Advanced Engineering Informatics 32 (2017) 224–236

Contents lists available at ScienceDirect

Advanced Engineering Informatics

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

Full length article

Selection of target LEED credits based on project information and climatic factors using data mining techniques

M.A. Jun, Jack C.P. Cheng *

Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong

ARTICLE INFO

Article history:
Received 18 May 2016
Received in revised form 9 March 2017
Accepted 13 March 2017

Keywords:
Adaboost Decision Tree
Climate factors
Decision support system
LEED
Random Forest
SVM

ABSTRACT

Developed by the United States Green Building Council, Leadership in Energy and Environmental Design (LEED) is a credit-based rating system that provides third-party verification for green buildings. Selection of target credits is important yet challenging for LEED managers because various factors such as target certification grade level and building features need to be considered on a case-by-case basis. Local climatic factors could affect the selection of green building technologies and hence the target credits, but currently there is no research suggesting target LEED credits based on climatic factors. This paper presents a methodology for the selection of target LEED credits based on project information and climatic factors. This study focuses on projects certified with LEED for Existing Buildings (LEED-EB). Information of 912 projects and their surrounding climatic circumstances was collected and studied. 55 classification models for 47 LEED-EB credits were then constructed and optimized using three classification algorithms - Random Forests, AdaBoost Decision Tree, and Support Vector Machine (SVM). The results showed that Random Forests performed the best in most of the 55 classification models. With a combination of the three algorithms, the trained classification models were used to develop a web-based decision support system for LEED credit selection. The system was tested using 20 recently certified LEED projects, and the results showed that our system had an accuracy of 82.56%.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

Developed by the United States Green Building Council (USGBC), Leadership in Energy and Environmental Design (LEED) is one of the major rating systems for evaluating green building performance. LEED is a credit based rating system, in which each credit evaluates one aspect of building performance and grades relative credit points. The totally achieved LEED credit point determines the award of different LEED grade levels, ranging from Platinum, Gold, Silver, to Certified. The latest version of LEED system is LEED 2009, which is sometimes referred to as LEED v2009 or LEED v3. LEED 2009 has separate rating systems for different project types, such as LEED for New Construction, LEED for Existing Buildings, LEED for Schools, LEED for Commercial Interiors, and LEED for Healthcare. All these LEED 2009 systems have a maximum credit point of 110, which include 100 base points, 6 additional points for Innovation in Design, and 4 additional points for Regional Priority. The 100 base points are grouped in five credit categories: (1) Sustainable Sites, (2) Water Efficiency, (3) Energy and Atmosphere, (4) Materials and Resources, and (5) Indoor Environmental Quality. Buildings are required to achieve 80 points or above to be certified as LEED Platinum, 60–79 points for Gold, 50–59 points for Silver, and 40–49 points for Certified [1].

Selection of target LEED credits is often an important yet challenging problem for LEED project managers due to limited budgets, tight project schedules [2]. Researchers have attempted to analyze the relationships between decision making factors and LEED credits. For example, Ma and Cheng [3] studied the influence of related LEED credits on the problem of target credit selection. Madanayake and Ruwanpura [4] explored a credit selection strategy for LEED projects based on multiple criteria like cost, productivity and environmental impact into consideration. Juan, Gao and Wang [5] addressed a strategy for the selection of green building technologies by optimizing the tradeoff between cost and possible achievements in various green building rating systems. Castro-Lacouture, Sefair, Flórez and Medaglia [6] studied possible influential factors and explored an optimal strategy for material selection in LEED projects. Zhao and Lam [7] conducted research to analyze the influence of city-level factors over the local LEED building markets in east coast cities in the United States. In addition to these aspects, climatic factors like sunshine percentage could also affect the

* Corresponding author.
E-mail address: cejcheng@ust.hk (J.C.P. Cheng).

http://dx.doi.org/10.1016/j.aei.2017.03.004
1474-0346/© 2017 Elsevier Ltd. All rights reserved.

selection of target credits [8]. Climate consideration is one important reason why LEED managers have different preferences for target LEED credits for projects in different locations even with the same target level [9,10]. However, few literatures have studied local weather and suggest target LEED credits taking climatic factors into account.

In addition, LEED managers frequently meet new types of projects with different project characteristics and requirements. Therefore, it would be helpful if LEED managers could refer to other similar certified green building cases when planning and designing LEED projects. To learn from the past data and discover hidden knowledge for further decision making, researchers often use data mining technologies. However, there were not many researches on using data mining techniques to learn from the past and suggest target LEED credits.

This study therefore aims to explore the relationship between the achievement of LEED credits and the project information, as well as the surrounding climatic factors, and to develop a methodology for target credit selection based on the relationships using data mining techniques. The first part of our study has already been published [8]. It explored and numerically verified the relationship between climate factors and LEED credits. This paper presents the second part of this study, which is to develop a methodology for target credit selection based on the relationships between climate factors and also the project information using data mining techniques. The proposed methodology for LEED-EB 2009 credit selection could provide users with a list of suggested target credits for their LEED projects according to the project location, the target grade level, the owner type, the gross floor area, and the total property area. For experienced users, the suggested target credits can help remind them not to miss the credits they could have achieved. For novice, the list of target credits can help them have an idea on how similar projects select target credits. Since retrofitting existing buildings has a high potential impact to the built environment [11], increasing numbers of building owners have started to pursue the certification of LEED for Existing Buildings (LEED-EB). Therefore, LEED-EB was selected as the focus of this study.

To investigate the relationship between LEED credit selection and the climatic factors, 912 LEED-EB 2009 certified projects were studied. The climatic conditions of the 912 projects were collected from the U.S. National Climate Data Center. By setting the climatic factors and other project attributes as the variables and the credit achievement as the class, classification models were then built for each LEED-EB credit. Each credit has a binary classification model to predict whether or not the project should pursue. 55 credits (including the High/Low credits) result in 55 classification models. Each model will be built and tested using three classifiers – Random Forests, AdaBoost Decision Tree, and Support Vector Machine, and select the optimal one. The optimized models were then incorporated into a web-based decision support system (DSS) as the engine. Finally, the DSS was tested and verified with 20 recent LEED-EB 2009 certified projects. As for the structure of the paper, Section 1 introduces the background. Section 2 presents the methodology framework and the algorithms used. Section 3 and Section 4 describe the data collection and preprocessing processes. Section 5 presents and discusses the data mining results, whereas Section 6 shows the DSS development and validation. Section 7 concludes the whole work.

2. Methodology

2.1. Data mining and classification

In order to study the suggestions on target credit selection based on climatic factors through historic data, data mining techniques were used in this work. Data mining is a computational process for discovering knowledge from a dataset and transforming such knowledge into an understandable structure for further use [12]. Several studies have used data mining techniques to discover "hidden" knowledge in LEED or green buildings. Kavousian, Rajagopal and Fischer [13] studied the energy consumption factors in green homes using classification techniques. Kim, Stumpf and Kim [14] addressed energy efficient building design using pattern recognition techniques. By using association rule mining techniques, Cheng and Ma [2] explored the LEED credit bundles that could be achieved using the same green building technologies. However, these studies did not consider climatic factors surrounding each LEED project.

In this study, classification techniques in data mining were used because to suggest which credit to select is a predictive problem. Classification is a predictive technique aiming to assign a new observation to a specific class, based on a training dataset containing observations (or instances) whose class assignment is known [15]. Classification techniques have been used in various civil and infrastructure engineering problems such as categorization of vehicle types, classification of traffic flow patterns and freeway operating conditions, road defect detection, and identification of structural modal parameters [16–25]. However, classification techniques are seldom used in green building design and construction studies.

As shown in Fig. 1, the methodology framework of this research is divided into two parts: data preparation and classification. In data preparation, credit achievements of 912 LEED-EB 2009 certified projects were collected from the USGBC website. Connecting through project ID, we also collected project information of the 912 LEED-EB certified projects from the official LEED project directory provided by the USGBC. In addition, climate information surrounding the 912 projects was obtained from the United States National Climate Data Center (NCDC) climate database, which provides data of 15 climatic factors such as temperature and precipitation. In classification, the LEED credit achievement was set as the target class while the surrounding climatic factors and other projects attributes were set as the variables. Multi-point LEED credits were preprocessed and separated. 55 classification models were eventually built for the 47 LEED credits. Each classification model will be trained using three different algorithms, and the best one will be picked. The predicted results of the models become the suggestions from the decision support system to help selection of target credits for LEED managers.

Three classification techniques namely Random Forests, AdaBoost Decision Tree, and Support Vector Machine (SVM) were used in this study to build and optimize the classification models for each LEED credit, due to their high prediction accuracy for many classification problems in general [15,26]. These classification techniques will be described below.

2.2. Random Forests

The Random Forests algorithm is based on the Classification and Regression Tree (CART). CART is an algorithm for decision tree learning, which is a classification method commonly used in data mining. The goal of decision tree learning is to create a model that predicts the value of a target variable based on several input variables. Each interior node corresponds to one of the input variables, for example owner type, target level or climate factors. There are edges to the branches for each of the possible values of that particular input variable. Each leaf represents a value of the target variable given the values of the input variables represented by the path from the root to the leaf [27], for example cases with target level higher than silver, or cases with sunshine percentage higher than 50%, etc.

Download English Version:

https://daneshyari.com/en/article/4911046

Download Persian Version:

https://daneshyari.com/article/4911046

Daneshyari.com