

Human Activity Recognition in Smart Homes

Helnaz Soltani

June 09, 2020

Galvanize DSI

Project Overview



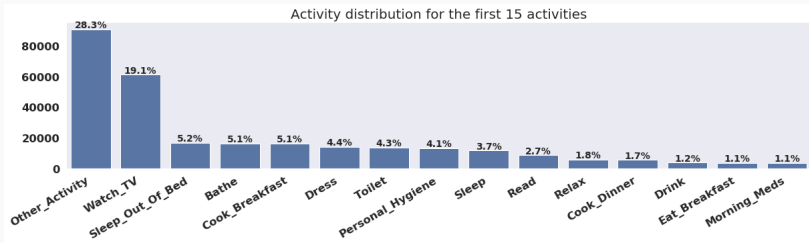
- Objectives:
 - To map a sequence of sensor events to a corresponding activity label
- Applications:
 - Health care
 - Energy saving
 - Security

Dataset

- Public source: UCI repository
- 30 datasets for 30 smart homes with different floorplans:
 - 7M+ observations
 - 37 features
 - Multivariate classification with 45 classes
- Data reflects info from motion sensors, door sensors, and light sensors.
- 1 floorplan for EDA and initial modeling:
 - 320k+ observations
 - 37 features
 - Multivariate classification with 35 classes

Explanatory Data Analysis

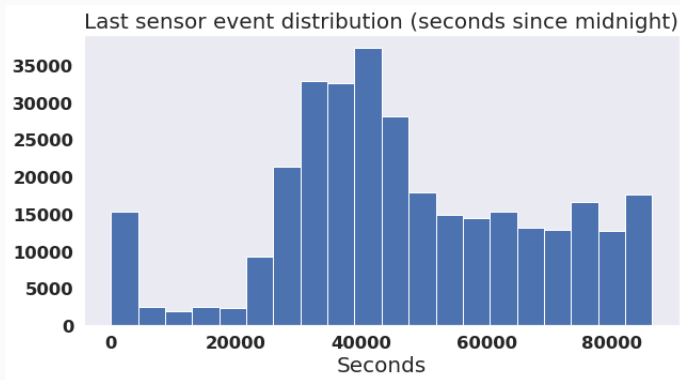
- A great majority of activities belong to 'Other Activity'.
- The rest of activities (which are not shown in the below plot) have less than 1% distribution.



- The data is pretty imbalanced.

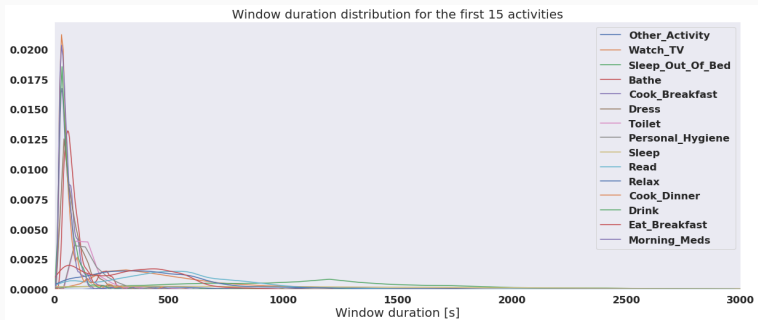
Explanatory Data Analysis

- The sensors record less activity during the normal sleeping time.



Explanatory Data Analysis

- The activities with the highest majority seem to have smaller window duration.

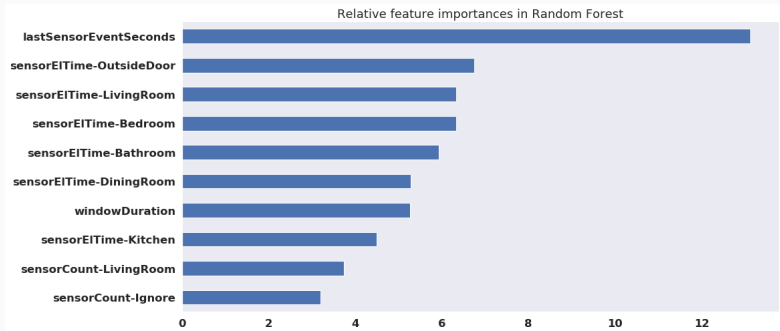


Machine Learning Modeling

- Model selection
- Performance Metrics
 - Accuracy
 - Confusion matrix
 - Elapsed run time
- Resampling dataset
- Model tuning
- Dimensionality reduction (PCA and tSNE)
- Generalizing Decision Tree to the combined 30-smart-home datasets with 7M+ observations:
 - Accuracy: 92.3%
 - Elapsed run time: 294 sec

	Accuracy	Elapsed Time[s]
Logistic Regression	0.662	86.73
Decision Tree	0.953	8.67
Random Forest	0.973	83.05
Gradient Boosting	0.840	4646.25
XGBoosting	0.966	973.75
Ada Boosting	0.410	67.64
Multi-layer Perceptron	0.887	106.59

Feature Importance (10 tops)



- The number of seconds since last sensor activation is the main contributor to activity prediction.
- The sequence of sensor records are pretty important.

Conclusions

- Concluding Remarks
 - The predictive model is able to recognize human activity in a smart home with 97% accuracy.
 - The prediction error is for the categories that have not been defined clearly.
 - The raw data was processed fairly well by domain experts which affects the model performance.
- Future Plans
 - To generalize all models to the combined 30-smart-home dataset which makes my predictive model independent of floorplan
 - To approach the problem unsupervisingly
 - To generalize the model to wearable sensors such as smart devices
- [Github Repo](#)

Thank you!