Making Machine Learning Classification of Sensor Data Easier

Dan Tasse, Anind Dey, CMU

Classifying sensor data is useful

With RapidABC data: classify child as "engaged" or not, or other higher-level concept.

In other domains, behavioral and otherwise:

Activity recognition
Stress detection
Cognitive load estimation
Gesture recognition

ML provides a good way to classify sensor data

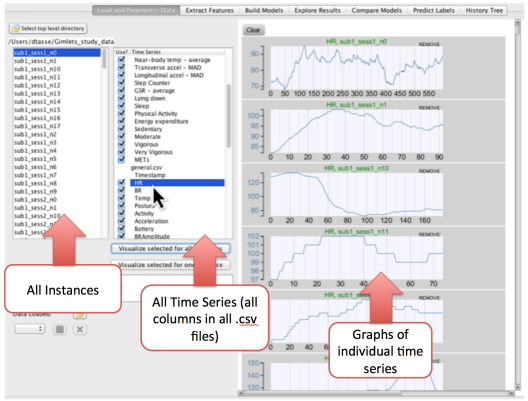
Algorithms like support vector machines and naïve Bayes classifiers can "learn" to classify data samples, based on labeled training data.

However, it is difficult

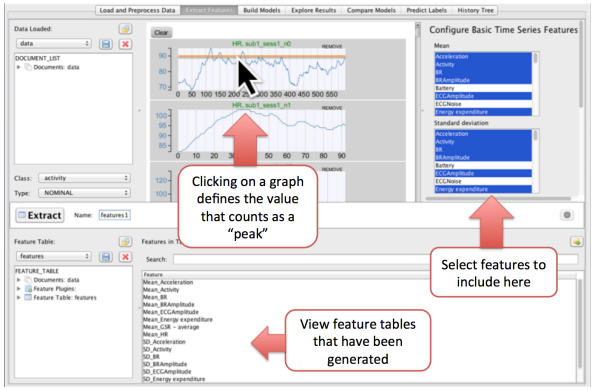
Sensor data is voluminous (many samples per second), multidimensional, and unintelligible (hard to visualize and understand).

The machine learning classifier pipeline is complicated, and it's hard to understand the output of all the classifiers you've built.

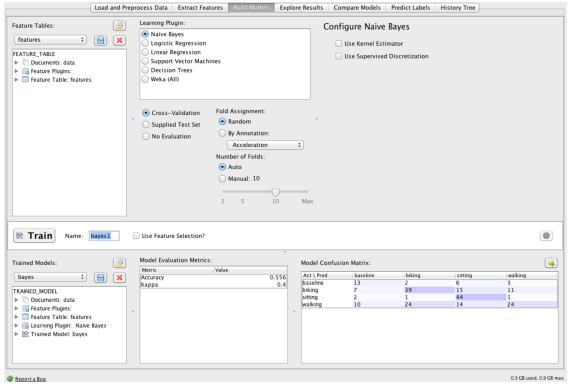
We built a probe called Gimlets to try to make it easier



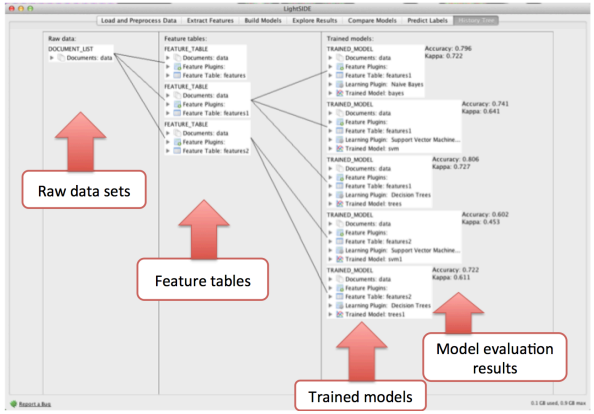
Load and visualize raw data



Generate features easily; build complex features interactively by clicking on graphs



Build and evaluate classifiers



Understand which features and models you tried

Findings from user study:

- participants found our features useful
- some people want to build an accurate classifier, some just want to understand their data
- people want the tool to be smarter
- while this enables intermediate users to build models, it can also help experts explore more options