Poster Abstract: Investigating Occupancy Profiles Using Convolutional Neural Networks

Preprint

June Young Park,[1] Jianli Chen,[1] Xin Jin,[1] and Zoltan Nagy[2]

1 National Renewable Energy Laboratory
2 University of Texas at Austin

Presented at the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BuildSys 2019)
New York, New York
November 13–14, 2019

NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated by the Alliance for Sustainable Energy, LLC

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Contract No. DE-AC36-08GO28308

Conference Paper
NREL/CP-5500-74892
March 2020

Poster Abstract: Investigating Occupancy Profiles Using Convolutional Neural Networks

Preprint

June Young Park, [1] Jianli Chen, [1] Xin Jin, [1]
and Zoltan Nagy[2]

1 National Renewable Energy Laboratory
2 University of Texas at Austin

Suggested Citation
Park, June Young, Jianli Chen, Xin Jin, and Zoltan Nagy. 2020. *Poster Abstract: Investigating Occupancy Profiles Using Convolutional Neural Networks: Preprint*. Golden, CO: National Renewable Energy Laboratory. NREL/ CP-5500-74892. https://www.nrel.gov/docs/fy20osti/74892.pdf.

NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated by the Alliance for Sustainable Energy, LLC

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Contract No. DE-AC36-08GO28308

Conference Paper
NREL/CP-5500-74892
March 2020

National Renewable Energy Laboratory
15013 Denver West Parkway
Golden, CO 80401
303-275-3000 • www.nrel.gov

NOTICE

This work was authored in part by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding provided by the U.S. Department of Energy Office of Energy Efficiency and Renewable Energy Solar Technologies Office. The views expressed herein do not necessarily represent the views of the DOE or the U.S. Government.

Cover Photos by Dennis Schroeder: (clockwise, left to right) NREL 51934, NREL 45897, NREL 42160, NREL 45891, NREL 48097, NREL 46526.

Poster abstract: Investigating occupancy profiles using convolutional neural networks

June Young Park
juneyoungpark@utexas.edu
NREL, UT-Austin

Jianli Chen, Xin Jin
{jianli.chen;xin.jin}@nrel.gov
NREL

Zoltan Nagy
nagy@utexas.edu
UT-Austin

ABSTRACT

In this paper, we implement a convolutional neural networks (CNN)-based autoencoder to investigate occupancy profiles. We used American Time Use Survey (ATUS) data, which contained 191,558 schedules with binary occupancy information. Our results suggest that the trained filters provide an important insight into occupancy profiles (i.e., dominant and distinct patterns), and the latent space compresses the profiles with representative information.

KEYWORDS

convolutional neural networks, autoencoder, occupant behavior

ACM Reference Format:
June Young Park, Jianli Chen, Xin Jin, and Zoltan Nagy. 2019. Poster abstract: Investigating occupancy profiles using convolutional neural networks. In *Proceedings of The 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BuildSys '19).* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3360322.3360989

1 INTRODUCTION

Occupancy information is essential for building simulation and control problems [4]. For this reason, researchers have recently explained the importance of data-driven approaches for occupant behavior modeling [1]. With a very large data set (e.g., [3]), extensive manual feature engineering is often required, which is cumbersome and error-prone. In this paper, we adopt a CNN-based autoencoder to investigate the ATUS data (191K occupants with 15-minute intervals). Our approach automatically: (1) calculates the convolution filters, which contain critical information of occupancy schedule; and (2) compresses the input schedule into a lower dimension with informative representations. We visualize the trained filters and demonstrate the procedure of our autoencoder with a test occupancy schedule.

2 OCCUPANCY SCHEDULE DATA SET

As noted, we used ATUS data [3], in which respondents reported their activities (e.g., sleeping, working, cooking, and so on) and the corresponding timeline (15-minute interval) for the previous day. We preprocessed the original data set (2013–2017) into an occupancy profile format. For example, if the responses are related to home activity, then we considered such activity to be present; otherwise, we considered responses to be absent. In brief, our data set contains 191,558 participant schedules as 96 data points with binary occupancy information (0: absence, 1: presence).

Table 1: Autoencoder architecture (Decoder is the mirror of encoder; the last layer uses a sigmoid activation function)

Layer	Details	Shape
Input	-	(1,96,1)
First convolution	ReLU with same padding	(1,96,48)
First pooling	Max pooling with (1,4)	(1,24,48)
Second convolution	ReLU with same padding	(1,24,24)
Second pooling	Max pooling with (1,4)	(1,6,24)
Encoded	-	(1,6,6)

3 CNN-BASED AUTOENCODER

To analyze our large amount of data, we employed a CNN-based autoencoder. Typically, an autoencoder is used for dimensionality reduction in an unsupervised manner, and it takes the same input and output data to learn important representations of data. Table 1 details our autoencoder architecture. It has two encoding and decoding layers with a bottleneck in the middle to reduce the dimension of the input data. This reduction could filter out the noise and preserve only important feature-related information. Generally, occupancy profiles have continuous presence and absence patterns, which can be compressed by the proposed bottleneck layer. Note that the dimension of the latent space is important and can be further studied for the various objectives of other autoencoders.

As a learning component, we used CNN [2], which is actively researched in computer vision. We investigated the trained filters in the convolution layers to understand the learning process. For example, we intentionally selected 1d convolution filters (1, 4) to investigate which subsequences are critical in occupancy schedules. We used a rectified linear unit activation function (ReLU) to consider nonlinearity in feature extraction and a sigmoid activation function

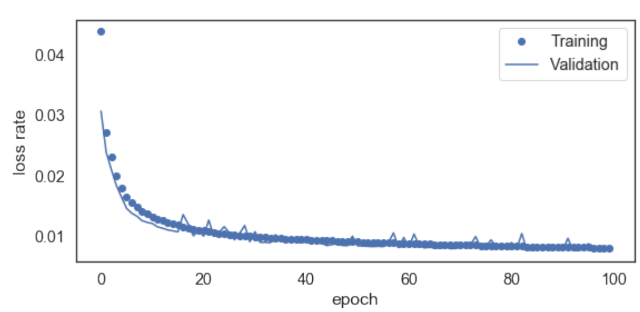

Figure 1: Training and validation loss

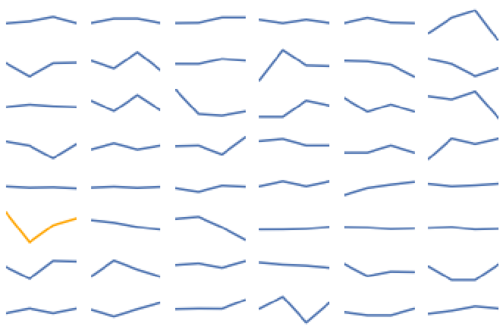

Figure 2: 48 trained convolution filters in the first layer

in the last decode layer. The RMSprop [5] was used to optimize the weights of the autoencoder with 100 epochs and a 128 batch size.

4 RESULTS

We assigned 191K schedules to the training (60%) and validation (20%) set. The rest (20%) are considered to be the final test set, which is not used in the training phase. Fig. 1 indicates that both the training and validation loss rates smoothly decreased. In addition, the overall error rate on the test set (41K schedules) is 0.0119.

Fig. 2 shows all 48 trained filters in the first convolution layer. The majority have a *relatively steady* shape. Note that the filter size is (1, 4), which is a 1-hour subsequence with a 15-minute interval. This suggests that the occupancy schedules have a number of 1-hour subsequences (i.e., either continuous presence or absence), and the autoencoder learns occupancy profiles by such filters.

To demonstrate the autoencoder process, we sampled one occupancy schedule from the test set. As shown in Fig. 3, the encoder compresses the input into the latent space, and the decoder generates the output with the same input dimension. The similarity

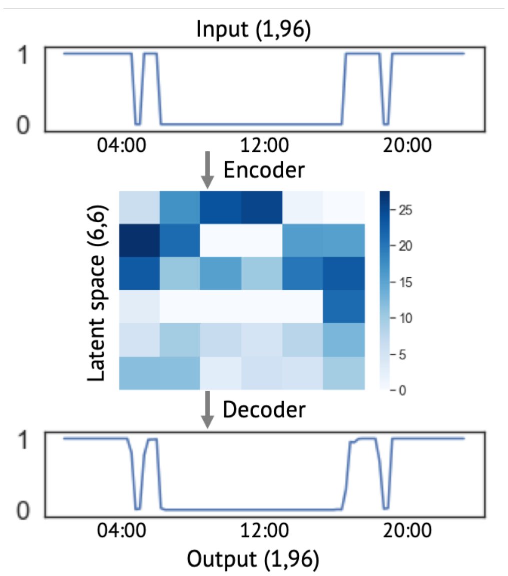

Figure 3: Input, output, and latent space of the test schedule

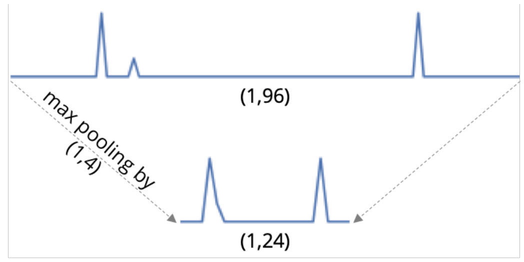

Figure 4: Feature map and pooling result of the test schedule with the highlighted filter in Fig. 2

between the input and output profile suggests that the latent space preserves the informative representation of the input.

We also detail the first convolution and pooling layer. The highlighted filter (orange) in Fig. 2 is very important, because it activates significantly to the next layer. Fig. 4 illustrates the feature map and pooling result from the sample input schedule. The filter has a *sharp down and up* shape and convolves similar subsequences in the input test schedule, resulting in two high peaks and one low peak in the feature map. Also, the max pooling compressed the feature map with the informative representation.

5 FUTURE WORK

The autoencoder could be improved by tuning the parameters. Also, we should evaluate the feature importance systematically. For applications, we could use the latent space for clustering schedules and apply the autoencoder to analyze various behavior types.

ACKNOWLEDGMENTS

This work was authored in part by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding provided by the U.S. Department of Energy Office of Energy Efficiency and Renewable Energy Solar Energy Technologies Office. The views expressed herein do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.

REFERENCES

[1] Mikkel Baun Kjærgaard, Bing Dong, Salvatore Carlucci, Flora D Salim, Junjing Yang, Clinton J Andrews, and Omid Ardakanian. 2018. Data-driven occupant modeling strategies and digital tools enabled by IEA EBC annex 79. In *Proceedings of the 5th Conference on Systems for Built Environments*. ACM, 188–189.

[2] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*. 1097–1105.

[3] U.S. Bureau of Labor Statistics. 2017. American Time Use Survey Releases. "https://www.bls.gov/tus/". Accessed: 2019-6-3.

[4] June Young Park and Zoltan Nagy. 2018. Comprehensive analysis of the relationship between thermal comfort and building control research—A data-driven literature review. *Renewable and Sustainable Energy Reviews* 82 (2018), 2664–2679.

[5] Tijmen Tieleman and Geoffrey Hinton. 2012. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. *COURSERA: Neural networks for machine learning* 4, 2 (2012), 26–31.