Detecting Anomalous Sensor Events in Smart Home Data for Enhancing the Living Experience

Vikramaditya Jakkula and Diane J. Cook
School of Electrical Engineering and Computer Science
Washington State University,
Pullman, WA 99164-2752
cook@eecs.wsu.edu

Abstract

The need to have a secure lifestyle at home is in demand more than ever. Today's home is more than just four walls and a roof. Technology at home is on the rise and the place for smart home solutions is growing. One of the major concerns for smart home systems is the capability of adapting to the user. Personalizing the behavior of the home may provide improved comfort, control, and safety. One of the challenges of this goal is tackling anomalous events or actions. This work proposes using machine learning techniques to address this issue of detecting anomalous events or actions in smart environment datasets. The approaches are validated using real-world sensor data captured from a smart home testbed.

Introduction

Smart homes are built by adding intelligent and adaptive behavior to home automation systems. This additional capability gives the user of smart home new tools to sense and adapt to their personal needs. As the population continues to age, providing technology to maintain independent living and support the aging in place concept to healthcare is now even more important. Smart home tools are also geared to address the increased cost of healthcare by reducing the load on care providers while finding ways to prevent medical emergencies. Given the costs of nursing home care and the importance individuals place on remaining in their current residence as long as possible, use of technology to enable individuals with cognitive or physical limitations to remain in their homes longer should be more cost effective and promote a better quality of life. A range of intelligent systems built for providing healthcare and wellness enables people to live at home with an improved overall quality of life (Cook 2004).

A notable challenge to the deployment of these systems is designing anomaly detection algorithms which can improve existing techniques by identifying, and possibly filtering, rare and unexpected events. Detection of unusual events is an important issue in smart home research. However, this is a challenging task when designing an effective and computationally reasonable solution. This work demonstrates tools aimed at building a solution that detects abnormal behavior in sensor data collected in a smart home.

The role of anomaly detection is to identify rare (anomalous) events in large datasets. Classical approaches to detection of unexpected events utilize a set of expert-defined rules to detect anomalous events. Anomaly detection has grown beyond simple rules by including statistical analysis and advanced machine learning techniques.

Anomaly detection offers many benefits to smart home research, as summarized in Table 1. The most common ones include identifying rare, unexpected events which may indicate a situation of concern or interest. Additionally, filtering such events helps to improve learning algorithms, such as activity recognition, by reducing noise in the dataset. This process will also benefit adapting the smart home to the resident. Anomaly detection also plays a major role in reminder systems. Filtering anomalous events will improve a prompting system's performance. Additionally, when anomalous events are noted the user may be informed of the unexpected situation through either audio, video or message prompts. Imagine a prompting system which identifies anomalous activity and provides prompts to take correct action when required.

The need for a robust anomaly detection model is essential for a prediction model for any intelligent smart home to function in a dynamic world. For a smart environment to perform anomaly detection, it should be

capable of applying the limited experience of environmental event history to a rapidly changing environment, where event occurrences are related by temporal relations. For example, if we are monitoring the well being of an individual in a smart home and the individual has not opened the refrigerator all day as they normally do, this should be reported to the individual and the caregiver. Similarly, if the resident turned on the bathwater, but has not turned it off before going to bed, the resident or the caregiver should be notified, and the smart home could possibly intervene by turning off the water.

Anomalous event detection has unique facets and possesses a number of challenges. Out of the various open issues to address when attempting anomalous event detection, this work focuses on the problem of whether a given sensor event is anomalous in nature by using a single class support vector machine. To validate this approach, experimental data was collected from real world settings with human subject participants. The experiment conducted is detailed in the experiment evaluation section and the results observed are presented. We believe that this approach to anomaly detection performs well and should enable smarter home service provision.

Table 1. Benefits of Anomaly Detection to Smart Sensor Data

- Standardization of Smart Sensor Datasets
- Feedback to Learning Models
- Promote difference between standard data and raw data
- Reminder systems and prompting system performance improvement
- Evaluation of human lifestyles & improvement suggestions

Related Work

Recent advancements in multiple technology domains have positioned smart environments as feasible tools for assisted living, work spaces and other living spaces. This has been accomplished by advancing sensor technology, artificial intelligence, data mining, and machine learning techniques. Today, smart environments are equipped with a wide variety of sensors including motion, temperature, pressure sensors, and other intrusive/non-intrusive sensors, that allow the system to collect data on inhabitant activities and environmental situations and to later use them for automating the home.

There have been a number of smart environment research projects, such as CASAS (Deleawe 2010), MavHome (Cook 2003), the Gator Tech Smart House (Helal 2005), the iDorm (Doctor 2005), Duke smart home, and the Georgia Tech Aware Home (Abowd 2004). These university projects continue to grow alongside the industrial sector. Anomaly detection in the monitoring of elderly people to facilitate independent living and automatic adjustment of lifestyles in smart homes are known research topics and solid solutions to the anomaly detection problem are in high demand.

Anomaly detection is a relatively new field which is currently being approached and explored in smart home research. Some past approaches include a temporal-based approach where temporal relations (Jakkula 2008) identified and probabilistic models were built to evaluate and identify anomalies (Jakkula 2007). An RFID-based approach was also experimented with, for human behavior modeling and anomaly detection for elderly care. This approach presented a system for RFID data collection and preprocessing, clustering for anomaly detection, and promising experimental results (Hsu 2010). Neural network-based approaches are also investigated where predicted values are used to inform the caregiver when anomalous behavior is predicted in the near future (Ahmad 2011). Anomaly detection is used for other domains as well. One example is prior work on identifying anomalous video data to fight crime (Goldgof 2009). Additionally, there are conceptual studies done and use cases reported for abnormal events in smart environment context (Tran 2010).

Environmental sensing

A smart environment may be defined as a system that collects data about the inhabitants of a living space and the environment in order to model and adapt the environment. This allows the space to adapt to the residents and meet the goals of safety, security, cost effectiveness, and comfort. In an environment that is equipped with sensors to detect motion, temperature, and other conditions, sensed events can be captured and associated with a time stamp. The history of observed sensor events reflects activities that occur in the environment and can be used to discover frequent recurring activity patterns, to recognize activities of daily living, identify suspicious states, and to predict resident actions.

The data used for this work's experimentation was collected from real smart home test beds, with more details available in experimentation section below. The data was later stored into a database and later annotated by a human to provide a ground truth.

The sensor data collected by this system is expressed by several features, as summarized and illustrated in Table 2. There are five fields represented as follows include Date, Time, Sensor ID, Message and Annotation.

Table 2. An example for data collected from smart environment

Date – Time – Sensor – Message – Annotation – Annotation State
2009-06-15 17:07:52.312001 D031 OPEN Enter_Home begin
2009-06-15 17:07:54.921001 M006 ON
2009-06-15 17:07:58.828001 M006 OFF
2009-06-15 17:08:00.218001 M015 ON
2009-06-15 17:08:00.562001 D031 CLOSE Enter_Home end
2009-06-15 17:08:04.515001 M015 OFF

The data for the experiment is collected from three different test bed set ups with real residents and the activity is recorded as mentioned above. The resulting dataset possesses millions of data points.

Also to note is that only non-intrusive sensors were used to collect the data. The goal of these systems is to be as non-intrusive as possible and by using passive, low profile sensors the smart home is designed to allow the residents to live in their home as normally as possible.

The test beds used for this work contained a living room, dining area and kitchen. The activity level for each of the smart test beds is illustrated by Figures 1. The illustration presents activity occurrences, activity density, sensor frequency distribution, and activity time distribution for the test beds B1 which is part of the experiment.

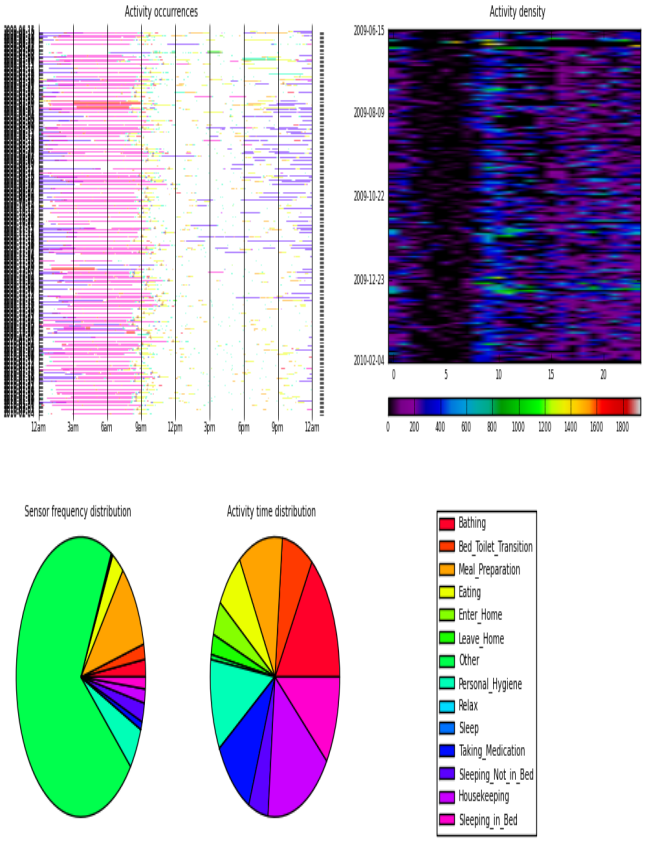

Figure 1. Smart test bed Code name "B1" activity pattern

Methodology

The One Class Support Vector Machines (OCSVM) is quite popular for anomaly detection problems. Suppose that a dataset has a probability distribution P in the feature space. The goal would be to find a "simple" subset S of the feature space such that the probability that a test point from P lies outside S and is bounded by some a priori specified value (Schölkopf 2001).

Supposing that there is a dataset drawn from an underlying probability distribution P, one needs to estimate a "simple" subset S of the input space such that the probability that a test point from P lies outside of S is bounded by some a prior specified v ∈ (0, 1) . The solution for this problem is obtained by estimating a function f which is positive on S and negative on the complement S. The algorithm can be summarized as mapping the data into a feature space H using an appropriate kernel function, and then trying to separate the mapped vectors from the origin with maximum margin (see Figure 2).

$$f(x) = \begin{cases} +1 & \text{if } x \in S \\ -1 & \text{if } x \in \overline{S} \end{cases}$$

In our context, let x1, x2, . . . , xn be training examples belonging to one class X, where X is a compact subset of RN. Let Φ : X → H be a kernel map which transforms the training examples to another space. Then, to separate the data set from the origin, one needs to solve the following quadratic programming problem (Schölkopf 2000):

$$min \frac{1}{2}\|w\|^2 + \frac{1}{vl}\sum_{i=1}^{l}\xi_i - \rho$$

subject to

$$(w \cdot \Phi(x_i)) \geq \rho - \xi_i \quad i = 1, 2, ..., l \quad \xi_i \geq 0$$

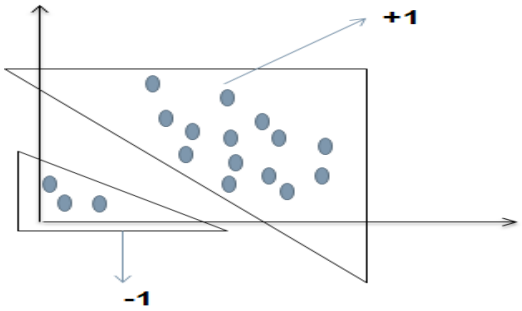

Figure 2. One Class Support Vector Machine

Once transformed to a different space, the data points which are closer to the origin are identified as anomaly and reported. For this experiment we use LIBSVM via weka tool (Mark 2009) with default parameters.

Experiment Evaluation

The experiment done in this work consisted of parsing the data, training the learning algorithm and testing it against test data. After the testing data is passed through the classifier observations about its performance and capabilities are made.

The core component of this experiment is evaluating the one class support vector machine as an anomaly detection tool. This technique has been successfully applied in various domains and had good results (Varun 2009). The training data consists of annotated data while the testing data consists of annotation free data. Annotated data is clean data without anomalies and is cleared of any outliers by using Interquartile range filter available via the Weka tool. (Figure 3)

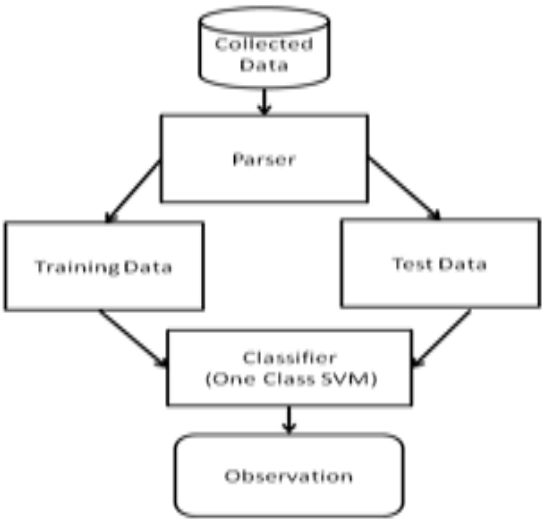

Figure 3. Experimentation Process

The test data consists of annotation free samples. The data used for the experimentation is created by a parser tool which improves the dimensionality of the dataset by introducing additional attributes. These attributes include the daily count of sensor occurrence for a particular day, monthly count of sensor occurrence for a particular month, and yearly count of sensor occurrence for a particular year.

Results and Discussion

The LIBSVM algorithm available via Weka was used for this work. This is an integrated tool for support vector classification and regression which can handle one-class SVM using the Sholkopf algorithms. The standard parameters of the algorithm were used.

The training data consists of annotated data. We test using the data with positive samples and report the findings. We assume the test set to be positive samples observe the false positives and report the observations. A positive sample consists of no annotations and negative samples are anomalies. Ground truth is having anomaly identified in the dataset during data collection or annotation phase. Type I error are false positives which help us identify the normal class being misclassified as anomaly. For the experimentation the RBF kernel with default parameters was used.

The evaluation metrics used for this experimentation include precision, recall and F-measure and type I, type II errors (Wikipedia 2011) (Varun 2009). Generally both Type I and Type II errors are used for performance evaluation, however we use type I errors as a performance measure for this experimentation, and table 3 shows the observations on the test set run of the experiment.

Table 3. Experiment Observation on Test Set

		Test Set
B1	Type I	0.5
	Type II	-
	Precision	1
	Recall	1
	F- Measure	1
B2	Type I	0.4
	Type II	-
	Precision	1
	Recall	1
	F- Measure	1
B3	Type I	0.5
	Type II	-
	Precision	1
	Recall	1
	F- Measure	1

Given that the ground truth is not provided for testing, we train the SVM with positive samples and provide positive samples to test; hence type II errors are not considered. Positive samples are those with no annotations and with no ground truth provided they are being considered as positive samples. We should note that the results reported are observations directly from weka and type I errors are observations calculated based on classifier performance on test set. We should note that anomalies can also be caused by erroneous readings due to sensor failure, but we do not include these scenarios in our current experimentation.

The results are presented in table 3, and lead way to the next steps where we increase the dimensionality and vary the kernels and hyper-parameters to observe the performance with ground truth-based anomalous datasets.

In future work, the plan is to extend this work by introducing multiple one class support vector machines. This would provide a SVM for each annotation to catch anomalies. This approach is a major step for anomaly detection in smart sensor datasets.

Conclusion

In today's world living smarter is more meaningful than ever. Long lasting and sustainable living is possible thanks to technology in everyday life. A robust anomaly detection framework is a niche area, and the resulting tools from this research area may be used to enhance the overall experience in a smart home setting by maximizing user adaptation, identifying issues in lifestyle, raising alerts, enhancing reminder system, and assist prompting systems. The approach in this paper is an initial step towards anomaly detection in smart home data which looks promising. Some future steps would include increasing the dimensionality of the SVM and to evaluate the use of the multiple one class support vector machines approach where we build an one class support vector machine for each annotation and see if an event is anomalous or not. Anomaly detection adds value to smart home systems and has immense potential for a smarter living framework.

Acknowledgement

Our sincere thanks to Aaron Crandall, for his time and effort in the form of review and feedback. This work was supported by Bosch Research LLC.

References

D. Cook and S. Das. 2004. Smart Environments: Technology, Protocols and Applications. Wiley Series on Parallel and Distributed Computing. Wiley-Interscience.

S. Deleawe, J. Kusznir, B. Lamb, and D. Cook. 2010. Predicting air quality in smart environments. Journal of Ambient Intelligence and Smart Environments, 2(2):145-154.

D. Cook, M. Youngblood, I. Heierman, E.O., K. Gopalratnam, S. Rao, A. Litvin, and F. Khawaja. 2003. Mavhome: an agent-based smart home. In Proceedings of the First IEEE International Conference on Pervasive Computing and Communications, pages 521-524.

S. Helal, W. Mann, H. El-Zabadani, J. King, Y. Kaddoura, and E. Jansen. 2005. The gator tech smart house: A programmable pervasive space. Computer, 38(3):50-60.

F. Doctor, H. Hagras, and V. Callaghan. A fuzzy embedded agent-based approach for realizing ambient intelligence in intelligent inhabited environments. 2005. IEEE Transactions on Systems, Man and Cybernetics,Part A: Systems and Humans, 35(1):55-65.

G. Abowd and E. Mynatt. 2004. Smart Environments: Technology, Protocols, and Applications, chapter Designing for the human experience in smart environments. pages 153-174. Wiley.

Vikramaditya R. Jakkula, Diane J. Cook, and Aaron S. Crandall. 2007. Temporal pattern discovery for anomaly detection in smart homes. Proceedings of the the 3rd IET International Conference on Intelligent Environments (IE 07), Germany.

Hui-Huang Hsu, and Chien-Chen Chen. 2010. RFID-based human behavior modeling and anomaly detection for elderly care. Mobile Information Systems. Volume 6 Issue 4 341-354.

Vikramaditya Jakkula and Diane J. Cook. 2008. Anomaly Detection Using Temporal Data Mining in a Smart Home Environment. Methods of Information in Medicine,Smart Homes and Ambient Assisted Living special issue.

Ahmad Lotfi, Caroline Langensiepen, Sawsan M. Mahmoud, and, M. J. Akhlaghinia. 2011. Smart homes for the elderly dementia sufferers: Identification and prediction of abnormal behavior. Journal of Ambient Intelligence and Humanized Computing. Springer-Verlag.

Goldgof, D.B., Sapper, D., Candamo, J., and Shreve, M. 2009. Evaluation of Smart Video for Transit Event Detection/ Report No. 2117-7807-00 prepared by National Center for Transit Research, for Florida Department of Transportation and Research and Innovative Technology Administration.

An C. Tran, Stephen Marsland, Jens Dietrich, Hans W. Guesgen, and Paul Lyons. 2010. Use cases for abnormal behaviour detection in smart homes. In Proceedings of the Aging friendly technology for health and independence, and 8th international conference on Smart homes and health telematics (ICOST'10), Yeunsook Lee, Z. Zenn Bien, Mounir Mokhtari, Jeong Tai Kim, Mignon Park, Jongbae Kim, Heyoung Lee, and Ismail Khalil (Eds.). Springer-Verlag, Berlin, Heidelberg, 144-151.

B. Schölkopf, J. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson. 2001. Estimating the support of a high-dimensional distribution. Neural Computation, 13, 1443-1471.

B. Schölkopf, A. Smola, R. Williamson, and P. L. Bartlett. 2000. New support vector algorithms. Neural Computation, 12, 1207-1245.

Varun Chandola, Arindam Banerjee, and Vipin Kumar. 2009. Anomaly detection: A survey. ACM Comput. Surv. 41, 3, Article 15 (July 2009), 58 pages. DOI=10.1145/1541880.1541882 http://doi.acm.org/10.1145/1541880.1541882

Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten . 2009. The WEKA Data Mining Software: An Update; SIGKDD Explorations, Volume 11, Issue 1.

Wikipedia.2011.http://en.wikipedia.org/wiki/Receiver_operating_characteristic. Retrieved March 26, 2011.