Conference Paper Session 19: Occupancy Sensors and Schedules

Inferring Occupant Counts from Wi-Fi Data (KC-19-A041)

Zhe Wang

Lawrence Berkeley National Laboratory

zwang5@lbl.gov

Learning Objectives

- **Explain how to protect privacy when using Wi-Fi data for occupant counts inference**
- Apply occupants' biological responses in building controls for thermal comfort improvement
- Apply multi-agent deep reinforcement learning algorithms in building controls
- Define the major factors that affect the occupancy schedule in a residential building
- Design different types of building systems based on the occupant characteristics
- **Identify important WiFi-related features to infer occupant counts.**

Acknowledgements

LBNL team: Tianzhen Hong, Mary Ann Piette

This research was supported by the Assistant Secretary for Energy Efficiency and Renewable Energy, Office of Building Technologies of the United States Department of Energy, under Contract No. DE-AC02-05CH11231.

The authors appreciate the technical support from Wanni Zhang (LBNL), and data collection support from Michael Smitasin (LBNL), Baptise Ravache (LBNL), Bruce Nordman (LBNL), Han Li (LBNL), and Sang Hoon Lee (LBNL).

Outline/Agenda

- Background
- Methodology
  - Feature engineering
  - Algorithm comparison
- Results and conclusions

Background

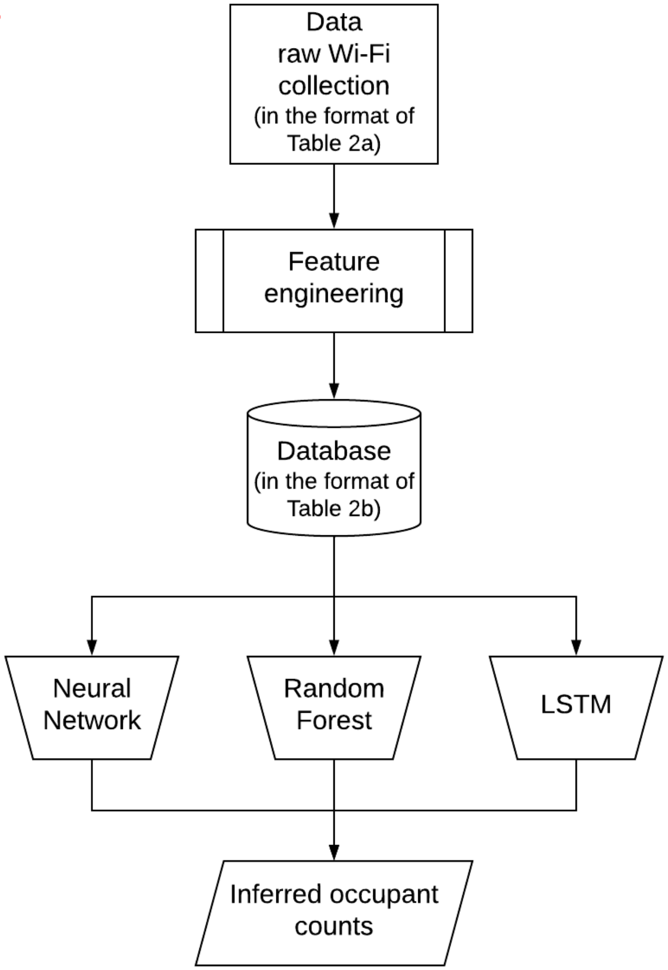
- U.S. buildings consume 40% of primary energy and account for a third of total carbon emission
- Occupancy information could help to reduce energy while enhancing comfort

Resolutions	Definition	Application
Occupancy statues	Whether space is occupied or not	Lighting, HVAC schedule
Occupant counts	How many people are in a space	Control optimization, e.g. MPC, DCV
Identity	Who they are	Personalized thermal environment management
Activity	What they are doing	

- Current occupant sensing technologies (CO2 sensor, camera-based, infrared-based) are *expensive* or *labor-intensive*, and might raise *privacy concerns*

Objectives

- Occupant counts detection through **Wi-Fi**
  - Leveraging existing infrastructure
  - Trade-off between accuracy and privacy
    - MAC address-based approach: accurate but has privacy concerns
    - Connection counts-based approach: protect privacy but not accurate
- Research objective
  - Propose a new approach to detect occupant counts through Wi-Fi, which is *non-intrusive*, *accurate*, and *free of privacy concerns*
  - Keys: *features* and *algorithms*



Method - Feature engineering

- Different type of devices have different **mapping relations** between the *Wi-Fi connection counts* and *occupant counts*

Type of devices		Type of owners		Mapping rules of Wi-Fi connection counts and occupant counts
	*Always connected*		*Office appliances*	*Could not be used to infer occupant counts*
	*Long-term connected*		*Inhabitants*	*Each occupant averagely has two devices (cellphone and computer) connected with Wi-Fi*
	*Short-term connected*		*Visitors*	*Each occupant averagely each has one device (cellphone) connected with Wi-Fi*
	*Occasionally connected*		*Passerby*	*Does not locate in the target area, should not be counted*

Method - Feature engineering

- Cluster the devices based on their connection time
- The clustering could be done locally with a simple script (*edge computing*)
  - Protect privacy: No need to upload MAC address
  - Reduce the size of data cached and transmitted
  - Maintain the major information for occupant counts inference

Time	Shuffled Device_ID	AP_ID
...		
20180521_0000	dfd6bafb68c1cd1f1e2d9190ca9d55f0	ap135-4206w
20180521_0000	e6c1fe930c6d2c2f2e2d9d69fc0abeda	ap135-3103
...		
20180521_0000	dd464552ecc1208e94a955bffee1f749	ap135-4110
20180521_0010	dfd6bafb68c1cd1f1e2d9190ca9d55f0	ap135-4206w
20180521_0010	e6c1fe930c6d2c2f2e2d9d69fc0abeda	ap135-3103
...		

(a) Raw data collected

Time	Target zone	Device_type	Device_count
...			
20180521_0000	Zone 1	Short term (less than 1h per day)	0
...			
20180521_0000	Zone 1	Long term (more than 12h per day)	20
20180521_0000	Zone 2	Short term (less than 1h per day)	0
...			
20180521_0000	Zone 2	Long term (more than 12h per day)	15
20180521_0010	Zone 1	Short term (less than 1h per day)	0
20180521_0010	Zone 1	Long term (more than 12h per day)	21
...			

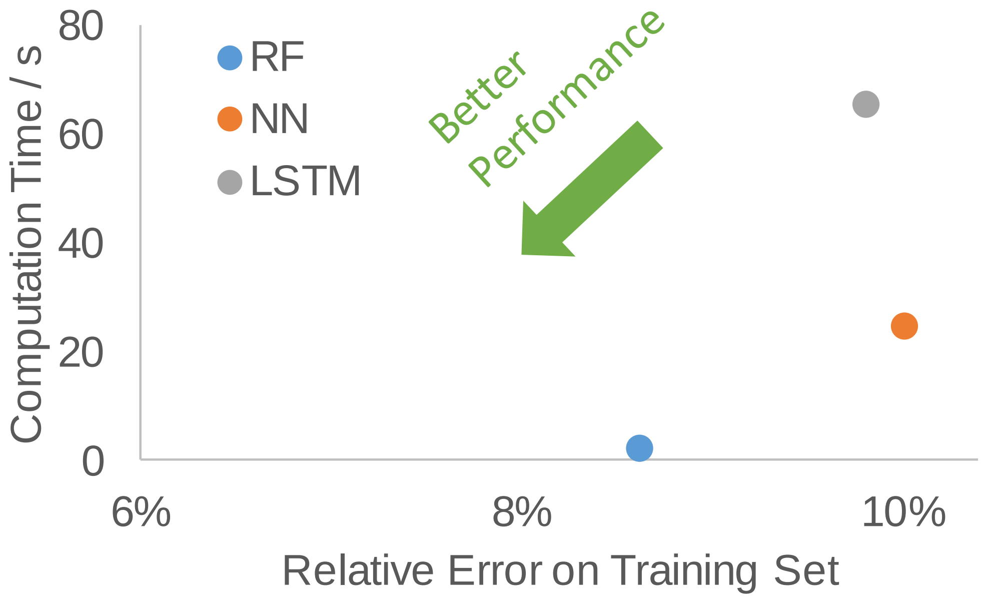
(b) data input to the machine learning algorithm

Method - Algorithm comparison

- Three types of algorithm have been compared
  - *Random Forest*
  - *Deep Neuron Network (DNN) - based regression*
  - *Recurrent Neuron Network: Long Short Term Memory*
- Two metrics
  - Accuracy: CV(RMSE)
  - Computational complexity
- Testbed
  - An office building in Berkeley, CA

Results – Algorithm comparison

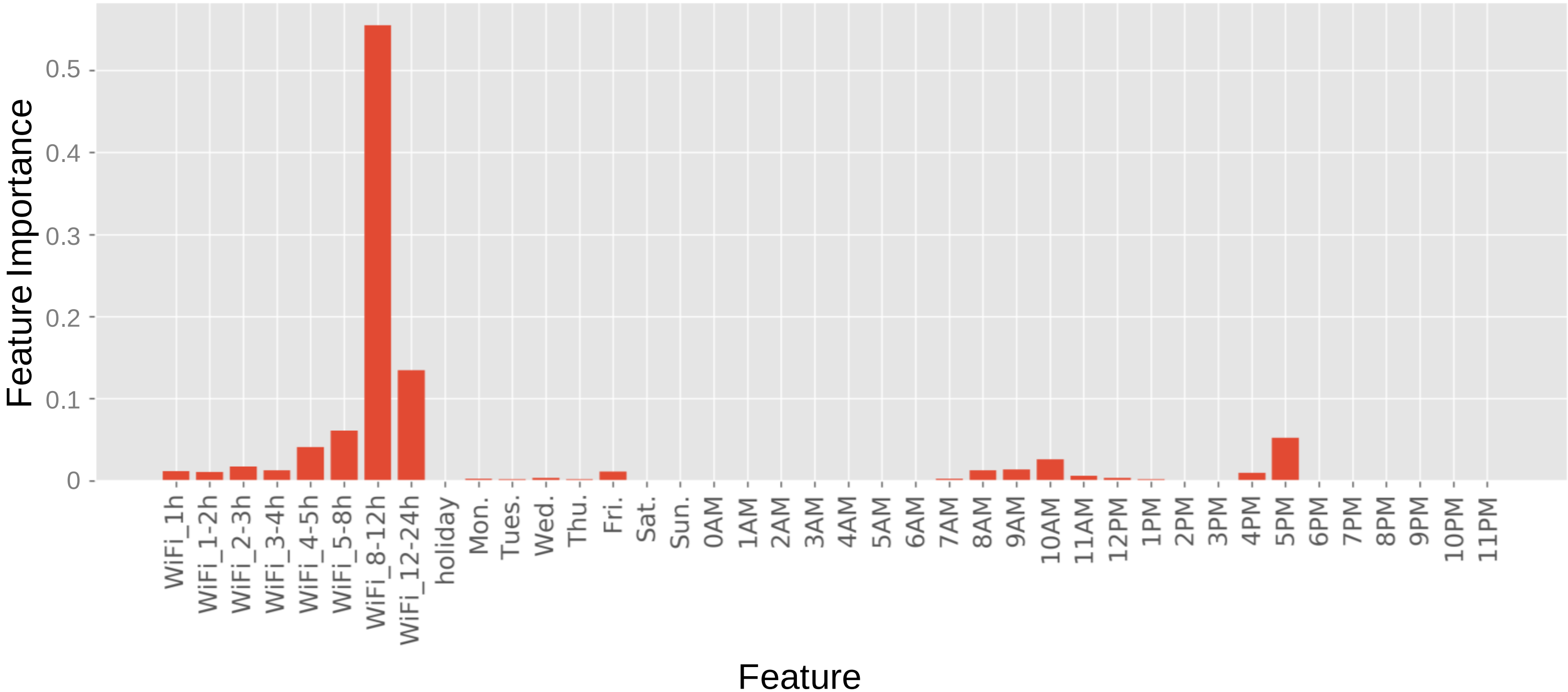
- Random Forest outperforms the other two
- The *sequential information* does not really help



	Random Forest (RF)	Neural Network (NN)	LSTM
RMSE on the training set	1.20	2.63	2.21
RMSE on the testing set	3.95	4.62	4.52
Computation time[a]	2.38s	24.86s	65.61s

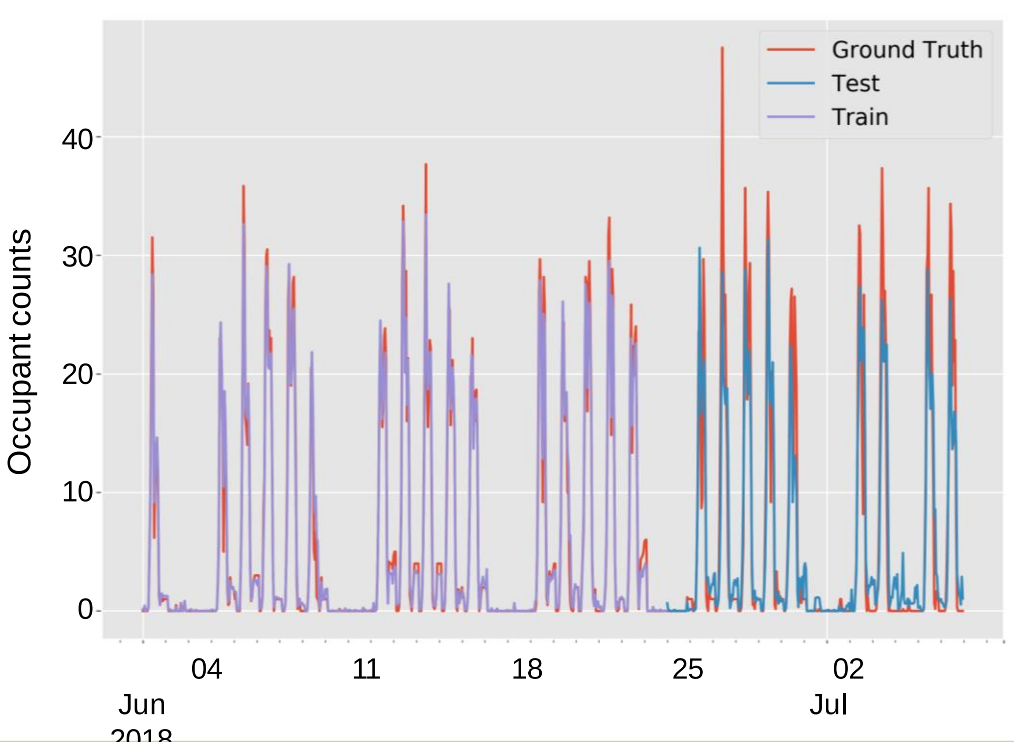
Results - Feature importance

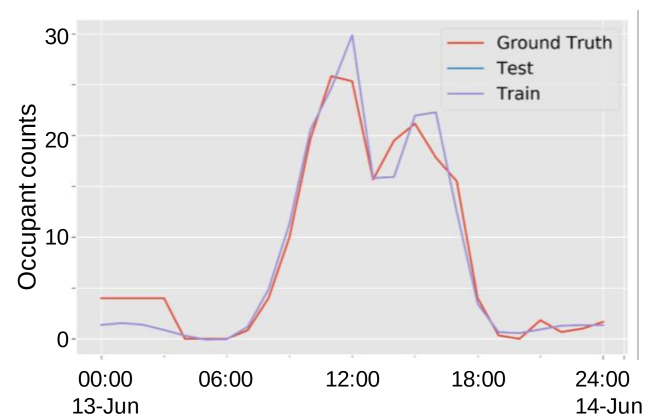
- Connection counts of *long time connected devices (8-12 h)* are the most important features
- Time-related features are less important than WiFi-related features

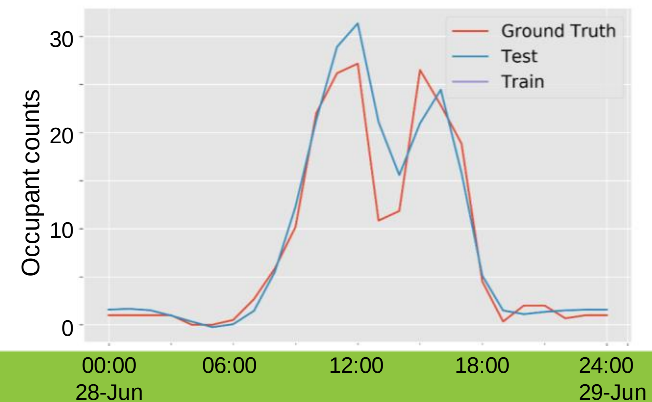


Results - Accuracy

- RMSE is 4 in a space with average occupancy of 22–27 people and peak occupancy of 48–74 people
- Delivering competitive results compared with other approaches







Conclusions

- Inferring occupant counts are important and challenging
- We proposed a new approach to enhance occupant counts detection through *feature engineering*
- This feature engineering approach and different ML algorithms have been tested in an office building testbed
- Our approach is non-intrusive and accurate

Bibliography

Wang, Z., Hong, T., Piette, M.A. and Pritoni, M., 2019. Inferring occupant counts from Wi-Fi data in buildings through machine learning. Building and Environment. 158: 281-294

Questions?

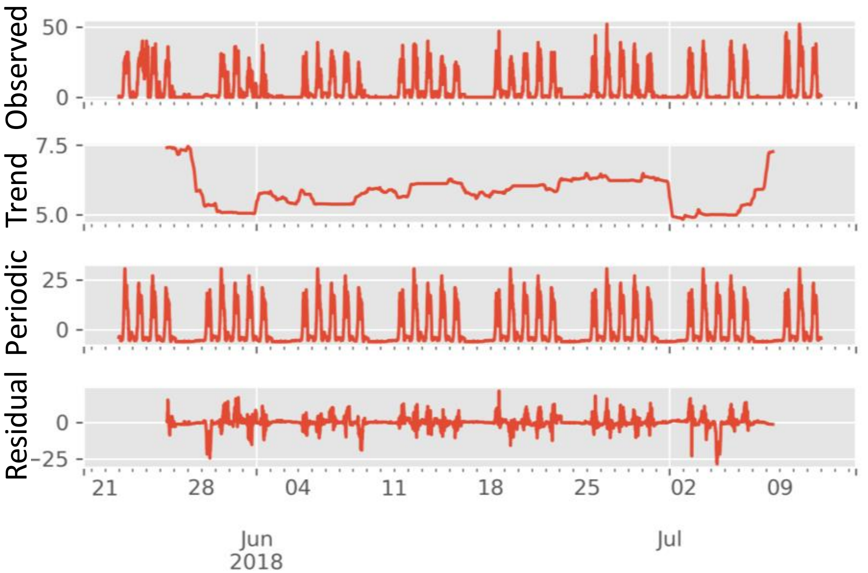
Zhe Wang

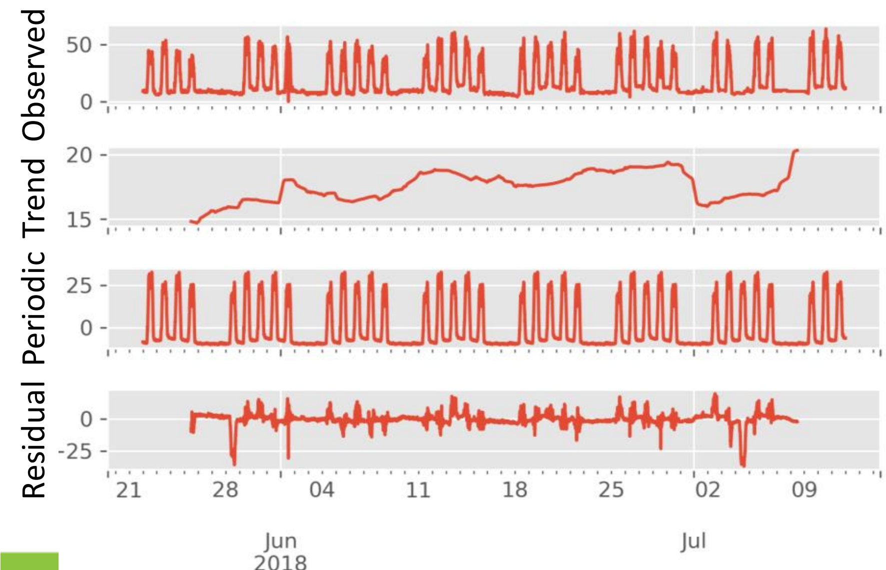
zwang5@lbl.gov

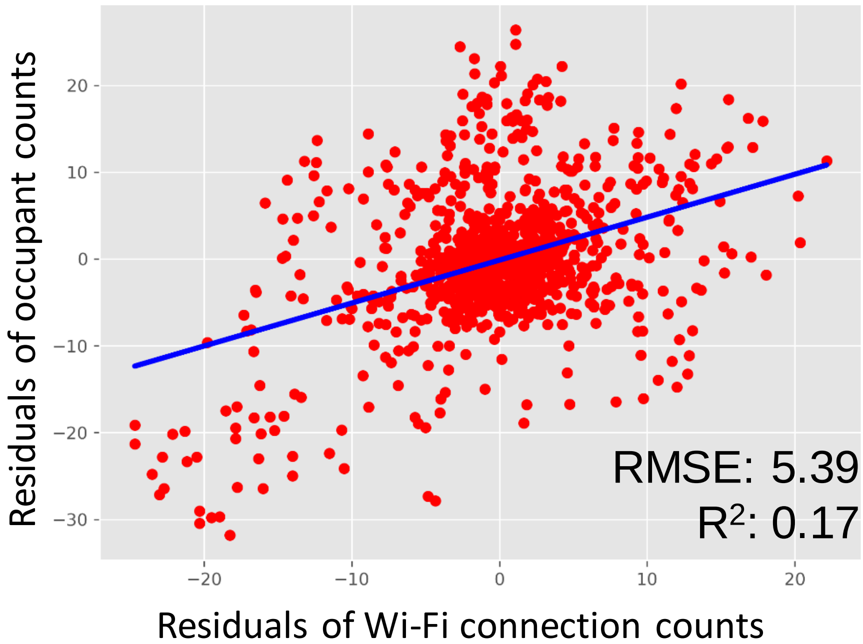
Tianzhen Hong

thong@lbl.gov

Backup slides: time series decomposition







The information of WiFi counts alone is inadequate to predict occupant counts