Learning generative models for classification-recognition of human trajectories using semi-supervised EM algorithm *

Jacinto C. Nascimento
ISR/IST
jan@isr.ist.utl.pt

Mário A. T. Figueiredo
IST/IST
mario.figueiredo@lx.it.pt

Jorge S. Marques
IT/IST
jsm@isr.ist.utl.pt

Abstract

This work presents a semi-supervised EM algorithm for learning generative models for classification/recognition of human trajectories, with application to surveillance. The classifier is based on switched dynamical models, each model describing a specific motion regime. We present a semi-supervised modified version of the classical Baum-Welch algorithm, which is able to take into account a subset of known model labels. The experimental results shows the effectiveness of the present approach in both synthetic and real data. It is shown as well, that the classifier learned with semi-supervision leads to a higher classification accuracy than the fully unsupervised version. This abstract describes the work presented in [2].

(a) (b)

Figure 1. (a) Examples of activities (Browsing, leaving) unrolled in the context of our application; (b) sensors located in the scenario.

1 Introduction

The main feature of this paper is the use a bank of switched dynamical systems to describe the trajectory of a pedestrian in a video sequence. We have previously used this type of models for activity recognition in [2]. The novelty in this paper lies on the use of a semi-supervised learning framework, in which some of the model labels are observed. More specifically, we show how the classical Baum-Welch (BW), that is, the expectation-maximization (EM), algorithm for learning the parameters of a switched dynamical model can be modified to incorporate the observation of some labels.

The context of application of the present work is the recognition of typical human activities in a shopping center. The activity classes considered in this paper are *"passing"*, *"entering"* (a shop), *"exiting"* (a shop), and *"browsing"*; see Fig. 1(a) for an illustration. The trajectories in each activity class are described by a switched dynamical model of several motion regimes (such as "moving left", "stopped"). When only trajectories are observed, the parameters of these switched dynamical models are estimated by an EM algorithm, which in this case coincides with the BW algorithm. When some model labels are observed (*e.g.*, obtained manually) the EM/BW algorithm has to be modified; this modification is the main topic of this paper.

2 The Model

It is assumed that the human motion represented by the trajectory of a person mass center in the video sequence. The evolution of the mass center, $\mathbf{x}_t$ is modeled by a bank of switched dynamical models, following the equation

$$\mathbf{x}_t = \mathbf{x}_{t-1} + \mathbf{T}_{k_t} + w_t, \tag{1}$$

where $k_t \in \{1, \ldots, m\}$ is the label of the active model at time instance t; $\mathbf{T}_{k_t}$ is the mean displacement which depends on the active model; and $w_t \sim \mathcal{N}(0, \mathbf{R}_{k_t})$ is a white Gaussian noise with zero mean and covariance matrix $\mathbf{R}_{k_t}$.

We assume that the sequence of model labels $\mathbf{k} = (k_1, \ldots, k_N)$ is a sample of a Markov chain, with $(m \times m)$ transition matrix $\mathbf{B} = [B(i,j)]$ and initial distribution $\boldsymbol{\pi}$.

*This work was partially supported by EU IST Program, under project IST-2001-37540 (CAVIAR) and FCT ISR/IST plurianual funding, POSC, FEDER.

The sequence $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_N)$ is observed and $\mathbf{k} = (k_1, \ldots, k_N)$ is partially hidden: we observe the labels at some regions of the image. See Fig. 1(b) where the small areas correspond to "sensors" where we know the model labels. For each activity, the model parameters are those of an classical HMM: $\boldsymbol{\theta} = (\mathbf{T}_1, \mathbf{R}_1, \ldots, \mathbf{T}_m, \mathbf{R}_m, \mathbf{B}, \boldsymbol{\pi})$.

Finally, we consider $\mathbf{v} = (v_1, \ldots, v_N)$ the binary sequence indicating whether k_t is visible ($v_t = 1$) or hidden ($v_t = 0$).

3 Model Parameter Estimation

3.1 The EM Algorithm

To learn the parameters of the model presented in the previous section, we use a Baum-Welch-type algorithm [3]. Just as the standard BW algorithm is an instance of EM [1] to estimate the parameter of an HMM (which the switched dynamical model is), the algorithm herein presented is the EM algorithm for the semi-supervised learning setting above described.

The standard BW algorithm, which assumes that all elements of the model labels are hidden, defines a set of "weights" $w_{t\ i}$, where $w_{t\ i}$, that is, the current estimate of the probability that at time t of the sequence, the active model is i. Similarly, the BW algorithm also defines transition weights $w_{t\ ij}$. These weights are the only information which is needed to compute the Q-function.

In our scenario, we assume that if $v_t = 1$, then k_t is not hidden, but an observed label. This requires defining new modified "weights" $\overline{w}$ as follows

$$\overline{w}_{t\ i} = \begin{cases} w_{t\ i} & \Leftarrow \quad v_t = 0 \\ \delta(i - k_t) & \Leftarrow \quad v_t = 1, \end{cases} \qquad (2)$$

where is δ is the Kronecker delta function, i.e., $\delta(a-b) = 1$, if $a = b$, and zero otherwise. Notice that if $v_t = 1$, then k_t is an observed variable. Similarly,

$$\overline{w}_{t\ ij} = \begin{cases} w_{t\ ij} & \Leftarrow \ (v_{t-1} = 0) \wedge (v_t = 0) \\ \delta(i - k_{t-1})\delta(j - k_t) & \Leftarrow \ (v_{t-1} = 1) \wedge (v_t = 1) \\ \langle \delta(i - k_{t-1}) w_{t\ ij} \rangle_j & \Leftarrow \ (v_{t-1} = 1) \wedge (v_t = 0) \\ \langle \delta(j - k_t) w_{t\ ij} \rangle_i & \Leftarrow \ (v_{t-1} = 0) \wedge (v_t = 1), \end{cases}$$

In the M-step the Q-function is simply obtained as in the standard BW algorithm, but using the semi-supervised weights $\overline{w}_{t\ i}$ and $\overline{w}_{t\ ij}$ instead of $w_{t\ i}$ and $w_{t\ ij}$. The parameter estimates are updated according to standard rules of the BW algorithm (see [3] for full details).

4 Experimental Results

The proposed approach was also tested in real data collected in the context of CAVIAR project: 64 video sequences hand labeled with the ground truth. These sequences include indoor shopping center observations of individual and groups of pedestrians.

Fig. 2 shows several real trajectories, i.e, evolution of the centroid of the bounding box, as well as the corresponding activity classifier output. To assess the classification accuracy, we have classified 51 trajectories, using the MAP criterion. Using the parameter estimates obtained with semi-supervision, the accuracy obtained was was 90.0%.

5 Conclusions

In this work we have presented a semi-supervised framework for modeling and recognition of human trajectories was presented. The activities considered here are: *"passing"*, *"entering"* (a shop), *"exiting"* (a shop), and *"browsing"*. We have shown how to modify the classical Baum-Welch (BW) algorithm to take into account a subset of known model labels, leading to a semi-supervised BW algorithm. The experimental results illustrates the effectiveness of the present approach.

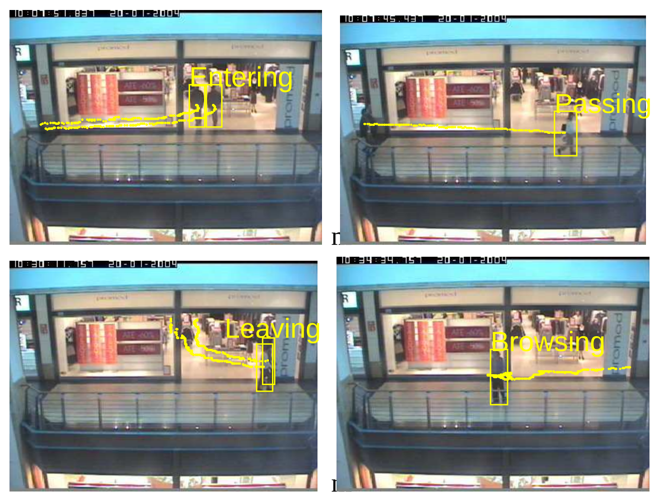

Figure 2. Several synthetic activities: (a) entering, (b) leaving, (c) passing, (d) browsing.

References

[1] A. Dempster, N. Laird, and D. Rubin. Maximum likelihood estimation from incomplete data via the EM algorithm. *Jour. Royal Statist. Soc. (B)*, 39:1–38, 1977.

[2] J. C. Nascimento, M. A. T. Figueiredo, and J. S. Marques. Segmentation and classification of human activities. In *British Machine Vision Conf., Int. Workshop on Human Activity Recognition and Modelling*.

[3] L. R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. of the IEEE*.