Data Mining Office Behavioural Information from Simple Sensors

Samuel J. O'Malley, Ross T. Smith and Bruce H. Thomas
School of Computer and Information Science
University of South Australia
Mawson Lakes Boulevard, Mawson Lakes, South Australia, 5095

oymsj001@mymail.unisa.edu.au ross.smith@unisa.edu.au bruce.thomas@unisa.edu.au

Abstract

This paper discussed the concept of using three simple sensors to monitor the behavioural patterns of an office occupant. The goal of this study is to capture behaviour information of the occupant without the use of invasive sensors such as cameras that do not maintain a level of privacy when installed. Our initial analysis has shown that data mining can be applied to capture re-occurring behaviours and provide real-time presence information to others that occupy the same building.

Keywords: Digital Foam, Data Mining, Apriori Algorithm, Non-invasive, Ambient Display, Market Basket Analysis

1 Introduction

This paper explores the concept of using simple sensors incorporated into an office environment to capture behavioural information that provides presence information to co-workers through an ambient display. Two sensors are used to capture the state information of doors when they are opened and closed, in conjunction with a pressure sensitive cushion on the desk chair which we refer to as a 'seat sensor'. We present how these three simple sensors can be set up to log information to determine physical presence and behavioural habits by data mining the sensor information. By itself this data means nothing, but by data mining the information over a long period of time we can start to see reoccurring patterns in the occupant's behaviour.

Sensors capture data from the physical world and convert the information to analogue or digital data. Electronic sensors can capture temperature, humidity, light, and sound amongst others. We consider the complexity of a sensor in terms of the data complexity. For example, a 1.3 megapixel webcam at 24 FPS generates approximately 20 gigabytes of information per hour. Although complex sensors are very powerful for capturing data, this paper shows how useful information can also be captured from simple sensors that can be used to capture everyday behavioural patterns. By capturing the state of these three sensors we can start to build a behavioural presence model using data mining.

2 Background

Sensors come in many different forms and new types are still being discovered. In this paper we are interested in capturing two events; firstly the state of a door to identify when it is open or closed and secondly to identify when an office seat is occupied.

Data mining provides a mechanism of discovering useful information by analysing captured data to provide a summary of the patterns that occur. The Apriori algorithm is commonly used on massive supermarket transaction databases to discover association rules, such as what products are commonly purchased together. This is called market basket analysis (Chen et al., 2005).

3 Simple Sensors Behaviour Concept

While a camera in the office would provide us with all the information we needed to monitor an occupants presence, it can be very invasive. By utilising simple sensors such as magnetic door sensors and a seat sensor, the only information recorded is that which could be obtained by walking past the office and looking in. By employing the state information of these three sensors, we can discover behavioural information such as the most likely time that the occupant is available as a pattern observed over a period of time.

From analysing the layout of the office in Figure 2, we noted that there are three possible locations for simple sensors that provide the most meaningful information. The office we are interested in shares a reception area with another office. As either door can be locked to outsiders, we chose to place magnetic door switches on both the inner and outer doors. Finally we decided to place a sensor on the desk chair to determine if the occupant is seated or not. These three states can give us an insight as to the occupant's availability.

3.1 Data Mining Sensor Information

The goal of data mining the sensor information is to discover patterns in the office occupant's behaviour. This will give us an insight into the most likely times that the occupant is available. We used the Apriori algorithm (Agrawal and Srikant, 1994) to generate a list of association rules. In order to apply the Apriori algorithm we logged the data as transactions, where a new item is added to the log when a state changes. An example of a useful association rule is "**!**inner → **!**seated 97%" which translates to: 97 percent of the time, if the inner door is shut then the occupant is not seated.

3.2 Ambient Display of Information

A visualisation we employed to present the status information to others in the same building but in a different room follows a traffic light metaphor (see Figure 1). The purpose of the Traffic Light visualisation is to inform the current availability of the occupant. This visualisation utilises the behavioural knowledge gained from preliminary data logging, based on four weeks of data, and attempts to determine whether the occupant is available or not using the state changes. The traffic light uses ambient lights to display and communicate information without distracting the user from their primary task. We employ the philosophy of Ishii et al. used in their ambientRoom system. The ambientRoom uses simple lights to convey information such as the number of hits on a website (Ishii and Ullmer, 1997).

Figure 1: Traffic Light Visualisation

4 Implementation

We employ magnetic door switches and a foam based pressure 'seat' sensor for the implementation of our system. These sensors are monitored using a micro-controller that captures when the state of the sensors changes. Magnetic door sensors are used to acquire state information from two doors located in the office, the state identified is either open or closed. These sensors are very simple, the switch is located on the door frame and a magnet is placed on the door. When the door is shut the magnet pulls the switch into an open position. This information is identified by a micro-controller and communicated to the computer as a string of comma delimited integers over the serial port.

We decided to create a cushion which features four sensors made from a material called Digital Foam to detect when the occupant was seated. Digital Foam is a conductive foam material that changes resistance when it is compressed (Smith, 2009), this material has also been used in an input device for free-form 3D modelling which records the depth and location of finger presses on its surface (Smith et al., 2008). By reading the resistance values with the microcontroller, we can detect if the cushion is being compressed. Including four sensors enables us to potentially differentiate between a person and inanimate objects, such as a stack of books, placed on the chair.

There are two modules to the software architecture, Data logging, and visualisation. The Data logging software reads the sensor information from the serial port and converts it into a meaningful format. This information, along with the time and date, is stored in a file which can later be used for data mining. This information is sent across the network using TCP/IP in a standard client-server configuration. The visualisation software then presents the data using a variable floor plan diagram on a wall mounted plasma screen, and via the ambient Traffic Light visualisation. This diagram features animated doors and seat state changes.

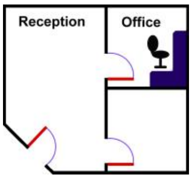

Figure 2: Office Floor Plan Visualisation

5 Results

The data used to generate the results was gathered over four weeks of constant logging. Each log entry contains the time of day which is: early morning (before 9am), morning (9am to midday), afternoon (until 5pm) or evening (until midnight), as well as the day of the week. In this paper we are focusing on finding the associations in the data rather than graphing behaviour over time because we have only recorded four weeks of data. In the future when we have logged months or even years of data then we will use graphs to visualise the occupant's reoccurring behavioural patterns over time.

Table 1: Top Results from the Apriori Algorithm

Rule	Association Rule	Confidence
1	$EMorning \rightarrow !seated$	100%
2	$Evening \rightarrow !seated$	94%
3	$Morning \rightarrow !seated$	72%
4	$Afternoon \rightarrow !seated$	66%
5	$EMorning \;\&\; !inner \rightarrow !seated$	100%
6	$Evening \;\&\; !inner \rightarrow !seated$	97%
7	$Afternoon \;\&\; !inner \rightarrow !seated$	92%
8	$!inner \rightarrow !seated$	97%
9	$!outer \rightarrow !seated$	96%
10	$!outer \;\&\; !inner \rightarrow !seated$	97%
11	$seated \rightarrow outer \;\&\; inner$	89%
12	$Monday \rightarrow !seated$	6%
13	$Tuesday \rightarrow !seated$	87%
14	$Wednesday \rightarrow !seated$	69%
15	$Thursday \rightarrow !seated$	91%
16	$Friday \rightarrow !seated$	95%
17	$Wednesday \rightarrow inner$	79%

Key: !inner = Inner Door Shut, inner = Inner Door Open

The association rules generated from the Apriori algorithm give us some useful insights into the office occupant's behaviour, **Table 1** is a list of the most relevant association rules that we have selected.

Rules 1 to 4 show us that even if the door is open before 9am, it is very unlikely that the occupant is seated yet, and Afternoon is the most likely time that the occupant is seated.

Rules 5 to 7 can be interpreted as showing that the Afternoon is the most likely time for the occupant to be working with the office door shut.

Rules 8 to 11 show 10 cases in four weeks of working with the inner door shut, 11 cases with the outer door shut, and 5 cases with both shut. This shows that it is a standard behaviour of the occupant to work with the door shut; however it is not done all the time. This means that if either door is shut but the occupant is seated, we display a yellow light on the Traffic Light to signify that the occupant is busy. Compared to a green light which shows the occupant is available, and a red light to show unavailability.

Rules 12 to 16 show the days on which it is unlikely for the occupant to be seated, Monday is very likely that the occupant is seated and could possibly mean that Monday is the day where the occupant sits for the longest time without leaving the office. If we count how much activity occurs on each day then Monday and Wednesday are the most active days, with Thursday and Friday being the least active. Together these two rules would suggest that Monday is the busiest day in the office. However the scope of this system is only the activity which occurs inside the office, Thursday and Friday could be equally as busy days with many meetings and commitments outside the office.

Rule 17 shows that 79% of Wednesday's activity was with the Inner Door open. This could mean that this is a day where the secretary and the other office occupant were not present, so that when the occupant left his office he locked the outer door rather than both doors.

6 Conclusion

What we hope to achieve with the analysis is to provide a summary of the behavioural information over time. From these preliminary results we can already see glimpses of what we could potentially learn from a larger dataset. Given the system is kept running, the data mining will allow new patterns to be revealed and potentially re-occurring behaviours over many years could be identified. Providing co-workers with this information where appropriate will allow them to select the most appropriate time to find the occupant. In the future we could also include other data such as public holiday dates, semester times and conference submission deadlines to allow us to generate an even more detailed behavioural presence model.

7 References

AGRAWAL, R. & SRIKANT, R. 1994. Fast Algorithms for Mining Association Rules in Large Databases. *Proceedings of the 20th International Conference on Very Large Data Bases.* Morgan Kaufmann Publishers Inc.

CHEN, Y.-L., TANG, K., SHEN, R.-J. & HU, Y.-H. 2005. Market basket analysis in a multiple store environment. *Decision Support Systems,* 40**,** 339-354.

ISHII, H. & ULLMER, B. 1997. Tangible bits: towards seamless interfaces between people, bits and atoms. *Proceedings of the SIGCHI conference on Human factors in computing systems.* Atlanta, Georgia, United States: ACM.

SMITH, R. T. 2009. *Digital foam: a 3D input device*.

SMITH, R. T., THOMAS, B. H. & PIEKARSKI, W. 2008. Digital foam interaction techniques for 3D modeling. *Proceedings of the 2008 ACM symposium on Virtual reality software and technology.* Bordeaux, France: ACM.