LSTMs For Indoor Sensors

Colin Warn, Meghna Dutta, Julianna Froderberg, Alireza Ghods & Aaron Crandall
Gerontechnology-Focused Summer Research Program (GSUR)

Motivation

- Predictive models that forecast future indoor building status can help organizations mitigate their carbon footprint and efficiently allocate resources.

Hypothesis

- Long Short-Term Memory (LSTM)[1] computing algorithms can be used for accurately forecasting next-day indoor sensor values at mixed use, high occupancy facilities.

Methodology

- Datasets were collected using the CASAS Smart Home in a Box[2] system of sensors, and from NOAA data. Dataset sizes ranged from a few to several years in length. Facilities used in study included Frank Innovation Zone, Palouse Discovery Science Center.
- The Keras LSTM library was implemented as the machine learning model. The model took in the past 7-days (lookback) worth of multivariate sensor data and predicted the following day's value for each sensor in the facility being modeled.
- LSTM model hyperparameters were as follows: A single LSTM layer with 700 memory units. Activation function was 'tanh'. This was followed by a single dense layer. Optimizer was Adam. Loss was calculated using 'Mean Squared Error'.

Example Input/Prediction Data Row

Date	Outdoor (°C)	Indoor (°C)	Activity Level	Zigbee Level
2017-04-06	19.16	25.29	1096.0	892.0

- ➢ Activity Level – Number of overall building motion sensor activations per day.
- ➢ Zigbee Level – Number of sensor reports. Describes how many sensors are online.
- ➢ Indoor Temperature – Daily mean building indoor temperature in degrees Celsius.
- ➢ Outdoor Temperature – Daily mean NOAA Pullman outdoor temperature in degrees Celsius.

Data Exploration

Datasets were explored to extract interesting features for our model. An example dataset exploration for one of our facilities, the Frank Innovation Zone, is shown below.

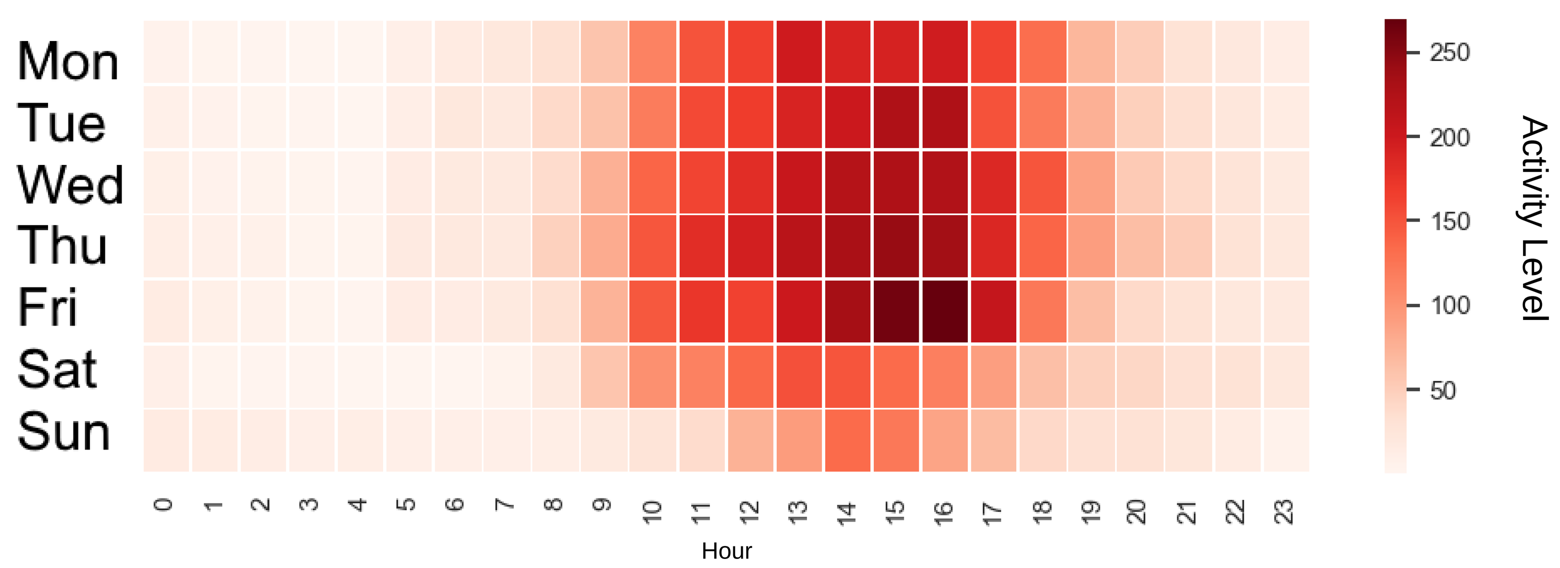

Figure 1. *Activity level as a function of weekday and hour. Day of the week and hour are strong predictors of activity levels.*

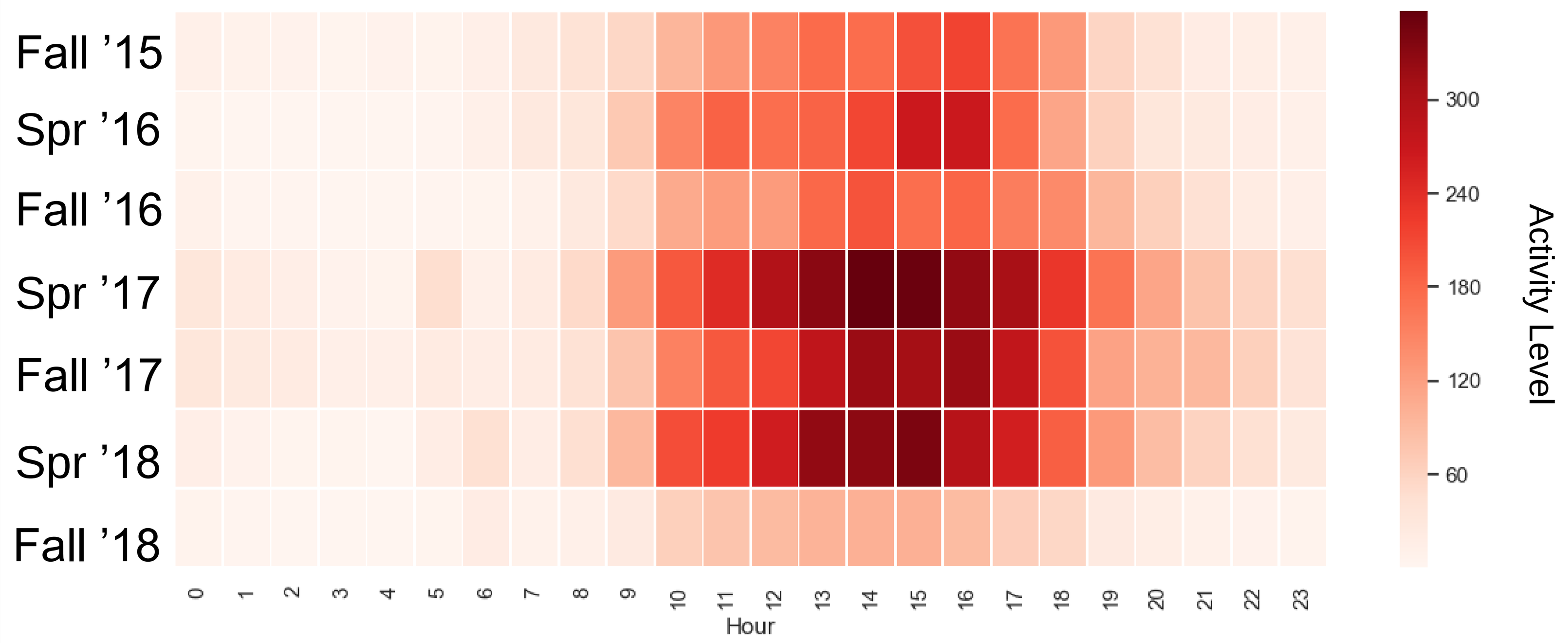

Figure 2. *Activity level as a function of semester and hour. School semester and hour are strong predictors of activity levels. Fall 2018 had multiple sensors offline.*

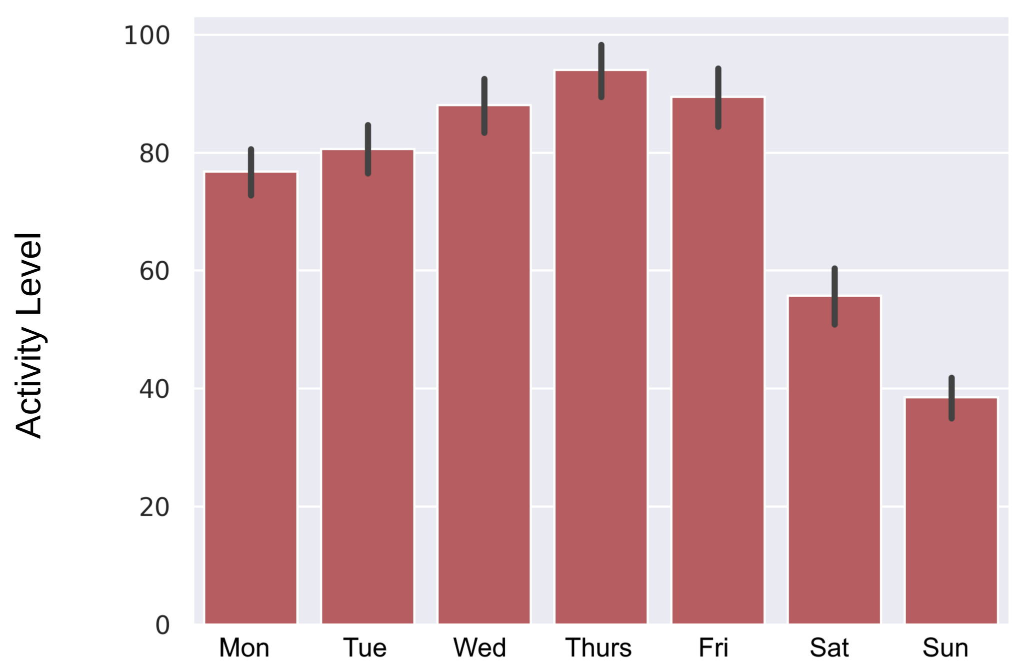

Figure 3. *Activity level as a function of weekday. Weekdays experience higher activity levels than weekends. Error bars are one standard deviation.*

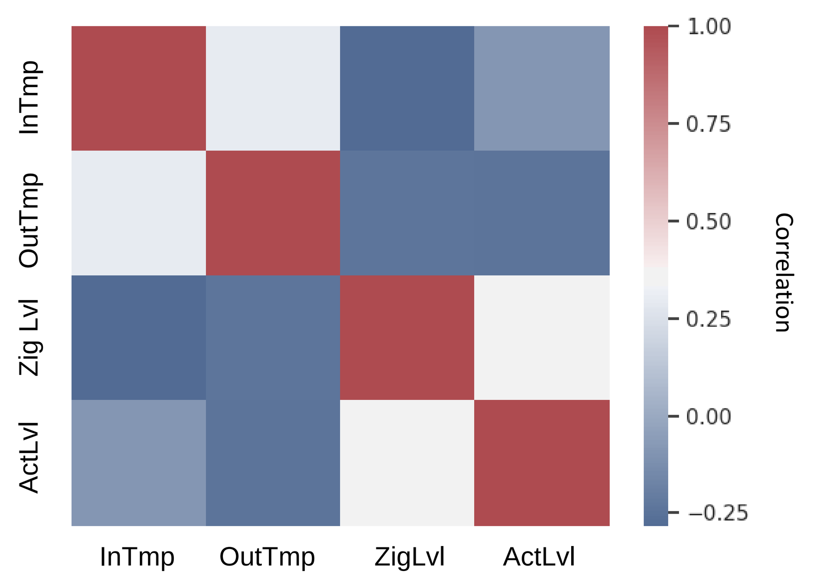

Figure 4. *Feature correlation matrix. Outdoor temperature was modestly correlated with indoor temperature, and Zigbee level did well in determining if activity level was low due to sensors being dead.*

For simplicity, our model features just included activity levels, Zigbee levels, indoor and outdoor temperatures.

Results

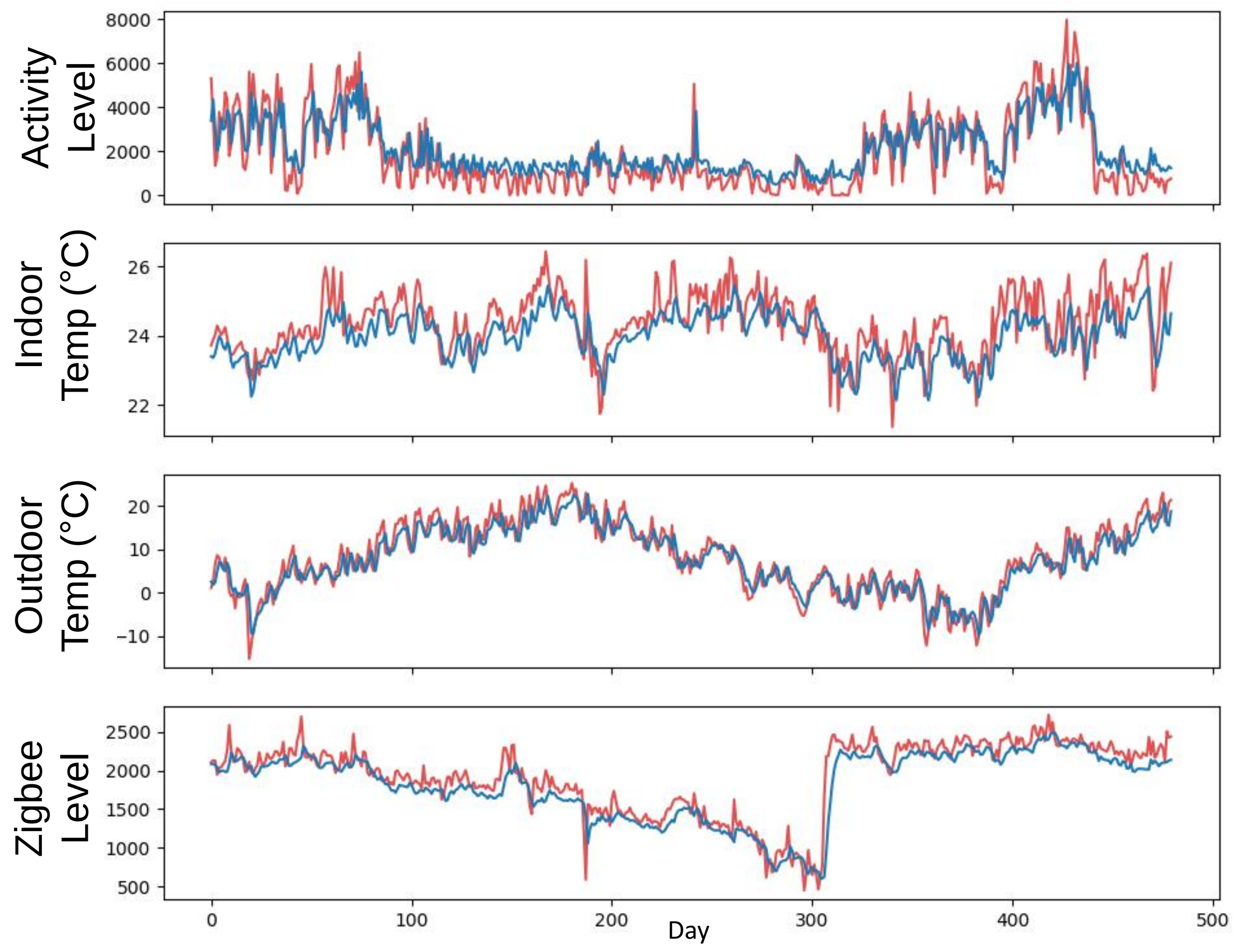

Figure 5. *Test set graph of next-day predicted values from plotted over actual value. Red signifies actual value; blue signifies predicted value.*

Variable	RMSE Score
Activity Level	7%
Indoor Temperature (°C)	4-7%
Outdoor Temperature (°C)	6-8%
Zigbee Level	3-5%

Figure 6. *Percent error for a given next day sensor prediction. Score was calculated using the following formula:*

$$RMSE\ Score = \frac{Normalized\ RMSE}{Max\ -Min} * 100\%$$

Conclusions and Future Work

- LSTMs can be applied to predict certain next-day indoor sensor values with an accuracy of 4-8%.
- Future work should explore the effects of other features on model accuracy. Rainfall and snowfall levels were examples of some features that looked promising.

Acknowledgements and References

Thank you to the National Institute of Health, grant #R25AG046, for providing the funding for this research. Special thanks to Brian Thomas, Jason and Bryan Minor for their help in answering questions relating to our research.

[1] Gers, Felix A., Jürgen Schmidhuber, and Fred Cummins. "Learning to forget: Continual prediction with LSTM." (1999): 850-855
[2] Cook DJ, Crandall AS, Thomas BL, Krishnan NC. CASAS: A Smart Home in a Box. *Computer (Long Beach Calif)*. 2013;46(7):10.1109/MC.2012.328. doi:10.1109/MC.2012.328