Big-data driven building retrofitting: An integrated Support Vector Machines and Fuzzy C-means clustering method

Weizhuo Lu\textsuperscript{a}, Kailun Feng\textsuperscript{a, b, c, d}

\textsuperscript{a} Department of Civil, Environmental and Natural Resources Engineering, Luleå University of Technology
\textsuperscript{b} Department of Construction Management, Harbin Institute of Technology, Harbin 150009, China
\textsuperscript{c} Key Lab of Structures Dynamic Behavior and Control of the Ministry of Education, Harbin Institute of Technology, Harbin 150090, China
\textsuperscript{d} Key Lab of Smart Prevention and Mitigation of Civil Engineering Disasters of the Ministry of Industry and Information Technology, Harbin Institute of Technology, Harbin 150090, China

Abstract:

It has become a mainstream to use physical models to quantify expected energy savings from alternative retrofit methods and technologies. However, they are not suitable for predicting energy use of buildings when detailed and specified input parameters are unavailable. The overall purpose of the research is to support the stakeholders in taking decisions on refurbishments options when not all of physical information is available, in order to achieve the Swedish Energy Agency’s measurements of near-zero energy buildings. The research will transfer big data from Swedish Energy Performance Certificates for building retrofitting. A Support Vector Machines and Fuzzy C-means clustering (SVM-FCM) integrated machine learning algorithm is used directly to extract the case-specific knowledge from EPC big data regarding building characteristics and energy saving of retrofit measures. It enables to prioritize retrofit measures and compute their expected energy savings for buildings. This proposed data driven method is an attempt of taking advantage of big data for practical building retrofit selection.

1. Introduction

A large share of buildings in Sweden was constructed in the so-called Million Homes Program between 1965 and 1975. During the next 10 years, the Million Homes Program buildings will reach a 50-year service life, which presents the opportunities for mitigating energy use. However, when there are some unknown/uncertainty parameters in existing buildings, e.g. the precise component size and exact insulated performance, identifying the most effective retrofit measures and predicting the expected energy savings remain difficult in practice [1-3]. Older buildings further add difficulty because of undocumented system changes, deterioration of components over time and incomplete design documentation.

Two main approaches have been taken for predicting energy-savings from a set of retrofit technologies: physical-based approaches and data-driven approach. Examples of physical-based approaches include EnergyPlus, eQuest, and Ecotect, which aim to represent a building as operated; that is, the model should capture the building systems as-installed, as-operated, and as-used [4]. These types of models require detailed building parameters such as building construction details; operation schedules; HVAC design information; and climate, sky, and solar/shading information [5]. However, such detailed data may not always be available [6], especially for older buildings with deterioration of components over time and incomplete design documentation.
To address the challenge of some unknown/uncertainty parameters in existing buildings, in previous studies the uncertainty parameters have usually been set to default values or Monte Carlo analysis has been applied. However, a single set of default values clearly cannot cover all possible values of unknown/uncertainty parameters, and may lead to unreliable results and decisions [7]. In Monte Carlo analyses, the probability distributions have been used to obtain values for unknown/uncertainty parameters, rather than using default values. However, knowledge of the probability distributions of the unknown/uncertainty parameters is required. It is highly difficulty to determine the probability distributions of unknown/uncertainty parameters [8], especially as they will be influenced by the characteristics of individual buildings.

We are seeing increasing availability of big data on actual building energy use and building characteristics, therefore, there is an emerging opportunity to effectively utilize these big data to develop a data-driven approach. Data-driven building prediction does not require full detailed building parameters, alternatively, it learns from historical/available data for predictions [5, 9]. To do so, this research will use big data from Swedish Energy Performance Certificates (EPC). The database and machine learning algorithms are used directly to extract the case-specific knowledge regarding building characteristics and actual building energy use in order to compute the expected energy savings of alternative retrofit methods and technologies. This method represents energy savings by prediction intervals, a value range that future performance will locate in certain confidence, e.g. 95 percent.

Therefore, the research will propose a data-driven and data-learning method for assessing energy saving, without using default values or requiring knowledge of unknown/uncertainty parameters’ probability distributions. The method is expected to provide a robust retrofit solution by computing prediction intervals of expected energy savings and prioritizing retrofits methods.

2. Data-driven method for retrofit strategy selection

2.1 Knowledge from Energy Performance Certificates

To obey the Energy Performance Building Directive (EPBD), all European Union members are obligated to pursue energy efficient buildings in national realm. In Sweden, Energy Performance Certificates (EPC) is developed to follow this European Communities directive. EPC of a building is conducted by a trained expert. The EPC includes basic building properties and energy consumption by sensors or estimation. Therefore, EPC provides an overall energy performance of the building for the stakeholder, for instance, house owners, renters, possible buyers. The purpose of EPC is to make buildings to be economical with energy and to motivate the building stakeholder to make the building energy efficient. In this way EPC helps reduce harmful emissions to the climate and to our living environment [10].

According to current Swedish regulations, all newly-built buildings, rental, and not rental buildings, such as multi-dwelling blocks and detached and semi-detached houses must have EPC. Only a few types of buildings can have exceptions. At present, about 550,000 buildings have been certificated by EPC, which involves detached houses, apartments and public, commercial or industrial facilities [11]. EPC has become a big data source that comprises massive information of building energy performance.

In a standard EPC, it has four sections of information related to the certificated building. The first is building identification information, such as address, cities, and etc.. The second is the building properties, such as building type, building area and use, number of floors and stairs. The third is the energy performance that includes energy consumption and contribution from ventilation, heating, cooling, lighting, watering and so on. The last major information is the recommendations by experts for building’s cost-effective retrofit measures and predicted energy savings. Therefore, in the EPC big
data, it contains knowledge about the connection between building properties and energy savings by specific retrofit measures. However, this connection is implied into big data and a proper method is required to extract and model it. This research will reveal this connection by machine learning and extract it for any building’s retrofit selection. Table 1 shows the building properties of each EPC sample, and Table 2 shows the suggested retrofit strategies of each EPC sample. The building properties and suggested retrofit strategies are combined as features, and the corresponding expected energy savings are regarded as labels.

Table 1. The building property of an EPC sample.

| No. | Building property                                             | No. | Building property                                             |
|-----|--------------------------------------------------------------|-----|--------------------------------------------------------------|
| 1   | Building category                                            | 28  | Is type of ventilation system F with recycling?              |
| 2   | Building type                                                | 29  | Available air conditioning systems with nominal cooling power greater than 12 kW? |
| 3   | Location type                                               | 30  | District heating (#1) kWh                                    |
| 4   | Construction year                                            | 31  | Heating by oil (#2) kWh                                      |
| 5   | Heated area (> 10 °C except storage room) m²                 | 32  | Heating by natural gas (#3) kWh                              |
| 6   | Obtained ways of heated area m²                             | 33  | Heating by firewood (#4) kWh                                 |
| 7   | Number of basement floors heated (> 10 °C except storage room) m² | 34  | Heating by tile / pellets / briquettes (#5) kWh              |
| 8   | Storage room area m²                                        | 35  | Heating by other biofuel (#6) kWh                            |
| 9   | Number of floors above ground                               | 36  | Heating by electricity (waterborne) (#7) kWh                 |
| 10  | Number of stairs                                            | 37  | Electricity (direct acting) (#8) kWh                         |
| 11  | Number of residential apartments                            | 38  | Electricity (airborne) (#9) kWh                              |
| 12  | Average outside wind flow, l/s,m²                           | 39  | Ground heat pump (electricity) (#10) kWh                     |
| 13  | Available electrical power for heating and water (> 10 W/m²) | 40  | Heat pump exhaust air (electricity) (#11) kWh                |
| 14  | Estimated energy use is indicated without measurable consumption or normal year is not corrected | 41  | Heat pump-air / air (electricity) (#12) kWh                   |
| 15  | Is there solar heat?                                        | 42  | Heat pump-air / water (electricity) (#13) kWh                |
| 16  | Is there a solar cell system?                                | 43  | Energy for above heating and hot water                       |
| 17  | Normal year adjusted value (degree days) kWh                 | 44  | Of which energy for hot water preparation                   |
| 18  | Normal Year Adjusted Value (Energy Index)                    | 45  | District cooling (#14) kWh                                   |
| 19  | Reference value 1 (according to new building requirements) kWh/m²/year | 46  | Solar collector area m²                                     |
| 20  | Reference value 2, min (statistical range) kWh/m²/year      | 47  | Solar cell area m²                                          |
| 21  | Reference value 2, max (statistical range) kWh/m²/year      | 48  | Building electricity (#15) kWh                               |
| 22  | EPC Version                                                 | 49  | Household electricity (#16) kWh                              |
| 23  | Is there a requirement for regular ventilation control in the building? | 50  | Operational electricity (#17) kWh                            |
| 24  | Is type of ventilation system FTX?                          | 51  | Comfort cooling electricity (#18) kWh                         |
| 25  | Is type of ventilation system F?                            | 52  | The building's energy use (#7-13, #15-18) kWh                |
| 26  | Is type of ventilation system FT?                           | 53  | The building's energy use (As defined in the Swedish Housing Agency's building rules) (#1-15, 18) kWh |
| 27  | Is type of ventilation system natural ventilation?           | 54  | The electricity that is included in the building's energy use (#7-13, 15, 18) kWh |
Table 2. The retrofit strategy of an EPC sample.

| No. | Retrofit strategy                      | No. | Retrofit strategy                           |
|-----|----------------------------------------|-----|---------------------------------------------|
| 1   | New radiator valves                    | 18  | Energy efficient lighting                   |
| 2   | Adjustment of heating system           | 19  | Insulation of pipes and ventilation ducts  |
| 3   | Time / need control of heating system  | 20  | Replacement / installation of heat pump     |
| 4   | Cleaning and / or aeration of heating  | 21  | Replacement / installation of energy efficient heat source |
| 5   | Maximum indoor temperature limit       | 22  | Replacement / completion of ventilation system |
| 6   | New indoor sensor                      | 23  | Recovery of ventilation heat                |
| 7   | Replacement / installation of pressure controlled pumps | 24  | Other action on installation               |
| 8   | Other action on heating system         | 25  | Additional insulation of attic ceiling / roof |
| 9   | Adjustment of ventilation system       | 26  | Additional insulation walls                |
| 10  | Timing of ventilation system           | 27  | Additional insulation basement / ground     |
| 11  | Need control of ventilation system     | 28  | Installation of solar cells                |
| 12  | Replacement / installation of speed controlled fans | 29  | Installation of solar heating              |
| 13  | Other action on ventilation            | 30  | Change to energy efficient windows / window doors with inner window |
| 14  | Time / need control of lighting        | 31  | Complement window / window doors with inner window |
| 15  | Time / need control of cold            | 32  | Sealing windows / window doors / exterior doors |
| 16  | Other action on lighting, cooling      | 33  | Other measure (construction)               |
| 17  | Hot water saving measures              |     |                                             |

2.2 SVM-FCM based knowledge learning

Machine learning has been developed to be a set of statistical learning algorithms technology that can learn from data and make accurate prediction [12]. In all type of algorithms, Support Vector Machines (SVM) with a wide applicable property has been utilized in many fields. SVM takes a balance of high learning efficiency and high accuracy [13]. Motived by its advantages, SVM is employed in this research as the core learning algorithm for knowledge extraction and modeling.

A Support Vector Machines and Fuzzy C-means clustering integrated method (SVM-FCM) is established to establish the building retrofit selection model. The overall framework of SVM-FCM method is shown as Figure 1. SVM-FCM method establishes a model that provides prediction intervals of energy savings for various retrofit strategies. The prediction intervals, which contain the effects of uncertain building parameters, enable a robust building retrofit selection.

In the first stage, the data of EPC is divided into three sections naming training, validation, and testing sets. For the training set, FCM is used to cluster EPC samples into clusters that have similar building properties. According to statistic theory, the samples with similar features have similar performances [14]. The FCM iterative algorithm is performed as Eq.1, which should subject to Eq.2 and 3.
EPC database

Baseline prediction model

SVM

Sample clustering

Prediction intervals of each cluster and each sample

SVM

Mappings of building property and prediction intervals of retrofits’ energy saving

Retrofits selection

Figure 1. Overall framework of SVM-FCM method for retrofit selection

\[ J_w = \sum_{i=1}^{N} \sum_{c=1}^{C} (\mu_{i,c} \| x_i - m_c \|^2), \]  
(1)
\[ \mu_{i,c} = \frac{1}{\sum_{k=1}^{C} (\| x_i - m_k \| \| x_i - m_k \| ^2)}, \]  
(2)
\[ m_c = \frac{\sum_{i=1}^{N} (\mu_{i,c} x_i)}{\sum_{i=1}^{N} (\mu_{i,c})}, \]  
(3)

In the above equations, \( \| \cdot \| \) represents the Euclidean norm; \( x_i \) represents the inputs of the i-th EPC sample; \( \mu_{i,c} \) represents the membership grade of the i-th sample to the c-th cluster, and is defined by (2); and \( m_c \) represents the centre of the c-th cluster, defined by (3).

SVM is used for two parts, baseline prediction model and mappings of building property with prediction intervals of retrofits’ energy saving. In the baseline prediction model, the combined building properties and suggested retrofit measures are features. And the expected energy saving for each year when adopts retrofit measures are labels. In the second mapping model, the features are same as baseline prediction model, but the labels are calculated prediction intervals. SVM is trained to the regression models for above connections. Based on the above prediction intervals model, any building with only basic known building parameters need to be retrofitted can prioritize alternative retrofit measures.

For the prediction intervals calculations, the prediction intervals of cluster should be obtained as prerequisite for and prediction intervals of each sample. The prediction intervals of each cluster is selected to be the residual error of samples that are located at \( \alpha/2 \) and \( 1-\alpha/2 \) of total membership grade of the sample in that cluster when the samples are sort to be ascending order by residual error. And the prediction intervals of each sample are calculated as Eq.4, where \( u_k \) is the prediction intervals of sample k, and c represents the cluster and \( U \) is the prediction intervals of cluster c. More details on prediction intervals calculation can be refer at research by [15].

\[ u_k = \sum_{c=1}^{C} (\mu_{k,c} U_c), \]  
(4)
3. Results and discussion

3.1 Overall validation

In this research, 39,870 buildings located in Stockholm, Sweden in EPC database are selected as big data for building retrofit analysis. The SVM-FCM integrated algorithms are programming in Matlab2017a platform. Buildings are divided as training, validation, and testing set as 70, 15, and 15 percent of all data. In the final testing stage, the accuracy ratios of established prediction intervals that correctly cover testing samples are 91.5 percent (see Figure 2). It implies that the established model can correctly predict any building’s energy savings under different retrofit strategies in more than 90 percentages. In Figure 2, the red cross represents actual energy savings of each sample in testing set, and the lines represent obtained upper and lower prediction intervals by proposed SVM-FCM method. The results in Figure 2 validate the proposed SVM-FCM integrated method.

![Figure 2](image)

Figure 2. Testing results of SVM-FCM method.

3.2 Retrofit application

After validation of proposed method, three buildings in Stockholm, Sweden in EPC database are used as case examples. The basic properties of these buildings can be seen in Table 3. Some energy related important parameters are not provided, such as U-value of building components, window wall ratios, building wall materials, envelop sharp and etc.. In total 33 single alternative retrofit strategies (see Table 2) are analyzed in this application, the results of predicted energy savings can be seen in Figure 3 and Table 4.
Table 3. Building property of three case buildings.

| Property No. | Building #1 | Building #2 | Building #3 | Property No. | Building #1 | Building #2 | Building #3 |
|--------------|-------------|-------------|-------------|--------------|-------------|-------------|-------------|
| 1            | 321         | 400         | 320         | 28           | 0           | 0           | 0           |
| 2            | 2           | 4           | 2           | 29           | 0           | 0           | 0           |
| 3            | 3           | 2           | 2           | 30           | 184873      | 0           | 193593      |
| 4            | 1931        | 1958        | 1952        | 31           | 0           | 46000       | 0           |
| 5            | 1462        | 2056        | 1186        | 32           | 0           | 0           | 0           |
| 6            | 2           | 1           | 2           | 33           | 0           | 0           | 0           |
| 7            | 1           | 0           | 1           | 34           | 0           | 0           | 0           |
| 8            | 0           | 0           | 0           | 35           | 0           | 0           | 0           |
| 9            | 6           | -1*         | 2           | 36           | 0           | 10000       | 0           |
| 10           | 1           | -1*         | 11          | 37           | 0           | 25000       | 0           |
| 11           | 23          | -1*         | 11          | 38           | 0           | 86000       | 0           |
| 12           | -1*         | -1*         | -1*         | 39           | 0           | 0           | 0           |
| 13           | 0           | 0           | 0           | 40           | 0           | 11000       | 0           |
| 14           | 0           | 0           | 0           | 41           | 0           | 0           | 0           |
| 15           | 0           | 0           | 0           | 42           | 0           | 0           | 0           |
| 16           | 0           | 0           | 0           | 43           | 184873      | 178000      | 193593      |
| 17           | 229158      | 186458      | 226372      | 44           | 45472       | 5000        | 36882       |
| 18           | 223576      | 196351      | 222933      | 45           | 0           | 0           | 0           |
| 19           | 110         | 100         | 110         | 46           | 0           | 0           | 0           |
| 20           | 107         | 181         | 135         | 47           | 0           | 0           | 0           |
| 21           | 130         | 223         | 165         | 48           | 20100       | 12500       | 3248        |
| 22           | 2010        | 2010        | 2010        | 49           | 0           | 0           | 0           |
| 23           | 1           | 1           | 1           | 50           | 0           | 0           | 0           |
| 24           | 0           | 1           | 0           | 51           | 0           | 2000        | 0           |
| 25           | 1           | 1           | 0           | 52           | 20100       | 146500      | 3248        |
| 26           | 0           | 1           | 0           | 53           | 204973      | 192500      | 196841      |
| 27           | 0           | 0           | 1           | 54           | 20100       | 146500      | 3248        |

*Note: * represents this property is not given.

Figure 3. Predicted energy savings of retrofit strategies for case buildings.
Table 4. The predicted energy savings of each retrofit strategy (kWh).

| No. | Retrofit strategy | Building #1 | Building #2 | Building #3 |
|-----|-------------------|-------------|-------------|-------------|
| 1   | New radiator valves | [3685, 10733] | [0, 15718]  | [12054, 30086] |
| 2   | Adjustment of heating system | [11780, 19105] | [0, 16732]  | [18272, 36236] |
| 3   | Time / need control of heating system | [38403, 45451] | [10664, 28410] | [2010, 20042] |
| 4   | Cleaning and / or aeration of heating | [2398, 9446] | [0, 14432]  | [2088, 20120] |
| 5   | Maximum indoor temperature limit | [12447, 19495] | [0, 12102]  | [0, 13159]   |
| 6   | New indoor sensor | [7302, 14350] | [0, 11637]  | [23971, 42003] |
| 7   | Replacement / installation of pressure controlled pumps | [7775, 14823] | [0, 16067]  | [0, 13356]   |
| 8   | Other action on heating system | [3273, 10321] | [0, 13467]  | [9189, 27221] |
| 9   | Adjustment of ventilation system | [5683, 12731] | [18581, 36328] | [0, 17752] |
| 10  | Timing of ventilation system | [14342, 21389] | [33505, 51251] | [12619, 30650] |
| 11  | Need control of ventilation system | [7650, 14698] | [14720, 32467] | [5269, 23300] |
| 12  | Replacement / installation of speed controlled fans | [174, 7222] | [0, 12208]  | [1163, 19195] |
| 13  | Other action on ventilation | [12920, 19968] | [7207, 24953] | [0, 14149] |
| 14  | Time / need control of lighting | [0, 4881] | [0, 10913]  | [0, 8742]   |
| 15  | Time / need control of cold | [0, 3584] | [0, 8570]   | [0, 8742]  |
| 16  | Other action on lighting, cooling | [0, 4671] | [0, 9656]   | [0, 8742]  |
| 17  | Hot water saving measures | [6267, 13315] | [0, 13022]  | [6898, 24929] |
| 18  | Energy efficient lighting | [0, 5075] | [0, 10455]  | [0, 10332] |
| 19  | Insulation of pipes and ventilation ducts | [0, 3584] | [0, 8570]   | [8071, 26102] |
| 20  | Replacement / installation of heat pump | [293767, 300815] | [78969, 96716] | [163070, 181101] |
| 21  | Replacement / installation of energy efficient heat source | [27311, 34359] | [0, 13737]  | [98407, 116356] |
| 22  | Replacement / completion of ventilation system | [20457, 27505] | [9502, 27248] | [6588, 24620] |
| 23  | Recovery of ventilation heat | [11391, 18439] | [60809, 78556] | [30162, 48194] |
| 24  | Other action on installation | [969, 8017] | [0, 13003]  | [27621, 45652] |
| 25  | Additional insulation of attic ceiling / roof | [7066, 14114] | [649, 18396] | [3682, 21713] |
| 26  | Additional insulation walls | [69040, 76088] | [1847, 19593] | [63178, 81210] |
| 27  | Additional insulation basement / ground | [2920, 9968] | [0, 14953]  | [0, 8742]  |
| 28  | Installation of solar cells | [10414, 17462] | [8295, 26041] | [8146, 26177] |
| 29  | Installation of solar heating | [0, 4211] | [0, 11520]  | [68350, 86382] |
| 30  | Change to energy efficient windows / window doors with inner window | [2007, 9055] | [0, 14041]  | [4086, 22118] |
| 31  | Complement window / window doors with inner window | [0, 4696] | [0, 9681]   | [0, 12515] |
| 32  | Sealing windows / window doors / exterior doors | [0, 3584] | [0, 8570]   | [14587, 32619] |
| 33  | Other measure (construction) | [0, 3584] | [0, 8570]   | [0, 8742]  |

If setting the lower level of prediction intervals as benchmark (conservative strategy), Figure 3 shows that for case building #1 the best three retrofit strategies are strategy No.20, No.26, and No.3; for case building #2 the best three retrofit strategies are strategy No.20, No.23, and No.10; and for case building #3 the best three retrofit strategies are strategy No.20, No.21, and No.29. Therefore, the results indicate that the buildings with different properties have different suitable retrofit strategies. And proposed SVM-FCM integrated method enables prioritizing retrofits for buildings.

Although replacement/installation of heat pump (strategy No.20) is the most energy saving measure for all case buildings, the cost or unavailability may prevent it to be adopted. To achieve the similar energy savings, some second best or third best strategies that are available can be adopted. The proposed method is more feasible that can provide alternative options in the situation that best energy measure is not available or cost impractical. In EPCs, in general the experts only provide one or two
retrofit strategies. It is the same case that the proposed method is more feasible that the house owner can compare different strategies.

4. Conclusion

The retrofitting becomes an important part to improve the energy efficiency of buildings, especially for old buildings that reach their service life. This research introduced a big-data driven method SVM-FCM, which is suitable for the situation that some detail physical information is not available thus physical model method is not applicable. The application of the developed data-driven method in three actual buildings in Stockholm demonstrated that it can provide accurate prediction intervals of various retrofits’ energy savings. The results enable prioritizing and providing the most effective retrofit measure or measure combination.

To apply proposed data-driven method, a reliable big data source is extremely curial. In Sweden, Energy Performance Certificates can be a very suitable data source. Then, a proper feature and label selection and data preparation is conducted. For a reliable and fast learning, SVM-FCM method requires the data to be normalized. After running the data learning algorithms, the retrofitting model is developed. For any building that has same characteristics, such as in same city or region no matter outside or inside the big data, developed model can be used to support retrofit measures selection.

In this research, only part of buildings in Stockholm is taken into account. A wider application is required to test the further applicability of proposed method. Meanwhile, after establishment of the retrofit model that only considers energy savings, a comprehensive objective such as combination of required cost, retrofit suitability, and energy savings is valuable and more practical attempt in future work.

Acknowledgments

This research was supported by China Postdoctoral Science Foundation (grant no. 2020M670918) and the Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning (Formas).

References:

[1] Collinge WO, DeBlois JC, Landis AE, Schaefer LA, Bilec MM (2015). Hybrid Dynamic Empirical Building Energy Modeling Approach for an Existing Campus Building. Journal of Architectural Engineering, 22(1).

[2] Tian W, Yang S, Li Z, Wei S, Pan W, Liu Y (2016). Identifying informative energy data in Bayesian calibration of building energy models, Energy and Buildings, 119, 363-376.

[3] Nagpal S, Mueller C, Aijazi A, Reinhart CF (2019). A methodology for auto-calibrating urban building energy models using surrogate modeling techniques, Journal of Building Performance Simulation, 12(1), 1-16.

[4] Heo Y, Choudhary R, Augenbroe GA (2012). Calibration of building energy models for retrofit analysis under uncertainty, Energy and Buildings, 47, 550-560.

[5] Amasyali K, El-Gohary NM (2018), A review of data-driven building energy consumption prediction studies, Renewable and Sustainable Energy Reviews, 81(1), 1192-1205.
[6] Mathew PA, Dunn LN, Sohn MD, Mercado A, Custudio C, Walter T (2015). Big-data for building energy performance: Lessons from assembling a very large national database of building energy use, Applied Energy, 140, 85-93.

[7] Tian W, Yang S, Zuo J, Li Z, Liu Y (2017). Relationship between built form and energy performance of office buildings in a severe cold Chinese region. Building Simulation, 10, 11-24.

[8] Rezaee R, Brown J, Augenbroe G, Kim, J (2015). Assessment of uncertainty and confidence in building design exploration. AI EDAM, 29(4), 429-441.

[9] Walter T and Sohn MD (2016). A regression-based approach to estimating retrofit savings using the Building Performance Database, Applied Energy, 179, 996-1005.

[10] The National Board of Housing, Building and Planning of Sweden (Boverket), Energy Performance Certificate – three steps towards benefits, https://www.boverket.se/sv/om-boverket/publicerat-av-boverket/publikationer/2010/energy-performance-certificate/ <2020-4-2>

[11] Yan Z and Wang J (2014). Robust model predictive control of nonlinear systems with unmodeled dynamics and bounded uncertainties based on neural networks, IEEE Transactions Neural Network Learning System, 25, 457-469.

[12] Johansson T, Vesterlund M, Olofsson T, Dahl J (2016). Energy performance certificates and 3-dimensional city models as a means to reach national targets–A case study of the city of Kiruna. Energy Conversion and Management, 116, 42-57.

[13] Maji S, Berg AC, Malik J (2008). Classification using intersection kernel support vector machines is efficient. 2008 IEEE conference on computer vision and pattern recognition, 1-8.

[14] Shrestha DL and Solomatine DP (2006). Machine learning approaches for estimation of prediction interval for the model output. Neural Networks, 19(2), 225-235.

[15] Feng K, Lu W, Wang Y (2019). Assessing environmental performance in early building design stage: An integrated parametric design and machine learning method, Sustainable Cities and Society, 50.