

# DATA STORAGE AND DATA MINING OF BUILDING MONITORING DATA WITH CONTEXTS

Nobuyoshi Yabuki<sup>1\*</sup>, Yoshihiro Yoshida<sup>2</sup>, Shota Tsukamoto<sup>3</sup>, and Tomohiro Fukuda<sup>4</sup>

<sup>1</sup> *Professor, Division of Sustainable Energy and Environmental Engineering, Osaka University, Suita, Japan*

<sup>2</sup> *CTI Engineering, Co., Ltd., Tokyo, Japan*

<sup>3</sup> *Graduate Student, Division of Sustainable Energy and Environmental Engineering, Osaka University, Suita, Japan*

<sup>4</sup> *Associate Professor, Division of Sustainable Energy and Environmental Engineering, Osaka University, Suita, Japan*

\* *Corresponding author ([yabuki@see.eng.osaka-u.ac.jp](mailto:yabuki@see.eng.osaka-u.ac.jp))*

**ABSTRACT:** In order to improve the energy efficiency of buildings, many sensors have been being installed in buildings for monitoring temperature, humidity, electricity consumption, etc. In the Building Energy Management Systems (BEMS), it seems that visualization of the monitoring data is the main objective and that managers still have to find problems from the visualized data with drawings of buildings as contexts. As the number of sensors and amount of data increase, it will be difficult to find problems. Although, data mining would be an effective means to automate this process, two problems exist. The first is despite the necessity of contextual data of the surrounding environment of sensors to discover meaningful knowledge, it is difficult to define the data model to store monitoring data and contexts beforehand and it is very difficult to change the data model afterward. The second problem is that simple data mining techniques are not effective for time-series and massive monitoring data. Thus, in this research, first, a flexible and variable data modeling technique, named A&A method, was proposed to incorporate various and vague contextual data into the sensor data model. Next, Data Mining for Sensor data of buildings using Contexts (DMSC) method was developed to support building managers to discover useful knowledge from a large amount of sensor data for better building energy management. To check the feasibility of the proposed methods, we stored actual monitoring data in our laboratory at Osaka University with the developed data model and executed data mining, using DMSC method.

**Keywords:** *Data Mining, Knowledge Discovery, Context, Monitoring Data, Building, Sensor, Energy Efficient Building*

## 1. INTRODUCTION

In order to discover useful knowledge from a large amount of data observed by many sensors to be installed in the near future, data mining seems to be an effective technique on the surface. However, simple application of data mining technique to sensor data may not be so successful as expected because sensor data are mostly mere numerical values. Thus, contextual data should be incorporated in the database for data mining as well as sensor data [1]. However, it is difficult to define the data model of sensors and contexts beforehand and it is even more difficult to change the data model once the data storage is begun.

Second problem is that the effective well-established data mining technique has not been developed yet for discovering useful and meaningful knowledge from a large amount of sensor data. Therefore, the objective of this research is to develop a flexible and variable data modeling technique and to develop a data mining technique incorporating contextual data with sensor data.

## 2. A&A METHOD

Since input data for most data mining algorithms are formulated as tables, relational database in which all data are stored as multiple tables seems to be suitable for data

mining. However, sensor data are stored in various formats. Furthermore, it is difficult to make a data model which can be applied to all environments where sensors are installed because the sensor application situations are numerous and various. Hence, the sufficient flexibility is required for a data model for data mining.

To enhance the flexibility of relational data model, “Addition & Absorption (A&A) method” was proposed in this research. In this method, hypothetical tables are generated linking the essential tables such as “Element”, “Attribute”, and “Value”. A&A method allows us to add a new table without changing the schema of the data model. Addition of a new table can be done by absorbing the new information. Thus, this technique is named Addition and Absorption (A&A) method. Fig. 1 shows the A&A data modeling.

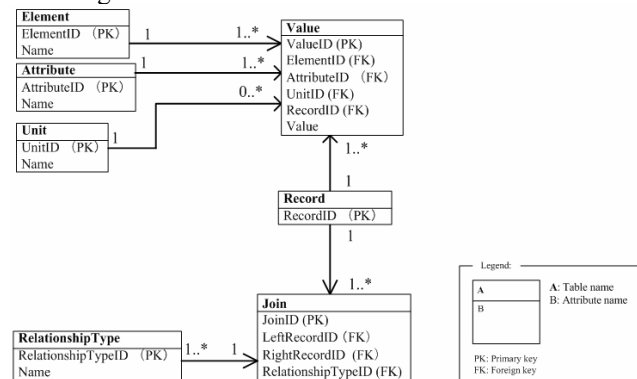


Fig. 1 A&A Data Modeling.

### 3. DMSC METHOD

In order to discover effective knowledge from a large amount of building sensor data, Data Mining for Sensor data of buildings using Contexts (DMSC) method was developed in this research. In this method, first, sensor data and contextual data are collected and stored making a data warehouse. Next, preprocessing such as processing of missing values, interpolation of outliers, reduction of dimensions, discretizing numerical attributes, selecting sensors and contexts is done. Then, decision tree technique is employed for pattern discovery. Finally, the result is evaluated by the user and if the result is not satisfactory, new data may be added or more preprocessing is done until some useful knowledge can be found. The flow of this method is shown in Fig. 2

Contexts can be classified as static contexts and time-series contexts. As for static contextual data, simple clustering can be used while for time-series data, decision tree technique can be more effective and useful than other methods. In addition, as time-series data tend to become very large, Piecewise Aggregate Approximation (PAA), Dynamic Time Warping (DTW) and Angular Metrics for Shape Similarity (AMSS) have been added.

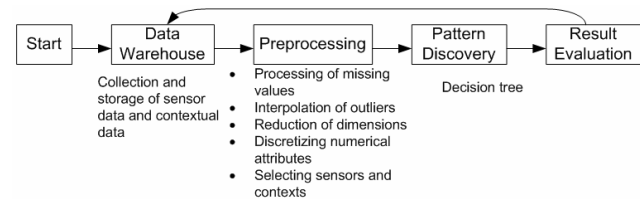


Fig. 2 Flow of the DMSC Method

### 3. EXPERIMENTS

In order to verify the proposed methods, measurement experiments were done at our laboratory of Osaka University from November 27, 2010 to February 3, 2011 using 13 wireless sensors which can observe temperature, humidity, illumination intensity, 6 sensors for electric current, 4 web cameras for storing contexts. DMSC method was applied to the sensor data and contextual data. Some useful knowledge was found in this experiment.

### 4. CONCLUSION

In this paper, A&A method and DMSC method were proposed to store sensor and contextual data of buildings flexibly and to discover useful knowledge from a large amount of sensor data. The result of the experiments showed the feasibility of the proposed methods. Further work includes enhancement of the data mining method and more experiment with larger number of sensors.

### REFERENCES

- [1] Yabuki, N., “An Intelligent Framework of Knowledge Discovery from a Large Amount of Data in SHM”, *World Forum on Smart Materials and Smart Structures Technology*, p.280, CD-R, 5p, 2007.