The Predictor for Urban Buildings’ Hourly Electricity Consumption

Shubing Shan1, Buyang Cao2 and Zhiqiang Wu2

1 School of Software Engineering, Tongji University, China
2 College of Architecture and Urban Planning, Tongji University, China
E-mail: 1410786@tongji.edu.cn

Abstract: The accurate prediction approach of urban buildings’ electricity consumption is an important foundation for smart urban energy management. It provides the decision basis for electricity deployment at peak times. This paper presents a knowledge graph of urban building electricity consumption called ECKG. It provides an effective way to obtain the influencing factors of buildings’ electricity consumption. In addition, in order to improve the accuracy of prediction, this paper proposes the logarithmic electricity consumption gravity model and error correction model based on data selection. We use 17520 hours’ electricity consumption of a five-star hotel building in Shanghai, China as the study case, and 8 common models as benchmarks to conduct the comparisons. Our approach outperforms all benchmarks in terms of average accuracy.

1. Introduction
Building energy consumption accounts for 30-45% of global energy consumption and buildings’ electricity consumption is a major part of building energy consumption [1]. According to the latest statistics from the World Bank, the global urbanization population has increased by about 3.9% and the per capita electricity consumption has increased by 1100 kWh (kilowatt hour) in the last decades [2]. The electricity consumption of urban buildings has risen sharply worldwide. Precise forecasting hourly electricity consumption of urban buildings plays an important role in optimizing the use of energy in urban buildings and realizing energy-saving operations. It can help electricity supply departments improve their deployment strategies and avoid electricity shortages at peak times. This is necessary to formulate a reasonable urban energy production plan and reduce carbon emissions.

Nowadays, many researches of forecasting urban buildings’ electricity consumption focus on three topics, including the factors that influence the electricity consumption of urban buildings, predicting methods of urban building electricity consumption and data collecting methods of influencing factors of urban building electricity consumption. The influencing factors of building electricity consumption are mainly divided into internal factors such as occupancy rate, architectural attributes such as...
window-to-wall ratio, area, while external factors such as weather and climate [3][4]. The prediction methods are including macroscopic and microscopic methods [5][6]. The macro-prediction methods treat all the buildings in an area as a building group and use the macro factors such as climate and economy etc. to predict the electricity consumption of the building group. The microscopic methods use microscopic factors such as building attributes and weather etc. to predict the electricity consumption of a single building. Data collecting methods include traditional manual statistical methods and automatic acquisition methods using hardware devices such as sensors [7][8].

However, the above-mentioned researches still face two challenges. It does not integrate the influencing factors that have been studied in the past and establish the relationship between these factors. This leads many studies to use prior subjective experience to obtain influencing factors. In addition to using some classic methods to predict electricity consumption, new methods are needed to get better predictions. To overcome these obstacles this paper proposes an innovative approach for predicting the electricity consumption of urban buildings. It employs the factors that influence the electricity consumption of urban buildings as knowledge and builds a knowledge graph of building electricity consumption. Then, it uses a new method to predict the electricity consumption of a building.

The contributions of this paper are as follows:

- A new knowledge graph. We propose the knowledge graph of building electricity consumption called ECKG and describe its construction process.
- We propose the logarithmic electricity consumption gravity model (LE_GRA) upon the gravity model and error correction model based on data selection (ECSDS).
- We employ the real data of a hotel building in Shanghai, China to illustrate the forecasting process and validate the model. The commonly applied 8 prediction models are employed as the benchmarks to evaluate the proposed model.

This paper is organized as follows. Section 2 describes the methodology. In section 3, it presents a case study to prepare the data following by the section for comparative analyses and discussions. The paper concludes with some remarks.

2. Methodology

Our method includes two parts, a knowledge graph of urban building electricity consumption and a predictor for the electricity consumption of urban buildings.

2.1 Knowledge graph of urban building electricity consumption

There are three steps in constructing the knowledge graph of urban building electricity consumption: extracting target words and constructing a co-word matrix; converting the co-word matrix into a knowledge similarity matrix and clustering these knowledge; establishing the relationship of knowledge.

- Extracting target words and constructing a co-word matrix

We use keyword rules to extract target words from each reference and use co-word analysis to count the number of occurrences of each two target words in the same reference as co-word occurrences. Each element in the co-word matrix represents the number of occurrences of two words. The form of co-word matrix is as follows:

- Constructing knowledge similarity matrix and clustering knowledge
We treat each target word in the co-word matrix as a knowledge. However, the word frequency in the co-word matrix varies significantly, which is the burden to cluster knowledge and it needs to be normalized. For this, we use the Ochiai coefficient method to convert the co-word matrix to the knowledge similarity matrix, as shown in formula 1. According to Ochiai coefficient a similarity’s value will be between 0 and 1. Then, we use hierarchical clustering to cluster knowledge based on similarities and form various knowledge clusters. The knowledge in the same cluster all together usually solves one problem.

\[
\text{Ochiai coefficient} = \frac{\text{the frequency of the two words co-occurrence}}{\sqrt{\text{the frequency of the 1st word}} \cdot \sqrt{\text{the frequency of the 2nd word}}}
\]

- Establishing the relationship of knowledge

The relationships of knowledge include the one within the knowledge cluster and the one crossing different clusters. We establish the relationship between a pair of knowledge one by one in a knowledge cluster. We create a link for a knowledge to another one that is most similar. If two knowledge have already been established the relationship, we choose to associate the knowledge with the next most similarity, until the relationship of all the knowledge in the cluster has been established. More relationships indicate that this knowledge is more important in the cluster. We compare the similarities of all knowledge in the two knowledge clusters and select the two most similar knowledge to establish the relationship or link.

2.2 LE_GRA model

Since the gravitational model is suitable for analyzing the flow changes between two different locations and has good generalization, it is widely used in various areas such as immigration, transportation, tourism, etc., even though its parameters and variables need to be appropriately tuned [9] [10] [11]. We can revise a gravity model for the purpose of predicting electricity consumptions. The electricity consumption of a building is affected by many factors, and the electricity consumption over a period of time can be derived from the difference in meter readings at two points in time. If different points in time are treated as different “locations” in time, and the increase or decrease of electricity consumption is viewed as a type of flow change, then above mentioned difference (in meter readings) reflects the flow change between different locations. Therefore, electricity consumption can be determined by the gravity model with proper modifications. The electricity consumption gravitational model is presented in equation (2) and the associated parameters are as follows. \( E \) is the electricity consumption; \( x \) is the influence factors affecting the electricity consumption such as temperature; \( Z \): a special event factor such as the number of days for the building renovation and/or major activities in the building etc.; \( n \): the number of factors; \( m \): the number of special events; \( a, b, c \): coefficients; \( d \): the duration between two points in time (hours).

\[
E_{ij} = a_0 x_1^{a_1} x_2^{a_2} \cdots x_n^{a_n} d_i^{b_1} Z_2^{b_2} \cdots Z_m^{b_m}
\]

(2)
The electricity consumption represented by equation (2) will vary with the influencing factors, and the proportion of these changes may be different. If the dispersion of \( E \) gradually increases (decrease) with the increase (decrease) of \( x \), then the collected data is heteroscedastic [12]. To address this issue, the logarithmic transformation is one of the effective means to reduce the heteroscedasticity [13]. Therefore, in order to mitigate the heteroscedasticity effect in the dataset, we take the logarithm on both sides of equation (2) and obtain the logarithmic electricity consumption gravity model.
LE_GRA, as shown in equation (3). The meanings of the parameters in equation (3) are the same as those in equation (2). As described above, the influencing factors in our study are temperature, humidity, and wind speed. There is no influencing factor for any special event. Substituting these relevant influencing factors into equation (3) we will be obtaining equation (4).

$$\ln E_{ij} = \ln a_0 + \sum_{k}^{n} a_k \ln x_k + \sum_{k}^{m} b_k \ln z_k + \ln \frac{1}{d_{ij}}$$  (3)

$$\ln E_{ij} = \sum_{k}^{n} a_k \ln x_k + \bar{Y}, \quad (\bar{Y} = \ln a_0 + \ln \frac{1}{d_{ij}})$$  (4)

2.3 ECSDS model

We extract some training data similar to the data distribution of the testing set upon 8 statistical characteristics including mean, median, mode, variance, range, skewness, and coefficient of variation. Then, the extracted training data is recorded as $D$. Based on the data set $D$, Johansen is used for co-integration test to determine the co-integration relationship, and the VEC model is constructed to correct the prediction results. Since the data distribution of $D$ is similar to the test set, it is beneficial to construct a correction model with stronger correction ability. The above is the error correction process of the ECSDS model.

3. Study case

3.1. A miniature knowledge graph of urban building electricity consumption

We use the collected data as the key word source to extract the target words from 50 papers. We extracted 30 target words as 30 knowledge of urban building electricity consumption. Figure 1 illustrates the knowledge graph using 30 knowledge based on above described steps. The result plots a miniature knowledge graph of urban building electricity consumption called ECKG. Each color in figure 1 represents a cluster and each cluster indicates a knowledge subgraph with its own color. The size of a node illustrates the number of that knowledge being used.

3.2. Data preparation

We took a five-star hotel building in Shanghai, China as the study case, and collected its hourly electricity consumption data and relevant influencing factors’ data from September 2013 to September 2015. We take the most used knowledge (larger size of the node) and next related knowledge in ECKG as the impact factors of urban building electricity consumption in the case study. However, since the architectural properties of the hotel building and the surrounding environment showed in the diagram will not change in the short term and the behaviors of the guests staying presented in the graph at hotel are also difficult to obtain, we have extracted weather, temperature, humidity, wind speed, occupancy, time, solar radiation intensity from the ECKG as the substitutes for prediction. The collected data is listed as follows and we apply a normalization processing to all data for the better prediction accuracy.
3.3. Partitioning datasets

We divide the 24-month of data into training, testing datasets. The short-term prediction period of building electricity consumption is from 1 to 6 months [14]. In order to fully evaluate the predictive capability of a model, we set various test datasets to fit the short-term prediction. Data for 1-, 2-, 3-, 4-, 5- and 6-month are used as the testing sets to evaluate the short-term prediction.

4. Prediction Results

In order to evaluate the qualities of the prediction results, we utilize MAPE (Mean Absolute Percentage Error) [15] illustrated by equation (5), where $A_t$ is the actual value and $F_t$ is the forecast value.

$$M = \frac{100\%}{n} \sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right|$$

(5)

In the computational experiments, we employ 8 commonly applied forecasting models including Ordinary Least Squares Regression(OLSR), Bayesian Ridge Regression(BRR), Stochastic Gradient Descent Regression(SGDR), Support Vector Regression(SVR), Nearest Neighbors Regression(NNR), Decision Tree Regression(DTR), Multi-layer Perceptron Regression(MLP), Autoregressive Integrated Moving Average model(ARIMA) as the benchmarks. Some applications of these models can be found in [16] [17]. In this case the proposed LE_GRA model can be evaluated comprehensively.
Figure 2. The comparison of each model’s predicted MAPEs in various experiments.

- Observations of the prediction results

Figure 2 shows the MAPE for each model’s prediction results in various experiments (dataset divisions of various experiments are mentioned in section 3.3). This helps us to observe the generalization capabilities of benchmarks and the proposed model. The MAPEs of the predicted results of all models including the LE_GRA are oscillating in all experiments (A to F) (i.e. the time periods of testing data range are from 1- to 6-month). In the time period of testing data is 2-month, the prediction results of these models are the worst comparing to other experiments except ARIMA. However, the predicted results of LE_GRA appear to be less fluctuant than the predictions of 8 benchmarks in various experiments. The prediction results of the LE_GRA model are significantly better after being corrected by ECSDS.

We use four metrics including maximum MAPE, minimum MAPE, average MAPE and variance to evaluate the predictions of each model in different experiments. Figure 3 presents the metrics of prediction results obtained by 8 benchmarks and LE_GRA. In terms of the minimum MAPE, LE_GRA model is ranked fourth among all models. However, LE_GRA model performs the best among all models considering the maximum MAPE and average MAPE. The maximum MAPE of LE_GRA is 4.85% lower than the second one and the average MAPE of LE_GRA is 0.41% lower than the second one. This demonstrates that the proposed LE_GRA model has overall better accuracy. After the prediction of the LE_GRA model was corrected by ECSDS, the maximum MAPE decreased by 3.59%, the minimum MAPE was 4.99%, and the average MAPE decreased by 1.77%.
**Figure 3.** Metrics of prediction results of 8 benchmarks and LEGRA

5. Conclusions

Our approach proposed in the paper applies the knowledge graph to obtain the suitable factors that impact the electricity consumption of buildings, which not only solves the above problems, but also avoids the subjectivity of prior experience. We propose an ensemble prediction model called LEGRA with the purpose of producing better predictions for electricity consumption. According the comprehensive computational experiments, it reveals that the proposed LEGRA performs better than and 8 benchmarks with respect to the average and maximum prediction accuracy. The computational experiments validate that the LEGRA model is capable to deliver better results for short-term forecasts. ECSDS has good error correction effect.

6. Reference

[1] D.A. Asimakopoulos et al., "Modelling the energy demand projection of the building sector in Greece in the 21st century," *Energy and Building*, vol.49, 2012, pp.488–498.

[2] The World Bank, "Indicators," 2017, http://data.worldbank.org/indicator.

[3] R.V. Jones, A. Fuertes and K.J. Lomas, "The socio-economic, dwelling and appliance related factors affecting electricity consumption in domestic buildings," *Renewable and Sustainable Energy Reviews*, vol.43, March 2015, pp.901–917.

[4] G.M. Huebner et al., "Explaining domestic energy consumption-The comparative contribution of building factors, socio-demographics, behaviours and attitudes," *Applied Energy*, vol.159, December 2015, pp.589–600.

[5] B. Liddle, "Impact of population, age structure, and urbanization on carbon emissions/energy consumption: evidence from macro-level, cross-country analyses," *Popul Environ*, vol.35, no.3, March 2014, pp.286–304.

[6] Y.T. Chae et al., "Artificial neural network model for forecasting sub-hourly electricity usage in commercial buildings," *Energy and Buildings*, vol.111, January 2016, pp.184–194.

[7] R.K. Jain et al., "Forecasting energy consumption of multi-family residential buildings using support vector regression: Investigating the impact of temporal and spatial monitoring granularity on performance accuracy," *Applied Energy*, vol.123, June 2014, pp.168–178.

[8] M.S. Guil, S. Patidar, "Understanding the energy consumption and occupancy of a multi-purpose academic building," *Energy and Buildings*, vol.87, January 2015, pp.155–165.

---

**Table:**

| EXPERIMENT ID | MAXIMUM MAPE | MINIMUM MAPE | AVERAGE MAPE |
|---------------|--------------|--------------|--------------|
| OLSR          | 38.85%       | 19.37%       | 25.31%       |
| SVR           | 38.85%       | 19.01%       | 26.66%       |
| SGDR          | 39.08%       | 19.01%       | 26.66%       |
| SGDR          | 42.48%       | 19.48%       | 26.66%       |
| BRR           | 38.82%       | 19.48%       | 26.66%       |
| NNR           | 39.08%       | 19.48%       | 26.66%       |
| DTR           | 38.82%       | 19.48%       | 26.66%       |
| MLPR          | 39.08%       | 19.48%       | 26.66%       |
| ARIMA         | 41.01%       | 24.10%       | 24.10%       |
| LE_GRA        | 42.74%       | 13.05%       | 23.43%       |
| ECSDS         | 99.99%       | 27.68%       | 33.97%       |
[9] Ramos, R., & Suriñach, J. “A Gravity Model of Migration Between the ENC and the EU”. *Tijdschrift voor economische en sociale geografie*, 108(1), pp.21-35, 2017.

[10] Sayed, M. A., Rahman, M. M., Zaber, M. I., & Ali, A. A. (2017, December). “Understanding Dhaka city traffic intensity and traffic expansion using gravity model”, *In Computer and Information Technology (ICCIT) 2017 20th International Conference of IEEE*, Dhaka Bangladesh, 2017, pp. 1-6.

[11] Becker K, Terekhov I, Gollnick V. A global gravity model for air passenger demand between city pairs and future interurban air mobility markets identification, in *2018 Aviation Technology, Integration, and Operations Conference*. Atlanta, Georgia, USA, 2018: 2885.

[12] Klein A G, Gerhard C, Büchner R D, et al. “The detection of heteroscedasticity in regression models for psychological data”, *Psychological Test and Assessment Modeling*, 58(4), 567, 2016.

[13] Chatterjee S, Hadi A S. “Regression analysis by example”, *John Wiley & Sons*, 2015.

[14] Makridakis, S., Spiliotis, E., & Assimakopoulos, V. “Statistical and Machine Learning forecasting methods: Concerns and ways forward”, *PloS one*, 13(3), 2018.

[15] Moy, Ronald L, Li-Shya Chen, and Lie Jane Kao. “Multiple Linear Regression,” *Study Guide for Statistics for Business and Financial Economics*. Springer, Cham, pp.223-240, Dec.2015.

[16] Pedregosa F, Varoquaux G, Gramfort A, et al. “Scikit-learn: Machine learning in Python”, *Journal of machine learning research*, 12(Oct), pp.2825-2830, 2011.

[17] Brockwell P J, Davis R A, Calder M V. “Introduction to time series and forecasting”. *New York: springer*, 2002.

Acknowledgement
This research is supported in part by China Intelligent Urbanization Co-Creation Center for High Density Region and Tongji Architectural Design(Group) Co., Ltd. under grant No. CIUC20150011; and in part by the China Scholarship Council under grant No.201706260039.