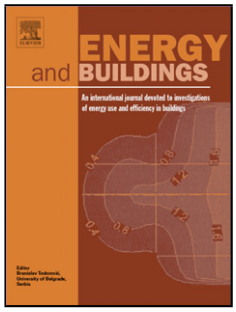

Title: Estimating Occupancy In Heterogeneous Sensor Environment

Author: Manar Amayri Abhay Arora Stephane Ploix Sanghamitra Bandhyopadyay Quoc-Dung Ngo Venkata Ramana Badarla

PII: S0378-7788(16)30622-3
DOI: http://dx.doi.org/doi:10.1016/j.enbuild.2016.07.026
Reference: ENB 6857

To appear in: *ENB*

Received date: 27-11-2015
Revised date: 8-7-2016
Accepted date: 12-7-2016

Please cite this article as: Manar Amayri, Abhay Arora, Stephane Ploix, Sanghamitra Bandhyopadyay, Quoc-Dung Ngo, Venkata Ramana Badarla, Estimating Occupancy In Heterogeneous Sensor Environment, *<![CDATA[Energy & Buildings]]>* (2016), http://dx.doi.org/10.1016/j.enbuild.2016.07.026

Estimating Occupancy In Heterogeneous Sensor Environment

Manar Amayri[a], Abhay Arora[b], Stephane Ploix[a], Sanghamitra Bandhyopadyay[c], Quoc-Dung Ngo[d], Venkata Ramana Badarla[b]

[a]*G-SCOP lab / Grenoble Institute of Technology, Grenoble, France*
[b]*Indian Institute of Technology, Jodhpur, India*
[c]*Indian Statistical Institute, Kolkota, India*
[d]*Posts and Telecommunications Institute of Technology, Hanoi, Vietnam*

Abstract

A general approach is proposed to determine the common sensors that shall be used to estimate and classify the approximate number of people (within a range) in a room. The range is dynamic and depends on the maximum occupancy met in a training data set for instance. Means to estimate occupancy include motion detection, power consumption, CO_2 concentration sensors, microphone or door/window positions. The proposed approach is inspired by machine learning. It starts by determining the most useful measurements in calculating information gains. Then, estimation algorithms are proposed: they rely on decision tree learning algorithms because these yield decision rules readable by humans, which correspond to nested if-then-else rules, where thresholds can be adjusted depending on the living areas considered. In addition, the decision tree depth is limited in order to simplify the analysis of the tree rules. Finally, an economic analysis is carried out to evaluate the cost and the most relevant sensor sets, with cost and accuracy comparison for the estimation of occupancy. C45 and random forest algorithms have been applied to an office setting, with average estimation error of 0.19-0.18. Over-fitting issues and best sensor sets are discussed.

Keywords: human behavior, building performance, activities recognition, office buildings, machine leaning, data mining.

1. Introduction

Recently, research about building has turned to focusing on occupant behavior. Most of these works deal with the design stage: the aim is to represent the diversity