Constrained Clustering of Pecan-Street Electricity Consumption by Pairwise Similarity Matrix Completion

WEN Zhenduo 118010323, LI Zhuoyu 118010159

March 14, 2021

Contents

1 Abstract

In this project, by constructing a binary similarity matrix S, with $S_{i,j} = 1$ indicates consumer i and j lie in the same cluster and vice versa, we transform the constrained K-means clustering problem into a matrix completion problem, which utilizes the prior knowledge and avoids the order-sensitive issue in original constrained clustering algorithm.

Keywords–Constrained Clustering, Sparse Matrix Completion, Marginal System Impacts, Nuclear Norm Minimization

2 Literature Review

2.1 Constrained K-means Clustering

Constrained K-means Clustering is a classical semi-supervised clustering method introduced by Kiri Wagstaff, Claire Cardie, Seth Rogers, Stefan Schroedl, et al. [4] This method is proposed in a sense of properly applying any prior knowledge which traditional K-means Clustering method fails to utilize. This method considers two types of pairwise constraints:

- **Must-link**: Constrains to specify that two data points must be in the same cluster.
- **Cannot-link**: Constraints to specify that two data points cannot be in the same cluster.

The algorithm is a modified version of K-means Method incorporated with the above pairwise constraints:

1. Let $C_1, ..., C_k$ be the initial cluster centers.
2. For each data point d_i in D to be clustered, assign it to the closest cluster C_j such that no Must-Link or Cannot-Link is violated.
3. For each cluster C_i, update the cluster center by the mean of all data point d_i within.
4. Iterate between 2. and 3. until convergence.
5. Return $\{C_1, ..., C_k\}$.

In each iteration, the above algorithm continues down the sorted list of clusters by distance and put data point d_i into a cluster when current cluster is legal to hold d_i. So the constrains are never broken.

In a sense that this algorithm utilizes prior knowledge, Wagstaff et al. shows that constrained clustering alleviates the ill-posed problem of K-means clustering due to the latter's unsupervised nature. However, constrained clustering is stilling solving a non-convex optimization problem, leading to the result that it is sensitive to the initialization of clusters. Moreover, this algorithm is sensitive to the order of assigning data-points into clusters. If a poor decision is made early on, the algorithm may encounter the case that some data point d_i has no legal clusters to place in.

2.2 Marginal System Impact

The marginal system impact of each user is first proposed by Yu et al. [5]. They first define Consumer i's daily demand profile as:

$$Pf_i = \frac{L_i}{|L_i|_1} = \left(\frac{l_{i,1}}{\sum_h l_{i,h}}, \ldots, \frac{l_{i,H}}{\sum_h l_{i,h}} \right)$$

where consumer i's daily load profile is $\underline{l}_i = (l_i^1, ..., l_i^T)$. Also the total cost of grid is:

$$C_T(L) = \sum_{t=1}^{T} C(L_t)$$

where $L = (L_1, ..., L_T)$ is the total energy consumption of grid in each period, which is the sum of individual consumers in each period t with $L_t = \sum_i l_{i,t}$.

The Consumer i's marginal system feature impact on $\phi_j(L)$ is defined as:

$$MFI_{i,j} = \lim_{\Delta \to 0} \frac{\phi_j(L + \frac{\Delta l_i}{||l_i||_1}) - \phi_j(L)}{\Delta}$$

Correspondingly, the marginal system-cost impact of consumer i's daily demand is

$$MCI_i = \lim_{\Delta \to 0} \frac{C_T(L + \frac{\Delta l_i}{||l_i||_1}) - C_T(L)}{\Delta}$$

A remarkable conclusion raised by Yu et al. is that if L_i and L_k are linearly correlated, consumers i and k share the same MFIs and MCI. Further more, they state that consumers in the same profile share the same MFIs and

MCI. Their work thus provides a powerful tool to determine the "distance" in terms of power consumption clustering. Such kind of application of MFIs and MCI is also used in Jingshi Cui, Haoxiang Wang, Chenye Wu, and Yang Yu's work. [3] They proposed an optimal greedy algorithm for consumer clustering based on MCI_i.

2.3 Low-Rank Matrix Completion via Convex Optimization

Low-Rank Matrix Recovery is already a classic and matured algorithm. It is discussed in details in Emmanuel J Cand'es and Benjamin Recht's work. [1] They show that, if we observed certain m amount of entries from the observed matrix M, with great possibility, the original matrix X can be perfectly recovered by the following optimization problem:

$$\begin{aligned} &\text{minimize } ||X||_* \\ &\text{subject to } X_{i,j} = M_{i,j},\ (i,j) \in \Omega \end{aligned}$$

where Ω is the observable set, $||X||_*$ is the nuclear density of X.

They also prove that, if matrix $X_{m\times n}$ has row and column spaces that are incoherent with the standard basis of $R^{\max(m,n)}$, then the above nuclear norm minimization can recover X from a relative small random sample observation of X.

This inspires us in a sense that, if we manage to express the clustering of consumers in form of a sparse matrix, given certain condition satisfied, we can obtain the "real" underlying clustering via nuclear norm minimization.

3 Method

In this project, we proposed a semi-supervised clustering method based on low-rank matrix completion. The gist is to construct a Similarity Matrix S, with $S_{i,j} = 1$ if consumer i and j are in the same cluster and $S_{i,j} = 0$ otherwise. Then the observable part of S is obtained based on MCIs: We determine a positive value ρ, such that $|MCI_i - MCI_j| < \rho$ implies i and j fall in the same cluster, and $|MCI_i - MCI_j| > K * \rho$ implies implies i and j do not fall in the same cluster for some positive number K. This gives the Must Link and Cannot Link in constrained clustering. With properly chosen ρ, K, it is possible to perfectly recover S. In the sense that finding the perfect partition of data is equivalent to recovering the similarity matrix, we transform the clustering problem into a matrix completion problem. Similar transformation of clustering problem is also proposed by Yudong Chen, Ali Jalali, Sujay Sanghavi, and Huan Xu. [2]

Our algorithm is composed by the following steps:

1. Compute MCI_i for each consumer i.

2. Set $\rho, K, m > 0$. Loop through the data set:

   if $|MCI_i - MCI_j| < \rho$: set $O_{i,j} = 1$ store (i, j) into Ω,

   else if $|MCI_i - MCI_j| > K * \rho$: set $O_{i,j} = 0$, store(i, j) into Ω,

   do nothing otherwise.

   End loop if certain m pairs of consumers have conducted the above procedure.

3. Solve the optimization problem:

   minimize $||S||^*$, s.t. $S_{i,j} = O_{i,j}, (i, j) \in \Omega$

The rest part of this project will be the theoretical analysis about choosing criteria of MCI distance: ρ, K and the maximum size of observations: m, and the implementation of our algorithm on Pecan-Street data set.

References

[1] Emmanuel J Candès and Benjamin Recht. Exact matrix completion via convex optimization. *Foundations of Computational mathematics*, 9(6):717–772, 2009.

[2] Yudong Chen, Ali Jalali, Sujay Sanghavi, and Huan Xu. Clustering partially observed graphs via convex optimization. *The Journal of Machine Learning Research*, 15(1):2213–2238, 2014.

[3] Jingshi Cui, Haoxiang Wang, Chenye Wu, and Yang Yu. Robust data-driven profile-based pricing schemes. *arXiv preprint arXiv:1912.05731*, 2019.

[4] Kiri Wagstaff, Claire Cardie, Seth Rogers, Stefan Schroedl, et al. Constrained k-means clustering with background knowledge. In *Icml*, volume 1, pages 577–584, 2001.

[5] Yang Yu, Guangyi Liu, Wendong Zhu, Fei Wang, Bin Shu, Kai Zhang, Nicolas Astier, and Ram Rajagopal. Good consumer or bad consumer: Economic information revealed from demand profiles. *IEEE Transactions on Smart Grid*, 9(3):2347–2358, 2017.