Analysis of the Related Credits in LEED Green Building Rating System Using Data Mining Techniques

Jun Ma[1] and Jack C.P. Cheng[2]

[1]Department of Civil and Environmental Engineering, the Hong Kong University of Science and Technology; email: jmaae@ust.hk
[2]Department of Civil and Environmental Engineering, the Hong Kong University of Science and Technology; email: cejcheng@ust.hk

ABSTRACT

Like many green building rating systems, LEED (Leadership in Energy and Environmental Design) certifies green buildings as different grades according to the number of credit points the buildings have achieved. LEED project consultants often attempt to maximize the number of credit points to be achieved strategically with limited budgets and resources. For example, some green building technologies and features can be used to achieve multiple credits with little additional effort. Therefore, some applicants have studied the relationship and similarity in scope among the credits in green building rating systems, thereby finding ways to achieve multiple related credits. Some green building guides, such as the official LEED reference guide, also give suggestions on related credits. However, there has been a lack of study testing the strength of these suggestions. This paper aims to evaluate the strength of these suggested relationships by using data mining techniques. A database of certified green building projects was constructed based on the data from the USGBC website. The credits achieved by these projects were analyzed using association rule mining techniques. The strength of the suggested credit rules were identified according to the calculated outcomes. The results show that some suggested credit rules are related and commonly co-occur.

INTRODUCTION

Leadership in Energy and Environmental Design (LEED) is the most widely used green building rating system that provides credit-rating based third party verification of green building projects. It consists of many different prerequisites and credits addressing different aspects of building performance, such as water efficiency and indoor environmental quality. Buildings that satisfy the prerequisites earn points to achieve different levels of certification. There are many sub-systems in LEED for different building types, among which LEED for New Construction and Major Renovations (LEED-NC) is the one that evaluates the green performance of new constructions or existing buildings with major renovations. Since LEED-NC is the most commonly pursued LEED system, it was selected as the focus of this study. In the version 3 of LEED-NC (LEED-NC v3), there are 7 categories of prerequisites and

credits: (1) Sustainable Sites (SS), (2) Water Efficiency (WE), (3) Energy and Atmosphere (EA), (4) Materials and Resources (MR), (5) Indoor Environmental Quality (IEQ), (6) Innovation in Operations (IO), and (7) Regional Priority (RP). These different categories have different numbers of credits, and each credit evaluates the building performance and gives relevant points. Buildings are certified with LEED-NC v3 according to the following scale: 40–49 points for Certified, 50–59 points for Silver, 60–79 points for Gold, and 80 points and above for Platinum (USGBC, 2008).

LEED managers often attempt to achieve as many credits as possible with limited budgets and resources. Some green building technologies and features can be used to achieve multiple credits with little additional effort. For example, implementing a stormwater collection system not only helps achieve the LEED credit SSc6.1 (Stormwater Design, Quantity Control), but also saves potable water from landscaping by using the rainwater and thus helps achieve WEc1 (Water Efficient Landscaping). In this case, SSc6.1 and WEc1 can be seen as related credits. Project managers may thus pay more attention to these two credits when designing green building projects. Therefore, study of the relationship and similarity among credits in LEED rating systems is becoming more important. However, little research has been conducted to address this issue. (Thomas, 2008) studied the related credits of LEED-NC v2.0 using data mining techniques, but his results are not up-to-date now as LEED-NC was updated to the new version v3 in 2009. (Cheng & Ma, 2013) explored the credit bundles of LEED for Existing Building using association rule mining techniques, but this study only considered the correlation relationships rather than the directional relationships, and was therefore not so practical.

In this study, the relationships among credits of LEED-NC v3 were explored and identified using data mining techniques in two aspects. The first is to test the strength of the suggested “related rules” proposed by U.S. Green Building Council (USGBC), which are published in the official LEED Reference Guide for Green Building Design and Construction (the Guide for short) (USGBC, 2009). Since USGBC is the “rule maker”, their suggestions on the related credits are considered as reasonable and practical; however, some of the suggestions are weak and thus seldom followed by LEED managers. This study aims to explore the strength of these suggested rules through information on certified projects. Among the suggested rules, the ones reveal high performance in association rule analysis are viewed as strong credit rules. Besides this aspect, new rules are also discovered during the data mining process.

DATA MINING

Data Collection. Projects certified with LEED-NC v3 were collected from the USGBC website (USGBC, 2013). A C++ based tool was built to help collect the data. Consider that different countries may have different green building policies, market influences and technology preferences for LEED projects, only those located in the United States were considered in this study. Table 1 summarizes the collected LEED projects in the United States.

Table 1. Summary of the Collected LEED Projects in the United States.

Level	Certified	Gold	Platinum	Silver	Grand Total
Projects	457	372	84	472	1385

Preprocessing. Credits within "Innovation in Design" and "Regional Priority" categories were excluded in this study, because many work tasks performed to achieve these credits are unique and cannot be generalized for data mining. This results in 46 credits left in this study. These 46 credits were transformed in a way that is suitable for association rule analysis, leaving out as little information as possible. There are two kinds of credits in the remaining 46 credits – binary credits and multi-point credits. For the binary credits, 0/1 is used to represent whether the project has achieved the credit. For the multi-point credits such as EAc1 (Optimize Energy Performance) (1-19 points) and EAc2 (On-site Renewable Energy) (1-7 points), higher points reveal a better building performance. The multi-point credits were divided into two credit items – one uses 0/1 to represent whether the project achieved the credit; another one uses high/low to represent whether or not the project has achieved points higher than average. After preprocessing, 46 credits were transformed into 55 credit items. Fig.1 provides an excerpt from the project-credit database.

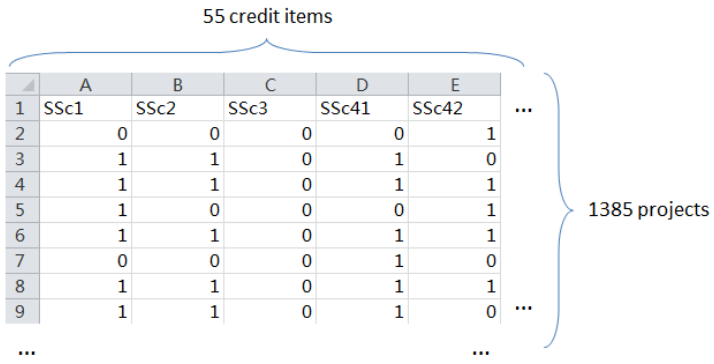

Fig.1. Excerpt from the credit achievement database.

Association Rule Mining. Association rule mining is a popular method for discovering interesting relations between variables in a database. By using different measures, strong rules are detected for discovering regularities between products in large-scale transaction data. For example, the rule {onions, potatoes} => {burger} found in the sales data of a supermarket would indicate that if a customer buys onions and potatoes together, he or she is likely to also buy hamburger meat (Agrawal, Imieliński, & Swami, 1993). In this study, similar relationships between LEED-NC v3 credits were explored using association rule mining techniques.

Support and Confidence. Support and confidence are the basic measures in association rule analysis. They are given by the following two equations (Cheng & Ma, 2013).

$$\text{Support}(\text{A}, \text{B}) = \text{P}(\text{A} \cap \text{B}) = \frac{Support\_Count(A \,\&\, B)}{Total\ Cases} \quad (1)$$

$$\text{Confidence}(\text{A} => \text{B}) = \text{P}(\text{B}|\text{A}) = \frac{Support\_Count\ (A \,\&\, B)}{Support\_Count\ (A)} \quad (2)$$

where A and B represent two credit items, and the Support Count represents the count of occurrence. It can be seen from the equations that Support refers to the probability of the occurrence of a certain patterns, while Confidence refers to the conditional probability of the occurrence given the occurrence of another credit item.

A higher support value means a higher occurrence possibility of the two credit items, but this may also result from the combination of easy credits (high achievement rates) rather than strongly related credits. Therefore Confidence, which measures the reliability of the inference made by a rule, is required to enhance the evaluation.

Confidence Difference to Prior (Difference). People sometimes enhance the strength of the rules with another measure called Confidence Difference to Prior (Cheng & Ma, 2013). This compares the posterior and the prior confidence of an association rule. Here the prior means the former item of a rule, while the posterior means the latter item of a rule. It is given by the following equation.

$$\text{Difference}(\text{A} => \text{B}) = |\text{Confidence}(\text{A} => \text{B}) - \text{Confidence}(\text{B})| \quad (3)$$

where the Confidence of B means the Confidence of All => B, which equals the Support of B. This measure compares the posterior and the prior Confidence of an association rule (Borgelt, 2004). The former should differ considerably from the latter if the rule is valued.

RESULTS AND DISCUSSIONS

Suggested Credit Rules. The Guide, which was published by USGBC, provides detail instructions on how to pursue each credit including potential technology, design procedure and criteria, etc. Besides, it also provides information about the related credits. For example, it states SSc8 (Light Pollution Reduction) is related to IEQc6.1 (Controllability of Systems – Lighting) because automatic occupancy controls to shutoff interior perimeter lighting should be coordinated with occupant controllability objectives. In this study, the strength of these suggested rules were analyzed by detecting their statistical relationships based on the certified 1385 projects.

On one hand, all the 239 suggested credit rules were manually extracted from the USGBC Guide. On the other hand, Support, Confidence and Difference were calculated (Equations 1 to 3) for all the possible credit rules. Note that since there are 55 credit items, pairwise combination of the credit items will generate 54*55=2970 rules.

Evaluation of the Strength. The experiments and calculations were coded with R (3.0.2). Based on the data mining results, the measurement values and ranks of the 239 suggested rules out of 2970 rules for each association rule criterion were obtained. The outcome is partly shown in Table 2. Because of the limited space, only the top 10 rules out of 239 suggested rules are presented. The rank column means the rule's rank out of the totally 2970 rules.

Table 2. Top 10 rules out of 239 suggested credit rules based on three criteria.

Rules	**Support**	Rank	Rules	**Confidence**	Rank	Rules	**Conf. Difference**	Rank
{WEc1} => {EAc1}	0.8592	1	{EAc2_high} => {EAc1}	1.0000	3	{IEQc8.1} => {IEQc8.2}	0.3673	6
{EAc1} => {WEc3}	0.8390	4	{IEQc7.2} => {IEQc7.1}	0.9988	7	{MRc3} => {MRc1.1}	0.3074	7
{IEQc7.1} => {EAc1}	0.8245	7	{EAc2} => {EAc1}	0.9974	8	{SSc5.1} => {SSc5.2}	0.2888	9
{EAc1} => {IEQc7.1}	0.8245	8	{IEQc2} => {EAc1}	0.9825	13	{MRc7} => {IEQc4.4}	0.2651	12
{SSc7.2} => {EAc1}	0.7798	13	{IEQc7.1} => {EAc1}	0.9811	16	{SSc6.1} => {SSc62}	0.2502	14
{EAc1} => {SSc7.2}	0.7798	14	{EAc6} => {EAc1}	0.9808	17	{MRc6} => {MRc5}	0.2412	16
{WEc3} => {WEc1}	0.7560	17	{EAc5} => {EAc1}	0.9801	19	{SSc4.1} => {SSc2}	0.2238	20
{IEQc3.1} => {IEQc4.2}	0.7552	19	{IEQc6.1} => {EAc1}	0.9782	20	{MRc6} => {MRc4}	0.2203	23
{IEQc4.2} => {IEQc3.1}	0.7552	20	{IEQc8.1} => {EAc1}	0.9780	22	{SSc62} => {SSc6.1}	0.2191	24
{IEQc4.1} => {IEQc4.2}	0.7408	23	{IEQc6.2} => {EAc1}	0.9774	25	{SSc2} => {SSc4.1}	0.2112	25

*Note: For detail description of the credits, please refer to the LEED-EB v2009 system (USGBC, 2008).

The results in the Support table (the left-side section in Table 2) may not reveal strong rules since high support values may be resulted from commonly achieved credits rather than related credits. For example, for the top rule {WEc1} => {EAc1}, WEc1 was achieved 1218 (87.9%) times in the 1385 projects while EAc1 was achieved 1352 (97.6%) times. Therefore, these two credits co-occur frequently and the high support value may not reflect the direct relationship between these two credits.

The Confidence value means the conditional probability, and it reflects the probability of the occurrence of one credit given the occurrence of another credit. Fig.2 shows a histogram of all the suggested rules grouped by their confidence values. It can be seen that the largest proportion of the confidence interval is (0.9, 1).

However, Confidence sometimes is misleading. For example, as shown in the middle section in Table 2, EAc1 is the posterior item in 9 out of 10 rules. Statistically, this is because the achievement rate of EAc1 is too high (97.61%), which results in that the confidence of EAc1 related rules become significantly higher than the other rules. Although these rules make practical sense from an engineering aspect, the Confidence values could be misleading for ranking purposes.

The third table is based on the Difference measure, which offsets the bias of the Confidence by deducting the Confidence of the prior item. In this case, those highly ranked rules based on Confidence that rely on the high achievement rate of the prior item become less weighted. The top rules of the third table are valid. For example, the second rule (ranked 7) shows that MRc3 (Materials Reuse) contributes to MRc1.1 (Building Reuse – Maintain Existing Walls, Floors and Roof). In fact, the development of a comprehensive reuse management plan that evaluates the

anticipated materials saved could help determine whether a project meets the requirement of MRc1.1.

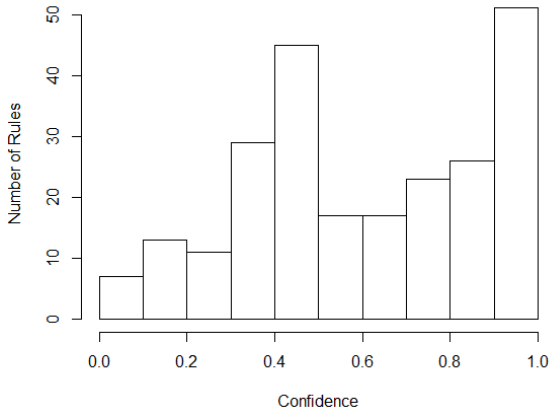

Fig.2. Confidence value of the suggested rules.

As a result, rules with Difference > 0.2 were selected as the strong rules. In addition, rules addressing the same aspect of performance such as {IEQc8.1} => {IEQc8.2} (Daylight and View) and {SSc6.1} => {SSc6.2} (Stormwater Design) were treated as easy rules and were filtered out in the outcome. Table 3 shows parts of the discovered strong rules.

New Rules. According to the data mining results, many rules reveal a higher performance than that of the suggested rules. Table 4 lists the top 10 rules out of total 2970 rules. After eliminating the easy rules and duplicated rules that can be ignored, there are 4 interesting rules left. They can be summarized into two rules – one is {EAc2} => {EAc1_high} and the other is {MRc1} => {SSc6.2}.

The first rule between EAc2 (On-site Renewable Energy) and EAc1 (Optimize Energy Performance) is reasonable. The amount of energy-use saved from the implementation of on-site renewable energy can help a lot in earning high points in EAc1. As a result, achieving EAc2 will contribute to achieving high EAc1.

Table 3. Examples of the Discovered Strong Rules.

Rank	Rules	Confidence	Conf. Difference	Practical Suggestion
7	{MRc3} => {MRc1.1}	0.4706	0.3074	The development of a comprehensive reuse management plan that evaluates the anticipated materials saved will help determine whether the project meets the requirement of MRc1.1.
MRc1.1 (Building reuse - maintain existing walls, floors and roof) MRc3 (Materials reuse)				
12	{MRc7} => {IEQc4.4}	0.8687	0.2651	When specifying mixed FSC materials, determine whether the finished product will be free of urea-formaldehyde in order to contribute IEQc4.4.
MRc7 (Certified wood) IEQc4.4 (Low-emitting materials - composite wood and agrifiber products)				
16	{MRc6} => {MRc5}	0.9091	0.2412	Rapidly renewable materials like cork or bamboo plywood may come from distant sources and may help achieve MRc5.
MRc6 (Rapidly renewable materials) MRc5 (Regional materials)				
20	{SSc4.1} => {SSc2}	0.8295	0.2238	Sites close to existing public transportation infrastructure tend to be in more densely developed areas, including previously developed areas. This help achieve SSc2.
SSc4.1 (Alternative transportation - public transportation access) SSc2 (Development density and community connectivity)				

23	{MRc6} => {MRc4}	0.9394	0.2203	Try to keep the project materials costs used in MRc6 to be consistent with those used in MRc4.
MRc6 (Rapidly renewable materials) MRc4 (Recycled content)				

Table 4. Top 10 Rules out of Total 2970 Rules.

Rank	Rules	Conf. Difference	Comments
1	{MRc12} => {MRc11}	0.8136	Easy Rule
2	{EAc2_high} => {EAc2}	0.7263	Ignored
3	{EAc2_high} => {EAc1_high}	0.4594	*Maybe Interesting*
4	{EAc2} => {EAc2_high}	0.4331	Ignored
5	{MRc12} => {SSc62}	0.4239	*Maybe Interesting*
6	{IEQc81} => {IEQc82}	0.3673	Easy Rule
7	{MRc3} => {MRc11}	0.3074	Suggested
8	{EAc2} => {EAc1_high}	0.3054	*Maybe Interesting*
9	{SSc51} => {SSc52}	0.2888	Easy Rule
10	{MRc11} => {SSc62}	0.2869	*Maybe Interesting*

The second rule is out of expectation because there is no obvious relationships between MRc1 (Building Reuse) and SSc6.2 (Stormwater Design – Quality Control). One possible hidden relationship may be that both MRc1 and SSc6.2 address the concept of "reuse". SSc6.2 emphasizes the reuse of rainwater while MRc1 focuses on the reuse of building materials. Therefore, it is reasonable that projects pursing MRc1 may be more interested in reusing materials or natural resources and thus pursuing SSc6.2.

CONCLUSIONS

Using data mining techniques, this study evaluated the strength of the credit rules suggested by USGBC in its official LEED Reference Guide for Green Building Design and Construction. The results show that some credit rules like {MRc7} => {IEQc4.4} not only reveal reasonable relationships in engineering aspects but also have a strong relationship in statistical analysis. On the other hand, some credit rules were seldom or never followed in LEED certified projects, although the credits in these rules are related conceptually.

Besides evaluating the USGBC suggested credit rules, this study also discovered new strong rules like {EAc2} => {EAc1_high} and {MRc1} => {SSc6.2}. In the future, these newly discovered rules will be evaluated one by one considering the potential hidden relationships.

ACKNOWLEDGEMENT

The authors would like to acknowledge the support by the Hong Kong Research Grants Council, Grant No. 611513. Any opinions and findings are those of the authors, and do not necessarily reflect the views of the Hong Kong Research Grants Council.

REFERENCE

Agrawal, Rakesh, Imieliński, Tomasz, & Swami, Arun. (1993). *Mining association rules between sets of items in large databases*. Paper presented at the Proceedings of the 1993 ACM SIGMOD international conference on Management of data, Washington, D.C., USA.

Borgelt, Christian. (2004). Finding association rules/hyperedges with the apriori algorithm: Working Group Neural Networks and Fuzzy Systems, Otto-von-Guericke-University of Magdeburg, Universitätsplatz.

Cheng, Jack CP, & Ma, Lucky J. (2013). A Study of the Relationship between Credits in the LEED-EB&OM Green Building Rating System. *IACSIT International Journal of Engineering and Technology*, *5*(4), 5.

Thomas, Benjamin J. (2008). *Mining Association Rules Between Credits in the Leadership in Energy and Environmental Design for New Construction (LEED-NC) Green Building Assessment System.* (Master), DTIC Document.

USGBC, US Green Building Council. (2008). LEED rating systems. *Retrieved September*, *11*, 2008.

USGBC, US Green Building Council. (2009). *LEED Reference Guide For Green Building Design And Construction*: U.S. Green Building Council.

USGBC, US Green Building Council. (2013). LEED for New Construction Projects Directory. Retrieved Aug 15, 2013, from http://www.usgbc.org/projects/new-construction