Occupancy Detection: A Data Mining Approach

Dhaifallah Alghamdi

Abstract- Lighting and cooling places cost money and businesses pay a lot on providing energy to vacant areas. Having an accurate model that detects occupancy and helps switch on or off light and other resources would result in energy saving and cost reduction. This paper studies data containing a number of attributes and uses data mining techniques to have a precise model that enables smart sensors to classify correctly and react accordingly. It applies a number of data mining algorithms and then selects the best model among them.

Index Terms— Occupancy-detection, Data Mining, KNN, Naïve Bays, Support Vector Machine (SVM) and Adaptive Boosting.

1 INTRODUCTION:

In this paper, sophisticated Data Mining techniques were implemented to achieve a higher overall classification accuracy. Occupancy detection is essential in energy saving and recent studies show that accurate detection can save up to 42% of energy. Having a system that can detect physical presence without a visual recording instrument is interesting. Nowadays, the advancement of digital sensors is a big motive to come up with an extremely accurate computational model that enables occupancy recognition. This research presents a strong computational model that can be used in the advanced sensors in order to save energy and reduce costs related to it.

2.2 Data Collection

The data was obtained through the University of California, Irvine. It was found on the online machine learning repository which contains a lot of datasets available for researchers to study and present their findings.

3. ANALYSIS

The data consists of five attributes which are time, temperature, light, CO2 measurements and humidity rate. The output of the data is binary which represents occupancy status.

3.2 Algorithms Used

Four complex data mining algorithm were used to find the best classification model with the maximum accuracy.

3.2.1 Mathematical Expressions

Following are brief description of each technique:

Naive Bayes

It is a supervised classification algorithm that is based on the original theorem which is called Bayesian theorem. It basically assumes independence between features regardless of any possible correlation.

Support Vector Machine (SVM)

It is a supervised learning model which can be used for classification or regression.

K-nearest Neighbor

In the last century, Fix and Hodges (Silverman & Jones, 1989) has made a contribution to knowledge discovery by coming up with a non-parametric classification algorithm which is called K-nearest neighbor (KNN). It compares the data point to its k closest neighbors and based on that the output value is determined.

Adaptive Boosting

This algorithm has made a contribution to knowledge discovery in 2003. It is also called AdaBoost for short and is based on the concept of creating an accurate classification by combining weak rules. This algorithm, assigns equal weights to all inputs at the initiation of the machine learning process. Then, it learns gradually and changes the weights to lead ultimately to the accurate model.

3.2.2 Results of Each Algorithm

3.2.2.1 Naive Bayes

Using Naïve Bayes, the following classification errors were found:

Table 1: Classification Error of Naïve Bayes

Data File	Overall Classification Error	Average Error
1	0.04	0.0525
2	0.04	
3	0.06	
4	0.07	

As shown above the average error of this algorithm is approximately 6%.

3.2.2.2 SVM

Using SVM Algorithm, the following classification errors were found:

Table 2: SVM Classification Error

Data File	Overall Classification Error	Average Error
1	0.071	0.0585
2	0.066	
3	0.056	
4	0.041	

As shown above the average error of this algorithm is approximately 6%.

3.2.2.3 K-nearest neighbor

Using K-nearest neighbor the following classification errors were found:

Table 3: KNN's Classification Error

Data File	Overall Classification Error	Average Error
1	0.005	0.005
2	0.005	
3	0.006	
4	0.004	

KNN model average error is less than 1%. It shows that KNN is considered as one of the best classifiers in data mining.

3.2.2.4 Adaptive Boosting Algorithm

• *Author Dhiafallah Alghamdi is currently pursuing a master's degree program in Industrial engineering in University of Illinois at Chicago, US, PH +13312501561, E-mail: dhaif27@gmail.com*

Using Adaptive Boosting Algorithm, the following classification errors were found:

Table 4: Classification Error of Adaptive Boosting

Data File	Overall Classification Error	Average Error
1	0.005	0.0035
2	0.002	
3	0.003	
4	0.004	

AdaBoosting algorithm yields the lowest average error among the four algorithms of less than 0.5%.

4. Conclusions

In this paper, a number of sophisticated data mining models were applied to the targeted data.

Algorithm	Overall Classification Accuracy	Average Error
Naïve Bayes	94.75	0.005
SVM	94.15	
K-NN	99.5	
Adaboosting	99.65	

Naïve Bayes and SVM resulted in an accuracy of around 94%. The accuracy attained wasn't satisfied enough so K-NN and Adaboosting algorithms were applied on the same data which yielded more than 99% average accuracy with a slight preference to Adaboosting. The comparison that took place in this paper, helps select the best model which is the one related to Adaboosting and use it in the artificial intelligence part in order to come up with an advanced technology that enables occupancy detection at an accurate level.

Ultimately, the purpose of this paper is to have precise occupancy detection that helps reduce business cost and save energy. What was achieved in this study can be taken further to reach the ultimate goal.

References:

Jon Hutchins, A. I. (NA). MODELING COUNT DATA FROM MULTIPLE SENSORS: A BUILDING OCCUPANCY MODEL. *Department of Computer Science University of California, Irvine CA 92697–3425* , 4.

Luis M. Candanedo, V. F. (2016). Accurate occupancy detection of an office room from light, temperature, humidity and CO2 measurements using statistical learning models. *Energy and Buildings* , 12.

Zhang, H. (NA). The Optimality of Naive Bayes. *Computer Science University of New Brunswick* , 6.