

# 基于最大信息系数的用户光伏出力分解技术

潘可达<sup>1</sup>, 李镇豪<sup>2</sup>

<sup>1</sup>广东工业大学自动化学院, 广东 广州

<sup>2</sup>广东电网有限责任公司清远供电局, 广东 清远

收稿日期: 2022年3月29日; 录用日期: 2022年4月20日; 发布日期: 2022年4月29日

## 摘要

近年来, 分布式光伏系统(Distribution Photovoltaic Systems, DPVSs)由于相关政策的支持、低安装成本和易操作性, 得到了越来越多用户的青睐。然而大量小于10 kWp的小规模DPVSs是安装于表后侧(Behind-the-Meter, BTM)的, 光伏发电信息并未被单独计量, 这导致了光伏发电对表前侧(配电网侧)的不可见性。若配网侧仅以净负荷数据进行能量管理, 将无法进行最优的配电网调度和需求响应策略的实施, 导致一系列的安全经济问题。为解决上述问题, 本文提出了一种针对小规模DPVS的BTM净负荷分解方法, 该方法可在不依赖于天气数据和光伏设备物理模型的情况下, 对BTM光伏发电进行估计。所提算法在社区用户的DPVS通常表现出近似的光伏发电特性这一假设的情况下, 利用邻近用户的净负荷来提取光伏发电信息作为彼此间相互的光伏代理。最终的光伏发电分解结果将根据最大信息系数(Maximal Information Coefficient, MIC)进行确定。实验结果表明, 所提分解算法在不依赖气象数据的情况下拥有较高的分解精度。

## 关键词

表后系统, 光伏发电分解, 相关性分析

# Maximal Information Coefficient Based Residential Photovoltaic Power Generation Disaggregation

Keda Pan<sup>1</sup>, Zhenhao Li<sup>2</sup>

<sup>1</sup>Department of Electrical Engineering, Guangdong University of Technology, Guangzhou Guangdong

<sup>2</sup>Qingyuan Power Supply Bureau, Guangdong Power Grid Co., Ltd., Qingyuan Guangdong

Received: Mar. 29<sup>th</sup>, 2022; accepted: Apr. 20<sup>th</sup>, 2022; published: Apr. 29<sup>th</sup>, 2022

## Abstract

Due to policy support, low cost and easy applicability, distribution photovoltaic systems (DPVSs) are increasingly popular among residential community. However, small-scale DPVSs of less than 10 kWp are always installed behind the meter (BTM), without metering the photovoltaic (PV) power generation separately, which results in the invisible of the PV power generation. Only access of net load data can result in non-optimal distribution network control and optimization, leading to a series of energy management problems. In order to solve the aforementioned problems, this paper proposes a BTM net load disaggregation method focusing on small-scale DPVSs, with only net load data of residential users in a community, without relying on weather data and models assumption. Considering that community users' DPVSs usually exhibit approximate output characteristics, neighboring net load is used to extract PV power generation information as mutual proxies. After obtaining approximate PV proxy data by subtracting composite power of inter-users, Maximal Information Coefficient (MIC) is performed to obtain final PV power generation disaggregation results. Testing results show that the proposed method achieves considerable disaggregation accuracy in the absence of weather data.

## Keywords

Behind-the-Meter System, Photovoltaic Power Generation Disaggregation, Correlation Analysis

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

不可再生能源的紧缺和环境恶化问题是推动传统化石燃料能源依赖型电力系统转型的主要因素。目前, 冰岛、挪威、哥斯达黎加、巴西和加拿大等国已经分别实现了 100%、97%、93%、76% 和 62% 的可再生能源并网[1]。光伏(Photovoltaic, PV)发电考虑到较低的度电成本, 以及可从太阳获得大量的可再生资源特性, 具有巨大的发展前景[2]。为实现光伏发电的就地消纳, 目前大量的小型分布式光伏发电设备(Distributed Photovoltaic Systems, DPVSs)都是安装于用户屋顶的[3]。根据 2019 年欧洲太阳能发电(Solar Power Europe 2019)统计报告, 屋顶光伏的装机容量大约在 44 GW 至 76.5 GW 之间[4]。然而绝大部分的屋顶光伏系统都是安装于表后侧(Behind-the-meter, BTM)的[5], 且由于成本问题, 并未安装有独立计量能量消耗和光伏发电量的表计, 这将导致能量管理难以有效实施、储能设备容量无法合理确定以及继电保护难以正确设定等问题的出现。此外, 用户由于电力信息隐私的问题, 不愿公开住宅能源信息的细节, 也将导致电网运营商由于缺乏实际的负荷消耗和光伏发电信息, 难以进行最优能量调度的问题。综上所述, 设计一种可用于估计 BTM 系统下的 PV 发电量的方法是十分有意义的。

目前有关研究 BTM 系统下 PV 发电分解方法的文献, 可被划分为基于模型假设的分解方法和基于数据驱动的分解方法。以文献[6]和[7]为代表的基于模型假设的方法的共同特点是, 分解的 PV 发电量是根据假设的 PV 系统几何参数和 PV 出力特性进行求解的, 这将导致太阳辐照度与 PV 发电量之间具有过度的传递一致性。因此, 一旦假设 PV 系统模型与实际情况不一致, 对 PV 发电量的估计误差就会持续存在。

基于数据驱动的分解方法则可以避免由于模型假设不合理所引入的误差, 但此类分解方法有更高的外部数据需求(如温度数据、辐照度数据等)。部分数据驱动的方法依赖于能量代理的设置来分解净负荷。

如在文献[8]中, 通过使用支持向量回归(Support Vector Regression, SVR)模型估计待分解用户的 PV 设备装机容量后, PV 发电量的分解结果由所设置的单位容量的 PV 代理的光伏发电量, 乘以估计的装机容量获得。在文献[9]中, 以在馈线侧单独计量所得的用户群的负荷消耗数据和 PV 发电数据作为额外特征变量, PV 发电量的分解结果通过搭设线性回归模型计算获得。

综上所述, 无论所用分解方法是基于模型假设还是基于数据驱动的, 若想实现 BTM 系统下净负荷的分解, 必须获得至少一种外部变量用于对待分解的用户 PV 发电量或负荷消耗量进行推导。对于基于模型假设的方法来说, 用于推导的外部变量多为气象数据, 而对于数据驱动的方法, 用于推导的外部变量多为所设代理的负荷消耗或光伏发电数据。

对于外部变量如气象数据来说, 若其数据采集地点与待分解用户的净负荷数据采集地点不一致, 将引入时间或地理空间的误差。而对于外部变量如代理站点的光伏发电量或负荷消耗量数据来说, 代理站点的假设和代理站点的能量数据采集, 将极大增加经济成本和时间成本。上述两点都是不利于 BTM 光伏发电量分解的实际实施的。因此, 本文提出了一种基于数据驱动的方法, 在仅使用居民用户的净负荷数据的情况下, 对 BTM 光伏发电量进行分解, 以解决使用上述外部变量可能带来的多种问题。

## 2. BTM 光伏发电量分解算法介绍

### 2.1. 使用社区用户总净负荷筛选 PV 代理站点

由于处于相同的地理位置, 同一社区内的天气条件是相似的, 如太阳辐射、环境温度等。因此, 太阳能光伏板接受到的单位面积辐照度、工作温度应是近似的, 也就意味着整个社区内不同的光伏板的 PV 发电量曲线的形态是近似的, 只是由于装机容量的不同, 在幅值上有差异。因此每个用户的光伏设备都可看作是其它用户的 PV 代理站点。

即使用户间的光伏设备可以作为彼此的 PV 代理站点, 但由于用户均采用净负荷计量的方式, 也无法获得实际的光伏发电量信息用作 BTM 光伏发电量的分解。虽然实际的光伏发电量无法获得, 但用户间的光伏发电量差值是有可能获得的。假设两个用户具有近似的负荷消耗量和负荷消耗行为, 通过对这两个用户的净负荷作差, 以抵消其中的负荷消耗信息, 能获得近似的光伏发电量差值, 由于用户间光伏设备装机容量不同, 所导致的光伏发电量的差值。考虑到辐照度与 PV 发电量之间的高度的线性关系, 此时获得的 PV 发电量差值曲线的形态, 应与社区内其余用户的 PV 发电量曲线的形态近似。那么该 PV 发电量差值曲线也可看作一个特殊的 PV 代理站点的 PV 发电量结果。

净负荷定义为负荷消耗减去 PV 发电量, 假设一个社区内共有  $n$  个用户, 则对于用户  $i$ , 其负荷消耗如下式(1)所示:

$$C_{i,t} = D_{i,t} - P_{i,t}, i \in n \quad (1)$$

为使用户之间的负荷消耗的幅值尽可能相似, 社区中每个用户的净负荷都需要缩放至相同数量级。由于在夜间, 光伏发电量为 0, 根据式(1)可知, 此时光伏发电量为  $D$  为 0, 即负荷消耗  $C$  与净负荷  $D$  在此时相等。为避免缩放时净负荷中光伏发电量带来的影响, 各用户的净负荷将根据夜间耗电量进行缩放, 如式(2)所示。

$$C'_{i,t} = \frac{C_{i,t}}{\frac{1}{m} \sum_{t_1=1}^m C_{i,t_1}}, i \in n, t_1 \in T_{night} \quad (2)$$

式中,  $T_{night}$  表示夜间时段, 分子  $\frac{1}{m} \sum_{t_1=1}^m C_{i,t_1}$  表示, 夜间的净负荷均值(由于夜间无光伏发电量, 与夜间的负

荷消耗均值相等),  $m$  代表夜间时段时刻点的数量。

则经过缩放后, 用户  $i$  和用户  $j$  的净负荷差值如式(3)所示。

$$C'_{i,t} - C'_{j,t} = (D'_{i,t} - D'_{j,t}) - (P'_{i,t} - P'_{j,t}); i \in n, j \in n, i < j \quad (3)$$

为便于表示, 将式(3)改写为式(4)。

$$\Delta C'_{ij,t} = \Delta D'_{ij,t} - \Delta P'_{ij,t} \quad (4)$$

由于进行了缩放, 在假设用户用电行为近似的情况下, 式(4)中的负荷消耗差值  $\Delta D'_{ij,t}$  可以看作小的噪声。则式(4)可近似改写为:

$$\Delta C'_{ij,t} \approx -\Delta P'_{ij,t} \quad (5)$$

且由于上文提及的, 社区用户间较近的地理距离以及辐照度与 PV 发电量之间的高度相关性, 可认为用户  $i$  和用户  $j$  在  $t$  时刻光伏设备接受到的单位辐照度  $I_{i,t}$  和  $I_{j,t}$  满足如下关系:

$$I_{i,t} = I_{j,t} \quad (6)$$

设用户 PV 装机容量  $\alpha$ , 单位辐照度  $I$  和 PV 发电量  $P$  满足以下关系:

$$P = \varepsilon \cdot \alpha \cdot I \quad (7)$$

式中  $\varepsilon$  为“辐照度-PV 发电量”转换系数, 假设不同用户的  $\varepsilon$  相等, 则式(6)可改写为如下形式:

$$\frac{P'_{i,t}}{\alpha_i} = \frac{P'_{j,t}}{\alpha_j} \quad (8)$$

式中,  $\alpha_i$  和  $\alpha_j$  分别代表用户  $i$  和用户  $j$  的 PV 装机容量。令用户  $i$  和用户  $j$  间缩放后的 PV 发电量差值  $\Delta P'_{ij,t} = P'_{i,t} - P'_{j,t}$ , 则根据式(8)可推导得如下式。

$$P'_{i,t} = \beta_{ij} \Delta P'_{ij,t} \quad (9)$$

$$\beta_{ij} = \alpha_i / (\alpha_i - \alpha_j), \quad \alpha_i \neq \alpha_j \quad (10)$$

从式(8)中可知, PV 发电量差值与 PV 发电量之间有非常强的线性相关性, 这也解释了前文所说 PV 发电量差值与 PV 发电量之间有近似的曲线形态。对式(9)同除  $\alpha_i$  可得:

$$\frac{P'_{i,t}}{\alpha_i} = \frac{\beta_{ij}}{\alpha_i} \Delta P'_{ij,t} \quad (11)$$

则  $\Delta P'_{ij,t}$  可以看作, 装机容量为  $\frac{\alpha_i}{\beta_{ij}}$  的 PV 设备在  $t$  时刻的 PV 发电量, 也即是本文所提的 PV 代理站点的 PV 发电量。

式(9)和式(11)解释了为什么 PV 发电量差值可作为 PV 代理站点的 PV 发电量。同样地, 对于同一个社区内的不同用户, 根据式(4)至式(11), 通过匹配不同用户的净负荷进行相减, 可以求得多个 PV 发电量差值  $\Delta P'_{ij,t}$  作为不同 PV 代理站点的 PV 发电量。

然而, 用户的用电行为是具有多样性的, 也即式(4)所提及的负荷消耗差值  $\Delta D'_{ij,t}$  造成的小噪声。因此有必要对不同用户匹配所得的净负荷差值  $\Delta C'_{ij,t}$  进行筛选, 使其具有足够的代表性, 以满足式(5)。

对于拥有  $n$  个用户的社区, 来说用户间的组合模型有  $(n^2 - n)/2$ , 也即可以获得  $(n^2 - n)/2$  个净负荷差值  $\Delta C'_{ij,t}$ 。本文将以最大信息系数(Maximal Information Coefficient, MIC)从中筛选出最优的  $\Delta C'_{ij,t}$ 。MIC 是一种用于分析变量间相关性的方法, MIC 数值越大, 两个变量间的相关性越高。MIC 的计算方法如文献[10]所述, 此处不再赘述。令  $C'_{n,t}$  为社区中  $n$  个用户的总净负荷量, 通过计算  $\Delta C'_{ij,t}$  和总净负荷量  $C'_{n,t}$  的

MIC 系数, 选择其中 MIC 系数最大时所对应的  $\Delta C'_{ij,t}$  作为最合适的 PV 代理站点的 PV 发电量。式如(12)所示。

$$R_1 = \arg \max_{i,j} \text{MIC}(C'_{n,t}, \Delta C'_{ij,t}) \quad (12)$$

也就是说, 所提方法希望找出与总净负荷量  $C'_{n,t}$  相关性尽可能高的  $\Delta C'_{ij,t}$ , 因为总净负荷量  $C'_{n,t}$  中的总 PV 发电量的曲线形态应与各用户的 PV 发电量形态一致, 如果  $\Delta C'_{ij,t}$  中可包含尽可能少的负荷消耗差值  $\Delta D'_{ij,t}$  带来的噪声, 则净负荷差值  $\Delta C'_{ij,t}$  所呈现出的曲线形态将更近似于总净负荷量  $C'_{n,t}$  中的总 PV 发电量的曲线形态。至此, 最佳的 PV 代理站点可通过筛选获得。

## 2.2. 使用最大信息系数估计用户 PV 发电量

为便于表示, 将筛选所得最佳 PV 代理站点的 PV 发电量结果表示为  $\Delta P_t$ 。由式(9)可知, 用户  $i$  经缩放后的 PV 发电量  $P'_{i,t}$  可由代理站点的 PV 发电量  $\Delta P_t$  乘以转换系数  $\beta$  获得, 如式(13)所示。

$$P'_{i,t} = \beta_i \Delta P_t; i \in n \quad (13)$$

对于式(13)中的系数  $\beta_i$ , 同样可以通过 MIC 进行参数搜寻获得, 如式(14)所示。

$$R_2 = \arg \min_{\beta_i} \text{MIC}(C'_{i,t} + \beta_i \Delta P_t, \Delta P_t) \quad (14)$$

根据式(1)中净负荷、负荷消耗与光伏发电量之间的关系可知, 式中  $C'_{i,t} + \beta_i \Delta P_t$  可表示在转换系数  $\beta_i$  下估计的用户负荷消耗值。此处需注意, 式(14)中, 是希望求得 MIC 最小时的转换系数  $\beta_i$ 。这是因为希望估计所得的用户负荷消耗值  $C'_{i,t} + \beta_i \Delta P_t$ , 能尽可能不包含光伏发电量信息  $\Delta P_t$ , 也即两者之间的相关性最小。设定待优化参数  $\beta_i$  最小可取值为  $\beta_{i,\min}$ , 最大可取值为  $\beta_{i,\max}$ , 步长为  $\Delta \beta_i$ , 通过枚举的方式以获得  $R_2$  取值最小时的  $\beta_i$  值, 即为最优转换系数。

最终, 用户  $i$  的 PV 发电量分解结果可通过经缩放后的 PV 发电量  $P'_{i,t}$ , 根据式(2)的反缩放获得。如式(15)所示。

$$\hat{P}_{i,t} = P'_{i,t} \cdot \frac{1}{m} \sum_{t_1=1}^m C_{i,t_1} \quad (15)$$

所提社区用户 PV 发电量分解算法的伪代码如下:

---

### 算法: BTM 系统下社区用户 PV 分解算法

---

**输入:** 社区中共  $n$  个用户各自的净负荷  $C_{i,t}, i \in n$ 。

**输出:** 用户  $k$  的光伏发电量分解结果  $\hat{P}_{k,t}$ 。

- 1: **for**  $i=1:n$  **do**
  - 2: 对用户  $i$  的净负荷  $C_{i,t}$  根据(2)进行缩放以获得  $C'_{i,t}$ 。
  - 3: **end for**
  - 4: 计算所有用户的总净负荷  $C'_{n,t} = \sum_{i=1}^n C'_{i,t}$ 。
  - 5: **for**  $i=1:n$  **do**
  - 6: **for**  $j=i+1:n$  **do**
  - 7: 根据式(3)~(5)计算用户  $i$  和用户  $j$  之间的净负荷差值  $\Delta C'_{ij,t}$ 。
  - 8: 计算  $C'_{n,t}$  和  $\Delta C'_{ij,t}$  之间的最大信息系数  $\text{MIC}(C'_{n,t}, \Delta C'_{ij,t})$ 。
  - 9: **endfor**
  - 10: **endfor**
-

---

```

11: 根据式(12)获得  $R_1$  取最大值时的  $\Delta C_i$ 。
12: for  $\beta_k = \beta_{k,\min} : \Delta\beta_k : \beta_{k,\max}$  do
13: 计算  $C'_{k,t} + \beta_k \Delta P$  和  $\Delta P$  之间的最大信息系数  $\text{MIC}(C'_{k,t} + \beta_k \Delta P, \Delta P_t)$ 。
14: endfor
15: 根据式(14)获得  $R_2$  取最小值时的转换系数  $\beta_k$ 。
16: 根据式(13)计算  $P'_{k,t} = \beta_k \Delta P_t$ 。
17: 根据(15)反缩放  $P'_{k,t}$  以获得光伏发电量分解结果  $\hat{P}_{k,t}$ 。
返回  $\hat{P}_{k,t}$ 。

```

---

### 3. 算例分析

#### 3.1. 数据集及评价指标

为验证所提 BTM 系统下光伏发电量分解算法的有效性, 采用了由 Pecan Street Dataport 提供的美国德克萨斯州奥斯汀市以及纽约州伊萨卡岛的开源数据进行仿真[11]。经过数据预处理后, 奥斯汀市数据集包含了共 24 个居民用户, 计量时间为 2018 年 1 月 1 日至 2018 年 12 月 31 日。伊萨卡岛数据集包含了共 18 个居民用户, 计量时间为 2019 年 1 月 5 日至 2019 年 10 月 31 日。两组数据的采样频率均为每 15 分钟 1 个点。本文设日间时段为 06:30 至 17:30。

分解结果采用均方根误差(Root Mean Square Error, RMSE)和变化系数(Coefficient of Variation, CV)作为精度评价指标。RMSE 和 CV 的计算式如下:

$$\text{RMSE} = \sqrt{\frac{1}{T} \sum_{t=1}^T (\hat{P}_{i,t} - P_{i,t})^2} \quad (16)$$

$$\text{CV} = \frac{1}{N} \sum_{d=1}^N \sqrt{\frac{\sum_{t=1}^T (\hat{P}_{i,d,t} - P_{i,d,t})^2}{\sum_{t=1}^T P_{i,d,t}}} \quad (17)$$

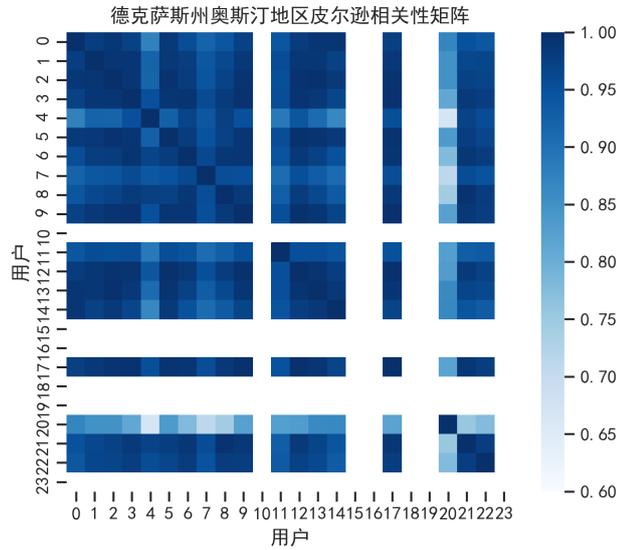
式中,  $\hat{P}_{i,d,t}$  和  $P_{i,d,t}$  为添加了日期索引的  $P_{i,t}$ ,  $d$  表示日期索引。

#### 3.2. 社区内用户的光伏发电量相关性分析

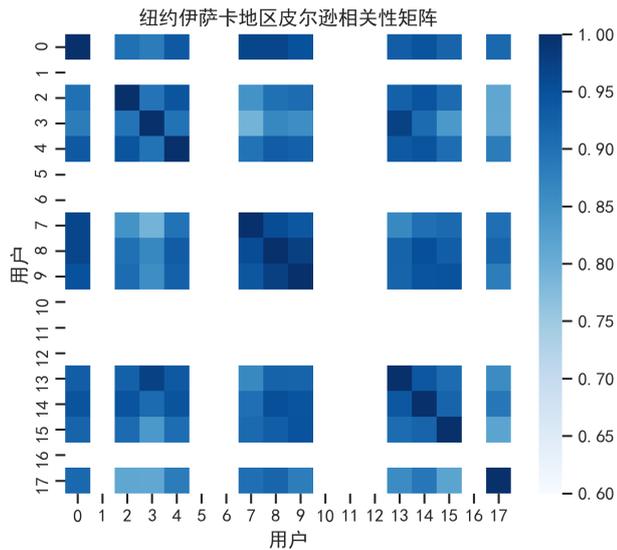
所提光伏发电量算法可实现的关键点在于社区内不同的光伏设备的 PV 发电量曲线的形态是近似的。为验证假设的合理性, 使用皮尔逊相关系数(Pearson Correlation Coefficient, PPC)来评估不同用户间光伏发电量曲线的线性相关性[12]。图 1 和图 2 以合理图的方式, 展示了奥斯汀市用户和伊萨卡岛用户的 PPC。图中横轴和纵轴表示为用户编号。PPC 的取值范围为-1 到 1, PPC 取值越接近于 1, 说明两变量间的正相关性越强。

从图 1 和图 2 中可以看出, 社区用户之间的光伏电量都具有非常强的线性相关性, PPC 数值基本大于 0.9。图 1 和图 2 中呈白色的区间表示该用户并未安装光伏设备, 因此其 PV 装机容量  $\alpha$  为 0。根据图中所呈现的相关性可知, 社区内不同的光伏设备的 PV 发电量曲线的形态是近似的这一假设是合理的。

Pecan Street Dataport 提供的用户数据集中, 并未保证用户是处于同一社区内的。这将在一定程度上影响用户间的 PV 发电量曲线的相关性。若采集的用户数据在地理位置上比较集中, 如同在一馈线下, 相关系数会更高, 这将更适用本算法所提 PV 发电量分解场景。



**Figure 1.** Pearson correlation coefficient matrix for customer PV generation in Austin  
**图 1.** 奥斯汀市用户光伏发电量皮尔逊相关系数矩阵



**Figure 2.** Pearson correlation coefficient matrix for customer PV generation in Ithaca  
**图 2.** 伊萨卡岛用户光伏发电量皮尔逊相关系数矩阵

### 3.3. 所提 PV 分解算法性能分析

在本节中, 所提 PV 分解算法将与文献[6]和[11]中提出的分解算法进行分解精度对比。需要说明的是, 文献[6]基于光伏物理模型假设的分解算法(为简便, 下称为方法 1)和文献[13]基于数据驱动的分解算法(为简便, 下称为方法 2)在建模时都需要用到太阳辐照度和温度等外生变量。相较之下, 所提分解算法在建模时只需要社区中相同类型用户的净负荷数据, 极大地降低了数据的依赖性与获取成本, 有较好的实际应用价值。方法 1 和方法 2 建模所用到的气象数据来自于美国太阳辐射数据库(National Solar Radiation Database, NSRDB) [14]。

为说明所提算法的 PV 发电量分解有效性, 以奥斯汀数据集以及伊萨卡岛数据集中的 2 号用户为例, 进行分解结果的分析。表 1 和表 2 展示了 2 号用户不同 PV 发电量分解算法下, 全年的日平均 RMSE 和 CV。

**Table 1.** Daily RMSE and CV of various disaggregation methods of user #2 in Austin, Texas of a whole year

**表 1.** 奥斯汀市 2 号用户在不同 PV 发电量分解算法下全年的日平均 RMSE 和 CV

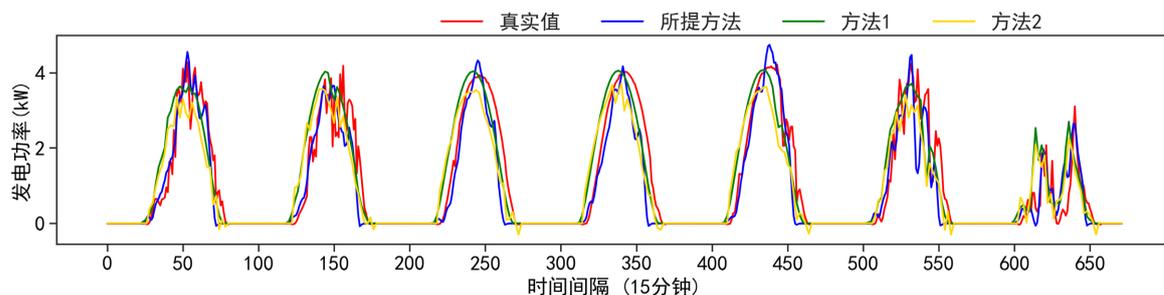
评价指标	方法 1	方法 2	所提算法
RMSE (kW)	0.550	0.543	0.540
CV (%)	8.537	7.870	8.235

**Table 2.** Daily RMSE and CV of various disaggregation methods of user #2 in Ithaca, New York of a whole year

**表 2.** 伊萨卡岛 2 号用户在不同 PV 发电量分解算法下全年的日平均 RMSE 和 CV

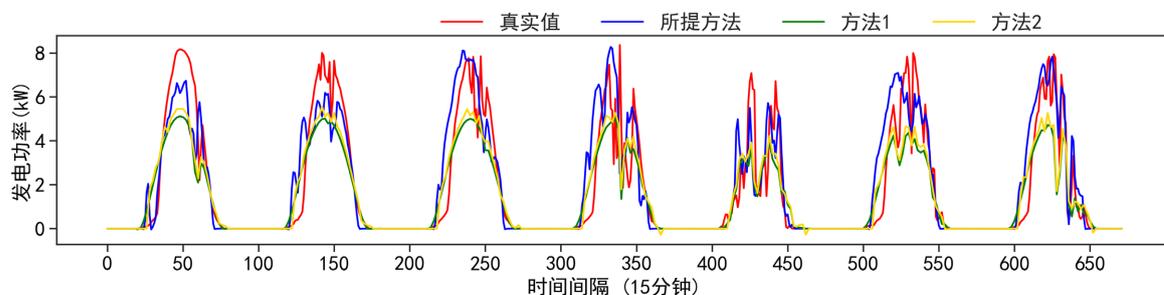
评价指标	方法 1	方法 2	所提算法
RMSE (kW)	1.006	0.926	1.070
CV (%)	7.434	6.939	9.556

从表 1 和表 2 可以看出, 三种算法具有相近的分解精度, 所提分解算法在奥斯汀市的数据集上取得了最好的 RMSE 分解结果, 但总体来说三种方法的分解误差差异很小。但是, 所提算法可以在不依赖外生的太阳辐照度和温度数据的情况下获得与其他两种方法相近的结果, 这在实际应用中具有巨大的优势。分析表 3 可知, 在伊萨卡岛数据集中, 所提算法分解结果略差于其他两种方法。其中的原因可能是纽约数据集的居民用户比奥斯汀数据集少, 因此难以搜寻到具有类似负荷消耗行为的用户, 但所提算法在如此低的用户数据量以及外生数据的需求下, CV 仅为 9.556% 是可以接受的。图 3 和图 4 展示了 2 号用户在不同 PV 发电量分解算法下一周的 PV 发电量分解结果。



**Figure 3.** Disaggregation results for one week for customer #2 in Austin of various PV generation disaggregation methods

**图 3.** 多种 PV 光伏发电量分解方法下奥斯汀市 2 号用户一周的分解结果



**Figure 4.** Disaggregation results for one week for customer #2 in Ithaca of various PV generation disaggregation methods

**图 4.** 多种 PV 光伏发电量分解方法下伊萨卡岛 2 号用户一周的分解结果

由图 3 和图 4 可以看出, 所展示的光伏发电量分解结果是针对单个用户的小型分布式发电设备的, 因此真实的光伏发电曲线会出现较多的毛刺和波动。方法 1 和方法 2 的 PV 发电量分解曲线相对来说会更加光滑, 这是因为两种方法均采用了外生的辐照度数据作为数据进行光伏发电量的分解的。外生辐照度数据是由 NSRDB 于临近气象站通过辐照度监测仪采集所得, 一般架设在较为宽阔的地带, 因此较少出现阴影遮挡的状况。但正是由于外生数据采集地点与待分解用户的 PV 发电地点不一致, 导致外生气象数据无法真实的反映待分解用户所在地点的气象状况, 引入了额外的误差。所提算法由于不依赖于外生气象数据, 因此可更加真实的反映待分解用户的实际光伏发电量情况, 分解结果也会与真实值一致, 具有更多的毛刺, 更能反映光伏发电量因如云层运动造成的阴影遮蔽等因素所导致的能量波动情况。

### 3.4. PV 发电量分解消融实验

数据集中的居民用户数量是所提算法的一个重要参数。从理论上来说, 社区用户的规模越大, 越有可能匹配具有相似用电行为的用户, 最终的 PV 发电量分解精度将会更高。为说明上述推断的合理性, 本节进行了 PV 发电量的消融实验, 分析在不同社区用户规模下的分解结果变化。

本实验采用奥斯汀市的用户数据集, 实验对象仍然为 2 号用户, 通过有放回的随机抽样方式, 选取该数据集中的其他用户, 形成用户数量分别为 8、16、24 的 3 个新数据集。表 3 展示了所包含用户数量不同的数据集下, 2 号用户的分解结果。

**Table 3.** Disaggregation results for user #2 under different data sets with different number of users  
**表 3.** 用户数量不同的数据集下 2 号用户的分解结果

评价指标	数据集中用户数量		
	8	16	24
RMSE (kW)	0.766	0.658	0.540
CV (%)	12.334	8.936	8.235

分析表 3 可知, 当数据集中用户数量下降, PV 发电量的 RMSE 和 CV 随之增加, 即 PV 发电量的分解精度下降。当数据集中用户量为原数据的三分之一时, PV 发电量的 RMSE 由 0.540 kW 上升到了 0.766 kW, CV 由 8.235% 上升到了 12.334%。造成精度下降的原因可解释为, 随着数据集中用户数量的下降, 可匹配的最优的具有相似用电行为的用户, 可能为包含有所有用户的原数据集中的次优匹配结果。因此, 所提算法若想获得更高的分解精度, 可考虑使用包含有尽可能多的用户的数据集, 并尽可能保证所有用户处于同一社区内或同一馈线下。

## 4. 结论

本文提出一种 BTM 系统下的 PV 发电量分解技术, 该技术最大的优势在于能够在不依赖于任何外生气象数据, 仅使用社区内其余用户的净负荷数据的情况下, 实现用户的 PV 发电量分解。因为较低的外生数据依赖性, 所提 PV 发电量分解算法具有较好的实际应用价值。通过使用美国德克萨斯州奥斯汀市以及纽约州伊萨卡岛的开源数据进行仿真, 结果发现, 与已有的依赖于外生气象数据的基于物理模型假设的和基于数据驱动的分解算法相比, 本文所提 PV 发电量分解算法在更少的数据量需求下, 仍获得了较高的分解性能, 与其余两种分解算法的求解结果接近。同时, 如果打算使用本文所提算法获得更高的分解精度, 应考虑使用包含有尽可能多用户的数据集, 并尽可能保证所有用户处于同一社区内或同一馈线下。

## 参考文献

- [1] Johnson, B., Denholm, P., Kroposki, B. and Hodge, B. (2017) Achieving a 100% Renewable Grid. *IEEE Power and Energy Magazine*, **15**, 61-73.
- [2] Lai, C.S., *et al.* (2017) Levelized Cost of Electricity for Photovoltaic/Biogas Power Plant Hybrid System with Electrical Energy Storage Degradation Costs. *Energy Conversion and Management*, **153**, 34-47. <https://doi.org/10.1016/j.enconman.2017.09.076>
- [3] Uzum, B., Onen, A., Hasanien, H.M. and Muyeen, S.M. (2020) Rooftop Solar PV Penetration Impacts on Distribution Network and Further Growth Factors—A Comprehensive Review. *Electronics*, **10**, 55. <https://doi.org/10.3390/electronics10010055>
- [4] Schmela, M., *et al.* (2018) Global Market Outlook for Solar Power/2018-2022.
- [5] 穆程刚, 丁涛, 曲明, 贺元康, 王永庆, 陈天恩. 基于区块链的表后微网系统及其点对点能量块交易模型设计[J]. 中国电机工程学报, 2021, 41(20): 6927-6941. <https://doi.org/10.13334/j.0258-8013.pcsee.202157>
- [6] Wang, Y., Zhang, N., Chen, Q., Kirschen, D.S., Li, P. and Xia, Q. (2017) Data-Driven Probabilistic Net Load Forecasting with High Penetration of Behind-the-Meter PV. *IEEE Transactions on Power Systems*, **33**, 3255-3264. <https://doi.org/10.1109/TPWRS.2017.2762599>
- [7] Kabir, F., Yu, N., Yao, W., Yang, R. and Zhang, Y. (2021) Joint Estimation of Behind-the-Meter Solar Generation in a Community. *IEEE Transactions on Sustainable Energy*, **12**, 682-694. <https://doi.org/10.1109/TSTE.2020.3016896>
- [8] Li, K., Wang, F., Mi, Z., Fotuhi-Firuzabad, M., Duić, N. and Wang, T. (2019) Capacity and Output Power Estimation Approach of Individual Behind-the-Meter Distributed Photovoltaic System for Demand Response Baseline Estimation. *Applied Energy*, **253**, Article ID: 113595. <https://doi.org/10.1016/j.apenergy.2019.113595>
- [9] Tabone, M., Kiliccote, S. and Kara, E.C. (2018) Disaggregating Solar Generation behind Individual Meters in Real Time. *BuildSys 2018—Proceedings of the 5th Conference on Systems for Built Environments*, Shenzhen, 7-8 November 2018, 43-52. <https://doi.org/10.1145/3276774.3276776>
- [10] 赵雅雪, 王旭, 蒋传文, 张津瑋, 周子青. 基于最大信息系数相关性分析和改进多层次门控 LSTM 的短期电价预测方法[J]. 中国电机工程学报, 2021, 41(1): 135-146+404. <https://doi.org/10.13334/j.0258-8013.pcsee.191731>
- [11] Pecan Street|Dataport. <https://dataport.pecanstreet.org/#>
- [12] 张华, 龙呈, 胡思洋, 廖凯, 高艺文. 基于层次聚类法与皮尔逊相关系数的配电网拓扑校验方法[J]. 电力系统保护与控制, 2021, 49(21): 88-96. <https://doi.org/10.19783/j.cnki.pspc.210075>
- [13] Sossan, F., Nespoli, L., Medici, V. and Paolone, M. (2018) Unsupervised Disaggregation of Photovoltaic Production from Composite Power Flow Measurements of Heterogeneous Prosumers. *IEEE Transactions on Industrial Informatics*, **14**, 3904-3913. <https://doi.org/10.1109/TII.2018.2791932>
- [14] NSRDB Data Viewer. [https://maps.nrel.gov/nsrdb-viewer/?aL=x8CI3i%255Bv%255D%3Dt%26ozt\\_aP%255Bv%255D%3Dt%26ozt\\_aP%255Bd%255D%3D1&bL=clight&cE=0&IR=0&mC=4.740675384778373%2C22.8515625&zL=2](https://maps.nrel.gov/nsrdb-viewer/?aL=x8CI3i%255Bv%255D%3Dt%26ozt_aP%255Bv%255D%3Dt%26ozt_aP%255Bd%255D%3D1&bL=clight&cE=0&IR=0&mC=4.740675384778373%2C22.8515625&zL=2)