Hidden CRFs for Human Activity Classification from RGBD Data

Avi Singh & Ankit Goyal
IIT-Kanpur

CS679: Machine Learning for Computer Vision

April 13, 2015

Overview

Problem Statement

Input

An RGBD video with a human subject performing some day-to-day activity like drinking water.

Output

A classification label as to what that activity is.

Application: Assistive Robotics

Popular 3D Features for human activity recognition

- 3D silhouettes
- Skeletal joints or body part tracking
- Local Spatio-temporal features
- Local 3D occupancy features
- 3D optical flow

Models for Structured Prediction

- Hidden Markov Models (HMM)
- Maximum Entropy Markov Models (MEMM)
- Conditional Random Fields (CRF)

Hidden Markov Model (HMM)

- Generative model
- Predicts the (hidden) state of a system from the visible output
- Have been traditionally used in temporal pattern recognition such as speech recognition

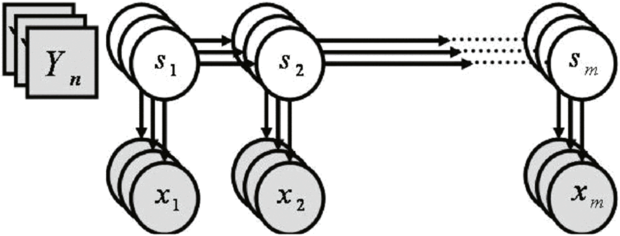


Figure: Hidden Markov Model

Problems with HMM

- Requires enumeration of all possible observation sequences.
- Requires the observations to be independent of each other.
- Generative approach for solving a conditional problem leading to unnecessary computations.

Maximum Entropy Markov Model (MEMM)

- Conditional model.
- Unlike HMM, uses a set of overlapping features.
- Uses the maximum entropy framework to fit a set of exponential models that represent the probability of a state given an observation and the previous state.
- More efficient training algorithms than HMM or CRF

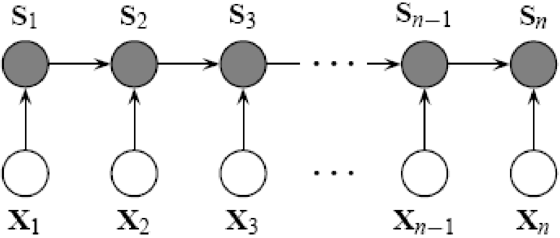


Figure: Hidden Markov Model

Problems with MEMM

Label Bias Problem

- States with low-entropy transition distributions "effectively ignore" their observations. States with lower transitions have "unfair advantage",
- Since training is always done with respect to known previous tags, so the model struggles at test time when there is uncertainty in the previous tag.

Conditional Random Fields (CRF)

Activity Classification

Problem
Features
Models
HMM
MEMM
CRF
HCRF
Our Solution
Features
Modeling
Results
Future Scope
References

- Conditional Models.
- Uses a single exponential model for joint probability of entire sequence of labels given the observation sequence.
- Avoids the limitations of MEMM.

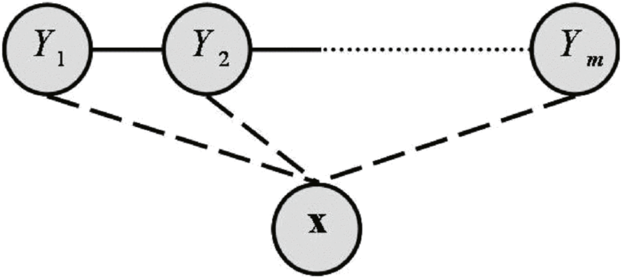


Figure: Conditional Random Field Model

Activity Classification

Problem

Features

Models

HMM

MEMM

CRF

HCRF

Our Solution

Features

Modeling

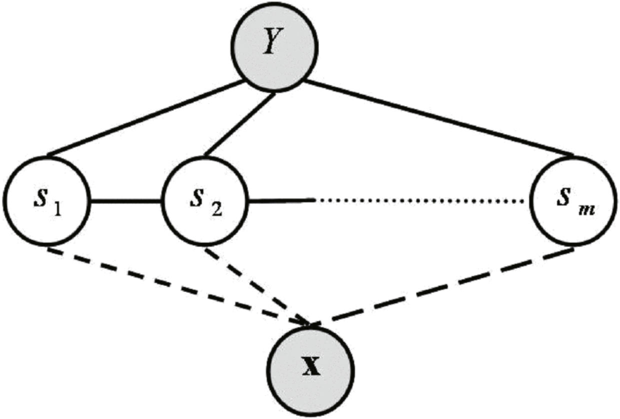
Results

Future Scope

References

Hidden Conditional Random Fields

- An extension of CRF.
- Discriminative latent variable model for classifying of a sequence of observations.



Hidden Conditional Random Fields

- Aim : To predict the label y from $\mathbf{x}$. Each y is a member of a set Y of possible labels and each vector $\mathbf{x}$ is a vector of local observations $\mathbf{x} = \{x_1, x_2, \ldots, x_m\}$. And $\mathbf{h} = \{h_1, h_2, \ldots, h_m\}$ is the vector of latent states.
- A conditional probabilistic model with θ as parameter is defined:

$$P(y, \mathbf{h} \mid \mathbf{x}, \theta) = \frac{e^{\Psi(y,\mathbf{h},\mathbf{x};\theta)}}{\sum_{y',\mathbf{h}} e^{\Psi(y',\mathbf{h},\mathbf{x};\theta)}}, \qquad (1)$$

- This model gives $P(y \mid \mathbf{x}, \theta)$:

$$P(y \mid \mathbf{x}, \theta) = \sum_{\mathbf{h}} P(y, \mathbf{h} \mid \mathbf{x}, \theta) = \frac{\sum_{\mathbf{h}} e^{\Psi(y,\mathbf{h},\mathbf{x};\theta)}}{\sum_{y',\mathbf{h}} e^{\Psi(y',\mathbf{h},\mathbf{x};\theta)}}. \qquad (2)$$

Hidden Conditional Random Fields

- The objective function while training is chosen as:

$$L(\theta) = \sum_i \log P(y_i \mid \mathbf{x}_i, \theta) - \frac{1}{2\sigma^2}\|\theta\|^2. \qquad (3)$$

- Ψ is the potential function and takes the following form:

$$\Psi(y, \mathbf{h}, \mathbf{x}; \theta) = \sum_j \phi(x_j) \cdot \theta(h_j) + \sum_j \theta(y, h_j) + \sum_{(j,k)\in E} \theta(y, h_j, h_k), \qquad (4)$$

- $\theta^* = \arg\max L(\theta)$ is calculated from the training example using quasi-Newton gradient descent.
- Exact methods for the inference and parameter estimation are tractable.

Features

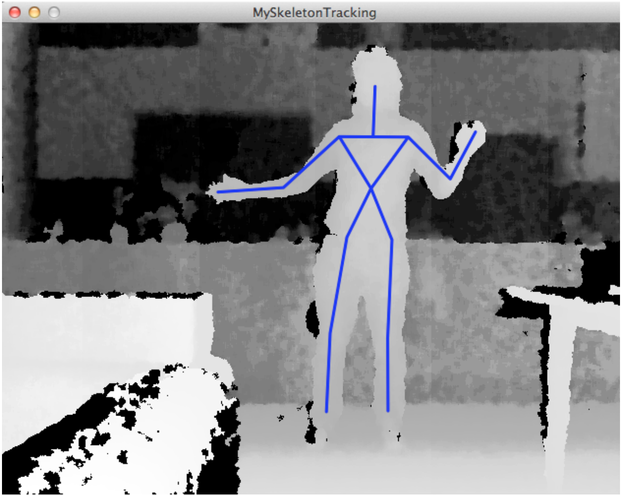


Figure: Joint Coordinates from Single Depth Image

Preprocessing

- X,Y,Z coordinates of 20 body joints extracted from depth image
- Coordinates originally in Kinect frame transformed to the person's frame of reference
- A normalization technique used to account for changes in body part size
- The motion information is captured by calculating the difference in coordinates between successive frames.

Our Model

We used Hidden-state Conditional Random Fields. Some hyperparameters:

- Number of hidden states: 10
- Window Size: 1
- Optimizer: Broyden Fletcher Goldfarb Shanno algorithm

(MSR Daily Activity 3D dataset)

- 10 subjects
- 16 activities: drink, eat, read book, call cellphone, write on a paper, use laptop, use vacuum cleaner, cheer up, sit still, toss paper, play game, lie down on sofa, walk, play guitar, stand up, sit down
- Each subject performs each activity twice (sitting/standing)
- 10x16x2 = 320 RGBD videos + skeleton data

Reduced Dataset

Due to computational restrictions, we used a subset of the original dataset.

We used 6 activities, 10 subjects, and one video for every subject (sitting position)

Training/Testing

We did 5-fold cross validation.
48/12 split for training/testing

All the testing was performed in "new person" setting.

Final Accuracy Obtained: 71.67%

Confusion Matrix

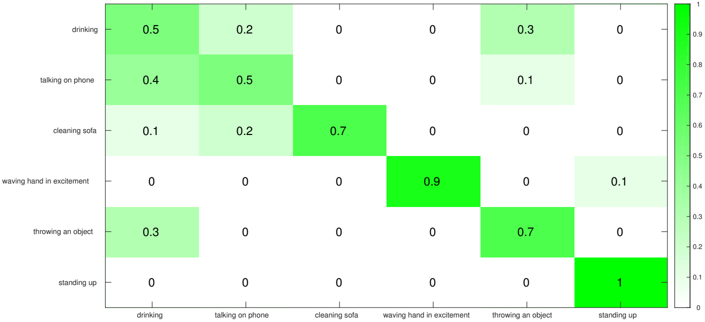


Figure: The Confusion Matrix

Future Scope

- Including more features from the skeleton data, especially those having nonlinear relation to joint coordinates.
- Including features to model interaction between different objects in the scene and the human.
- Bayesian optimization for the hyper-parameters like number of hidden sates.
- Exploring other models like Structured SVM.

References

A. Quattoni, S. Wang, L.-P Morency, M. Collins, T. Darrell;
Hidden-state Conditional Random Fields
PAMI 2005

Sy Bor Wang, Ariadna Quattoni, Louis-Philippe Morency, David Demirdjian, Trevor Darrell
Hidden Conditional Random Fields for Gesture Recognition
CVPR 2006

John Lafferty, Andrew McCallum, Fernando C.N. Pereira
Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data
ICML 2001

Andrew McCallum, Dayne Freitag, Fernando C.N. Pereira
Maximum Entropy Markov Models for Information Extraction and Segmentation
ICML 2000

References

Lawrence M. Rabiner
A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition

Jaeyong Sung, Colin Ponce, Bart Selman and Ashutosh Saxena
Unstructured Human Activity Detection from RGBD Images
ICRA 2012

Jamie Shotton, Andrew Fitzgibbon, Mat Cook, Toby Sharp, Mark Finocchio, Richard Moore, Alex Kipman, Andrew Blake
Real-Time Human Pose Recognition in Parts from Single Depth Images
CVPR 2011

J.K. Aggarwal, Lu Xia
Human activity recognition from 3D data: A review
Pattern Recognition Letters 2014

The End