Sustainable Cities and Society 5 (2012) 37–43

Contents lists available at SciVerse ScienceDirect

Sustainable Cities and Society

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

Influential factors analysis on LEED building markets in U.S. East Coast cities by using Support Vector Regression

Jie Zhao*, Khee Poh Lam

Center for Building Performance and Diagnostics, School of Architecture, Carnegie Mellon University, Pittsburgh, PA 15213, USA

ARTICLE INFO

Keywords:
LEED
Green building market
Support Vector Regression

ABSTRACT

Building industry is closely related to current energy and environmental issues. Several green building codes and rating systems addressing the problems have been developed. Leadership in Energy and Environmental Design (LEED) rating system is recognized as one of the effective and widely adopted commercial building standards. LEED buildings were investigated in several green city and green building studies but only used as instances in static matrices. These studies were not able to answer the question why a particular city favors LEED. However, in this paper, three commonly used machine learning algorithms – Linear Regression, Locally Weighted Regression and Support Vector Regression (SVR) – are compared and SVR is used to investigate, discover and evaluate the variables that could influence LEED building markets in U.S. East Coast cities. Machine learning models are first created and optimized with the features of city geography, demography, economy, higher education and policy. Then SVR model identifies the key factors by dynamic self-training and model-tuning using the dataset. Via optimization, the correlation coefficient between the model's prediction and actual value is 0.79. The result suggests that population and policy can be important factors for developing LEED buildings. It is also interesting that higher education institutions, especially accredited architecture schools could also be driving forces for LEED commercial building markets in East Coast cities.

1. Introduction

Energy and environmental problems have become a critical concern for our society. The building sector, including commercial and residential has great impact on the issues. According to U.S. Department of Energy (DOE) Building Energy Data Book, in 2010 the total energy consumed by commercial and residential buildings accounted for 41% of total U.S. primary energy use, which has exceeded the consumption of the industrial (30%) and transportation sectors (29%) (Department of Energy, 2011a). Meanwhile, the building industry has negative impact on the environment, such as soil erosion, water pollution, construction waste, and ecological damage.

To address these problems, green building design, which primarily focuses on sustainable site, energy efficiency, water efficiency, environmentally friendly material, indoor environment quality, and other environmental aspects, has been well recognized as a next generation building design method. Numerous green building standards, codes, and evaluation systems have been developed globally. Leadership in Energy and Environmental Design (LEED) is one of the most popular rating systems. LEED was developed by the U.S. Green Building Council (USGBC) in 2000 and "it provides building owners and operators with a framework for identifying and implementing practical and measurable green building design, construction, operations and maintenance solutions" (U.S. Green Building Council, 2011b).

According to the USGBC project directory, the U.S. LEED building market has been growing dramatically since LEED was established in 2000. Fig. 1 shows the total number of LEED certified projects in the U.S. from 2000 to 2011. The average annual growth rate for the past 12 years is 103%.

Despite the economic recession which began in 2008, according to McGraw-Hill Construction, the value of green building construction increased 50% from 2008 to 2010 and represented 25% of total new construction activity in 2010 (Bernstein, 2010). Particularly, the author also mentioned that "LEED may impact growth of larger commercial green building market" by pointing out that, LEED projects certified between 2004 and 2009 have higher average value than non-LEED projects. In a nutshell, LEED has become an essential part and could be a driving force for the building industry in the U.S.

However, in different cities, the quantity of LEED certified projects varies significantly. It is worthwhile discovering the key factors that affect the popularity of LEED buildings in various places. By uncovering, investigating and advocating the influential

* Corresponding author at: MMCH-415 (IW), Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh 15213, USA. Tel.: +1 412 608 0510; fax: +1 412 268 6129.
E-mail address: jayzhao@cmu.edu (J. Zhao).

2210-6707/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.scs.2012.05.009

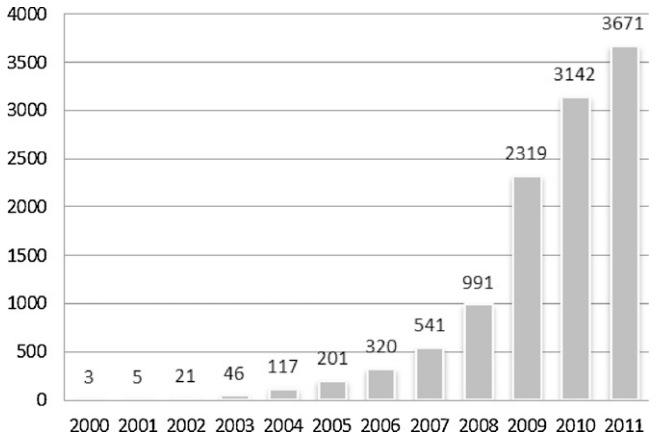

Fig. 1. Number of new LEED certified projects in the U.S. from 2000 to 2011.

features, LEED projects could be promoted through effective methods. In this research authors do not differentiate between platinum, gold, silver, and certified LEED projects. One reason is that data mining requires large dataset. However, the dataset for different levels of LEED certification is skewed. For example there are only 659 LEED Platinum buildings out of 11,377 LEED overall certified buildings for the entire U.S. over 12 years. It is nearly impossible to find general patterns for different cities with a small dataset. It is necessary to only consider quantity of all certified LEED buildings rather than separate them by levels of certification. In addition, this study tries to find reasons why people in a city favor LEED buildings. Instead of focusing on a few "deep green" projects that can achieve LEED Platinum awards, looking at a bigger all inclusive picture and encouraging more people to be involved in LEED activities will eventually raise the bar for the whole society.

2. Related work

Several green building and green city ranking studies were conducted based on the information of LEED projects. "2011 Green Building Opportunity Index (GBOI)" is one of them (BetterBricks & Cushman & Wakefield, 2011). In this research, office markets in the 25 biggest cities in the U.S. have been ranked by data analysis in the following 6 categories: office market conditions, investment outlook, green adoption and implementation, mandates and incentives, state energy initiatives, and green culture. The purpose of this research is to discover the "best" city for green building investment. GBOI ranked those cities based on 6 categories of indicators that help reveal the investment value of green buildings in the cities. Among the 6 categories, "green adoption and implementation" uses LEED statistics data as indices. Specifically, total floor area of LEED certified CBD (Central Business District) private sector commercial office space, total floor area of LEED certified as a percentage of inventory, the number of accredited LEED professionals per capita, and the number of accredited LEED EBOM (Existing Building Operation and Management) specialists per capita have been examined as variables. The study presented that San Francisco has the highest score over the 6 categories, which implies potentially, the city potentially have the highest return when investing on green buildings. However, the authors did not provide sufficient information on the matrices or algorithms they used for the ranking. So it is difficult to determine how the variables are weighted and what the key factors are.

The Economist Intelligence Unit (EIU) sponsored by Siemens conducted a study named "US and Canada Green City Index" (Economist Intelligence Unit, 2011). In this study, 27 major cities in the U.S. and Canada have been ranked based on information in the following categories: CO_2, energy, land use, buildings, transportation, water, waste, air and environmental governance. In the building category, the number of LEED certified (platinum, gold and silver) buildings per 100,000 people has been used as one of the 3 equally weighted indicators. The other two indicators are energy efficient building standards and energy efficient building incentives. The research assessed and ranked the cities by using a relatively simple matrix containing 31 equally significant indicators determined by experts.

Likewise, SustainLane, an online community, conducted a study on the U.S. sustainable city ranking of 50 largest U.S. cities (SustainLane, 2009). One criterion for the sustainable city ranking is green building. In this category, the numbers of both LEED certified and registered buildings in each city have been analyzed. The numbers were normalized on a per 100,000 people basis and received a weight of 1.0. The other attributes were given a weight of 0.5, 1.0 or 1.5. Similar to the EIU study above, the factors are manually weighted by experts.

It is evident that the previous research relating LEED building data with building markets and city green index only simply treated LEED building data as one of many indicators in a static matrix that is manually generated. Although those variables are intuitively associated with the popularity of LEED buildings in a city, these studies manually defined weights or matrices over those potential factors and failed to provide a scientific method to demonstrate the relevance and/or causal relationship between the factors and the number of LEED buildings. Thus arguably they neglect the question why these so-called green cities have the ability to attract developers to adopt green practice. The following questions have not been but should be addressed. What are the relative importance of the features in the ranking process? Why does a certain feature receive a weight of 0.5 and not 1.5? What are the influential features associated with awareness of LEED buildings in a city? To answer these questions, in this paper, instead of pre-defining a static matrix for potential factors, a machine learning method is applied to build a regression model to examine the potential factors, distribute weights accordingly and investigate the ones that have strong influence on the number of LEED buildings in a city.

3. Data collection and cleaning

3.1. Class

In the numeric machine learning field, there are two types of variables; (1) "attribute," which represent the learning variables used to build, train and test model prediction; (2) "class, (or prediction result)" which is a variable that should be predicted and tested. The performance of machine learning model is typically evaluated by comparing true class values with predicted class values of thousands of instances. In order to conduct data mining, the first step is to acquire considerable amount of data, including both true class and attribute values. In this case, the objective is to learn what factors can influence the number of LEED certified commercial buildings in the U.S. East Coast cities. There are two reasons why only East Coast cities are considered in this study. First, from the economic, social and cultural standpoint, East Coast cities are significantly different from cities other parts of the U.S. These cities typically have longer history, more universities, higher population density, and higher average income (City-Data, 2012). About 36% of U.S. population lives in the East Coast (National Oceanic and Atmospheric Administration, 1998). The buildings in East Coast cities usually have smaller living area and limited surrounding space (East-West Property Advisors, 2010). They also feature older and more traditional architectural styles, while many cities in the West Coast only have 50 years of building history (Dan, 2009). In