# Activity Recognition
## Through Active and Semi-Supervised Learning in Mobile Devices

By Matt Barren

# Activity Recognition

- Utilizing sensors to determine user's current state
  - Accelerometers, heart rate, etc.
- Perform appropriate actions based on the current activity state

# Typical Activity Application Development

- Training based upon an initial dataset of users
  - Calibration
- Classify activities with a static classifier

***Limitations***

- Not unique to user
- No dynamism
- Potentially requires a large set of labeled data

# What can be gained?

**Ideally - Classification is unique and responsive to user data**

- **Classification unique to user**
- **Increase accuracy of classifier through updating**
- New classifiers based on user
- Increased efficacy in metadata

**Opportunities due to..**

- Variety of sensors in mobile devices
- Computational power in mobile devices

# Overview of Discussion

**Active and Semi-Supervised Learning**

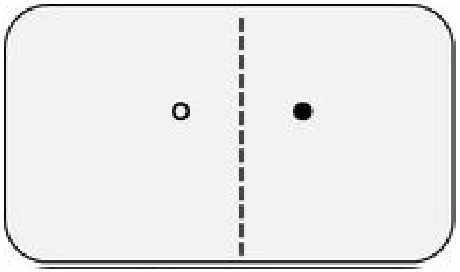
- Overview of Semi-Supervised Learning and General Learning Model
- Semi-Supervised Learning Models
- Experimentation and Evaluation
- Results
- Conclusions of Work

- Briefly Examine Nonparametric Discovery of Human Routines
- Conclusion

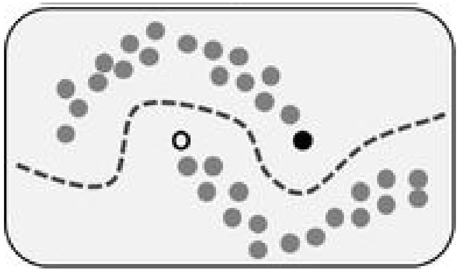
# Semi-Supervised Learning Models

- What is meant by “semi”-supervised?
  - Train with small quantities of labeled and large quantities of unlabeled
- How to select unlabeled data to use?
  - “Confidence!”
- Why use semi-supervised learning?
  - Infeasibility of labeling
  - Iterative learning

Labeled Data with two classes

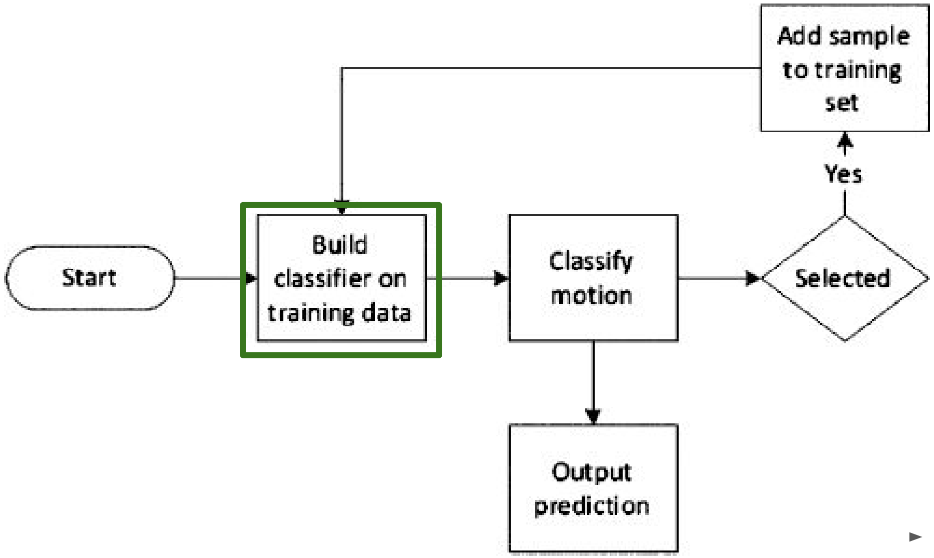


Unlabeled Data with two classes



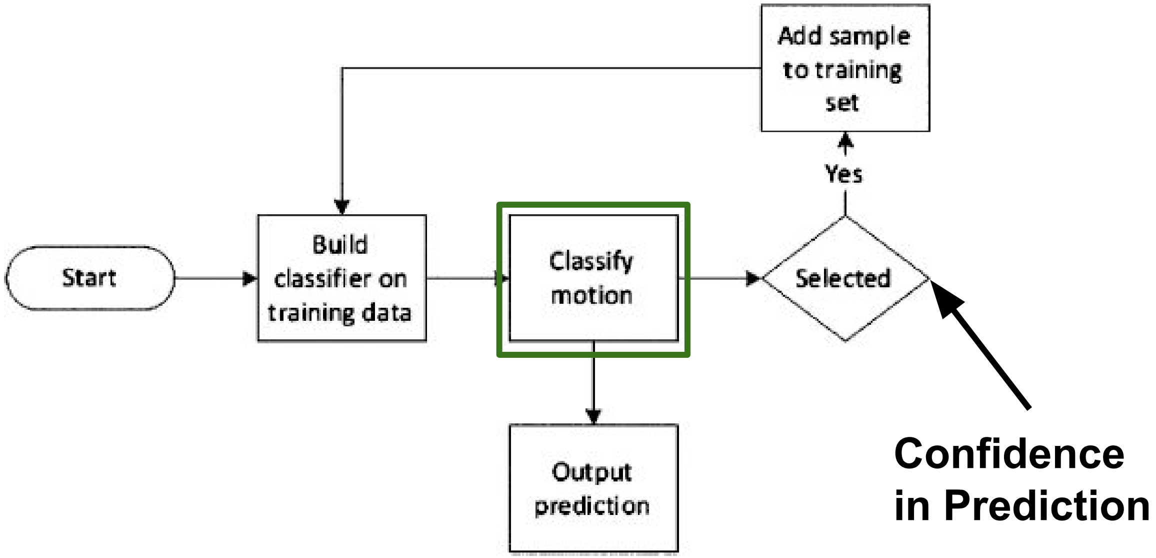
# General Learning Model

- Initial training set to build a classifier(s)



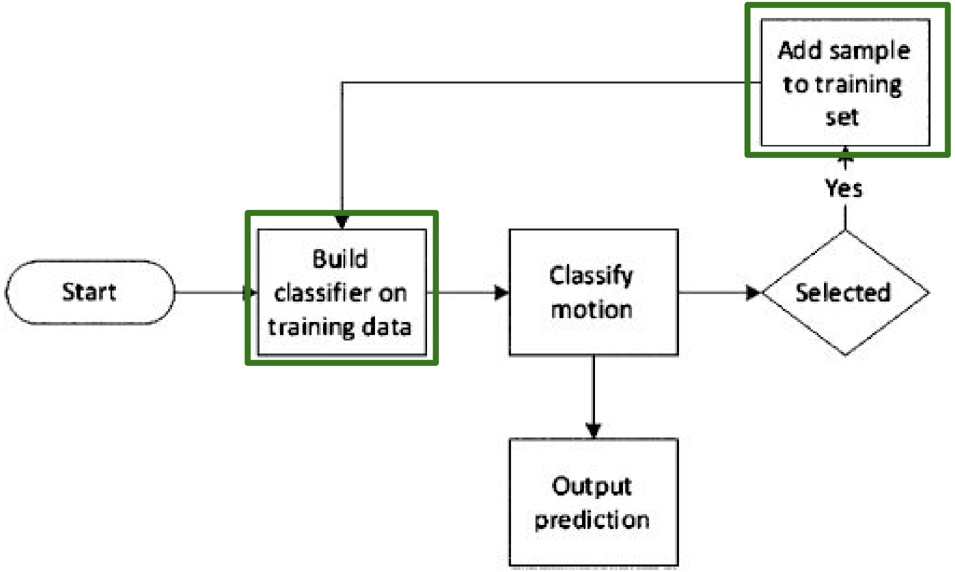
# General Learning Model

- Classify observed motion
  - If confidently predicted, add to the training set



# Continuous Improvement of Activity Classification

- Update the model based on the new training set
  - New iteration



# Self-Learning

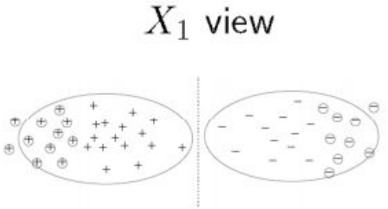
- Single classifier
- Start with a “seed” set of labeled training data
- Use “seed” set to build initial classifier
- Classify unlabeled data
    - **Confidently** classified samples are added to “seed” set

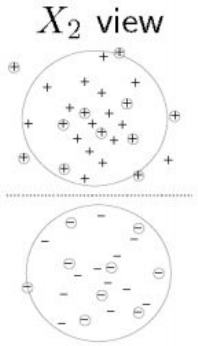
**Benefits:** Simple, continues to expand/tailor training pool

**Limitations:** Negatively impacted by “confidently” mislabeled data, requires larger quantities of labeled data compared to other semi-supervised learners

# Co-Learning

- Use two classifiers
  - Different views (or feature sets)
  - Must not be perfectly correlated
- Train with same data
- Predict unlabeled data
  - **Confident** predictions added to the training set
  - Retrain with additional data



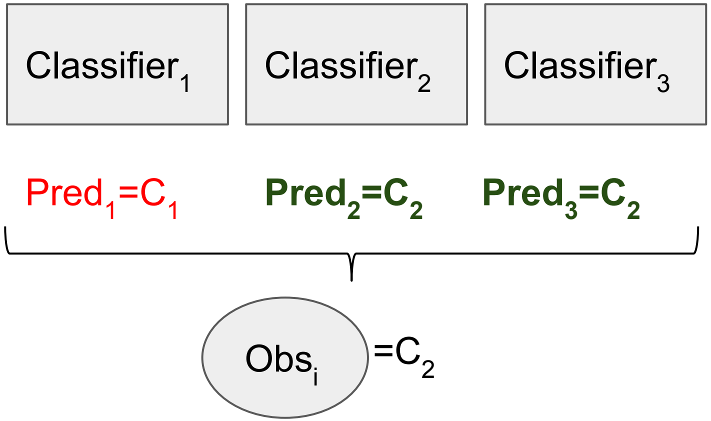




**Benefits:** Confidence is disaggregated to two classifiers

**Limitations:** Requires views to be splittable such that they are independent and can classify

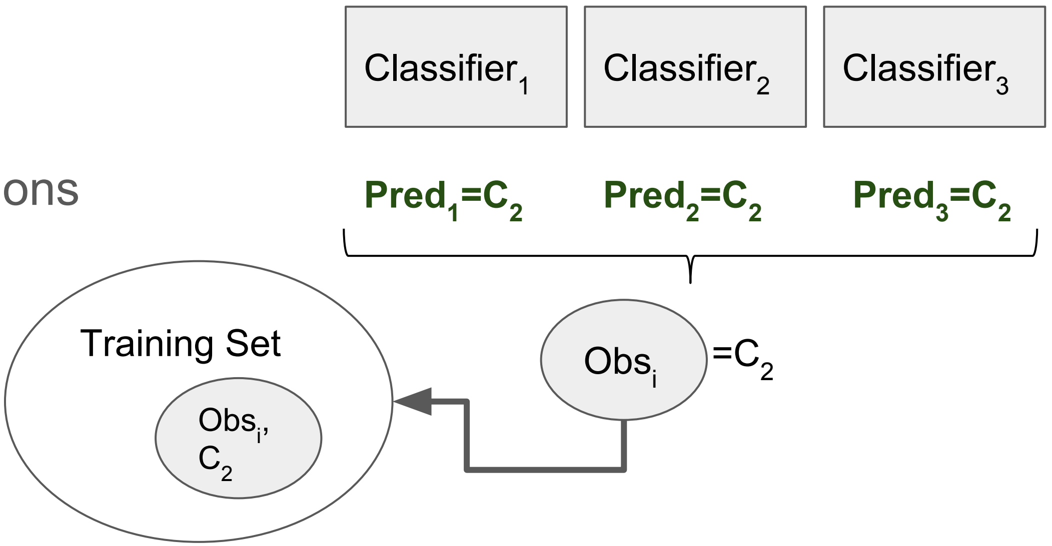
# En-Co-Training

- Ensemble learning
  - Same data view, three different classifiers
- Prediction based on majority voting
  - Eases prediction and confidence to classification label



# En-Co-Training

- Update training set
  - Consensus classifications
- Update classifiers
  - Occurs periodically

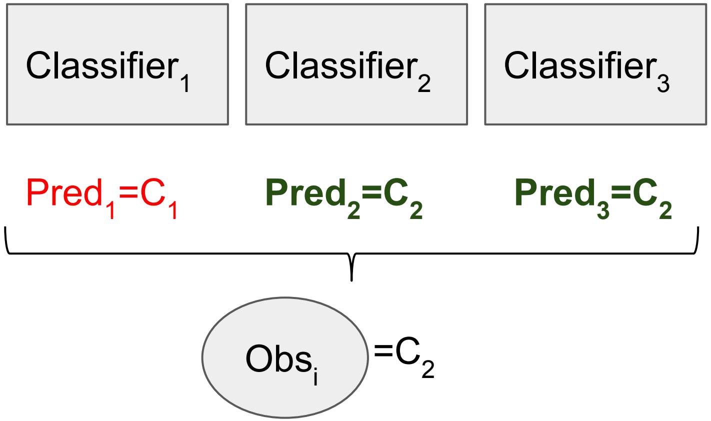



**Benefits:** Ease of prediction determination, democratic classification

**Limitations:** Redundant training on common data set

# Democratic Co-Learning

- Ensemble Learning
  - Same views, three different classifiers
- Initial training on same data set
- Prediction based on majority voting

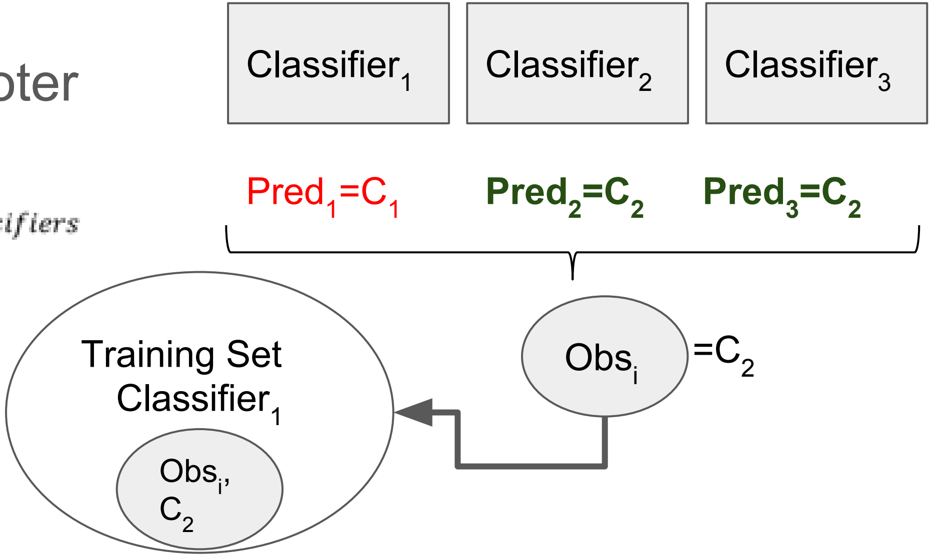


# Democratic Co-Learning

- Update training set of minority voter
  - Priority selection of points

$$Priority[obs_i] = \sum Majority\ Classifiers - \sum Dissenting\ Classifiers$$

- Update classifiers
  - Occurs periodically




**Benefits:** Separate training pools, priority training

**Limitations:** Without priority will have a quickly growing training set

# Active Learning

- Train a classifier on labeled data
- Balance user interruption with classifier accuracy
  - Choose samples of interest for user to label **manually**
    - **Uncertainty Sampling,** Dissenting Committee, Expected Model Change, Expected Error Reduction
  - Update training set with user input

**Benefits:** Exact labeling of priority data points

**Limitations:** Requires user input and feedback

# Experiment

- Classification of **idle**, **walking**, and **running**
- Build Base Classifier
    - 17 participants, 30 minutes per activity
- 15 participants for unlabeled data

- **Classifiers**
    - Self-learning - C4.5 Decision tree
    - Active learning - C4.5 Decision tree
    - Co-learning - C4.5 Decision tree, naive Bayes, and Support Vector Machine(SVM)
        - SVM used sequential minimal optimization algorithm

# Classifier vs. Static Classifier

TABLE I
PERCENTAGE CHANGE FROM BASE CLASSIFIER WITH 480 NEW DATAPOINTS OVER EIGHT ITERATIONS AND A CONFIDENCE INTERVAL OF 95%

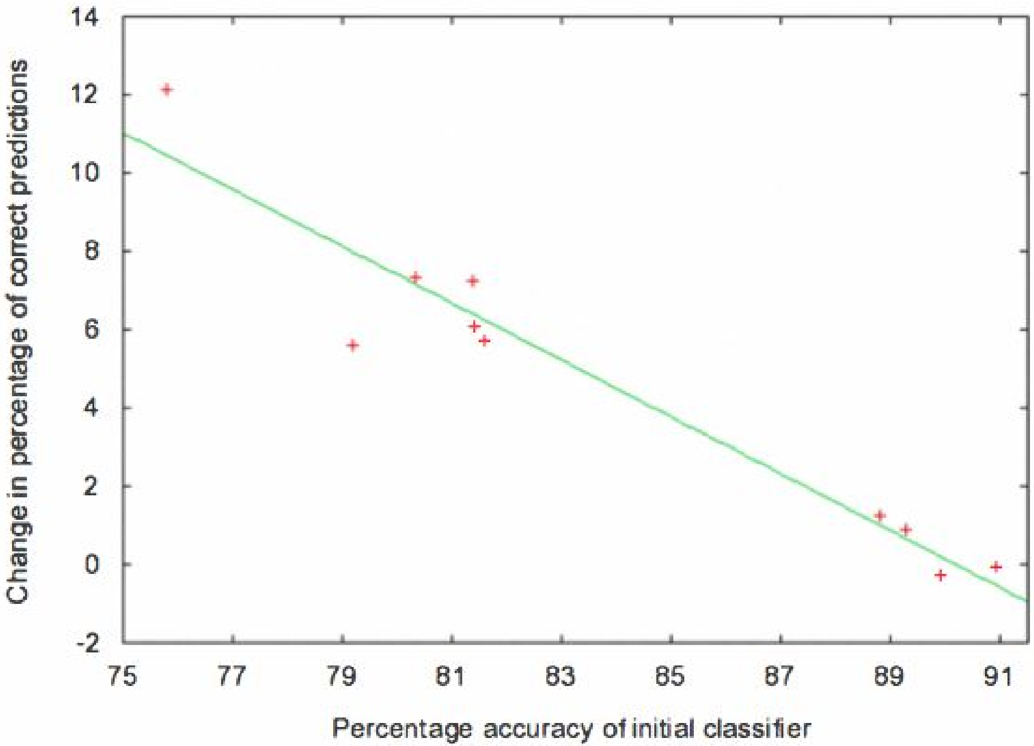
| | Self-Learning | Active Learning | En-Co-Training | | Democratic Co-learning | |
|---|---|---|---|---|---|---|
| Unlabeled | DT only | DT only | DT only | Democratic | DT only | Democratic |
| 50% | −1.27% ± 2.07% | 2.15% ± 2.85% | −0.91% ± 2.15% | −0.34% ± 2.67% | −2.06% ± 3.08% | −0.63% ± 2.85% |
| 55% | −5.35% ± 5.66% | 3.17% ± 4.87% | −6.64% ± 6.46% | 0.67% ± 0.66% | −1.46% ± 3.14% | 0.38% ± 0.87% |
| 60% | 3.31% ± 4.41% | 17.13% ± 7.95% | 5.53% ± 5.29% | 13.05% ± 7.20% | 14.38% ± 8.31% | 15.07% ± 8.00% |
| 65% | 0.05% ± 0.28% | 12.38% ± 7.28% | 0.88% ± 1.66% | 6.34% ± 3.43% | 8.59% ± 8.08% | 10.48% ± 6.34% |
| 70% | 0.17% ± 0.54% | 9.35% ± 6.41% | 0.04% ± 0.58% | 5.04% ± 3.14% | 7.99% ± 5.76% | 8.41% ± 5.82% |
| 75% | 3.31% ± 4.41% | 9.79% ± 6.44% | 1.65% ± 6.51% | 6.69% ± 4.61% | 9.03% ± 6.31% | 9.12% ± 6.31% |
| 80% | −0.02% ± 0.03% | 1.48% ± 2.31% | −0.01% ± 0.03% | 1.14% ± 0.80% | 0.54% ± 1.40% | 1.03% ± 1.11% |
| 85% | 1.38% ± 1.87% | 8.77% ± 6.57% | 0.23% ± 0.55% | 5.45% ± 3.51% | 7.80% ± 6.40% | 8.84% ± 6.12% |
| 90% | −0.63% ± 0.89% | 3.13% ± 4.50% | 0.10% ± 1.54% | 1.41% ± 1.56% | 0.51% ± 2.15% | 1.02% ± 1.95% |
| 95% | −1.74% ± 1.33% | 8.90% ± 5.03% | 1.82% ± 2.79% | 6.27% ± 4.08% | 8.72% ± 6.48% | 8.97% ± 6.56% |

**“Unlabeled”** refers to the quantity of unlabeled data used

$\mu = (\mu_{classifer} - \mu_{base\ classifier})$, 95% confidence interval

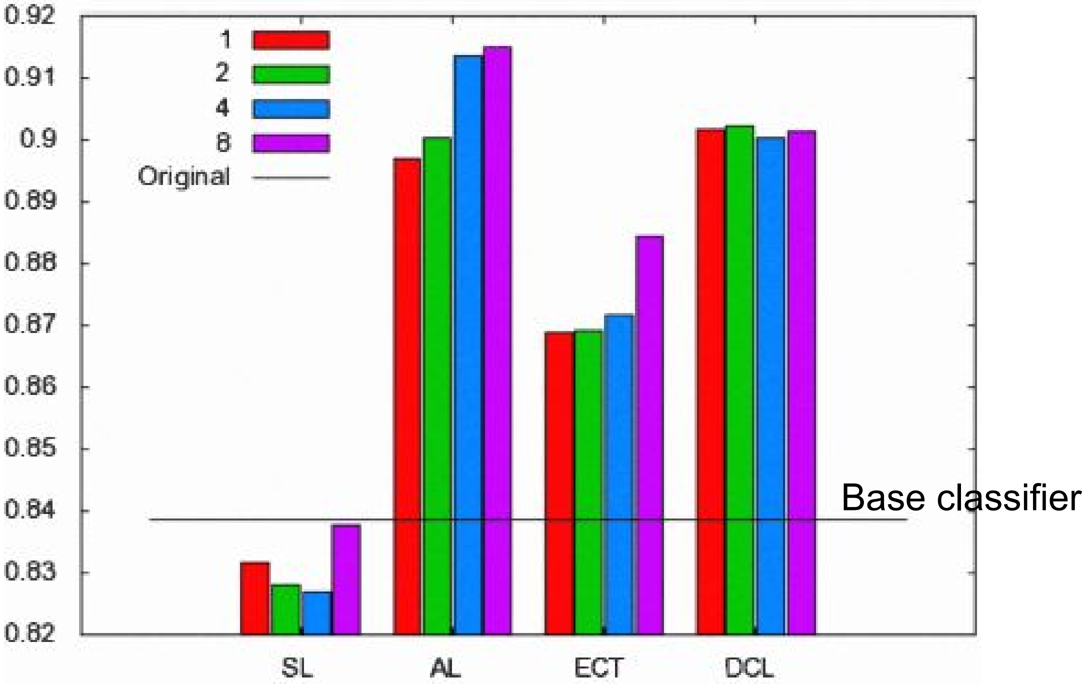
# Correlation Between Initial Classifier

- Each data point is the average percent difference from the base classifier for the quantity of unlabeled data
- Unlabeled data used increases from 50%-95% in increments of 5%



# Updating the Model

**Iterations** - number of instances the model was updated based on new training data



# Conclusions About Learning Model Study

- Democratic Co-Learning(DCL) and Active Learning have comparable results
  - DCL avoids patient interaction
- Limitations on responsiveness to new trends and new classifications
  - Change in user behavior
- Study is limited to recognizing 3 activities

# Future Works

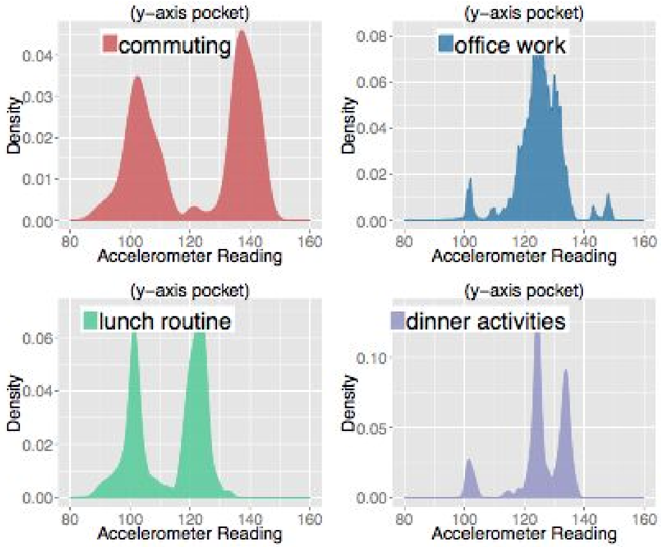
## Nonparametric Discovery of Human Routines from Sensor Data

- Extract low level features of sensor data
  - I.e. accelerometer or GPS
- Build higher level features, “artificial words”, from composite of low-level feature set
  - Dirichlet Process Gaussian Mixture Model (DPGMM)
- Examine set of “artificial words” in a time window to build a routine
  - Hierarchical Dirichlet Process (HDP)

# Future Works

## Nonparametric Discovery of Human Routines from Sensor Data

- Extract low level features of sensor data
  - I.e. accelerometer or GPS

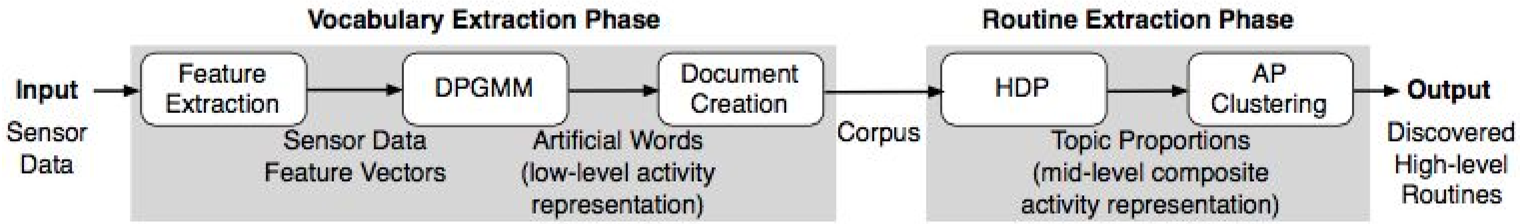



(a) Density distributions of mean of accelerometer data (y-axis pocket) from the daily routine dataset

# Future Works

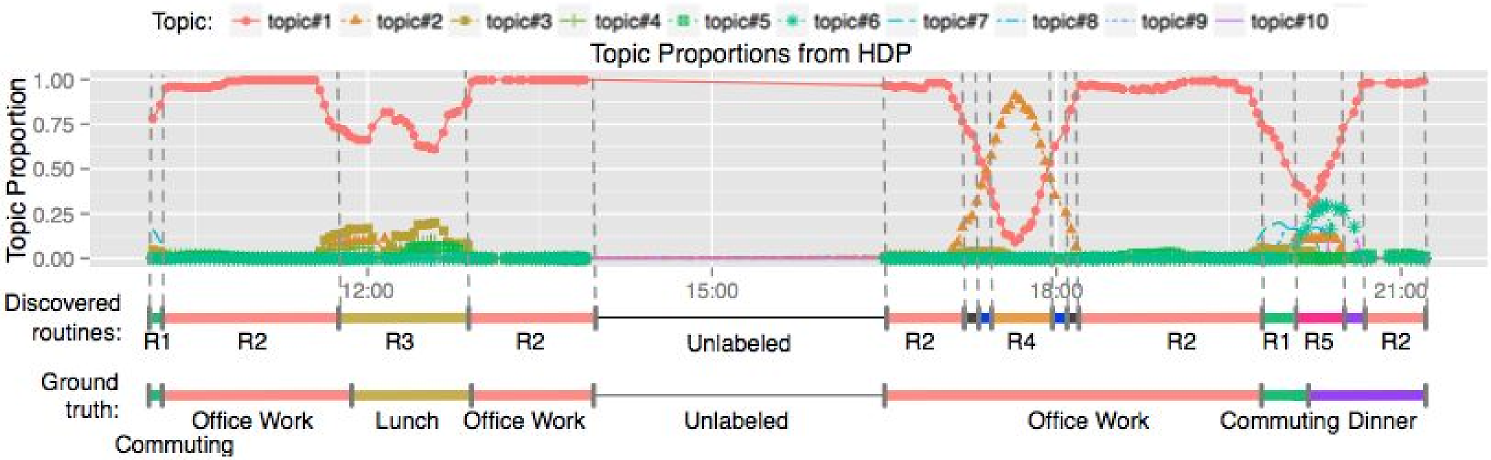
## Nonparametric Discovery of Human Routines from Sensor Data

- Build higher level features, “artificial words”, from composite of low-level feature set
  - Dirichlet Process Gaussian Mixture Model (DPGMM)



# Future Works
## Nonparametric Discovery of Human Routines from Sensor Data

- Examine set of “artificial words” in a time window to build a routine
  - Hierarchical Dirichlet Process (HDP)



# References


1. Longstaff, B., Reddy, S., & Estrin, D. (2010). Improving activity classification for health applications on mobile devices using active and semi-supervised learning. *Proceedings of the 4th International ICST Conference on Pervasive Computing Technologies for Healthcare*. doi:10.4108/icst.pervasivehealth2010.8851
2. Sun, F., Yeh, Y., Cheng, H., Kuo, C., & Griss, M. (2014). Nonparametric discovery of human routines from sensor data. *2014 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. doi:10.1109/percom.2014.6813938
3. Divvala, S. K. (n.d.). *Co-Training & Its Applications in Vision*. Lecture.
4. Semi-supervised learning - Wikipedia, the free encyclopedia. (n.d.). Retrieved October 19, 2016, from https://en.wikipedia.org/wiki/Semi-supervised_learning