A Data Driven Approach for Smart Lighting

Hongyu Guo, Sylvain Letourneau, and Chunsheng Yang

National Research Council of Canada
1200 Montreal Road, Ottawa, Ontario
{hongyu.guo,sylvain.letourneau,chunsheng.yang}@nrc-cnrc.gc.ca

Abstract. Smart lighting for commercial buildings should consider both the overall energy usage and the occupants' individual lighting preferences. This paper describes a study of using data mining techniques to attain this goal. The lighting application embraces the concept of Office Hotelling, where employees are not assigned permanent office spaces, but instead a temporary workplace is selected for each check-in staff. Specifically, taking check-in workers' light requirements as inputs, a collective classification strategy was deployed, aiming at simultaneously predicting the dimming levels of the shared luminaries in an open office sharing light. This classification information, together with the energy usages for possible office plans, provides us with lighting scenarios that can both meet users' lighting comfort and save energy consumption. We compare our approach with four other commonly used lighting control strategies. Our experimental study shows that the developed learning model can generate lighting policies that not only maximize the occupants' lighting satisfaction, but also substantially improve energy savings. Importantly, our data driven method is able to create an optimal lighting scenario with execution time that is suitable for a real-time responding system.

1 Introduction

Smart lighting, which aims to improve on both the overall energy usage and the occupants' individual lighting comfort, has been identified as a potential market of 4.5 billion dollars in revenue by 2016[1]. Such smart lighting is of importance, not only for the "green" concept in terms of energy efficiency, but also for "personalized" office space.

Recent research has shown that buildings consume one-third of the total primary energy in the U.S., and of which, lighting, in particular, accounts for about 30% [9,10]. To cope with this increasing operational expenditure, modern lighting systems aim to be designed to minimize the energy consumption. Equally important, modern lighting also needs to take into account the occupants' lighting preferences. Studies have indicated that lighting comfort, for example, can dramatically impact workers' moods and thus productivity [8,13,14]. This is especially true under the context of Office Hotelling, where a company does not assign permanent office spaces for employees; instead it selects a temporary

[1] http://www.nanomarkets.net/

A. Moonis et al. (Eds.): IEA/AIE 2014, Part II, LNAI 8482, pp. 308–317, 2014.

workspace for each check-in staff. As a result, in addition to minimizing energy saving, introducing personalized lighting for occupants in commercial buildings is also of great importance [4].

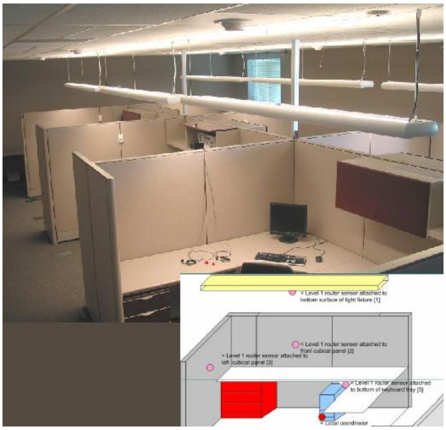

Fig. 1. The mock-up smart lighting office with six cubicles

This paper discusses a study of using a data driven approach to attain the above goals. Specifically, we apply recent data mining techniques to generate lighting scenarios for an open office sharing light, within the context of Office Hotelling. Figure 1 depicts the demonstration laboratory being set up for this application. This laboratory includes six cubicles, and sensors were installed in various positions of each cubicle in order to measure the environmental data such as temperature and light level. The sensor positions are shown at the bottom-right corner of Figure 1. The nine (9) shared lights are on the ceiling, and can be adjusted by either the computer in each cubicle or the center control system installed. The lighting policy generating unit here takes aim at creating lighting scenarios that not only minimize energy consumption but also satisfy users' light requirements, based on occupants' lighting preferences.

To generate a lighting scenario, the light requirements for the six desks are first obtained and used as inputs for the smart lighting system. Next, the dimming levels for the nine lights on the ceiling are determined by a machine learning classification model. By doing so, such classification information will be able to provide us with lighting scenarios that can both match users' preferred lighting and save energy consumption, provided that we have the energy usages for possible office plans. To this end, to obtain the various energy usages of potential office arrangements, we shuffle the workplaces of the employees, which is a practical approach within the office hotelling context where workers typically have different offices each time they check in. In this way, an energy saving lighting scenario, for instance, could be assigning closer offices to workers with similar lighting preferences. When compared with four other alternative lighting control strategies, our study shows that the developed data driven learning model can generate, within reasonable responding time, lighting policies that not only maximize the occupants' lighting satisfaction, but also substantially improve energy savings.

The rest of the paper is organized as follows. Section 2 outlines our prediction task and challenges. Next, in Section 3, we discuss our modeling approach. This is followed by an empirical study in Section 4. Finally, Section 5 concludes the paper and outlines our future research directions.

2 Data Mining Task and Challenge

2.1 Task Description

Constrained by the lighting preferences of the check-in employees in the office, the aim of the smart lighting system is to generate an energy saving lighting scenario. Also, the system has to be able to create the optimal lighting scenario in a reasonable execution time which is suitable for a real-time responding system.

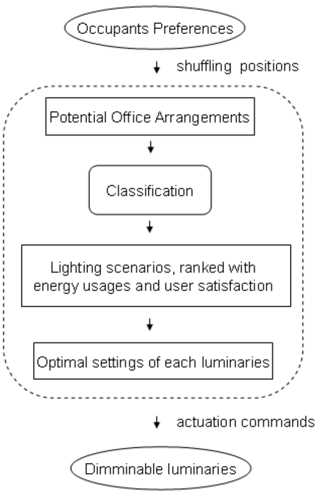

Fig. 2. Framework of the smart light control system

Figure 2 depicts the framework of our smart lighting control system. In detail, the input of this lighting system is the check-in occupants' light preferences. With such input, we can have 720 different positioning scenarios for the system. That is, we can assign an occupant to any one of the six desks in the room. In our approach, before initiating the classification system, we shuffle the positions of the individuals, and then use the shuffled scenarios as inputs. Subsequently, the classification system generates the output for each of the 720 positioning arrangements. In this way, we then can rank them in a specific order. Hence, the automatic system is able to choose the one that both saves energy and satisfies users' preferences. By doing so, for the checked-in occupants in the room, the classification system is able to generate multiple position and lighting arrangements, each with different preference-matched level and energy usage. As highlighted in Figure 2, the optimal one is then chosen to actuate the luminaries.

2.2 Modeling Setup

Practically, it is a common approach to use a Synthetic Imaging System for light design [1]. For instance, the RADIANCE system is often used by domain experts to simulate different lighting scenarios and to foresee the effects of tailored configurations. In our studies, the laboratory as depicted in Figure 1 was simulated using the RADIANCE system, where each of the configurable light sources has three dimming levels, i.e. Low, Median, and High. By doing so, the light on the ceiling of the room can be controlled by the RADIANCE system, and the illuminance values on each of the desks in the room can be accurately measured in Lux. Consequently, the data for the machine learning task was produced through this simulated environment. To this end, the collected data include all the combinations of the lighting levels on the ceiling, and the resulting measurements on the desks. In total, 19683 instances were collected, each is composed of six attributes, corresponding to the illuminances on the desks, as well as nine label sets, which reflect the nine luminaries' dimming levels.

2.3 Modeling Challenge

From a data mining perspective, this task can be mapped into a multi-target classification problem. That is, using six inputs $\boldsymbol{X}$ (i.e., occupants' light requirements on the desks) to predict the nine outputs $\boldsymbol{Y}$ (i.e., the dimming levels of the nine light sources on the ceiling): $\boldsymbol{Y} = f(\boldsymbol{X})$. It is worth noting that, in such applications, correctly classifying all target variables of an instance is required. As a result, one needs to consider a classifier has classified an instance correctly only if all target variables of that instance are correctly determined (i.e., "exact match"). In other words, the overall accuracy here refers to the "exact match" accuracy. Consequently, the main aim for such classification algorithm is to take aim at achieving higher "exact match" accuracy through learning a function f that maps X to Y.

To deal with such multi-target tasks, one straightforward approach is to learn a binary or multi-class classifier for each set of labels, and then each trained learner independently assigns a corresponding label for the test object. However, such an approach tends to result in poor predictive accuracy in terms of correctly classifying all labels simultaneously. This is because there is a large number of possible labels for each object to be classified, as discussed previously. For example, as observed in our experiments, a decision tree learning method [12] can achieve an average accuracy of 85% over nine (9) independent classifiers against this lighting application. Nevertheless, the predictive accuracy in terms of simultaneously predicting all correct labels for the nine label sets was only 22.57%. Unfortunately, as mentioned earlier, in such problems, simultaneously predicting the correct labels for all label sets is of importance. For example, imagine that we correctly predict eight out of the nine light sources, but misclassify one of them.

In such scenario, the resulting luminance values of the six desks will be very different from that of correctly classifying all nine light sources. This is because the misdetermined light source contributes its light to all the six desks. That type of incorrect determination will result in the dissatisfaction of all occupants in the open offices.

3 Modeling Methodology

To address the above mentioned modeling challenge, we deployed a state-of-the-art multi-target learning strategy, as presented by Guo and Létourneau in [5]. As reported in [5], when compared with several popular multi-target classification algorithms, the so-called Iterative Multi-target Classification (IAMC) approach can meaningfully enhance the "exact match" accuracy. Instinctively, the IAMC method benefits from being able to not only employ many accurate, mature single-target learning approaches to model each of the target attributes, but also utilize an iterative learning strategy to exploit the relationships among multiple related target attributes, thus achieving higher accuracy, when compared with other popular learning strategies for multiple targets problems [5].

Due to the IAMC method's superior predictive performance in terms of "exact match", we adapt this strategy for our smart lighting application. In particular, we significantly improve the IAMC method's predictive accuracy in our lighting application through integrating an ensemble strategy, namely the AdaBoost approach. Next, we will discuss the IAMC method and our extension in detail.

Algorithm 1. The Training of the IAMC Algorithm

Input: Object set with X attributes and Y labels, and a single target method f.
Output: Classification model $Y = f(X)$

1: **Training begins**
2: **for** each $y_i \in Y$ **do**
3: Build a model f_i^s using X only;
4: Apply AdaBoost strategy to model f_i^s; obtaining predictive accuracy $\epsilon^s(i)$;
5: **end for**
6: **for** each $y_i \in Y$ **do**
7: Build a model f_i^r using $X \cup (Y \backslash Y_i)$
8: Apply AdaBoost strategy to model y_i^r; obtaining predictive accuracy $\epsilon^r(i)$
9: **end for**

3.1 Iterative Approach For Multi-target Classification

The IAMC method includes two phases: training and inference. The training stage constructs two collections of single-target classifiers, while the inference stage aims at exploiting the relationships among target attributes through these constructed classifiers. Specifically, as depicted in Algorithm 1, the IAMC method firstly constructs two collections of classifiers: one utilizes the descriptive attributes only while the other is augmented with provided target attributes in the training data. Next, as descripted in Algorithm 2, these two collections of classifiers are used for iterative inference, as follows. The first collection is used to initiate the iterative process, where all values of the target attributes in the test data set are unknown. The second one is then deployed to continue the inference procedure until the process stops. In each iteration, the current target attribute estimates, resulting from the previous iteration, are used to enhance the learning models. The above iterative process repeats until all of the labels have stabilized or a pre-set number of iterations have been reached. As stated in Algorithm 2, the IAMC outputs the labels of the last iteration.

Note that a detailed description of the IAMC algorithm falls beyond the scope of this paper. Interested readers are referred to [5] for more discussions on this strategy.

Algorithm 2. The Joint Inference of the IAMC Algorithm

generate descending ordering O based on prediction improvement $\epsilon^r(i) - \epsilon^s(i)$
for each object t in the test set **do**
    obtain y_i using f_i^s;update y_i in the test set
end for
repeat
    for each object t in the test set **do**
        for each $y_i \in O$ **do**
            compute y_i using f_i^r; update y_i in the test set
        end for
    end for
until pre-set threshold number of iterations have elapsed or all labels have stabilized

3.2 Improving the IAMC Strategy

Recall from Algorithm 1 that, the IAMC method requires a single-target learning method as input. Our studies show that the accuracy of this single-target learning method has a significant impact on the overall "exact match" accuracy of the IAMC approach. Basing on this observation, we meaningfully improve the predictive accuracy of the IAMC method in our application with a boosting strategy, as will be discussed next.

In a nutshell, the IAMC method falls in the learning framework of collective classification. As pointed out by Neville and Jensen in [7], one of the necessary conditions for the success of collective classification is that the system must be able to make some initial inferences accurately. Following this thought, we intend to improve the individual classifiers' prediction before initiating the collective inference procedure. Particularly, we look into the AdaBoost ensemble method, which have been proven to be able to meaningfully improve predictive accuracy of a high error classifier, namely the so-called weak classifier [3]. The underlying principle of the Boosting strategy states that by using a weak learning algorithm several times on a sequence of carefully constructed training examples, the weak learning algorithm can be converted into an algorithm with a predictive performance that surpasses the original weak algorithm [11]. While learning, the Boosting algorithm first focuses on the production of a series of dependent classifiers, in which each classifier is better able to predict hard examples for which the previous classifier performance was poor [11]. The outputs of these classifiers are then combined using weighted voting in the final prediction of the model.

In the implementation of our lighting application, the single-target learning method in the IAMC strategy as depicted in Algorithms 1 and 2 are replaced by a Adaboost ensemble. That is, the f function there is replaced by a H function as defined as following.

$$H(x) = sign(\sum_{m=0}^{M} \alpha_m h_m(x))$$

Here, h_m is the m-th classifier of the series of M dependent classifiers in the boosting ensemble, and α_m represents the corresponding weight of the classifier h_m.

Promisingly, our experimental studies, as will be presented in the next section, suggest that boosting the performance of individual learners before the collective inferences can significantly improve the collective classification's prediction as measured by the "exact match" metric. Therefore, in our lighting application, we apply the AdaBoost algorithm [3] to boost the single-target learners' accuracy before deploying them for the IAMC strategy.

4 Experiments

4.1 Predictive Accuracy Achieved

In this experiment, we present our evaluations on using the C4.5 decision trees [12] and the Artificial Neural Networks [2] as the single-label learning methods of the IAMC approach. The C4.5 decision tree learner was used due to its de facto standard for empirical comparisons. Also, Artificial Neural Networks were chosen because they have proven to be surprisingly successful in many real-world knowledge discovery applications [6]. Each of these experiments produces results using 10-fold cross validation. In addition, the number of iterations for the collective inferences was heuristically set to 20 for each experiment.

We compared the predictive accuracy, in terms of simultaneously predicting all correct labels, obtained by the three approaches, namely 1) the straightforward approach that learns a classifier for each set of labels, and then each trained learner independently assigns a corresponding label for the test object (noted as the Intrinsic model), 2) the collective classification strategy IAMC, and 3) the collective classification strategy with the AdaBoost method applied (noted as BoostIAMC). We presented the accuracies obtained by the three approaches with decision trees and neural networks as single-label learning methods in Tables 2 and 1, respectively. In these two tables, we also described the performance improvement of both the IAMC algorithm and the BoostIAMC method over that of the Intrinsic strategy. The statistic significance of these results was examined using a paired t-test.

Table 1. Accuracy obtained by the Intrinsic, IAMC, and BoostIAMC methods using decision trees as the single-label learning method, along with the prediction improvement of both the IAMC and BoostIAMC strategies over that of the Intrinsic approach ($p < 0.001$ in the paired t-test)

	Predict all lights simultaneously	Accuracy Improvement
Intrinsic Model	22.57%	
IAMC	41.37%	18.80%
BoostIAMC	68.33%	45.76%

Table 2. Accuracy obtained by the Intrinsic, IAMC, and BoostIAMC methods using neural networks as the single-label learning method, along with the prediction improvement of both the IAMC and BoostIAMC strategies over that of the Intrinsic approach ($p < 0.001$ in the paired t-test)

	Predict all lights simultaneously	Accuracy Improvement
Intrinsic Model	29.37%	
IAMC	64.23%	34.86%
BoostIAMC	84.86%	55.49 %

Results as shown in Tables 2 and 1 indicate that the BoostIAMC approach can statistically and significantly increase the predictive accuracy of the Intrinsic models in terms of simultaneously predicting all of the correct labels. That is, the experimental results suggest that the AdaBoost strategy and the collective inference technique were successfully employed in this lighting application. For example, when decision trees were applied as single-label learners, the accuracy obtained by the Intrinsic method was very low in terms of simultaneously predicting all of the correct labels. The accuracy was only 22.57%. In this case, the collective inference process increased its accuracy to 41.37%. Furthermore, this prediction was improved by applying the AdaBoost approach before the collective classification. As a result, the final predictive accuracy achieved by the BoostIAMC algorithm was 68.33%.

When considering deploying the neural network algorithm, the accuracy for the Intrinsic and IAMC methods were 29.37% and 64.23%, respectively. Significantly, this accuracy was improved to reach as high as 84.86% when the AdaBoost algorithm was employed before the collective inferences, as achieved by the BoostIAMC approach. Results from these two tables have also shown that these prediction improvements were statistically significant. The results indicate that the p-values achieved by the paired t-test were less than 0.001.

These results indicate that the BoostIAMC model with neural networks applied was very promising. The final accuracy of this model against all luminaries was 84.86%, which was more than 16% higher than that of applying decision trees as the single-label learning methods. Thus, in our smart lighting application as described in Section 2.3, we deployed a BoostIAMC strategy with neural network algorithms as its single-target learning methods for our classification task.

4.2 Comparison with Alternative Control Strategies

In this section, we compare our classification system with four other lighting control strategies which are commonly used in the lighting domain, namely all-on, all-off, half-on, and exhaustive search. The all-on model refers to a configuration of the office where all of the luminaries are wired to be turned on all together. On the contrary, the all-off model will turn off all of the lights in the room at the same time. In contrast, the half-on model sets each light to its Median dimming level. When the exhaustive search model is applied, the system will exhaustively search all of the possible lighting scenarios in the light data set and then choose the best-matched one as its output.

We present the comparison results in Table 3, where the energy consumption, average discrepancy, and response time are described for each strategy. We calculated the discrepancy of a lighting scenario by comparing each desk's illuminance value, generated by the lighting strategy, with its true value obtained. The measurement of the discrepancy was computed using the following Euclidean Distance function:

$$Discrepancy = \sqrt{\sum_{i=1}^{n}(p_i - q_i)^2}$$

Here p_i is the illuminance value of the i^{th} desk in the room, resulting from the lighting scenario generated by the lighting strategy; q_i is the illuminance value of the i^{th} desk measured, given the current lighting scenario.

Table 3. Energy consumption required and satisfaction discrepancy measured, along with the execution time needed, to generate a lighting scenario

	Energy Consump.	Average Discrep.	Response Time (sec.)
All ON Mod.	100%	1454.44	0
ALL OFF Mod.	0%	1460.26	0
Half ON Mod.	50%	490.13	0.0
Exhaustive Search	50%	0	20
Data Mining Mod.	50.04%	4.93	2

The results, as shown in Table 3, indicate that the all-on, all-off, and half-on models produced a much larger gap between the occupants' light preferences and the ones they would receive if such a lighting system was employed than that of the exhaustive search and data mining models. For example, the average discrepancy was less than 5 for the two latter strategies, compared to that of over 490 for the former two approaches.

When comparing the exhaustive search with the data mining models, the results suggest that the response time required for the exhaustive search model was large. It required 20 seconds to generate a lighting scenario, compared to that of only 2 seconds needed from the data mining strategy. These observations imply that the exhaustive search approach is not suitable for a real-time responding environment.

5 Conclusions

Smart lighting needs to take into consideration both the overall energy usage and the occupants' lighting comfort. This paper describes a data-mining approach to attain this goal. Specifically, taking check-in workers' light requirements as inputs, a collective classification strategy was deployed to simultaneously predict the dimming levels of the shared luminaries. This classification information, together with the energy usage for potential office plans, provides us with lighting scenarios that take into account both the users' preferred lighting and energy consumption. We evaluate our method against four other lighting control strategies. Our study shows that our method can generate lighting policies that both maximize the occupants' satisfaction and improve the overall energy savings.

Our future work will include the dynamic lighting settings, such as people entering and leaving the room.

References

1. RADIANCE Synthetic Imaging System, Lawrence berkeley national laboratory, Berkeley (2006), `http://radsite.lbl.gov/radiance/` (retrived March 2009)
2. Duda, R., Hart, P., Stork, D.: Pattern Classification. Wiley-Interscience (2000)
3. Freund, Y., Schapire, R.E.: Experiments with a new boosting algorithm. In: ICML 1996, pp. 148–156 (1996)
4. Gu, Y.: The Impacts of Real-time Knowledge Based Personal Lighting Control on Energy Consumption, User Satisfaction and Task Performance in Offices. PhD thesis, Carnegie Mellon University (2011)
5. Guo, H., Létourneau, S.: Iterative classification for multiple target attributes. J. Intell. Inf. Syst. 40(2), 283–305 (2013)
6. Mitchell, M.T.: Machine Learning. McGraw Hill, New York (1996)
7. Neville, J., Jensen, D.: Iterative classification in relational data. In: AAAI Workshop on Learning Statistical Models from Relational Data, 13-20 (2000)
8. Newsham, G., Veitch, J.: Individual control over office lighting: Perceptions, choices and energy savings. Construction Technology Updates (1998)
9. Online. The Interlaboratory Working Group on Energy-Efficient and Clean-Energy. In: Scenarios for a Clean Energy Future: Interlaboratory Working Group on Energy-Efficient and Clean-Energy Technologies (2000), `http://www.nrel.gov/docs/fy01osti/29379.pdf`
10. Online. U.S Department of Energy, Building Technology Program. Energy Solution for Your Building (2000), `http://www.eere.energy.gov/buildings/info/office/index.html`
11. Opitz, D., Maclin, R.: Popular ensemble methods: An empirical study. Journal of Artificial Intelligence Research 11, 169–198 (1999)
12. Quinlan, J.R.: C4.5: programs for machine learning. Morgan Kaufmann Publishers Inc., USA (1993)
13. Wen, Y.-J., Agogino, A.: Control of wireless-networked lighting in open-plan offices. Lighting Research and Technology 43, 235–248 (2011)
14. Wen, Y.-J., Bonnell, J., Agogino, A.M.: Energy conservation utilizing wireless dimmable lighting control in a shared-space office. In: Proceedings of the 2008 Annual Conference of the Illuminating Engineering Society (2008)