Predict Commercial Building Type By Clustering Energy Consumption

Hongyi Tang
Civil and Environmental Engineering
Carnegie Mellon University
thongyi@andrew.cmu.edu

Weijian Li
Civil and Environmental Engineering
Carnegie Mellon University
weijian1@andrew.cmu.edu

ABSTRACT

Understanding the energy consumption pattern of commercial buildings can help a lot for energy saving. However, As the types of commercial building are variable, it is difficult to find a certain energy consumption pattern for every commercial building. But for a certain kind of commercial building, the energy consumption pattern can be stressed out.

In this paper, we demonstrated cluster analysis on the dataset from US Energy Information Administration[1] on commercial building energy consumption to predict the building type of a certain building. The typical energy consumption patterns of all the selected building types are distinguished out by cluster analysis. After comparing the cluster assignment result with the building type data in the dataset, we found that our model can predict 86.3% of the building type among 2 building types, 59.7% among 3 building types, 49.5% among 4 building types and 37.5% among 5 building types.

CCS Concepts

• **Information systems→ Database management system engines**

Keywords

Commercial Building, Energy Consumption, Cluster Analysis.

1. INTRODUCTION

Commercial Buildings now consume 60%[2] of the total electricity in US. So, the energy used in commercial building worth a lot of research and analysis in terms of energy saving. Once the energy consumption pattern is found, it is much easier to design the energy saving police of a certain building. Commercial building usages varies a lot such as office, education, warehouse, mall and so on. Distinguishing a certain building type among all those commercial buildings is difficult but helpful for energy saving policy designing. To do that, the unique energy consumption characteristics should be located. Cluster analysis is the task of grouping a set of data which is exactly what we want to achieve.

In reference [3], cluster analysis is used to find out the energy demand distribution of buildings during a whole day. And 8 patterns of energy demand are presented in their work. Similar to what they did, we used different energy consumptions rather than hour of a day to conduct cluster analysis and to find the energy consumption patterns. Different from what they got, our energy consumption pattern is energy structure rather than energy consumption in different time period in one day.

Once we find out the way to predict the building type, we can design the energy saving plan for this building by its energy consumption pattern rather than its nominal building usage. And we can find the main energy consumption part of the building. After that, energy saving policy can be designed to mainly focus on this energy consumption.

2. Proposed Approach

By cluster analysis, the selected building would be divided into several clusters by their energy consumption patterns. After that, we can compare the cluster result with the building type data in the dataset to evaluate the efficiency of the cluster system. If the system can predict or distinguish correctly for 80% the building type, the predicting method is thought to work.

After that, more cluster analysis on more building types and more energy consumption types can be demonstrated to test the efficiency of the cluster analysis method.

3. Dataset

In this paper, commercial building energy consumption survey (CBECS) dataset[1] is used for cluster analysis.

In the CNECS dataset, over 6000 buildings' energy consumption data is recorded. And there are 35 kinds of energy consumption for each building. Among all these attributes in the dataset, we selected annual electricity consumption (ELBTU), annual natural gas consumption (NGBTU), electricity ventilation consumption (ELVNBTU) and natural gas heating consumption (NGHTBTU) to conduct the cluster analysis. And due to the different sample number for different building type, 2 types of building with relatively more sample number and thought to have different energy consumption pattern were selected which are office and inpatient health care. More building types would be added in later in discussion section.

Before the cluster analysis, all the energy consumption data need to be normalized by divided by area in square foot of the building. After that, the box graph of all the data would be plotted out to show the visually consumption pattern of different building type. For example, inpatient health care may have more natural gas heating consumption for the night consumption. And then, cluster analysis can be demonstrated by using these four energy consumption types to distinguish the building type. At last, the building type information of the dataset would be compared with the cluster assignment result to evaluate the efficiency of the method.

4. Result

4.1 Selected Data Size

Before plotting the box graph, the data was pretreated. After dividing all the data by the area of the building, any the data with a zero number in any selected attribute is wiped out. Because any

zero number in the energy consumption would influence the energy consumption pattern dramatically. And the numbers of data sample left are presented in table 1.

Table 1 Number of Samples for Building Type

Type	Office	Inpatient Health Care
Number	714	283

4.2 Visually Energy Consumption Pattern

After pretreating, the box graph for different building type is plotted out in figure 1 to show the visually consumption pattern for every building type.

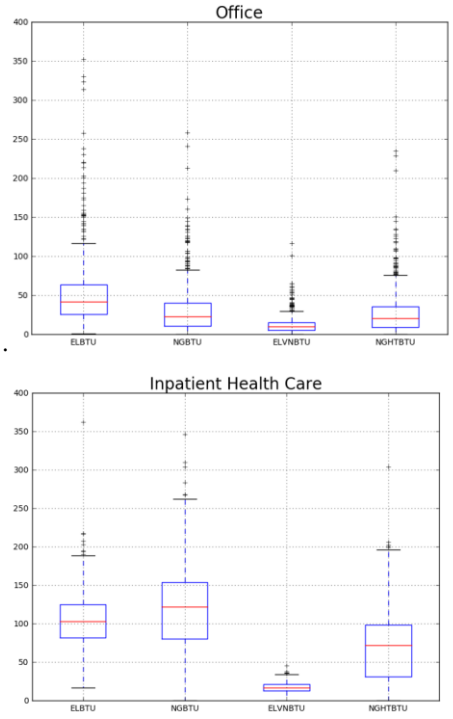

Figure 1. Box Graph of Energy Consumption

From figure 1, we found that the biggest difference for energy consumption pattern between office and inpatient health care is that for inpatient health care, annual natural gas consumption (NGBTU) is more than annual electricity energy consumption (ELBTU) while it is the other way for office building. This is mainly because in winter night, inpatient health care still need natural gas heating while office is basically vacant at night.

4.3 Cluster Analysis

Using kmean function in sklearn of python, setting the cluster number as the building type number which is 2. The cluster result is presented in figure 2.

In figure 2, all the data in the same cluster is plotted in the same figure. The dark line in the middle refers to the centroid which can be understood as the core line of the cluster. The trend of the centroid line shows the energy consumption pattern of buildings in that cluster.

Similar to what in the box graph, the main difference between the two centroid is that for the second one, annual natural gas consumption is higher than the natural electricity consumption while it is the other way for the first one. So, when there are only two building types, the cluster analysis can point out the difference in terms of energy consumption pattern perfectly.

4.4 System Efficiency Analysis

Although cluster analysis can point out the pattern difference, it is not sure that all the data samples are assigned to the correct subset. So, the cluster assignment and the original building type information was compared to see whether cluster analysis assign all the data samples correctly. And 860 samples were assigned correctly to the same subset which means that cluster analysis assigned 86.3% of all the samples (997 samples) of office and inpatient health care correctly.

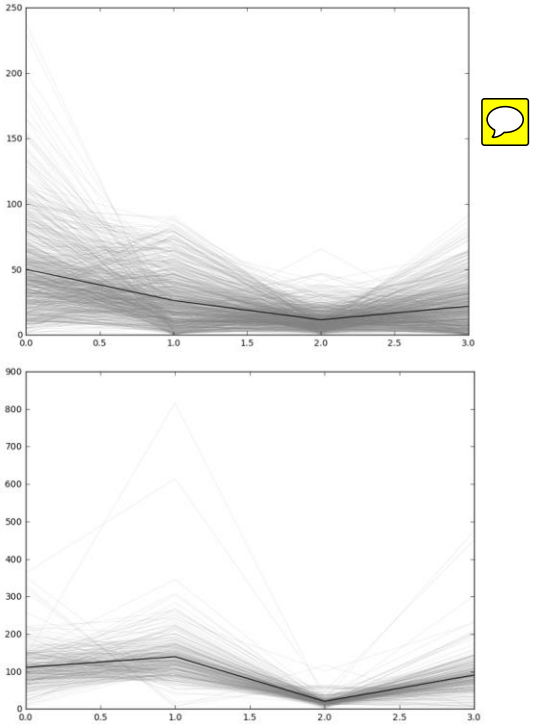

Figure 2. Cluster Result

5. Validation

As talked above, about 86.3% samples are assigned correctly which is a good result. But considering that there are only two building type and these two building types are carefully picked so that the difference between them are obvious and significant, the efficiency of 86.3% cannot be a perfect answer.

So, in the discussion part, more building types are going to be picked out to see whether this method works for more building types.

6. Discussion

To test the efficiency of the cluster predicting system, more trails were demonstrated.

6.1 More Building Type

First, more building types were selected in such as service, public assembly and education buildings. And the numbers of all the data sample are presented in table 2. Cluster analysis with building type number of 3, 4 and 5 are all tested, but we only list the result of 5 in detail here.

Table 2 Number of Samples for Building Type

Type	Office	Inpatient Health Care	Service	Public Assembly	Education
Number	714	283	244	226	471

6.2 Visually Energy Consumption Type

The box graph of added building type is plotted in figure 3. Now, the energy consumption pattern is hard to tell. Some types may share the same pattern but have different value scale on certain energy consumption. For example, service and education buildings are really similar to each other.

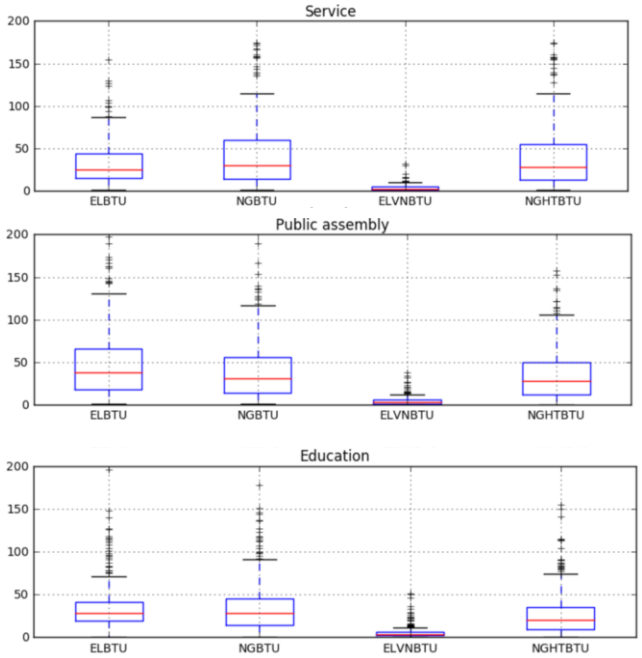

Figure 3. Box Graph for Added Building Type

6.3 Cluster Analysis Result

Cluster analysis is shown in figure 4. This figure is the result of several re-clustering. That means after the first time of cluster analysis, the data was divided into 5 clusters but there was a cluster which only contained one digit number of samples. So, those samples were treated as weird data sample and were deleted before the second cluster analysis was demonstrated. That process was like a loop. After several times cluster analysis, the number of the samples in each is quite reasonable. After this loop like cluster analysis, the cluster assignment is fit to the building type by comparing with the boxplot result. The final cluster analysis is shown in table 3.

It seems like that cluster analysis find out the energy consumption pattern to distinguish all the building type between each other. However, the assignment result is not so good. Only 727 data samples were assigned to the right subset, which means that cluster analysis only assign about 38% of the total data sample correctly.

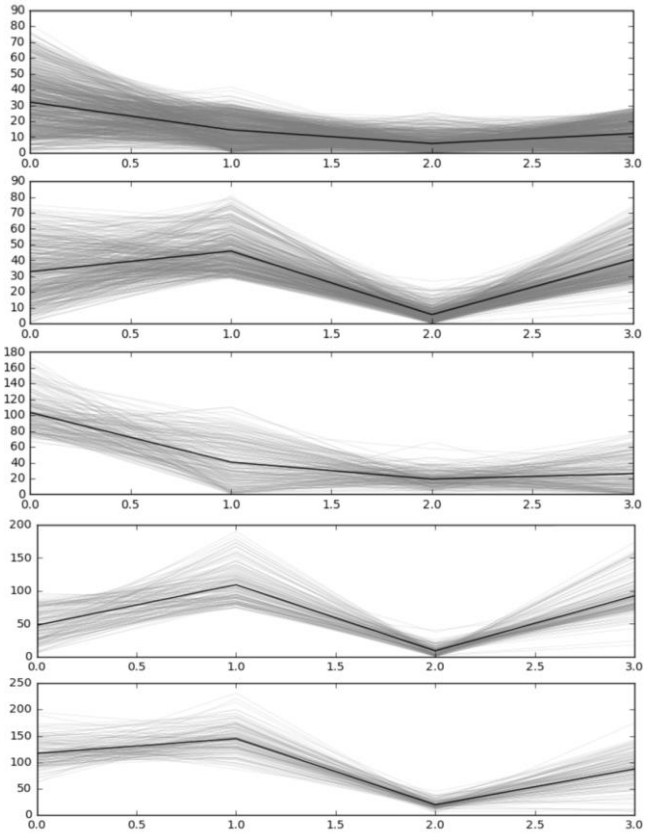

Figure 4. Cluster Analysis Result

Table 3 Number of Samples in Each Cluster

Type	Office	Inpatient Health Care	Service	Public Assembly	Education
Number	854	140	165	213	477

6.4 More Trail

The result of cluster analysis for 5 building types using four energy consumption type is bad. And the similar cluster analysis was demonstrated for 3 and 4 building types, too. The result is presented in table 4. The more the building type, the lower the correct percentage is.

Table 4 Correct Percentage for Cluster Analysis

Type Number	2	3	4	5
Correct Percentage	86.3%	59.7%	49.5%	37.5%

7. Conclusion and Future Work

In this paper, we presented our model for predicting commercial building type by clustering energy consumption. Our evaluation showed that this model works well when there are only two building types, but not ideal with the increase of building type. To be more specific, about 86% samples are assigned correctly in two building types model, however, only 38% in five building types model, which is hard to accept. So, from the tested above, we made our conclusion that this system can only predict building pattern when there are only two building types. The following aspects of our test can be improved in future work.

Although it looks like our system works well in two building type, some tips are worth noting. First, building types for which we used in our system are carefully selected, in the absence of data that more pairs of building types are analyzed, we cannot say that our system is good for predicting two building type pattern. Second, energy types we used in our system are not enough to reflect the features of building type. For example, office is seldom used for cooking, so energy for cooking should be added to our data to do analyze, which is good for differentiate office from other building type. As we mentioned before. There are 35 kinds of energy consumption for each building, but we only chose 4 kinds, which are not enough to get a reasonable result. So, in future, we will add more energy types to test our system.

8. REFERENCES

[1] "Energy Information Administration (EIA)- Commercial Buildings Energy Consumption Survey (CBECS) Data". *Eia.gov*. N.p., 2016. Web. 9 Dec. 2016.

[2] "Energy Information Administration (EIA)- Commercial Buildings nergy Consumption Survey (CBECS)". *Eia.gov*. N.p., 2016. Web. 9 Dec. 2016.

[3] "Analyzing Energy Usage on a City-scale using Utility Smart Meters",*Srinivasan Iyengar, Stephen Lee, David Irwin, and Prashant Shenoy* ,University of Massachusetts Amherst