

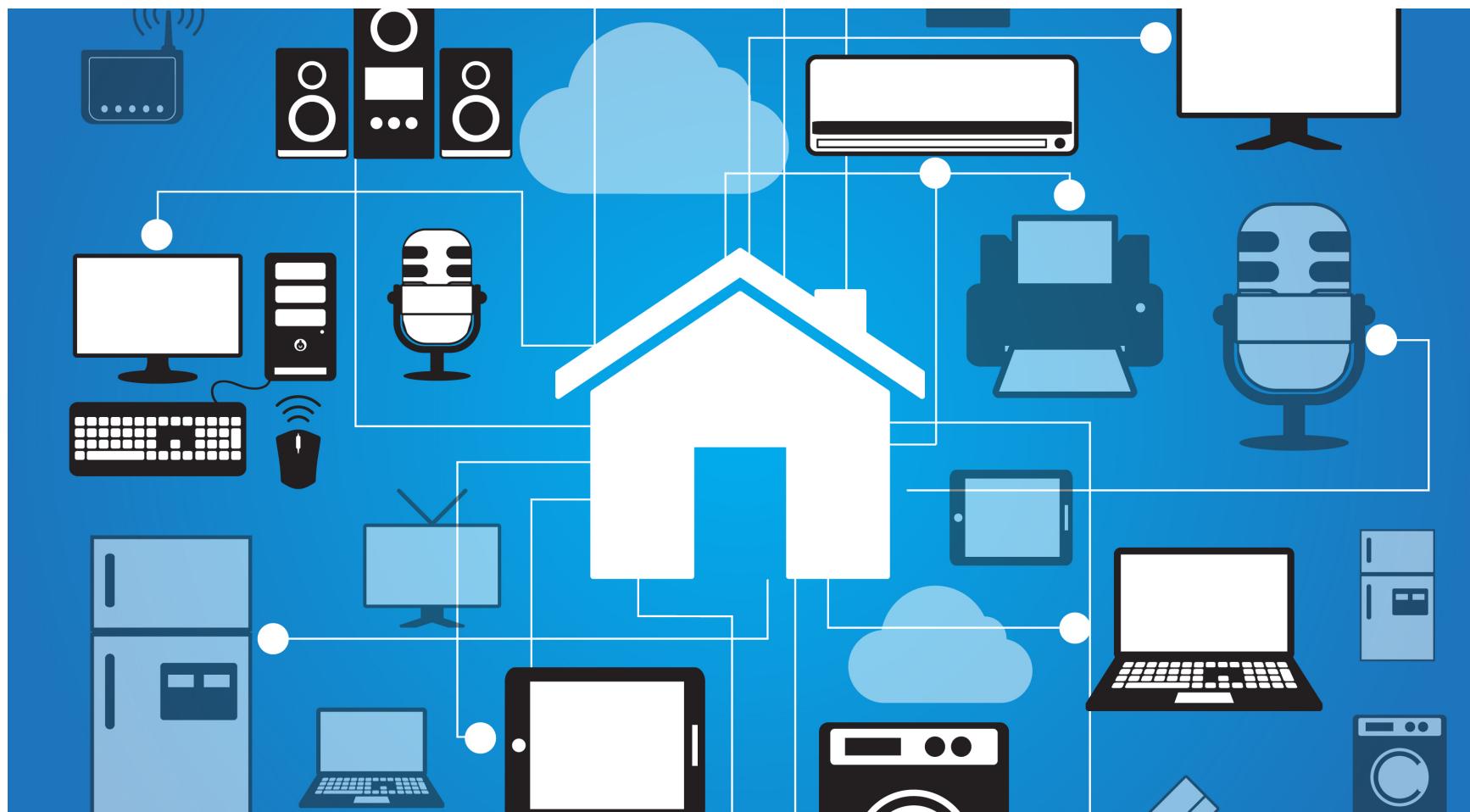


Activity recognition in Smart Home



By Nataliya Nadtoka, Data Scientist
<https://www.linkedin.com/in/nataliya>
co-organizer SF Data Mining
May 18th 2015

IOT of connected home



Dataset

<http://courses.media.mit.edu/2004fall/mas622j/04.projects/home/>

- 22 Every day Activities
- 2 weeks of sensor data in March/April 2003



Sensors in the smart home placed in:

- Bathroom
- Foyer
- Kitchen
- Living room
- Bedroom
- Porch
- Office/study

Some activities had less than five examples...

Activity	Number of Examples
Other	1
Washing hands	1
Going out for entertainment	1
Lawnwork	1
Going out for shopping	2
Putting away dishes	2
Putting away laundry	2
Putting away groceries	2
Watching TV	3

The activities we used:

Activity	Number of Examples
Preparing dinner	8
Washing dishes	8
Cleaning	9
Going out to work	12
Preparing breakfast	14
Preparing a snack	15
Preparing a beverage	15
Preparing lunch	17
Bathing	18
Doing laundry	19
Dressing	24
Grooming	37
Toileting	84

Main Challenge: Data format

Toileting,4/1/2003,11:52:1,11:58:50

100,137

Toilet Flush,Freezer

11:55:43,11:56:2

16:35:49,11:56:13

Going out to work,4/1/2003,12:11:26,12:15:12

81,139,140

Closet,Jewelry box,Door

12:12:29,12:13:27,12:13:45

12:13:0,12:13:35,12:13:48

Preparing lunch,4/1/2003,11:21:17,11:38:22

140,137,131,53,84,131

Door,Freezer,Toaster,Cabinet,Drawer,Toaster

11:23:4,11:23:55,11:24:8,11:34:59,11:35:4,11:35:12

11:23:7,11:24:3,11:24:14,11:35:1,11:35:7,11:35:22



95% of time spent
on
data cleanup
and
feature engineering

Sensor Features

- Sensor activation duration
- Order of sensor activation
- Relative order
- Seconds from the start of activity
- Seconds from the activity start normalized by the total activity duration in seconds
- ...

External features

- Day of week
- Time of day
 - 4am-10am
 - 10am-4pm
 - 4pm – 10pm
 - 10pm-4am



Potential feature candidates (not used here):
sensors locations, national holidays, ...

Dataset

578 features

78 sensors

280 examples

13 activities

Lots of NaNs

Machine Learning

- Decision tree (full/reduced features): 0.52
with full 0.55 with partial
- Logistic Regression (full features): 0.5
- KNN (full features): 0.53
- Random forest (full features/partial features)
0.73 with full features, 0.57 with partial

WORKED

DID NOT WORK

- Taking time for feature engineering
- Random Forest

- Native Bayes
- Reducing number of features for Random Forest
- Logistic Regression

If I only had more time...

- More features for Random Forest
- Better feature engineering
- More time on understanding results/refining the features



Conclusion

More Data and clever
feature engineering beats
better algorithms



