

Understanding the indoor environment through mining sensory data - a case study

Article

Accepted Version

Wu, S. and Clements-Croome, D. (2007) Understanding the indoor environment through mining sensory data - a case study. *Energy and Buildings*, 39 (11). pp. 1183-1191. ISSN 0378-7788 doi: <https://doi.org/10.1016/j.enbuild.2006.07.011> Available at <https://centaur.reading.ac.uk/12452/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1016/j.enbuild.2006.07.011>

Publisher: Elsevier

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Understanding the indoor environment through mining sensory data-A case study

Shaomin Wu ^{a,*}, Derek Clements-Croome ^b

^a Centre for Resource Management and Efficiency, Sustainable Systems Department, Cranfield University, Bedfordshire MK43 0AL, UK

^b School of Construction Management and Engineering, The University of Reading, RG6 6AW Reading, United Kingdom

Received 10 June 2006; accepted 21 July 2006

* Corresponding author. Tel.: +44 1234 750111 x 2270; fax: +44 1234 751671. E-mail address: shaomin.wu@cranfield.ac.uk (S. Wu).

Abstract

A wireless sensor network (WSN) is a group of sensors linked by wireless medium to perform distributed sensing tasks. WSNs have attracted a wide interest from academia and industry alike due to their diversity of applications, including home automation, smart environment, and emergency services, in various buildings. The primary goal of a WSN is to collect data sensed by sensors. These data are characteristic of being heavily noisy, exhibiting temporal and spatial correlation. In order to extract useful information from such data, as this paper will demonstrate, people need to utilise various techniques to analyse the data. Data mining is a process in which a wide spectrum of data analysis methods is used. It is applied in the paper to analyse data collected from WSNs monitoring an indoor environment in a building. A case study is given to demonstrate how data mining can be used to optimise the use of the office space in a building.

© 2007 Elsevier B.V. All rights reserved.

Keywords: Wireless sensor network; Data mining; Clustering; Indoor environment

1. Introduction

The increasing miniaturisation of radio frequency (RF) devices and microelectro-mechanical systems (MEMS), along with the advances in wireless technologies, has generated a great deal of research and application interest in the area of wireless sensor networks (WSNs), which provide a promising infrastructure for gathering information about parameters of the physical world.

WSNs have found a wide spectrum of exciting applications [1,2], some of which are applications in various buildings. These are *home automation*, *indoor environmental monitoring* and *emergency services*:

1.1. Home automation

Networking various home appliances, such as vacuum cleaners, micro-wave ovens, and refrigerators [3], with wireless medium, has been dreamt of for many years. Embedded sensors inside such appliances can interact with each other, and with the external network via the internet or satellites. They allow users to manage home devices locally and remotely more easily.

1.2. Indoor environmental monitoring and emergency services

Indoor physical parameters can be monitored by different sensors that assist the occupants in managing their thermal comfort, light comfort, efficient operation, and work productivity [4–6]. The signals provided by the monitoring system give a dynamic picture of the state of the indoor environment, thus, in principle, allowing for efficient real-time diagnostics of system and component malfunctions and operation anomalies, an off-line analyse.

There are two main applications of WSNs to indoor environmental monitoring: collecting information on environmental physical parameters in order to better control environmental systems such as heating, ventilation, and air-conditioning (HVAC), and emergency services such as fire and smoke detection. The sensor signals in these scenarios are usually used for decision making or triggering of actuators in real-time.

An important reason that people prefer WSNs to wired sensor networks in buildings is: WSNs are easy and cheaper to install [7]. In new residential buildings such networks can now be implemented easily and with very little cost by low power, wireless sensors. Upgrading a WSN based on new industry standards (e.g., IEEE 802.15.4 and ZigBee) is easily carried out simply by adding extra sensors. However, WSNs have some disadvantages. For example, they are limited in power, computational capacities, and memory; and they are prone to failures [8].

Research on the application of WSNs in building mainly focuses on the real-time control, which might aim to reduce energy consumption or improve well-being of occupants (e.g., [9,10]).

- Energy consumption can be reduced by using sensors to monitor occupant behaviour. Building energy performance is currently understood as dependent upon [11]: urban geometry, building design, systems efficiency, and occupant behaviour. Occupant behaviour can be monitored by sensors such as occupancy sensors and motion sensors (e.g., [12]). For example, when no occupancy is detected by occupancy sensors for a given time period (for example, 15 min), the HVAC system and lighting system can be turned off. When occupancy is again detected, the HVAC system and lighting system resumes operation as set by the user preference (e.g., [13]).
- Occupant well-being can be measured by various comfort indexes, among which thermal comfort and light comfort have drawn most attention. WSNs have been applied in buildings to optimise these comfort indexes (e.g., [9,10]).

However, there is little research on analysing datasets collected from sensors located in different parts of a building, or off-line analysis, to improve post-occupancy evaluation. This may be due to the fact that WSNs are still a new technology, and there are few datasets collected from the real world. Hence, little attention has been attracted. Nevertheless, analysing sensory data can allow facility managers, and/or building designers, to gain an in-depth understanding of the distribution of some physical parameters – e.g., temperature, humidity, and light – in the whole building, which can improve their post-occupancy evaluation and future design work.

On the other hand, *environments for working in are becoming more fluid to meet changing work patterns* [14], which means there is a need to understand both working patterns and detailed environmental information. Collecting and then processing sensory data of indoor environmental variables – such as temperature, humidity, light – through WSNs in a building draw a dynamic picture of the state of the indoor environment in the building, which can be helpful for improving productivity under an environment with the changing work patterns.

As the size of sensory data collected from a building might be so huge that the traditional statistical analysis technique is not able to deal with it, data mining can find its application. Data mining is a process which involves a variety of data analysis approaches. It has its roots in statistics and computer science, and has found applications in banking, insurance, manufacturing and many other industries. This paper proposes to use data mining to analysis sensory data. A case study shows how data mining can be applied in finding patterns of physical parameters in an indoor environment in a building. The patterns can be useful in the post-occupancy evaluation.

The paper is structured as follows. Section 2 offers an introduction to data mining, and reviews the needs of data mining techniques in analysing sensory data. Section 3 studies a dataset collected from a WSN in a building. Section 5 presents conclusions.

2. Data mining

This section offers a brief introduction to data mining.

2.1. What is data mining?

Data mining, as defined by Fayyad et al., and Piatetski-Shapiro and Frawley [15,16], is the non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data. It solves problems by analysing data that already exists in databases. Data mining is a synonym for another popularly used term, knowledge discovery in databases (KDD).

Techniques used in data mining are categorised into two classes:

2.1.1. Predictive algorithms

These algorithms are usually to build a mapping function based on a set of input and output observations, for example, to build a model mapping staff's income based on their gender, educational level, and age. The techniques for building such models include regression modelling, decision trees, neural networks, K -nearest neighbour, and Bayesian learning algorithms.

2.1.2. Descriptive algorithms

These algorithms can be used for exploratory data analysis to discover individual patterns, such as associations, clusters, and other patterns that can be of interest to the user.

Related research areas include database technology and data warehouses, statistics, machine learning, pattern recognition and soft computing, text and web mining and visualisation.

The process of a typical data mining project might the follow steps: business understanding, data understanding, data preparation, modelling, model evaluation, and model deployment as shown below.

2.2. Data mining operations

A data mining project was initially carried out in different ways with each data analyst based on his/her own experience and way of approaching the problem often through trial-and-error. Later, people introduced standardised data mining processes, among which two processes are usually used in industries: SEMMA from the SAS institute, and CRISP-DM from SPSS company. SEMMA stands for sample, explore, modify, model, and assessment. The SAS data mining tool, SAS enterprise miner, has corresponding modules for the five processing steps. CRISP-DM stands for cross-industry standard process for data mining. The CRISP-DM was intended to be independent of the choice of data mining tools, industry segment, and the application/problem to be solved. It defines the crucial steps of the knowledge discovery process.

Although in most data mining projects, several iterations of individual steps or step sequences need to be performed, these basic guidelines are very useful both for the data analyst and the client in need for problem solution.

The individual steps of the CRISP-DM process [17] are the following:

- (1) *Business understanding*: To define business and data mining objectives, and the business and data mining evaluation criteria.
- (2) *Data understanding*: In this step, data miners become familiar with the data and the application domain, by exploring and defining the relevant prior knowledge.
- (3) *Data preparation*: In this step, through data cleaning and pre-processing, data miners create the relevant data subsets, find useful variables, and generating new variables.
- (4) *Modelling*: This is the most important step of this process, concerned with choosing the most appropriate data mining tools (from the available tools for summarization, classification, regression, association, clustering), and searching for patterns and models of interest.
- (5) *Evaluation*: The modelling results are interpreted, analysis and evaluation of results.
- (6) *Deployment*: In this phase, the produced models are put into action.

2.3. Data mining on sensory data

The majority of research on sensory data are focused on real-time dynamic data. Bontempi and Le Borgne [18] introduce an adaptive approach to mining sensory data. Elnahrawy and Nath [19] present a framework for cleaning and querying noisy sensors. However, their aims are on processing real-time sensory data, which is different from the off-line data analysis with regard to their constraint problems and data resources. Constraint problems like time, bandwidth and calculation capability are main factors that might impact the design of the algorithms for real-time data processing, whereas such problems might not exist for off-line data analysis. Data resources for the real-time data processing are the data collected until a time point, whereas those for the off-line data analysis are from an entire time period.

Other research on sensory data focuses on developing data mining algorithms to deal with various problems (for example, Kulakov and Daveev [20]).

However, little research has been found on using data mining to analyse sensory data for improving building performance that is related to occupant's benefits.

Potential use of data mining

Both the descriptive and predictive algorithms might find their applications in mining the sensory data from an indoor environment. Below we show some examples how data mining can be used.

2.3.1. Predictive algorithms

Predictive models can be built to estimate physical parameters (say, temperature and humidity) at a location where no sensor is placed, to predict a failure to occur, to predict occupant behaviour, etc.

2.3.2. Descriptive algorithms

Descriptive algorithms can be served to find the relationship between different variables, for example, occupant's behaviour and energy consumption, usage patterns and a certain failure mode, indoor air quality and energy consumption. Section 3 shows how clustering or association rule discovery algorithms can be used to assess the distribution of certain physical parameters.

3. Case study

In this case study, we are mainly interested in the distributions of three parameters, temperature, humidity, and light, in a building.

Information about temperatures in houses is of importance in assessing the value of various energy conservation measures and gives an indication of the standards of thermal comfort enjoyed by the occupants.

Temperature and relative humidity can affect comfort and indoor air quality. Changing thermostat settings or opening windows to try to control temporary fluctuations in temperature can worsen comfort problems and also have an adverse effect on other parts of the building.

For our case study, we borrow a sensory data that was created by the Intel Berkeley Research Lab, which deployed 54 wireless sensors in their lab for collecting information of temperature, humidity, and light from 28th February and 5th April 2004. A log of about 2.3 million readings collected from these sensors, along with the health of sensors¹ can be found in [21]. The data are sampled every 31 s.

3.1. Data understanding

The sensors under study were arranged in the lab according to the diagram shown in Fig. 1. The original dataset has 2,313,682 observations, and eight variables: *date*, *times-tamp*, *epoch*, *moteID*, *temperature*, *humidity*, *light*, and *battery voltage*. Epoch is a monotonically increasing sequence number from each mote. Two readings from the same epoch number were produced from different motes at the same time. MoteID's, ranging from 1 to 54, are identities of wireless sensors. Temperature is in degrees Celsius. Humidity is temperature corrected relative humidity, ranging from 0 to 100%. Light is in Lux. Voltage is expressed in volts, ranging from 2 to 3.

Fig. 2 shows the number of observations collected from each mote. Motes 5 and 28 have never transmitted any data. Mote 22 is the busiest one (with 23,206 observations), whereas mote 15 is the idlest one (with only 722 observations).

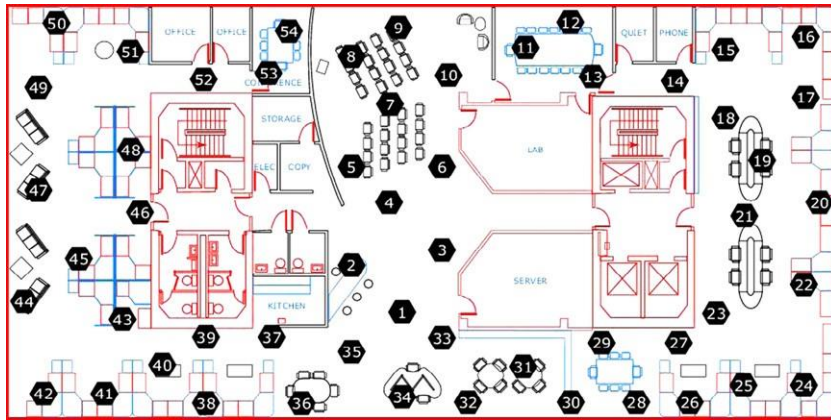


Fig. 1. Wireless sensors installed in the Intel Berkeley Research Lab [21].

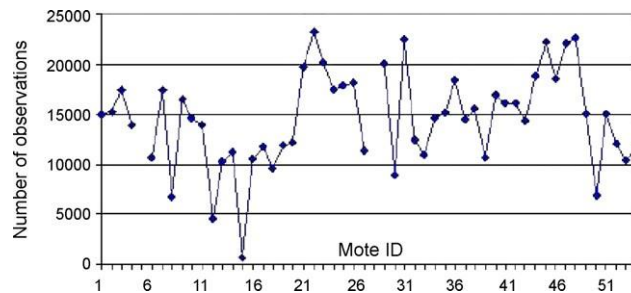


Fig. 2. Number of observations from each mote.

¹ Here, health of sensors means the voltage of the batteries attached to the wireless sensors.

3.2. Invalid observations

The original observations are sampled every 31 s in a whole day. In this paper, we would like to only focus on the change of the physical variables within working hours. If one is concerned with the indoor environment within working hours, and assume that working hours are from 8:00 to 18:00, then 1,365,866 observations are ignored, and 947,816 left.

The following steps are conducted to remove observations that are not valid for various reasons.

- Invalid motes: The 54 motes are labelled from 1 to 54 in the dataset, that is 1 ;; moteID ;; 54. Hence, if we search all of the observations with moteID bigger than 54 or smaller than 1, 3695 observations can be found and removed, and 944,121 observations left.
- Invalid temperature: Assume the temperature in the building ranges from 10 to 40 8C. Under such a condition, 169,428 observations are found and removed, and 774,693 observations are left.
- Invalid humidity: As indicated in the website [21], the humidity falls in 0–100%. All of the observations satisfy this condition, and hence no observation is removed from the dataset.
- Invalid light: Set light > 0, then 23,654 observations in which light is negative are found and removed and 751,039 observations are left.
- Invalid voltage: Set 2 ;; voltage ;; 3, then 167 observations are removed and 750,872 observations remains.

Having conducted the above cleansing procedures, we found the data collected from 26th March and 5th April 2004 have been removed. We have an impression that the original dataset has a large amount of invalid observations, or in a data mining terminology, the dataset is very dirty.

All of the above cleansing procedures check the validation of one variable. As the number of removed observations is so big, we doubt there might be outlying observations, or called outliers, in the dataset. In the following subsection, we concentrate on removing outliers.

3.3. Removal of outliers

Fig. 3 shows the number of the observations collected within the rest days, from 28th February to 25th March (denoted as days 1–27 in the figure).

From Fig. 3, the last 2 days – days 26 and 27 (i.e., 24th and 25th March) – collects a small amount of observations.

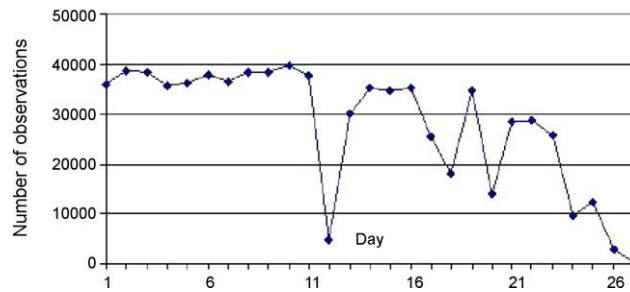


Fig. 3. Frequency.

Table 1
Statistics of observations on 25th March 2004

	Mean	Min	Max
Temperature	34.81	28.22	39.90
Humidity	48.94	44.52	52.05
Light	830.56	195.04	1847.36

- Observations collected on 25th March. There are only 218 observations collected on the last date (25th March), and all of these observations are from mote 9. The basic statistics of temperature, humidity, and light on 25th March is shown in Table 1. As all of these observations are from one mote, which is not representative enough, we remove them from the dataset. Having been removed these observations, the dataset has 750,654 observations left.
- Observations collected on 10th and 24th March. There 4823 observations on 10th March, and 1827 observations on 24th March. As they are from a number of different motes, we keep them in the dataset.

3.3.1. Using clustering algorithms to discover outliers

There might be outliers within the rest of the observations. We can use a clustering algorithm to detect the possible outliers. The SAS system (<http://www.sas.com/>) is used to implement the clustering process.

There are a dozen of clustering algorithms in the SAS system. Among the clustering algorithms, the K -means algorithm is usually used for large datasets. As our dataset is so big, we select the K -means clustering algorithm to detect outliers. The K -means clustering algorithm² have been widely used in data mining and statistical data analysis (see [22] for more detailed discussion).

We change the number of clusters, K , from 15 to 50. When $K = 16$, both temperature and humidity are abnormally big in the smallest cluster. Hence, we set the number of clusters to be 16, there are 506 observations in the smallest dataset. These observations are collected from motes 18, 19, 22, 25, 38, 40, 41, 42, and 44. The basic statistics of the 506 observations are shown in Table 2. From the table, the mean temperature is high, we therefore consider the 506 observations to be outliers.

After the outliers have been removed, there remain 750,148 observations.

Table 2 Basic statistics of outliers

Variable	N	Mean	S.D.	Minimum	Maximum
Temperature		34.81		28.22	39.90
Humidity		48.94		44.52	52.05
Light		830.56		195.04	1847.361

3.4. Clustering

It is very hard to estimate the temperature distribution of a place in a day. For example, given four days, March 1–4, we can find the temperature distributions collected by mote 22 are different (see Figs. 4 and 5). They showed different overall patterns. For example, the temperature on 1st March are increasing dramatically from 8:00 am to 14:00 pm, whereas the temperature on 2nd March changes with a different pattern comparing to that on 1st March.

Cluster analysis is performed using the Ward algorithm [23] in the SAS system. By comparing the clustering results with different numbers of clusters using the statistic R^2 , we find that it is best to cluster them into four clusters. Table 3 shows basic statistics of the four clusters. Cluster 1 has the smallest size, highest temperature and highest humidity. Cluster 2 has higher temperature, and lowest humidity. Cluster 3 has lowest temperature, and dimmest light. Cluster 4 has the largest size of observations, and brightest light.

Figs. 6 and 7 show the percentage of observations of a mote belonging to a certain cluster.

Fig. 7 shows that all motes belong to Cluster 4 with high percentages. From Fig. 6, more than 20% of observations from motes 21, 22, 23, 24, 25, 27 belong to Cluster 2, which has a higher temperature, lower humidity and median illuminance. More than 20% observations from motes 7, 10, 13, 14, 46, 48, 51 and 52 belong to Cluster 3, which has lower temperature, higher humidity, and lower light. According the locations of these motes, along with the findings, the building managers can adjust the functionality of the office. For example, staffs prefer a lower temperature can work in the space where motes 7, 10, 13, 14, 46, 48, 51 and 52 are located.

Investigating the distribution of the clusters with time of a day, we can draw four histograms shown as Figs. 8 and 9 (the X-axis in the figure is decimal, which causes holes in the figure). The values of the X-axis in these figures are time ranged from 8:00 am to 18:00 pm. From the figures, Cluster 2 is fairly uniformly distributed with time. In Cluster 3, the number of observations are decreasing before 13:00 pm and increasing after that time. One reason might be the office becomes warmer during the working time, and cooler when time is close to 18:00 pm. From these figures, Clusters 3 and 4 have quite strong trends with time.

²The K -means clustering algorithm classifies n observations into K clusters by assigning each observation to the cluster whose average value is nearest to it by some distance measure (usually Euclidean). The algorithm computes these Assignments iteratively, until reassigning points and re-computing averages (over all points in a cluster) produces no changes.

Table 3 Basic statistics of each cluster

Cluster	Variable	N^* Maximum	Mean	S.D.	Minimum	Maximum
1	Temperature	8,309	31.94	3.81	24.90	39.98
	Humidity	8,309	46.09	5.01	34.85	57.67
	Light	8,309	735.79	495.67	22.91	1847.07
2	Temperature	41,195	31.02	2.16	26.24	39.96
	Humidity	41,195	24.05	4.47	14.41	36.82
	Light	41,195	660.17	490.94	33.94	1847.07
3	Temperature	100,298	20.62	1.52	14.50	25.36
	Humidity	100,298	45.88	3.25	38.52	60.53
	Light	100,298	402.50	307.55	5.89	1847.07
4	Temperature	600,346	24.10	2.63	15.70	34.62
	Humidity	600,346	36.23	5.19	18.76	54.70
	Light	600,346	925.62	596.00	5.89	1847.07

* In the table, N is the number of observations.

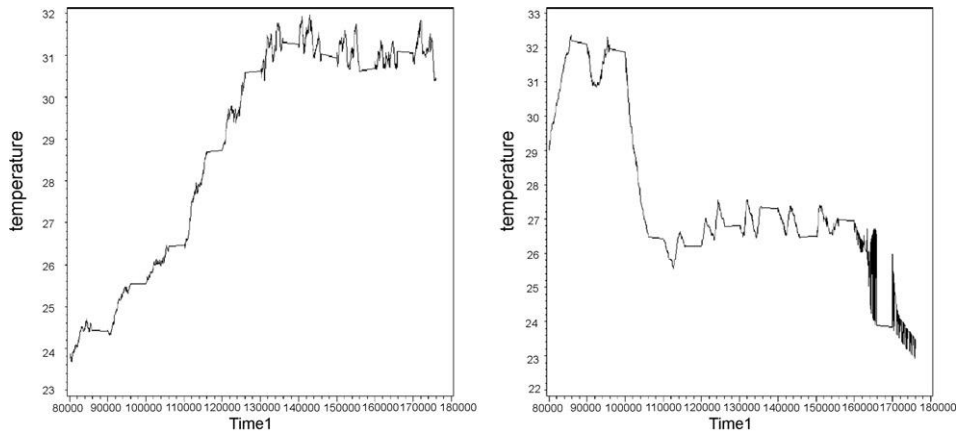


Fig. 4. Temperature distribution on 1st and 2nd March.

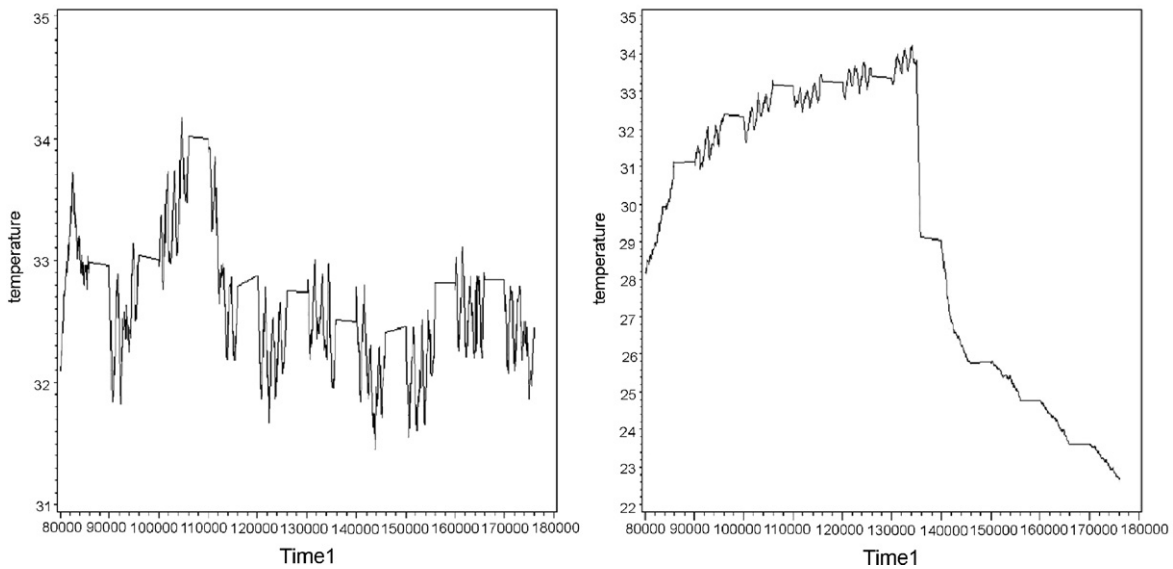


Fig. 5. Temperature distribution on 3rd and 4th March.

4. The relationship between temperature and humidity

The Pearson correlation coefficients between variables temperature, humidity, and light are shown in Table 4. From this table, temperature and humidity have a strong negative relationship, whereas light has little relationship with the other two variables.

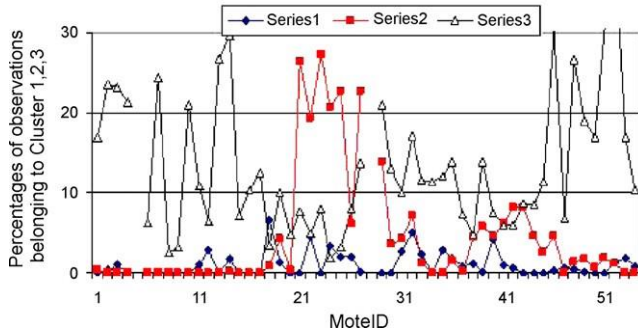


Fig. 6. Percentages of observations belonging to Clusters 1, 2, 3.

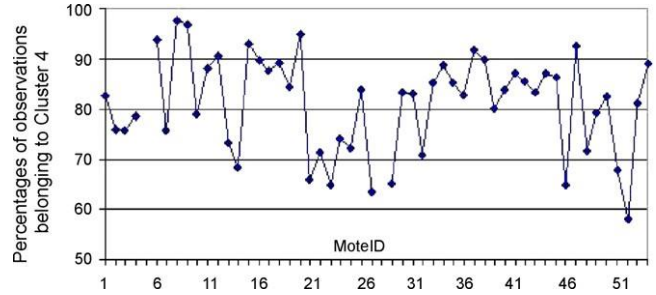


Fig. 7. Percentages of observations belonging to Cluster 4.

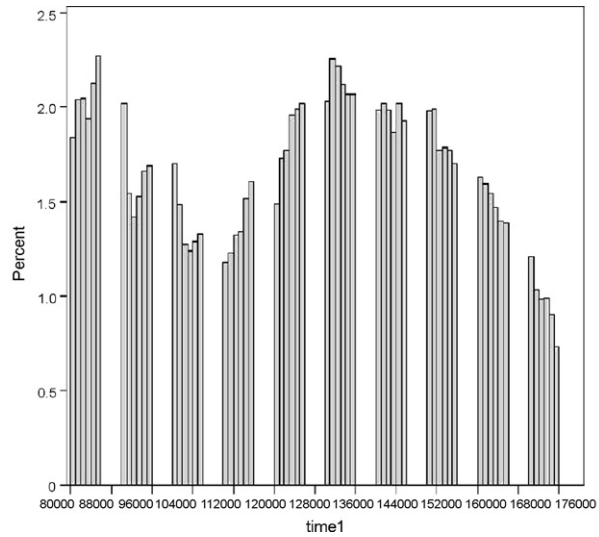
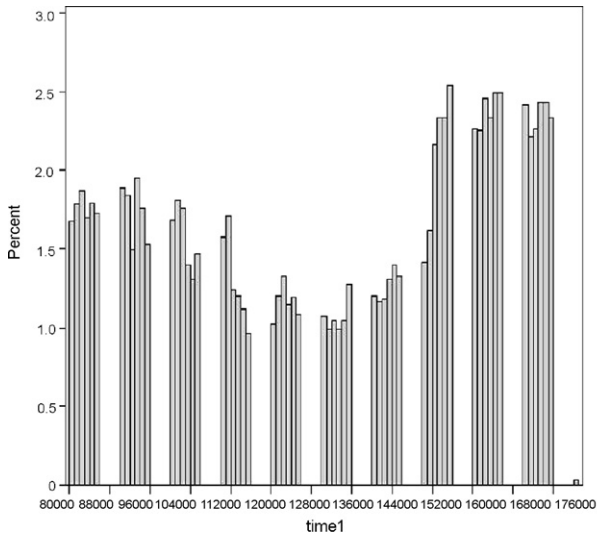


Fig. 8. Clusters 1 and 2.

Table 4
Pearson correlation coefficients

Pearson	Temperature	Humidity	Light
Temperature	1.00000	-0.63082	0.12820
Humidity	-0.63082	1.00000	-0.04479
Light	0.12820	-0.04479	1.00000

4.1. Discussion

Section 3 demonstrates that a high percentage of observations in the sensory data collected from the Intel Berkeley Research Lab is invalid, which means the sensory data are noisy.

Commonly, sensory data is subject to several different sources of errors, which can be broadly classified as either systematic errors (bias) or random errors (noise). The main sources of noise are

- noise from external sources,
- random hardware noise,
- inaccuracies in the measurement technique (i.e., readings are not close enough to the actual value of the measured

phenomenon),

- various environmental effects and noise, and
- imprecision in computing a derived value from the underlying measurements (i.e., sensors are not consistent in measuring the same phenomenon under the same conditions).

Consider the characteristics possessed by the sensors, the following two methods can be applied to improve the reliability of the data.

4.1.1. Temporal redundancy

The sensory data exhibits temporal correlation, which means two consecutive observations from a sensor are correlated. In order to improve the reliability of readings from a sensor, the sampling rate can be increased. For example, we can read temperature every 15 s instead of 30 s, and average two consecutive readings, and then transmit the mean of the two readings. The disadvantage of this approach is it increases energy consumption of the sensor.

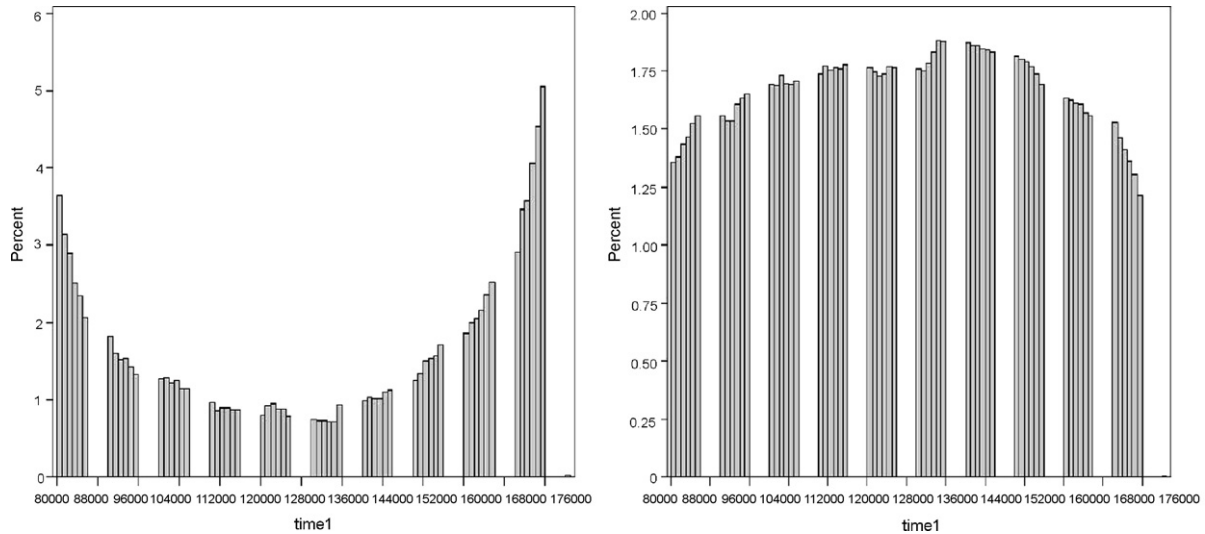


Fig. 9. Clusters 3 and 4.

4.1.2. Spatial redundancy

The sensory data also exhibits spatial correlation, which means readings from two neighbour sensors at a time point are correlated. Combining (for example, averaging, weighted averaging, etc.) readings from neighbour sensors is a good way to improve the reliability of sensory data.

Section 3 mainly uses clustering algorithms to analysis data, and for the entire time period (from 28th February to 5th April). Another two angles to proceed data mining for the sensory data are

- Other data mining algorithms can be used to analyse the sensory data. For example [24], built predictive models to predict temperature of a space where no sensor is occupied.
- Data mining can be conducted from a multi-dimensional angle. From example, we can look at the temperature distribution at a specified time period, or a specified room.

5. Conclusions

As the business environment becomes increasingly more competitive, it is essential that all available resources are used optimally and effectively. The need to evaluate various comfort indexes in different part of a building are becoming increasingly important, as those indexes directly and/or indirectly, affect people's working productivity in the building.

Using wireless sensor networks to collect data about the indoor physical parameters is a promising approach as wireless sensors is becoming cheaper, and they are easy to install. Analysing data collected by the wireless sensor networks can provide a whole picture of various distributions of indoor physical parameters.

Data mining provides a variety of techniques that aim to analyse large datasets in order to find interesting patterns.

This paper applies data mining to analysis data collected from wireless sensor networks. To our best understanding, this is the first paper on using data mining to analyse sensory data for understanding the distributions of interesting parameters in an indoor environment. The case study shows the following two characteristics:

- (1) The sensory data are very noisy. Hence, more effort should be made in data preparation and data cleansing phase when conducting data mining.
- (2) Interesting patterns can be found using data mining techniques, as shown in the case study of this paper.

This paper only studied analysing the off-line dataset. It will be interesting to measure different aspects of considerations from a facilities manager point of view. Our future work will be focusing on analysing the real-time dynamic data.

Acknowledgements

The authors would like to thank Dr. Penny Noy for her helpful comments and suggestions, which have resulted in a couple of improvements in the paper.

This work is one of outcomes of the research project IDCOP supported by EPSRC. We would like to thank the Intel Berkeley Research Lab for their effort in collecting the data used in the case study.

References

- [1] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, E. Cayirci, Wireless sensor networks: a survey, *Computer Networks* 38 (4) (2002) 393–422.
- [2] T. Arampatzis, J. Lygeros, S. Manesis, A survey of applications of wireless sensors and wireless sensor networks, *Proceedings of the IEEE International Symposium on Intelligent Control* 1–2 (1) (2005) 719–724.
- [3] E.M. Petriu, N.D. Georganas, D.C. Petriu, D. Makrakis, V.Z. Groza, Sensor-based information appliances, *IEEE Instrumentation and Measurement Magazine* December (2000) 31–35.
- [4] J. Granderson, A.M. Agogino, Y. Wen, K. Goebel, Towards demand responsive intelligent daylighting with wireless sensing and actuation, *IESNA, Proceedings of 2004 Annual IESNA conference*, 2004.
- [5] J.S. Sandhu, A.M. Agogino, A.K. Agogino, Wireless sensor networks for commercial lighting control: decision making with multi-agent systems, in: *Proceedings of the AAAI Workshop on Sensor Networks*, 2004.
- [6] V. Singhvi, A. Krause, C. Guestrin, Intelligent light control using sensor networks, in: *Proceedings of the 3rd ACM Conference on Embedded Networked Sensor Systems*, San Diego, 2005.
- [7] W.M. Healy, Lessons learned in wireless monitoring, *Ashrae Journal* 47 (2005) 54–62.
- [8] C. Perkins, *Ad Hoc Networking*, Addison-Wesley, 2001.
- [9] M.C. Mozer, The neural network house: an environment that adapts to its inhabitants, in: M. Coen. (Ed.), *Proceedings of the American Association for Artificial Intelligence Spring Symposium on Intelligent Environments*, AAAI Press, Menlo Park, California, 1998, pp. 110–114.
- [10] D. Kolokotsa, G.S. Stavrakakis, K. Kalaitzakis, D. Agoris, Genetic algorithms optimized fuzzy controller for the indoor environmental management in buildings implemented using plc and local operating networks, *Engineering Applications of Artificial Intelligence* 15 (5) (2002) 417–428.
- [11] C. Ratti, N. Baker, K. Steemers, Energy consumption and urban texture, *Energy and Buildings* 37 (7) (2005) 762–776.
- [12] M. Mysen, S. Berntsen, P. Nafstad, P.G. Schild, Occupancy density and benefits of demand-controlled ventilation in Norwegian primary schools, *Energy and Buildings* 37 (12) (2005) 1234–1240.
- [13] V. Garg, N.K. Bansal, Smart occupancy sensors to reduce energy consumption, *Energy and Buildings* 32 (1) (2000) 81–87.
- [14] D. Clements-Croome, *Creating the Productive Workplace*, Taylor and Francis, 2005.
- [15] U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, *From Data Mining to Knowledge Discovery*, *Advances in Knowledge Discovery and Data Mining*, AAAI/MIT Press, 1996.
- [16] G. Piatetsky-Shapiro, W. Frawley, *Knowledge Discovery in Databases*, MIT Press, Cambridge, MA, 1991.
- [17] SPSS, CRISP-DM Process, <http://www.crisp-dm.org>, 2006.
- [18] G. Bontempi, Y. Le Borgne, An adaptive modular approach to the mining of sensor network data, in: R. Pedersen, E. Jul (Eds.), *First International Workshop on Data Mining in Sensor Networks*, Newport Beach, CA, 2005.
- [19] E. Elnahrawy, B. Nath, Cleaning and querying noisy sensors, in: *Wireless Sensor Networks and Applications Workshop*, San Diego, California, USA, 2003.
- [20] A. Kulakov, D. Davcev, Data mining in wireless sensor networks based on artificial neural-networks algorithms, in: R. Pedersen, E. Jul (Eds.), *First International Workshop on Data Mining in Sensor Networks*, Newport Beach, CA, 2005.
- [21] Intel Berkeley Research Lab, <http://db.lcs.mit.edu/labdata/>, 2004.
- [22] I.H. Witten, E. Frank, *Data Mining—Practical Machine Learning Tools and Techniques*, Morgan Kaufmann Publishers, San Francisco, CA, 2005 .
- [23] J.H. Ward, Hierarchical grouping to optimize an objective function, *Journal of American Statistical Association* 58 (301) (1963) 236–244.
- [24] Y. LeBorgne, G. Bontempi, Round robin cycle for predictions in wireless sensor networks, in: *Proceedings of the 2nd International Conference on Intelligent Sensors, Sensor networks and Information Processing*, 2005.