Enabling Active Learning for Real time Activity Recognition

Tianyi Wu[1], Anahita Hosseini[2]
Christine King, PhD[2], Majid Sarrafzadeh, PhD[2]
[1]Dept. Computer Science, University of California, Los Angeles, USA
[2]Dept. Electrical Engineering, Xi'An Jiaotong University, China

Introduction

- Real-time personalized activity recognition system requires an online learning model with user adaptation in a streaming environment [1].
- While manually labeling stream of data is time consuming, semi-supervised learning helps reduce the work load by classifying with unlabeled data [2].
- Active learning aims at querying users for the least amount of labels when classification is most uncertain to refine class boundaries and adapt to new changes [3].

Method

- Offline Model Training
    - Graph-based semi-supervised learning
    - K-means cluster based representation
    - Random forest for similarity comparison

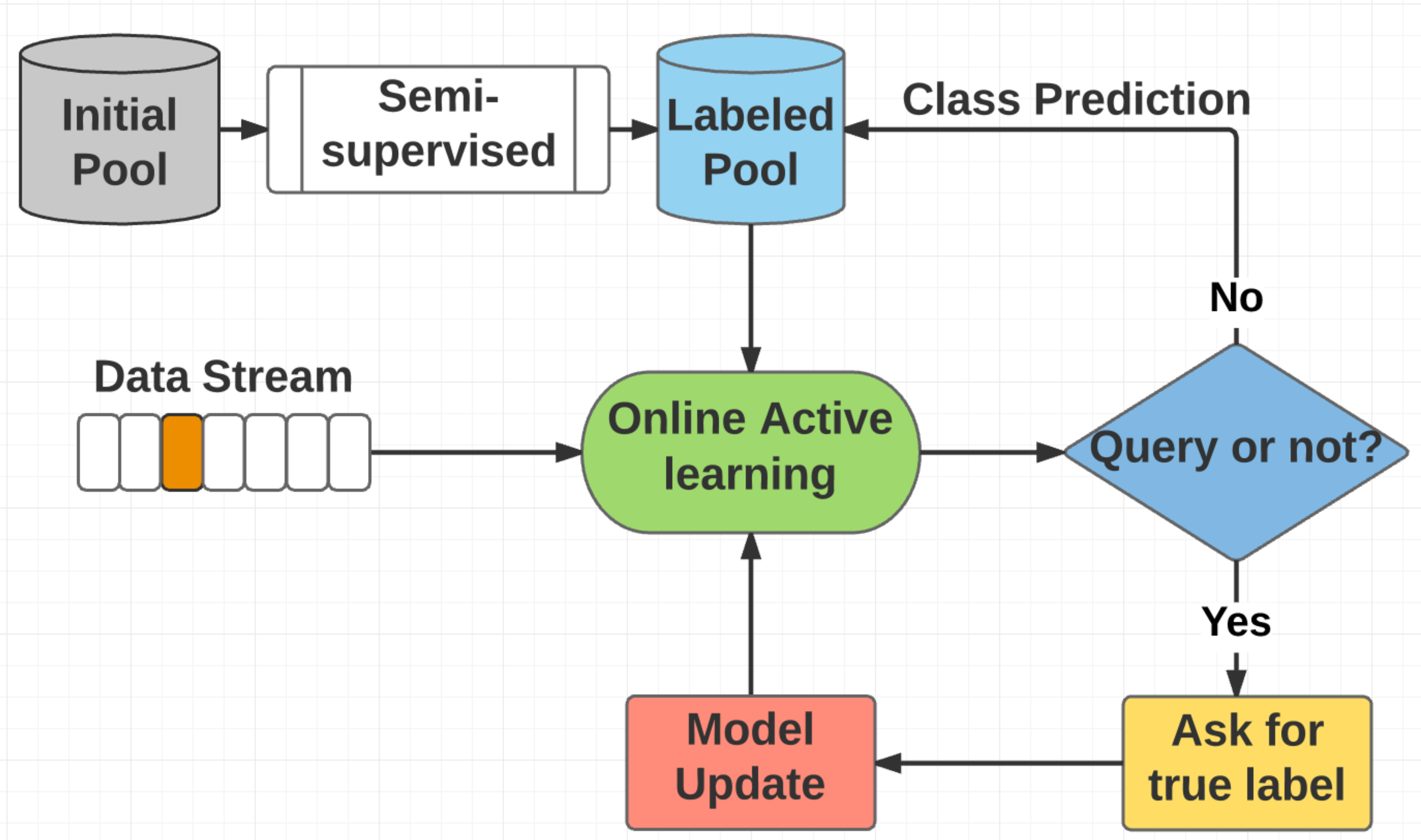

Fig 1. Stream-based Active Learning System Design

- Online Active Learning Query Strategy
    - Max-min threshold
    - Disagreement between classifiers
    - Time based dissimilarity

Algorithm Design

Algorithm 1: Random Forest Similarity Measurement
Input: New data **x**, old data **y**, Training Model **RF**
Initialization: count = 0
For each tree in **RF** estimators :
  if tree.predict(**x**) = =tree.predict(**y**) **then** :
    count = count +1
end for
  return count / total number of estimators

Results

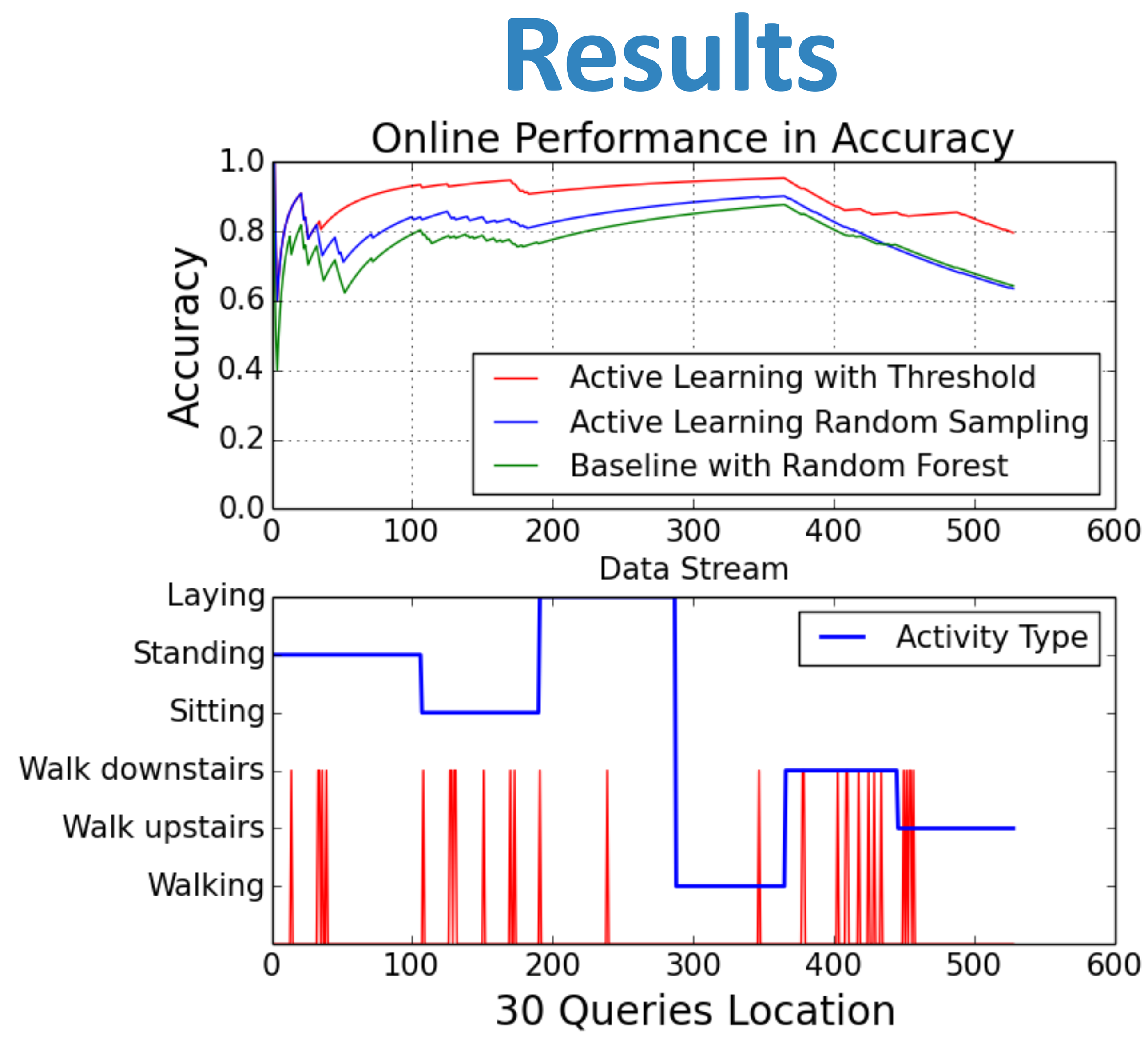

Fig 2. Real time Simulation for One User in HAPT Dataset[4]

Offline Training Samples	Active learning Accuracy	Random Selection Accuracy	Online Response Time(s)
50	0.7885	0.7207	0.332

Table 1. Average Online Performance across 10 Users

Conclusion

Our proposed active learning system with a multi-threshold query strategy further improves the accuracy for activity recognition by 6.8% compared to baseline random selection. Combining stream-based active learning with offline semi-supervised learning and a random forest classifier can efficiently reduce the labeling costs and improve validation for uncertain activities. Future work will focus on testing the framework with smartwatches to assess the feasibility of active learning and apply user adaptation over time.

Acknowledgement

This study was funded by the National Institute of the Biomedical Imaging and Bioengineering (NIBIB) Los Angeles Pediatric Research using Integrated Sensor Monitoring Systems (PRISMS) Center: The Biomedical REAl-Time Health Evaluation (BREATHE) Platform, NIH/NIBIB U54 award EB022002.

References

[1] N.C. Krishnan, D.J. Cook, Activity recognition on streaming sensor data, *Pervasive Mob. Comput.* 10 (Feb) (2014), 138–154.
[2] Germain Forestier , Semi-supervised learning using multiple clusterings with limited labeled data , *Information Science*, 361-362 (2016), 48-65.
[3] Sicheng Xiong, Javad Azimi, and Xiaoli Z. Fern, Active Learning of Constraints for Semi-Supervised Clustering, *IEEE Transactions on Knowledge and Data Engineering*, VOL. 26, NO. 1 (2015), 43-54
[4] Lichman, M. (2013). UCI Machine Learning Repository. [http://archive.ics.uci.edu/ml] Irvine, CA: University of California, School of Information and Computer Science.