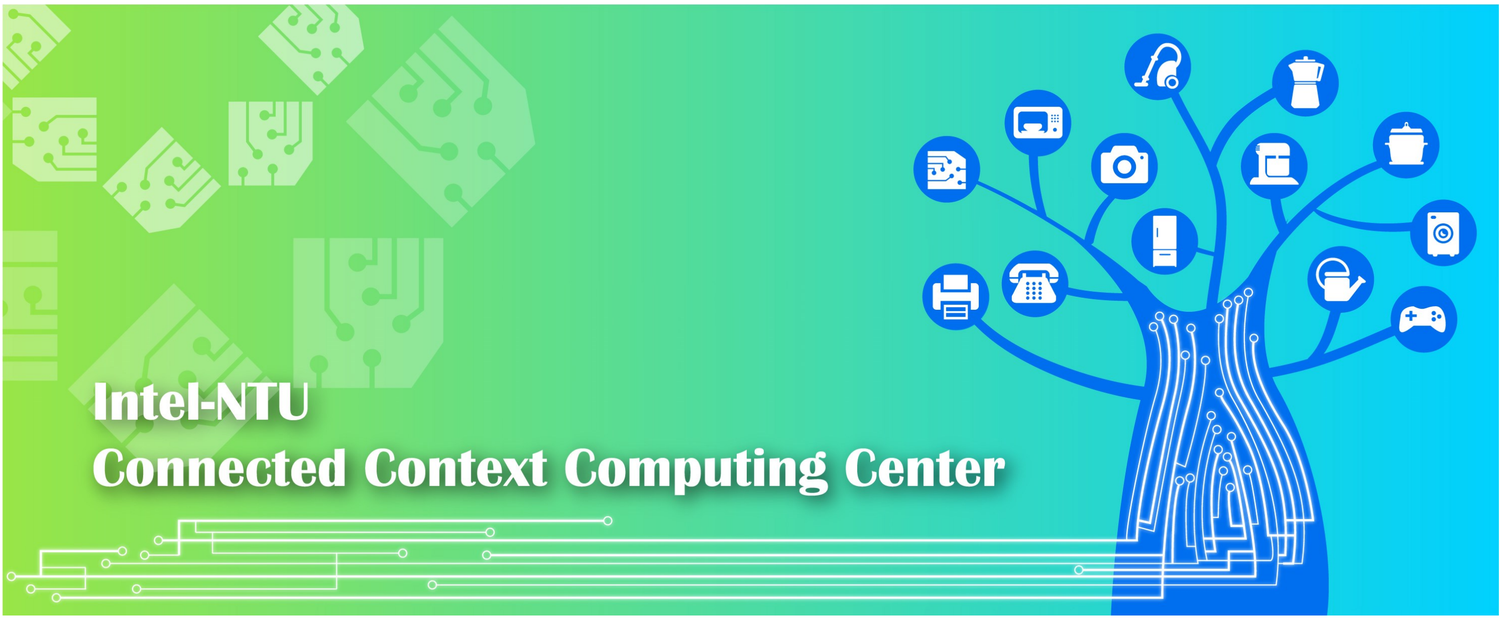

M2M-based Anticipatory Reasoning for Contexts

Newsletter No. 52 October 2015

Introduction

With the support of innovative technologies like Internet of Things (IoT), heterogeneous sensors are pervasively deployed in the smart environments, and Big Data are generated real-time in our daily life from all sorts of human activities, environment changes, and interactions between humans and environments. One of the most important smart environments generating such Big Data is smart homes, and the analytics of these Big Data can help to make smart homes human-centric by providing context-aware services. To be specific, for human-centric smart homes to provide context-aware services, one of the most important analytics for Big Data from smart homes is to understand human needs by recognizing contexts (e.g. user activities).

In general, data need to be preprocessed (e.g. labeling) in a supervised way, whereby smart homes can apply some artificial intelligence (AI) techniques to derive satisfactory results based on the analytics mentioned above. Conventionally, these preprocessing procedures are done manually, which thus requires lots of human efforts. However, since these Big Data from smart homes are generated real-time with huge quantity, it is difficult for humans to preprocess these data efficiently. As a result, it also makes smart homes difficult to obtain useful information from Big Data to become human-centric. Given that the conventional supervised way is non-preferable, we here propose an M2M-based approach for reasoning the contexts, namely, an unsupervised method to deal with these Big Data from smart homes. Without asking humans to label the (context) data and to specify their preferences, the hereby proposed method can automatically discover the underlying contexts (e.g. user activities) from the Big Data generated in smart homes.

System Architecture

The proposed system architecture is shown in Fig. 1. First, the data collected with each sensor in a smart home are clustered by the proposed Density-enhanced Affinity Propagation (DAP) for quantization as well as noise removal. Next, all the sensors are grouped according to their locations (e.g. living room, study room, etc.) in a home environment and data from the sensors deployed at the same location are combined and then clustered by DAP to find important signatures for each location (i.e. room-level potential contexts), represented in terms of specific combinations of data from all sensors involved in that location. Finally, these signatures from each location are combined and then clustered by DAP to find important signatures across locations in a smart home (i.e. home-level potential contexts). Next, we will describe the problem formulation.

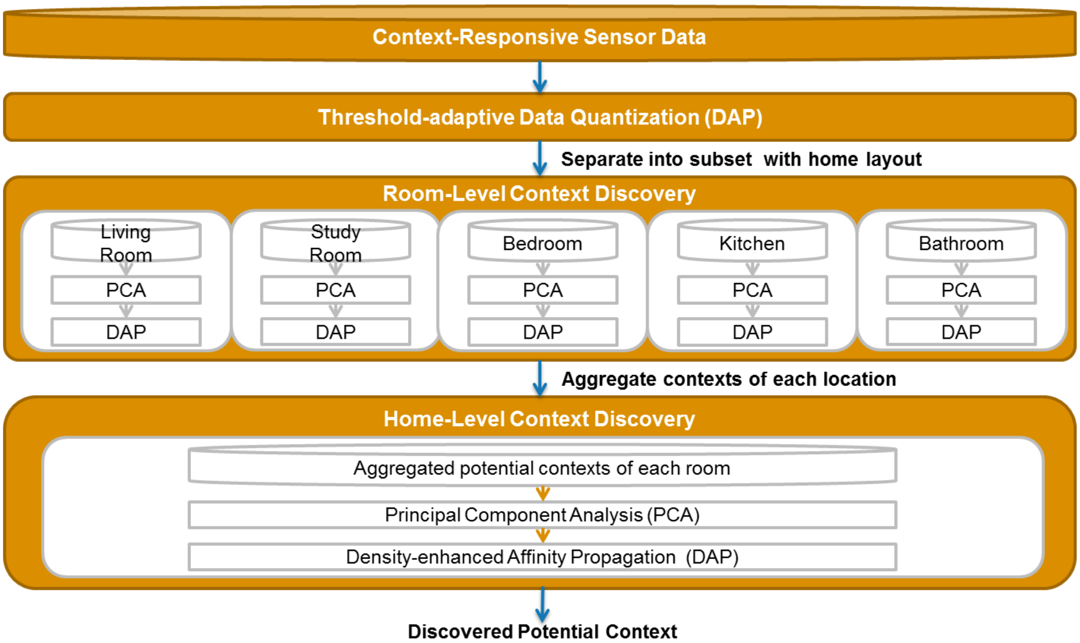

Fig. 1. System Architecture of M2M-based Anticipatory Reasoning for Contexts.

Problem Formulation

Assume that there are X kinds of contexts (e.g. activities) in a smart home, denoted as c1, c2, ..., cX, and they are composed of K mixture context components to produce sensor readings. Each component is parameterized by θk with weight πk, where ∑πk=1. Given N observations {z1, ..., zN} from a smart home, then p(z)=∑k πk * f(z|θk), where f(z|θk) is the mixture context component parameterized by θk. With M sensors deployed in a smart home, each observation is further extended as zi={z(i,1), ..., z(i,M)}, parameter for each component over different sensors is also extended as θi=(θ(i,1), ..., θ(i,M)), and p(zm)=∑km π(m,km) * f(zm|θ(m,km)) with ∑π(m, km)=1.

From the viewpoint of Topic Model, each sensor has its own Km characters represented by {δ(m,1), ..., δ(m,Km)}. Furthermore, each observation is an artificial word (δ(1), …, δ(M)) where δ(i) is one of (δ(i,1), ..., δ(i,Ki)). Each artificial word is generated from a major word Wi, and each major word Wi is generated by one of K mixture components. And finally, a context is a combination of one or multiple artificial major words from K mixture components.

The proposed system architecture is to first cluster data within each sensor to find Km characters, then to compose characters as words and cluster them as major words to find K mixture components, and finally to compose major words as potential contexts.

Experimental Results

In the experiment to evaluate our method to do context discovery, we have collected a single -user dataset that contains 3 individual users with 16 hour of daily life routines.in our smart home with 7 defined single-user activities as shown in Table I. The number of instances is 54083.

Dataset	Number of Activities	Number of Instances	List of Activities
Single-User	7	54083	Reading, Sleeping, Studying,UsingNoteBook, UsingPC, WatchingTV, PlayingKinect

TABLE I. DATASETS USED IN EXPERIMENT

Table II shows the confusion matrix of the discovered contexts. There are 10 contexts (each of which is a potential activity) are discovered with 7 single-user activities. It can be observed that almost all the user activities are clustered with 100% precision as some context. Two activities, which are "Watching TV" and "Playing Kinect" are confused. About 21% of "Watching TV" instances are clustered as "Playing Kinect" activity due to the reason that TV is used in both activities, and all these data instances result in the activation of the same sensor in both cases except for the sensor for Kinect. Therefore, when a user switches between these two kinds of activities, sometimes the sensor for Kinect is not activated instantly, and this sensing delay results in the wrong result of clustering from the viewpoint of human beings, but not the viewpoint of data. The other activity instances are clustered to different clusters with 100% precision. For the activity of "Sleeping", there are 2 clusters belonging to the activity of "Sleeping" with 100 precision. The activity of "Using PC" is clustered into 3 clusters with 100% precision. The reason that an activity is clustered into different clusters is that the user habit for an activity (e.g. "Sleeping" or "Using PC") may slightly differ.

TABLE II. THE CONFUSION MATRIX OF SINGLE-USER ACTIVITY

Activity	Potential Activity									
	1	2	3	4	5	6	7	8	9	10
Reading	99.92	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Sleeping	0.00	100.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Studying	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00
UsingNB	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00
UsingPC	0.00	0.00	0.00	0.00	0.00	100.00	100.00	100.00	0.00	0.00
WatchingTV	0.08	0.00	0.00	0.00	0.00	0.00	0.00	0.00	75.95	21.71
PlayingKinect	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	24.05	78.29

Summary

In this work, we propose an unsupervised nonparametric method to do analytics with a framework for human-centric smart homes to automatically discover contexts. The models resulting from the proposed analytics can be used to build models to identify the potential contexts. Our proposed analytics method is suitable for Bid Data analysis for human-centric systems like smart homes, since it does not require users to assign a specific number as the cluster parameter to train the models. The experiment has a promising result and has demonstrated the effectiveness of out proposed method in a real living environment.

About the Authors

Dr. Li-Chen Fu, Distinguished Professor

Dept. Electrical Engineering and Dept. Computer Science & Information Engineering, NTU

Dr. Chao-Lin Wu , Assistant Research Fellow

Intel-NTU Connected Context Computing Center

INC MOST Intel-NTU Connected Context Computing Center

BL-7C, Barry Lam Hall, 1 Roosevelt Road, Sec. 4
National Taiwan University, Taipei, 10617 Taiwan
Tel: +886-2-3366-1467 http://ccc.ntu.edu.tw

Acknowledgments

This work was supported in part by the Ministry of Science and Technology, National Taiwan University, and Intel Corporation under Grants MOST 103-2911-I-002-001, NTU-ICRP-104R7501, and NTU-ICRP-104R7501-1