Development of an ANN-based Building Energy Model for Information-Poor Buildings Using Transfer Learning

Ao Li[1], Fu Xiao[1*], Cheng Fan[2], Maomao Hu[3]

[1] Department of Building Services Engineering, The Hong Kong Polytechnic University, Hong Kong, China

[2] Department of Construction Management and Real Estate, Shenzhen University, Shenzhen, China

[3] Oxford e-Research Centre Department of Engineering Science, University of Oxford, UK

Abstract

Accurate building energy prediction is vital to develop optimal control strategies to enhance building energy efficiency and energy flexibility. In recent years, the data-driven approach based on machine learning algorithms have been widely adopted for building energy prediction due to the availability of massive data in building automation systems (BASs), which automatically collect and store real-time building operational data. For new buildings and most existing buildings without installing advanced BASs, there is a lack of sufficient data to train data-driven predictive models. Transfer learning is a promising method to develop accurate and reliable data-driven building energy prediction models with limited training data by taking advantage of the rich data/knowledge obtained from other buildings. Few studies focused on the influences of source building datasets, pre-training data volume, and training data volume on the performance of the transfer learning method. The present study aims to develop a transfer learning-based ANN model for one-hour ahead building energy prediction to fill this research gap. Around 400 non-residential buildings' data from the open-source Building Genome Project are used to test the proposed method. Extensive analysis demonstrates that transfer learning can effectively improve the accuracy of BPNN-based building energy models for information-poor buildings with very limited training data. The most influential building features which influence the effectiveness of transfer learning are found to be building usage and industry. The research outcomes can provide guidance for implementation of transfer learning, especially in selecting appropriate source buildings and datasets for developing accurate building energy prediction models.

Keywords: Building energy prediction; data-driven approach; transfer learning; neural network; information poor buildings

1. Introduction

Building construction and operations account for 36% of the global final energy use and 39% of energy-related carbon dioxide (CO_2) emissions in 2017 (IEA 2018). In China, the building sector represents nearly 16% of the total global final energy consumption in buildings (IEA 2015). In Hong Kong, buildings are responsible for over 90% of electricity use in 2017 (EMSD 2019), which are the primary users in power grids and significantly influence the supply-demand balance and grid reliability. Building energy prediction models are widely used in evaluating building design alternatives (Asadi et al. 2014), developing energy efficient optimal control and diagnosis strategies (Li and Wen 2014), and developing the demand and supply management in power grids (Xue et al. 2014). Building energy prediction is an essential basis for decision making towards reducing building energy consumption and CO_2 emissions as well as enhancing the building-grid eco-system.

Two basic approaches have been developed for building energy prediction, i.e., physical-based and data-driven approaches (Amasyali and EI-Gohary 2018). The physical modeling is mainly based on engineering methods and dedicated building energy simulation tools, which usually requires detailed information and in-depth understanding of the building and building energy systems. With the dramatic increase in the scales and complexity of modern buildings, the physical-based modeling approach become increasingly time-consuming and computationally heavy, which make it unfavorable for online optimization and control applications. As a result, the data-driven approach based on building operation data and intelligent machine learning algorithms is attracting increasing interests (Fan, Sun et al., 2019; Fan, Wang et al., 2019; Fan et al., 2017; Rahman et al., 2018). However, most research and applications on data-driven prediction mainly deal with information-rich buildings with sufficient high-quality building operational data to train the data-driven models (Amasyali and EI-Gohary 2018). Without sufficient training data, the performance of the data-driven models may significantly deteriorate, owing to under-fitting or local minimum (Goodfellow et al. 2016). For those information-poor buildings, such as new buildings with very little historical data, and existing buildings without installing advanced building automation systems (BASs), the development of data-driven energy prediction models remains a big challenge.

Transfer learning is a machine learning technique initially motivated by difficulties in accessing a large amount of training data for training models in some applications, such as image recognition and natural language processing (Silver et al. 2013). Transfer learning aims at applying knowledge gained in solving one problem (i.e., the source task) to a different but related problem (i.e., the target task) (Weiss et al. 2016). The concept of transfer learning and related algorithms have been widely used in many fields, including software engineering (Ma et al. 2012), voice processing (Hu et al. 2015), image processing (Li et al. 2014), and natural language processing (Hosseinzadeh et al. 2016). Hu et al. (2015) implemented transfer learning-based Logistic Regression (LR) classifiers to improve mispronunciation detection performance. The shared hidden layers of this neural network-based classifier for extracting useful speech features were pre-trained using training data. The new LR classifier streamlined training multiple individual classifiers separately by learning the common feature representation via the shared hidden layer. The proposed method showed a 7.4% improvement of the precision and recall rate than the conventional model. Shin et al. (2016) adopted transfer learning to fine-tune convolutional neural network (CNN) models pre-trained using natural image datasets and then used the tuned/target models to classify medical images. The target models retained the structure of the pre-trained CNN models and showed superior in detecting some diseases. Previous research on transfer learning has proved that, if appropriately implemented, it has the following advantages: (1) reducing the amount of training data required for the development of the target model; (2) saving time for constructing and training models; and (3) improving prediction performance.

In recent years, several studies have focused on applying transfer learning to data-driven building energy prediction considering insufficient training data and taking advantage of additional datasets from other buildings (Ribeiro et al. 2018; Hooshmand and Sharma 2019; Perera et al. 2019). Ribeiro et al. (2018) developed a neural network model for electricity consumption prediction of a newly built school (i.e., the target building) by transferring knowledge learned from the three years' data of the other four similar schools in the same area (i.e., the source buildings). There were only one-month data available in the target building. The authors proposed a novel transfer learning method, namely Hephaestus, to remove seasonal and trend effects, and prepare time-independent features as model inputs. The results showed that prediction accuracy increased by up to 11.2% using transfer learning compared with the model that was trained using the one month of data of the target building. Hooshmand and Sharma (2019) proposed a transfer learning-based framework for short-term electricity load forecasting. A CNN model was designed to predict the next 24 hours of electricity demand of target building using its energy consumption data in the past four weeks. The CNN model was pre-trained by public repository datasets which include data of 370 different buildings (i.e., the source buildings) and then fine-tuned by four months' data from the target building, which showed lower error than other baseline cases. Fan et al. (2020) proposed a transfer learning-based methodology for 24-hour ahead building energy demand prediction. A public-available dataset composed of more than 400 different buildings was used in this research. He tested the methodology by developing a pre-trained model using data retrieved from 80% of the buildings, and evaluating the pre-trained model on target domain formed by the other 20% buildings. The results showed that approximately 15% to 78% of prediction errors can be reduced by the transfer learning-based methodology, compared with standalone models.

It is observed that previous studies on transfer learning-based building energy prediction usually transferred the knowledge, such as model structures and parameters. The source dataset is usually composed of, based on engineering experience, one or a small number of buildings with similar building industry and scale and in the same climate area to the target building (Ribeiro et al. 2018), or a great number of buildings of different scales and types (Hooshmand and Sharma 2019; Fan et al. 2020). It is well known that building energy/electricity consumption is influenced by multiple factors, including building design, location, usage, weather conditions, occupant behavior, energy systems, operation strategies and etc. In principle, the more similar the source buildings are to the target building, the better the prediction results can be obtained. Selecting sources buildings in transfer learning based on this principle would require a substantial amount of information about the sources buildings and the target building. However, it is not well studied in previous research whether the knowledge learnt from source buildings, whose usages, scales and locations are different from the target building, are valuable and transferrable to the target building, for example, using datasets from residential buildings to pre-train the prediction model of a non-residential building. And there are few experiences and guides on how to select source dataset for pre-training building energy prediction models. To well leverage transfer learning for building energy prediction model development, the guides on how to select dataset for pre-training models are in urgent need.

This study aims to develop an ANN-based building energy model by transferring the knowledge about energy consumption learned from information-rich buildings to buildings with limited operational data. As Keogh and Kasetty (2003) highlighted, the contribution of many studies would have been dwarfed by the variance that would have been observed by testing on many real-world datasets, or the variance that would have been observed by changing minor (unstated) implementation details. Using a large, consistent benchmark dataset to test different data-driven and machine learning algorithms can generate more generic comparisons of accuracy, speed, and ease-of-use, and consequently provide more convincing general conclusions. This paper tests implement transfer learning on one open-source benchmarking dataset, the Building Data Genome Project (Miller and Meggers 2017b). Annual energy consumption data of total 404 buildings are used as source data separately. The knowledge about the Back Propagation Neural Network (BPNN) models developed using the data from 404 buildings is transferred to the BPNN models developed for the target buildings, which are assumed to have only several days' operation data. The influence of different building factors on the performance of transfer learning is discussed with the aim to share experiences and provide useful guide in selecting proper source buildings and source datasets in using transfer learning for building energy prediction.

The remaining part of the paper is constructed as follows. Chapter 2 introduces the research outline and gives a brief overview of transfer learning. Chapter 3 presents the performance of ANN-based building energy consumption prediction models, and analyses the influence of different factors (pre-train data volume, train data volume, and building features). Chapter 4 concludes the paper.

2. Methodology

2.1 Research outline

The prediction task in this research is to predict one-hour ahead building energy consumption by using previous 24-hour energy consumption data. This type of prediction was widely used in building energy management and building-grid interactive management (Zhao and Magoules 2012). The benchmark database used in this research is developed by the Building Data Genome Project (Miller and Meggers 2017b). It contains hourly electrical power consumption data and weather data of more than four hundred non-residential buildings in America and Europe. The buildings in this database are mainly college and primary/secondary school buildings. Brief building information is also recorded, including its area, primary use type (e.g., office, laboratory, classroom and dormitory) and building industry (e.g., education and government). The datasets in this database have been used to predict building use type, performance class and operations strategy (Miller and Meggers 2017a), and load forecasting (Nichiforov et al. 2018).

As shown in Fig.1, the research methodology consists of the following parts: 1) Using the datasets from the information-rich buildings (i.e., source dataset) to pre-train a base model; 2) Using the training dataset from the information-poor building (i.e., target training dataset) to fine-tune the pre-trained model; 3) Using the test dataset from the information-poor building (i.e., target test dataset) to validate the target model. In principle, transfer learning can work with a diversity of predictive machine learning algorithms, like CNN (Hooshmand and Sharma 2019), RNN (Fan, Sun et al. 2020) and LR (Hu et al. 2015), this study selects back-propagation neural network (BPNN) to develop the prediction model considering its simplification and popularity in building energy prediction. To fulfill the aim of this research, i.e. to investigate the influences of source building type, scale and usage, pre-training data volume, and training data volume on the performance of the transfer learning, a large number prediction models (i.e., n in Fig. 1) are developed using different pairs of source and target buildings extracted from the Genome database. The influences of multiple factors concerned on transfer learning are then analyzed based on performance evaluation of all models.

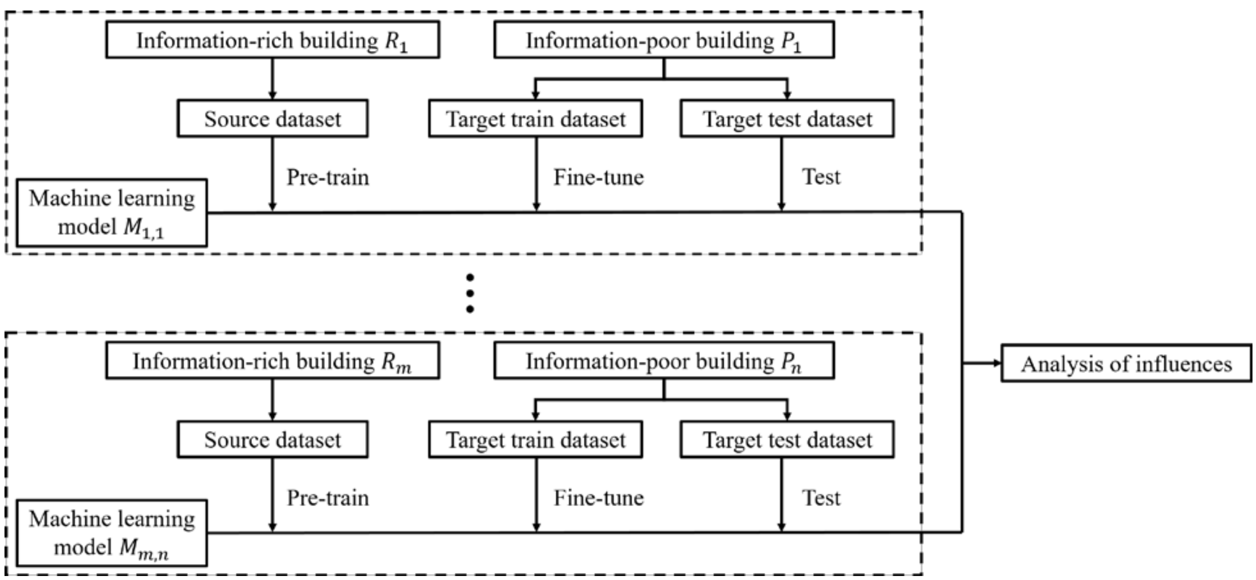


Fig.1 Research outline

Before the data are used to develop prediction models, they are pre-processed as shown in Fig.2, including data cleaning, data transformation and normalization. Data cleaning methods are implemented to detect outliers and fill missing values in order to improve data quality. After data cleaning, data transformation methods are needed to transform the data format, for example transferring numerical data (e.g., building type, location) into categorical data and prepare proper data format for the next stage, since some machine learning algorithms can only deal with categorical data. Data normalization is to eliminate the influence of the scales of data/measurement. The min-max normalization methods, as shown in Eq. (1), is adopted in this research to adjust the values of power consumption from buildings with different scales considering that larger buildings have larger values of power consumption.

$$y_i = \frac{x_i - min(x)}{max(x) - min(x)} \tag{1}$$

where $x = \{x_1, x_2, \dots, x_n\}$ and y_i is the normalized data of x_i.

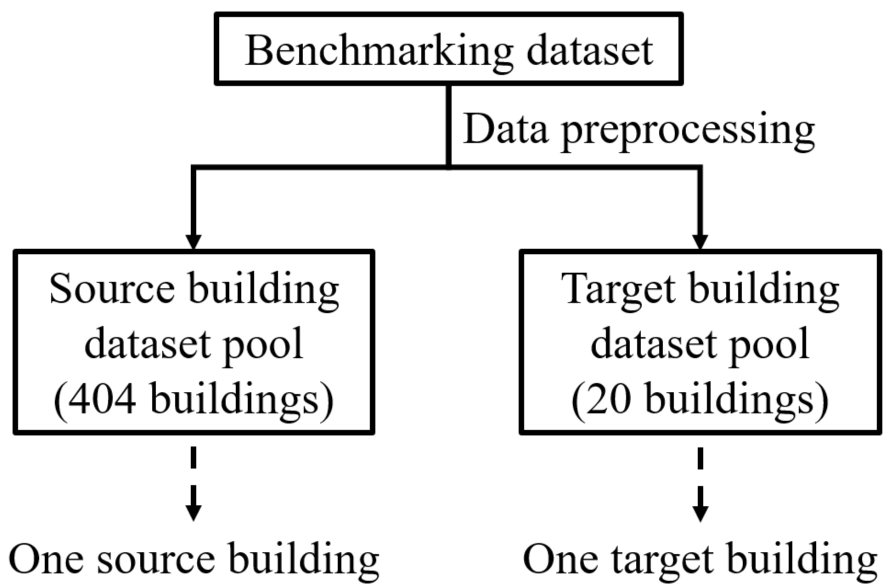


Fig.2 Data preparation

After data are pre-processed, 404 buildings in total are selected to form the pool of source buildings, which provide the source datasets. In addition, 20 buildings of different scales, primary usages, and weather conditions are selected as target buildings. Their information is shown in Table 1. These buildings are considered as information-pool buildings by assuming that only several days' energy data are available for developing the prediction models. For each case of developing prediction models, one source dataset is extracted from the source building dataset pool, while the target train dataset and target test dataset are extracted from the target building dataset pool. As a result, for each target building, 404 prediction models are developed corresponding to each source building.

Table 1. Information of 20 target buildings

Building type	Building No.	Location	Vintage	Outdoor Temperature (℃)	Relative Humidity (%)	Size (m^2)
Office (OF)	OF-1	Phoenix, USA	1930s	17.9	41.9	6,892
	OF-2	Los Angeles, USA	1960s	5.52	67.2	21,948
	OF-3	London, EU	Pre 1100s	13	76.8	122,223
	OF-4	Chicago, USA	NA	0.6	60.9	14,636
	OF-5	Zurich, EU	NA	5.3	84.1	9,244
Primary Classroom (PC)	PC-1	New York, USA	NA	11.2	73.7	8,070
	PC-2	New York, USA	NA	7.9	79.3	8,747
	PC-3	New York, USA	NA	10.4	74.7	21,652
	PC-4	New York, USA	NA	7.3	76.3	10,877
	PC-5	London, EU	Pre 1910s	12	86	1,597
Dormitory (DO)	DO-1	Phoenix, USA	1960s	17.9	41.9	10,340
	DO-2	Phoenix, USA	1930s	17.9	41.9	3,051
	DO-3	New York, USA	1970s	6.9	78.8	6,100
	DO-4	London, EU	1960s	13	76.8	6,181
	DO-5	Chicago, USA	NA	0.6	60.9	16,495
University Laboratory (UL)	UL-1	Phoenix, USA	1950s	17.9	41.9	8,480
	UL-2	Phoenix, USA	2010s	17.9	41.9	30,403
	UL-3	Los Angeles, USA	1960s	5.5	67.2	11,608

UL-4	New York, USA	1980s	6.9	78.8	6,930
UL-5	London, EU	1990s	13	76.8	7,715

Note: The average outdoor dry-bulb temperature and relative humidity within the sampling time period

2.2 Transfer learning

Transfer learning aims to improve the learner (e.g. the prediction model) in a target domain using the knowledge from other domains and learning tasks (Weiss et al. 2016; Pan and Yang 2009).

Definition. Given a source domain D_S and a learning task T_S, a target domain D_T and a learning task T_T, transfer learning aims to improve the learning of the target predictive function $f_T(\cdot)$ in D_T using the knowledge in D_S and T_S, where $D_S \neq D_T$, or $T_S \neq T_T$ (Weiss et al., 2016).

From the definition above, a domain D is defined as a pair $D = \{\chi, P(X)\}$, where χ is a feature space with n dimensions; X is a learning sample such that $X = \{x_1, ., x_n\} \in \chi$, and $P(X)$ is the marginal is the marginal probability distribution of X. A task T is defined as $T = \{Y, P(Y|X)\}$, where Y represents the label space and $P(Y|X)$ represents the conditional probability of Y given X. Given a source domain D_s (the information-rich buildings in this research) and a corresponding source task T_s (building energy consumption prediction), a target domain D_t (the information-poor building) and a target task T_t (building energy consumption prediction), transfer learning aims to learn the target conditional probability distribution $\mathrm{P}(Y_t|X_t)$ in D_t with the help of knowledge learnt from D_s and T_s, where D_s and D_t, T_s and T_t are not identical.

2.3 Artificial neural network

The back-propagation neural network (BPNN) was initially developed by Rumelhart et al. (1986) as a solution to the problem of training multi-layer perceptron (MLP). In the area of building energy consumption prediction, BPNN is one of the most popular neural network models because of its simple architecture yet powerful problem-solving ability (Bourdeau et al. 2019). The prediction task concerned in this study is to use the previous 24-hour building power consumption to predict the power consumption in the next hour. Therefore, the input layer of the BPNN contains 24 nodes, corresponding to the previous 24-hour power consumption. The output layer contains one node, i.e., the next hour power consumption. Considering the volume of the source and target dataset, a four-layer BPNN model structure was found to be suitable. The first hidden layer contains 24 nodes, and the second hidden layer contains 12 codes. Rectified Linear Unit (ReLU) and Adam algorithm (Kingma and Ba 2014) are selected as the activation function and optimizer respectively. The rectifier is an activation function defined as the positive part of its argument, as shown in Eq. (2).

$$f(x) = \max(0, x) \quad (2)$$

ReLU was first introduced to develop a dynamical network by Hahnloser et al. in 2000 with strong biological motivations and mathematical justifications (Hahnloser et al. 2000). ReLU is the most popular activation function for deep neural networks in 2017 (Ramachandran et al. 2017). Adam is an adaptive learning rate optimization algorithm which finds individual learning rates for each parameter (Kingma and Ba 2014). It has several attractive benefits, including easy-to-implement, computationally efficient and little memory requirements. And this optimization algorithm released in Dec 2014 has been used in around 23% of papers according to a survey in 2017 (Karparthy 2017). The loss function used in network work training is the mean squared error.

2.4 Implementation strategies for transfer learning

Network-based transfer learning is one category of transfer learning (Weiss et al. 2016). The usual network-based transfer learning approach is to train a base neural network and then copy its first n layers to the first n layers of a target network. The remaining layers of the target network are then randomly initialized and trained using the target train dataset in the target task. There are two methods to handle the first *n* layers in the target network (Yosinski et al. 2014). The first method is to back-propagate the errors in the target network/model into the base (copied) features to fine-tune them to suit the target task. The second method is to freeze the transferred layers, which means that they do not change during training in the target task. Usually, the selection of these two methods depends on the overfitting problem (Yosinski et al. 2014). In this research, the model is fine-tuned using the former way, considering that the number of features is rather small so that overfitting is not a problem.

After fine-tuning, the model will be validated by the target test dataset. It is assumed that transfer learning works well when the source domain is related or similar to the target domain because the knowledge learned from the source domain may be applicable to the target domain due to the relevance or similarity. However, when the source domain is not closely related or similar to the target domain, the target model could be negatively impacted, which is the so-called negative transfer (Weiss et al. 2016). The consequence of negative transfer is that the target model's prediction accuracy is lower than that of the base model trained by using only the target building dataset. The negative transfer ratio is defined as shown in Eq. (3).

$$\text{Negative transfer ratio} = \frac{Number\ of\ transfer\ learning\ cases}{404\ (the\ number\ of\ source\ building\ pool)} \times 100\% \tag{3}$$

3. Results and discussions

As explained in 2.1, a pair of one source building and one target building is used in each prediction task, and 8080 (404×20) tasks in total are performed. The results are analyzed and compared to obtain more generic conclusions regarding the applicability and performance of transfer learning in building energy prediction, taking into account the volumes of source data and target data, different building locations, scales, and usages. Performance indicators used in this research include Mean Absolute Percentage Error (MAPE) and Mean Square Error (MSE). MAPE expresses average absolute error as a percentage, whereas MSE measures the average of the squares of the errors. MAPE and MSE are calculated as by Eq. (4) and Eq. (5), respectively.

$$\mathrm{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{4}$$

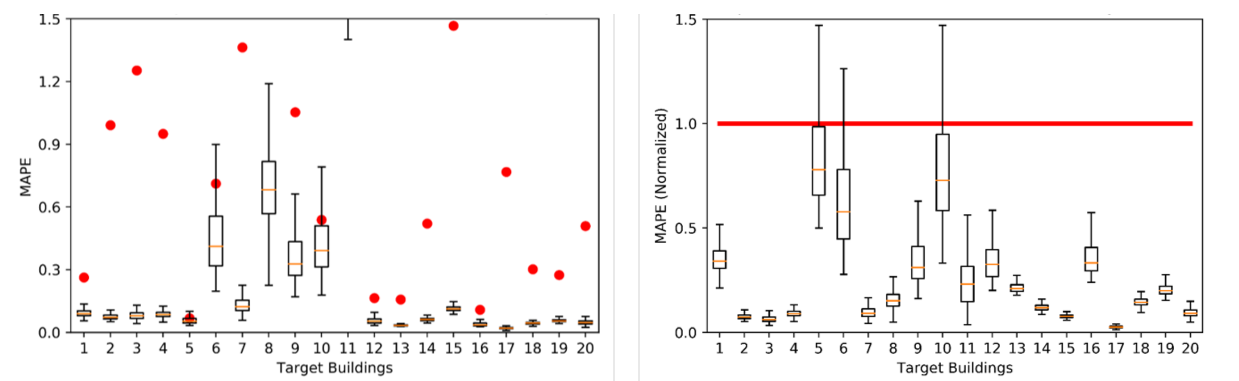
$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{5}$$

where y_i is the actual energy consumption, $\hat{y}_i$ is the predicted energy consumption, and N is the number of observations.

3.1 The effects of transfer learning

To test the effects of transfer learning on the model performance, the data volumes of source dataset for pre-training, target train dataset and target test dataset are set at 20, 2 and 20 days, respectively, the performance of the 8084 models developed for 20 target buildings without and with transfer learning was compared in Fig.3. The horizontal axis represents the 20 target buildings. The red points in Fig.3-a represent the accuracy of models developed using only the target 2-day train dataset and without using transfer learning, which serves as the baseline cases. In most baseline cases, the models' performance is undesirable, with MAPE higher than 50%. The boxplots represent the performance of the models of the 20 target buildings, which are pre-trained by source datasets from 404 sources buildings and then trained/fine-tuned by the target train data by adopting transfer learning. Each boxplot (or vertical column in the figure) represents the 404 models of the corresponding target building (shown on the x-axis). Remarkable improvements in prediction accuracy can be seen after applying transfer learning. For every target building, the average model performance of all 404 cases is calculated and shown in Table.2. For 13 target buildings, the average MAPE is lower than 10%, and the average MAPE for target building 7 and 15 are 13% and 12% respectively, which are very close to 10%.

To better investigate the influences of transfer learning, the prediction accuracy of cases using transfer learning is normalized by the accuracy of the corresponding baseline case. Fig.3.b provides the normalized results. As can be seen from the figure, most source and target building pairs can help improve the prediction performance. And the negative transfer phenomenon only occurs for three target buildings. These results suggest that when the available training data is somewhat limited, the mechanism of transfer learning works well in most source-target building pairs. In such cases, the effects of weight initialization outperform the impact of data distribution differences.



(*a*) (*b*)

Fig.3 (*a*) Model performance of different source-target building pairs

(*b*) Normalized model performance (by baseline cases)

Table 2. Model performance with and without using transfer learning

Target building	Performance without transfer learning (MAPE)	Performance with transfer learning (MAPE)
1	26%	9%
2	99%	8%
3	125%	8%
4	95%	9%
5	7%	6%
6	71%	46%
7	136%	13%
8	447%	7%
9	105%	37%
10	54%	42%
11	375%	929%
12	16%	6%
13	16%	3%
14	52%	6%
15	147%	12%
16	11%	4%
17	77%	2%
18	30%	44%
19	27%	6%
20	51%	5%

3.2 The effects of train data volume

When the available data from target building increases, the model performances in terms of normalized MAEP are shown in Fig.4. The twenty black polylines in the figure represent the average MAPE (normalized by baseline cases) of all source-target building pairs for twenty target buildings. And the green line shows the overall average prediction accuracy of the 8040 models when using different train data volumes, as indicated on the x-axis. When the volume of training dataset increases, transfer learning still shows a positive contribution to the prediction task, but the proportion of improvement, which is the model using transfer learning over the model without using transfer learning, shows a gradual decline. Fig.5 presents the negative transfer ratio with different train dataset volume. An overall slight uptrend of the negative transfer ratio can be observed as the volume of the train dataset increases. The most likely cause of the aforementioned phenomenon is that, when the train data volume is small, the pre-trained dataset provides somewhat reasonable weights and biases for the neural network (the weight initialization process); when the train data volume increases, the difference of the data distribution between the source dataset and target dataset plays an increasingly important role, which negatively influences the prediction accuracy. In general, the less available training data, the better the effect of transfer learning.

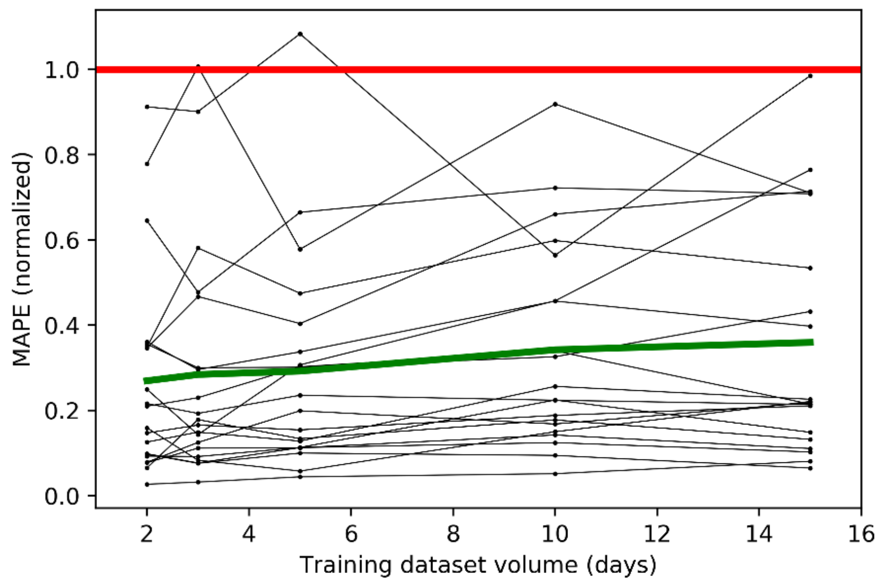


Fig.4 Effect of transfer learning with different train data sample

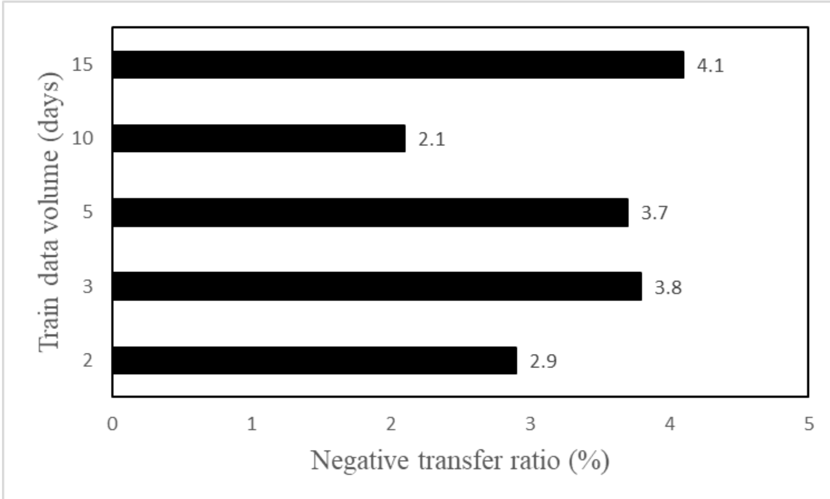


Fig.5 Negative transfer ratio with different train data sample

3.3 The effects of pre-train dataset volume

The impact of different pre-train data samples (train dataset are still retrieved come from 2 days) is studied in this chapter. The model performance, normalized performance by baseline cases and negative transfer ratio with 10 days, 20 days, and 30 days' pre-train data are presented in Table.3, Fig.6 and Fig.7, respectively. It can be seen that when the pre-train data sample increases from 10 days to 30 days, the average prediction accuracy is slightly increased, and the negative ratio shows a slight drop. The consistency of the source dataset may have been an important influential factor; thus, the model cannot learn much more new knowledge from the increased pre-train data.

Overall, this difference is rather small, compared to the performance difference with and without transfer learning. When the target building has limited training data, transfer learning can achieve great accuracy improvement without using much pre-train data. It is worth mentioning that this generalization of this conclusion remains uncertain.

Table 3. Model performance with different pre-train dataset volumes (train data sample: 2 days)

Target building	Without transfer learning	10 days	20 days	30 days

1	0.26	0.1	0.09	0.09
2	0.99	0.08	0.08	0.08
3	1.25	0.09	0.08	0.08
4	0.95	0.09	0.09	0.09
5	0.07	0.07	0.06	0.06
6	0.71	0.47	0.46	0.46
7	1.36	0.14	0.13	0.13
8	4.47	0.71	0.7	0.7
9	1.05	0.37	0.37	0.37
10	0.54	0.44	0.42	0.42
11	37.5	9.84	9.29	9.29
12	0.16	0.06	0.06	0.06
13	0.16	0.03	0.03	0.03
14	0.52	0.06	0.06	0.06
15	1.47	0.12	0.12	0.12
16	0.11	0.04	0.04	0.04
17	0.77	0.02	0.02	0.02
18	0.3	0.05	0.44	0.04
19	0.27	0.06	0.06	0.06
20	0.51	0.05	0.05	0.05

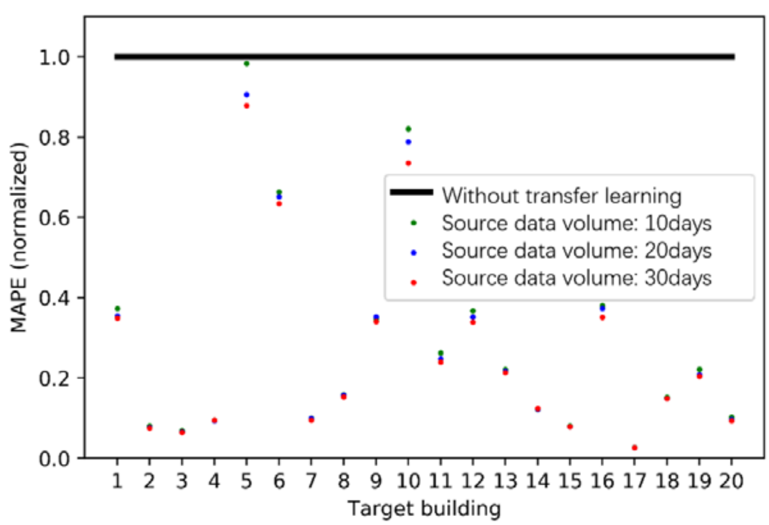


Fig.6 Model performance (MAPE) with different pre-train data volume (train data volume: 2 days)

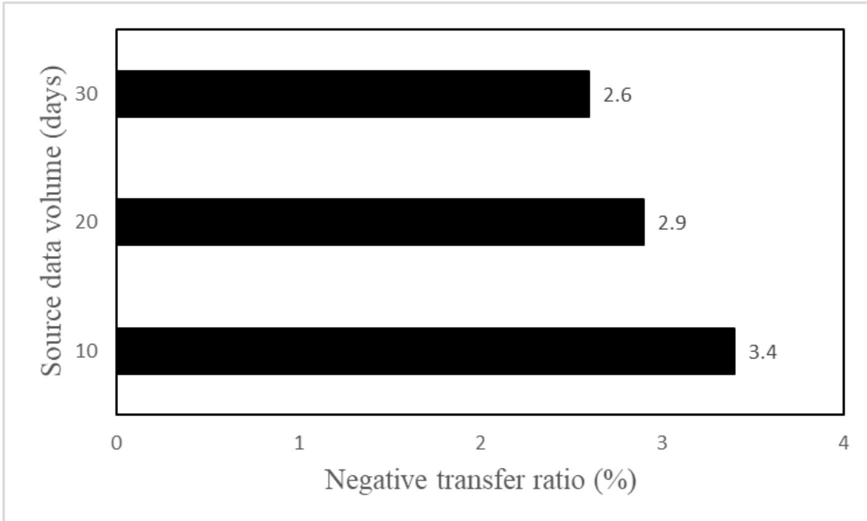


Fig.**7** Negative transfer ratio with different pre-train data volume (train data volume: 2 days)

3.4 The effects of different building features

In this section, the impact of different building features, including usage, industry, outdoor dry-bulb temperature, and relative humidity, on the performance of transfer learning, will be analyzed both individually and collectively.

In this study, for particular target building, the source-target building pair producing the most accurate model using transfer learning is considered the most suitable source-target building pair. In other words, the knowledge (i.e., the BPNN structure, as explained in Section 2.3) learned from the source building is most valuable in developing the BPNN model for the target buildings using transfer learning. The source-target building pairs producing the worst accurate model using transfer learning can be defined as the most unsuitable source-target building pair. The top 3 suitable and unsuitable source-target building pairs are summarized in Table.4 and Table.5 (the building names follow their usage). From Table.4, it can be found that for eight target buildings (marked bold), using their own data as the source data to pre-train the BPNN model can achieve the highest accuracy improvement. This finding validates the logical rationality of transfer learning's mechanism, as no other building has a more similar data distribution as the target building than its own.

For top3 unsuitable source-target building pairs, interestingly, the dormitory buildings take up a large share in Table 5. To dig out the underlying causes, the energy profiles of dormitory buildings are analyzed. Fig.8 provides several examples of the typical dormitory energy profile. The dormitory energy consumption patterns are very different from other types of buildings. The peak and valley hours of the load curve are in the mid-night and noontime, respectively. The knowledge learned from this type of energy profile can easily mislead the model for buildings with conventional load curves, whose peak and valley hours are at noon and night, respectively.

Table 4. Top3 suitable source-target building pairs

Target building	**OF-1**	**OF-2**	**OF-3**	**OF-4**	**OF-5**
Top1	UL-84	**OF-2**	**OF-3**	UL-73	UL-82
Top2	UL-68	UL-95	UL-68	UL-76	UL-64
Top3	**OF-1**	UL-88	UL-5	UL-65	UL-78
Target building	**DO-1**	**DO-2**	**DO-3**	**DO-4**	**DO-5**
Top1	OF-142	UL-85	UL-78	UL-76	**DO-5**
Top2	UL-73	UL-93	UL-70	UL-69	UL-65
Top3	UL-84	UL-61	UL-81	UL-74	UL-88
Target building	**PC-1**	**PC-2**	**PC-3**	**PC-4**	**PC-5**
Top1	UL-74	**PC-2**	**PC-3**	UL-74	UL-5
Top2	UL-90	PC-27	UL-74	UL-95	PC-93
Top3	UL-95	UL-65	OF-87	**PC-4**	PC-71
Target building	**UL-1**	**UL-2**	**UL-3**	**UL-4**	**UL-5**
Top1	UL-67	UL-95	UL-86	UL-93	**UL-5**
Top2	UL-80	UL-60	UL-5	UL-69	UL-68
Top3	UL-78	UL-68	UL-74	UL-62	UL-93

Table 5. Top3 unsuitable source-target building pairs

Target building	**OF-1**	**OF-2**	**OF-3**	**OF-4**	**OF-5**
Top1	DO-19	DO-56	DO-26	UL-54	OF-34
Top2	DO-49	DO-63	UL-54	OF-135	PC-75
Top3	PC-22	DO-8	PC-49	UL-4	OF-119
Target building	**DO-1**	**DO-2**	**DO-3**	**DO-4**	**DO-5**
Top1	DO-58	DO-20	DO-51	DO-66	OF-46
Top2	UL-24	DO-51	DO-58	UL-21	DO-17
Top3	DO-63	DO-49	DO-56	DO-51	DO-49
Target building	**PC-1**	**PC-2**	**PC-3**	**PC-4**	**PC-5**
Top1	UL-29	DO-37	OF-66	UL-55	DO-43
Top2	UL-4	DO-24	DO-43	OF-75	UL-10
Top3	UL-24	OF-23	DO-37	PC-73	OF-109
Target building	**UL-1**	**UL-2**	**UL-3**	**UL-4**	**UL-5**
Top1	UL-80	DO-58	DO-63	DO-43	DO-56
Top2	OF-116	DO-56	DO-56	OF-112	OF-33
Top3	OF-116	DO-39	UL-24	PC-8	OF-76

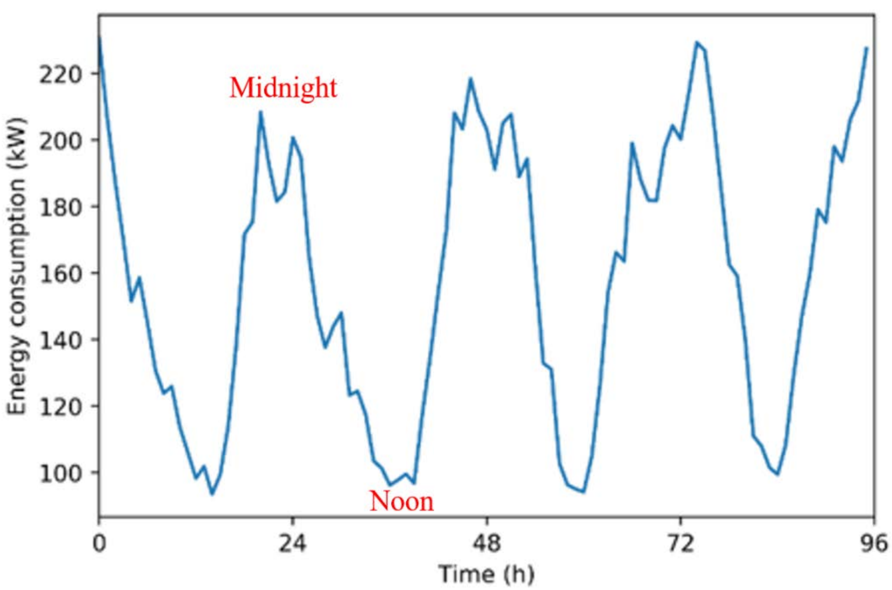


Fig.8 Typical dormitory energy consumption profile

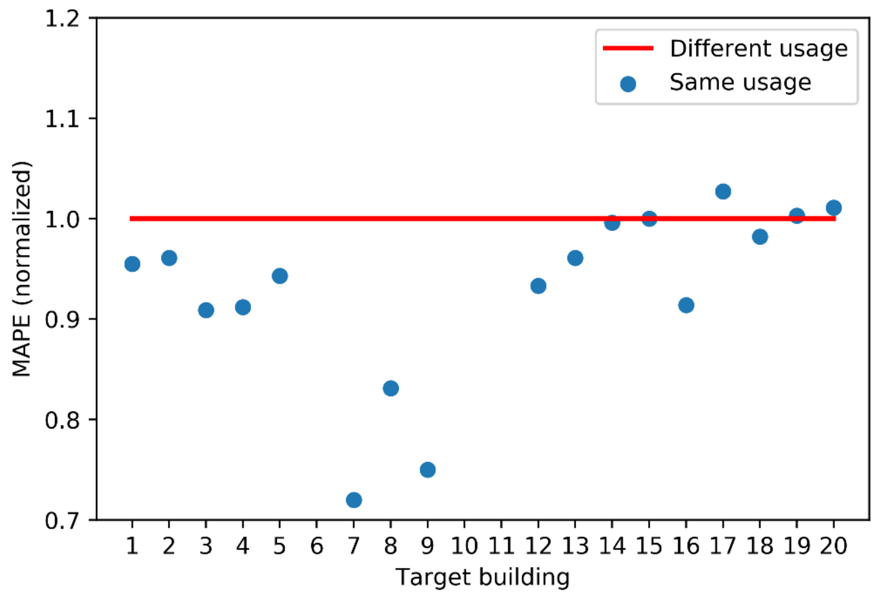


Fig.9 Impact of same/different usage between source and target building

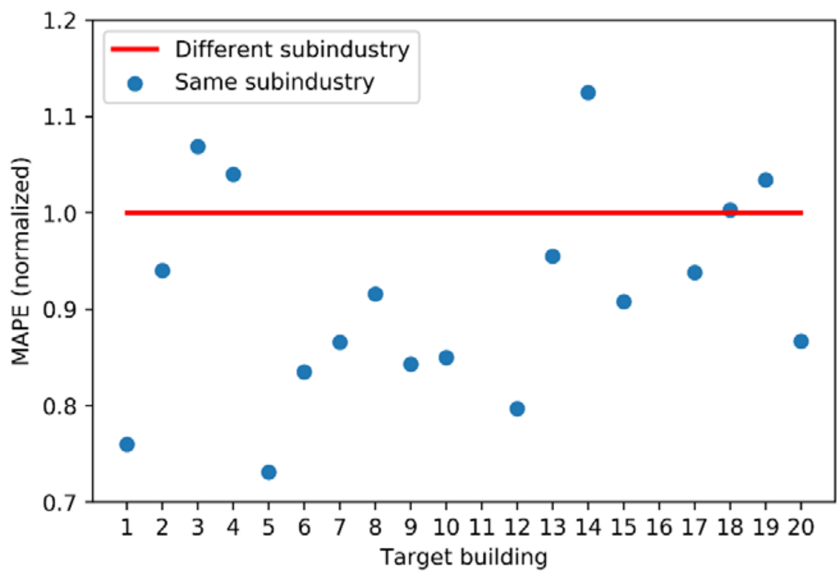
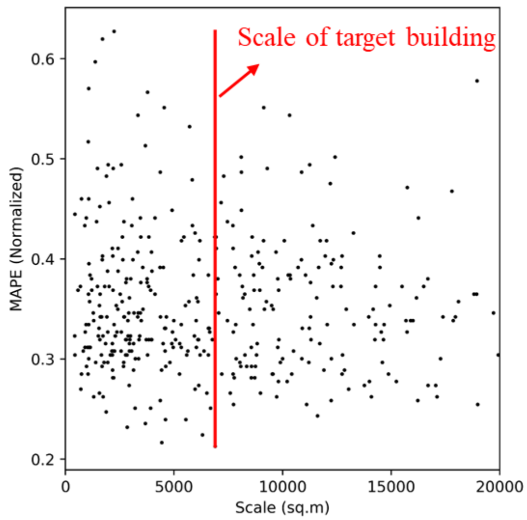
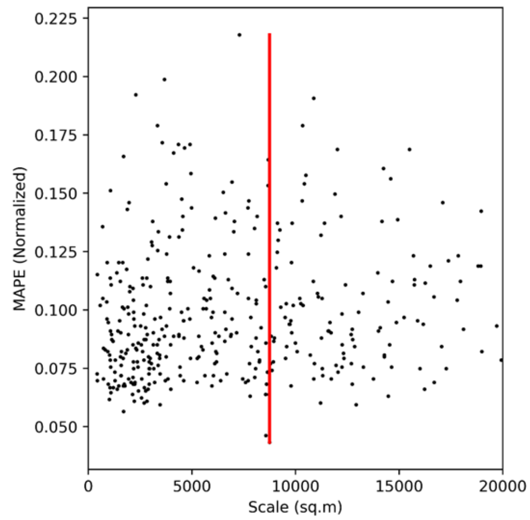


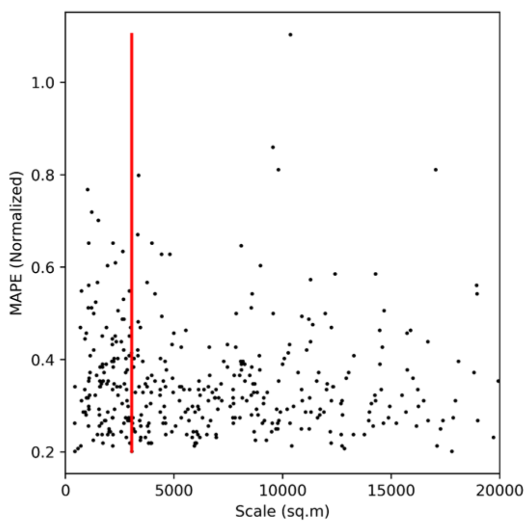
Fig.10 Impact of same/different subindustry between source and target building



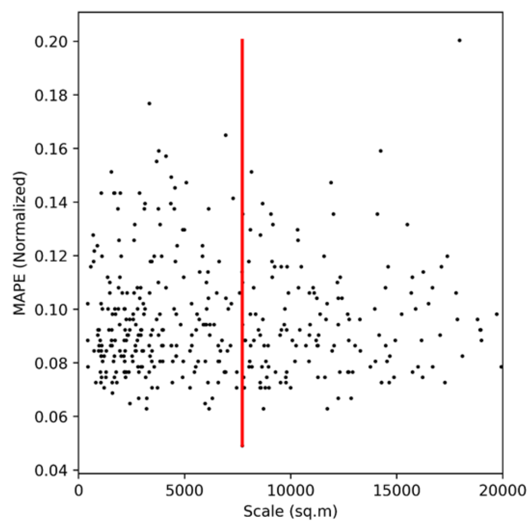
(*a*) Target building: OF-1



(*b*) Target building: DO-1

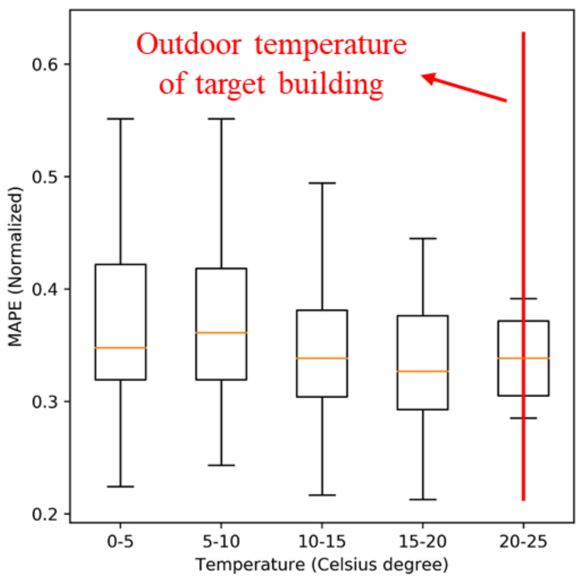


(*c*) Target building: PC-1

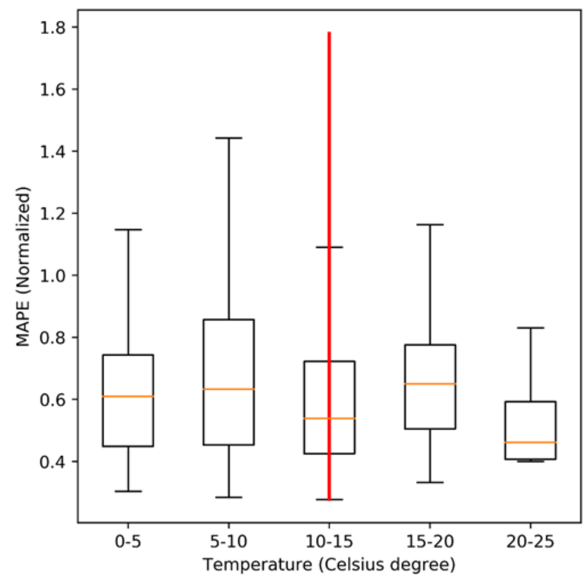


(*d*) Target building: UL-1

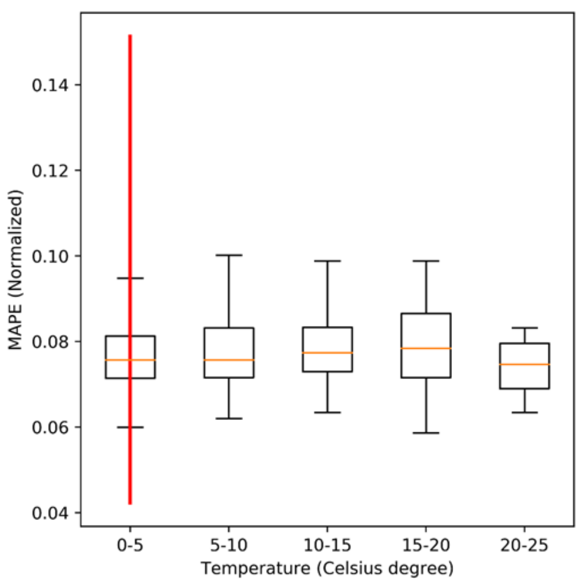
Fig.11 Model performance under source data of different building scale



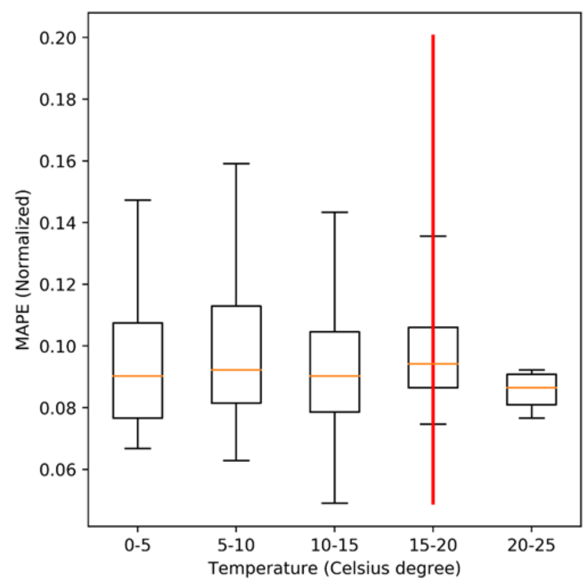
(*a*) Target building: OF-1



(*b*) Target building: DO-1

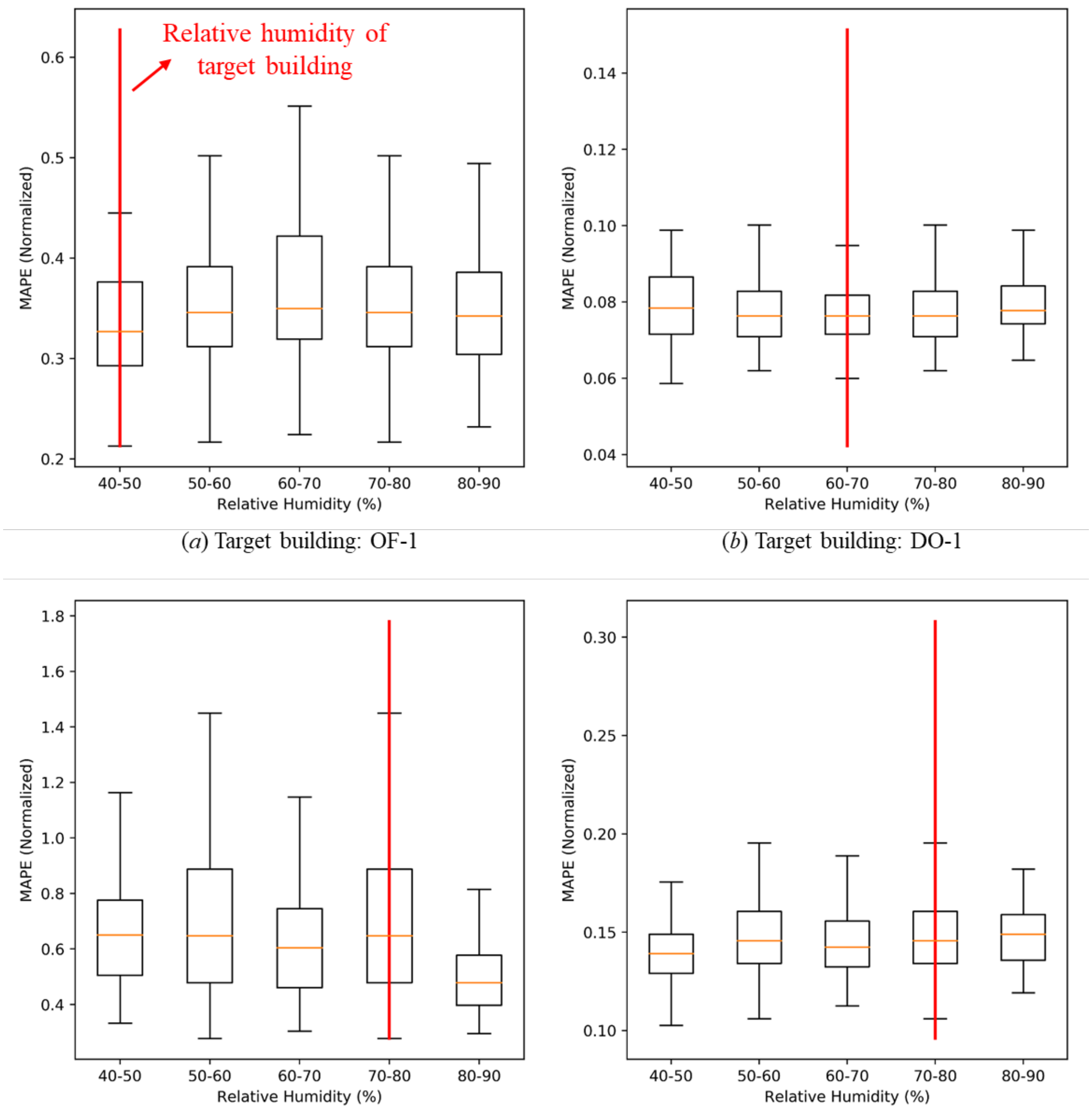


(*c*) Target building: PC-1



(*d*) Target building: UL-1

Fig.12 Model performance under source data of different outdoor dry-bulb temperature



(*a*) Target building: OF-1

(*b*) Target building: DO-1

(*c*) Target building: PC-1

(*d*) Target building: UL-1

Fig.13 Model performance under source data of different relative humidity

Table 6. Kendall coefficient of different building features

Target building	Building type	Industry	Scale	Temperature	Humidity	Euclidean distance
1	0.07	0.03	-0.04	-0.06	0.06	0.11
2	0.06	-0.18	0.02	-0.07	-0.07	-0.02
3	0.18	0.11	-0.02	-0.05	-0.03	0.24
4	0.19	-0.01	-0.11	0.02	0.02	0.17
5	0.06	0.04	-0.07	-0.02	0.04	0.05
6	**0.32**	**0.32**	0.13	-0.08	-0.16	0.20
7	**0.39**	**0.39**	0.1	-0.12	-0.18	0.17
8	**0.28**	**0.27**	0.08	-0.06	-0.08	0.14
9	**0.22**	**0.21**	0.11	-0.08	-0.08	0.21
10	**0.39**	**0.38**	0.14	-0.12	-0.19	0.38
11	**-0.27**	**-0.24**	0.12	-0.08	-0.19	0.01
12	-0.02	0.11	-0.08	-0.02	0.1	0
13	0.18	0.08	0.02	-0.02	0.04	0.04
14	0.01	0.13	-0.06	-0.02	0.08	-0.04
15	0.06	0.08	-0.09	0.02	0.09	-0.01
16	-0.01	0.03	-0.03	-0.05	0.07	-0.02
17	-0.01	-0.09	-0.05	0.01	0	0.10
18	0.1	0	-0.02	-0.09	0.06	0.01
19	0.06	0.03	0	0.01	0.01	0.05
20	-0.04	**-0.21**	-0.02	-0.07	-0.09	0.15

The impacts of different building features on transfer learning performance are analyzed, and the results are presented in Fig.9–13. In Fig.9 and Fig.10, the performance of models trained by source-target building pairs of same/different usage and subindustry is shown. In Fig.11, the red lines represent the building scale of the target building and the black points represent the scale of the source building. In Fig.12 and Fig.13, the average value (within the sample dataset) of outdoor dry-bulb temperature and relative humidity are categorized into 5 subsets. The red lines represent the target buildings' outdoor dry-bulb temperature (or relative humidity).

For each target building, the mean MAPE of models trained by source buildings of the same usage/subindustry is normalized by the mean MAPE of different usage/subindustry. It can be seen that, in most cases, the source-target building pairs of the same usage (or subindustry) get better results compared to the building pairs of different usage (or subindustry). However, the influence of source-target building pairs' outdoor climates seems relatively small. And the Kendall correlation coefficients of different building features are calculated and presented in Table 6. The Kendall rank correlation coefficient, also referred to Kendall's tau coefficient, is a statistic used to measure the ordinal association between two measure quantities (Kendall 1938). The Kendall τ coefficient can be determined using Eq. (6).

$$\tau = \frac{(number\ of\ concordant\ pairs) - (number\ of\ discordant\ pairs)}{\frac{n(n-1)}{2}} \tag{6}$$

The coefficient must be in the range [-1,1]. If the agreement between the two rankings is perfect (i.e., the two rankings are the same), the coefficient has value 1. If the disagreement between the two rankings is perfect (i.e., one ranking is the reverse of the other), the coefficient has value −1. If two rankings are independent, the coefficient is expected to be approximately zero.

It can be summarized from the aforementioned results, among these five features, the building's primary usage and industry show the most significant influence on the effect of transfer learning. That is, if the source dataset has the same building usage (or subindustry) as the target building dataset, the implementation of transfer learning may bring greater accuracy improvement. This phenomenon is easy to explain that buildings with the same usage have a higher probability of having similar energy consumption patterns. However, no significant and consistent correlation was found between the other three features and the transfer learning-based model accuracy. This may conflict with traditional domain expertise that the weather condition and building scale all have substantial impacts on the building energy consumption. A possible explanation is that the building energy consumption data in this study are all normalized to adjust the values in different scales. The weather condition and building scale affect the energy profile more in terms of scale (absolute magnitude), while the building usage and industry affect the energy profile more in terms of shape (pattern). Therefore, the effects of these factors on transfer learning may be insignificant.

Based on the individual and general result analysis, it seems that the similarity of the load profile is the most influential factor in the effect of transfer learning. To further validate this assumption, the Euclidean distance between the source building and target building's average energy consumption profile is calculated by Eq. (7).

$$\text{Euclidean distance} = \sqrt{\sum_{t=1}^{24}\left(E_{source,t} - E_{target,t}\right)^2} \tag{7}$$

The Kendall correlation coefficient between the energy profile Euclidean distance and the transfer learning-based model accuracy is shown in Table 6. The results show that, for more than one-third of the target buildings, the model pre-trained by more similar source building energy consumption profiles can lead to higher accuracy.

4. Conclusions

Building energy consumption prediction plays a crucial role in evaluating different building design alternatives, developing energy efficiency-optimal control and diagnosis strategies, and developing the demand and supply management in power grids. For information-poor buildings such as new-built buildings with limited historical data, or already-built buildings with under-developed building automation systems, the energy consumption prediction task remains a significant challenge. This paper implemented transfer learning to improve energy consumption prediction accuracy for a target building with limited available data, with the help of additional data from other buildings. A three-layer BPNN model is developed and tested using a large public benchmark dataset – Building Data Genome Project. The effects of transfer learning with different source building data samples, different target building data samples, different source-target building pairs are investigated and compared. To sum up, the contributions of this study are summarized as follows: 1) When the available training data is very limited, transfer learning can increase the prediction accuracy with most source datasets, no matter the sources buildings are similar or not. 2) The less the available training data are, the more accuracy improvement (compared with the baseline model) transfer learning can bring about to building energy prediction modeling; 3) When the target building has limited training data, increasing pre-train data samples from 10 days to 30 days make little difference; 4) The most influential building features on the transfer learning are the building usage and industry (which has the most significant effect on the building energy consumption pattern), compared to outdoor dry-bulb temperature, relative humidity, and building scale. When selecting the information-rich buildings as source buildings, it is recommended to pay attention to the building usage and industry.

References

[1] Amasyali, K., & El-Gohary, N. M. (2018). A review of data-driven building energy consumption prediction studies. Renewable and Sustainable Energy Reviews, 81, 1192-1205.

[2] Asadi, S., Amiri, S. S., & Mottahedi, M. (2014). On the development of multi-linear regression analysis to assess energy consumption in the early stages of building design. Energy and Buildings, 85, 246-255.

[3] Bourdeau, M., qiang Zhai, X., Nefzaoui, E., Guo, X., & Chatellier, P. (2019). Modeling and forecasting building energy consumption: A review of data-driven techniques. Sustainable Cities and Society, 48, 101533.

[4] EMSD (2019), Hong Kong Energy End-use Data 2019, Available at http://www.emsd.gov.hk.

[5] Fan, C., Sun, Y., Xiao, F., Ma, J., Lee, D., Wang, J., & Tseng, Y. C. (2020). Statistical investigations of transfer learning-based methodology for short-term building energy predictions. Applied Energy, 262, 114499.

[6] Fan, C., Sun, Y., Zhao, Y., Song, M., & Wang, J. (2019). Deep learning-based feature engineering methods for improved building energy prediction. Applied Energy, 240, 35-45.

[7] Fan, C., Wang, J., Gang, W., & Li, S. (2019). Assessment of deep recurrent neural network-based strategies for short-term building energy predictions. Applied Energy, 236, 700-710.

[8] Fan, C., Xiao, F., & Zhao, Y. (2017). A short-term building cooling load prediction method using deep learning algorithms. Applied energy, 195, 222-233.

[9] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT press.

[10] Hahnloser, R. H., Sarpeshkar, R., Mahowald, M. A., Douglas, R. J., & Seung, H. S. (2000). Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. Nature, 405(6789), 947-951.

[11] Hooshmand, A., & Sharma, R. (2019, June). Energy predictive models with limited data using transfer learning. In Proceedings of the Tenth ACM International Conference on Future Energy Systems (pp. 12-16).

[12] Hosseinzadeh, H., Razzazi, F., & Kabir, E. (2016). A weakly supervised large margin domain adaptation method for isolated handwritten digit recognition. Journal of Visual Communication and Image Representation, 38, 307-315.

[13] Hu, W., Qian, Y., Soong, F. K., & Wang, Y. (2015). Improved mispronunciation detection with deep neural network trained acoustic models and transfer learning based logistic regression classifiers. Speech Communication, 67, 154-166.

[14] IEA (2015), Building Energy Use in China, OECD/IEA, Paris.

[15] IEA (2018), World Energy Statistics and Balances 2018, OECD/IEA, Paris.

[16] Karparthy, A. (2017). A peek at trends in machine learning. Medium. com.

[17] Kendall, M. G. (1938). A new measure of rank correlation. Biometrika, 30(1/2), 81-93.

[18] Keogh, E., & Kasetty, S. (2003). On the need for time series data mining benchmarks: a survey and empirical demonstration. Data Mining and knowledge discovery, 7(4), 349-371.

[19] Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

[20] Li, X., & Wen, J. (2014). Review of building energy modeling for control and operation. Renewable and Sustainable Energy Reviews, 37, 517-537.

[21] Li, W., Duan, L., Xu, D., & Tsang, I. W. (2013). Learning with augmented features for supervised and semi-supervised heterogeneous domain adaptation. IEEE transactions on pattern analysis and machine intelligence, 36(6), 1134-1148.

[22] Ma, Y., Luo, G., Zeng, X., & Chen, A. (2012). Transfer learning for cross-company software defect prediction. Information and Software Technology, 54(3), 248-256.

[23] Miller, C., & Meggers, F. (2017a). Mining electrical meter data to predict principal building use, performance class, and operations strategy for hundreds of non-residential buildings. Energy and Buildings, 156, 360-373.

[24] Miller, C., & Meggers, F. (2017b). The Building Data Genome Project: An open, public data set from non-residential building electrical meters. Energy Procedia, 122, 439-444.

~~[25] Mozafari, A. S., & Jamzad, M. (2016). A SVM-based model-transferring method for heterogeneous domain adaptation. Pattern Recognition, 56, 142-158.~~

~~[26] Nam, J., Pan, S. J., & Kim, S. (2013, May). Transfer defect learning. In 2013 35th international conference on software engineering (ICSE) (pp. 382-391). IEEE.~~

[27] Nichiforov, C., Stamatescu, G., Stamatescu, I., Făgărăşan, I., & Iliescu, S. S. (2018, June). Intelligent load forecasting for building energy management systems. In 2018 IEEE 14th International Conference on Control and Automation (ICCA) (pp. 896-901). IEEE.

[28] Pan, S. J., & Yang, Q. (2009). A survey on transfer learning. IEEE Transactions on knowledge and data engineering, 22(10), 1345-1359.

[29] Perera, A. T. D., Wickramasinghe, P. U., Nik, V. M., & Scartezzini, J. L. (2019). Machine learning methods to assist energy system optimization. Applied energy, 243, 191-205.

[30] Rahman, A., Srikumar, V., & Smith, A. D. (2018). Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks. Applied energy, 212, 372-385.

[31] Ramachandran, P., Zoph, B., & Le, Q. V. (2017). Searching for activation functions. arXiv preprint arXiv:1710.05941.

[32] Ribeiro, M., Grolinger, K., ElYamany, H. F., Higashino, W. A., & Capretz, M. A. (2018). Transfer learning with seasonal and trend adjustment for cross-building energy forecasting. Energy and Buildings, 165, 352-363.

[33] Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. nature, 323(6088), 533-536.

[34] Shin, H. C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., ... & Summers, R. M. (2016). Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE transactions on medical imaging, 35(5), 1285-1298.

[35] Silver, D. L., Yang, Q., & Li, L. (2013, March). Lifelong machine learning systems: Beyond learning algorithms. In 2013 AAAI spring symposium series.

[36] Weiss, K., Khoshgoftaar, T. M., & Wang, D. (2016). A survey of transfer learning. Journal of Big data, 3(1), 9.

[37] Xue, X., Wang, S., Sun, Y., & Xiao, F. (2014). An interactive building power demand management strategy for facilitating smart grid optimization. Applied Energy, 116, 297-310.

[38] Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks?. In Advances in neural information processing systems (pp. 3320-3328).

[39] Zhao, H. X., & Magoulès, F. (2012). A review on the prediction of building energy consumption. Renewable and Sustainable Energy Reviews, 16(6), 3586-3592.