Glimpsing into the Future: An Accurate WiFi-Based Client Count Prediction Algorithm

Huy Tran
Portland State University
hptran@pdx.edu

Santosh Pandey
Cisco Systems
sanpande@cisco.com

Abhishek Bhattacharyya
Cisco Systems
abhisbha@cisco.com

Preethi Natarajan
Cisco Systems
prenatar@cisco.com

Nirupama Bulusu
Portland State University
nbulusu@pdx.edu

1. INTRODUCTION

Many retailers and government enterprises can greatly streamline their operations management and expenditures if they can accurately predict the anticipated number of people at various zones within their premises at any given time [1, 2]. For example, a large retailer like Wal-mart can optimize personnel costs given accurate predictions of clients at various checkout counters.

The problem we address is to predict client counts (i.e. the median number of people in ten-minute intervals) in different floor zones of a building several hours ahead, given the floor map, historical client counts, and current client counts. In a floor, Cisco's routers record the received signal strength (RSSI) emitted from mobile devices. Cisco's software uses this data to estimate client locations. We use the location data to generate client counts in different zones over time.

Our vision is to enable anticipatory resource deployment by harnessing advances in statistical modeling.

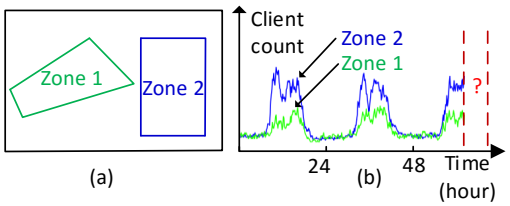

Figure 1: Our problem is to predict client counts several hours ahead given (a) a floor map and (b) historical client counts and current client counts of floor zones

There are two main challenges in solving this problem. First, the time-series of client counts are noisy because they are generated from estimated location data, whose error range is about three to five meters. Second, we have limited information with respect to location data and map information.

There are two approaches to predict client counts: direct count prediction and location-based count prediction. In the first approach, works such as [3] use statistical models and machine learning algorithms to exploit temporal correlation of counts in a single time-series to perform prediction. In our work, we further exploit temporal correlation of client counts in multiple time-series. The second approach is to first predict client locations [4] before generating predicted client counts, potentially compromising user location privacy.

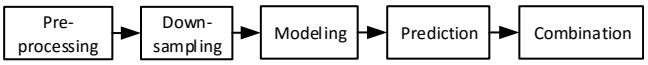

Figure 2: Our prediction algorithm

Our algorithm consists of five steps (Figure 2). Pre-processing maps device locations to floor zones to generate time-series of client counts. Down-sampling calculates the medians of the client counts in every ten-minute interval to create down-sampled time-series of the client counts. Modeling develops statistical models to extract temporal correlations of the client counts. Prediction predicts the client counts several hour ahead by using the models and naive approaches. Combination selects and combines predicted values of these approaches to output the final results.

We evaluated our algorithm with a real dataset containing client locations in a Cisco office building having ten different zones over five weeks. Our results show that, for an eight-hour look-ahead window, our algorithm reduces the mean absolute percentage prediction error [5] by nearly 36% compared to a naive approach.

2. REFERENCES

[1] Why queuing is the new battleground for retailers. http://www.ipsos-retailperformance.com/resources/blog/why-queuing-is-the-new-battleground-for-retailers/.

[2] Wal-mart cuts workers' hours at some stores. http://www.wsj.com/articles/wal-mart-reduces-worker-hours-at-some-stores-to-lower-expenses-1441046019.

[3] J. N. Manjunatha, K. R. Sivaramakrishnan, Raghavendra Kumar Pandey, and M Narasimha Murthy. Citation prediction using time series approach kdd cup (task 1). In *ACM SIGKDD*, 2003.

[4] Christian Koehler, Nikola Banovic, Ian Oakley, Jennifer Mankoff, and Anind K. Dey. Indoor-alps: An adaptive indoor location prediction system. In *ACM UbiComp*, 2014.

[5] Mean absolute percentage error. https://en.wikipedia.org/wiki/Mean_absolute_percentage_error.