Statistical Inference for Intelligent Lighting: A Pilot Study

Aravind Kota Gopalakrishna, Tanir Ozcelebi, Antonio Liotta, and Johan J. Lukkien

Abstract. The decision process in the design and implementation of intelligent lighting applications benefits from insights about the data collected and a deep understanding of the relations among its variables. Data analysis using machine learning allows discovery of knowledge for predictive purposes. In this paper, we analyze a dataset collected on a pilot intelligent lighting application (the *breakout* dataset) using a supervised machine learning based approach. The performance of the learning algorithms is evaluated using two metrics: *Classification Accuracy* (CA) and *Relevance Score* (RS). We find that the breakout dataset has a predominant *one-to-many* relationship, i.e. a given input may have more than one possible output and that RS is an appropriate metric as opposed to the commonly used CA.

1 Introduction

Many intelligent applications such as the next-generation networked systems [10] involve collecting data from different sources, organizing them and then perform data analysis. Data analysis here refers to the process of acquiring information from the collected data and making conclusions and decisions based on these that are useful for the application. Data analysis depends on the type of the data collected, either quantitative or qualitative [1]. Quantitative data can be analyzed using statistical operations such as frequency distributions, central tendency (using mean,

Aravind Kota Gopalakrishna · Tanir Ozcelebi · Johan J. Lukkien
System Architecture and Networking (SAN), Department of Mathematics and Computer Science, Intelligent Lighting Institute (ILI), Eindhoven University of Technology, The Netherlands
e-mail: {a.kota.gopalakrishna,t.ozcelebi,j.j.lukkien}@tue.nl

Antonio Liotta
Electro-Optical Communications (ECO), Department of Electrical Engineering, Eindhoven University of Technology, The Netherlands
e-mail: a.liotta@tue.nl

D. Camacho et al. (eds.), *Intelligent Distributed Computing VIII*,
Studies in Computational Intelligence 570, DOI: 10.1007/978-3-319-10422-5_3

median and mode), correlation and regression. Analyzing qualitative data involves examining the data collected through surveys, interviews and observations. Data analysis helps to understand the relationship between input and output, thereby enabling to make decisions upon designing and implementing desired applications. In this paper, we analyze the data collected for the purpose of developing an intelligent lighting application, where users are dynamically presented with the light settings of their preferences in various contexts.

Intelligent lighting [6] is an application that makes use of contextual information such as user identity, type of activity, influence of external light, time of the day and more to provide a suitable lighting to its users. The pilot setup for intelligent lighting is a particular part of an office space, known as the *breakout* area. A breakout area is an area where office employees can have informal meetings or some time for personal retreat. The breakout area implementation for intelligent lighting [6] contains numerous connected lighting elements and sensors. The challenge here is to develop an intelligent application that learns from the data collected and uses this knowledge in the future to predict a suitable light setting for a given scenario. To design such a system, it is necessary to understand how the input parameters and output light settings are related. Machine learning is a data analysis technique that can be used to discover knowledge from the data for predictive purposes such as intelligent lighting.

In this paper, we investigate the nature of the breakout dataset and present the insights gained into the properties that enables to make better design decisions towards implementing intelligent lighting. The pilot setup of the breakout area is discussed. Subsequently, the details of the breakout dataset and how the data are collected and organized into the breakout dataset are explained. The dataset is processed using supervised machine learning algorithms. The prediction performance is evaluated using two metrics: *Classification Accuracy* (CA) and *Relevance Score* (RS) [7]. CA measures how precise the light prediction is, given a certain environment state. In contrast, RS gives the measure of how relevant a light prediction is for a given state of the environment. Analysis of the results show that the breakout dataset has one-to-many relationships, i.e. a given input (i.e. a context or a state of the environment) may have many possible output light settings. Furthermore, when it comes to intelligent lighting applications, RS is more appropriate than the common performance metric CA.

The paper is organized as follows. In Section 2, the pilot setup of the breakout area, the description of the breakout dataset such as the number of samples, output class distribution and user-sample distribution, and the means in which data are collected and processed are discussed. The experiments performed and the insights from the results are discussed in Section 3. The paper is concluded in Section 4.

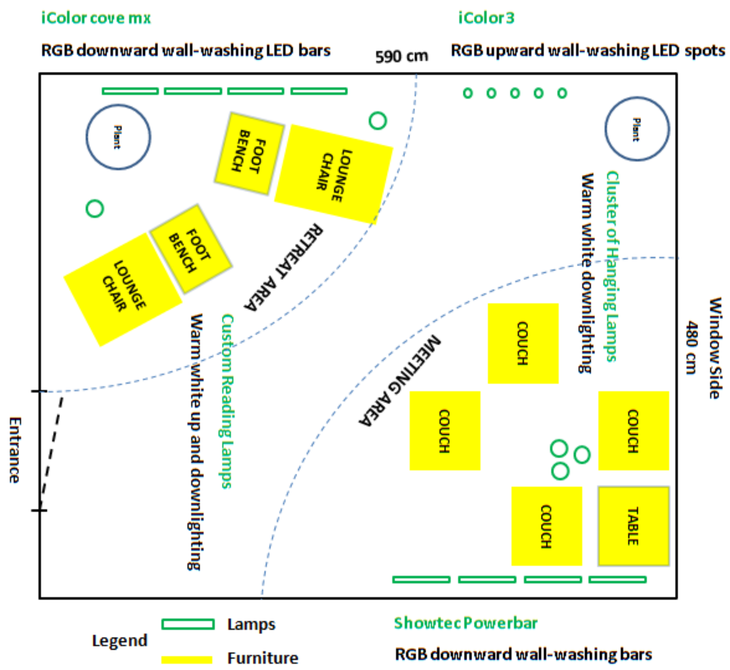

Fig. 1 Floor plan of the breakout area

2 Pilot Setup and Data Collection

In this section, we discuss the pilot implementation setup of the breakout area where the data have been collected, provide a description of the breakout dataset and explain process in which data have been collected.

2.1 Pilot Setup

Figure 1 shows the floor plan of the breakout area. Opposite to the entrance is a wall with windows and blinds, which allows for controlling the external light influence. Furthermore, the area is divided into two spaces dedicated to different purposes: meeting area for informal meetings and the retreat area for personal retreat and relaxation. However, the users of the breakout area are not restricted to use a specific area for a specific activity. For example, user A may choose to use either the meeting area or the retreat area for *relaxation*. Given an intelligent lighting application, the desired light settings in an area may depend on user identity, type of activity in the area and external light influence (sunlight), an area the user may choose for the activity and many more features. The lighting system in the breakout area

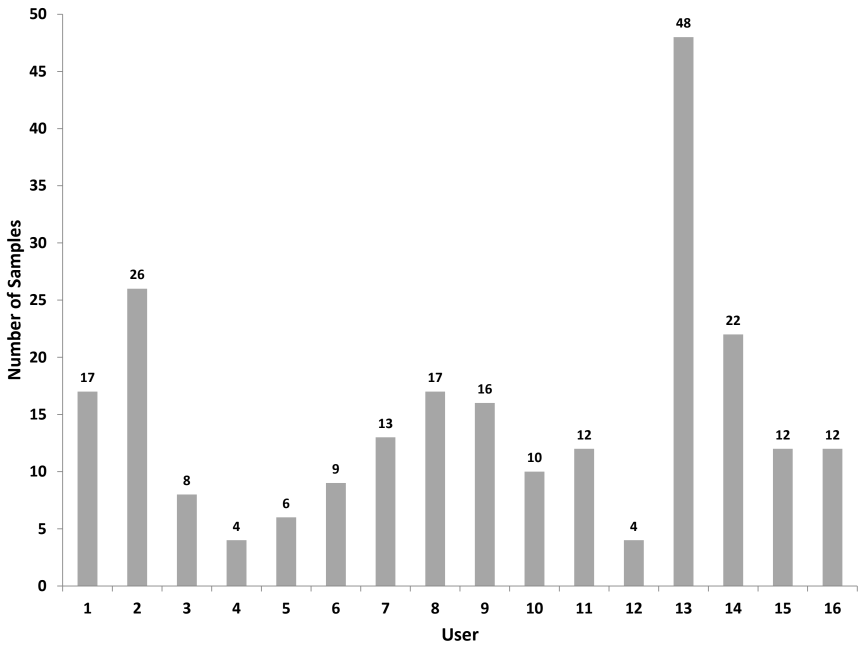

Fig. 2 The distribution of samples in the breakout dataset

contains two types of lights: colored wall-washing lights for creating an atmosphere and white *down-lights* [12] for illuminating areas of user tasks. The sensor system for monitoring the breakout area contains Passive Infra Red (PIR) sensors for monitoring movements, sound pressure sensors for monitoring sound volume intensity and light sensors for measuring external light influence.

Table 1 List of input features considered, that influences user's choice of light selection

	Feature	Type of the feature	Possible Values
1.	User-Identity (UID)	Categorical	U1, U2, U3, U4, ...
2.	Type of Activity (ToA)	Categorical	Active_Group, Active_Alone, Relax_Group, Relax_Alone
3.	Area of Activity (AoA)	Categorical	Meeting, Retreat
4.	Intensity of Activity (IoA) in the other subarea	Categorical	$0, 1, 2, \ldots, 10$
5.	Time of the Day (ToD)	Numeric	$\in [0, 24)$, e.g. 10.5 for 10:30am
6.	External Light Influence (ExLI)	Categorical	VeryHigh, High, Low, VeryLow

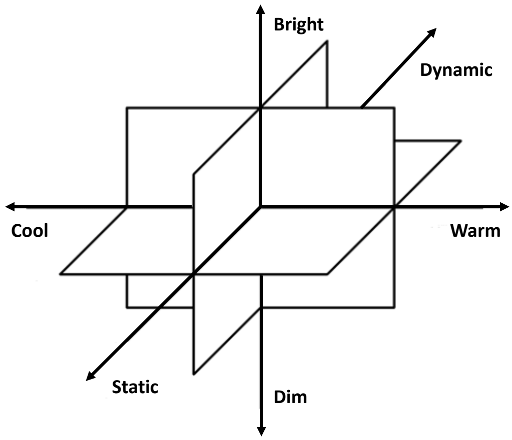

Fig. 3 Possible Output Light Combinations

2.2 Description of the Breakout Dataset

The breakout dataset for intelligent lighting consists of 236 samples collected as discussed in Section 2.3. The dataset does not have any missing data. We select six input features that may influence a user's choice in selecting one of the pre-defined light settings for a given context as summarized in Table 1. Figure 2 shows the number of users and the numbers of data samples collected per user. We consider eight output light settings to support users' activities as shown in Fig. 3. The class distribution of these eight light settings over the 236 samples is presented in Fig. 4.

2.3 Data Collection

Among the mentioned features in Table 1, AoA, ToD, IoA and ExLi are gathered implicitly from the breakout area through sensor monitoring. The features UID and ToA are acquired explicitly from the users via the breakout application installed on their smart phones.

The data samples for the breakout dataset were collected using two methods. In the first method, we created various contexts in the breakout area with different ExLI, IoU values, in which the participants were asked to select a light setting that they prefer for the activities listed in Table 1. In the second method, the participants were allowed to use the breakout area on-demand for six weeks. During this six-week period, all interactions of the users with the system (i.e. activities and selected light settings) as well as the sensor readings were logged.

In order to learn users' preferences of light settings in a particular context, collected data samples should contain the values for features UID and ToA. On the

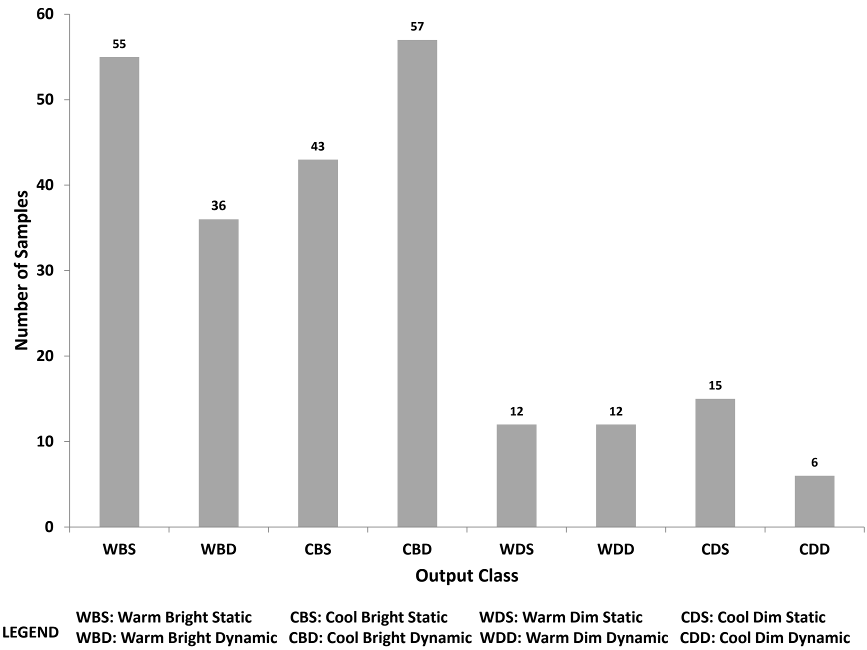

Fig. 4 The distribution of output class in the breakout dataset

other hand, when there are no users in the breakout area, then the data samples collected contain no entries for these features. The breakout dataset is then obtained by performing data cleaning in two steps. Firstly, those samples that do not contain feature values for UID and ToA are filtered out. Secondly, those samples that belong to users' free explorations of different light settings for a particular context are removed.

3 Experiments and Discussion

We use supervised learning algorithms to analyze the breakout dataset by investigating the prediction performance using two different metrics: *Classification Accuracy* (CA) and *Relevance Score* (RS). The following six rule-based prediction models in WEKA [8] are considered: DecisionTable [9], JRip [3], Nearest Neighbor with generalization (NNge) [11], PART [4], ConjunctiveRule [2] and Ridor [5]. The prediction performance of the prediction models on the breakout dataset are computed using 10-fold cross-validation.

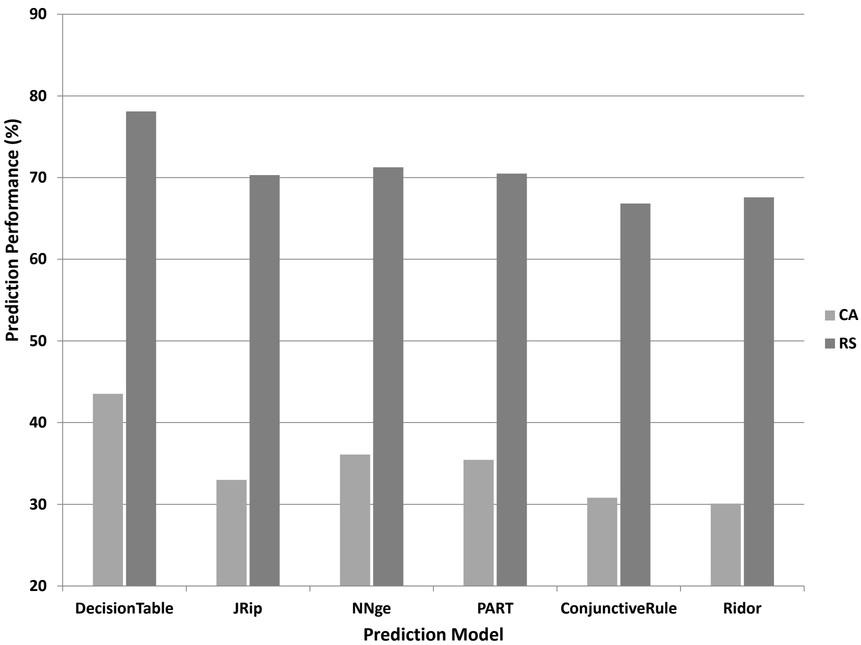

Fig. 5 Prediction performance vs. Prediction model for six rule-based prediction models with 8-output light settings

3.1 *Analysis of Prediction Performance with 8-Outputs*

Figure 5 presents the prediction performance of the six rule based classification models considering CA and RS as metrics. It can be seen that CA values are very low for all the considered prediction models, compared to RS values. This is because the CA metric measures how accurate the prediction is for a sample, i.e. the predicted outcome is compared to the actual outcome. If the predicted and actual outcomes do not match, then the CA metric scores a zero. Since users are not consistent in selecting a particular light setting for a given context, the average CA for a lighting application is typically low. The inconsistency comes from the fact that it is very difficult (indeed impossible) to consider the full set of input features (context) that determine a user's light setting choice. Furthermore, some contextual information, such as a user's mood, can not be monitored easily. Instead, a learning algorithm takes only a part of all relevant input features (i.e. an observed context) into account. Since, multiple light settings can satisfy a user in a given observed context, the nature of the breakout dataset i.e. the input-output relationship is *one-to-many*. The RS metric measures how relevant the predicted outcome is, for a given context based on the information computed from the dataset [7]. The RS metric does not score a zero when there is a mismatch between the predicted and actual outcome and thus gives higher performance.

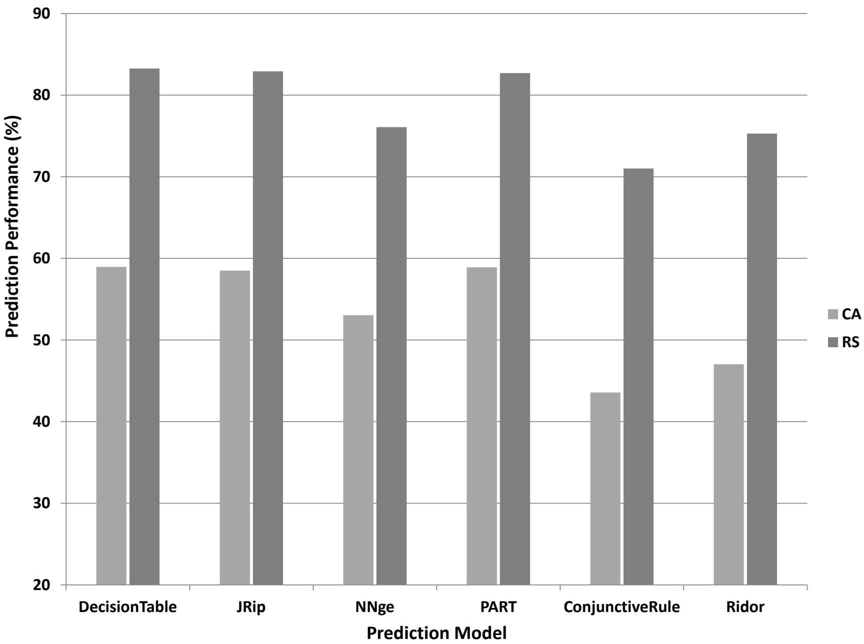

Fig. 6 Prediction performance vs. Prediction model for six rule-based prediction models with 4-output light settings

3.2 *Analysis of Prediction Performance with 4-Outputs*

In this experiment, static and dynamic light settings are combined into 4-output classes. This is done because the users could not differentiate much between the static and dynamic settings [12]. Figure 6 presents the performance values of the six rule based classification models considering CA and RS as metrics. It can be seen that since the output space is reduced, the CA values improve as compared to the results of the 8-output dataset. However, the RS values do not improve much.

3.3 *On Implementing Intelligent Lighting*

Table 2 shows the standard deviation of the prediction performance for the six prediction models computed using 10-fold cross-validation. The values show that there is a high degree of inconsistency in the prediction performance for both the metrics considered. This means that the performance of the prediction models that use supervised learning approach varies significantly with different training and test sets.

From this study, we find that the use of supervised learning algorithms for implementing intelligent lighting with a metric such as CA is inappropriate considering the nature of the breakout dataset. The RS metric is better as it evaluates the prediction performance from a different perspective. Furthermore, the selection of the learning approach to implement intelligent lighting depends on the objectives to be achieved such as whether the system should learn and adapt continuously or not. Supervised learning algorithms are trained on a fixed dataset. These algorithms does not adapt as the input-output relationships change in time due to dynamic factors such as changing user preferences and changing lengths of daytime in different seasons. Therefore, it is necessary to explore other learning techniques such as online learning and analyze their prediction performance.

Table 2 Standard deviation of the prediction performance for the six rule-based prediction models

Prediction Model	Std Dev (CA)	Std Dev (RS)
Decision Table	7.78	5.45
JRip	7.80	4.74
NNge	8.55	5.90
ConjunctiveRule	8.48	5.23
PART	6.88	7.77
Ridor	11.38	8.08

4 Conclusion

Data analysis of an intelligent lighting application leads to insights into the relations among various input features as well as suitability of different performance metrics and performance limitations. In designing such applications, such insights help in deciding upon a certain sensor modality and learning algorithm. By means of statistical analysis of a dataset collected from a pilot implementation named the breakout area, we were able to infer that the breakout dataset has a *one-to-many* input-output relationship unlike most available real-world datasets. This means that more than one output may be satisfying for a given input context. The experiments were performed using six rule-based prediction models and two performance evaluation metrics: Classification Accuracy (CA) and Relevance Score (RS). We find that the CA is not an appropriate metric for applications such as intelligent lighting having *one-to-many* input-output relationship and that the RS is most appropriate performance metric. As a future work, we will investigate other learning techniques such as online learning and reinforcement learning and analyze their prediction performance.

Acknowledgements. This work has been supported under the Smart Context Aware Services project (SmaCS) through the Point One grant No. 10012072. We would also like to thank Serge Offermans, Intelligent Lighting Institute (ILI), Eindhoven University of Technology for the pilot implementation of the *Breakout* area.

References

1. Analyzing and Interpreting Data. Technical report, Wilder Research (2009)
2. Clarke, P., Niblett, T.: The CN2 Rule Induction Algorithm. In: Machine Learning, pp. 261–283 (1989)
3. Cohen, W.H.: Fast Effective Rule Induction. In: Proceedings of the Twelfth International Conference on Machine Learning, pp. 115–123. Morgan Kaufmann (1995)
4. Frank, E., Witten, I.H.: Generating Accurate Rule Sets Without Global Optimization. In: Proceedings of the Fifteenth International Conference on Machine Learning, pp. 144–151 (1998)
5. Gaines, B.R., Compton, P.: Induction of ripple-down rules applied to modeling large databases. Journal of Intelligent Information Systems 5(3), 211–228 (1995)
6. Gopalakrishna, A.K., Ozcelebi, T., Liotta, A., Lukkien, J.J.: Exploiting Machine Learning for Intelligent Room Lighting Applications. In: Proceedings of the 6th IEEE International Conference on Intelligent Systems, pp. 406–411 (2012)
7. Gopalakrishna, A.K., Ozcelebi, T., Liotta, A., Lukkien, J.J.: Relevance as a Metric for Evaluating Machine Learning Algorithms. In: Perner, P. (ed.) MLDM 2013. LNCS (LNAI), vol. 7988, pp. 195–208. Springer, Heidelberg (2013)
8. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The WEKA Data Mining Software: An Update. SIGKDD Explorations 11(1) (2009)
9. Kohavi, R.: The Power of Decision Tables. In: Lavrač, N., Wrobel, S. (eds.) ECML 1995. LNCS, vol. 912, pp. 174–189. Springer, Heidelberg (1995)
10. Liotta, A.: The Cognitive Net is Coming. IEEE Spectrum 50(8), 26–31 (2013)
11. Martin, B.: Instance-based Learning: Nearest Neighbor with Generalization. Technical Report, University of Waikato (1995)
12. Offermans, S., van Essen, H., Eggen, B.: Exploring a Hybrid Control Approach for enhanced User Experience of Interactive Lighting. In: Proceedings of the 27th International BCS Human Computer Interaction Conference, pp. 1–9 (2013)