Analysis of an energy efficient building design through data mining approach

Hyunjoo Kim a,⁎, Annette Stumpf b,1, Wooyoung Kim c,2
a Department of Civil and Environmental Engineering, California State University, United States
b U.S. Army Corps of Engineers, Engineering Research and Development Center, United States
c Construction & Economy Research Institute of Korea, South Korea

ARTICLE INFO

Article history:
Accepted 3 May 2010

Keywords:
Energy simulation
Data mining
Classification
Factor selection
Energy design process

ABSTRACT

Incorporating energy efficiency and sustainable green design features into new/existing buildings has become a top priority in recent years for building owners, designers, contractors, and facility managers. This paper intends to address why delivery of an energy efficient building is not just the result of applying one or more isolated technologies. Rather, it can best be obtained using an integrated whole building process throughout the entire project development process, which leads building designers to generate a large amount of data during energy simulations. The authors observed that even a simple energy modeling run generated pages of data with many different variables. The volumes of energy modeling data clearly overwhelm traditional data analysis methods such as spreadsheets and ad-hoc queries with so many factors to be considered. An integrated or whole building design process involves studies of the energy-related impacts and interactions of all building components, including the building location, envelope (walls, windows, doors, and roof), heating, ventilation and air conditioning (HVAC) system, lighting, controls, and equipment, which shows why it is so difficult to find the correlation between different systems. The objective of this research is to develop an energy efficient building design process using data mining technology which can help project teams discover important patterns to improve the building design. This paper utilizes the data mining technology to extract interrelationships and patterns of interest from a large dataset. Case study revealed that data mining based energy modeling help project teams discover useful patterns to improve the energy efficiency of building design during the design phase. The method developed during this research could be used to guide designers and engineers through the process of completing an early design energy analysis based on energy simulation models.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

Building energy simulation programs are in use throughout the building energy community. Energy modeling programs provide users with key building performance indicators such as thermal loads, energy use and demand, temperature, humidity, and costs. The A/E/C industry is embracing energy simulation programs, so building designers are currently dealing with a large amount of data generated during energy simulations. From our experience, even a simple energy modeling run generated pages of data with many different variables. Examples of those variables include but are not limited to the estimated energy costs or savings in terms of building orientation, HVAC system, lighting efficiency and control, construction of roof and walls, glazing type, water usage, day-lighting, etc. Such volumes of data clearly overwhelm traditional data analysis methods such as spreadsheets and ad-hoc queries with so many factors to be considered. It is difficult to find the best correlation/combination of different energy systems during the building design process.

The objective of this research is to develop a process which can help project teams discover useful patterns to improve energy efficient building design. This paper utilized data mining technology, which is a data analysis process that combines different techniques from machine learning, pattern recognition, statistics, and visualization, to automatically extract concepts, interrelationships and patterns of interest from a large dataset. By applying data mining technology to the analysis of energy efficient building designs one can identify valid, useful, and previously unknown patterns out of energy simulation modeling.

This paper presents the necessary steps to develop the data mining approach such as 1) requirement identification, 2) energy simulation, 3) data mining, and 4) refinement. In order to establish a process, a case study was conducted with an on-going design project. Then detailed steps showing how energy analysis tools were used early in the design process are presented.
2. Literature review

2.1. Energy modeling

For the past 50 years, a wide variety of building energy simulation (BES) analysis tools have been developed, enhanced, and applied throughout the building energy community. Examples of these tools are BLAST, EnergyPlus, eQUEST, TRACE, DOE2, and ECOTECT. Building energy simulation tools are complex applications which require a great deal of time to learn [1,2]. Additionally, energy efficient building design is not just the result of applying one or more isolated design guidelines. Rather, it requires an integrated whole building process throughout the entire project development process. Whole building design considers the energy-related impacts and interactions of all buildings components, including the building location, envelope (walls, windows, doors, and roof), its heating, ventilation, and air conditioning (HVAC) systems, lighting, controls, and equipment.

Several research papers describe energy analysis as a holistic evaluation [3]. Dahl et al. [4] and Lam et al. [5] showed that decisions made early in a project have a strong affect on the life cycle costs of a building.

In this research, energy-efficient concepts, technologies and building elements will be developed and evaluated using energy modeling tools. Building designers could increase energy efficiencies in their building designs by using energy simulation tools in the context of a whole building approach.

2.2. Data mining implementation for pattern discovery

Recognizing the complexity of the search algorithms and the size of the data being analyzed when identifying useful patterns in energy modeling data, this research utilized the data mining technology, which can be considered an interdisciplinary field involving concepts from machine learning, statistics, mathematics, high-performance computing, and visualization. Fayyad et al. [6] define data mining as the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data. John [7] defines it as the process of discovering advantageous patterns in data.

In this research, a pattern is an expression of describing facts in a subset of a set of facts. The expression is called a pattern if it is simpler than the enumeration of all facts in the subset of facts. A useful pattern means that building designers may increase the energy efficiencies of their buildings through the pattern found. One discovered pattern might be “building insulation that has a pattern of decreasing energy use by 50%”. Patterns might also be much more complex, taking into account several different building components such as walls, windows, doors, and roof and specifying the conditional probability of improving energy efficiency. We can estimate the payoff of such a strategy by assuming an effectiveness rate. If the estimated payoff from the strategy was sufficiently high, it would be implemented to increase energy efficiency, and its results could be measured.

3. Case study

In this section, a sequence of three different steps in the data mining process is outlined, as shown in Fig. 1. The first step is to identify the project requirements. This step is challenging since projects have constrained budgets, schedules, and resources. It is essential that all building stakeholders— including owners, designers, engineers and contractors—have a clear understanding of problem definition and participate in identifying a set of design alternatives early in the project planning process. The second step, energy simulation, is where a large amount of data is generated. Examples of those variables include estimated energy costs or savings in terms of building orientation, HVAC systems, lighting efficiency and control, roof and wall construction, glazing type, etc. Such volumes of data clearly overwhelm the traditional data analysis methods such as spreadsheets and ad-hoc queries with many factors to be considered. It is difficult to find the best correlation/combination of different energy systems during the building design process. The third step is the data mining process where we develop an overall data analysis mechanism that can be applied to find patterns that explain or predict any behaviors resulting from the energy simulations.

3.1. Requirement definition

The energy efficient design process begins when the occupants needs are assessed and a project budget is established. Then, the proposed building is located on the site, and programmed spaces are carefully arranged to reduce energy use for heating, cooling, and lighting. Building heating and cooling loads are minimized by optimizing the building form and designing energy efficient building elements—windows, walls, and roofs. Taken together, they form the basis of integrated, whole building design.

3.1.1. Building description

The project described in this paper is a new Community Emergency Service Station (CESS) facility. This building provides fire fighting, medical and police support services for a residential neighborhood. The station consists of offices, training rooms, physical training, day room, kitchen, dormitory area, apparatus room, decontamination room, storage area/rooms, latrines, communication and electrical closets, and a mechanical room. The size of the proposed new facility is approximately 808 m² as shown in Fig. 2. The apparatus room was sized for a fire truck, military police car, and ambulance. Space for hose drying, lockers and a work bench were also required in the apparatus room. The building is occupied seven days a week for 24 h a day. A summary of the building model parameters and the thermal loading are presented in Table 1. Fig. 3 shows the rooms to be located in each CESS facility.

Requirement identification for the CESS facility was accomplished using a four day charrette process. A charrette is held at the beginning of a project where a group of designers, engineers, and contractors may draft solutions to design problems. For our case study, the charrette took place in the early stage of design and also included stakeholders outside of the design/build team. Each participant presented his/her work to the full group. The charrette served as a way of quickly generating solutions while understanding and integrating the interests of different groups of people.

Energy modeling in the charrette presented several challenges. First, it was essential to provide energy analysis results so we could identify energy-saving improvements while the design was being modified. In addition, energy modeling usually involves the time consuming process of re-entering all the building data, geometry and parameters to conduct an energy analysis.

3.1.2. Energy modeling

The energy simulations were completed using Autodesk® Green Building Studio® plus eQUEST 3.63. Fig. 4 shows the graphical image...
of the energy model built in eQUEST. Green Building Studio® was used to build a three-dimensional energy estimation model and eQUEST 3.63 to fine-tune the simulations. Green Building Studio® and eQUEST annual energy analysis results are depicted as cost per year for comparison purposes. Individual energy analysis runs are often performed to identify the annual energy cost for each building feature or system under consideration, or to determine the effectiveness of applying multiple features or systems to a design.

This section describes the modeling assumptions used in the baseline and proposed (energy efficient) models. The baseline building envelope features are modeled using ASHRAE Standard 90.1-2004 steel frame wall construction, roof insulation door and fenestration types. The energy savings for the insulation are small and show an increase with the bigger R-values. The energy savings were significant ranging between 2 and 5% of the total energy consumption. Fig. 5 and Table 2 show electricity consumption of the baseline model in kWh due to the regional situation that the CESS is located where natural gas supply is not available. Based on electricity usage only, the estimated annual energy consumption is 116,270 kWh as shown in Table 2.

Table 1
Baseline building model parameters.

<table>
<thead>
<tr>
<th>Building component</th>
<th>Baseline building model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of users</td>
<td>14 people</td>
</tr>
<tr>
<td>Area</td>
<td>808 m²</td>
</tr>
<tr>
<td>Floors</td>
<td>1</td>
</tr>
<tr>
<td>Fenestration type</td>
<td>Standard 90-1-2004</td>
</tr>
<tr>
<td>Wall construction</td>
<td>Steel frame</td>
</tr>
<tr>
<td>Wall insulation</td>
<td>Standard 90.1-2004 steel frame</td>
</tr>
<tr>
<td>Roof insulation</td>
<td>Metal frame with code compliant insulation</td>
</tr>
<tr>
<td>Temp. set points</td>
<td>70 ° F heating; 75 ° F cooling—set back/up to 55 ° F heating; 91 ° F cooling</td>
</tr>
</tbody>
</table>

Fig. 2. Proposed building floor plan.

Fig. 3. Thermal zoning for the community services station energy model.
3.1.3. Data analysis

Given the simulation data obtained in the previous section, our data analysis goal is to develop an overall data analysis process that can be applied to find patterns for energy-related behaviors of a construction project. During data analysis, the feature subset selection was first used to calculate the relevance of features. Then, useful patterns were extracted from the data sets. When comparing different building components and equipment, all the energy costs were estimated and presented in dollars at the rate of $0.07 per kWh (which is a local rate in the region) throughout the data analysis.

3.1.3.1. Feature subset selection. The technique of feature subset selection was used to find which building elements (windows, walls, doors, roof and HVAC systems) are most likely to improve energy efficiency significantly.

The feature subset selection algorithm conducts a search for a good subset using the induction algorithm as part of the evaluation function. The accuracy of the induced classifiers is estimated using accuracy estimation techniques. There are several induction algorithms (Quinlan [8], and Domingos and Pazzani [9]). This research conducted the feature subset selection using C4.5 decision tree which is a well-known algorithm in the machine learning community.

3.1.3.2. Roof construction. Roof construction includes structural frames, materials (wood, steel, polystyrene, screed, concrete, plaster) and various insulations. The purpose of this section is to figure out which combination of roof construction materials and method would produce the lowest annual energy cost using computer-generated simulations. A total of 127 energy simulation runs were conducted to study roof construction alternatives and estimate the total annual energy cost as shown in Fig. 6 (Max. $13,133, Min. $7,134, AVG $7,934, Median. $7,667 and STDIV $1,002.35).

Table 3 shows different factors considered when energy-modeling of roof construction alternatives. According to the Factor Subset Selection, the depth of insulation was the most significant factor when calculating energy estimations of roof construction. But it is also interesting to find that adding air space to the roof was almost as important as increasing the depth of insulation (The significances of insulation and air space in roof construction are 83% and 79% respectively in Feature Subset Selection). Constructing and adding a ceiling to the roof turned out not significant as other factors such as roof materials, or insulation. The best option is either Wood Frame Roof with Super High Insulation or Metal Frame Roof with Super High Insulation. The worst option is Cool Roof—R15 continuous insulation over roof deck. Also it was found that increasing R value in Cool Roof continuous insulation over roof deck produced good results for that location. Table 3 shows different factors considered during the energy simulations.

3.1.3.3. Wall construction. Walls are constructed using various forms and materials for different types of buildings. Exterior walls are important to protect the building from external environmental

![Fig. 4. Energy modeling in eQuest.](image)

![Fig. 5. Electric consumption (kWh) for baseline model.](image)
effects such as heat and cold, sunlight, rain and snow, and sound. Walls in this research don’t contain doors and windows, which provide for controlled passage of environmental factors and people through the wall line. The doors and windows were considered later on in the research. A total number of 96 energy runs were performed for wall construction alternatives. Fig. 7 shows the annual energy cost results (Max. $10,973, Min. $7,565, Avg $8,251, Median $8,000, and STDIV $647.19).

The best result came from “Frame Wall with Super High Insulation combined with brick veneer” ($7,565) and the worst option was “Massive Wall” or metal frame) without insulation ($10,973). According to the Feature Subset Selection algorithm, the significant factors were Insulation (84%), Materials (81%), Air Space (79%), U-value (68%), and Construction method (45%). In wall construction, combined materials (concrete + brick veneer + insulation or face brick + tiles + insulation) also produced good energy efficiencies.

Table 4 shows five different factors considered in the energy simulations.

3.1.3.4. HVAC. Twelve HVAC systems below were modeled as improvements to the baseline system:
• 14 SEER/0.9 AFUE Split/Packaged <5.5 ton,
• 12SEER/0.9 AFUE Split/Packaged 5–11 ton,
• EER Packaged VAV 84.8 % boiler heating,
• Central VAV, HW Heat, Chiller 5.96 COP, Boilers 84.5 efficient,
• 4 Pipe Fan Coil System, Chiller 5.96 COP, Boilers 84.5 efficient,
• Central VAV, Electric Resistance Heat, Chiller 5.96 COP,
• 14 SEER/8.3 HSPF Split/Packaged Heat Pump,
• 12 SEER/7.7 HSPF Split Package Heat Pump,
• 2-Pipe Fan Coil System, Chiller 5.96 COP, Boilers 84.5 efficient,
• 12 SEER/8.3 HSPF Packaged Terminal Heat Pump (PTHP),
• 11 SEER Packaged Terminal Heat Pump,
• 17 SEER/0.85 AFUE Split/Package <5.5 ton.

Fig. 8 shows different performances of each HVAC option (AVG $8,100.5, Max. $9,156, Min. $6,746, Median $8,006, STDIV $802.56). All twelve systems showed different performances. The most energy efficient option was given from 17 SEER/0.85 AFUE Split/Packaged <5.5 ton. In general, Heat Pumps were more efficient than other types of equipment such as Fan Coil Units or Packaged

![Fig. 6. Estimated annual energy costs of roof construction alternatives.](image)

![Fig. 7. Estimated annual energy costs of wall construction.](image)

### Table 2
Electric consumption (kWh X 000) for baseline model.

<table>
<thead>
<tr>
<th></th>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
<th>May</th>
<th>Jun</th>
<th>Jul</th>
<th>Aug</th>
<th>Sep</th>
<th>Oct</th>
<th>Nov</th>
<th>Dec</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space cool</td>
<td>0.16</td>
<td>0.24</td>
<td>0.43</td>
<td>0.70</td>
<td>1.32</td>
<td>1.78</td>
<td>2.15</td>
<td>2.34</td>
<td>1.62</td>
<td>0.98</td>
<td>0.45</td>
<td>0.24</td>
<td>12.41</td>
</tr>
<tr>
<td>Heat reject</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Refrigeration</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Space heat</td>
<td>0.76</td>
<td>0.60</td>
<td>0.17</td>
<td>0.08</td>
<td>0.01</td>
<td>0.00</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0.01</td>
<td>0.18</td>
<td>0.58</td>
<td>2.39</td>
</tr>
<tr>
<td>HP supplement</td>
<td>0.09</td>
<td>0.15</td>
<td>0.00</td>
<td>0.00</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0.00</td>
<td>0.08</td>
<td>0.33</td>
<td>–</td>
</tr>
<tr>
<td>Hot water</td>
<td>0.45</td>
<td>0.42</td>
<td>0.46</td>
<td>0.47</td>
<td>0.44</td>
<td>0.38</td>
<td>0.39</td>
<td>0.38</td>
<td>0.35</td>
<td>0.39</td>
<td>0.37</td>
<td>0.43</td>
<td>4.94</td>
</tr>
<tr>
<td>Ventilation fans</td>
<td>0.05</td>
<td>0.06</td>
<td>0.05</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>1.15</td>
</tr>
<tr>
<td>Pumps &amp; auxiliary</td>
<td>0.67</td>
<td>0.61</td>
<td>0.61</td>
<td>0.59</td>
<td>0.65</td>
<td>0.67</td>
<td>0.71</td>
<td>0.70</td>
<td>0.68</td>
<td>0.60</td>
<td>0.59</td>
<td>0.67</td>
<td>7.76</td>
</tr>
<tr>
<td>External usage</td>
<td>0.66</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
<td>7.76</td>
</tr>
<tr>
<td>Misc. equip.</td>
<td>1.96</td>
<td>1.77</td>
<td>1.96</td>
<td>1.98</td>
<td>2.01</td>
<td>1.89</td>
<td>2.01</td>
<td>2.01</td>
<td>1.89</td>
<td>2.01</td>
<td>1.84</td>
<td>1.96</td>
<td>23.30</td>
</tr>
<tr>
<td>Task lights</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Area lights</td>
<td>4.10</td>
<td>3.68</td>
<td>4.03</td>
<td>4.15</td>
<td>4.12</td>
<td>3.76</td>
<td>4.12</td>
<td>4.14</td>
<td>3.83</td>
<td>4.22</td>
<td>3.72</td>
<td>4.12</td>
<td>47.97</td>
</tr>
</tbody>
</table>

### Table 3
Different factors considered during energy simulations of roof construction.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Data type</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Material</td>
<td>Non-numeric</td>
<td>Wood, asphalt, asphalt shingles, concrete (lightweight/heavyweight), stone, fiberboard, gyp board, metal r bätt insulation, etc.</td>
</tr>
<tr>
<td>Insulation</td>
<td>Numeric</td>
<td>1–12 (inches)</td>
</tr>
<tr>
<td>Construction method</td>
<td>Non-numeric</td>
<td>Regular/super-insulated</td>
</tr>
<tr>
<td>U-value</td>
<td>Numeric</td>
<td>0.02–0.5948</td>
</tr>
<tr>
<td>Air space</td>
<td>Non-numeric</td>
<td>Yes/No</td>
</tr>
<tr>
<td>Ceiling</td>
<td>Non-numeric</td>
<td>Acoustical ceiling, suspended ceiling</td>
</tr>
</tbody>
</table>
VAV units. This research was deliberately limited to analyze the annual energy cost of 12 HVAC systems since it would be too difficult in a research paper to try to identify all specific HVAC systems. However, further research could be undertaken to analyze more HVAC options to provide a broader pattern of HVAC systems.

3.1.3.5. Rotation. The proposed building was rotated at every 15° for a total of 12 times to determine the most energy efficient building orientation. Even though the best option was to have the longest wall face due south, there were small differences between each rotation ($8,094–$8,195). Fig. 9 shows the estimated energy costs of each orientation (Max. $8,195, Min. $8,094, Avg $8,100.3, STDEV $3.54).

3.1.3.6. Summary. Comparison of annual energy costs for each design alternative was conducted in the six categories of building elements. We studied 12 different options in HVAC, 127 in roof construction, 88 in wall construction, and 12 orientations.

Compared to the baseline building performance, HVAC options provided the greatest impact on annual energy costs, with estimated $1,507 savings using a 17 SEER/0.85 AFUE Split/Pkgd as shown in Table 5. On the other hand, building orientation had the least impact on energy costs ($11–17 per year) for this particular structure.

4. Conclusions and further research

Utilizing data mining based energy modeling technology, this research conducted an energy modeling process where project teams may utilize energy simulations and see the results early in the design process. To date, energy modeling tasks are usually conducted later in the design process due to its time consuming data entry. The case study revealed that data mining based energy modeling helps project teams discover useful patterns to improve the energy efficiency of building design during the design phase. The method developed during this research could be used to guide designers and engineers through the process of completing an early design energy analysis based on energy simulation models. There is an essential need to conduct further research with energy simulations early in the design process, such as during charrettes, to see to it that project teams can design an ultra-efficient building that takes advantage of free site energy for day-lighting, natural ventilation and passive solar heating. Once the building orientation and form is optimized, then high efficiency mechanical and lighting systems can be integrated to minimize annual energy consumption and work towards a “Net Zero” building. (A Net Zero building generates as much energy as it uses over a year.)

The integration of energy analysis tools with Building Information Modeling (BIM) Tools is a promising development because it would reduce the duplicative work of recreating a representation of the building design to conduct energy analysis. Instead, project delivery...
teams could apply these integrated tools to obtain better feedback on how their design decisions affect the building’s annual energy use. Next, the application of a data mining technique could really help project delivery teams sort through results of numerous energy analysis runs generated when designing a building to select the most cost effective alternative and system. Finally, the comparison between estimated computer results and real energy consumption would provide us a better understanding and reflection on the actual performance. The collection of actual energy and water consumption data is being requested by the US Green Building Council for projects being certified using the new version of the LEED (Leadership in Energy and Environmental Design) rating tool (LEED v.3) [10].

References