

Applying Logistic Regression for Activity Recognition

Fahd Al-Bin-Ali

Computer Science Department
University of Arizona
Tucson, AZ 85721 USA
albinali@cs.arizona.edu

Nigel Davies

Computing Department
Lancaster University
Lancaster LA1 1WB UK
nigel@comp.lancs.ac.uk

ABSTRACT

Intelligent environments typically involve sophisticated learning models that are difficult for end-users to control and understand, but usable and deployable systems need not be complicated for end-users. Numerous research projects have employed complex learning methods for rather simple pattern recognition problems without realizing that using complex techniques comes with a high cost to end-users; explaining how system decisions are made and extracting causal relationships becomes more difficult. We argue that using simple learning models, with simple relationships between inputs and responses, is necessary for the development of more intimate systems that can engage users in richer interactions. We report on a series of experiments that explore the use of simple logistic regression for identifying activities in domestic environments.

INTRODUCTION

Attempting to build systems that can engage users in dialogues must allow for the translation of observed data to cause-effect relationships. Typically, statistical techniques have been used to determine these relationships. However, such methods cannot be used reliably to extract causal relationships especially in dynamic ubiquitous environments. Numerous reasons account for this limitation.

First, conventional statistical methods depend on covariation, which does not necessarily imply causality. Second, in statistical analysis, there is an implicit assumption that datasets reflect enough discriminating elements that can allow for the extraction of causes and effects: a highly questionable assumption in ubiquitous spaces. Finally, statistical methods often require stable

models, where as, in ubiquitous environments, data sources (e.g. sensors) often change.

Therefore, intelligent environments cannot rely exclusively on learning models but should request assistance from end-users. This suggests employing simple learning techniques that can be explained to end-users; to allow users to assist systems in their decisions. In this paper, we use simple Logistic Regression (LR) [2], as one approach for classifying activities in domestic environments. We give an overview of our system and we report promising experimental results.

SYSTEM OVERVIEW

Our system consists of binomial and multinomial logistic modules and 5 smart-its sensor boards to collect information from the environment. A smart-its board [1] is built around a microchip PIC microprocessor (PIC18F252) with 14 inputs for binary sensors and analog-to-digital conversion units that allow five analog sensors to be attached. Attached to every board are the following sensors: two dual-axis accelerometers, a capacitive proximity sensor, a temperature sensor, a light intensity sensor, an infrared sensor and a radiometrix BIM3 chip that transmits and receives data wirelessly over a short range. The sensor boards were deployed in a domestic space and users specified their activities using wireless handhelds to train the system.

EXPERIMENTAL RESULTS

Initially, we performed stepwise forward selection on all sensor data to identify influential sensors. The stepwise procedure is based on a statistical algorithm that checks for the statistical significance of a variable based on a fixed rule. In more detail, sensor errors are assumed to follow a

binomial distribution (for dichotomous responses) and a multinomial distribution (for the polychotomous responses), and significance is assessed via the likelihood ratio chi-square test. Thus, the statistically most significant sensor is the one with the greatest change in the log-likelihood of the model with and without the sensor. We used this approach to build our models. Subsequently, we used these models for classifying the activities.

Our first experiment characterizes the performance of LR for dichotomous situations (i.e. classifying 2 activities such as cooking and watching TV). Figure 1 shows the proportion of the incorrectly classified activities using the regression formula for different cutoff values. For most parameter estimation iterations, we found that LR performs best on our data with a cutoff value ranging from 0.6 to 0.8. However, early iterations show high standard errors for the estimated coefficients and high type I error probability (i.e. the probability that the model considers a sensor significant while in fact it is not), thus suggesting a suboptimal model. Later parameter estimation iterations significantly reduce the standard error and the type I error probability, thus resulting in a more accurate model that performs well with cutoffs ranging from 0.3 up to 0.9.

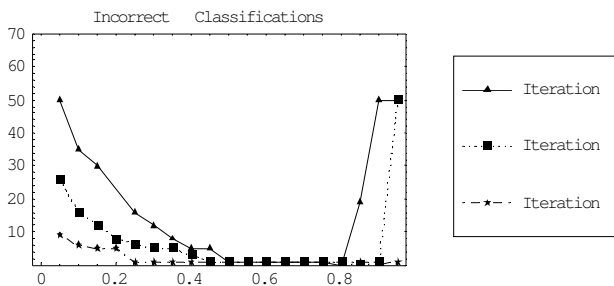


Figure 1: Incorrect Classifications (Dichotomous Response)

Our second experiment characterizes the performance of LR for the polychotomous case with 5 activities including bathing, cooking, watching TV, sleeping and no activity. Figure 2 shows the percentage of correct classifications for 3 incremental models that we selected using

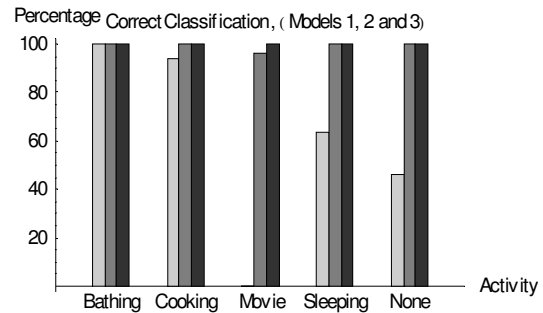


Figure 2: Correct Classifications for Different Models (Polychotomous Response)

forward selection. The first model uses the bathroom light intensity sensor only to classify all activities. Not surprisingly with one sensor, the regression formula performs poorly with overall accuracy of 60.8%. We also note that the model fails to classify any of the movie activities. The second model adds the second most significant sensor: the kitchen light intensity sensor. This significantly improves the classification accuracy reaching 98%. Finally, adding the bedroom light sensor further improves the classification accuracy of the regression formula resulting in 99.2% overall accuracy.

CONCLUSION

This paper proposes the use of simple learning models for classification problems in intelligent environments. Simple models represent direct and clear relationships between inputs and responses. End-users can potentially understand how systems reached certain conclusions and can alter systems' behavior when needed.

Our experiments show that logistic regression can be used for identifying activities in simple setups. Our future plans include assessing the performance of logistic regression in more complex setups and establishing the validity of our hypothesis.

REFERENCES

1. Holmquist, L.E. et. al. "Smart-Its Friends: A Technique for Users to Easily Establish Connections between Smart Artefacts," Ubicomp, Atlanta, September, 2001.
2. Hosmer, D., and Lemeshow, S.: *Applied Logistic Regression*. John Wiley & Sons, 2nd Edition (2000).