Clustering of Complementary Electricity Consumers Based on Their Usage Patterns

Sheng-Ta Chen, *Chi-Lun* Liu, *Ming-Hung* Lee, *Min* Fung, and *Wei-Guang* Teng

Dept. of Engineering Science, National Cheng Kung University, No.1, University Road, Tainan City 701, Taiwan

Abstract. In the electricity market, the real-time balance of electricity generation and consumption is a main task. In view of this, power providers usually sign contracts with their critical consumers (i.e., usually large-scale industrial companies) for managing their capacity demands. On the other hand, aggregators group commercial and residential consumers, and integrate their demands to negotiate with power providers. With a proper grouping of numerous electricity consumers, aggregators help to ensure stable electric supply, and reduce the burden of managing many consumers. In this work, we thus propose a novel data clustering approach to group complementary consumers based on their usage patterns (i.e., daily electricity consumption curves.) Furthermore, we incorporate the technique of discrete wavelet transform to speed up the clustering process. Specifically, approximations reconstructed from only a few wavelet coefficients may precisely capture the shape of original usage patterns. Experimental results based on a real dataset show that our approach is promising in practical applications.

1 Introduction

Resulting from the recent widespread deployment of advanced metering infrastructure, concrete information of user consumption is more available from smart meters, which is responsible for recording the electricity so that user can manage their energy usage and communicate with the supply-side. This facilitates the development and revolution of the electricity market. For example, a role of utility companies is introduced in a modern electricity market. In other words, such an electricity market may be composed of one generation company, multiple utility companies, and thousands of consumers. The utility company plays the role of an intermediary agent between the generation company and consumers. The utility company purchases the electric power from the generation company at a wholesale price and then sells it to the consumers at a retail price to maximize its profits. In addition, the utility company plays the role of aggregator so as to group numerous small-scale (i.e., commercial and residential) consumers.

Note that different electricity consumers may have quite different usage patterns. For example, the usage pattern of a commercial consumer is usually determined by its opening hours. Also, the usage pattern of a residential consumer is closely related to the living style of its family members. When attempting to group appropriate consumers together, it is important to have their usage patterns complementary. Consequently, to utilize data clustering techniques, conventional similarity measures cannot be directly applied. We thus propose a new measure to quantify the degree of complementary between two consumers. It is also noted the time required for performing the clustering process grows dramatically with the number of consumers. In view of this, we propose to incorporate the technique of discrete wavelet transform to efficiently handle the usage pattern of every electricity consumer.

The rest of this paper is organized as follows. Preliminaries of recent electricity market are provided in Sec.2. Details of our approach are presented in Sec.3. Experimental results and corresponding discussion are illustrated in Sec.4. Finally, Sec.5 concludes this paper.

2 Preliminaries

2.1. Billing of electricity usage

Electricity, unlike other forms of energy (e.g., coal, gas, and oil,) may not be economically stored and must be generated on demand. In practical environments, critical electricity consumers (e.g., large industrial companies) have to consume their electricity according to the contract that they have signed with power providers. This means the electricity usage (usually measured in terms of kWh per 15 minutes) cannot exceed the capacity on the contract.

Note that if the peak demand does not exceed the contract capacity, a fixed capacity charge is levied, and if the peak demand exceeds the contract capacity, a penalty charge is levied. Therefore, signing an extremely low capacity on the contract can impose high capacity charges, while signing an extremely high capacity on the contract may result in an unnecessary basic capacity charge. Hence, to decide a proper contract capacity have

received significant attention from critical electricity consumers [1]. Furthermore, problems occur when consumers demand more than promises. Although power providers earn more money from the punishment, they should increase the amount of power supply emergently to meet the demand, which can cause a series of chaos for both users and providers. Also, while power providers oversupply electricity, they have to afford the cost by themselves. These problems are summarized in Fig. 1.
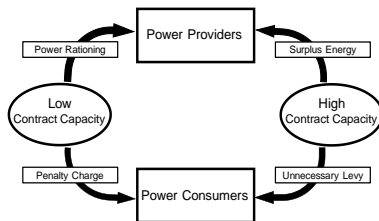

Fig. 1. Downsides of having an improper contract capacity.

2.2. Analysis of electricity consumption data

In the electricity market, the electricity real-time balance is a main task of the power system [2]. Traditionally, for large samples of households, only total household usage was available to a monthly time resolution, via utility bills. Studies of smaller samples of houses have been undertaken, primarily for electrical energy use, that provide data at a better temporal resolution (e.g., every 15 minutes), and, in some cases, by individual end use (e.g., air-conditioning unit, clothes dryer, stove). However, this kind of data collection comes at a high incremental cost for sub-metering equipment, which limited the number of houses studied and the time over which data was gathered [3].

In recent years, resulting from the widespread deployment of advanced metering infrastructure, concrete information of user consumption is more available from smart meters, which is responsible for recording the electricity so that user can manage their energy usage and communicate with the supply-side [4-5]. For instance, customers' energy use lifestyle can be understood by smart meter data [6], and customers can see their electricity use on a real-time basis. As a result, by supervising their electricity using behavior, they can reduce surplus energy consumption.

Load shapes derived from a smart meter reveal the magnitude and timing of electricity consumption. As shown in Fig. 2, not only energy providers but also consumers may be rewarded from data analytics of smart meter data.
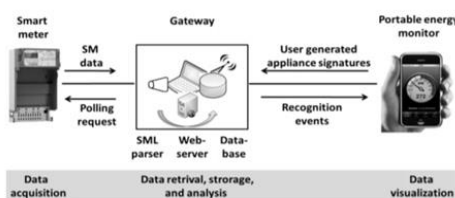
**Fig. 2.** Relationship between a smart meter and electricity. consumers [8].

Power providers can assess the impact of energy conservation and governmental policies on electricity usage, measure the impact of unusual weather (e.g., a heat wave,) and qualify the impact of projected climate change on electricity use and energy supplies [7].

Consequently, power providers may estimate the amount of energy they have to supply more precisely, that greatly reduces the probability of producing inadequate electricity for their consumers. Furthermore, the unnecessary cost of producing surplus power is saved. Energy providers can thus guarantee the quality and stability of electricity supplying. For instance, it has been shown that energy providers can identify usage patterns by electricity consumption data to predict future electricity consumption [7]. Moreover, after modeling the electricity usage, significant differences in consuming electricity between clusters of consumers can be identified.

Customers are allowed to make informed decisions regarding their energy consumption because of demand side management (DSM), which refers to the different initiatives intended to modify the time pattern and magnitude of the demand, or introducing advanced mechanisms for encouraging the demand-side to participate actively in the network optimization process [9]. Note that DSM plays a significant role in electricity markets [10], and it is a promising method to transform traditional power grid into a more reliably and economically operated smart grid [11-14].

2.3. Aggregators

For the stable operation of the power system, it is crucial to maintain the electricity real-time balance [2]. The role of utility companies, also known as aggregators, is introduced in a modern electricity market. As shown in Fig. 3, a bi-level electricity market may be composed of one generation company, multiple aggregators, and thousands of consumers. The aggregator plays the role of an intermediary agent between the generation company and consumers. The aggregator purchases the electric power from the generation company at a wholesale price and then sells it to the consumers at a retail price to maximize its profits.
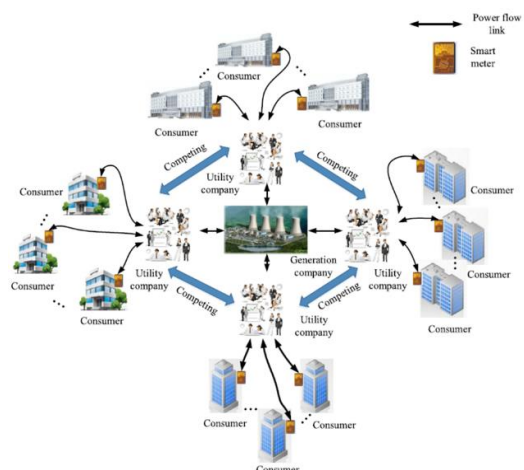
Fig. 3. A bi-level electricity market [2].

Aggregators cluster the consumers, sign contracts with them, and integrate their demands, which can ensure stable electric supply, and reduce the burden of managing many users. Moreover, the number of potential consumers participating in the demand response program, and the corresponding execution effectiveness both increase. To be specific, the consumers may respond to the price that improves the regulation capacities for their aggregator.

In summary, aggregators receive the wholesale price announced by the generation company and then set the optimal retail prices and regulation prices to maximize their profits. The consumers accept the optimal retail price sent by the utility company through the smart meters and then determine the optimal energy consumption. At the same time, the consumers provide regulation services to match supply with demand [2].

For example, there are some states in the USA adopting a *community electricity aggregation* program. With more than 200 such programs (or more than one million households,) the resulting benefits include reducing greenhouse gas, providing opportunities for new revenue, assisting with utility reform and creating job opportunities. In Germany, aggregators coordinate with transmission system operators, distribution system operators, balance responsible parties, energy retailers, and power consumers, so as to balance the amount of power between demand and supply. In the UK, aggregators cluster the power consumers, and sign contract with both power consumers and power suppliers. After that, they send the electricity data to the power suppliers, and if needed, they help the power suppliers to distribute the electricity.

3 Proposed scheme of complementary clustering

3.1. Data clustering

Data clustering is an unsupervised learning technique. In general, it targets on dividing data objects into different clusters such that the objects in the same cluster are more similar to one another than to those from different clusters. The similarity of two data objects is usually assessed according to a distance measure. Among several alternatives, hierarchical clustering algorithms provide a very simple and appealing way of displaying the organizational structure of the data as a dendrogram. Note that there exist many hierarchical clustering algorithms, and they may differ in aspects such as the adopted distance measure and whether the hierarchy (or dendrogram) is constructed in a top-down or a bottom-up manner.

Nevertheless, for our purpose of selecting appropriate consumers into the same cluster, the *complementarity* (rather than the similarity) of data objects is addressed. If two similar consumers are clustered together, their aggregated usage pattern becomes steeper. This means an additional cost of signing a higher contracted capacity to allow the usage of on-peak hours and a potential waste

in off-peak hours. Thus, complementary consumers rather than similar ones are desired to be clustered.
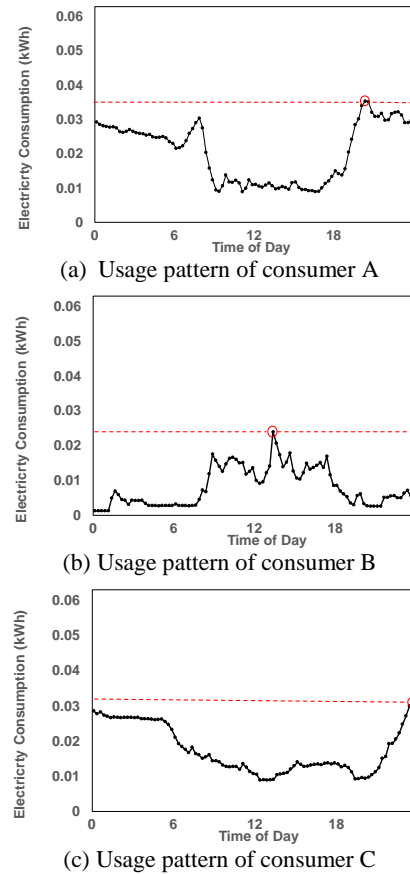

(a) Usage pattern of consumer A

(b) Usage pattern of consumer B

(c) Usage pattern of consumer C

Fig. 4. An example of different usage patterns.

For example, usage patterns of three different consumers (i.e., A, B, and, C) are shown in Fig. 4. When selecting two consumers to form a cluster, we can easily observe that the composite usage pattern in Fig. 5(b) is steeper than that in Fig. 5(a). In other words, it is more appropriate to group consumers A and B because their usage patterns are more complementary.
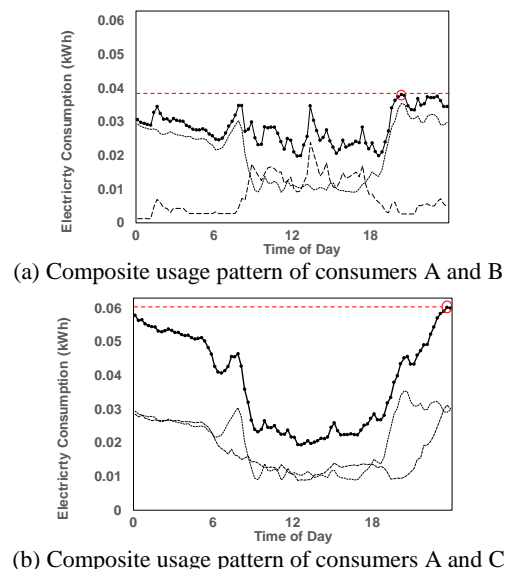

(a) Composite usage pattern of consumers A and B

(b) Composite usage pattern of consumers A and C

Fig. 5. Aggregation of different usage patterns.

In view of this, we need to propose a new measure to facilitate following clustering procedures because conventional distance measures cannot be directly utilized. Once the usage patterns of selected consumers are complementary enough, we expect the overall usage pattern of a cluster is smoother than that of individual consumer in this cluster.

3.2. Discrete wavelet transform

To reduce the computation complexity and thus the required execution time, the discrete wavelet transform is employed for processing time series. This technique is advantageous for its linear computational complexity. While the theory of wavelet is extensive, we conform ourselves to the preliminary wavelet transform in this paper. Namely, the Haar wavelet basis function (1) which is the simplest wavelet is explored and utilized for our approach.

$$H(t) = \begin{cases} 1, & 0 < t < \frac{1}{2} \\ -1, & \frac{1}{2} < t < 1 \\ 0, & otherwise \end{cases} \quad (1)$$

Note that $H(t)$ is the Haar wavelet's mother wavelet function.

Before and after the decomposition (as shown in Fig. 6,) it is observed that although the number of coefficients remains the same, decomposed coefficients tend to be distributed unevenly. Especially, it is reasonable to expect many decomposed coefficients to be very small in magnitude. With the purpose of obtaining a good approximation for the time series data using limited storage space, these small coefficients may be truncated with the effect of introducing only small errors in the reconstructed signal.
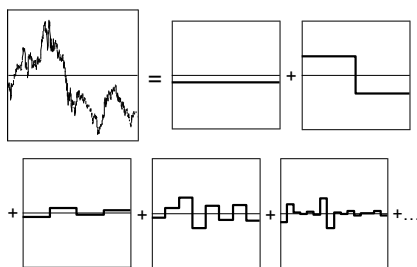
Fig. 6. Time series decomposition using the Haar wavelet.
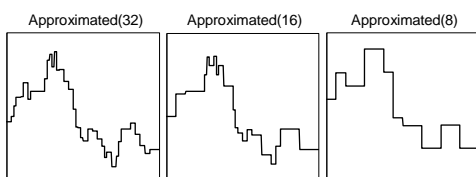
**Fig. 7.** Lossy reconstruction using different numbers of wavelet coefficients.

In general, there exists a tradeoff between the quality of reconstructed time series and the storage space for retained coefficients. Also, for the purpose of approximation, the reconstructed quality is in proportion to the number of significant coefficients stored. As shown in Fig. 7, an example of reconstruction using different numbers of wavelet coefficients is provided.

3.3. Progress value

In this paper, we define two indicators to facilitate the required data clustering process. Firstly, distance to the maximum value (DMV) is the sum of the differences between the maximum value of a load curve and the rest of the values. Therefore, the DMV is actually the size of the shaded area below the horizontal line and above the load curve in Fig. 8. Secondly, the progress value (PV) is the sum of the DMVs of the two consumers before being grouped, minus the DMV of the grouped consumer.

Note that as the DMV of the clustered consumer becomes smaller, which means the two consumers to be clustered are more complementary, the clustered load curve becomes flatter, and the PV value becomes larger, as shown in Fig. 4 and Fig. 5. In other words, if we can get a large value of PV, the most suitable couple of consumers to be clustered can be acquired.
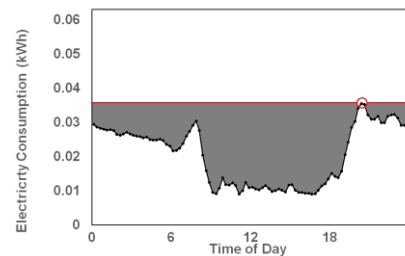
Fig. 8. DMV of an example load curve.

Integrating the methods above, first, by adopting the discrete wavelet transform, we simplify the original usage data into a series of coefficients, which not only decrease the execution time but similarly approximate the result from the original value. Then, we calculate the DMV of each consumer, and figure out the PVs of every pairs of consumers, which is the important standard and element while clustering consumers. At last, agglomerative, which is a "bottom up" approach, each observation starts in its own cluster, and pairs of clusters are merged as one moves up the hierarchy. Note that in each step of our approach, the pair with the highest PV value are merged.

4 Empirical studies

To implement the two approaches for comparison purposes, we develop corresponding programs using Python language. Our programs are run on a PC with Intel i7-7700HQ, 2.8GHz CPU and 8GBytes RAM.

Our testing dataset contains 1-year period usage data from 695 electricity consumers gathered by Bureau of Energy, Ministry of Economic Affairs, Taiwan. We adopt the data from June to September of that year, which is the summer time, to undergo complementary clustering.

In our approach, the goal of each itertation is to greedily find a pair of consumers having the highest PV value (and thus the lowest DMV value.) Note that the composite usage pattern of such a pair is flatter because the usage patterns of the two consumers are more compensate. Consequently, power consumers can utilize their contract capacity more effectively while they have a higher PV value. On the basis of above, we adopt the saving of contract capacity to measure the quality of our experiment, which can tell the amount of benefit that the aggregators get.

As shown in Fig. 4 and Fig. 5, when considering the two consumers A and B, their original contract capacity is around 0.035 and 0.024 kWh per 15 minutes, respectively. After grouping them together, the resulting contract capacity is around 0.038 kWh per 15 minutes, indicating a saving of 0.021 kWh per 15 minutes.

In our experiments, the original approach, the random approach, and the proposed approach are implemented for comparison purposes. Note that the hierarchical clustering technique is identically utilized for all these approaches. The only difference among these approaches is how we process the time series of all consumers. In the original approach, we simply feed original time series of all consumers into the clustering algorithm. In the random approach, consumers are randomly selected to form clusters, i.e., the time series of a consumer is not considered. In the proposed approach, we conduct the discrete wavelet transform on all the time series, and retain only a few wavelet coefficients to represent original time series for the clustering algorithm.

The performance of saving the overall contract capacity is shown in Fig. 9. Though merging the consumers according to PV value can be effective, while the peak value of two consumers are matched, it is unavoidable encountering the situation of PV value equals to 0. Wavelet(k) means that we retain only k wavelet coefficients to approximate the original time series. Note that using more wavelet coefficient saves more contract capacity than the random approach. In other words, aggregators have a measurement to select consumers to join them can get more benefits rather than generously accept an arbitrary consumer to be one part of them.
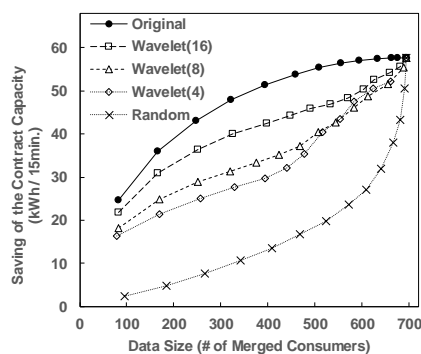
Fig. 9. Comparison of Clustered Size based on Saving of the Contract Capacity.

It is found that the results of the experiment using wavelet(16) are close with the results of original approach. Though there are similar results, our approach simplifies the process with wavelet transform, so that the execution time of the original approach is much longer than that of our approach. The execution time of utilizing the original and the proposed approach is shown in Fig. 10. With the data size of 500 consumers, it takes approximately 20,000 seconds for the original approach while the execution time of our approach is less than 5,000 seconds. Generally speaking, the original approach is with the best quality but requires a lot of time. When facing a large number of electricity consumers, the original approach may not be a good option.
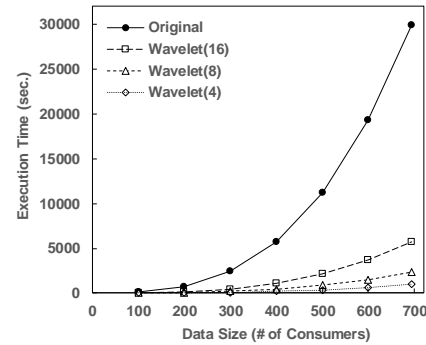
Fig. 10. Comparison of execution time.

5 Conclusions

In this work, we have focused on the problem of grouping appropriate electricity consumers based on their usage patterns. Note that complementary rather than similarity of two consumers are addressed. Consequently, we have revised the distance measure for our purposes of finding two complementary electricity consumers. Moreover, we have incorporated the technique of discrete wavelet transform for better efficiency in terms of the execution time when conducting the clustering technique. Experimental results have shown that our approach is promising in practice.

References

1. C. Y. Chen, C. J. Liao, *A Linear Programming Approach to the Electricity Contract Capacity Problem* (Appl. Math Model., **35**, 4077, 2011)

2. K. Ma, C. Wang, J. Yang, Z. Tian, and X. Guan, *Energy Management Based on Demand-Side Pricing: A Super Modular Game Approach* (IEEE Access. **5,** 18219, 2017)

3. M. Salani, A. Giusti, G. Di Caro, A. E. Rizzoli and L. M. Gambardella, *Lexicographic Multi-Objective Optimization for the Unit Commitment Problem and Economic Dispatch in a Microgrid* (*Proc. IEEE PES Eur. Conf. ISGT*, 2011)

4. I. Atzeni, L. G. Ordóñez, G. Scutari, D. P. Palomar, and J. R. Fonollosa, *Demand-Side Management via Distributed Energy Generation and Storage Optimization.* (IEEE Trans. Smart Grid, **4**, 866, 2013)

5. C. Greer, et al, *NIST Framework and Roadmap for Smart Grid Interoperability Standards, Release 3.0.* (*Nat. Inst. Standards Technol. Tech. Rep.,* 2014)

6. J. Kwac, J. Flora, and R. Rajagopal, *Household Energy Consumption Segmentation Using Hourly Data.* (IEEE Trans. Smart Grid, **5**, 420, 2014)

7. D. Le Comte, H. Warren, *Modeling the Impact of Summer Temperatures on National Electricity Consumption.* (J. Appl. Meteor, **20,** 1415, 1981)

8. M. Weiss, A. Helfenstein, F. Mattern, T. Staake, *Leveraging Smart Meter Data to Recognize Home Appliances* (*IEEE Proc. PerCom.*, 190, 2012)

9. P. Samadi, A-H. Mohsenian-Rad, R. Schober, V. W. S., Wong and J. Jatskevich, *Optimal Real-time Pricing Algorithm Based on Utility Maximization for Smart Grid.* (*Proc. IEEE Int. Conf. Smart Grid Commun.*, 415, 2015)

10. T. Lvogenthiran, D. Srinivasan and T-Z. Shun, *Demand Side Management in Smart Grid Using Heuristic Optimization* (IEEE Trans. Smart Grid, **3**, 1244, 2012)

11. W. Lee, L. Xiang, Schober, R., and V. W. S. Wong, *Direct Electricity Trading in Smart Grid: A Coalitional Game Analysis.* (IEEE J. Sel. Areas Commun. **32** 1398, 2014)

12. S. Maharjan, Q. Zhu, Y. Zhang, S. Gjessing, and T. Başar, *Demand Response Management in the Smart Grid in a Large Population Regime* (IEEE Trans. Smart Grid, **7**, 189, 2016)

13. P. Samadi, H. Mohsenian-Rad, R. Schober, and V. W. S. Wong, *Advanced Demand Side Management for the Future Smart Grid Using Mechanism Design.* (IEEE Trans. Smart Grid, **3**, 1170, 2012)

14. W. T. Li, C. Yeun, N. Hassan, W. Tushar, C. K. Wen, K. Wood, K. Hu, X. Liu, *Demand Response Management for Residential Smart Grid: From Theory to Practice* (IEEE Access. **3**, 2431, 2015)