "Aligning Spatio-Temporal Signals on a Special Manifold" Proc. 11 European Conf Computer Vision (ECCV '10), 2010.

Author Profile

Vanita Babane received the B.E. and M.E degrees in Computer Engineering from MBES COE Ambajogai and VIT Pune in 2005 and 2013, respectively. Currently working as an Assistant Professor of Computer Engineering Department in RMD SSOE Pune, India

Poonam Sangar Research Scholar RMD Sinhgad School of Engineering, University of Pune. She received B.E. in Computer Engineering from All India Shri Shivaji Memorial Society Institute of Information Technology, University of Pune in 2012. Currently pursuing M.E (Comp. Engg) from RMD SSOE, Pune, University of Pune