Energy & Buildings 252 (2021) 111435

Contents lists available at ScienceDirect

Energy & Buildings

journal homepage: www.elsevier.com/locate/enb

A general multi-source ensemble transfer learning framework integrate of LSTM-DANN and similarity metric for building energy prediction

Xi Fang [a], Guangcai Gong [a,*], Guannan Li [b], Liang Chun [a], Pei Peng [a], Wenqiang Li [a]

[a] College of Civil Engineering, Hunan University, Changsha 410082, PR China
[b] School of Urban Construction, Wuhan University of Science and Technology, Wuhan 430065 PR China

ARTICLE INFO

Article history:
Received 4 May 2021
Revised 3 July 2021
Accepted 5 September 2021
Available online 9 September 2021

Keywords:
Multi-source
Ensemble learning
Transfer learning
Similarity metric
Building energy prediction

ABSTRACT

Transfer learning can improve building energy prediction performance by utilizing the knowledge learned from source domain. However, most studies focus on the single-source transfer learning and may lead to model performance degradation when there exists large domain shift between the single source domain and target domain. To address this issue, this study proposes a multi-source ensemble transfer learning (Multi-LSTM-DANN) framework integrate of LSTM-DANN neural network and similarity metric, which can enhance the prediction performance of target building power consumption by using multi-source building data (domain). LSTM-DANN is first used to extract the domain invariant features between each pair of source domain and target domain. Then maximum mean discrepancy (MMD) is applied to metric the distance between each pair of the extracted domain invariant features. Finally, the reciprocal of MMD is used as similarity metric index to calculate the regression weight and prediction value of the proposed Multi-LSTM-DANN model. Experiments with different number of source domains are conducted to demonstrate the effectiveness of the proposed Multi-LSTM-DANN framework. Results demonstrate that most multi-source transfer learning models can enhance the prediction performance of the target building power consumption compared to the corresponding single-source transfer learning models. The proposed Multi-LSTM-DANN framework can provide guiding significance for the application of multi-source building data in the future.

1. Introduction

With the introduction of carbon peaks and carbon neutral targets, the energy development situation will undergo profound changes, and the pace of energy form transformation and structural adjustment will be further accelerated. Non-fossil renewable energy sources such as wind energy, solar energy, nuclear energy and biomass energy will gain broader development space [1]. The coordinated plan, optimized operation and collaborative management of various heterogeneous energy subsystems will become a trend. In this process, building energy consumption prediction plays an important role for building energy system management, and is of great significance to the online control and optimization of building energy systems. Currently, machine learning has been widely used in the field of building energy data analysis [2–5]. Accurate and reasonable building energy consumption prediction is the key foundation for the building energy efficiency and control, and it is also an important prerequisite for formulating relevant building energy saving measures.

The existing approaches for building energy consumption prediction can be mainly divided into two categories: physical modelling and data driven modelling [4,6]. Physical models also known as white-box models heavily rely on thermal dynamic characteristics of simulated buildings. Some commercial simulation software such as Energyplus, DOE-2, TRANSYS calculate the building energy consumption by the detailed physical models [4,7–9]. The physical models have the advantage of clear physical meaning. However, the thermal property parameters of building envelope and personnel activity schedule may not be available in the practical application. For most engineering designs and simulations, it is difficult to set the accurate parameters of the physical models, which may lead to poor prediction performance. Data driven models do not require detailed data information about the building and automatically capture the relationship between building energy consumption and historical data. With the popularization of building energy management systems and the development of computer technology, data driven models such as support vector regression (SVR), long short term memory (LSTM), autoencodes,

* Corresponding author.
E-mail address: gcgong@hnu.edu.cn (G. Gong).

https://doi.org/10.1016/j.enbuild.2021.111435
0378-7788/

random forest, boosting trees and simplified online methods have attracted great attention in the field of building energy consumption prediction [4,10–16].

The satisfactory performance can be obtained by using advanced machine learning algorithms for building energy predictions. However, these traditional machine learning algorithms generally need large-scale historical data to achieve a better prediction performance. The difficulty is that the data are not always sufficient, or when the data is sufficient and the data trained for one specific task cannot be directly generalized to different related tasks [17–19]. To address these issues, the learning model trained with specific building data should adapt to different building data distribution.

Transfer learning known as domain adaptation can effectively overcome the above data distribution discrepancy (domain shift) problems. As a kind of machine learning, transfer learning has a wide range of applications in computer vision, image classification, document classification and sentiment analysis due to this advantages [20,21]. In recent years, researchers have carried out some studies in the field of building energy prediction based on transfer learning methods, which also have a positive effect on the improvement of energy prediction performance. Fan et al. [22] investigated the potentials of transfer learning based methodology for building energy prediction using 507 non-residential buildings. Grolinger et al. [23] proposed a similarity-based chained transfer learning approach with recurrent neural network (RNN) for building energy prediction. Gao et al. [24] proposed a sequence-to-sequence model based on a transfer learning framework to improve prediction performance of the building energy consumption. Mocanu et al. [25] introduced an unsupervised reinforcement transfer learning method for cross building energy prediction to improve the energy prediction accuracy. Li et al. [26] presented a transfer learning method based on ANN model for building energy prediction and focused on the influence of data volume on the transfer performance. Ribeir et al. [27] proposed a transfer learning method for cross-building energy prediction with seasonal and trend adjustment. Fang et al. [28] employed a long short term memory and domain adversarial neural network (LSTM-DANN) based transfer learning strategy to overcome the data shortage problem of building energy prediction in different scenarios.

Although these studies have some good effects on the prediction performance improvement, most of the above studies focus on a single-source transfer learning for building energy prediction. However, the application of single-source transfer learning methods may get some unsatisfactory results when there exists large domain shift between source domain and target domain. Generally, data from multi-source domains with different data distributions can be collected in practical application scenarios. The knowledge and internal relationship learned from multi-source domains can be better used to assist the building energy prediction of target task. Therefore, transfer learning methods with multi-source domains should have more potential for the prediction performance improvement of target task. Due to this advantage and potential, multi-source transfer learning has gradually attracted the attention of researchers. Multi-source transfer learning methods have been widely used in some classification tasks [21,29–31], and the performance has also been improved to a certain extent compared to the single-source transfer learning methods. From the literature review, there are few studies on the multi-source transfer learning in the field of building energy consumption prediction. In addition, the commonly used fine-tune based single-source transfer learning method for building energy prediction cannot be directly extended to multi-source transfer learning tasks. Meanwhile, the effectiveness of the LSTM-DANN based single-source transfer learning method has been validated according to our previous work [28], and it can also be better extended to multiple sources for building energy prediction.

In this paper, inspired by our previous work of single-source transfer learning, a general multi-source ensemble transfer learning (Multi-LSTM-DANN) framework integrate of LSTM-DANN and similarity metric is proposed for building energy prediction. The goal of the proposed Multi-LSTM-DANN framework is aimed to employ multi-source domain data to create an efficient ensemble learning model to enhance the prediction performance of target building power consumption. LSTM-DANN is first used to extract the domain invariant features between each pair of source domain and target domain. Then maximum mean discrepancy (MMD) is applied to metric the distance between each pair of the extracted domain invariant features. Finally, the reciprocal of MMD is used as similarity metric index to calculate the regression weight and prediction value of the proposed Multi-LSTM-DANN model. Experiments are conducted to demonstrate the effectiveness of the proposed Multi-LSTM-DANN framework. To the best of our knowledge, it is the first time to apply a multi-source transfer learning method for building energy prediction based on deep neural network. The proposed multi-source ensemble transfer learning framework can provide guiding significance for the application of multi-source building data in the future.

2. Methodology

2.1. Overall flowchart of the research

This study was conducted as the following four steps shown in Fig. 1. It mainly consists of four steps, data collection and preparation, multi-source transfer learning model development, experiment setup and model parameter selection, model evaluation and building energy prediction. (1) Step 1: Data collection and preparation. The building dataset from the Building Data Genome Project are collected, and the raw missing values are filled and then are processed into suitable formats. (2) Step 2: Multi-source transfer learning model development. The Multi-LSTM-DANN framework integrate of LSTM-DANN neural network and similarity metric is proposed. (3) Step 3: Experiment setup and model parameter selection. A total of 15 different models including single-source and multi-source transfer learning models are conducted. The validation datasets are used to determine the near model optimal hyper-parameter. (4) Step 4: Model evaluation and building energy prediction. Experiments are conducted with different number of source domains to evaluate the model prediction performance of the proposed Multi-LSTM-DANN framework.

2.2. Multi-source transfer learning

Transfer learning has been widely studied for many years and the effectiveness has been verified by researchers. The single-source transfer learning model only considers a single-source domain and a target domain. The performance improvement of the transfer learning based methods largely depend on the data distribution between the source domain and the target domain. The application of single-source transfer learning methods may get some unsatisfactory results when there exists large domain shift between source domain and target domain. In addition, data from multi-source domains with different distributions are available in practical application scenarios. The knowledge and internal relationship learned from multi-source domains can be better used to enhance the prediction performance of target task in real-world application. Multi-source transfer learning methods can be classified into two main categories: the boosting based methods and the regularization based methods [21,31]. The boosting based

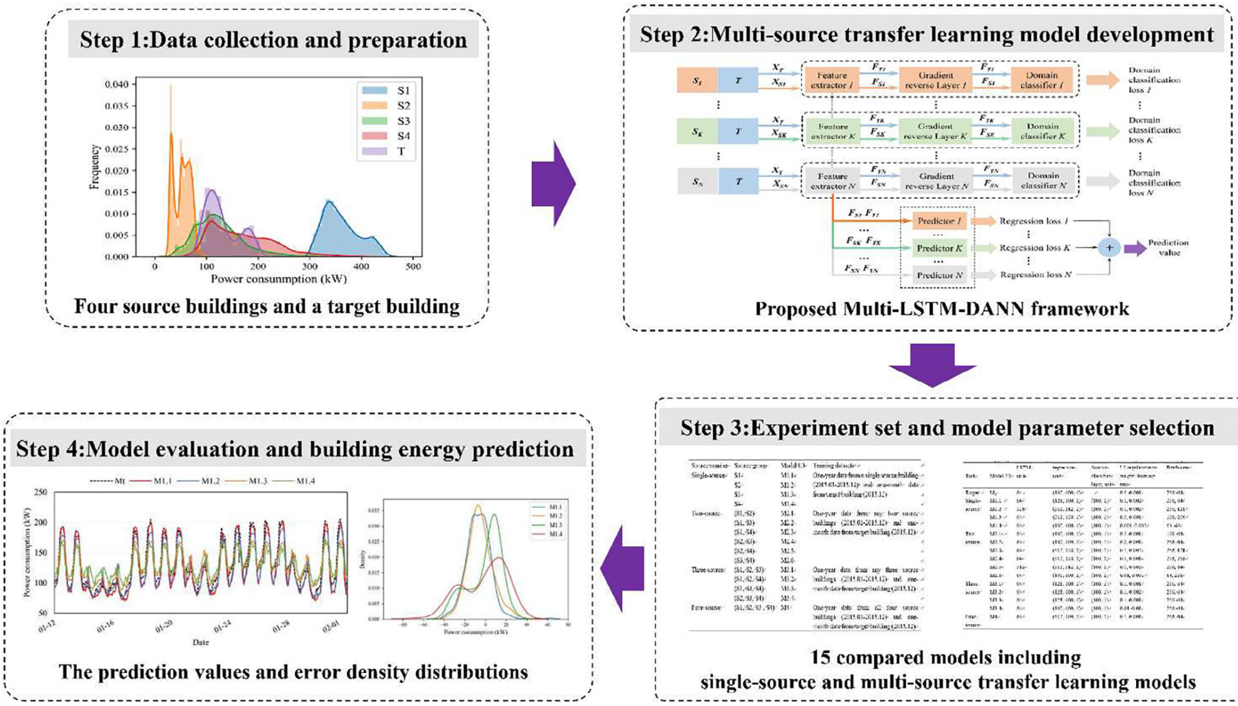

Fig. 1. Overall flowchart of the research.

methods construct a strong classifier by selecting a set of weak classifiers and making use of their respective weighted vote to obtain the final classification result. The regularization based methods attempt to apply the learning model with regularization term to solve the optimization problems. In this study, the proposed Multi-LSTM-DANN framework belongs to the boosting based method, which is aimed to construct a set of regression predictors to enhance the prediction performance of target building by using multi-source building data. The proposed multi-source transfer learning model attempts to find the domain invariant features between each pair of source domain and target domain. Domain invariant feature are common features between the source and target domain. Fig. 2 shows the diagram of the proposed multi-source transfer learning model (Take a two-source transfer learning model as an example).

2.3. LSTM-DANN

LSTM is a time series prediction algorithm proposed by Hochreiter & Schmidhuber [32], which has been widely used for building energy consumption prediction in recent years. DANN is an adversarial neural network based transfer learning method, and the superior performance of DANN has been demonstrated in many complex tasks [33]. Fang et al. [28] proposed a hybrid deep transfer learning strategy (LSTM-DANN) for short term building energy prediction based on LSTM and DANN. The hybrid LSTM-DANN structure mainly consists of feature extractor, domain classifier, and regression predictor. Feature extractor LSTM is employed to automatically extract temporal features between the source domain and target domain. DANN attempts to find domain invariant features between the source domain and target domain by adversarial

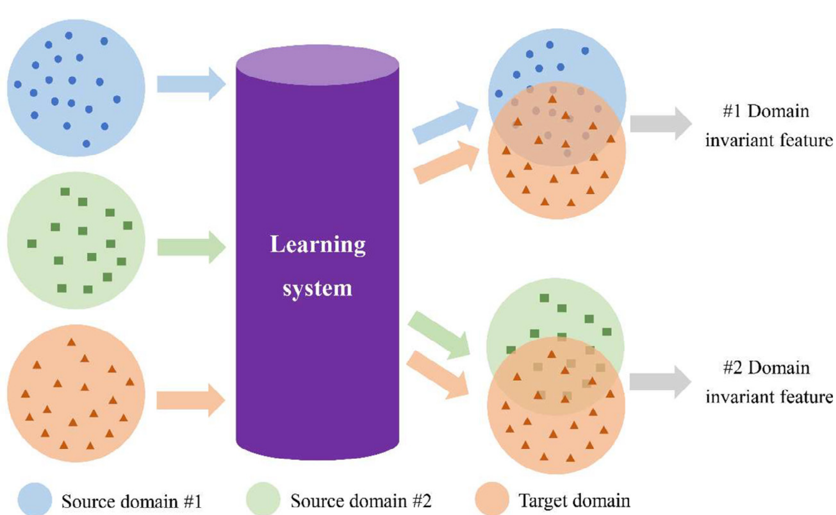

Fig. 2. Diagram of the proposed multi-source transfer learning model.

domain adaptation of feature extractor LSTM and domain classifier. The detailed principle of the proposed hybrid LSTM-DANN method can be seen in [28].

2.4. Maximum mean discrepancy

As there exists multi-source domains in multi-source transfer learning, it is necessary to choose an appropriate index to measure the data distribution differences between each pair of source domain and target domain. MMD is first used to construct two sample test and determine whether the two data distributions are the same [34]. In recent years, MMD has been widely used for similarity metric between the source domain and target domain in transfer learning [35–37]. MMD measures the distances of two distributions in the reproducing kernel Hilbert space and is a kernel based method. The principle of MMD is as follow. Given a source domain $X_s = \{(x_S^i)\}_{i=1}^{N_s}$ and a target domain $X_T = \{x_T^i\}_{i=1}^{N_T}$, the nonlinear function in the reproducing kernel Hilbert space is ϕ. MMD of *Xs* and *Xt* in the reproducing kernel Hilbert space can be defined as

$$MMD(X_s, X_t) = \| \frac{1}{N_s} \sum_{i=1}^{N_s} \phi(x_S^i) - \frac{1}{N_T} \sum_{i=1}^{N_T} \phi(x_T^i) \| \quad (1)$$

The smaller the value of MMD is, the more similar the two domains are. The proposed Multi-LSTM-DANN framework needs to measure the similarity of the extracted domain invariant features between each pair of source domain and target domain by MMD.

2.5. Similarity metric index

The proposed Multi-LSTM-DANN framework aims to learn knowledge from multi-source domains to construct an efficient ensemble model, which can enhance the prediction performance of building power consumption for the target task. LSTM-DANN is used to overcome the domain distribution discrepancy. The *N* source domain dataset and a target domain dataset are respectively processed by the LSTM-DANN to extract the domain invariant features between each pair of source domain and target domain, and the processed process can be expressed by Eq. (2).

$$\left(X'_{S_i}, X'_{T_i}\right) = LSTM\text{ - }DANN(X_{S_i}, X_T) \quad (2)$$

X_{S_i} is the *i-th* source domain dataset and X_T is the target domain dataset. X'_{S_i} is the *i-th* source domain invariant features extracted by LSTM-DANN and X'_T is the *i-th* target domain invariant features extracted by LSTM-DANN.

As there are multi-source domains, it is necessary to find a metric index to calculate the similarity between each pair of the extracted domain invariant features. The smaller MMD means the higher similarity between the extracted domain invariant features. Therefore, the reciprocal of MMD is used as the similarity metric index (SMI) between each pair of the extracted domain invariant features, which can be expressed as:

$$SMI_i = \frac{1}{MMD_i(X'_{S_i}, X'_T)}, i \in [1, N] \quad (3)$$

2.6. Proposed multi-source ensemble transfer learning framework

The proposed Multi-LSTM-DANN framework is aimed to learn a model that can transfer knowledge from multi-source domains with different data distributions, so as to enhance the prediction performance of target task. Given *N* source domains and a target domain are respectively defined as $S = S_1, S_2, \ldots, S_i, \ldots S_N\}$ and *T*. The proposed Multi-LSTM-DANN framework based on multiple LSTM-DANN neural networks is shown in Fig. 3.

LSTM-DANN is first used to find the domain invariant features between each pair of the source domain and target domain. Then the SMI is used to metric the similarity between each pair of the extracted domain invariant features. Finally, the SMI is applied to calculate the regression weight and the prediction value of the proposed Multi-LSTM-DANN framework is calculated. In this study, MMD is used to metric the domain invariant features between the source domain and target domain in the proposed multi-source ensemble transfer learning model. Then the regression weight can be calculated when there are more than one pair of domain invariant features in the multi-source transfer learning model. The regression weight *W* and prediction value *y* of the proposed Multi-LSTM-DANN framework can be respectively calculated by:

$$W_i = \frac{SMI_i}{\sum_{i=1}^{n} SMI_i}, i \in [1, N] \quad (4)$$

$$y = \sum_{i=1}^{N} W_i \hat{y}_i, i \in [1, N] \quad (5)$$

The proposed Multi-LSTM-DANN framework can be divided into the *N* predictors and an ensemble learner. The *N* predictors are combined by LSTM-DANN using *N* source domains and a target domain, and the ensemble learner employs multiple predictors to obtain an aggregated prediction value by the similarity metric index. The ensemble model usually performs better prediction performance and robustness than a single predictor.

The neural network structure of our proposed Multi-LSTM-DANN framework is shown in Fig. 4. The proposed Multi-LSTM-DANN framework is based on the previous hybrid transfer learning LSTM-DANN network structure, which consists of multiple feature extractors, domain classifiers and regression predictors. The feature extractors are LSTM layers, the domain classifiers and regression predictors are both fully connected layers. Specially, the gradient reversal layer (GRL) between feature extractors and domain classifiers is aimed to find the domain invariant features in the LSTM-DANN structure [33,38]. The main idea of the proposed Multi-LSTM-DANN framework is to train the multiple LSTM-DANN neural networks to extract domain invariant features between *N* source domains and the target domain, and then put the extracted domain invariant features to the *N* regression predictors for weighted ensemble based on the similarity metric index.

The training optimization loss of the neural network consists of the regression loss and domain classification loss. The regression loss of the energy prediction is defined as the mean square error:

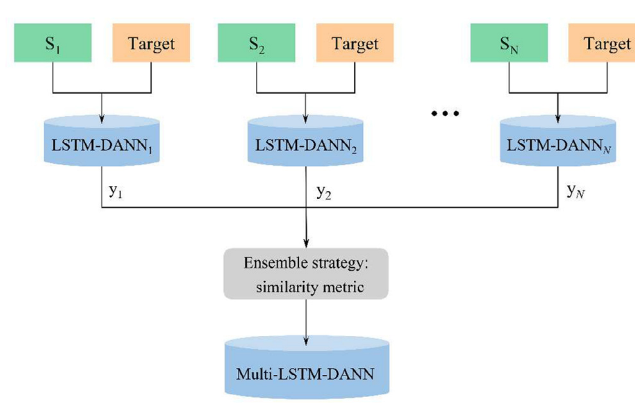

Fig. 3. The proposed Multi-LSTM-DANN framework.

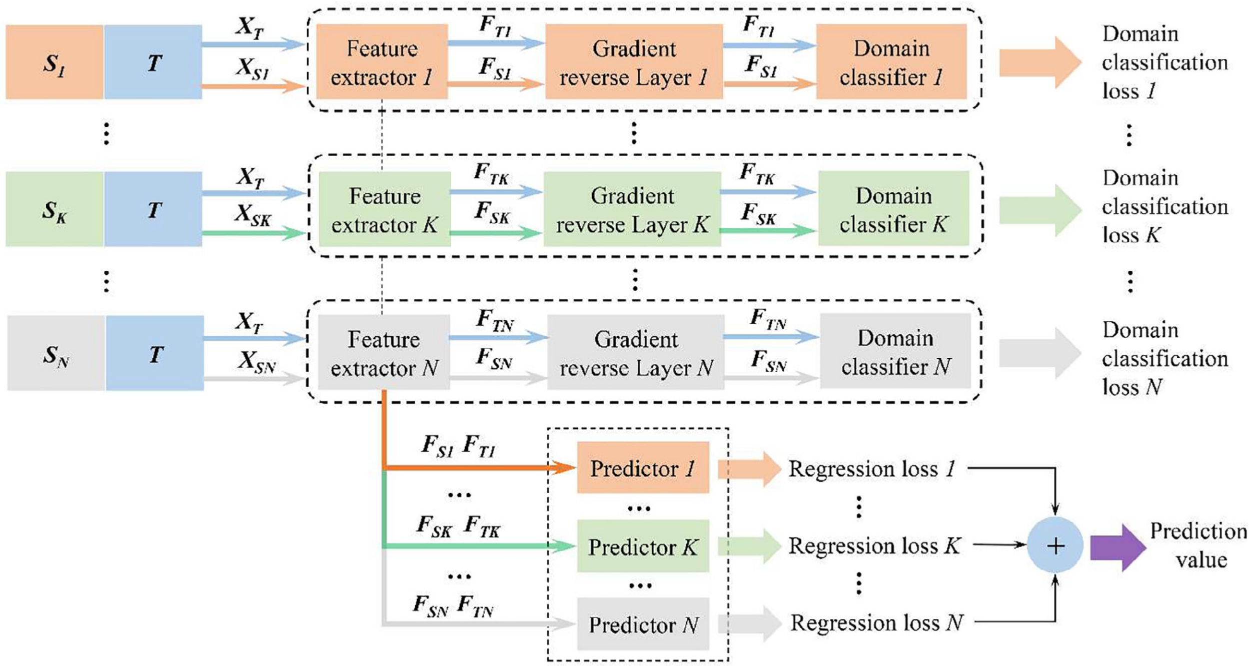

Fig. 4. The network structure of proposed Multi-LSTM-DANN framework.

$$\phi_y^i(\theta_f, \theta_y) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{6}$$

Where n is the batch size of the training data. y_i and $\hat{y}_i$ respectively denote the actual and prediction value of building energy.

The domain classification loss is defined as binary cross entropy:

$$\phi_d^i(\theta_f, \theta_d) = \frac{1}{n}\sum_{i=1}^{n}\left(d_i \log\frac{1}{\hat{d}_i} + (1 - d_i)\log\frac{1}{1 - \hat{d}_i}\right) \tag{7}$$

where d_i and $\hat{d}_i$ respectively denote the actual domain label (source: 0, target: 1) and the prediction domain label.

Then the final objective "pseudo-function" of the proposed Multi-LSTM-DANN framework can be optimized by the gradient descent, which can be expressed as:

$$\begin{aligned}\phi(\theta_f,\theta_y,\theta_d) = &\sum_{k=1}^{N}(\tfrac{1}{N_S}\sum_{i=1}^{N_S}\phi_{y,k}(G_y(G_f(x_i;\theta_f);\theta_y),y_i) + \tfrac{1}{N_T}\sum_{i=1}^{N_T}\phi_{y,k}(G_y(G_f(x_i;\theta_f);\theta_y),y_i)\\ &+\left(\tfrac{1}{N_S}\sum_{i=1}^{N_S}\phi_{d,k}((G_d(R(G_f(x_i;\theta_f);\theta_d),d_i\right) + \tfrac{1}{N_T}\sum_{i=1}^{N_T}\phi_{d,k}((G_d(R(G_f(x_i;\theta_f);\theta_d),d_i)))\end{aligned} \tag{8}$$

where N_S and N_T respectively represent the number of source and target domain data, $\theta_f, \theta_y, \theta_d$ denote the model parameter weights of feature extractors, regression predictors and domain classifiers, respectively. G_f, G_y, G_d respectively represent the feature extractors, the regression predictors and the domain classifiers. k represents the number of source domains.

3. Results and discussions

3.1. Data description and preprocessing

The building dataset from the Building Data Genome Project are collected to evaluate the proposed Multi-LSTM-DANN framework in this study [39]. The collected dataset mainly contains five types of buildings, including office, college classroom, primary classroom, college laboratory, and college dormitory. Table 1 shows the detailed information of the selected source and target buildings. These five office buildings located in America are selected for the following analysis, including one target building (T) and four source buildings (S1, S2, S3, S4). The target building collects data for two months from 2016.12 to 2017.01, and the other four source buildings collect data for one year from 2016.01 to 2016.12. All data are collected in an hour interval, and each data sample consists of five numeric variables and three categorical variables shown in Table 2, namely outdoor temperature, dew point temperature, wind speed, atmospheric pressure, power consumption, month type, day type, and hour type.

The raw time series data cannot be put directly into the model, and it should be preprocessed into a continuous and appropriate format first. As the raw building datasets have some missing val-

Table 1
Building information of the selected source and target buildings.

	Collecting period	Sample number	Usage type
Target building (#T)	2016.12–2017.01	1488 (62 days)	Office
Source building A (#S1)	2016.01–2016.12	8784 (366 days)	Office
Source building B (#S2)	2016.01–2016.12	8784 (366 days)	Office
Source building C (#S3)	2016.01–2016.12	8784 (366 days)	Office
Source building D (#S4)	2016.01–2016.12	8784 (366 days)	Office

Table 2
Modeling variables.

Variable	Type	Units/Range
Outdoor temperature (T_O)	Numeric	℃
Dew point temperature (Td)	Numeric	℃
Wind speed (V)	Numeric	kM/h
Atmospheric pressure (P)	Numeric	hpa
Power consumption (E)	Numeric	kW
Month type	Categorical	(Jan–Dec)
Day type	Categorical	(1st–31st)
Hour type	Categorical	(0:00–23:00)

ues, the linear interpolation method is first used to fill in the missing weather data. The missing building power consumption data is filled by random forest. Second, in order to make the training models interoperable between different datasets, the datasets should be rescaled to have the same shape and proportion between them. Then, the training datasets of target domain are rescaled to have the same proportion of source domain datasets. In this way, the batch size of the target domain training data and the source domain data can be set to the same value, which can speed up the training process of neural network. After recovering the missing values and rescaling the data samples, one-hot encoding should be used to encode the categorical variables into numerical values that can be processed in the models. Finally, the processed data are standardized by z-score normalization method.

Fig. 5 shows the frequency histogram of power consumption for the selected five buildings. Due to the difference in energy-consuming device, personal schedules and building envelope characteristics, each building has a different frequency histogram of power consumption, while the kernel density profile of power consumption for each building shows a similar distribution to some extent. Therefore, these source buildings have the potential for transfer learning by extracting useful knowledge to enhance the prediction performance of target building power consumption. In this study, the historical measurement data of the past 24 h are used to predict the target building power consumption in the next hour. This study is aimed to explore how the number of source buildings will affect the power consumption prediction performance of target buildings that lack of historical data. The prediction performance on the target building power consumption with different number of source buildings will be discussed in the later sections.

3.2. Experiment setup

In this study, all experiments are performed on an Intel Xeon(R) Gold 6226R processor with 128 GB RAM and a GeForce RTX 3090 GPU. We implement all experiments by Python 3.6 and deep learning framework pytorch 1.7. To evaluate the effectiveness of the proposed Multi-LSTM-DANN framework for building power consumption predictions, experiments are conducted with different number of source domains. In this study, different number of source domains for transfer learning are considered in the experiments, including a single-source, two-source, three-source and four-source transfer learning models respectively. Table 3 shows

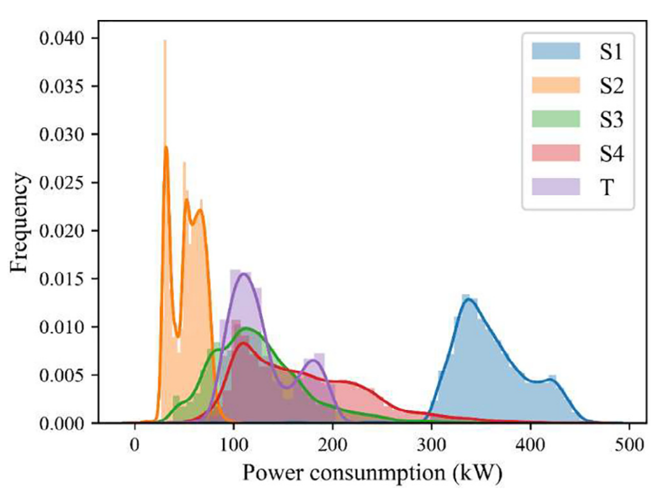

Fig. 5. Frequency histogram of power consumption for each building.

Table 3
Experiment setup summary of the compared transfer learning models with different number of source domains.

Source group	Source	Model ID	Training datasets
Single-source	S1	M1.1	One-year data from a single source building (2016.01–2016.12) and one-month data from target building (2016.12)
	S2	M1.2	
	S3	M1.3	
	S4	M1.4	
Two-source	(S1, S2)	M2.1	One-year data from any two source buildings (2016.01–2016.12) and one-month data from target building (2016.12)
	(S1, S3)	M2.2	
	(S1, S4)	M2.3	
	(S2, S3)	M2.4	
	(S2, S4)	M2.5	
	(S3, S4)	M2.6	
Three-source	(S1, S2, S3)	M3.1	One-year data from any three source buildings (2016.01–2016.12) and one-month data from target building (2016.12)
	(S1, S2, S4)	M3.2	
	(S1, S3, S4)	M3.3	
	(S2, S3, S4)	M3.4	
Four-source	(S1, S2, S3, S4)	M4	One-year data from all four source buildings (2016.01–2016.12) and one-month data from target building (2016.12)

Note: M1 represents the single-source transfer learning model, M2 represents the two-source transfer learning model, M3 represents the three-source transfer learning model, and M4 represents the four-source transfer learning model.

the experiment setup of the compared transfer learning models with different number of source domains. A total of 15 different transfer learning models (single-source: $C_4^1 = 4$, two-source: $C_4^2 = 6$, three-source: $C_4^3 = 4$, four-source: $C_4^4 = 1$) are conducted and the results of each model are averaged 10 times to reduce the random effect in the experiment. For each model, the mean and standard deviations of prediction performance metrics are calculated.

The data samples of target domain are divided into training, validation and test datasets respectively. The validation datasets are used to determine the near model optimal hyper-parameter and test datasets are used for evaluating the performance of the proposed multi-source transfer learning model. All source building dataset and the training dataset of target building are used for training, one-year data from source building (2016.01–2016.12) and one-month data from target building (2016.12) are for model training. Data from target building during January 01 to 10, 2017 are used for model hyper-parameter selection. The task of this study is to predict the power consumption of the target building from January 11 to 31, 2017.

In this study, the mean squared error (MSE), the mean absolute error (MAE), the mean absolute percentage error (MAPE), and the coefficient of variation of the root mean squared error (CV-RMSE) are selected to evaluate the prediction performance of target building power consumption. Four model performance evaluation metrics are respectively calculated by Eqs. (9)–(12).

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (9)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (10)$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{y_i} \quad (11)$$

$$CV-RMSE = \sqrt{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2/n/(\sum_{i=1}^{n}y_i/n)} \quad (12)$$

where n is the number of test target data samples, $\hat{y}_i$ is the prediction value of target building power consumption, y_i is the actual value of target building power consumption.

3.3. Model hyper-parameter selection

To obtain an optimal prediction performance by the proposed transfer learning model, the model hyper-parameter should be determined after the data preprocessing. The L2 regularization weight, the learning rate, the neural units of LSTM and regression layer, and the batch size of source domain and target domain are optimized by the validation datasets. The candidate hyper-parameter settings for model selection in the proposed transfer learning models are summarized in Table 4. This study follows some of the network structure and parameters used in [19,33,40]. The feature extractor is LSTM neural network with one layer. The domain classifier and regression predictor are three and two fully connected layers, respectively. Dropout is commonly used to avoid over-fitting in deep learning, which is set to be 0.5 in this study. The remaining hyper-parameters configuration of different transfer learning models for the experiment are optimized by the validation datasets, including the L2 regularization weight, the learning rate, the neural units of LSTM and regression layer, and the batch size of source domain and target domain shown in Table 5. The training epochs are set as 200 using the Adam optimizer.

3.4. Model performance evaluation and results analysis

In this section, experiments are conducted with different number of source domains to verify the effectiveness of the proposed Multi-LSTM-DANN framework. The prediction performance results of different transfer learning models for the target building power consumption are analyzed and compared. Figs. 6–9 respectively show the actual and prediction values of the target building power consumption using the transfer learning models with different number of source domains from January 11–31. The black dotted curve represents the actual value (Mt) of target building power consumption, and the other solid curves represent the prediction values using different transfer learning models (M1, M2, M3, M4). It can be seen that the prediction values of most transfer learning models have similar trend with the actual values. These prediction results verify the effectiveness of the transfer learning method. In addition, the prediction values also show large differences for different single-source and multi-source transfer learning models. The main reason is that the similarity between each pair of source domain and target domain is different, which lead to different prediction results of singel-source and multi-source transfer learning models. For different multi-source transfer learning models, the model prediction performance does not always increase as the number of source domains increase. This is mainly because if there are more pairs of source domain and target domain, it will be more difficult for the multi-source transfer learning neural network to simultaneously find the domain invariant features between between each pair of source domain and target domain, especially when there exists large domain shift between the source domain and target domain.

Figs. 10–13 respectively show the prediction error density distributions of transfer learning models with different number of source domains from January 11–31. It can be seen that the prediction errors of all transfer learning models vary between the positive and negative values. For different transfer learning models, the prediction error distributions are also different. It can be seen that the prediction errors are normally distributed if the transfer learning models show a good prediction performance. The prediction errors of M1.1, M1.2, M2.1, M2.2, M2.3, M3.1, M3.2, M3.3, and M4 are all approximately normally distributed, as these transfer learning models show the better prediction performance shown in Figs. 6–9. The prediction errors of other transfer learning models (M1.3, M1.4, M2.4, M2.5, M2.6, and M3.4) are distributed irregularly as these models simultaneously show a poor prediction performance. These results are also consistent with the actual model prediction error distribution.

Table 6 shows the performance evaluation metrics and standard deviation of different transfer learning models. For each experiment, the results are averaged over ten times to reduce the random effect, and the mean and standard deviations of performance metrics are reported. It can be seen that all the transfer learning models can improve the prediction performance of the target building power consumption compared to the model trained on the target only data (MT). These results verify the effectiveness of the single-source and multi-source transfer learning models. In addition, the improved prediction performance of target building power consumption using different source building combinations and different number of source buildings show great differences. For the single-source transfer learning models, M1.1 and M1.2 show better prediction performance, while the prediction performance of M1.3 and M1.4 decreases. For the two-source transfer learning models, M2.1, M2.2 and M2.3 also show better prediction performance, while M2.4, M2.5 and M2.6 show poor prediction performance. Among them, M2.3 show the best prediction performance of the all compared transfer learning models, and the four performance evaluation metrics (MSE, MAE, MAPE, and CV-RMSE) are all the smallest. For the three-source transfer learning models, M3.1, M3.2 and M3.3 show superior prediction performance and M3.4 show a poor prediction performance. For the four-source transfer learning model, the prediction performance of M4 is not very good. The main reason is the divergence between the source domain and target domain. The best model (i.e. M1.1, M2.3, and M3.1) prediction performance of each source group increases as the number of source domains increase from 1 to 3, while the model prediction performance (M4) shows a decrease trend when the number of source domains is 4.

In addition, the multi-source transfer learning model can improve the model prediction stability. It can be seen from Table 6 that the standard deviation of single-transfer learning model is greater than that of multi-source transfer learning model (except M2.6), which reveal the multi-source transfer learning model is more stable than the single-transfer learning models. The model stability is also important in the actual application.

To verify the superior performance of the multi-source transfer learning models compared to the single-source transfer learning models, the average metrics of corresponding single-source transfer learning models can be obtained according to the results of Table 6. For the two-source transfer learning models, compared to the average metrics of the two corresponding single-source transfer learning models, M2.4 shows a prediction performance degradation, and the other two-source transfer learning models all show performance improvements. For example, the MSE of M2.4 is 355.0748 while the average MSE of two corresponding

Table 4
The candidate hyper-parameter settings for model selection.

Hyper-parameter	Range
L2 regularization weight	{0.001, 0.002, 0.005, 0.01, 0.02, 0.05, 0.1, 0.2, 0.5}
Learning rate	{0.001, 0.002, 0.005, 0.01, 0.02, 0.05, 0.1}
neural units (LSTM, regression layer)	{64, 100, 128, 512}
Batch size (source domain, target domain)	{64, 128, 256}

Table 5
Hyper-parameter configuration of different transfer learning models for the experiment.

Source group	Model ID	LSTM unit	Regression unit	Domain classifierlayer, unit	L2 regularization weight, learning rate	Batch size
Target	MT	64,	(100, 100, 1)	–	0.5, 0.001	256, 64
Single-source	M1.1	64	(128, 100, 1)	(100, 1)	0.1, 0.002	256, 64
	M1.2	128	(512, 512, 1)	(100, 1)	0.1, 0.002	256, 128
	M1.3	64	(512, 128, 1)	(100, 1)	0.5, 0.001	128, 256
	M1.4	64	(100, 100, 1)	(100, 1)	0.001, 0.005	64, 64
Two-source	M2.1	64	(100, 100, 1)	(100, 1)	0.2, 0.001	128, 64
	M2.2	64	(100, 100, 1)	(100, 1)	0.2, 0.001	256, 64
	M2.3	64	(512, 128, 1)	(100, 1)	0.1, 0.002	256, 128
	M2.4	64	(512, 128, 1)	(100, 1)	0.5, 0.001	256, 256
	M2.5	512	(512, 512, 1)	(100, 1)	0.1, 0.002	256, 64
	M2.6	64	(100, 100, 1)	(100, 1)	0.01, 0.001	64, 256
Three-source	M3.1	64	(128, 100, 1)	(100, 1)	0.1, 0.001	256, 64
	M3.2	64	(128, 100, 1)	(100, 1)	0.1, 0.002	256, 64
	M3.3	64	(128, 100, 1)	(100, 1)	0.1, 0.001	256, 64
	M3.4	64	(100, 100, 1)	(100, 1)	0.01, 0.01	256, 64
Four-source	M4	64	(512, 100, 1)	(100, 1)	0.1, 0.001	256, 64

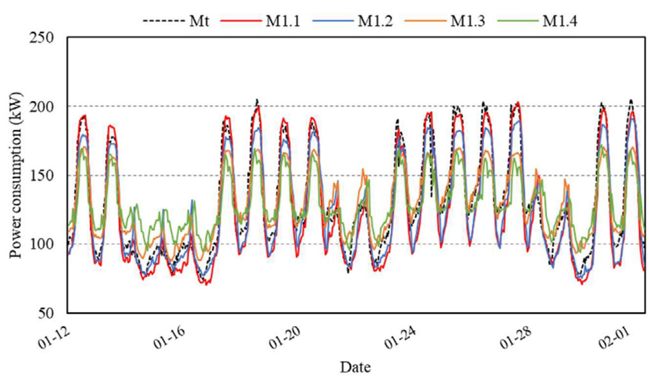

Fig. 6. The actual and prediction values of single-source transfer learning models from January 11–31.

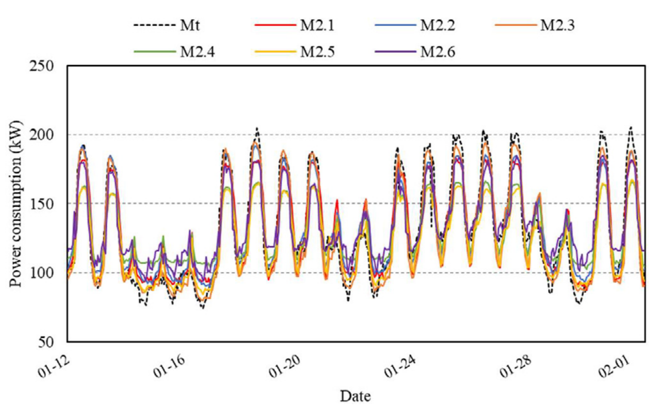

Fig. 7. The actual and prediction values of two-source transfer learning models from January 11–31.

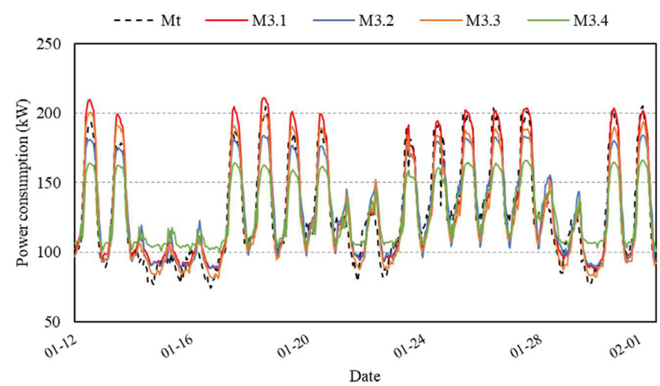

Fig. 8. The actual and prediction values of three-source transfer learning models from January 11–31.

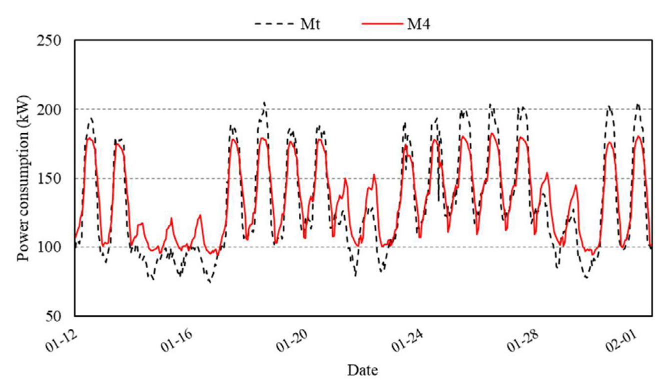

Fig. 9. The actual and prediction values of four-source transfer learning models from January 11–31.

single-source transfer learning models (M1.2 and M1.3) is 235.5260. For the three-source transfer learning models, compared to the average metrics of the three corresponding single-source transfer learning models, M3.4 similarly shows the prediction performance degradation, and the other three-source transfer learning models all show performance improvements. For the four-source transfer learning models, MAPE of M4 show an increase compared to the average value of the four corresponding single-source transfer learning models, and the other three metrics all show a decrease.

Fig. 14 shows the improvement percentages of performance evaluation metrics of multi-source transfer learning models over the average metrics of corresponding single-source transfer learning models. It can be seen that evaluation metrics (MAE, MSE, MAPE and CV-RMSE) of most multi-source transfer learning models have been improved compared to the average metrics of corresponding single-source transfer learning models. M2.4 and M 3.4 show the negative performance improvement percentages compared to the corresponding single-source transfer learning models, and the other multi-source transfer learning models (M1.1, M1.2, M1.3, M1.4, M2.1, M2.2, M2.3, M2.5, M2.6, M3.1, M3.2, M3.3, and M4) all show the positive performance improvement percentages.

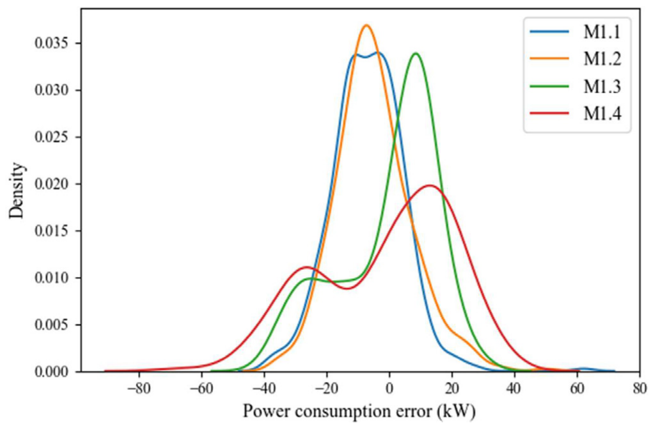

Fig. 10. The prediction error density distributions of single-source transfer learning models from January 11–31.

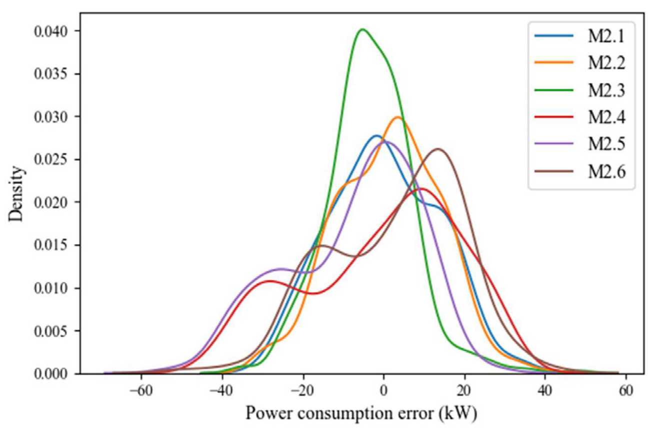

Fig. 11. The prediction error density distributions of two-source transfer learning models from January 11–31.

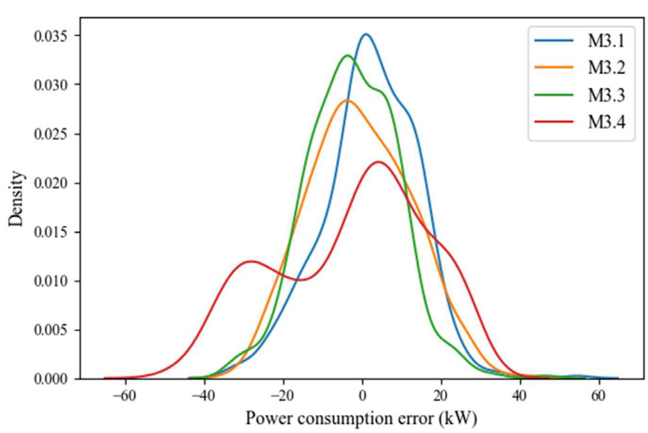

Fig. 12. The prediction error density distributions of three-source transfer learning models from January 11–31.

These results demonstrate the superiority of multi-source transfer learning models compared to the single-source transfer learning models. It should be noted that the prediction performance of multi-source transfer learning models cannot always be improved when the number of source building increases. For example, a source building data (S3) is added on the basis of Model1.2 (source building data: S2), while the prediction performance of Model2.4 (source building data: (S3, S3)) is worse than

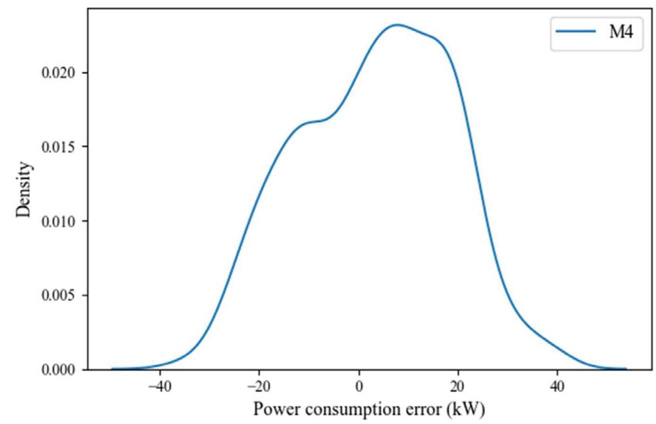

Fig. 13. The prediction error density distributions of four-source transfer learning model from January 11–31.

that of Model1.2. Similarly, the prediction performance of Model3.2 (source building data: (S1, S2, S4)) is worse than that of Model2.1 (source building data: (S1, S2)) when a source building data (S4) is added on the basis of Model2.1. This is mainly due to the increased distribution discrepancy between the added source building data and target building data, which lead to the model performance degradation. The positive improvement percentages of multi-source transfer learning models are almost over 10% compared to the average metrics of corresponding single-source transfer learning models. Specially, M2.3 show the highest performance improvement percentages of all compared transfer learning models. Overall, the three-source transfer learning models show better performance improvement percentages compared to two-source and four-source transfer learning models. M3.1, M3.2, M3.3 all show superior prediction performance.

To further discuss the relationship between MMD and the model prediction performance, Table 7–9 shows the results of MMD and weights of the multi-source (two-source, three-source, and four-source) transfer learning models. It should be noted that there is not an absolute negative correlation between the MMD and the model transfer performance for different models. In this study, MMD is used to metric the domain invariant features between different pairs of source and target domain within the same model. We can consider how to select the source domain from the perspective of ensemble learning, the ensemble model can show a good performance when the base learners have greater diversity. That is to say, the domain invariant features between different pairs of source and target domain should vary greatly. Therefore, the regression weights W (MMD) should vary greatly within the same model. In addition, the single-source transfer learning model M1.1 show the best performance. Therefore, the weights W1 should be larger, which can help to improve the model performance. It can be seen that M2.3 and M3.1 perform best for the two-source and three-source transfer learning models respectively. This is because the weight W1 and W2 of M2.3 vary greatly, and W1 accounts for a larger proportion 61.84%. Similarly, the weight W1, W2 and W3 of M3.1 vary greatly, and W1 accounts for a larger proportion 48.59%. Therefore, when selecting the source domains, while ensuring the similarity between the source domain and the target domain, it is necessary to maintain the divergence of different source domains, so as to effectively improve the efficiency of the multi-source ensemble transfer learning model.

Fig. 15 shows the t-SNE visualization of the original data features for the four source building data and the target building data. The different colors represent original source and target domain

Table 6
Performance evaluation metrics and standard deviation of different transfer learning models.

Source group	Model ID	MSE	MAE	MAPE	CV-RMSE
Target only	MT	683.2445 (46.3072)	17.9694 (0.4947)	0.1229 (0.0025)	0.2007 (0.0068)
Single-source	M1.1	180.7156 (31.3621)	10.3285 (1.0809)	0.0824 (0.0095)	0.1029 (0.0090)
	M1.2	184.0072 (29.9092)	10.8258 (1.1292)	0.0844 (0.0078)	0.1039 (0.0083)
	M1.3	287.0447 (22.2477)	13.8963 (1.1087)	0.1068 (0.0134)	0.1301 (0.0050)
	M1.4	435.6812 (35.5245)	16.7202 (1.0358)	0.1258 (0.0176)	0.1603 (0.0066)
Two-source	M2.1	175.5770 (11.9342)	10.7808 (0.3484)	0.0909 (0.0041)	0.1018 (0.0035)
	M2.2	192.5563 (16.4152)	11.3882 (0.5110)	0.0939 (0.0052)	0.1065 (0.0046)
	M2.3	139.1788 (13.6675)	9.1723 (0.4578)	0.0712 (0.0024)	0.0905 (0.0045)
	M2.4	355.0748 (16.4465)	15.6189 (0.3601)	0.1230 (0.0035)	0.1448 (0.0034)
	M2.5	286.8378 (28.1761)	13.0351 (0.5597)	0.0926 (0.0034)	0.1300 (0.0064)
	M2.6	313.0733 (62.0953)	14.5891 (1.7907)	0.1191 (0.0192)	0.1354 (0.0131)
Three-source	M3.1	153.5779 (12.2269)	9.8342 (0.5013)	0.0783 (0.0031)	0.0952 (0.0038)
	M3.2	189.6919 (24.3499)	11.2371 (0.7151)	0.0857 (0.0042)	0.1056 (0.0068)
	M3.3	163.1063 (13.5926)	10.1637 (0.4349)	0.0809 (0.0042)	0.0981 (0.0041)
	M3.4	343.0946 (23.0337)	14.8265 (0.6211)	0.1099 (0.0060)	0.1422 (0.0049)
Four-source	M4	223.2031 (19.6321)	12.5070 (0.6787)	0.1013 (0.0067)	0.1147 (0.0049)

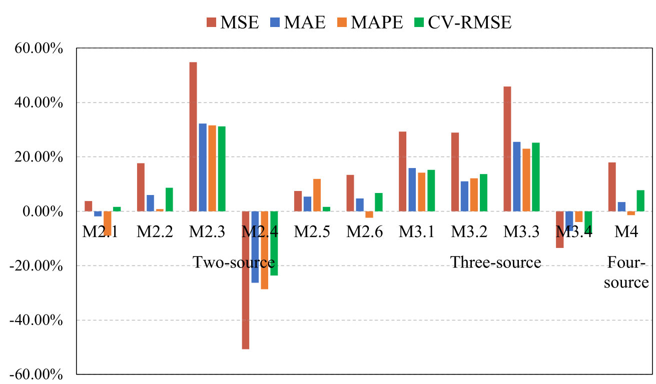

Fig. 14. The improvement percentages of performance evaluation metrics of multi-source transfer learning models over the average metrics of corresponding single-source transfer learning models.

data distribution by t-SNE visualization. It can be clearly seen that there exist obvious data distribution discrepancy between the four source building data and the target building data. The data distribution discrepancy is mainly due to the energy-consuming device, personal schedules and building envelope characteristics. Therefore, different source building data cannot be directly put into the same prediction model and should be first processed with data adjustment. The proposed LSTM-DANN method can effectively extract the domain invariant features between each pair of source domain and target domain. Different pairs of source building data and target building data can be processed by the proposed Multi-LSTM-DANN framework. Figs. 16–19 respectively show the t-SNE visualization of the domain invariant features for the single-source and multi-source (two-source, three-source, and four-source) transfer learning models extracted by the proposed Multi-LSTM-DANN framework. The different colors represent domain invariant feature distribution between different pairs of source and target domain. It can be seen that the extracted domain invariant features are distributed more concentrated compared to the original data features. These visualization results reveal that the proposed Multi-LSTM-DANN framework can well overcome the data distribution discrepancy of multi-source domains. Therefore, the model can be trained with the four source building data. In addition, the extracted domain invariant features between each pair of source domain and target domain are also different.

Table 10 and Fig. 20 shows the average performance evaluation metrics and standard deviations of each source group (single-source, two-source, three-source, and four-source) transfer learning models. It can be seen that the average performance evaluation metrics of each source group transfer learning models all show great improvements compared to the model trained on the target only data. Meanwhile, as the number of source domains increase from one to three, the performance evaluation metrics also decrease. When the number of source domains is four, performance evaluation metrics show an increase. Fig. 20 shows the improvement percentages of each group transfer learning models compared to the model trained on the target only data. The improvement percentages of MAE, MSE, MAPE and CV-RMSE are almost more than 15% compared to the model trained on the target only data. The improvement percentages achieve the maximum when the number of source domains is three. These results also demonstrate that the multi-source transfer learning models can enhance the prediction performance compared to the single-source transfer learning models when the number of source

Table 7
MMD and weights of the two-source transfer learning models.

Model ID	MMD_1	MMD_2	W_1	W_2
M2.1	0.0504	0.0284	0.3587	0.6413
M2.2	0.0645	0.0655	0.4853	0.5147
M2.3	0.1919	0.3228	0.6184	0.3816
M2.4	0.4984	0.6023	0.5519	0.4481
M2.5	0.0511	0.0579	0.5242	0.4758
M2.6	0.1712	0.2154	0.5077	0.4923

Table 8
MMD and weights of the three-source transfer learning models.

Model ID	MMD_1	MMD_2	MMD_3	W_1	W_2	W_3
M3.1	0.0506	0.1190	0.0871	0.4859	0.2079	0.3062
M3.2	0.0548	0.0473	0.0671	0.3319	0.3852	0.2829
M3.3	0.0543	0.0777	0.1401	0.4701	0.3492	0.1807
M3.4	0.0473	0.0754	0.0666	0.4236	0.2678	0.3086

Table 9
MMD and weights of the four-source transfer learning models.

Model ID	MMD_1	MMD_2	MMD_3	MMD_4	W_1	W_2	W_3	W_4
M4	0.0607	0.0933	0.1359	0.0899	0.3588	0.2358	0.1626	0.2428

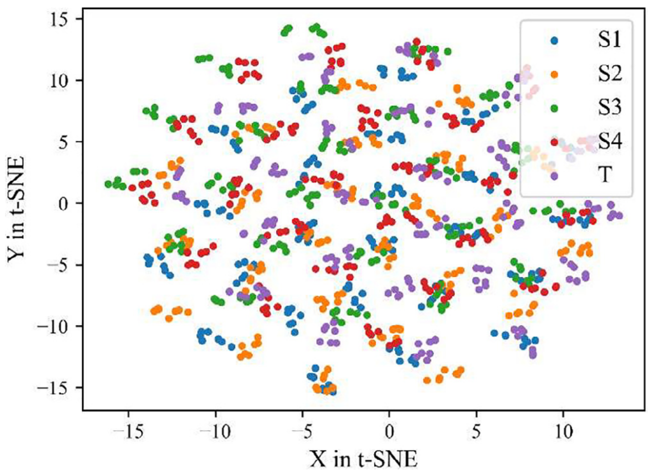

Fig. 15. The t-SNE visualization of the original data features for the four source building data and the target building data.

domains are suitable (not more than 3 in this study). However, the multi-source transfer learning models show a performance degradation when the number of source domains are too large (4 in this study).

3.5. Influence of source data volume on the model transfer performance

In Section 3.4, each source building collects one-year data for model performance evaluation and results analysis. In this section, the total source data volume is controlled to one year to discuss the influence of source data volume on the model transfer performance. For the two-source transfer learning model, half of the data per year (January–June) are used for each source domain. For the three-source transfer learning model, one third of the data per year (January to April) are used for each source domain. For the four-source transfer learning model, one quarter of the data per year (January to March) are used for each source domain. The scenario in Section 3.4 is defined as Scenario A, the scenario in this section is defined as Scenario B.

Figs. 21–24 shows the four performance evaluation metric (MSE, MAE, MAPE, and CV-RMSE) comparison of transfer learning models with different source data volume for the two scenarios. It can be clearly seen that the model performance gets worse when the amount of total source data volume is controlled to one year. Especially, the model performance of M2.6, M3.4, and M4 shows a significant degradation. Compared to the models trained by one-year source data volume in scenario A, the MSE of M2.6 increase from 313.07 to 573.54 when each source data volume decreases a half. The MSE of M3.4 increase from 343.09 to

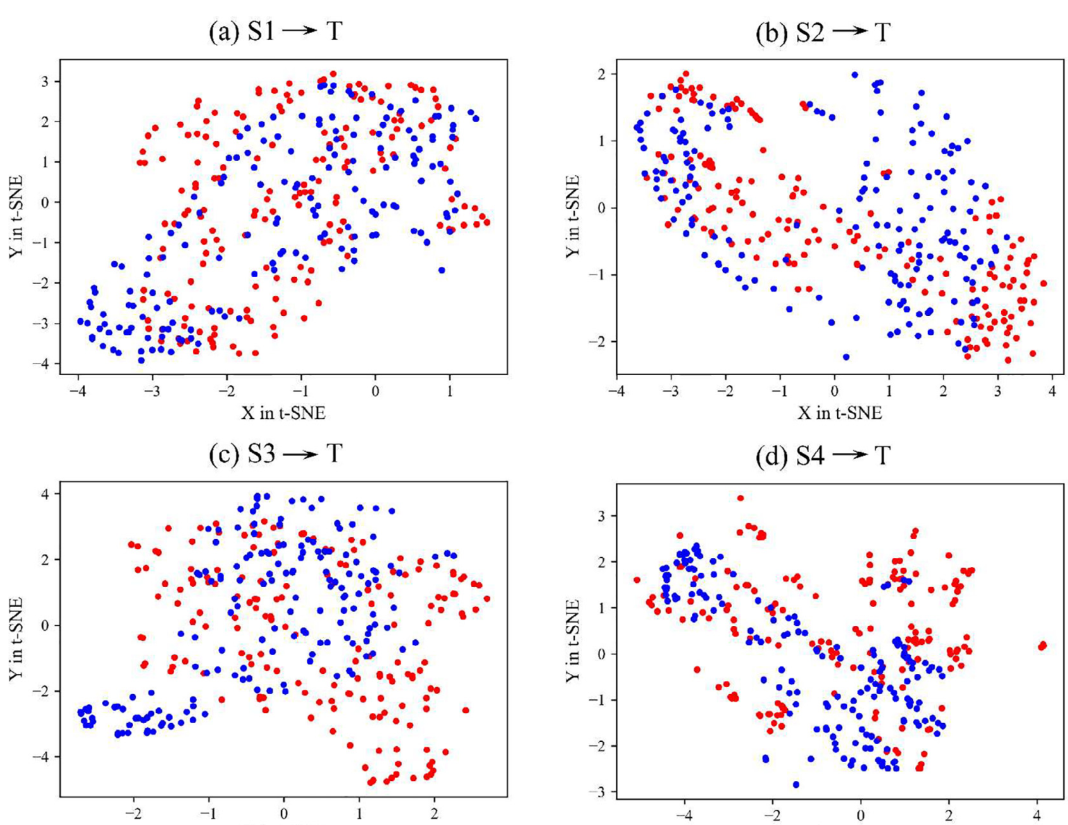

Fig. 16. The t-SNE visualization of the domain invariant features for the single-source transfer learning models.

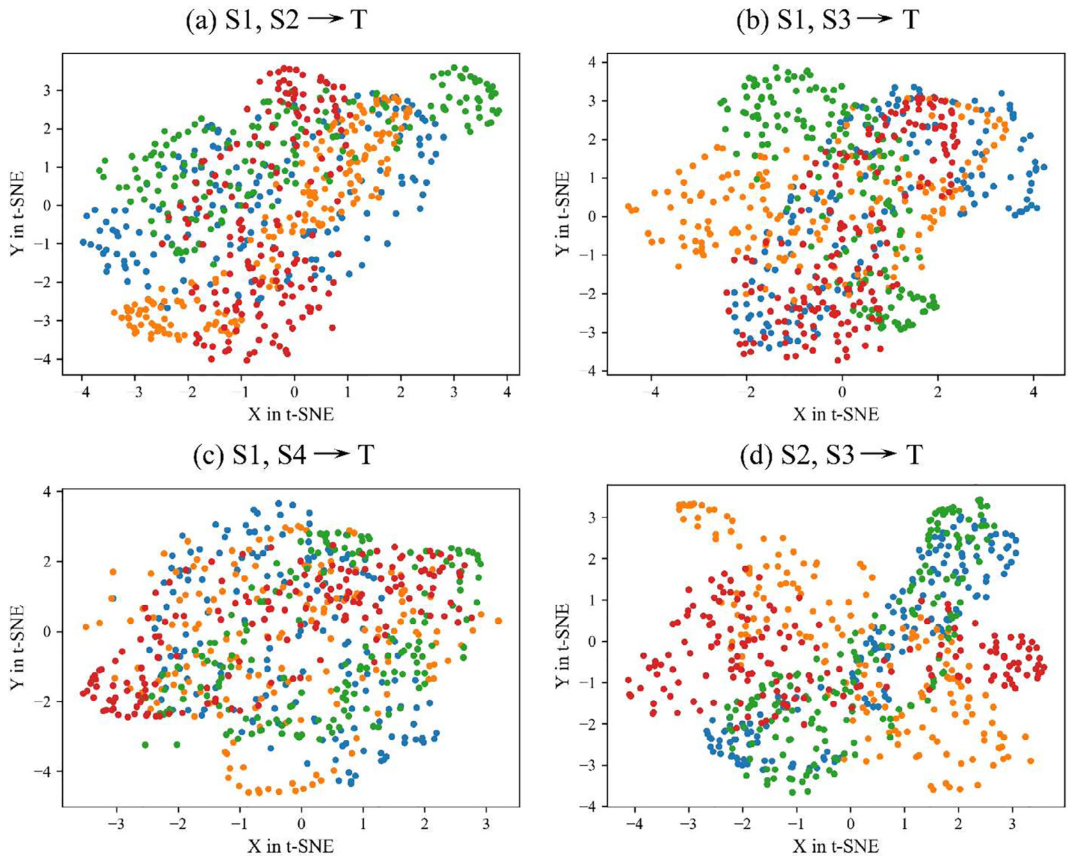

Fig. 17. The t-SNE visualization of the domain invariant features for the two-source transfer learning models.

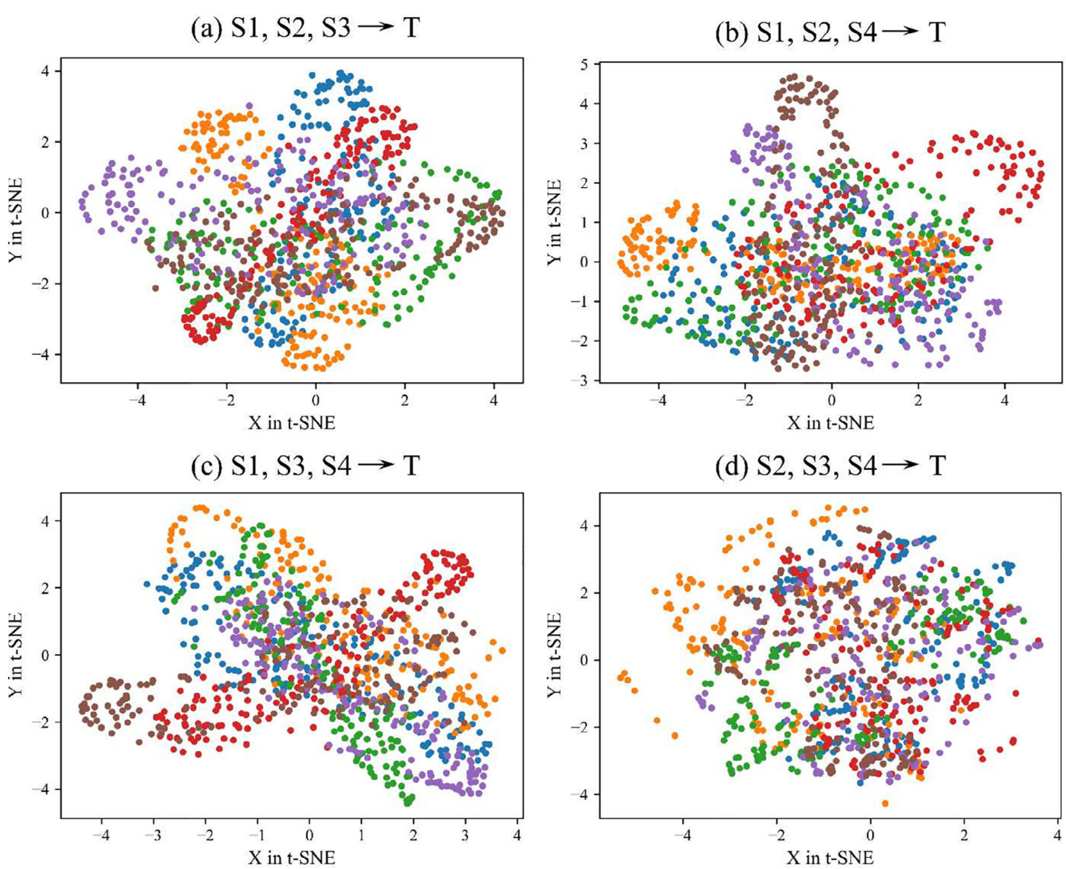

Fig. 18. The t-SNE visualization of the domain invariant features for the three-source transfer learning models.

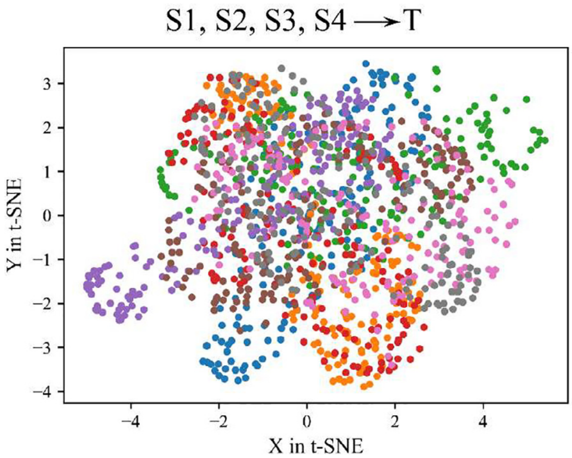

Fig. 19. The t-SNE visualization of the domain invariant features for the four-source transfer learning models.

557.79 when each source data volume decrease two-thirds. The MSE of M4 increase from 223.20 to 586.40 when the source data volume of each year decrease three-quarters. These results are reasonable as the training data decrease. This is mainly because some periodic characteristics are not reflected in the model, and some periodic information will be lost, which will reduce the knowledge learned from the source domain.

4. Conclusions

In this paper, a general multi-source ensemble transfer learning (Multi-LSTM-DANN) framework integrate of LSTM-DANN and similarity metric is proposed for building energy prediction. The proposed Multi-LSTM-DANN framework is aimed to create an efficient ensemble learning model to enhance the prediction performance of target task by using multi-source building data. LSTM-DANN is first used to extract the domain invariant features between each pair of source domain and target domain. Then MMD is applied to metric the distance between each pair of the extracted domain invariant features. Finally, the reciprocal of MMD is used as similarity metric index to calculate the regression weight and prediction value of the proposed Multi-LSTM-DANN model. Experiments are conducted to demonstrate the effectiveness of the proposed Multi-LSTM-DANN framework. Main conclusions can be obtained as follows:

(1) A general multi-source ensemble transfer learning (Multi-LSTM-DANN) framework is proposed for building power consumption prediction, which can effectively employ multi-source domain building data to enhance the prediction performance of target building power consumption.

Table 10
Average performance evaluation metrics and standard deviations of each source group transfer learning models.

Source group	MSE	MAE	MAPE	CV-RMSE
Target	683.2445	17.9694	0.1229	0.2007
Single-source	271.8622 (103.7947)	12.9427 (2.5736)	0.0998 (0.0178)	0.1243 (0.0234)
Two-source	243.7163 (78.8025)	12.4307 (2.2217)	0.0985 (0.0177)	0.1182 (0.0196)
Three-source	212.3677 (76.6268)	11.5154 (1.9808)	0.0887 (0.0125)	0.1103 (0.0188)
Four-source	223.2031	12.5070	0.1013	0.1147

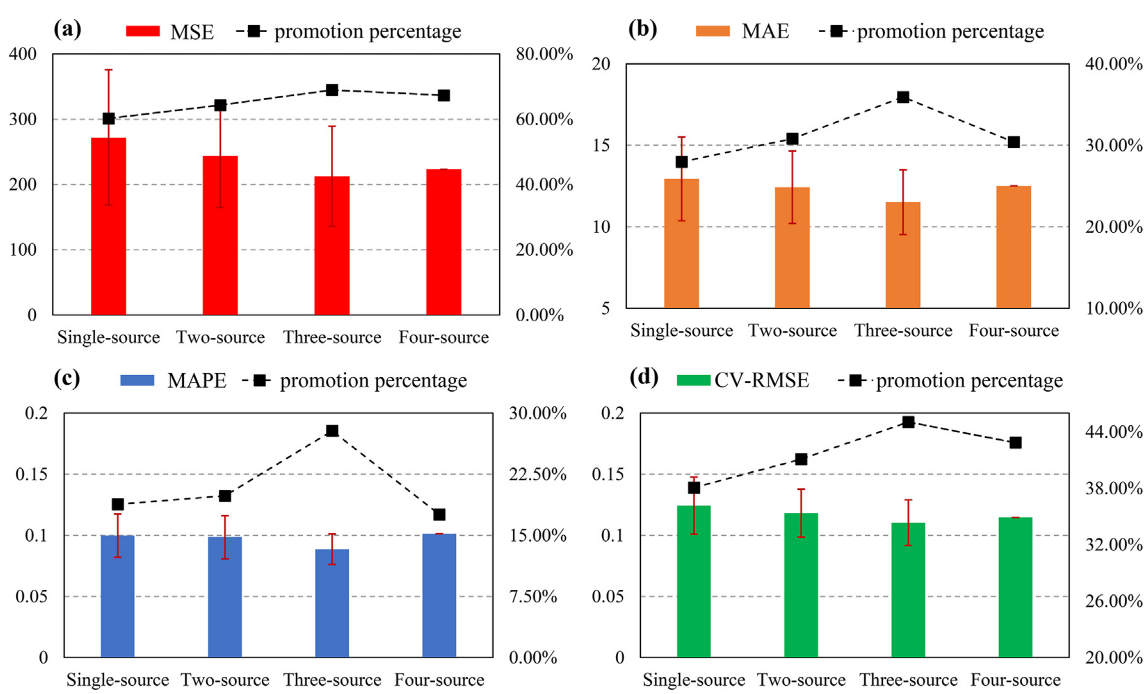

Fig. 20. Average performance evaluation metrics of each group transfer learning models and improvement percentages compared to the model trained on the target only data. (a) MAE; (b) MAPE; (c) RMSE; (d) CV-RMSE.

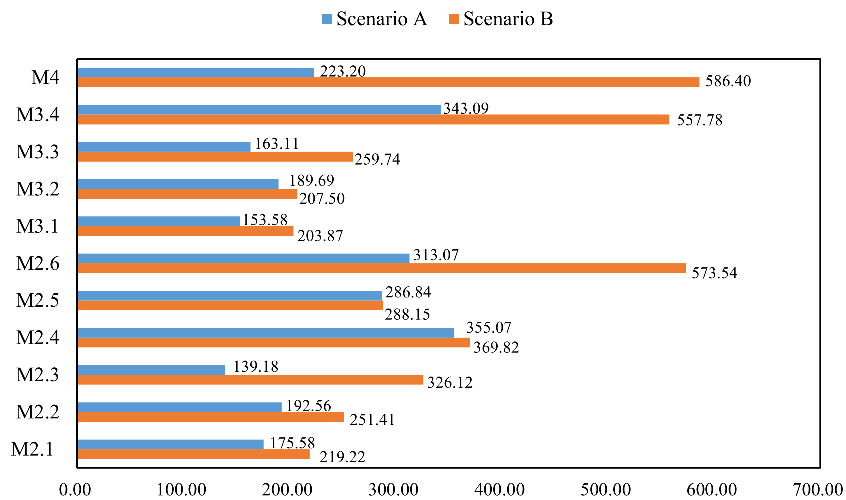

Fig. 21. Performance evaluation metric (MSE) comparison of transfer learning models with different source data volume.

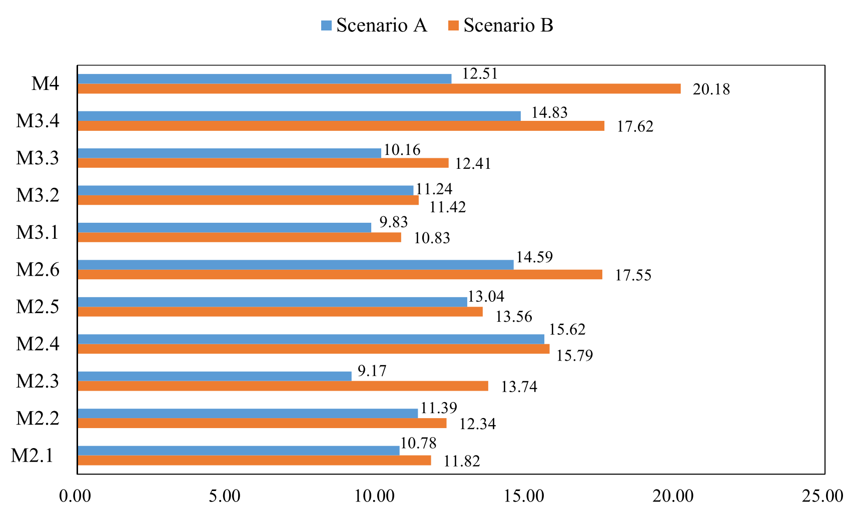

Fig. 22. Performance evaluation metric (MAE) comparison of transfer learning models with different source data volume.

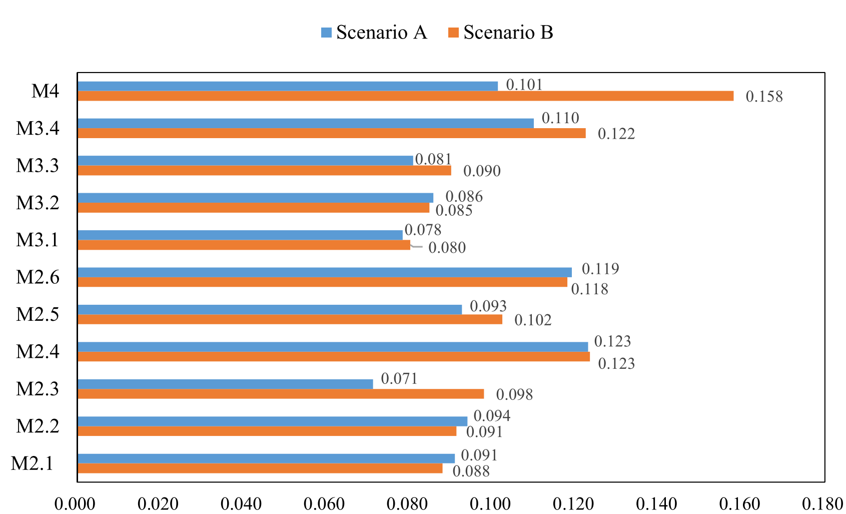

Fig. 23. Performance evaluation metric (MAPE) comparison of transfer learning models with different source data volume.

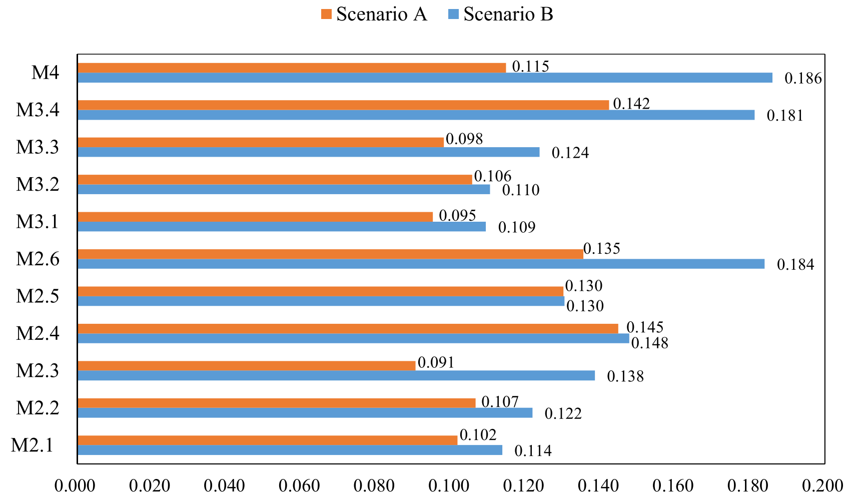

Fig. 24. Performance evaluation metric (CV-RMSE) comparison of transfer learning models with different source data volume.

(2) Single-source and multi-source transfer learning models can both improve the prediction performance of the target building power consumption compared to the model trained on the target only data. Most multi-source transfer learning models can enhance the prediction performance of the target building power consumption compared to the corresponding single-source transfer learning models.
(3) The improved prediction performance of target building power consumption using different source building combinations and different number of source buildings show great differences. There are an optimal combination and number of source domains for the performance promotion of multi-source transfer learning models. Three-source transfer learning models show an optimal prediction performance in this study.

The effectiveness of the proposed multi-source transfer learning framework has been validated by experiments. However, there are some points should be furtherly studied in the future work. For example, this study discusses the influence of source data volume on the model transfer performance. Due to the particularity of time series data, even if the time length of the same source domain is the same, the model transfer performance may vary depending on the time interval selected. The influence of the same time length for different time interval on the model transfer learning should be studied in the further work. In addition, the interpretation of transfer learning models can be an interesting and valuable research topic. This can help the users to understand what information is transferred in the black-box model based method, which can optimize the model performance further by the interpretation information.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The National Key Technology Support Program (Grant no. 2015BAJ03B01), the Hunan Provincial Innovation Foundation for Postgraduate Studies (Grant no. CX20190287), the Hunan Provincial Research and Development Plan of Key Areas (Grant no. 2020DK2003) and the Hunan Provincial Commercialization and Industrialization Plan of Scientific and Technological Achievements (Grant no. 2020GK2077) provided financial assistance for this study.

References

[1] X. Wu, C. Li, L. Shao, J. Meng, L. Zhang, G. Chen, Is solar power renewable and carbon-neutral: evidence from a pilot solar tower plant in China under a systems view, Renewable Sustainable Energy Rev. 138 (2021) 110655.
[2] Y. Zhao, T. Li, X. Zhang, C. Zhang, Artificial intelligence-based fault detection and diagnosis methods for building energy systems: advantages, challenges and the future, Renewable Sustainable Energy Rev. 109 (2019) 85–101.
[3] Z. Afroz, G.M. Shafiullah, T. Urmee, G. Higgins, Modeling techniques used in building HVAC control systems: a review, Renewable Sustainable Energy Rev. 83 (2018) 64–84.
[4] K. Amasyali, N.M. El-Gohary, A review of data-driven building energy consumption prediction studies, Renewable Sustainable Energy Rev. 81 (2018) 1192–1205.
[5] T. Ahmad, C. Huanxin, D. Zhang, H. Zhang, Smart energy forecasting strategy with four machine learning models for climate-sensitive and non-climate sensitive conditions, Energy 198 (2020) 117283.
[6] M. Bourdeau, X.Q. Zhai, E. Nefzaoui, X. Guo, P. Chatellier, Modeling and forecasting building energy consumption: a review of data-driven techniques, Sustainable Cities Soc. 48 (2019) 101533.
[7] H. Sha, P. Xu, C. Hu, Z. Li, Y. Chen, Z. Chen, A simplified HVAC energy prediction method based on degree-day, Sustainable Cities Soc. 51 (2019) 101698.
[8] W. Li, G. Gong, H. Fan, P. Peng, L. Chun, Meta-learning strategy based on user preferences and a machine recommendation system for real-time cooling load and COP forecasting, Appl. Energy 270 (2020) 115144.
[9] P. Wang, G. Gong, Y. Zhou, B. Qin, A simplified calculation method for building envelope cooling Loads in Central South China, Energies 11 (7) (2018) 1708.
[10] M. Cai, M. Pipattanasomporn, S. Rahman, Day-ahead building-level load forecasts using deep learning vs. traditional time-series techniques, Appl. Energy 236 (2019) 1078–1088.
[11] C. Fan, J. Wang, W. Gang, S. Li, Assessment of deep recurrent neural network-based strategies for short-term building energy predictions, Appl. Energy 236 (2019) 700–710.
[12] T.-Y. Kim, S.-B. Cho, Predicting residential energy consumption using CNN-LSTM neural networks, Energy 182 (2019) 72–81.
[13] J. Zheng, H. Zhang, Y. Dai, B. Wang, T. Zheng, Q.i. Liao, Y. Liang, F. Zhang, X. Song, Time series prediction for output of multi-region solar power plants, Appl. Energy 257 (2020) 114001.
[14] W. Li, G. Gong, H. Fan, P. Peng, L. Chun, X.i. Fang, A clustering-based approach for "cross-scale" load prediction on building level in HVAC systems, Appl. Energy 282 (2021) 116223.
[15] C. Fan, Y. Sun, Y. Zhao, M. Song, J. Wang, Deep learning-based feature engineering methods for improved building energy prediction, Appl. Energy 240 (2019) 35–45.
[16] Y. Sun, S. Wang, F.u. Xiao, Development and validation of a simplified online cooling load prediction strategy for a super high-rise building in Hong Kong, Energy Convers. Manage. 68 (2013) 20–27.
[17] Y. Chen, Z. Tong, Y. Zheng, H. Samuelson, L. Norford, Transfer learning with deep neural networks for model predictive control of HVAC and natural ventilation in smart buildings, J. Cleaner Prod. 254 (2020) 119866.
[18] J. Ma, J.C.P. Cheng, C. Lin, Y.i. Tan, J. Zhang, Improving air quality prediction accuracy at larger temporal resolutions using deep learning and transfer learning techniques, Atmos. Environ. 214 (2019) 116885.

[19] J. Ma, Z. Li, J.C.P. Cheng, Y. Ding, C. Lin, Z. Xu, Air quality prediction at new stations using spatially transferred bi-directional long short-term memory network, Sci. Total Environ. 705 (2020) 135771.
[20] S.J. Pan, Q. Yang, A survey on transfer learning, IEEE Trans. Knowl. Data Eng. 22 (10) (2010) 1345–1359.
[21] J. Li, W. Wu, D.i. Xue, P. Gao, Multi-source deep transfer neural network algorithm, Sensors 19 (18) (2019) 3992, https://doi.org/10.3390/s19183992.
[22] C. Fan, Y. Sun, F. Xiao, J. Ma, D. Lee, J. Wang, Y.C. Tseng, Statistical investigations of transfer learning-based methodology for short-term building energy predictions, Appl. Energy 262 (2020) 114499, https://doi.org/10.1016/j.apenergy.2020.114499.
[23] Y. Tian, L. Sehovac, K. Grolinger, Similarity-based chained transfer learning for energy forecasting with big data, IEEE Access 7 (2019) 139895–139908.
[24] Y. Gao, Y. Ruan, C. Fang, S. Yin, Deep learning and transfer learning models of energy consumption forecasting for a building with poor information data, Energy Build. 223 (2020) 110156.
[25] E. Mocanu, P.H. Nguyen, W.L. Kling, M. Gibescu, Unsupervised energy prediction in a Smart Grid context using reinforcement cross-building transfer learning, Energy Build. 116 (2016) 646–655.
[26] A.o. Li, F.u. Xiao, C. Fan, M. Hu, Development of an ANN-based building energy model for information-poor buildings using transfer learning, Build. Simul. 14 (1) (2021) 89–101.
[27] M. Ribeiro, K. Grolinger, H.F. ElYamany, W.A. Higashino, M.A.M. Capretz, Transfer learning with seasonal and trend adjustment for cross-building energy forecasting, Energy Build. 165 (2018) 352–363.
[28] X.i. Fang, G. Gong, G. Li, L. Chun, W. Li, P. Peng, A hybrid deep transfer learning strategy for short term cross-building energy prediction, Energy 215 (2021) 119208.
[29] Y. Zhu, F. Zhuang, D. Wang, Aligning domain-specific distribution and classifier for cross-domain classification from multiple sources, Proc. AAAI Conf. Artif. Intell. 33 (2019) 5989–5996.
[30] H. Zhao, S. Zhang, G. Wu, J. Costeira, J. Moura, G. Gordon, Multiple Source Domain Adaptation with Adversarial Training of Neural Networks, 2017.
[31] P. Huang, G. Wang, S. Qin, Boosting for transfer learning from multiple data sources, Pattern Recogn. Lett. 33 (5) (2012) 568–579.
[32] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780.
[33] Y. Gani, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, et al., Domain-Adversarial Training of Neural Networks, 2015.
[34] A. Gretton, K. Borgwardt, M. Rasch, B. Schölkopf, A.J. Smola, A kernel two-sample test, J. Mach. Learn. Res. 13 (2012) 723–773.
[35] F. Zhou, Z. Jiang, C. Shui, B. Wang, B. Chaib-draa, Domain generalization via optimal transport with metric similarity learning, Neurocomputing 456 (2021) 469–480.
[36] P. Zhao, T. Wu, S. Zhao, H. Liu, Robust transfer learning based on Geometric Mean Metric Learning, Knowl.-Based Syst. 227 (2021) 107227.
[37] S. Pan, J. Kwok, Q. Yang, Transfer Learning via Dimensionality Reduction, 2008.
[38] Y. Ganin, V. Lempitsky, Unsupervised Domain Adaptation by Backpropagation, 2014.
[39] C. Miller, F. Meggers, The Building Data Genome Project: An open, public data set from non-residential building electrical meters, Energy Procedia 122 (2017) 439–444.
[40] P.Rd.O. da Costa, A. Akçay, Y. Zhang, U. Kaymak, Remaining useful lifetime prediction via deep domain adaptation, Reliab. Eng. Syst. Saf. 195 (2020) 106682.