Energy and Buildings 156 (2017) 78–84

Contents lists available at ScienceDirect

Energy and Buildings

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

Replication Studies

Sampling for building energy consumption with fuzzy theory

Jinghong Qin, Jili Zhang*

Faculty of Infrastructure Engineering, Dalian University of Technology, Dalian 116024, China

ARTICLE INFO

Article history:
Received 28 April 2017
Received in revised form 4 August 2017
Accepted 16 September 2017
Available online 18 September 2017

Keywords:
Fuzzy theory
Information entropy
Sampling survey
Building energy consumption
Sample size

ABSTRACT

The foundation of energy saving is knowing the real status of building energy consumption. For various kinds and a great number of building energy consumption data, the fuzzy theory is applied for sampling. It would make data representational. Firstly, a fuzzy clustering method is used to classify the data set and then the samples are extracted from the subclass. A modified clustering algorithm based on entropy weight method is proposed. It can determine the number of the classification of data set. The simulation results indicate that the new method can directly determine the optimal sample size. This method is suitably applied for dynamic energy consumption data and is more accurate compared with the statistical method.

1. Introduction

Building energy consumption has been gradually derived and also sparked attention. It comprises 40% of total energy use and has been increasing demand for energy continuously during the last three decades in China [1]. It is an important topic to decrease the energy consumption. Building energy consumption monitoring platform (BECMP) is an innovative energy-saving technology [2–7]. It collects and analyzes energy consumption data in order to increase the energy utilization efficiency.

With the operation of data acquisition in the BECMP, a lot of same or similar data are accumulated. How to select samples from a lot of data and how many samples shall be taken are the problems to be solved immediately. The capacity of the samples shall be grasped. If the capacity $\boldsymbol{n}$ is too small, the problem estimation is not accurate and the problem checking will be not reliable; if the capacity is larger, it will cause the waster of manpower and material resources. Many scholars have investigated China's building energy consumption [8–30], and they analyzed the data basically by using the method of statistics. As to the sampling, the assumption of selection is subject to the random distribution. This method is not scientific. There is a great difference in the number of samples, from several, dozens to hundreds, which shows that scholars have a quite diverse view on a reasonable number of samples to be taken for residential energy consumption survey. Moreover, the data analyzed are fixed; corresponding to the different areas, seasons, building types, operation conditions of energy consumption and statistical items of energy consumption (dynamic data), the method used by them is basically not applicable anymore.

In this paper, sampling for building energy consumption with fuzzy methods are proposed for the first time. A modified clustering algorithm is introduced, which can classify the energy consumption data. After that, it can determine sample size in every subclass by using the statistical theory. Finally, samples in each classification will be merged to obtain a new sample set. This algorithm is simple and effective. The simulation results show that this method solved the sampling problem for building energy consumption.

The rest of the paper is organized as follows. In Section 1, an improved clustering algorithm is presented to amend the weakness of fuzzy C-means clustering (FCM) algorithm. In Section 2, sampling theory is described to determine the sampling capacity. In Section 3, two applications of the fuzzy method are described. In Section 4, the paper is concluded.

2. Modified fuzzy clustering algorithm

Generally, the input data is the vector with diverse dimensionality, and the data in each dimensionality will make different contribution to influence the result. For example, the temperature of an air-conditioned room is mainly influenced by the outdoor temperature, the temperature of cool air input in the room and a load of persons. The changes of input data will make diverse effects on the final temperature, and every dimensionality of input vector has dissimilar weights.

* Corresponding author.

http://dx.doi.org/10.1016/j.enbuild.2017.09.047
0378-7788/

2.1. Information entropy

Entropy was introduced into the information theory by C.E. Shannon and has been widely used in engineering technology at present [32,33]. The basic idea of entropy weight method is to determine the objective weight according to the variance degree of an index. Generally, the smaller the entropy E_j in the index is, the greater the variance degree of the index is and the more the information provided is, thus the greater the effect on comprehensive assessment is and the higher the weight is. On the contrary, the bigger the entropy E_j in an index is, the smaller the variance degree of the index is and the fewer the information provided is, thus the smaller the effect on comprehensive assessment is and the lower the weight is.

2.2. Method of entropy weight

2.2.1. Data standardization

The standardized processing is conducted for the data of each index. Assume that $\boldsymbol{k}$ indexes $X_1, X_2, ..., X_k$ are given, where $X_i = \{x_1, x_2, ..., x_n\}$. If the values after the standardized processing of each index are $Y_1, Y_2, ..., Y_k$, then $Y_{ij} = \frac{x_{ij}-\min(X_i)}{\max(X_i)-\min(X_i)}$.

2.2.2. Obtain the information entropy of each index [34,35]

According to the definition of entropy, the entropy of one set of data is $E_j = -\ln(n)^{-1}\sum_{i=1}^{n} p_{ij}\ln p_{ij}$, where $p_{ij} = Y_{ij}/\sum_{i=1}^{n} Y_{ij}$.

2.2.3. Determine the weight of each index

According to the computational formula of entropy, the entropy of each index is figured out as $E_1, E_2, ...E_k$. Compute the weight of each index by information entropy:

$$w_i = \frac{1-E_i}{k-\sum E_i}\ (i = 1, 2, \ldots, k) \tag{1}$$

Assume that n samples in the domain U are taken as one subset X

$$X = \{x_1, x_2, \ldots, x_n\}$$

where

$$x_n = (x_{n1}, x_{n2}, \ldots, x_{nm}) \tag{2}$$

then

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix}_{n\times m} \tag{3}$$

w_j is the weight of the factor of the attribute in j^{th} column. $(j = 1, 2, \ldots m)$

2.3. Fuzzy C-means clustering algorithm

FCM is the solving of the extreme value of following objective function [36]:

$$J_m = \sum_{i=1}^{N}\sum_{j=1}^{K} u_{ij}^m \|x_i - c_j\|^2 \tag{4}$$

Where $x_i\ (i = 1, \ldots, N)$ is clustering input data set, $c_j\ (j = 1, \ldots, K)$ is the center of every clustering, u_{ij} is a membership function and m is the weighting index describing the classification. The membership function meets

$$\sum_{j=1}^{K} u_{ij} = 1, 0 \le u_{ij} \le 1, \forall i, j \tag{5}$$

Update the membership function and cluster center in every iteration

$$u_{ij} = \frac{\|x_i - c_j\|^{\frac{-2}{m-1}}}{\sum_{k=1}^{K}\|x_i - c_k\|^{\frac{-2}{m-1}}} \tag{6}$$

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^m x_i}{\sum_{i=1}^{N} u_{ij}^m} \tag{7}$$

After the initialization of the cluster center, the membership function and center will be updated in every iteration until the change of objective function is lower than a preset threshold.

2.4. Modified FCM based on information entropy

A clustering algorithm on the base of information entropy is put forward to solve the sensitive problem of FCM on the initial number of clusters in Literature [31]. The main idea is to determine the number of clusters by combining the clustering and taking the degree of membership as the basis of entropy computation.

u_{ij} is fuzzy membership function that represented the degree of a random sample i belonging to j^{th} cluster. u_{ij} is a matrix, and the weight w_j of data in every column can be obtained in accordance with the method of entropy weight. On the basis of Literature [31], the paper adds the computation weight w_j to the column vector of fuzzy membership matrix u_{ij} to further optimize the clustering algorithm. The specific method is as follows:

1) The number range of cluster $[C_{\min}, C_{\max}]$ shall be determined by user.

2) During the process of the number of clusters increasing from $C_{\min}$ to $C_{\max}$, one membership matrix ($k \in [C_{\min}, C_{\max}]$) will be generated for one corresponding number of clusters k. Each membership matrix has corresponding entropy $H_k(x)$. If the row value of such membership matrix is set as the serial number of data point i, and the column value is set as the class number of cluster j, then $H_k(x) = \sum_{i=1}^{n} H_{ki}(x)$, where $H_{ki}(x)$ is the entropy of every data point in cluster affiliation, $H_{ki}(x) = -\sum_{j=1}^{k} w_j u_{ij} \log_2 u_{ij}$.

3) The construction of the membership matrix shall be iterated constantly. The iteration will adjust the classification of all data points simultaneously and modify the cluster center and enter the next iteration. When the data point affiliation caused by two adjacent iterations does not change anymore, $H_k(x)$ shall be calculated.

4) When k is increased from $C_{\min}$ to $C_{\max}$, the number of $H_k(x)$ generated is $C_{\max} - C_{\min}$. The number of clusters k corresponding to the minimum $H_k(x)$ is selected as the final number of clusters C.

5) Finally, the clustering result is obtained by FCM.