Efficient Multi-Instance Learning for Activity Recognition from Time Series Data Using an Auto-Regressive Hidden Markov Model

Xinze Guan, Raviv Raich, Weng-Keen Wong
School of EECS, Oregon State University
Email: {guan, raich, wongwe}@eecs.oregonstate.edu
Introduction

- Wearable sensors are everywhere
- Record human motion as a multivariate time series
Introduction

• Goal: physical activity recognition

From the Opportunity dataset (Chavarriaga et al. 2013)
Introduction

Physical activity recognition important for:

- Elder care
- Assistance with cognitive disabilities
- Health surveillance and research
Introduction

- High annotation effort to label training data

[Graph showing activity patterns with events such as 'Drink from cup', 'Open Fridge', 'Open Drawer', 'Close Drawer', 'Open Drawer', 'Close Drawer']
Introduction

• Stikic et al. (2011) proposed a weakly supervised approach based on multi-instance learning
• Trades off the ease of labeling with ambiguity in the labeling
• Our work builds on their approach
Methodology: MIL

Multi-instance Learning (Dietterich et al. 1997):

- Data made up of bags of instances
- Bags can be labeled **positive** or **negative**
Methodology: MIL for Time Series

Majority Labeling Scheme:
Bag labeled + if the majority of the time ticks belong to the activity of interest (eg. “Drink from Cup”)

Drink from cup

Open Fridge

Open Drawer

Close Drawer

Open Drawer

Close Drawer

Bag (+)

Bag (-)
Related Work

Structured MIL

- Relationship between instances in different bags (Deselaers and Ferrari 2010)
- Relationship between bags (Zhang et al. 2011)

Our work: models temporal dynamics between instances in a bag
Methodology: The Model
Methodology: The Model
Methodology: The Model
Methodology: The Model

\[ Y_b = MIL_{Softmax}(I_b^1, ..., I_b^{T_b}) \]
Methodology: The Model

\[ \text{Methodology: The Model} \]

\[ b=1: B \]

\[ Y_b \]

\[ I_b^2 \]

\[ I_b^3 \]

\[ \ldots \]

\[ I_b^{T_b} \]

\[ Z_b^2 \]

\[ Z_b^3 \]

\[ \ldots \]

\[ Z_b^{T_b} \]

\[ X_b^1 \]

\[ X_b^2 \]

\[ X_b^3 \]

\[ \ldots \]

\[ X_b^{T_b} \]
Methodology: The Model

\[ \omega \]

\[ \beta \]

\[ M \]

\[ \pi \]

\[ k=1:K \]

\[ A_k \]

\[ b=1:B \]

\[ Y_b \]

\[ k=1:K \]

\[ \theta_k \]

\[ \zeta \]

\[ \sum_k \]

\[ \Omega \]

\[ \phi \]

\[ \Theta \]

\[ \Psi \]

\[ \Lambda \]

\[ \Xi \]

\[ \Delta \]

\[ \Psi \]

\[ \Phi \]

\[ \Theta \]

\[ \Lambda \]

\[ \Xi \]

\[ \Delta \]

\[ \Psi \]

\[ \Phi \]

\[ \Theta \]

\[ \Lambda \]

\[ \Xi \]

\[ \Delta \]

\[ \Psi \]

\[ \Phi \]

\[ \Theta \]

\[ \Lambda \]

\[ \Xi \]

\[ \Delta \]

\[ \Psi \]

\[ \Phi \]

\[ \Theta \]

\[ \Lambda \]

\[ \Xi \]

\[ \Delta \]

\[ \Psi \]

\[ \Phi \]

\[ \Theta \]

\[ \Lambda \]

\[ \Xi \]

\[ \Delta \]

\[ \Psi \]

\[ \Phi \]

\[ \Theta \]

\[ \Lambda \]

\[ \Xi \]

\[ \Delta \]

\[ \Psi \]

\[ \Phi \]

\[ \Theta \]

\[ \Lambda \]

\[ \Xi \]

\[ \Delta \]

\[ \Psi \]

\[ \Phi \]

\[ \Theta \]

\[ \Lambda \]

\[ \Xi \]

\[ \Delta \]

\[ \Psi \]

\[ \Phi \]

\[ \Theta \]

\[ \Lambda \]

\[ \Xi \]

\[ \Delta \]

\[ \Psi \]

\[ \Phi \]

\[ \Theta \]

\[ \Lambda \]

\[ \Xi \]

\[ \Delta \]

\[ \Psi \]

\[ \Phi \]

\[ \Theta \]

\[ \Lambda \]

\[ \Xi \]

\[ \Delta \]

\[ \Psi \]

\[ \Phi \]

\[ \Theta \]

\[ \Lambda \]

\[ \Xi \]

\[ \Delta \]

\[ \Psi \]

\[ \Phi \]

\[ \Theta \]

\[ \Lambda \]

\[ \Xi \]

\[ \Delta \]

\[ \Psi \]

\[ \Phi \]

\[ \Theta \]

\[ \Lambda \]

\[ \Xi \]

\[ \Delta \]

\[ \Psi \]

\[ \Phi \]

\[ \Theta \]

\[ \Lambda \]

\[ \Xi \]

\[ \Delta \]

\[ \Psi \]

\[ \Phi \]

\[ \Theta \]

\[ \Lambda \]

\[ \Xi \]

\[ \Delta \]

\[ \Psi \]

\[ \Phi \]

\[ \Theta \]

\[ \Lambda \]

\[ \Xi \]

\[ \Delta \]

\[ \Psi \]

\[ \Phi \]

\[ \Theta \]

\[ \Lambda \]

\[ \Xi \]

\[ \Delta \]

\[ \Psi \]

\[ \Phi \]

\[ \Theta \]

\[ \Lambda \]

\[ \Xi \]

\[ \Delta \]

\[ \Psi \]

\[ \Phi \]

\[ \Theta \]

\[ \Lambda \]

\[ \Xi \]

\[ \Delta \]

\[ \Psi \]

\[ \Phi \]

\[ \Theta \]

\[ \Lambda \]

\[ \Xi \]

\[ \Delta \]

\[ \Psi \]

\[ \Phi \]

\[ \Theta \]

\[ \Lambda \]

\[ \Xi \]

\[ \Delta \]

\[ \Psi \]

\[ \Phi \]

\[ \Theta \]

\[ \Lambda \]

\[ \Xi \]

\[ \Delta \]

\[ \Psi \]

\[ \Phi \]

\[ \Theta \]

\[ \Lambda \]

\[ \Xi \]

\[ \Delta \]

\[ \Psi \]

\[ \Phi \]

\[ \Theta \]

\[ \Lambda \]

\[ \Xi \]

\[ \Delta \]

\[ \Psi \]

\[ \Phi \]

\[ \Theta \]

\[ \Lambda \]

\[ \Xi \]

\[ \Delta \]

\[ \Psi \]

\[ \Phi \]

\[ \Theta \]

\[ \Lambda \]

\[ \Xi \]

\[ \Delta \]

\[ \Psi \]

\[ \Phi \]

\[ \Theta \]

\[ \Lambda \]

\[ \Xi \]

\[ \Delta \]

\[ \Psi \]

\[ \Phi \]

\[ \Theta \]

\[ \Lambda \]

\[ \Xi \]

\[ \Delta \]

\[ \Psi \]

\[ \Phi \]

\[ \Theta \]

\[ \Lambda \]

\[ \Xi \]

\[ \Delta \]

\[ \Psi \]

\[ \Phi \]

\[ \Theta \]

\[ \Lambda \]

\[ \Xi \]

\[ \Delta \]

\[ \Psi \]

\[ \Phi \]

\[ \Theta \]

\[ \Lambda \]

\[ \Xi \]

\[ \Delta \]

\[ \Psi \]

\[ \Phi \]

\[ \Theta \]

\[ \Lambda \]

\[ \Xi \]

\[ \Delta \]

\[ \Psi \]

\[ \Phi \]

\[ \Theta \]

\[ \Lambda \]

\[ \Xi \]

\[ \Delta \]

\[ \Psi \]

\[ \Phi \]

\[ \Theta \]

\[ \Lambda \]

\[ \Xi \]

\[ \Delta \]

\[ \Psi \]

\[ \Phi \]

\[ \Theta \]

\[ \Lambda \]

\[ \Xi \]

\[ \Delta \]

\[ \Psi \]

\[ \Phi \]

\[ \Theta \]

\[ \Lambda \]

\[ \Xi \]

\[ \Delta \]

\[ \Psi \]

\[ \Phi \]

\[ \Theta \]

\[ \Lambda \]

\[ \Xi \]

\[ \Delta \]

\[ \Psi \]

\[ \Phi \]

\[ \Theta \]

\[ \Lambda \]

\[ \Xi \]

\[ \Delta \]

\[ \Psi \]

\[ \Phi \]

\[ \Theta \]

\[ \Lambda \]

\[ \Xi \]

\[ \Delta \]

\[ \Psi \]

\[ \Phi \]

\[ \Theta \]

\[ \Lambda \]

\[ \Xi \]

\[ \Delta \]

\[ \Psi \]

\[ \Phi \]

\[ \Theta \]

\[ \Lambda \]

\[ \Xi \]

\[ \Delta \]

\[ \Psi \]

\[ \Phi \]

\[ \Theta \]

\[ \Lambda \]

\[ \Xi \]

\[ \Delta \]

\[ \Psi \]

\[ \Phi \]

\[ \Theta \]

\[ \Lambda \]

\[ \Xi \]

\[ \Delta \]

\[ \Psi \]

\[ \Phi \]

\[ \Theta \]

\[ \Lambda \]

\[ \Xi \]

\[ \Delta \]

\[ \Psi \]

\[ \Phi \]

\[ \Theta \]

\[ \Lambda \]

\[ \Xi \]

\[ \Delta \]

\[ \Psi \]

\[ \Phi \]

\[ \Theta \]
Methodology: Parameter Estimation

Expectation-Maximization:

1. M-step:
   – Straightforward

2. E-step:
   – Requires computation of
     \[ P(I_b^t, Z_b^t, I_b^{t-1}, Z_b^{t-1} | X_b, Y_b) \]
   – If done naively: \( O(2^{T_b} k^{T_b}) \)
Methodology: Efficient Message Passing

\[ P(Y_b = 1 | I^1_b, ..., I^{T_b}_b) = \frac{\text{(# positive instances)} \times \exp(\omega)}{\text{(# positive instances)} \times \exp(\omega) + \text{(# negative instances)}} \]
Methodology: Efficient Message Passing

Replace $I_b^t$ with a counting variable $N_b^t$
Methodology: Efficient Message Passing
Methodology: Efficient Message Passing

- Replace the $N^t_b, Z^t_b$ nodes with a super-node $S^t_b = (N^t_b, Z^t_b)$
- Becomes an Auto-regressive Hidden Markov Model
Methodology: Efficient Message Passing

• Apply standard forward-backward message passing for ARHMM
• But can exploit a sparse transition matrix
• E-step computation is now $O(K^2T_b^2)$
Results: Algorithms

Using features from Stikic et al. (2011)

- miSVM (Andrew et al. 2003)
- DPMIL (Kandemir and Hamprecht 2014)
- miGraph (Zhou et al. 2009)

Using the raw time series:

- MARMIL (our NIPS workshop paper)
- ARHMM-MIL (ours)
Results: Experimental Setup

Datasets:

• Opportunity (Chavarriaga et al. 2013)
• Trainspotting1 (Berlin and Laerhoven 2012)
• Trainspotting2 (Berlin and Laerhoven 2012)
Results
Conclusion

• ARHMM-MIL models temporal dynamics between instances in a bag

• Generative model that can:
  – Predict bag and instance labels
  – Allow deeper analysis of data by decomposing it into AR processes
  – Allow you to sample data from it
Future Work

Multi-Instance Multi-Label Approach
Thank you!

This work is partially supported by NSF grant CCF-1254218

Poster Session: Tues Morning
Questions?