Mobile Phone-Derived Features for Stop Detection

Daniel Kronovet
Columbia University
dbk2123@columbia.edu

Berk Kapicioglu
Foursquare Labs
berk@foursquare.com

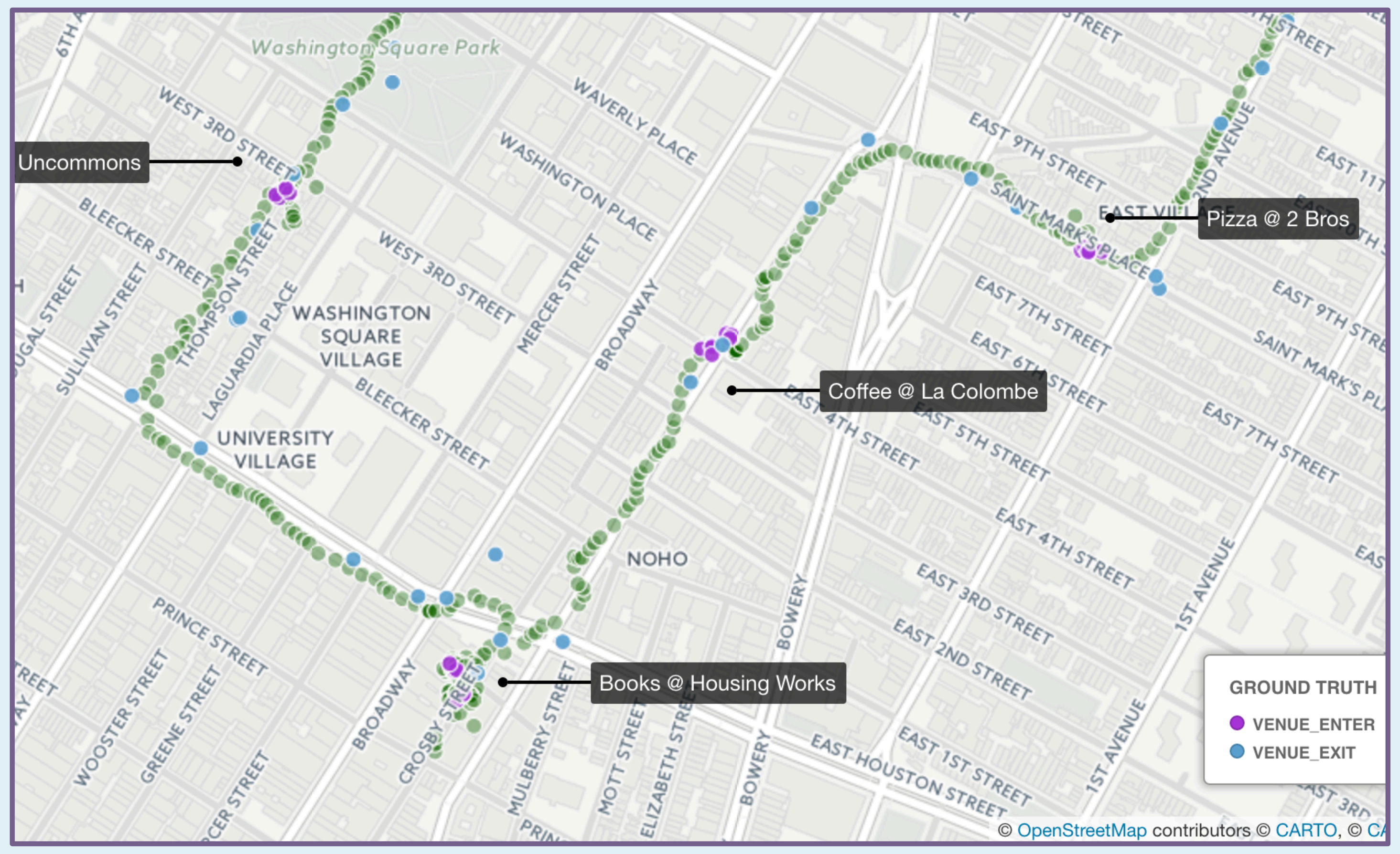
Lauren A. Hannah
Columbia University
lah2178@columbia.edu

Motivation

It is valuable to know how different types of people move through the world. The ubiquity of sophisticated mobile phones makes it possible to get location data in real-time, allowing for new kinds of analyses. This work uses these data to ask:

Is the user stopped at a venue, or in transit between venues?

Can we *quickly* and *accurately* distinguish a user browsing a shop from one stopped at a red light?



Data

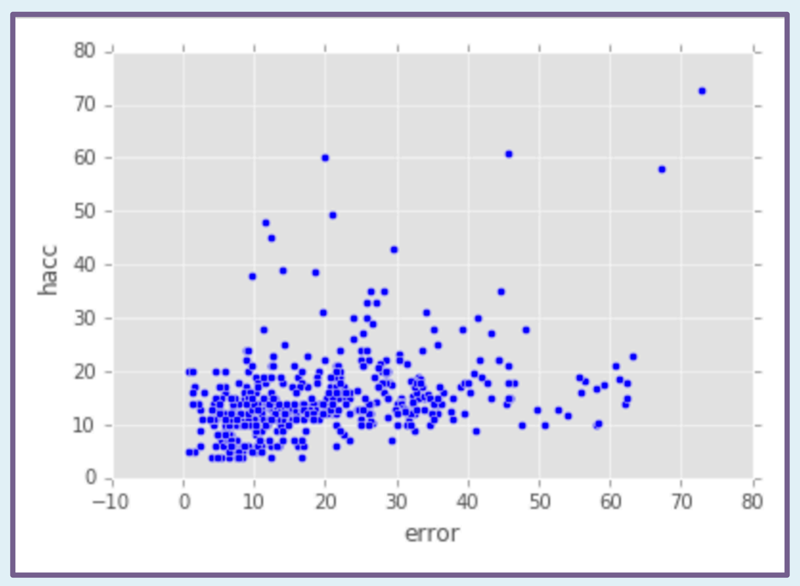
Our data consist of 18 time series, collected by a custom app, representing walking trips through lower Manhattan. Trips lasted ~60-120 minutes; location updates every ~60 seconds. N=1384.

Features include:

- Timestamp.
- Class label (Stop / No Stop).
- Latitude and Longitude.
- Measure of confidence in the given lat/lng ("HACC").
- Array of WiFi Access Points (AP), each containing:
  - A unique identifier.
  - Signal strength.
  - Signal frequency.

Caveats:

- Irregular update intervals.
- HACC weakly correlated with our estimate of error (rho = .354), untrusted.



Coordinate-based Features

We are ultimately interested in questions of (subjective) user intention; we approximate this by modeling (~objective) user speed.

- **Speed**: distance traveled in last X seconds / X.

Defining "distance":

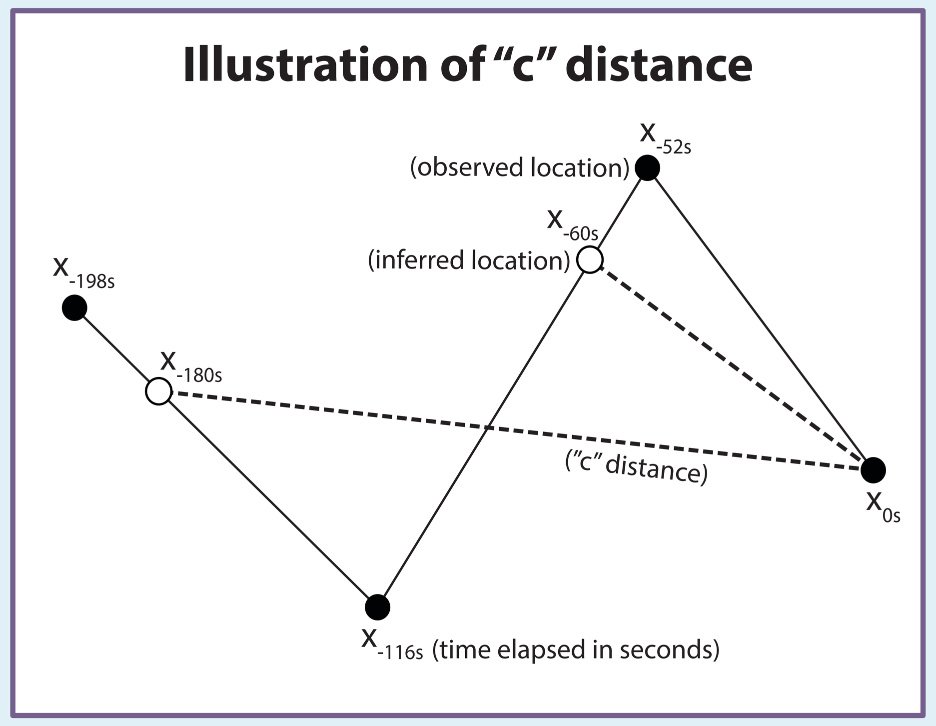
- Irregular intervals: interpolate lat/lng to infer per-second location.
- Euclidean ("c") distance is robust against measurement error.

Problem:

- Positioning imprecise.
- No "true" ground truth.
- Error hard to model.
- Reality is an illusion.

Solution:

- Coordinate smoothing!
- Multiple features.



Techniques:

- Naïve (no smoothing), rolling mean, rolling geometric median.

WiFi-based Features

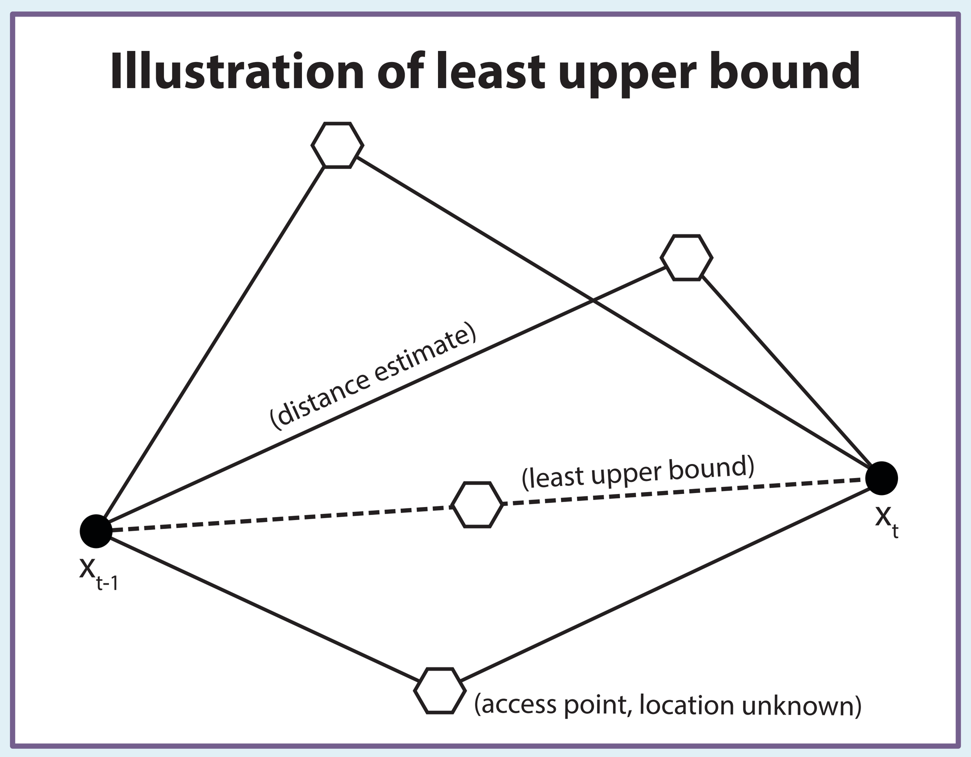
Knowledge of nearby WiFi APs allows for the exploration of a different notion of position and movement:

- **Number** of nearby APs.
- **Approximate distance** from an AP:

$$dist(k_t) = 10^{(27.55 - 20(\log_{10}(freq_{k_t})) - sigstrength_{k_t})/20}$$

- **"Least upper bound"** on distance:

$$lstup(t) = \min(\{dist(k_t) + dist(k_{t-1}), \forall k \in (K_t \cap K_{t-1})\})$$



Model Evaluation

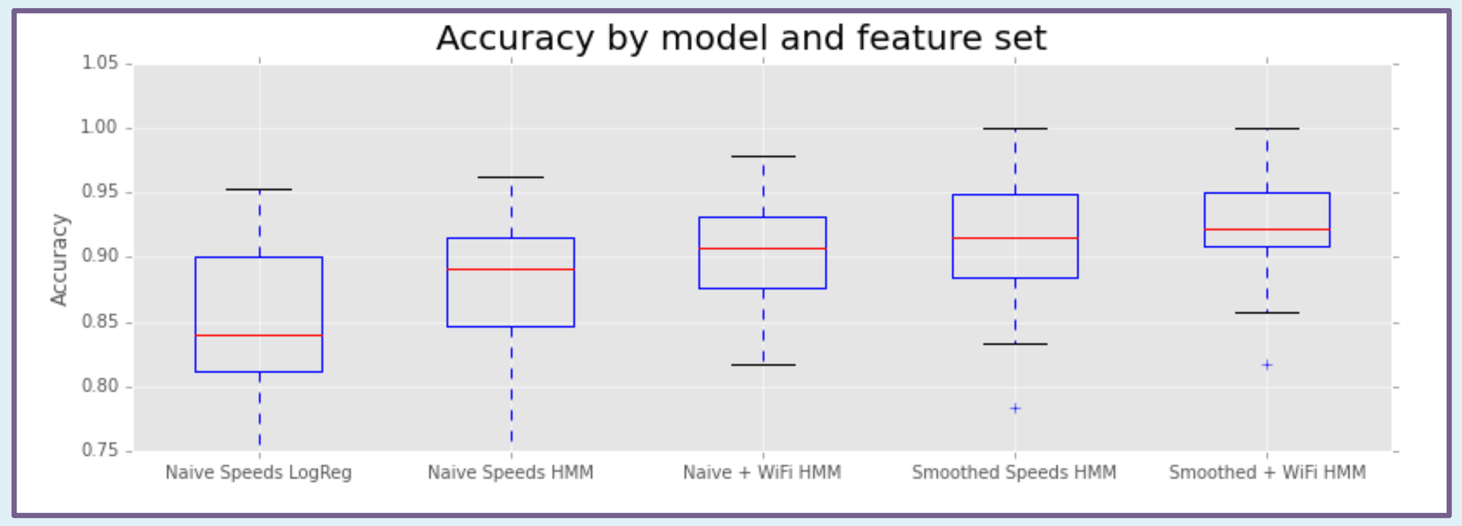
We evaluate features in the context of model performance, considering both **logistic regression** and **hidden Markov models**. We evaluate model performance using:

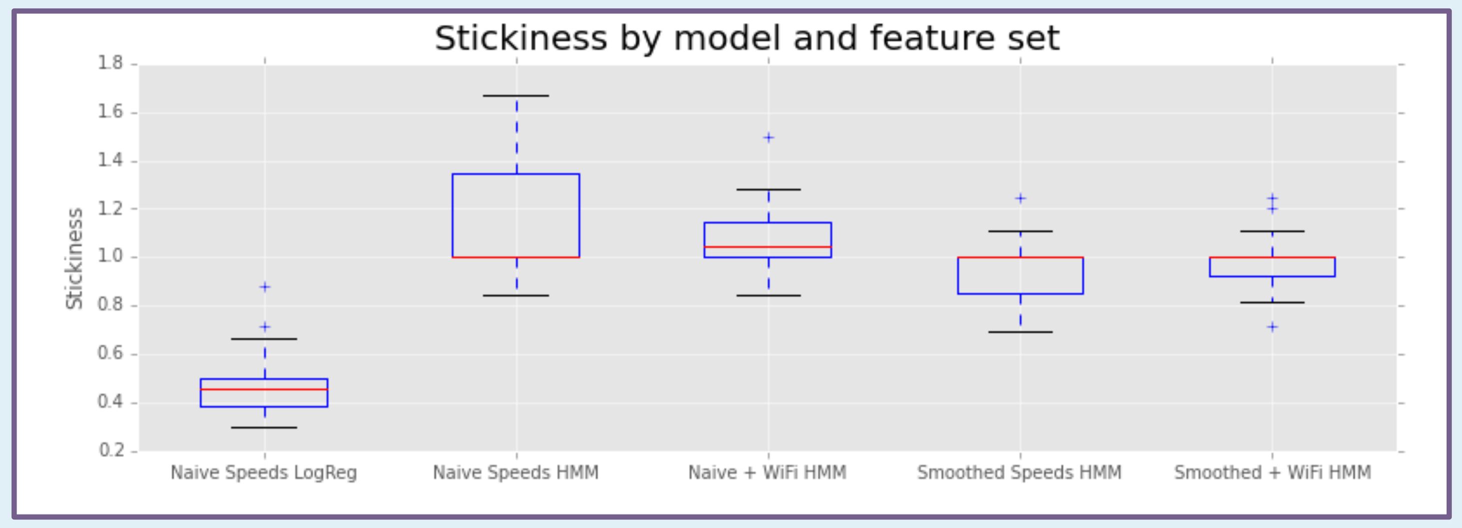
- **Accuracy**: correct predictions over all predictions.
- **Stickiness**: true state transitions over predicted transitions:

$$stk(y, \hat{y}) = \frac{trans(y)}{trans(\hat{y})} \qquad trans(y) = \sum_{i=1}^{|y|-1} |y_i - y_{i+1}|$$

Results

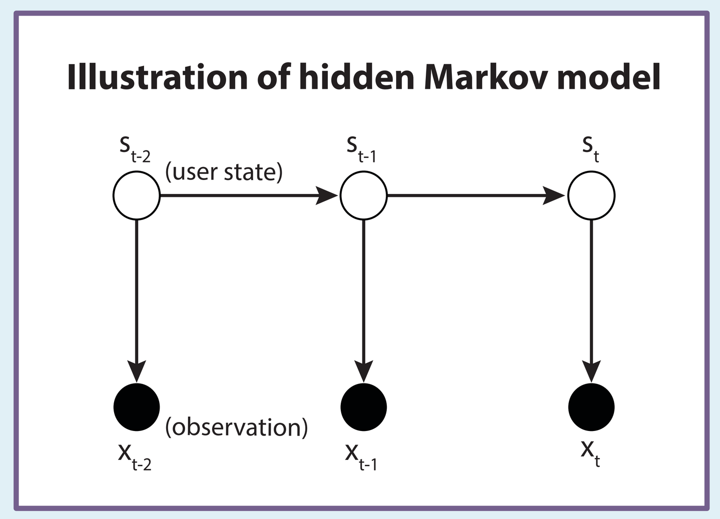
- Hidden Markov models are more accurate and sticky.
- Coordinate smoothing improves error tolerance.
- Mixing coordinate and WiFi features gives best performance.





Discussion:

- GPS can fail in urban areas.
- WiFi APs abundant in cities.
- HMMs model user intent.
- HMMs model state change.



Acknowledgements

We would like to thank Foursquare and the Pilgrim team for their support and assistance during Summer 2016.