DEVELOPING ENERGY EFFICIENT BUILDING DESIGN IN MACHINE LEARNING

Hyunjoo Kim[1], Annette Stumpf[2] and Richard Schneider[3]

[1] Assistant Professor, California State University at Fullerton, CA 92834, PH (657) 278-3867, hykim@fullerton.edu
[2] Principal Investigator, US Army ERDC (Engineer Research and Development Center), Champaign, IL 61826, PH (217) 373-4492, Annette.L.Stumpf@usace.army.mil
[3] Principal Investigator, US Army ERDC (Engineer Research and Development Center), Champaign, IL 61826, PH (217) 373-6752, Richard.L.Schneider@usace.army.mil

ABSTRACT

Building energy simulation programs have been developed, enhanced, and are in wide-spread use throughout the building construction community (Stumpf et al. 2009). Energy modeling programs provide users with key building performance indicators such as energy use and demand, temperature, humidity, and costs. As the A/E/C industry is embracing the technology of energy simulation programs, building designers are currently encountering a large amount of data generated during energy simulations. From our experience, even a simple energy modeling run generates hundreds of pages of data. Examples of building features simulated include the estimated energy costs in terms of building orientation, HVAC system, lighting efficiency and control, roof and wall insulation and construction, glazing type, water usage, day-lighting and so on. Such volumes of data are simply beyond human abilities to identify the best combination of building components (insulation, windows, doors, etc.) and systems (heating and cooling systems, ventilation, etc.) during the building design process. Evaluating building energy modeling outputs clearly overwhelms the traditional methods of data analysis such as spreadsheets and ad-hoc queries. This paper presents the analysis to develop the energy efficient solutions with the baseline and target energy estimations. Finally, energy efficient solutions are presented that enable the energy savings to be met in fifteen different climate zones in the United States.

INTRODUCTION

Main objective of this research is to develop a machine learning system which can help project teams discover useful patterns for more energy efficient building design and make efficient decisions to construct energy efficient buildings. This paper utilizes the technology of machine learning 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. One can identify valid, useful, and previously unknown patterns of energy simulation modeling, by applying machine learning technology to the analysis of energy efficient building designs (Fayyad et al. 1996).

In order to test the feasibility of the proposed approach, a prototype of the data mining framework was developed and tested with a dataset generated during the energy simulation modeling of a building. Then detailed steps and their results for energy analysis are presented in fifteen different climate zones of the United States.

RESEARCH METHODOLOGY

A big trend in A/E/C industry today is designing sustainable buildings. For the past 50 years, a wide variety of building energy simulation (BES) programs have been developed, enhanced, and are in use throughout the building energy community. Examples of the BES are BLAST, EnergyPlus, eQUEST, TRACE, DOE2, ECOTECT, and so on (Crawley et al. 2005). From our experience, even a simple energy modeling run generates hundreds of pages of data. Examples of building features simulated include the estimated energy costs in terms of building orientation, HVAC system, lighting efficiency and control, roof and wall insulation and construction, glazing type, water usage, day-lighting and so on. Such volumes of data are simply beyond human abilities to identify the best combination of building components (insulation, windows, doors, etc.) and systems (heating and cooling systems, ventilation, etc.) during the building design process. Evaluating building energy modeling outputs clearly overwhelms the traditional methods of data analysis such as spreadsheets and ad-hoc queries. This research utilized a data mining approach to analyze a large amount of data generated during energy simulations. (Soibelman and Kim, 2002).

Data Mining Approach for Energy Analysis

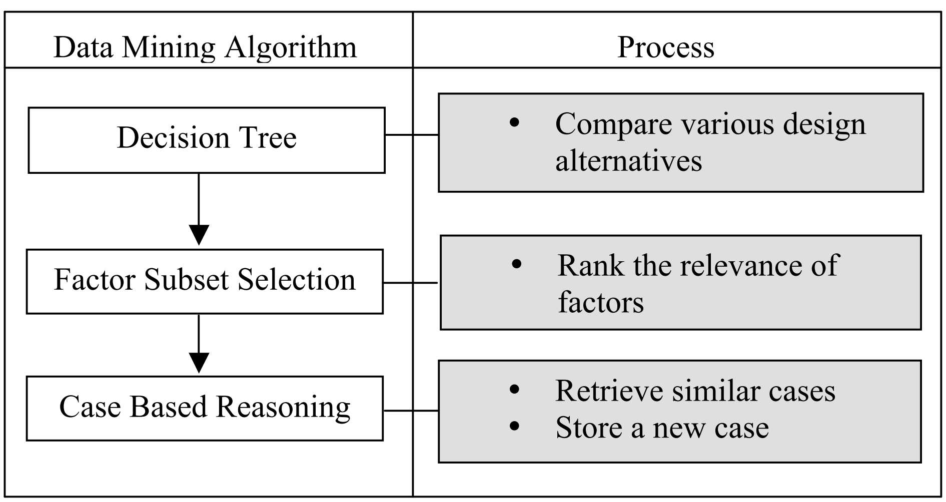

Figure 1. Energy Analysis Process

The proposed methods to be employed to find trends in a dataset are the data mining techniques of decision trees, case base reasoning and factor subset selections. For the initial design, a BIM (Building Information Modeling) model produced a building shell to generate multiple different design alternatives. The output of design alternative energy costs and fifteen different climate fuel units were analyzed using the data mining tools.

As shown in Figure 1, a Decision Tree process was employed to sort the design alternatives vs. annual energy costs. Then a factor subset selection program was used to find trends and rank the attributes by knowledge gained to give us an understanding of the importance of each factor. Lastly, Caspian, one of Case Based Reasoning tools, was utilized to retrieve the closest building and its location instance among previous cases in the fifteen different climate zones.

CASE STUDY

The problem domain is explored using various energy systems for an apartment at fifteen cities in the United States. Identifying the energy efficient design alternatives is not simple due to different climates, building types, and various design options. A dataset was built in fifteen different climate zones for an apartment building. The different configurations used to analyze the model were: rotation of the apartment building, HVAC systems, lighting efficiency, lighting control, roof type, wall construction, wall glazing type and wall glass amount. The energy estimation for these configurations must be analyzed through data mining against yearly temperature averages, monthly temperature averages, HDD (heating degree days), and CDD (cooling degree days). An BIM model of the apartment building was constructed as shown in Figure 2.

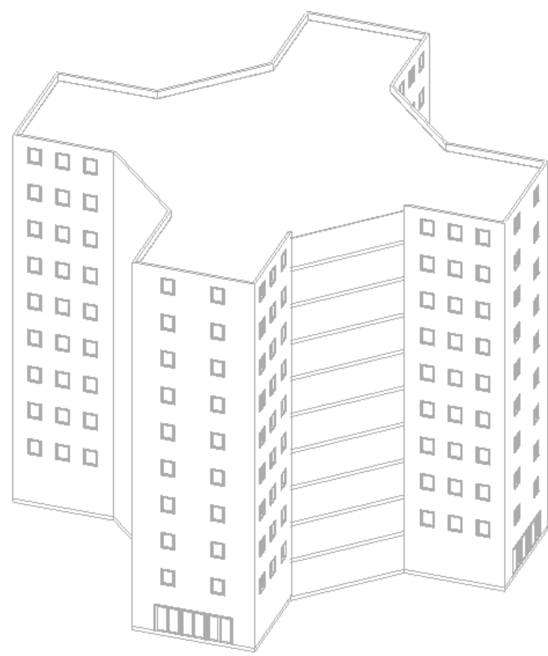

Figure 2. BIM model of an apartment building

Building Type	School Building
Area (SF)	71,994
Volume (CF)	575,955
Location	Fullerton, CA

Table 1. Building model and information

Decision Tree

As shown in 3, various design alternatives were simulated to find the most energy efficient options. From the decision tree algorithm, it was determined that rotation and wall systems materials have no real effect on the energy cost. The most important factor would be the HVAC systems and lighting controls because it can substantially lower the energy cost for the structure. This would be consistent with the design alternative information graphs in Figure 3. Decision Tree comprised of the design alternatives vs. annual energy costs to predict trends in data set. The following inputs were used for different design alternatives:

- Rotation - Input different rotation angles (-15 to 180) to the base building orientation and determine the estimated energy cost for each rotation of the building orientation.
- HVAC - Input different types of HVAC systems (SEER 14/8.3 HSPF Split Packaged Heat Pump, 17 SEER/0.85 AFUF Split/Pkdg, and so on)
- Lighting - Input different lighting efficiency design (sensors to 40% design reduction)
- Roof - Input different roofing material (i.e., wood frame or metal frame)
- Wall – Input different insulation type (metal frame with super insulation or code compliant insulation)

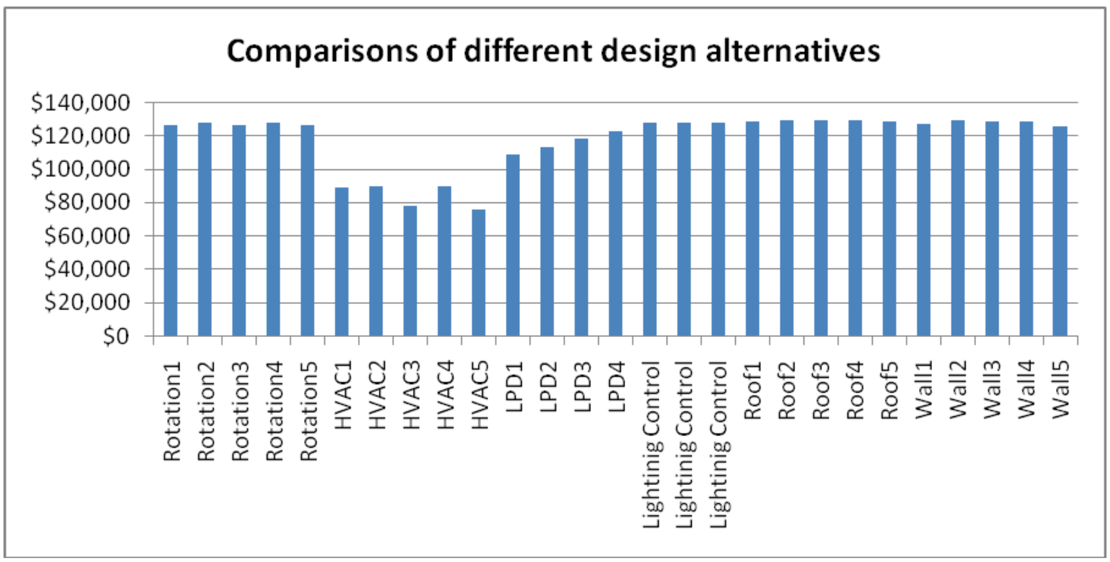

Figure 3. Comparisons of different design alternatives

Factor Subset Selection

The goal of factor subset selection is to understand the optimal subset of attributes and to make the data mining process more efficient by removing unimportant factors. With the amount of data available in the building industry, this data mining tool proved to be very useful to save time and computations. We utilized the WEKA (2009) data mining software to analyze the six (6) attributes. Figure 4 shows that in predicting estimated energy costs, locations, average temperature and heating degree days turned out to be important factors while annual lowest temperatures or humidity were relatively less important. We believe that the result was not highly comprehensive but can be improved with more cases.

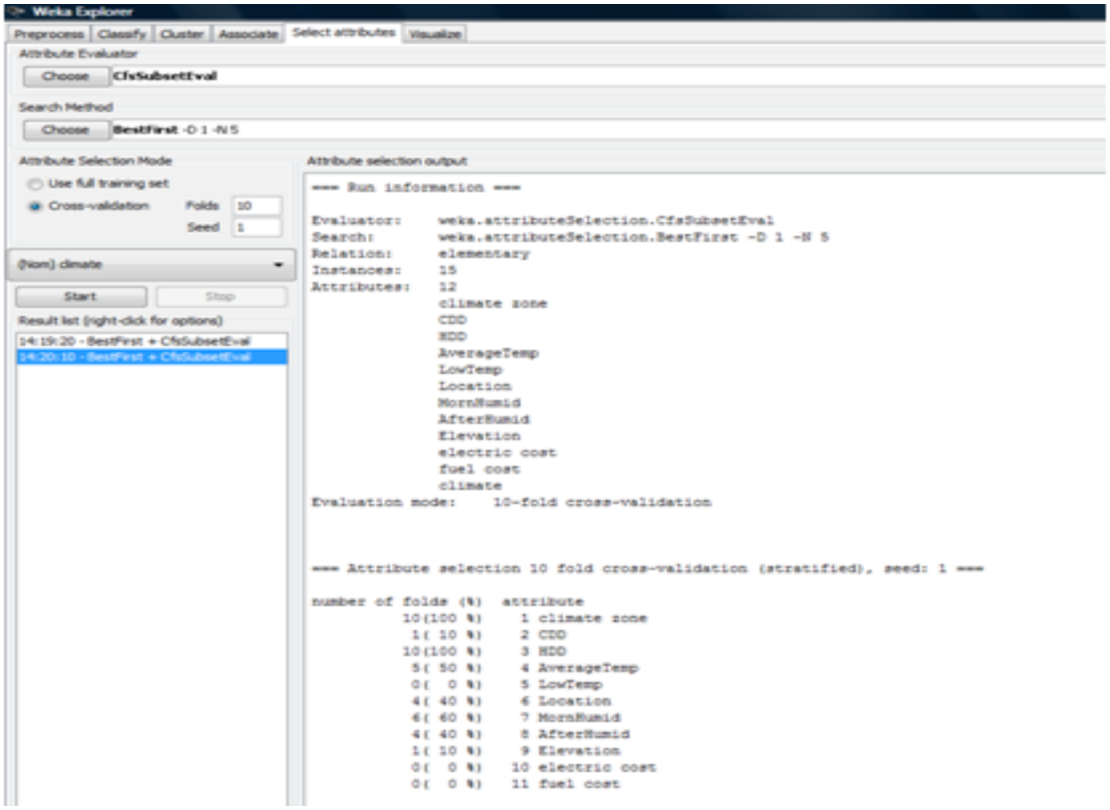

Figure 4. Result of Factor Subset Selection

Case Based Reasoning (CBR)

				Temp. Yearly Averages			Temp. Monthly Aveage			Annual Cost	
	Climate Zone	HDD	CDD	Average	Low	High	Lowest	Highest	Location	Baseline	Proposed
1	Miami Florida	200	9474	76	69	83	60	90	1A	$181,905	$102,070
2	Houston, Texas	1599	6876	69	58	79	41	94	2A	$153,076	$90,525
3	Phoenix, AZ	1350	8425	73	59	86	41	105	2B	$128,367	$76,232
4	Memphis, TN	3082	5467	62	53	72	32	92	3A	$149,750	$89,901
5	El Paso, TX	2708	5488	64	50	78	31	96	3B	$144,222	$87,292
6	San Francisco, CA	3016	2883	57	51	63	46	70	3C	$150,819	$80,477
7	Baltimore, MD	4707	3709	56	45	65	24	88	4A	$177,786	$113,455
8	Albuquerque, NM	4425	3908	57	43	70	23	92	4B	$122,513	$77,271
9	Seattle, WA	4908	1823	53	46	59	36	74	4C	$121,221	$66,578
10	Chicago, IL	6536	2941	49	40	59	13	84	5A	$136,567	$99,893
11	Colorado Springs, Co	6415	2312	49	36	62	17	85	5B	$108,512	$75,318
12	Burlington, VT	7771	2228	45	35	54	8	81	6A	$193,536	$189,447
13	Helena, MT	7699	1841	44	31	56	9	83	6B	$133,035	$98,101
14	Duluth, MN	9818	1536	39	29	48	-1	76	7A	$126,837	$105,281
15	Fairbanks, AK	13940	1040	27	17	37	-18	72	8A	$147,423	$117,471

Table 2. Annual energy costs in fifteen different climate zones

The CBR was used to retrieve similar case(s) and reuse the case to attempt to generate a pattern. The purpose of this section is to show that energy efficient solutions can be met in fifteen different climate zones. As expected, moving the base design model to different climate zones would have huge effects on the energy use of the building. Each of the fifteen climate zones and corresponding annual energy costs is illustrated in Figure 4. The range of baseline models based on ASHRAE 90.1-2004 specifications was from $115,432 in Colorado Springs, CO to $198,205 in Burlington, VT as shown in Figure 4. The range of annual energy costs with various combinations of design alternatives proposed was from a low of $66,578 to a high of $189,447 as shown in Figure 4, proposed model costs. The baseline model situated in Fullerton, California reported an annual energy cost of $121,250.

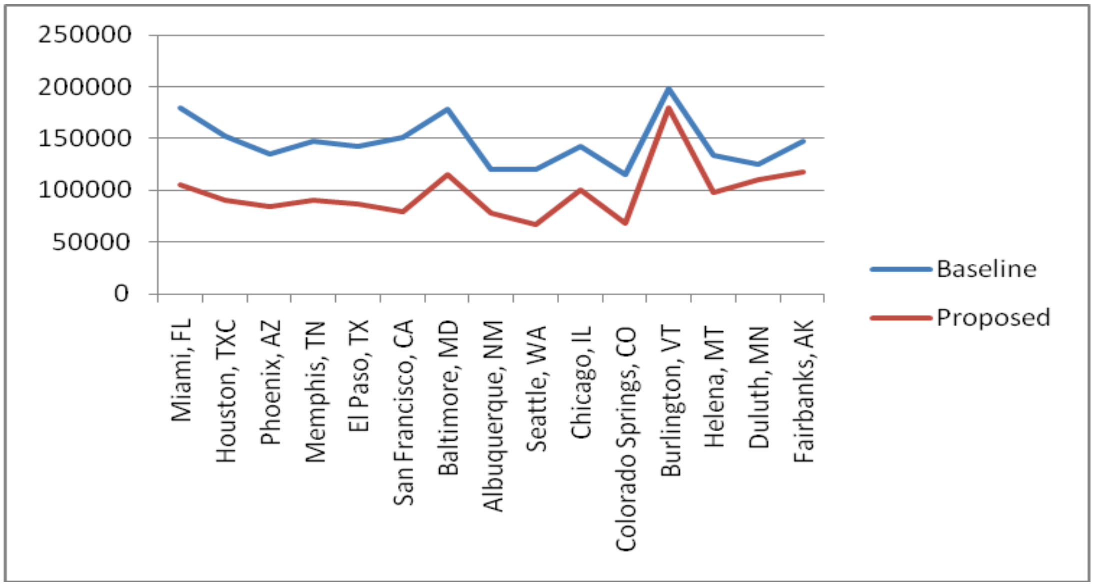

Figure 5. Baseline Model Costs of Apartment Complex in fifteeen different zones

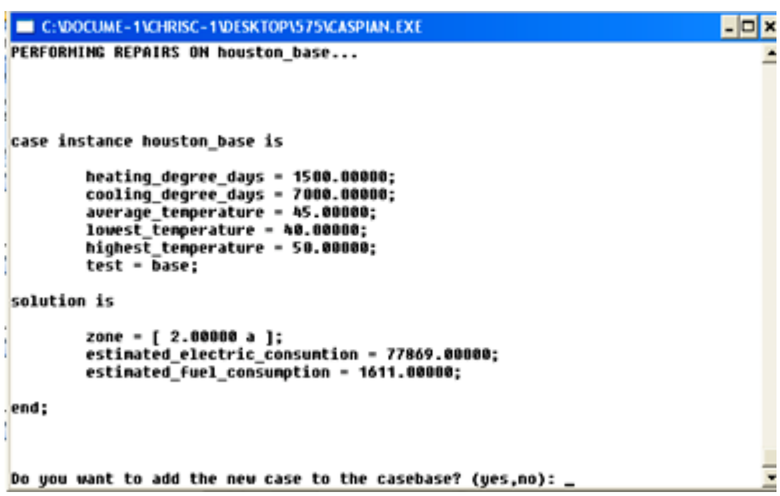

Figure 6. Case Based Reasoning system

Figure 5 shows a system, CBR in predicting two (2) output factors such as estimated energy costs and the location of a building based on five (5) input factors such as heating degree days, cooling degree days, average temparature, lowest temparature, and highest temperature. The CBR system used 30 different previous cases to predict the costs and location. The results were not as close as what was anticipated. As an example, the input of estimated cost requested 12,000 but the output shows a location with $173,919, which should be close to $100,000. It is believed that this was caused by insufficient number of inputs and outputs. This could be easily remedied in the future by adding more cases. Figure 5 shows an example of CBR system built in this research.

CONCLUSIONS

Advanced energy modeling techniques allowed us to generate a great amount of energy simulation results. Such volumes of data are simply beyond simple spread sheet or ad-hoc query to identify the best combination of building components during the building design process. From the energy analysis in this paper, we concluded that by using data mining techniques, we could identify patterns from a large amount of data and predict an estimated energy costs in different climate zones by utilizing data mining tools such as Decision Tree, Case Based reasoning and Factor Subset Selection. The Factor Subset Selection was used to determine which attributes are more relevant in predicting the estimated energy cost and its location. Decision Tree identified important patterns in various design alternatives in design process of a building. Lastly, Case Based Reasoning retrieved the most relevant case(s) from previously stored cases and determined its most likely climate zone.

REFERENCES

Crawley, D., Hand, J., Kummert, M. and Griffith, B. (2005), "Contrasting the capabilities of building energy performance simulation programs", Joint Report, Version 1.0

Fayyad, U., et al. (1996). "*Advances in knowledge discovery and data mining"*, AAAI Press/MIT Press, Cambridge, Mass., 1–34.

Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten (2009); The WEKA Data Mining Software: An Update; SIGKDD Explorations, Volume 11, Issue 1.

Soibelman, L and Kim, H. (2002), "*Data Preparation Process for Construction Knowledge Generation through Knowledge Discovery in Databases*", Journal of Computing in Civil Engineering, Vol. 16, January

Stumpf, A., Kim, H. and Jenicek, E. (2009), "*Development of Early Design Energy Analysis using BIMs (Building Information Models)*", ASCE Construction Research Congress