Designing Sensor Sets for Capturing Energy Events in Buildings

Tom Lovett[1], Elizabeth Gabe-Thomas[3], Sukumar Natarajan[1],
Matthew Brown[2], Julian Padget[2]
Dept. of Architecture and Civil Engineering[1], Dept. of Computer Science[2], Dept. of Psychology[3]
University of Bath, Bath, UK
{t.r.lovett,e.gabe-thomas,s.natarajan,m.brown,j.a.padget}@bath.ac.uk

ABSTRACT

We study the problem of designing sensor sets for capturing energy events in buildings. In addition to direct energy sensing methods, e.g. electricity and gas, it is often desirable to monitor energy use and occupant activity through other sensors such as temperature and motion. However, practical constraints such as cost and deployment requirements can limit the choice, quantity and quality of sensors that can be distributed within each building, especially for large-scale deployments. In this paper, we present an approach to select a set of sensors for capturing energy events, using a measure of each candidate sensor's ability to predict energy events within a building. We use constrained optimisation – specifically, a bounded knapsack problem (BKP) – to choose the best sensors for the set given each sensor's predictive value and specified cost constraints. We present the results from a field study of 4 UK homes with temperature, light, motion, humidity, sound and CO_2 sensors, showing how valuable yet expensive sensors are often not chosen in the optimal set.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous

Keywords

Energy use; sensing; intelligence; interaction; ENLITEN

1. INTRODUCTION AND PRIOR WORK

To tackle the problem of energy usage reduction in buildings, researchers have used sensing technology to capture and analyse buildings' energy use so that efficiency can be improved and methods of lowering energy demand can be explored, e.g. through changing occupants' energy-related behaviour. The first step in enabling behavioural change is the gathering and sensing of pertinent data. As such, key questions emerge about how best to approach energy sensing: what sensors should we use? How many do we need? How intrusive and costly is the installation? Direct energy sensing with electricity and gas sensors is commonplace [3], but direct sensing alone does not account for total energy use, nor does it allow for non-trivial analyses of the often individualistic causal factors involved in energy consumption. There are two key contributions in this paper:

- A method for assigning a value to a sensor in terms of its utility in capturing human activities that involve energy consumption in a building.
- A method for the selection of maximal value sensor sets subject to practical constraints such as budget and sensor quantities.

The closest work to the study in this paper is Zhang *et al.*'s study of feature selection for occupancy classification in office spaces [6]. Here, the authors explore the relative information gain – or uncertainty coefficients – as a value measure for a small range of sensors using intermittent ground truth gathered in an office environment. We use a different measure of sensor value in a domestic environment, but our results broadly support Zhang *et al.*'s, which show that sound and CO_2 sensors appear to be the most effective at detection; albeit for energy events in ours, and occupancy events in theirs. By incorporating sensor costs, however, we show that these sensors are not always the best ones to choose for maximising sensor value given a set of constraints.

2. APPROACH

2.1 Constrained Optimisation

The knapsack problem is a simple integer linear program that seeks to find the optimal combination of n distinct items that maximises the total value of a weight-constrained knapsack, given that each item has a value and a weight. More formally, given n distinct items, where each item i has a corresponding value v_i, number of copies x_i and weight w_i, and an overall weight constraint W, the knapsack problem seeks to:

$$\begin{aligned} &\text{maximise:} && \sum_{i=1}^{n} v_i x_i \\ &\text{subject to:} && \sum_{i=1}^{n} w_i x_i \leq W \quad x_i \in \{0, \ldots, c_i\} \end{aligned} \tag{1}$$

Where c_i is an upper bound on the number of copies of each item. c_i could be viewed as a sensor quantity limit, e.g. a stock limit. The above problem is a bounded knapsack problem (BKP), which does not restrict the items in the knapsack to one copy each.

e-Energy'14, June 11–13, 2014, Cambridge, UK.
ACM 978-1-4503-2819-7/14/06.
http://dx.doi.org/10.1145/2602044.2602080.

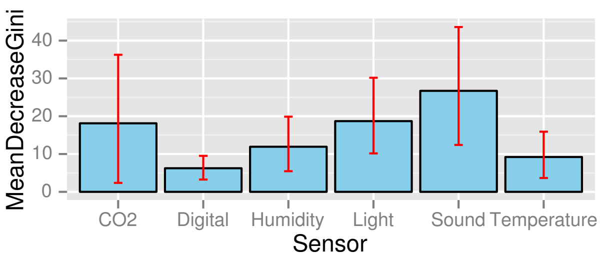

Figure 1: Mean Gini impurity decrease over all features for each sensor.

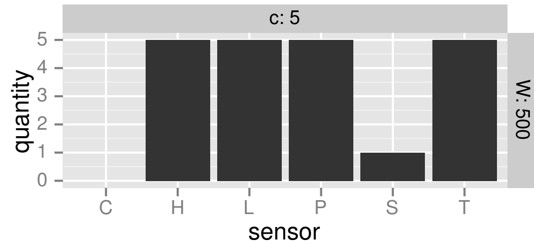

Figure 2: Example sensor set as output by the BKP algorithm for W=500 and c=5.

2.2 Defining Sensor Values and Costs

To perform feature selection, we use a random forest process on a set of extracted features. A random forest is an ensemble learning method that combines a set of decision tree classifiers, each of which is comprised of a random sample of input variables (in our case, extracted features). For brevity, we refer the reader to Breiman's description of the random forest method for a detailed overview [1]. We use random forests to measure the value of each extracted sensor feature using the average decrease in node impurities (Gini measure) from splitting the decision trees on that feature.

As with the choice of value measure, the choice of cost measure is likely to be context-dependent. An obvious choice is the financial cost of each sensor, but more complex cost functions could be designed that incorporate, for example, sensor energy costs, installation effort or sensor reliabilities. In addition to budgetary constraints, logical constraints can be introduced that restrict the chosen sensor set to particular subsets of the overall power set (all 2^n possible choices of sensor set from n sensors).

2.3 Field Study

In order to demonstrate how a sensor set for capturing energy events can be chosen, we present the results of a field study in a set of domestic buildings in the UK. We recruited 4 homes to be studied for 7 consecutive days. Within certain rooms in each home – each room common to each home – we installed the following sensors: temperature in $^\circ$C, light in lux, CO_2 in ppm, motion in $\{0, 1\}$ and sound level in dB, each sampled once per minute.

To capture a record of ground truth events in each home, we asked the primary occupant to record energy-related events around the home throughout the week in a diary study. To define the energy events, we used Oxford University's Multinational Time Use Study (MTUS) data [2], selecting domestic event codes that classify energy-consuming events around the home. We dismissed data during which the occupants did not log anything, i.e. the ground truth was unknown.

For each of the sensors, we calculated the following features: raw value at timestep k, y_k; first order difference: $\Delta(y_k) = y_{k+1} - y_k$; second order difference, $\Delta^2(y_k) = \Delta(y_{k+1}) - \Delta(y_k)$; and simple moving average over a m minute window.

3. RESULTS

For the random forest process, data is split .7 training, .15 validation and .15 test. Each forest consists of 500 trees, with 4 variables randomly sampled per split; no replacement. We used the *R* package "randomForest" [4] to run the random forest process with the aforementioned parameters. This package uses Breiman's approach [1].

For the BKP, we use Pferschy's $\mathcal{O}(nW)$ BKP algorithm described in [5]. For the moving average feature, we set m – the moving average window – to 20 minutes for each sensor. The sensor values for the BKP are set to the mean Gini decrease measures for each sensor. For the sensor costs, we use the approximate financial cost of the sensors in our study setup: 215 for CO_2, 20 for humidity, 16 for light, 115 for sound and 17 for temperature. The study participants logged 392 events in total over the 7 days (A = 119, B = 59, C = 77, D = 137).

Figure 1 shows the mean Gini impurity decrease for each sensor, averaged over the sensor's features. Figure 2 shows a set of example sensor sets output by the BKP algorithm for given weight constraints W and upper bounds on the sensor quantities c_i.

4. DISCUSSION AND CONCLUSION

The key implication of this work relates to the utilisation of environmental sensors as predictors of energy events in buildings. The sensors in our study are designed to measure a particular environmental property, e.g. temperature, rather than direct energy use – something that devices such as current clamps attached to electricity meters and plug power monitors do. The sensor values show that temperature, humidity, light, CO_2, sound and motion sensors are useful predictors of energy use, though their predictive values do vary both across sensors and between homes. The main limitations of our work relate to the context of sensing, the range of sensors and the study size.

In this paper, we presented a process for designing sensor sets to capture energy events in buildings. The key contributions lie in the use of random forests to produce a measure of sensor value *a priori*, and the implementation of a bounded knapsack problem (BKP) solver that chooses an optimum sensor set given a set of costs and values.

5. REFERENCES

[1] L. Breiman. Random Forests. *Machine Learning*, 45(1):5–32, 2001.

[2] K. Fisher and J. Gershuny. Multinational Time Use Study. Chapter 3: Activity Codes, July 2013. Retrieved from `http://www.timeuse.org/sites/ctur/files/858/mtus-user-guide-chapter-3.pdf`. Dec 2013.

[3] J. Froehlich, E. Larson, S. Gupta, G. Cohn, M. Reynolds, and S. Patel. Disaggregated end-use energy sensing for the smart grid. *Pervasive Computing, IEEE*, 10(1):28–39, 2011.

[4] A. Liaw and M. Wiener. Classification and Regression by randomForest. *R News*, 2(3):18–22, 2002.

[5] U. Pferschy. Dynamic programming revisited: improving knapsack algorithms. *Computing*, 63(4):419–430, 1999.

[6] R. Zhang, K. P. Lam, Y.-S. Chiou, and B. Dong. Information-theoretic environment features selection for occupancy detection in open office spaces. *Building Simulation*, 5(2):179–188, 2012.