Context Awareness Computing in Smart Spaces Using Stochastic Analysis of Sensor Data

Jae Woong Lee[1] and Sumi Helal[2]

[1] Department of Computer Science, SUNY Oswego, Oswego, New York
Jaewoong.lee@oswego.edu
[2] School of Computing & Communications, Lancaster University, Lancaster, UK
s.helal@lancaster.ac.uk

Abstract. In building a smart space, it becomes more critical to develop a recognition system which enables to be aware of contexts, since the appropriate services can be provided under the accurate recognition. As services satisfying for desires of individual human residents are more demanding, the necessity for more sophisticated recognition algorithms is increasing. This paper proposes an approach to discover the current context by stochastically analyzing data obtained from sensors deployed in the smart space. The approach proceeds in two phases, which is to build context models and to find one context model matching the current state space, however we mainly focus on the phase building context models. Experimental validation supports the approach and approved validity.

Keywords: Smart Spaces · Context Awareness Computing · Sensors · Conditional Probability Table · K-means Clustering · Principal Component Analysis

1 Introduction

During the last decade, we have been experiencing dramatic transformations in our paradigm of daily living life. Sensors are everywhere to monitor our activities and to enable to recognize contexts, and actuators and smart devices are operated to provide necessary and/or convenient services to us. In the smart space deployed with such devices, many chores are unawarely performed and complicated works are simplified. With outstanding achievements in internet technology, sensors and actuators are integrated into more intelligent devices, which is now known as Internet of Things (IoT) devices. The advent of ubiquitous sensors and IoT is continuously making our living space smarter.

The performance of smart spaces depends on various factors including sensor technology and IoT technology, but this paper addresses recognition of contexts. A smart space provides appropriate services to human residents after recognizing the current context. The recognition process demands accurate analysis of sensor data since the context is realized by the sensors attached in the space. It requires building well-structured profiles for contexts, which can be easily managed and efficiently manipulated.

This paper proposes an approach for context awareness computing utilized by stochastic analysis of sensor data. The key idea comes from two observations: sensors which are triggered in a given context are mostly related, and particular sensors are

important and contributive to become the context. Hence, a whole dataset can be divided into a certain number of groups, each of which contains related sensors. Due to the similarity, the structure of the context model developed in our context-driven simulation approach is adapted here with minor changes [1]. Note that the context model requires finding the important sensors.

The main framework of the approach consists of two phases: in the first phase, context models are built and then the context which describes the present state of the space most similarly is discovered among the models in the second phase. In the first phase, the context models are derived from given sensor datasets which were generated under supervised learning. In this learning phase, three methods used in machine learning and statistics are utilized. First, two methods of Conditional Probability Table (CPT) and K-means clustering enables to find k numbers of groups, each of which is declared as a context. Last, Principal Components Analysis (PCA) returns important and representative sensors per each context, which finally define the complete context model. Once the context models are built, one context will be discovered by comparing the present state space and all the context models. At this comparison step, two methods of calculating Euclidean distances and cosine similarity are utilized. Due to the page limit, and this paper addresses only the first phase of creating the context models.

This paper is organized as follows. In the section 2, we describe existing work related to context awareness computing and statistical analysis methods. And the principles of the proposed approach are overviewed in the section 3, which is followed by experimental validation. We conclude the paper with a short discussion and future work plan.

2 Related Work

There were many context models proposed in various areas – human computer interaction[2], context awareness computing [3], and activity recognition learning [4]. The proposals attempted to model contexts from the current state of the space which was observed through human senses or electronic sensors. In some research, the context models were defined in the form of rules. In Context-Aware Simulation System for smart home (CASS) [5], for instance, the system defined rules to describe certain conditions, detected the conflicts of rules, and provided the ability to control a character to move it. In Context-driven simulation approach [1], contexts were defined as abstract and representative state spaces by specifying related sensors and their status. Additionally, their context models defined the causality in between other contexts, which enabled to generate the entire daily living scenario.

The research commonly oriented the methods on matching the current state space with manually predefined context models [6]. Ontology facilitated efficient modeling and reasoning for context [7], however it still needed humans' efforts in configuration. To avoid the burdens and increase automaticity, research in activity recognition on deriving meaningful high-level information from low-level information could be used. The goal of the research is to cluster from collected sensor datasets. In CBARS [8], a supervised learning model was built first, and unsupervised learning for new data was applied for new activity recognition. The challenge was that CBARS needed a supervised learning model. AALO [9] addressed that challenge. AALO is an active recognization system that can accurately classify specified activities according to locations

and times in which the activities are performed. CBCE [10] proposed a method for combing multiple classifiers including Naïve Bayes (NB) models, hidden Markov models (HMMs), and conditional random fields (CRFs). This ensemble of classifiers can recognize activities in given sensor datasets, however, they do not provide a method to define abstract information of context to represent the other state spaces in a cluster.

3 Principles of Approach

The principal ideas in the defining context models are 1) to cluster sensor datasets into groups which have related sensors, and 2) to discover particular sensors and their values which can represent each group. The values of the sensors for each context are formed in context conditions, since the context begins only if the sensors have the values.

3.1 Context Model

Before we dive further into the details of our approach, it would first be helpful to define context. In the context-driven simulation approach [1], it is an abstracted state space envelope that represents consecutively occurring state spaces. A context is intended to represent an important and meaningful state space in the group of relevant state spaces with respect to activities. It is described by three properties: context conditions, which express conditions to enter the context, context activities, which are activities available in the context (for play back of some of them), and next contexts, which can be transitioned to after activities are performed. In context-awareness computing, context conditions only are needed to define a context.

3.2 Overall Approach

The approach proceeds in three steps, each of which utilizes a statistical method. First, the number of contexts is decided by Conditional Probability Table (CPT), and then meaningful and representative state spaces are defined as contexts by K-means Clustering. Finally, important sensors which contribute to become each context are discovered by using Principal Component Analysis.

Deciding the Number of Contexts by CPT. In order to find the number of contexts, we first capture the probability of consecutive occurrence of each pair of different sensors in the datasets. The idea is that sensors in a context are related and thus the occurrence probability of each other is fair high. In other words, if an occurrence probability of a pair of sensors is low, the sensors are not in a context. This probability can be accurately calculated from the frequency of occurrence of consecutive sensor events. These conditional probabilities are arranged as a ξ x ξ table (ξ being the number of sensors), which is called the Conditional Probability Table (CPT).

CPT is used as the probabilistic fingerprint of the entire dataset. A pair of sensor events with high conditional probability usually contains sensor events that are related and associated together. They could belong to the same context, and therefore are highly likely to occur together in this order. On the other hand, if the pair has low conditional

probability, its sensor events are considered unrelated and would rarely occur together. A pair of sensor events with low probability indicates the end of a context and the start of another. Therefore, we divided the dataset between sensor events e_i and e_{i+1}if the conditional probabilities of e_i and the first sensor event e_i satisfy the condition $p(e_i)$ ≤ θ * $p(e_i)$, where θ is a parameter that represents the extent to which ei relates to e_i. We set θ to 0.5 in the experiments. Using this method, we divided the dataset into k groups, which are considered as the number of contexts.

Defining the Contexts by K-means Clustering. By our observation, a meaningful state space is sufficiently distant from other meaningful state spaces, but could be close to other relevant yet non-meaningful state spaces. To find which state spaces are meaningful, all are partitioned into k clusters, in which each state space belongs to the cluster with the nearest mean. Therefore, the universal set of state spaces $S_U = \{S_1, \dots, S_i, \dots, S_\omega\}$ is divided into $\{\hat{S}_1, \dots, \hat{S}_i, \dots, \hat{S}_\omega\}$, where S_i is a state space and $\hat{S}_i$ is a cluster of state spaces. Each cluster $\hat{S}_i$ minimizes the sum of distances between the within-state space and the mean according to the following formula:

$$\arg\min_{S_U} \sum_{i=1}^{k} \sum_{S_t \in \hat{S}_j} \|S_t - \mu_i\|^2, \tag{1}$$

where S_i means a state space in cluster $\hat{S}_i$. After S_U is classified into k clusters, cluster centroids are considered meaningful state spaces and are candidates for contexts.

Discovering Context Conditions by PCA. The centroid in context is representative and meaningful, but has insufficient information to define the context. We observed that multiple sensors usually contribute to begin a context. Principal Components Analysis (PCA) enables to discover those important sensors. Once we find the relevant sensors via the stochastic analysis of sensors' high-dimensional data, the original dataset can be projected onto lower-dimensional data. The process is repeated for each cluster and the remaining data is used to build context conditions.

Principal components are sensors that show definite variance patterns that explicitly express the change of states. We want to know in which pattern the dataset is scattered. For this, a matrix of covariances (*cov*) is calculated first. In a ξ-dimensional dataset, covariance *cov* is calculated as

$$cov\,(\hat{s}^i, \hat{s}^j) = \frac{\sum_{k=1}^{\xi}(\hat{s}_k^i - \mu_i)(\hat{s}_k^j - \mu_j)}{(\xi - 1)}, \tag{2}$$

where $\hat{s}^i$ and $\hat{s}^j$ are the set of sensor values in dimensions i and j, repsectively; i and j are the sensor values in each dimension. The total covariances establish a $\xi \times \xi$ covaraince matrix R, shown in the equation 3.

From covariance matrix, we calculate the eigenvectors, each of which can conduct linear transformations of sensor data and characterize its variance; the eigenvalues then measure how well the sensor data is scattered. We choose the eigenvectors that show the most variant spread of data as principal components. If data is evenly scattered with

an axis transformed by an eigenvector (i.e., the data pattern is reconized explicitly), it is an important eigenvector, which means it's the desired principal component and has a high eigenvalue. The challenge is in determining the threshold for which eigenvalues are high enough to be acceptable. We propose threshold Θ_e for total eigenvalues of selected eigenvectors. In our approach, first the eigenvectors are sorted by eigenvalues in descending order; then, eigenvectors with higher values are chosen until the sum of corresponding eigenvalues exceeds Θ_e.

$$R\left(\hat{\mathrm{S}}\right) = \begin{bmatrix} cov\left(\hat{\mathrm{s}}^1, \hat{\mathrm{s}}^1\right) & \dots & cov\left(\hat{\mathrm{s}}^1, \hat{\mathrm{s}}^\xi\right) \\ \vdots & \ddots & \vdots \\ cov\left(\hat{\mathrm{s}}^\xi, \hat{\mathrm{s}}^1\right) & \dots & cov\left(\hat{\mathrm{s}}^\xi, \hat{\mathrm{s}}^\xi\right) \end{bmatrix} \quad (3)$$

Eigenvectors satisfying the condition establish a feature matrix. The original high-dimentional dataset is transformed into a low-dimensional dataset through the feature matrix. Context conditions are created by collecting all sensor and forming them in a range. For instance, the expected condition for s_i, those values are 1, 4, and 2.5, is $1 \leq s_i \leq 4$, which covers all the values.

4 Experimental Validation

To evaluate the performance of the proposed approach, we conducted a few experiments. For this experiment, we first obtained the context models from sensor datasets for 10 days by applying the approach. Then we synthesized sensor datasets by running Persim 3D [11] with the context models. Persim 3D adapted the context-driven simulation approach, and generated sensor data from a scenario which is described by the sequence of contexts. The key of the validation is to compare the actual dataset and its synthetic dataset and to show similarity in between. If the approach is valid and thus the contexts are correctly discovered, Persim 3D should generate the similar dataset as the actual dataset. Therefore, our validation goal was to compare the generated dataset with actual dataset.

For this purpose, we built statistic models of an actual dataset that apply a Bayesian network. The Bayesian network enables to calculate join probability distribution on the entire dataset by using CPT. If the occurrence probability of the simulated dataset is not similar to those of the actual dataset, it says that the dataset was not correctly generated. However, the probability based on sensor events becomes very low as multiplying probability of pairs of sensor events, thus we utilized Activity Playback Model which enables to describe a dataset in the activity level. It prevents the probability from diminishing. Table 1 shows the occurrence probabilities of each dataset. In most experiments, simulated datasets show high similarity and the average similarity is 70.95%.

5 Conclusion

The approach based on stochastic analysis of sensor data reduces humans' efforts in processing recognition of the current context and increases automaticity of the process. Through the experiments, it validly shows the good performance. Our next research

will concentrate on developing more efficient statistic methods which can improve the performance. We will also research on unsupervised learning methods, which thus are able to detect contexts without training. It will relate to the real-time recognition.

Table 1. Similarity of occurrence probability of generated dataset based on our approach against actual dataset. Note that different *k* is applied on Nov/05.

Dataset	Actual Dataset	# of Context (K)	Simulated Dataset	Similarity
Nov/03	0.150	5	0.124	82.55%
Nov/04	0.150	5	0.124	82.55%
Nov/05	0.033	4	0.234	70.75%
Nov/07	0.150	5	0.124	82.55%
Nov/10	0.150	5	0.124	82.55%
Nov/11	0.090	5	0.045	50.03%
Nov/13	0.090	5	0.045	50.03%
Nov/14	0.150	5	0.124	82.55%

References

1. Lee, J. W., Helal, A. Sung, Y., Cho, K.: A Context-driven Approach to Scalable Human Activity Simulation. In: ACM SIGSIM Conference on Principles of Advanced Discrete Simulation, pp. 373--378, ACM, New York (2013)
2. Fischer, G.: User Modeling in Human-Computer Interaction. J. of User Modeling and User-Adapted Interaction, 11(1-2), 65--86 (2001)
3. Salber, D., Dey, A., Abowd, G.: The Context Toolkit: Aiding the Development of Context-Enabled Applications. In: Conference on Human Factors in Computing Systems, pp. 434--441, ACM New York (1999)
4. Hasan, M., Roy-Chowdhury, A.; Context Aware Active Learning of Activity Recognition Models. In: IEEE International Conference on Computer Vision, pp. 4543--4551, IEEE Express, Washing DC. (2015)
5. Park, J., Moon, M., Hwang, S., Yeom, K.: CASS: A Context-Aware Simulation System for Smart Home. In: 5th ACIS International Conference on Software Engineering Research, Management & Applications, pp. 461--467, IEEE Express (2007)
6. Lee, J. W., Helal, S.: Inferring Context from Human Activities in Smart Spaces. In: 29th International FLAIRS Conference, pp. 695--701, AAAI Press, Florida (2016)
7. Wang, X.H., Zhang, D.Q., Gu, T., Pung, H.K.; Ontology Based Context Modeling and Reasoning Using OWL. In IEEE Annual Conference on Pervasive Computing and Communications Workshops, pp. 18--22, IEEE Washing DC., (2004)
8. Abdallah, Z. S., Gaber, M. M., Srinivasan, B., Krishnaswamy, S.: CBARS: Cluster Based Classification for Activity Recognition Systems. In: International Conference on Advanced Machine Learning Technologies and Applications, pp. 82--91. Springer, Berlin (2012)
9. Hoque, E., Stankovic, J.: AALO: Activity recognition in smart homes using active learning in the presence of overlapped activities. In: International Conference on Pervasive Computing Technologies for Healthcare, pp.139--146. IEEE Express (2012)
10. Jurek, A., Nugent, C., Bi, Y., Wu, S.: Clustering-based ensemble learning for activity recognition in smart homes. Sensors 14(7), 12285--12304 (2014)
11. Lee, J. W., Cho. S., Liu, S., Cho, K., Helal, S: Persim 3D: Context-driven simulation and modeling of human activities in smart spaces. IEEE T. on Automation Science and Engineering 12(4), 1243--1254 (2015)