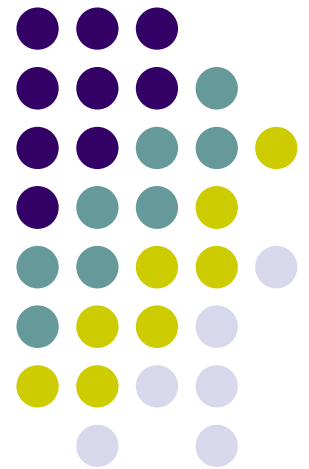
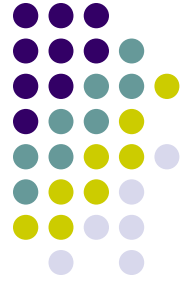


Challenges of Predicting Mobile User Patterns

Jeeyoung Kim
Ahmed Helmy

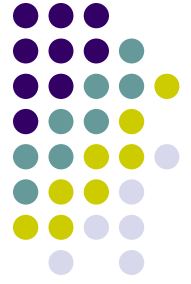




What are the Challenges?

- What is a “success”?
 - Definition (hit or miss? Or gray area allowed?)
 - Granularity (aggregation, different levels of prediction...)
- How close to “ground truth”?
 - Characteristic of the data
 - Never really can compare

Predictors used..



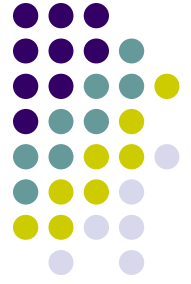
- Order k Markov Chain Predictor
- Lempel-Ziv Predictor
- Regression, other data mining techniques...

Predictors: Order-k Markov



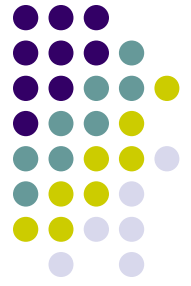
- Assumes location can be predicted from the current context which is the sequence of the k most recent symbols in the location history
- Markov model represents each state as a context, and transitions represent the possible locations that follow that context.
- We use Order 1, 2 and 3 predictors for our prediction

Predictors: Lempel-Ziv (LZ)



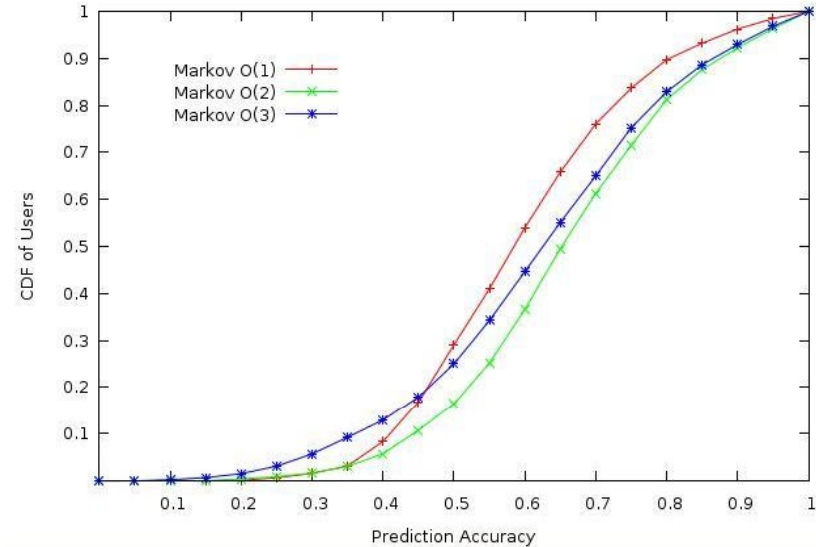
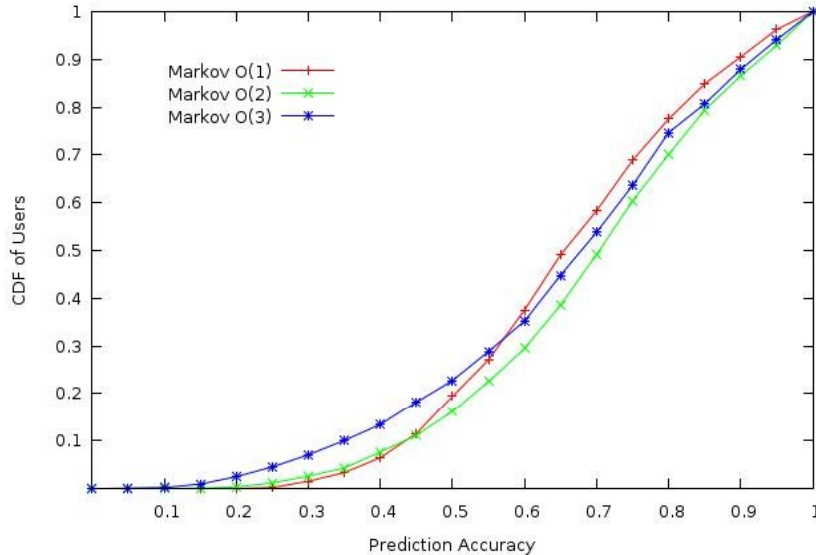
- Predicts in the case when the next symbol in the produced sequence is dependent on only its current state (but does not have to correspond to a string of fixed length)

Predictors

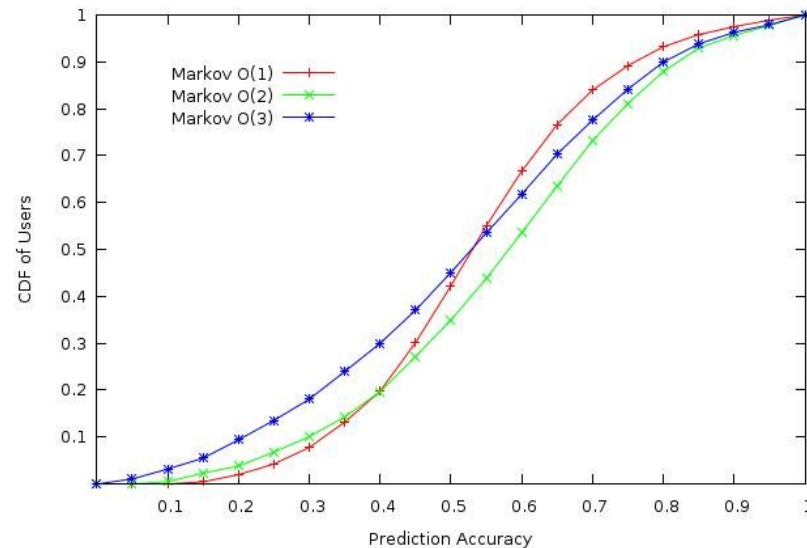


- Each predictor is run for each of the data sets
- The prediction accuracy is measured as the percentage of correct predictions of the next AP to visit, next building to visit (different granularities of prediction)
- AP level prediction results can be aggregated into building levels...(1 out of 5 cells hand-off prediction)

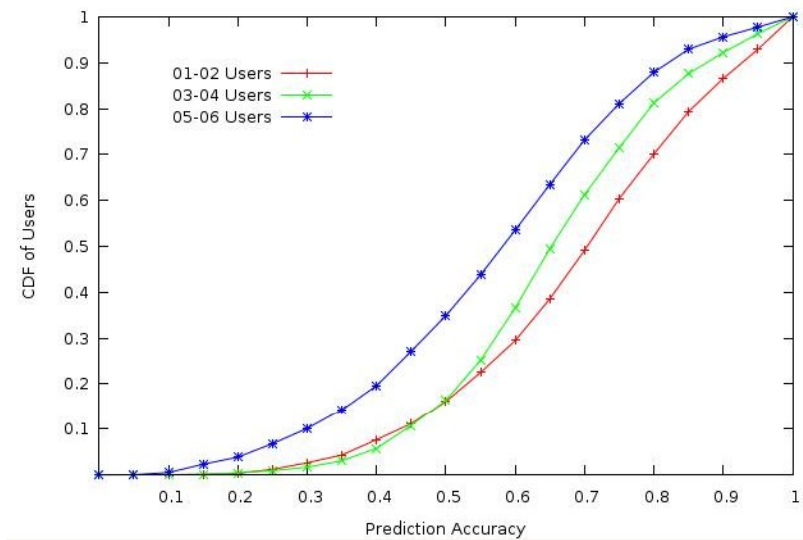
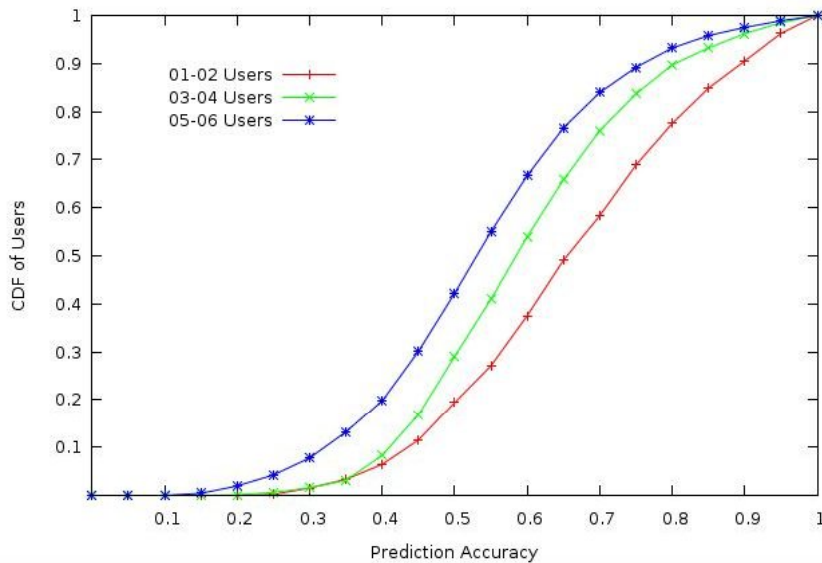
Markov Prediction Accuracies for each years trace



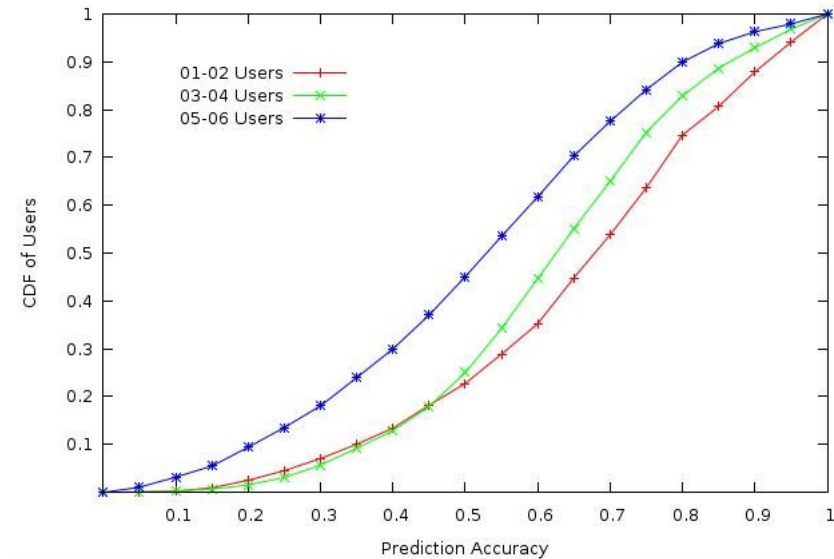
- Comparing the Markov order 1,2 and 3 predictors for each 1 year period (01-02, 03-04, 05-06)



Time evolution of predictability

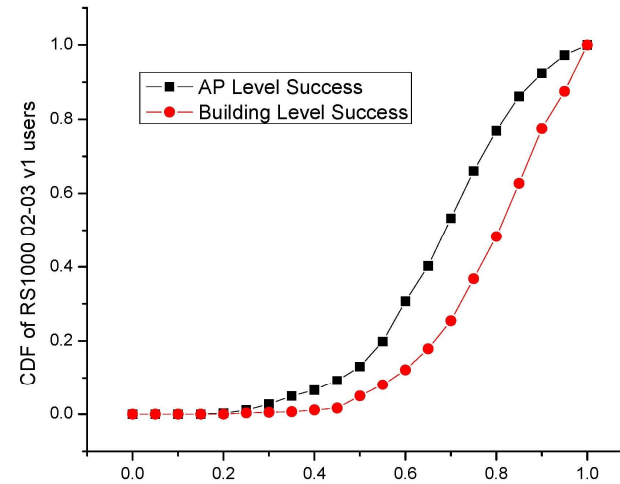
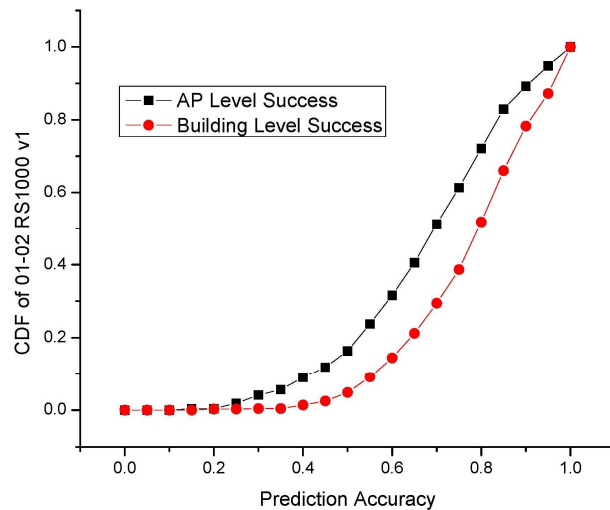


- Time evolution over time for the 3 years of each Markov order 1,2 and 3 predictors

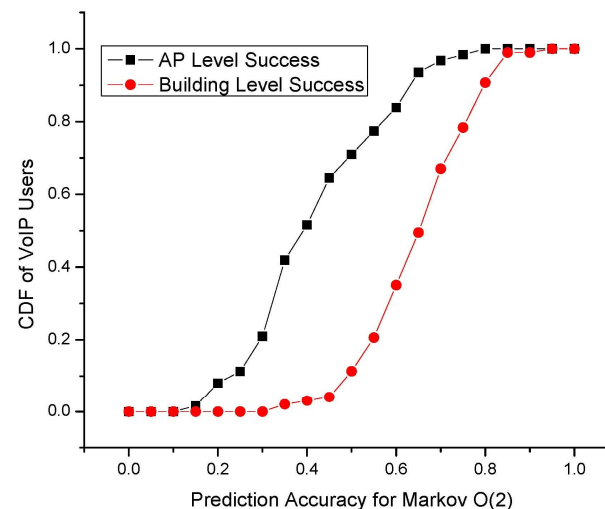




Aggregated prediction results



- Showing the drastic difference of AP level success and Building level success (01-02, 02-03, VoIP)





Conclusions

- Same predictors can differ in results regarding the granularity or the “definition of success”
- Due to the nature of the data itself, it is very difficult to compare it to the ground truth