Brief Article

The Author

November 2, 2015

1 Name of the project

HomeActivity: Recognizing activities in a home using sensor information.

2 Some description of the problem, and why it is interesting/challenging.

For our group project, we tackle the problem of activity recognition using available sensor information installed in homes. These sensors could be infrared motion detector sensors (to understand when a person entered a room or opened a fridge), to RFID sensors, to microphone based labeling of activities to record the ground truth. These datasets are usually available in the form of timestamp and the corresponding activity or sensor label. So, in this project we would like to predict a label (i.e. the activity) given a sequence of observations.

1. Interesting because it has applications that spans across various domains:

   (a) Healthcare
   - Long term monitoring of activities could provide us interesting insights on degenerative heath
   - Especially useful in monitoring health of elderly people

   (b) Energy savings
   - Understanding the pattern of activities could help us save energy.
   - For example, switching off AC/heater when no one is at home

   (c) From a Security perspective
   - Again, notifications can be sent to alert the home owners if there is an aberrant activity in the door entrance.

(d) Intelligent Homes

    -

2. Challenging

    (a) Sensor datasets could be noisy

      - the front doors may open and close multiple times; this doesn't may not indicate that the person has left the home.

      - sometimes the sensors itself may have false positives.

    (b) multiple labels may be mapped to a single sensor activity

      - opening of a fridge may mean both; make coffee and make cooking

      - some other information such as how long the person spent in the kitchen may provide some useful insight

    (c) Imbalanced class problem

      - machine learning algorithms and works best when the number of instances of each classes are roughly equal

      - where the total number of a class of data (positive) is far less than the total number of another class of data (negative)

      - In this problem, no activity is a perfectly reasonable label. So the prediction model may always output 'no activity' and be accurate 80% of the time.

3 Methods that might be explored to solve it.

We would like to explore the following techniques:

1. HMM -

2. CRF -

3. SVM -

 The datasets available to us are:

1. Kasteren dataset - has over a month long sensor information from 3 homes.

2. Tulum dataset - more than six month period from a single home

4 Timeline for finishing the project.

- 2.5 weeks for implementing these modeling methods.
- 1 week for analyzing with different dataset; and writing it up. -