ARTICLE IN PRESS

The Korean Institute of Communications and Information Sciences

Available online at www.sciencedirect.com

ScienceDirect

ICT Express ▮ (▮▮▮▮) ▮▮▮–▮▮▮

www.elsevier.com/locate/icte

A privacy-preserving sharing method of electricity usage using self-organizing map

Yuichi Nakamura[a,*], Keiya Harada[b], Hiroaki Nishi[b]

[a] School of Science and Technology, Graduate School of Keio University, Japan
[b] Department of System Design Engineering, Keio University, Japan

Received 1 December 2017; received in revised form 21 January 2018; accepted 22 January 2018
Available online xxxx

Abstract

Smart meters for measuring electricity usage are expected in electricity usage management. Although the relevant power supplier stores the measured data, the data are worth sharing among power suppliers because the entire data of a city will be required to control the regional grid stability or demand–supply balance. Even though many techniques and methods of privacy-preserving data mining have been studied to share data while preserving data privacy, a study on sharing electricity usage data is still lacking. In this paper, we propose a sharing method of electricity usage while preserving data privacy using a self-organizing map.

Keywords: Privacy preserving; Data sharing; Self-Organizing map

1. Introduction

Smart meters, instead of conventional power meters, are used to measure electricity usage. Smart meters are installed as essential devices for smart houses and home energy management systems. It allows power suppliers to acquire the electricity usage data of the target household. Analysis of the acquired data renders smarter services from power suppliers. For instance, the total amount of used electricity during peak times would be reduced by shifting the peak electricity usage in each household. The technique to adjust the electricity bill for electricity usage control is referred to as demand response (DR). The DR service is used to cut or shift the peak electricity usage to maintain the demand–supply balance at a period. However, when the company under contract only captures the smart meter data, it becomes difficult to manage the fine DR service in the target region. This is because all power companies in the region cannot share the usage data for achieving the electricity usage control of each household, considering the amount of lifestyle information acquired by the history of electricity usage data from the perspective of privacy preserving and data monopolism of companies.

Privacy preservation of households must be well considered when the data are shared to analyze electricity usage. Data analysts may recognize what types of electrical appliances were operated at a specific time and by which household. This technique is known as non-intrusive load monitoring (NILM) [1]. This information would be useful for criminals; for instance, a thief could predict the family structure and reduce the risk of meeting the people who are living in the target's household [2].

Homomorphic encryption is one of the ways of analyzing data while keeping a household data private from data analysts. The encryption provides a calculation while keeping values and equations encrypted [3,4]. However, this technique lacks in calculation speed. Even if the calculation cost is acceptable for data analysis, problems still arise when considering open data. Open data is the concept that some data should be published, or freely available, to further discover applications and knowledge [5,6]. The electricity usage data cannot be published while

* Corresponding author.
E-mail address: nakamura@west.sd.keio.ac.jp (Y. Nakamura).
Peer review under responsibility of The Korean Institute of Communications Information Sciences.

https://doi.org/10.1016/j.icte.2018.01.004
2405-9595/

preserving the privacy of households by encryption techniques only.

In this paper, we propose a safe and effective data sharing method of electricity usage while preserving data privacy. The method consists of a self-organizing map (SOM) [7]. The proposed method was evaluated using two types of data captured from smart meters in different smart grids. The evaluation results show the contribution of this study. The proposed method is adaptable in sharing any type of data while adjusting a special variable to maintain data accuracy according to the relevant analyses. In addition, the proposed method can reduce information loss (IL) compared to another conventional method that anonymizes the data of each power supplier without data sharing.

This paper is organized as follows: Section 2 presents previous works related to this study, and Section 3 describes the proposed method while proposing the SOM-based approach. The proposed method is evaluated in Section 4 with two datasets, and we conclude this study in Section 5.

2. Related works

2.1. Privacy-preserving data mining (PPDM)

Privacy-preserving data mining (PPDM) is a category of techniques that analyzes data while preserving data privacy [8]. PPDM includes several approaches as follows.

Homomorphic encryption: Homomorphic encryption is an approach of sharing data as mentioned in Section 1. This method provides encrypted calculation results without revealing the concerning values and equations [3,4]. However, this approach requires high calculation costs. In addition, the cost to manage encryption keys would be high when many smart meters are used. Guan et al. proposed a method to aggregate electricity usage data from smart meters in a smart grid using homomorphic encryption [9]. The method proposes a control center that is allowed to obtain the aggregated electricity usage data. The control center must pay a higher cost to use homomorphic encryption and to manage the encryption keys. Even if the costs are acceptable, several problems still exist for the DR services. The first problem is the data aggregation process. The control center in Guan's system cannot obtain any type of individual smart meter data, such as pattern or cluster, by classification. This feature makes it difficult to issue the appropriate control commands to each household. Furthermore, the method does not allow the sharing of electricity usage data among power suppliers and others.

Randomizing: Another approach is by randomizing values. This approach replaces values in the data with alternative values to prevent disclosing the original value. The alternative values are generated by adding noise to the original values, or by randomly selecting other records in the data. A data user uses the data while estimating the original value using statistic methods [10]. Although this approach has an advantage in its wide application for various types of data, it lacks in preserving privacy because the original data can be estimated. Kursawe et al. proposes some protocols to aggregate data in smart grids while preserving data privacy [11]. In this protocol, data providers add noise to data before the aggregation, and the aggregator obtains the sum of electricity usage by subtracting the noise. However, the aggregator only obtains the sum of the data [9].

Anonymization: Anonymization generalizes the unique record to prevent the identification of the corresponding data and to preserve the privacy. K-member clustering is one of the anonymization algorithms [12]. It creates clusters of similar records by generalizing records. This generalization maintains the number of records of each cluster greater than or equal to k. Subsequently, values in the same cluster are replaced with common values to satisfy the k-anonymity. The k-anonymity is a metric of anonymization that represents how many records have at least the same values in the items. It proves that more than k records will be matched to any kind of query.

Because of the generalization process, anonymization degrades data precision. IL is a metric that measures the degree of degradation. Although k-member clustering allows power suppliers to share each electricity usage data, the IL will be large, which occurs when the data include many records with similar values. The possibility of this situation becomes higher when the total number of records is small, i.e., when the IL of anonymized data of a small power supplier becomes large. Thus, a new methodology is required to reduce the IL.

2.2. Self-organizing map (SOM)

SOM is one of the unsupervised machine learning techniques based on artificial neural network and maps multidimensional records into two-dimensional nodes, which is called a map [7]. First, a SOM randomly fills the node weight vectors of nodes in a map as initial values. The learning part repeatedly proceeds the two-step executions below. The first step looks for the best matching unit (BMU) of each record. The BMU is the most similar node to a record. The second step is to update the weight vectors of the nodes in the neighborhood of the BMU and the BMU itself by pulling them closer to the input vector. These steps are repeated until the updated differences become less than a threshold value λ.

In this paper, the SOM-based data sharing method is used to preserve privacy. We previously proposed a method to collect electricity usage data that uses the SOM to share data while considering data privacy [13], which is a novel technique using the SOM for PPDM. The SOM is a suitable technique for sharing electricity usage among power suppliers. This is because the SOM does not require any supervisor data to create a map. Each power supplier is only required to prepare the relevant electricity usage data. Next, the SOM allows the use of any number of record dimensions, although all records must be of the same number of dimensions, i.e., the SOM is executable for any electricity usage data if all data are acquired in the same period. Finally, the SOM does not require the number of clusters or the number of records in each cluster. It can ignore the size of the record of each power supplier for data sharing, and this is its unique and superior feature [14]. For these reasons, the SOM is used to extract the features of electricity usage in the proposed method.

Please cite this article in press as: Y. Nakamura, et al., A privacy-preserving sharing method of electricity usage using self-organizing map, ICT Express (2018), https://doi.org/10.1016/j.icte.2018.01.004.