Available online at www.sciencedirect.com

ScienceDirect

Energy Procedia 75 (2015) 2934 – 2939

Energy Procedia

The 7th International Conference on Applied Energy – ICAE2015

A comparative study of clustering algorithms for electricity self-sufficient community extraction

Yoshiki Yamagata*[a, b], Daisuke Murakami[a], Kazuhiro Minami[b], Nana Arizumi[b], Sho Kuroda[c], Tomoya Tanjo[b], Hiroshi Maruyama[b]

[a]*National Instituе for Environmental Studies, 16-2 Onogawa, Tsukuba 305-8506, Japan*
[b]*Institute of Statistical Mathematics, 10-3 Midorimachi, Tachikawa 190-8562, Japan*
[c]*University of Tsukuba, 1-1-1 Tennodai, Tsukuba 305-8573, Japan*

Abstract

As a first step of implementing an energy resilient community cluster system, we apply clustering algorithms, including two graph partitioning algorithms and a stochastic optimization algorithm, to electricity self-sufficient community clusters in Yokohama city, Japan. The cluster extraction results are compared in terms of the resulting cost function values and the cluster shapes. The comparison result suggests usefulness of each algorithm in different points of views.

Peer-review under responsibility of Applied Energy Innovation Institute

Keywords: Vihicle to community system; Electricity; Graph partitioning; Simmulated annealing; Photovoltaic panels

1. Introduction

Urban resilience to natural hazards is increasingly important under climate change. Urban systems must be resilient both direct and indirect influences of natural disaster. Electricity blackout is a major indirect risk that stops urban systems (e.g., it can stop heating supply, and cease emergency service due to the blackout of the telecommunication systems). In Japan, a massive electricity blackout was happened under the 2011 great east japan earthquake, and it makes the general public be aware of its significant risks of depending on highly centralized energy sources.

To increase energy resiliency, electricity energy sources are decentralized by various ways, including electricity liberalization and use of renewable energy. In Japan, for example, Feed-in Tarrif (FiT) scheme was initiated in 2011 after the great earthquake, to increase the share of renewable energy use. Since

* Corresponding author. Tel.: +81-29-850-2545; fax: +81-29-850-2419.
E-mail address: yamagata@nies.go.jp.

Peer-review under responsibility of Applied Energy Innovation Institute
doi:10.1016/j.egypro.2015.07.595

renewable energies are typically charged into distributed energy storages [2] whose capacity is limited and excessive use can lead reverse energy flow [3], these storages must be distributed efficiently.

In such a background, [4] proposed the Vehicle to Community (V2C) system that aims an efficient use of photovoltaic powers. In this system, electricity generated from mass-adopted photovoltaic panels (PVs) is stored in the “cars not in use,” and shared within pre-determined local communities (see §2). This system would work well if the community clusters are determined in a sensible way. Unfortunately, the clustering problem is NP-hard, and the optimal solution cannot be solved in a polynomial time. Accordingly, a clustering algorithm of approximating the optimal community clusters is needed [5].

Towards an implementation of the V2C system, we perform a comparative analysis of clustering algorithms in Yokohama city, Japan, and discuss their efficiency in terms of clustering of energy sharing community.

2. The community clustering and storage affordability

To achieve efficient electricity sharing, each community must satisfy the followings: (i) Storage affordability: each community must have sufficient storage capacity to charge PV electricity; (ii) Electricity transmission cost within each community must be small for efficient electricity sharing.

The storage affordability in (i) can be defined as [*Storage capacity* − *Electricity surplus*]. This section discusses how to estimate the storage capacity and the electricity surplus in 250m grids in Yokohama, which we will analyze later.

To estimate the *storage capacity*, we must know the number of cars not in use, which are used as storages in the V2C system. We estimate it by simulating the daily movements of peoples in Yokohama city using an agent-based transportation simulator, MATSim (http://matsim.org/). The origin-destination trip data (source: Fourth Person Trip Survey in Tokyo Metropolitan Area) and the road-network data (source: the National Digital Road Map Database) are used as inputs. See [4] for more details.

The *electricity surplus* is evaluated by month with [electricity PV supply – electricity household demand]. The electricity PV supply in m-th month $PV_{i,m}$ (kWh/h) is estimated following [6] as

$$PV_{i,m} = I \times \tau \times roof_i^{PV} \times \eta_{pc} \times K_{m,pt} \times T \tag{1}$$

where i is an index of 250m grids, I is the total solar irradiance ($kWh/m^2/h$), which is calculated by METPV-2 database [7], τ is the array conversion efficiency (= 0.1), $roof_i^{PV}$ is the installation area in i-th grid (m^2), which is calculated based on the area of detached houses (See [4]), η_{pc} is the efficiency of power conditioner (= 0.95), $K_{m,pt}$ is the temperature correction coefficient set for each month m (e.g., May: 0.92; August: 1.00), and T is the performance ratio (=0.89). On the other hand, the monthly electricity household demand in i-th grid, $D_{i,m}$, is estimated by $F_i \times w_m$, where F_i is the total floor area in i-th grid, and w_m is the unit electricity demand in each month (Source: The Japan Institute of Energy, 2008).

Note that the storage capacity and electricity surplus are estimated under a maximum case scenario that all of cars not in use are replaced with electric vehicles (EVs), which allow us storing PV electricity, and that PVs are installed on rooftops of all detached houses. In other words, we analyze what if EVs and the PV capacities are increased sufficiently.

3. Clustering algorithms

This section introduces standard clustering algorithms, including graph partitioning algorithms and stochastic algorithms. These algorithms aim to identify community clusters with (i) storage affordability and (ii) circular-cluster shape, which allows us to share electricity efficiently.

This section represents the study region (Yokohama city) as a graph G = (V, E), where the nodes represent the 250m grids in the region and the edges represent neighbouring relations between grids. Each node $n_i \in V$ has the following attributes: a weight value $w(n_i)$ representing the storage affordability; the area $a(n_i)$; and the perimeter $p(n_i)$. Our objective is to identify self-sufficient community clusters or sub-graphs, which we express as G_1, G_2,.. G_n. Storage affordability, areas, and perimeters of G_i are described by $w(G_i)$, $a(G_i)$, and $p(G_i)$, respectively.

Because graph partitioning is typically performed by minimizing a cost function considering graph structure only, we also apply such cost functions and the storage affordability is considered by a constraint. In contrast, following typical studies of conducting stochastic optimization, in our stochastic optimization, the storage affordability is minimized explicitly, while the graph structure is considered by adding a penalty term.

3.1. Graph partitioning: Model

Our graph partitioning approach considers the following transmission cost (ii).

$$C_g = \alpha \sum_{i=1}^{n} |V(G_i)|^2 + \beta \left| E(G) \setminus \bigcup_{i=1}^{n} E(G_i) \right| \tag{2}$$

where α and β are coefficients such that $\alpha + \beta = 1$, and $V(G_i)$ and $E(G_i)$ are the sets of nodes and edges in G_i, respectively.

The first term of C_g represents the transmission cost within each subgraph G_i, and is thus proportional to the number of node pairs within each subgraph $|V(G_i)|^2$. The second term represents the transmission cost across different subgraphs, and is proportional to the number of edges connecting different subgraphs.

To consider the storage affordability (i), we add the following constraint for every subgraph G_i: $w(G_i) < k$, where k is a given threshold. It constraints each subgraph G_i having a reasonable level of storage affordability. The graph partitioning (community clustering) considering both the self-sufficiency (i) and the transmission cost (ii) is achieved by minimizing Eq.(2) under $w(G_i) < k$.

3.2. Graph partitioning: Recursive Coordinate geometric Bisection (RCB) algorithm

The RCB algorithm partitions considering only vertical or horizontal partitions [8]. The algorithm is summarized as follows (see Fig.1):

- 1) Divide nodes $V(G)$ into two parts by partitioning plane orthogonal to a coordinate axis. It could be vertical or horizontal.
- 2) For each subgraph, repeat the operation in 1) to make even smaller partitioning.

This is a simple and efficient algorithm. However, the RCB algorithm is an approximation algorithm, which divides an original clustering problem into smaller clustering problems recursively, its computed partitioning result might be distant from the global optimal solution.

3.3. Graph partitioning: Multi-Leveled Graph (MLG) algorithm

We next describe a more general graph partition algorithm that supports partitions of graph [9]. We show the procedure of the MLG algorithm below (see Fig.1):

- 1) Construct a coarser summarization of a graph by aggregating nodes and edges in the graph.
- 2) Perform a graph partitioning procedure on the summarized graph.
- 3) Propagate the computed partition back to the original graph.
- 4) Repeat this process as needed.

Since the MLG algorithm allows more complex-shaped clusters, it usually produces better clustering results than the RCB algorithm, although the former is more computationally expensive than the latter.

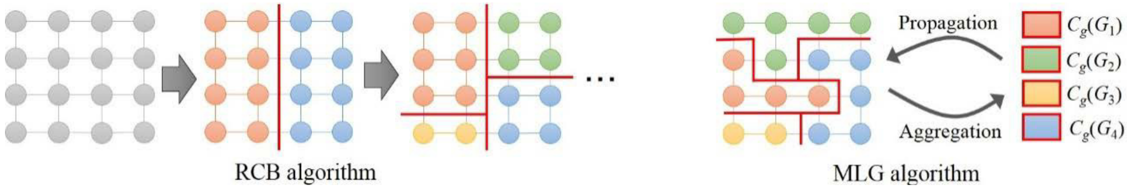


Fig. 1. Image of the RCB and MLG algorithms. The former iterates horizontal partitions while the latter iterates aggregation and propagation.

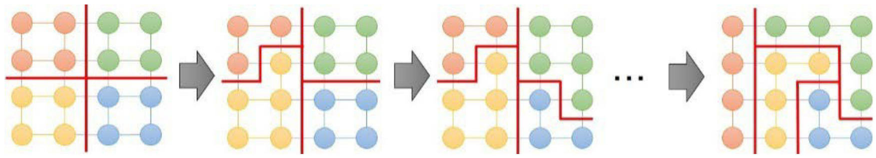


Fig. 2. Image of the SA algorithm. In each iteration, a local modification is accepted with a probability $T=\gamma^{k}T_0$, where $T_0 = 10^5$ and $\gamma = 0.9999$, and k is the number of iteration.

3.4. Stochastic optimization: Model

This stochastic optimization minimizes the following cost function:

$$C_s = \sum_{i=1}^{n} |w(G_i)| + \lambda \sum_{i=1}^{n} \left(1 - \sqrt{4\pi \frac{|a(G_i)|}{p(G_i)^2}} \right) \tag{3}$$

where λ is a given parameter. The first term evaluates the storage affordability (i) and the second term evaluates the circularity of G_i explaining the transmission cost (ii): if G_i is circular-shaped, average intra-subgraph distance is small, and, as a result, the transmission cost decreases.

3.5. Stochastic optimization: Algorithm

The optimization algorithm, which is called the simulated annealing (SA) [10], is described as follows:

- 1) Set initial M subgraphs, and calculate the cost $C_s = C_s(G_1, G_2, \ldots, G_m)$.
- 2) Detach a node n_i from a subgraph G_i, and merge it into another subgraph G'_i, then, calculate the cost $C'_s = C_s(G'_1, G'_2, \ldots, G'_m)$.
- 3) If $C'_s \le C_s$, the change of sub-graph is accepted, and, otherwise, it is accepted with a probability, which is defined by $\exp\{-(C_s - C'_s)/T\}$, where T is a parameter that controls the acceptance probability.
- 4) Iterate steps 2 and 3 alternately until C'_s converges.

To find the global optima, T needs be declined gradually across iterations. We define it by $T = \gamma^{k}T_0$, where $T_0 = 10^5$ and $\gamma = 0.9999$, and k is the number of iteration.

4. Application in Yokohama city, Japan

RCB, MLG, and SA algorithms are applied to the gridded monthly storage affordability data, which is constructed as discussed in §2, and 20 self-sufficient community clusters (G_i) are identified. Effectiveness of the resulting clusters are discussed in terms of (a) the storage affordability, (b) cluster shapes and these stability across months.

Fig.3 shows the clustering results. RCB clusters seem more stable across months than the other two. To quantify the stability, the similarity between clusters in months $m \in \{2,3,\cdots 12\}$ and m-1 are evaluated sequentially using the Jaccard's similarity coefficient (e.g., [10]), which approaches to 1 if two results are similar and 0 otherwise, and plotted in Fig.4. This figure shows that the RCB algorithm outperforms MLG and SA algorithms in terms of the stability (b). SA clusters are also stable, in particular in spring and autumn. The SA clusters are stable in the spring and the autumn, and not in the summer and the winter.

The average storage affordability and the maximum/minimum storage affordability in 20 clusters in each month are plotted in Fig.5 (average: red solid lines; maximum/minimum: red dash lines). This figure suggests that, on average, RCB, MLG, and SA clusters have positive storage affordability except for January and February, and December. Besides, in each result, all clusters have positive storage affordability (i.e., the minimum affordability is positive) in 6 months in spring and autumn, in which use of cooling or heating is limited. The result that PV electricity can cover whole household demand by community electricity sharing in these seasons is interesting. However, in each result, more than one clusters have negative affordability in 6 months (January, February, March, August, November, and December). The negative value of SA is the smallest in 4 in 6 months. SA might be preferable in terms of the storage affordability (a).

On the other hand, considering the feasibility and the operational cost, use of only one cluster set across months might be better. The clustering result in m'-th month is applied in all of the 12 months, and the average and maximum/minimum storage affordability in each cluster are evaluated for all $m' \in \{1,2,\cdots 12\}$ and plotted in Fig.5 (average: grey solid lines; maximum/minimum: grey dash lines). In this figure, storage affordability indicates similar values even if clusters given in a month is applied across months. Especially, clusters in March, which are plotted by black lines, provide smaller (in absolute value) minimum affordability in 9 in 12 months when SA is used, and, when the RCB is used, clusters in March is almost the same with the result when monthly clusters are applied.

In summary, considering (a) and (b) and the operational cost, use of SA/RCB clusters in March as energy-resilient community clusters might be a sensible way.

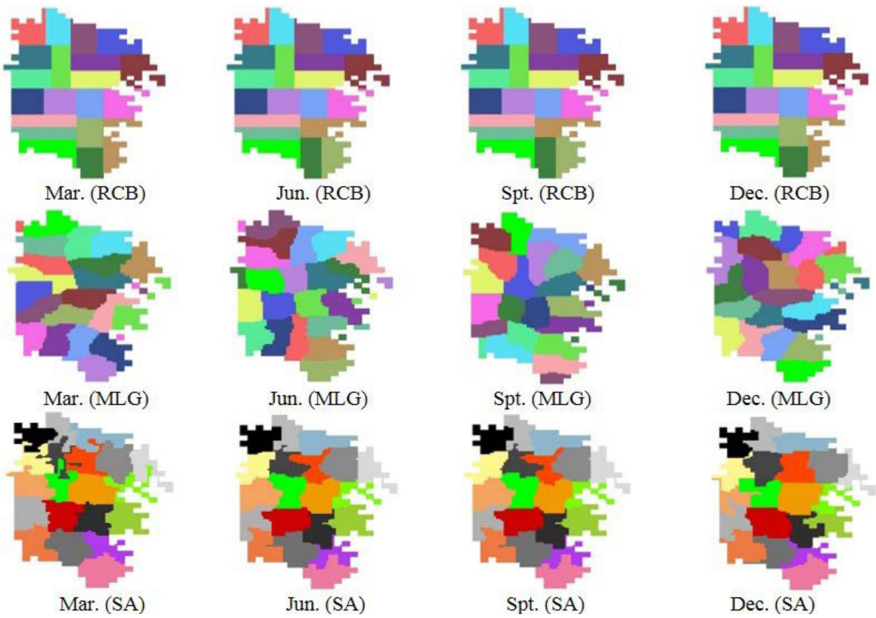


Fig. 3. Clustering results on four months.

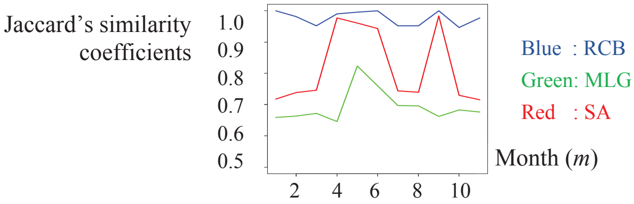


Fig. 4. Stability of clusters: Similarity between clusters in month *m* and *m*-1 are quantified by Jaccard's similarity coefficients.

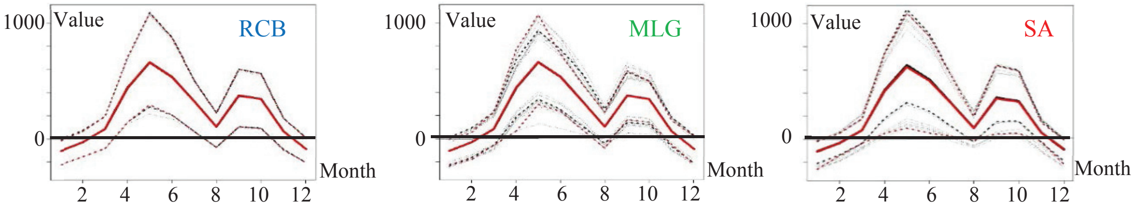


Fig. 5. Storage affordability in each month: Red represents results when monthly optimization results are applied in each month; black represents results when clusters optimized in March is used across months, and gray lines represent results when clusters optimized in a month (except for March) is used across months. Solid lines denote average storage affordability, and dash-lines denote maximum and minimum storage affordability in 20 clusters.

Acknowledgements

This study was funded by the "Research Program on Climate Change Adaptation" of the Ministry of Education, Culture, Sports, Science and Technology, Japan.

References

[1] Mohd A, Ortjohann E, Schmelter A, Hamsic N, Morton D. Challenges in integrating distributed Energy storage systems into future smart grid. *Proc. IEEE ISIE*, 2008; 1627–1632.

[2] Thomson M, Infield DG. Impact of widespread photovoltaics generation on distribution systems. *Renew Power Gener*, 2007; **1**: 33–40.

[3] Yamagata Y, Seya H. Proposal for a local electricity-sharing system: a case study of Yokohama city, Japan. *IET Intell Transp Syst*, 2013; 1–12.

[4] Yokoi T, Yamamoto Y, Tokai A, Morioka T. Development of decision support system to integrate block renewal and energy planning towards low-carbon city (in Japanese). *J Infrastruct Plan Manage G*, 2010; **66**: 17–34.

[5] Arizumi N, Minami K, Tanjo T, Maruyama H, Murakami D, Yamagata Y. A first step towards resilient graph partitioning for electrical grids. *Proc 7th Int Conf Inf Autom Sustain (ICIAfS)*, 2014.

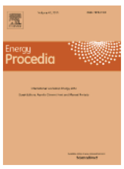
[6] Itagaki A, Okamura H, Yamada M. Preparation of meteorological data set throughout Japan for suitable design of PV systems. *Proc. Third World Conf. Photovoltaic Energy Conversion*, 2003; **2**: 2074–2077.

[7] Berger MJ, Bokhari SH. A partitioning strategy for nonuniform problems on multiprocessors. *Comput, IEEE Transact*, 1987; **100**: 570–580.

[8] Karypis G, Kumar V. Metis-unstructured graph partitioning and sparse matrix ordering system, version 2.0, 1995.

[9] Van Laarhoven PJ, Aarts EH. *Simulated annealing*. Springer,1987.

[10] Seifoddini H, Djassemi M. The production data-based similarity coefficient versus Jaccard's similarity coefficient. Comput Ind Eng, 1991; **21**; 263–266.



Biography

YOSHIKI YAMAGATA is a Principal Researcher of Center for Global Environmental Research, National Institute for Environmental Studies, Japan. He graduated from the University of Tokyo (PhD in System Science). His recent research topics include: Land use scenarios, resilient urban planning and International regime networks.