

Smart City Analytics: Prediction of Citizen Home Care

Casper Hansen, Christian Hansen, Stephen Alstrup & Christina Lioma
c.hansen@di.ku.dk

Takeaway: We can predict big increases in home care with limited data

Home care prediction

- The City of Copenhagen tracks a lot of information about the citizen's receiving home care
- Can we identify citizens with largely increased (>6 hours) home care need in the following 3 months?
 - Allows for preventive measures to be started
 - Facilitates planning of city resources
- Use an online ensemble of models trained on a monthly basis to predict 3 months ahead. Using either logistic regression or random forest as base learners
- Baselines
 - Predict an increase if it happened the same month last year (seasonal)
 - Predict in increase if it happened within the last 3 months (rapid increases)

Table 1: Our dataset of home care citizen records in Copenhagen from 2013 - 2017.

Category	Features
Basic	gender, age, zipcode, date, civil status of citizen
Living type	own residence, senior housing, assigned residence
Time	day, evening, night, weekday, weekend
Type	public, private
Health Care	generic activity, emergency care stay, dementia care, reoccurring visit, dental care, palliative care, personal care, practical help, rehabilitation, sick care
Feedback	citizen home, citizen not home, citizen hospitalized, other
Length	number of home care hours, number of large increases of at least 6 hours within the last 3 months
Financial	financial cost of service

Table 2: Cumulative information levels (IL)

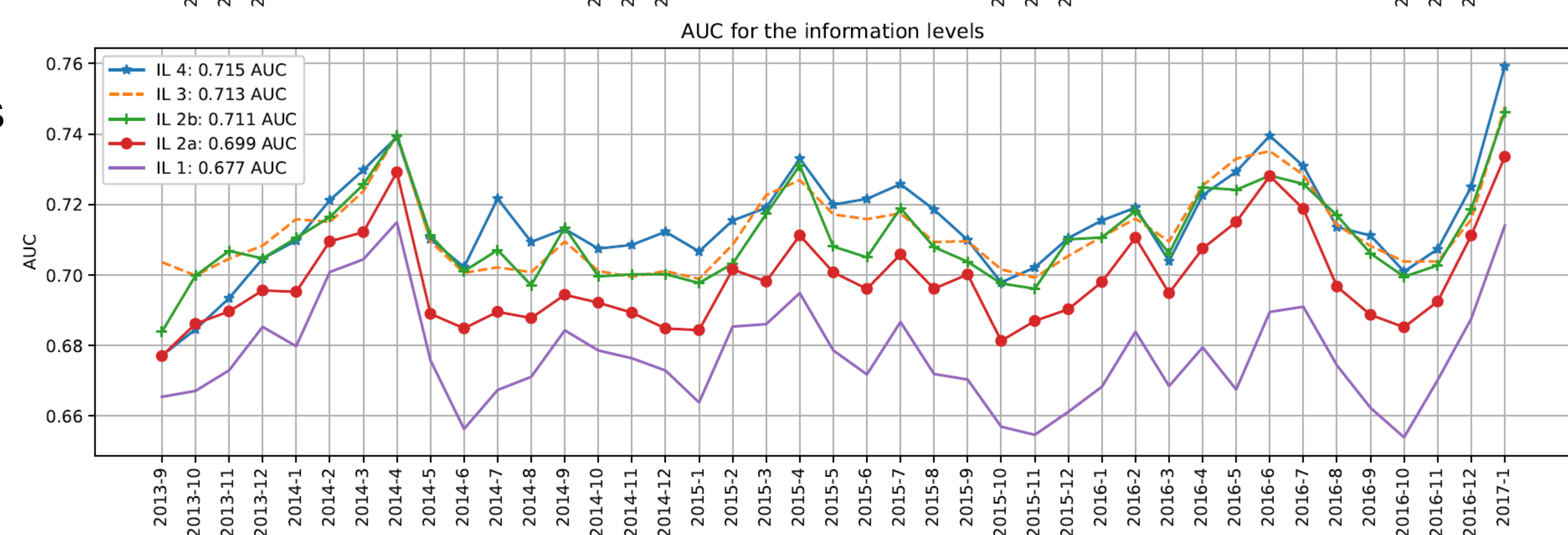
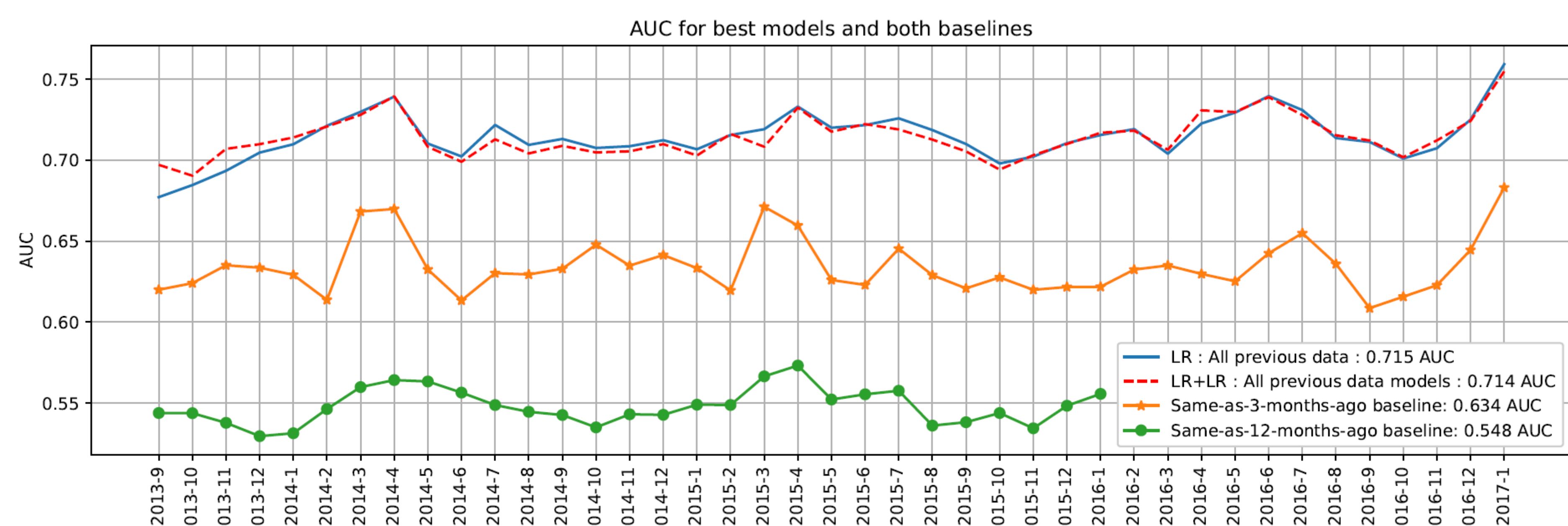
IL	Description
IL1	Basic, Length, and Living type categories from Table 1
IL2a	IL1 + binary Time category from Table 1
IL2b	IL1 + binary Health Care and Type categories from Table 1
IL3	IL2a+IL2b
IL4	Contains IL3 but with the full distribution (instead of binary) Time, Health Care, Feedback and Financial categories from Table 1

Data

- 27,775 citizens from 2013 to 2017 in Copenhagen
- A lot of features: basic living arrangements, provided home care services, time and feedback of services, and hospital admissions

Experiments

- Split data into monthly records with 3 months prior information in an aggregated format – 424,000 samples with 11% having increases above 6 hours
- Evaluate model on every month using AUC
- Partition features in cumulative information levels to see how few and what kind of features are needed



Findings

- 0.715 AUC. Similar performance using online ensemble vs model using just all the available data at once
- Significantly better than baselines
- Limited data (information level 2b) performs very similar to full city records. This kind of data is cheap to acquire!

Future work

- Can we use RNNs for better performance, by better utilizing the monthly sequences of data?
- With access to unstructured care taker text journals, can we utilize natural language processing to extract additional useful features?