Ranking of energy consumption objects using the principal components method

A. Perekrest  
Kremenchuk Mykhailo Ostrohradskyi National University,  
Ukraine  
E-mail: pksg13@gmail.com

O. Vovna  
Donetsk National Technical University,  
Ukraine  
E-mail: Oleksandr.Vovna@domtu.edu.ua

V. Ogar  
Kremenchuk Mykhailo Ostrohradskyi National University,  
Ukraine  
E-mail: Vita.Ogar@gmail.com

M. Kushch-Zhyrko  
Kremenchuk Mykhailo Ostrohradskyi National University,  
Ukraine  
E-mail: k.zh.mhl@gmail.com

I. INTRODUCTION

The development of information technology has led to an increase in the number of machine learning methods application areas and methodologies of its application [1-4]. These include managing the energy consumption of civil buildings and forecasting energy costs to ensure comfortable conditions. Machine learning methods help in the energy management systems parameters analyzing and support effective decision-making by energy managers [5].

There is a need to improve existing or find new approaches of energy consumption data analysis, in order to make decisions aimed at improving energy efficiency.

Implementation of solutions for heating systems automated monitoring and control allows to reduce the total heat consumption and accumulate large amounts of information about the operation of the system and the decisions made by the energy manager. Tasks related to the need to ensure regulatory sanitary and hygienic conditions in heated rooms, taking into account the influence of external factors [6, 7].

II. EXISTING SOLUTIONS ANALYSIS

The implementation of solutions aimed at controlling and monitoring energy consumption allows to accumulate significant amounts of information about the amount of resource consumed [4, 5]. The availability of meter readings gives an idea about heat consumption, which can affect management strategies, such as the allocation of non-obvious patterns of consumption of communal or any buildings - information that can not be obtained without the use of machine learning methods. The main problem you may face is the complexity, and often the inability, to obtain reliable and detailed data.

In previous works [8, 9] the structure of information flows for obtaining data from available sources was developed. Features of data depending on type and
III. PURPOSE OF THE WORK

Designing new features to compare the energy consumption of buildings leads to an increase in the dimensionality of the data and, consequently, an increase in the time to perform various actions on the data, which can be critical for larger samples over long periods of time. Thus, the aim of the study is to develop and test a model of information ranking of buildings according to their energy consumption using the method of principal components.

IV. RESEARCH MATERIALS

Three data sets were collected for this study:

– Energy consumption of educational buildings for the period from 2012 to 2016, indicating the building number, month, year, heat consumption (expressed in Gcal), electricity consumption (expressed in kWh). Number of data: columns - 5, rows - 420. Data format: case number - categorical variable, the rest of the data - numerical.

– Volumes of heat load, indicating the building number, heat load volumes (expressed in Gcal / h) and building volume. Number of data: columns - 3, rows - 7. Data format: case number – categorical variable, the rest of the data - numerical.

– Average monthly ambient temperature for the period from 2012 to 2016, indicating the year, month and average monthly temperature. Number of data: columns – 3, rows – 60. Data format: all data – numerical.

For comparative analysis of energy consumption of different buildings, the same indicators were used, for example, specific heat consumption (q1, kWh/m²), specific electricity consumption (q2, kWh/m²), etc.

Normative value of degree days of the heating period (GDOP):

$$Dd_H = (T_{in} - T_{out, norm}) \cdot Z_{norm},$$  (1)

where $Dd_H$ – normalized number of degree days; $T_{in} = 20\degree C$ – internal room temperature; $T_{out, norm} = -0.8\degree C$ – average normalized outdoor temperature; $Z_{norm} = 180$ – standard duration of the heating period.

The actual value of the degree-days of the heating period (HDD):

$$Dd = (T_{in, C} - T_{out, C}) \cdot Z,$$  (2)

where $Dd$ – the actual number of degree days; $T_{in, C}$, $T_{out, C}$ – internal room temperature; $z C$ – average outdoor temperature; $Z$ – the actual duration of the heating period.

HDD coefficient:

$$K_{Dd} = \frac{Dd_n}{Dd},$$  (3)

where $K_{Dd}$ – HDD coefficient; $Dd$ – the actual number of degree-days; $Dd_n$ – normalized number of degree-days.

Absolute heat consumption in kWh:

$$E_kWh = 1163E_Gkal,$$  (4)

where $E_kWh$ – absolute heat consumption, in kWh; $E_Gkal$ – thermal energy consumption, Gcal.

Specific heat consumption:

$$q1 = \frac{E_kWh}{V},$$  (5)

where $q1$ – specific heat consumption, kWh/m²; $E_kWh$ – thermal energy consumption, kWh; $V$ – volume of the building, m³.

Specific power consumption:

$$q2 = \frac{W_kWh}{V},$$  (6)

where $q2$ – specific electricity consumption, kWh/m²; $W_kWh$ – electricity consumption, kWh; $V$ – volume of the building, m³.

Total energy consumption:

$$E_{sum, kWh} = E_kWh + W_kWh,$$  (7)

where $E_{sum, kWh}$ – total energy consumption, kWh; $E_kWh$ – thermal energy consumption, kWh; $W_kWh$ – electricity consumption, kWh.

Specific total energy consumption:

$$q3 = \frac{E_{sum, kWh}}{V},$$  (8)

where $q3$ – specific total energy consumption, kWh/m³; $E_{sum, kWh}$ – specific energy consumption, kWh; $V$ – volume of the building, m³.

Specific heat consumption, reduced to the normative values of external and internal air temperature:

$$q1t = q1 \cdot K_{Dd},$$  (9)

$$DOI: 10.33042/2079-424X-2020-1-57-39-44$$
The purpose of PCA is to find linear combinations of variables containing the largest variance [10]. The linear combination has the following form:

\[ Y_1 = a_11X_1 + a_12X_2 + \cdots + a_{1p}X_p \]
\[ Y_2 = a_21X_1 + a_22X_2 + \cdots + a_{2p}X_p \]
\[ \vdots \]
\[ Y_k = a_k1X_1 + a_k2X_2 + \cdots + a_{kp}X_p \]

Wherein \( \sum_{i=1}^{p} a_{ii} = 1 \).

Dispersion of the first head component:

\[ \text{var}(Y_1) = a_1^2 \sum a_1. \]

where \( \Sigma \) – covariance (correlation) matrix.

The variance of the second and subsequent principal components is calculated similarly.

In this model, the vectors \( a'_i = (a_{i1}, a_{i2}, \ldots, a_{ip})' \), \( i = 1, p \) represent the eigenvectors of the covariance (correlation) matrix \( \Sigma \), whereas the variance of the \( i \)-th principal component is equal to the eigenvalue of the covariance matrix (correlation):

\[ \text{var}(Y_i) = \lambda_i. \]

The total variance of the sample is equal to:

\[ \sum_{i=1}^{p} \lambda_i. \]

Performing the PCA procedure allows to get the eigenvalues \( \lambda_i \) of the main components and the percentage of variance that they explain (Table I).

![Fig. 1. Correlation coefficients matrix](image)

| Table I. MAIN COMPONENTS VALUES |
|----------------------------------|
| comp | eigenvalue | percentage of variance | cumulative percentage of variance |
|------|------------|------------------------|-----------------------------------|
| comp 1 | 7.7733498 | 48.583463 | 48.58344 |
| comp 2 | 2.7178720 | 16.886699 | 65.57014 |
| comp 3 | 2.3300882 | 14.563051 | 80.13319 |
| comp 4 | 1.4952653 | 9.345408 | 89.47860 |
| comp 5 | 0.7287498 | 4.554685 | 94.03328 |
| comp 6 | 0.3633059 | 2.270815 | 96.30410 |
| comp 7 | 0.2747271 | 1.717044 | 98.02114 |
| comp 8 | 0.1548087 | 0.967554 | 98.98870 |
| comp 9 | 0.1130121 | 0.706335 | 99.69502 |
| comp 10 | 0.0288535 | 0.180346 | 99.87536 |
| comp 11 | 0.0190971 | 0.119356 | 99.99471 |
| comp 12 | 0.0008458 | 0.005286 | 100.00000 |

Obtained 4 main components, explaining 89.4786% of the dispersion.

Fig. 2 shows the variance of the weights of the main components.
The dotted line in fig. 2 – 6 corresponds to the expected value if the contribution of all variables is uniform. For the selected component, the contribution of any of the variables above the dashed line can be considered important.

As can be seen from fig. 3, the first main component is "loaded" by variables related to temperature and variables that characterize the absolute, specific and reduced to the normative values of external and internal air temperature heat consumption and total absolute energy consumption. Moreover, the analysis of factor loads shows that, with the internal and external environment temperature increase, the amount of heat consumption decreases respectively. This component should be called "Energy efficiency of the building by temperature", i.e. those structures are more energy efficient, which when varying the temperature have less variation in total energy consumption and heat consumption.

The second main component (Fig. 4) is related to the volume of the building, the amount of heat load, as well as the indicators of specific, reduced to the normative values of outdoor and indoor air temperature, electricity consumption and total energy consumption, specific electricity consumption. It can be called "Energy efficiency of a building by volume", i.e. those structures are more energy efficient that with the same volume and volume of heat load have less variation in this direction.

The third main component (Fig. 5) is formed by the values of the coefficient of degrees-days, the volume of the building, the volume of heat load, as well as specific indicators, reduced to the normative values of outdoor and indoor air temperature and total energy consumption. Since the second and third components contain the same variables that significantly affect the percentage of the described variance, we can say that this phenomenon is caused by an increase in the dimensionality of the data by creating parameters that depend on the volume and value of the degree-day ratio.

The fourth main component (Fig. 6) is specific and is formed by the values of absolute and specific power consumption.

Below (Fig. 7) a graph of the weights of the main components is represented.
Data were segmented by buildings (Fig. 8 – 10) by two seasons: non-heating (conditional name summer) and heating (conditional name winter) periods.

The convergence of the winter period segments centers to the origin indicates an increase in electricity consumption during the heating period and transition seasons, which may be caused by the active use of climate technology.

V. CONCLUSIONS

Methods of correlation analysis and PCA were used to reduce the dimensionality of informative features. Four main components describing about 90% of the sample variance were obtained.

The first main component is the variables related to temperature and the variables that characterize the absolute, specific and reduced to the normative values of outdoor and indoor air temperature heat consumption and total absolute energy consumption - “Energy efficiency of the building by temperature”.

The second main component is the volume of the building, the amount of heat load, the specific, reduced to the normative values of outdoor and indoor air temperature, electricity consumption and total energy consumption, the specific electricity consumption - “Energy efficiency of the building by volume”.

The third main component is the degree-day ratio, the volume of the building, the amount of heat load, the specifics, reduced to the normative values of outdoor and indoor air temperature of electricity consumption and total energy consumption.

We can conclude that the hit of the same variables to the second and third main components is due to the artificial creation of parameters from the original, which in turn caused an increase in the dimensionality of the data. To exclude this anomaly, at the stage of studying the correlation matrix it is necessary to exclude variables that are strongly correlated with each other. The fourth main component depends solely on the values of absolute and specific power consumption.

The application of cluster analysis allowed to identify homogeneous groups of energy-consuming objects by selected components.

REFERENCES

[1] C. M. Bishop Pattern Recognition and Machine Learning. NY: Springer. – 2006.
Рейтинг об’єктів енергоспоживання на основі методу основних компонентів

А. Л. Перекрест
КНУ ім. Михайла Остроградського, Україна

В. О. Огарь
КНУ ім. Михайла Остроградського, Україна

О. В. Вовна
ДНТУ, Україна

М. О. Кущ-Жирко
КНУ ім. Михайла Остроградського, Україна

Забезпечення комфортних умов в цивільних будівлях потребує виконання завдань моніторингу та прогнозування витрат енергетичних ресурсів, а також енергоефективного керування тепловими інженерними системами та їх обладнанням. Впровадження відповідних рішень з автоматизації та моніторингу дозволяє накопичувати значний обсяг даних. Для підвищення інформативності аналізу ефективності використання енергоресурсів при функціонуванні цивільних будівель розроблено модель їх інформаційного ранжування з використанням кореляційного аналізу та методу головних компонент.

Для прийняття початкових умов про електропожежнання та теплоспоживання навчальних корпусів університету на основі міждисциплінарної методології (CRISP-DM) визначені базові показники та отримана оцінка матриці коефіцієнтів кореляції їх взаємозв’язку. Певні дані (зовнішній об’єм та площа будівлі та середні значення температури для даного регіону за нормою) отримуються з технічної документації будівлі та доступні з відкритих джерел, інші (кількість спожитої теплової та електричної енергії, внутрішні температура у приміщенні) визначаються в процесі експлуатації і характеризують ефективність використання енергетичних ресурсів в будівлі. На початковому етапі проведено кореляційний аналіз взаємозв’язків основних параметрів, що характеризують будівлі та споживання ними енергетичних ресурсів. Для зниження розмірності ознакової множини даних та вивчення гомогенних груп об’єктів енергоспоживання використано метод головних компонент. Отримані чотири компоненти пояснюють біля 90 % дисперсії початкових даних та характеризують ефективність використання енергоресурсів по температурі, об’єму та коефіцієнту градусо-діб опалювального сезону. Отримані результати рекомендовано до впровадження в сучасних системах енергетичного моніторингу та муніципального енергетичного менеджменту в якості прикладних моделей для діагностування нештатних ситуацій та обґрунтованого прийняття управлінських рішень.

Ключові слова – будівлі; енергоспоживання; головні компоненти; машинне навчання; сегментація даних.