

DOI: 10.19783/j.cnki.pspc.231594

基于联盟合作博弈的风电数据定价方法

霍巍, 张耀, 赵寒亭, 王建学

(西安交通大学电气工程学院, 陕西 西安 710049)

摘要: 为发挥风电数据价值并给企业带来额外收益, 提出一种基于联盟合作博弈模型的风电数据定价方法。首先, 提出一种考虑第三方监管的风电数据交易流程与风电数据定价思路。其次, 通过数据交易获取邻近场站数据, 使用向量自回归模型提高风电预测精度, 从而在电力市场交易中降低不平衡成本, 提高风电场整体发电利润, 体现所交易数据的整体价值。然后, 将数据交易前后产生的超额利润(即使用交易数据后的发电利润增加值), 以联盟合作博弈的方式分配给数据售卖方作为交易数据的定价结果, 实现数据价值的定量衡量。最后, 考虑风电时空特性通过最大化数据价值确定数据交易对象, 实现风电数据的合理定价。算例分析结果表明, 所提方法能够提升风电预测精度、增加风场发电利润, 并实现风电数据定价的公平性与合理性。

关键词: 电力数据交易; 数据定价; 风电预测; 电力市场; 合作博弈

Wind power data pricing method based on an alliance cooperation game

HUO Wei, ZHANG Yao, ZHAO Hanting, WANG Jianxue

(School of Electrical Engineering, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: To give full play to the value of wind power data and bring extra profit to a wind farm, this paper proposes a wind power data pricing method based on alliance cooperation game theory. First, a wind power data trading process and pricing approach considering third-party supervision are proposed. Secondly, adjacent station data is obtained through data trading, and a vector autoregressive model is used to improve wind power forecasting accuracy, thereby reducing imbalance costs in electricity market transactions, increasing the overall generation profit of wind power plants, and reflecting the value of the traded data. Then, the excess profit generated before and after data trading (i.e., the increase in generation profit using traded data) is distributed to the data sellers as the pricing of the traded data using an alliance cooperation game, quantitatively measuring the value of the data. Finally, considering the spatio-temporal characteristics of wind power, the data trading parties are determined by maximizing the data value, achieving fair and reasonable data pricing. The case analysis results demonstrate that the proposed method can enhance wind power forecasting accuracy, increase wind farm generation profits, and achieve fairness and rationality in wind power data pricing.

This work is supported by the National Key Research and Development Program of China "Key Technologies on Intelligent Dispatch of Power Grid under 20% New Energy Integration Scenario" (No. 2022YFB2403500).

Key words: electric data trading; data pricing; wind power forecasting; electricity market; cooperative game

0 引言

近年来, 随着全球范围内能源问题与环境问题的日渐显著, 以风电和光伏为主的间歇性可再生能源发电受到全世界范围内的广泛关注并迅速发展^[1]。为提高风力资源利用率, 风电场需对未来发电量进行

准确预测, 以制定更合适的生产计划。众多研究显示, 风力发电的时空相关性意味着利用邻近风电场的的数据可以显著提升预测的准确度^[2-4], 从而为风电企业在电力市场交易中提供竞争优势, 增加其收入。然而, 现有研究往往忽视了一个重要事实: 不同风电场可能属于不同的发电公司, 它们之间由于商业竞争, 通常不愿意或不可能免费分享数据^[5]。解决上述阻碍的一种可行思路是通过数据交易(data trading)购买邻近场站的相关数据, 掌握周边地区风

基金项目: 国家重点研发计划项目资助“支撑20%新能源电量占比场景下的电网智能调度关键技术”(2022YFB2403500)

力资源的时空分布,从而使提高自身场站的预测精度成为可能。

数据定价(data pricing)在数据交易过程中扮演着核心角色,它是确保交易公正性的关键标准^[6]。合理公平的数据定价可以同时保证交易双方的利益,实现售卖方数据价值的创造,以及购买方数据价值的再利用;而不恰当的数据定价,不仅会造成双方利益损失或隐私泄露,还会破坏数据交易市场的健康发展。因此,在风电数据交易中寻找一种公平合理的数据定价方法,以推动风电预测精度的提高,已成为一个亟待解决的问题。

目前,已有众多针对数据定价理论方法的相关研究^[6]。本文根据衡量数据价值的主要因素将数据定价方法大致分为3种思路,但对于某一具体的定价方法可能同时具有不同的定价思路^[7]。第一、基于经济学的定价思路主要依赖于交易市场的建立和机制的设计,并以博弈论或拍卖等形式进行。文献^[8]基于博弈论以自身收益最大化为目标,并根据不同场景设计对应策略,计算出纳什均衡实现数据定价。拍卖作为目前最流行的数据定价机制之一,通常以市场驱动的模式,由交易双方自主竞价确定价格^[9]。第二、基于价值的定价思路更关注于数据的内在价值,例如基于隐私补偿的定价方法衡量了卖家提供的数据隐私损失成本,现阶段研究主要采用了差分隐私的计算方法^[10-11],也有研究提出了使用信息熵作为隐私损失的度量方法^[12];基于数据质量的定价方法通常使用多个指标在不同维度对数据质量进行刻画,以打分的形式给出数据质量决定的定价结果^[13-14]。第三、基于任务的定价思路更侧重于衡量购买方的数据使用效益,例如基于查询的定价方法通常根据买家的查询任务,灵活确定任意查询组合实现数据定价^[15-16];基于模型的定价方法主要是为了解决机器学习中难获取数据的困境。起初,解决这一难题的思路是直接对所售卖的数据进行定价^[17]。之后,文献^[18-19]则考虑售卖训练好的机器学习模型,并按照模型精度来确定数据价格。而文献^[20]提出了一种新的思路,即购买方直接向第三方提出预测任务并获得预测结果,而非获得数据或模型。

尽管大量研究表明,使用邻近风电场数据能显著提升风电预测精度,风电数据定价领域的相关研究却相对匮乏,目前尚未建立一套合理的定价模型以促进风电数据交易。因此,在“双碳”目标下,随着我国新能源装机容量的快速增长以及电力市场的持续深化改革^[21],研究一种可行的数据定价方法用于提高风力发电预测精度,为数据交易提供货币

化刺激,激励不同场站彼此交换商业数据,进一步发挥风电数据价值显得愈发重要。

针对上述问题,本文提出一种基于预测任务的风电数据定价方法,通过数据交易的方式,以更好地利用风力资源的时空相关性提升预测精度。本文所提的数据定价方法首先通过数据交易获取邻近风电场的商业数据,利用时空相关性得到更高精度的风电功率预测结果;然后使用更精确的预测结果从电力市场交易中获取更高的利润;最后基于合作博弈模型,采用 Shapley 值法与核仁法将超额利润分配给数据售卖方^[22],实现风电数据的合理定价。此外,考虑到风力资源具有时空相关性,本文还探讨了进行长期数据交易的可行性和必要性,并提出了筛选最佳数据合作对象的方法。

1 风电数据交易与定价思路

1.1 风电数据交易流程

风力发电的随机性要求风电场进行准确的风电功率预测,预测的精度也将影响风电场从电力市场交易中获得的总利润。一般而言,风电预测精度越高,风电场获得的总利润越多。考虑到风力资源的时空相关性,空间上邻近的风电场群在时间上可能呈现风力资源的超前或滞后现象^[23]。因此,在风电预测模型中,利用其他邻近场站的风电数据通常能显著提升预测精度,进而增加发电利润,使得相互邻近的风电场存在交换数据的动力。然而,鉴于不同场站间可能存在的商业竞争关系,这种数据交换需要通过数据交易的方式来实现。

本文所提出的基于预测任务的风电数据定价方法,需要依赖第三方搭建数据交易平台实现数据定价过程^[20],而数据购买方与售卖方作为价格的接受者不以任何形式直接参与定价过程,这也保证了数据交易的公平性。风电数据交易流程如图1所示。首先,参与交易的各方需向数据交易平台提供它们的风电数据信息,这标志着一次数据交易的开始。其次,数据交易平台担任核心角色,承担所有风电预测与数据定价任务。最后,数据交易平台向数据购买方提供风电预测结果,并将数据定价结果返还给所有交易参与方,完成数据交易。数据交易各方在数据的直接交换、预测任务和定价任务中均不直接参与,由此保证整个数据交易的公平性。

1.2 风电数据定价基本思路

数据定价是一种能充分体现其真实价值并能平衡买方效用和卖方收益的估值行为^[7]。同时,数据定价是数据交易中最重要的一环,使用合理的定价

方法能够使各方获利, 提高数据交易各方的积极性^[24]。本文基于预测任务对风电数据进行定价, 通过比较使用所交易风电数据前后的电力市场价值增量, 即购买数据带来的利润增加值, 来对风电数据进行合理定价。

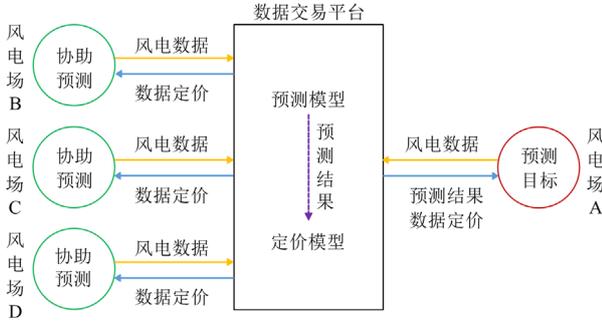


图 1 风电数据交易流程

Fig. 1 Wind power data trading flow

下面具体介绍本文提出的风电数据定价思路。首先风电场需要通过参与电力市场交易来确定市场出清电量并获得相应收入。本文采用了电力市场发展较为成熟的北欧电力市场运营模式, 风力发电商主要参与现货市场的交易^[25]。考虑到风电场是价格接受者(price taker), 因此风电场在日前市场(elspot)期间报量不报价, 它一般根据风电预测结果确定报量结果。然后经市场统一结算, 确定出清电价与出清发电量。风电场根据出清结果, 获得由日前市场支出的电能费用作为发电收入。在平衡调节市场(regulation market)期间, 由于风电具有随机性, 风电场通常会产生不平衡电量(即市场出清电量与实际发电量的差额), 需要支付不平衡成本, 作为实际发电量与出清发电量不符的惩罚。

在风电功率预测环节中, 风电场可以通过数据交易获取其他邻近场站的数据, 从而提高风电预测精度, 使日前市场的出清发电量更接近于实际发电量, 进而减小不平衡成本的惩罚, 提高风电场整体利润。本文将这部分“超额利润”(即利用其他场站数据后的发电利润增加值)用来评估所交易数据的价值。需要注意的是, 风电数据交易是“一对多”形式, 即一个风电场可以同时与多个风电场进行数据交易, 以便更好地掌握风力资源的时空分布特征。因此, 为了衡量每个售卖方的风电数据的真实价值, 本文运用合作博弈理论实现了对超额利润的分配, 将超额利润以合理公平的方式分配给所有数据提供者, 最终实现风电数据定价。

2 风电数据定价方法

2.1 考虑数据交易的风电概率预测方法

2.1.1 风电功率概率分布模型

由于风力资源具有不确定性, 因此可以采用“概率预测”来量化这种不确定性, 概率预测方法可以提供风电功率概率分布的预测结果。在使用装机容量对风电功率进行归一化处理, 风电功率将被限制在 0 与 1 之间, 具有双边界特点; 同时在针对单个风电场而非聚合区域时, 风电功率为 0 或 1 的概率往往并不等于零^[26]。如果将风电功率概率分布视为一种截断的无边界连续分布(如截断高斯分布)或有边界连续分布(如 Weibull 分布), 都忽视了风电功率停发或满发的潜在可能。为此, 本文采用一种离散-连续混合型概率分布来建模风电不确定性^[27], 其中连续部分服从 logit-normal 概率分布 $L(\mu, \sigma^2)$, 概率密度函数为

$$f(x) = \frac{1}{\sqrt{2\pi\sigma x(1-x)}} \exp\left\{-\frac{(l(x)-\mu)^2}{2\sigma^2}\right\} \quad (1)$$

式中: μ 与 σ 分别是 logit-normal 概率分布的位置参数和尺度参数; $l(x)$ 是 logit 变换函数, $l(x) = \ln\{x/(1-x)\}$, $x \in (0, 1)$ 。离散部分的概率质量 w_0 和 w_1 分别集中在 $x=0$ 和 $x=1$ 的两处边界上, 因此风电功率的概率分布 X 可以表示为

$$X \sim \delta_0 w_0 + \delta_1 w_1 + (1-w_0-w_1)L(\mu, \sigma^2) \quad (2)$$

式中, δ_x 为 x 处的狄利克雷函数, 边界上的概率质量 w_0 与 w_1 可由式(3)和式(4)计算得出。

$$w_0 = \Phi\left\{\frac{l(\eta)-\mu}{\sigma}\right\} \quad (3)$$

$$w_1 = 1 - \Phi\left\{\frac{l(1-\eta)-\mu}{\sigma}\right\} \quad (4)$$

式中: Φ 为标准正态分布的累积分布函数; η 表示风电功率量测精度, 当实际风电功率小于 η 或大于 $1-\eta$ 时, 可以将风电功率分别视为 0(停机)或 1(满发)。本文选取 η 为对应风电场的实际最小非零功率。

从上述公式可知, 风电功率概率分布模型的关键待估计参数为位置参数 μ 和尺度参数 σ 。在 t 时刻预测 $t+k$ 时刻风电功率 X_{t+k} 的概率分布时, 首先需要将历史风电功率时间序列 $\{x_t, x_t \in [0, 1]\}$, 经 logit 变换转化为新时间序列 $\{y_t, y_t \in \mathbb{R}\}$ 。然后使用第 2.1.2 节或第 2.1.3 节介绍的预测模型估计位置参数 $\hat{\mu}_{t+k}$ 和尺度参数 $\hat{\sigma}_{t+1}$, 最后再经过 logit 逆变换得到原始风电功率预测结果。

2.1.2 仅使用本站数据的自回归模型

自回归(autoregressive, AR)模型是经典的时间序列建模方法,也是风电预测传统方法之一。自回归模型以待预测变量的历史数据作为特征变量,将其进行线性组合得到预测结果^[23]。将 t 时刻待预测变量的时间序列 $\{y_t\}, y_t \in \mathbb{R}$ 建立为 p 阶的 AR 模型,则未来 $t+k$ 时刻的待预测变量 y_{t+k} 可表示为

$$y_{t+k} = \sum_{\tau=1}^{p_{AR}} a_{\tau} y_{t-\tau+1} + \varepsilon_{t+k} \quad (5)$$

式中: p_{AR} 是 AR 模型的阶数; a_{τ} 是滞后 τ 阶的自回归系数; $y_{t-\tau+1}$ 是待预测变量滞后 τ 阶的历史观测值; ε_{t+k} 是 AR 模型的预测误差,服从期望为零的正态分布。求解自回归模型可以得到位置参数预测结果,如式(6)所示。

$$\hat{\mu}_{t+k} = \sum_{\tau=1}^{p_{AR}} a_{\tau} y_{t-\tau+1} \quad (6)$$

式中, $\hat{\mu}_{t+k}$ 是自回归模型预测的未来 $t+k$ 时刻的位置参数。本文取 $k=1$,即只进行提前一步预测,方差预测结果可由模型的残差表示。

$$\hat{\sigma}_{T+1} = \sqrt{\frac{1}{T-l} \sum_{i=1}^T (y_i - \hat{\mu}_i)^2} \quad (7)$$

式中: T 是历史观测点的个数; $\hat{\sigma}_{T+1}$ 是预测未来 $T+1$ 时刻方差参数; l 是待预测变量的个数; $\hat{\mu}_i$ 是自回归模型预测 t 时刻的位置参数。之后所提的向量自回归模型的方差估计方法同上。

2.1.3 使用邻近场站数据的向量自回归模型

风电场通过数据交易可以获取邻近场站的相关数据,因此在风电功率预测模型中,需要引入邻近场站数据作为额外的特征变量。为此,本文采用向量自回归(vector autoregressive, VAR)预测模型,其作为自回归模型的一种变种模型^[28],可以利用变量之间的线性相关特征,以提升各个变量的预测准确性和可靠性。

首先将 M 个场站的时间序列组装成向量形式的时间序列 $\{y_t\}, y_t \in \mathbb{R}^M$, 其中既包括待预测场站也包括邻近场站。然后将其建立为 p 阶的 VAR 模型,则未来 $t+k$ 时刻的待预测变量 y_{t+k} 可表示为

$$y_{t+k} = \sum_{\tau=1}^{p_{VAR}} A_{\tau} y_{t-\tau+1} + \varepsilon_{t+k} \quad (8)$$

式中: p_{VAR} 是 VAR 模型的阶数; $A_{\tau} \in \mathbb{R}^{M \times M}$ 是滞后 τ 阶的向量自回归系数矩阵; $\varepsilon_{t+k} \in \mathbb{R}^M$ 是 VAR 模型的预测误差,服从期望为零的多元正态分布。求解向量自回归模型得到的位置参数预测结果如式(9)所示。

$$\hat{\mu}_{t+k} = \sum_{\tau=1}^{p_{VAR}} A_{\tau} y_{t-\tau+1} \quad (9)$$

式中: $\hat{\mu}_{t+k}$ 是向量自回归模型预测的未来 $t+k$ 时刻的位置参数; $y_{t-\tau+1}$ 是待预测变量滞后 τ 阶的历史观测值向量。

对比式(5)与式(8)可以看出,向量自回归模型除了使用待预测场站的自身数据,还使用了其他邻近场站的历史数据。因此与 AR 模型相比,VAR 模型可以提升预测精度,获得更准确的风电功率预测结果,从而体现邻近场站数据交易的价值。

在本文所提出的数据交易模式中,需要实现仅使用自身数据与使用邻近场站数据进行预测,为简化问题,仅采用了 AR 与 VAR 两种预测模型进行风电数据交易效益的对比,在实际的数据交易中,数据交易平台可以选择更有效的预测模型,实现数据效益最大化。

2.2 基于概率预测的风电最优投标策略

2.2.1 风电场参与电力市场交易的结算模型

为削减风电出力波动带来的影响,北欧电力市场为风电场商制定了相应的惩罚机制,对发电偏差引起的不平衡电量采用“双价格”结算机制^[29]:若风力发电不平衡不利于系统平衡(即加剧了系统不平衡程度),则采用惩罚性价格结算不平衡电量,反之,则采用常规价格结算。其不平衡结算方式如表 1 所示,其中 π_u 和 π_d 分别表示系统上调、下调服务价格, π_s 是日前市场出清价格。

表 1 双价格结算机制下的风力发电结算

Table 1 Settlement of wind power generation under bilateral pricing mechanism

| 发电方 不平衡类型 | 系统不平衡时段 | | |
|--------------|--------------|--------------|--------------|
| | 上调时段 | 下调时段 | 平衡时段 |
| 负不平衡 | 上调价格 π_u | 出清价格 π_s | 出清价格 π_s |
| 正不平衡 | 出清价格 π_s | 下调价格 π_d | 出清价格 π_s |

发电方负不平衡表示实际发电量小于出清电量;正不平衡表示实际发电量大于出清电量。根据不平衡电量的双价格结算方式,可以推导风电场参与市场交易后的利润函数。对于任意交易时段,假设 q 为日前市场中确定的出清电量, Q 为风电场实际发电量, π_c 表示单位发电成本,风电场利润函数 $H(q, Q)$ 可以写为

$$H(q, Q) = \begin{cases} \pi_s q - \pi_u (q - Q) - \pi_c Q, & \text{if } q > Q, \text{系统上调} \\ \pi_s q + \pi_s (Q - q) - \pi_c Q, & \text{if } q < Q, \text{系统上调} \\ \pi_s q - \pi_s (q - Q) - \pi_c Q, & \text{if } q > Q, \text{系统下调} \\ \pi_s q + \pi_d (Q - q) - \pi_c Q, & \text{if } q < Q, \text{系统下调} \end{cases} \quad (10)$$

式(10)表明风电场利润由日前市场出清收入、不平衡结算成本、机组发电成本三部分组成。进一步化简得

$$H(q, Q) = \begin{cases} \pi_w Q - \lambda_u \max(q - Q, 0), & \text{if 系统上调} \\ \pi_w Q - \lambda_d \max(Q - q, 0), & \text{if 系统下调} \end{cases} \quad (11)$$

$$\begin{cases} \lambda_u = \max(0, \pi_u - \pi_s) \\ \lambda_d = \max(0, \pi_s - \pi_d) \end{cases} \quad (12)$$

式中: λ_u 为上调超出部分价格; λ_d 为下调降低部分价格; $\pi_w = \pi_s - \pi_c$ 为风电场单位发电净利润。

2.2.2 基于最大化期望利润的风电报量策略

式(11)表示的风电场利润函数中存在两个随机变量, 即系统平衡状态(上调或下调)和未来风电功率。为了简化模型, 假设两个随机因素相互独立, 根据式(13)计算风电场期望利润。

$$\mathbb{E}_Q[H(q, Q)] = \pi_w \mathbb{E}[Q] - \lambda'_u \mathbb{E}[\max(q - Q, 0)] - \lambda'_d \mathbb{E}[\max(Q - q, 0)] \quad (13)$$

$$\begin{cases} \lambda'_u = p_u \lambda_u \\ \lambda'_d = p_d \lambda_d \end{cases} \quad (14)$$

式中: $\mathbb{E}_Q[H(q, Q)]$ 为风电场投标电量为 q 、发电功率服从概率分布 Q 时的期望利润; $\mathbb{E}[\cdot]$ 表示对概率分布求期望值; λ'_u 为上调超出部分修正价格; λ'_d 为下调降低部分修正价格; p_u 和 p_d 表示系统不平衡状态概率, 即上调状态概率和下调状态概率。

在参与电力交易时, 风电场需要在日前市场中确定投标电量, 考虑到风电优先出清, 因此风电场的投标电量与出清电量相等。本文以最大化风电期望利润为目标, 建立如式(15)所示的随机优化问题, 用于确定风电场的最优投标电量。

$$\max_q \mathbb{E}_Q[H(q, Q)] \quad (15)$$

由 2.1 节可知, 风电实际功率分布服从离散-连续型 logit-normal 概率分布, 具体形式见式(2), 然后将其代入式(13)并化简。由于期望利润可看成投标电量的函数, 为求解最大值问题即式(15), 分别计算期望利润关于出清电量的一阶导数与二阶导数, 分别如式(16)和式(17)所示。

$$\frac{d\mathbb{E}_Q}{dq} = (1 - w_0 - w_1)[\lambda'_d - F_L(q)(\lambda'_d + \lambda'_u)] + \lambda'_d w_1 - \lambda'_u w_0 \quad (16)$$

$$\frac{d^2\mathbb{E}_Q}{dq^2} = -(1 - w_0 - w_1)(\lambda'_d + \lambda'_u)f_L(q) \quad (17)$$

式中, $F_L(q)$ 与 $f_L(q)$ 分别为 logit-normal 分布的累积分布函数与概率密度函数。根据概率分布参数与市场结算参数的定义可知, 期望利润函数的二阶导数非正, 因此其为凹函数。对于全局可微的凹函数,

其全局最大值点的充要条件是该点的一阶导数为零, 令一阶导数即式(16)等于零, 可解得风电场期望利润取最大值时的最优投标电量为

$$q^* = \begin{cases} 0, & \gamma' \leq 0 \\ F_L^{-1}(\gamma'), & 0 < \gamma' < 1 \\ 1, & \gamma' \geq 1 \end{cases} \quad (18)$$

$$\gamma' = \frac{\gamma - w_0}{1 - w_0 - w_1} \quad (19)$$

式中: q^* 是风电场期望利润取最大值时的最优投标电量; F_L^{-1} 是 logit-normal 累积分布函数的反函数; γ' 是考虑风电概率分布的不平衡价格修正费用比; $\gamma = \lambda'_d / (\lambda'_d + \lambda'_u)$ 是不平衡价格费用比。

需要说明的是, q^* 是按装机容量进行归一化的投标电量, 因此 $q^* = 0$ 表示最优投标电量为 0, 而 $q^* = 1$ 表示最优投标电量为风电场装机容量。除上述两种极端情况外, 当 $0 < \gamma' < 1$ 时, 风电场的最优投标电量 q^* 等于风电功率发电量 Q 置信水平为 γ' 的分位数, 对应的期望利润为 $\mathbb{E}_Q[H(q^*, Q)]$ 。

2.3 基于合作博弈的超额利润分配模型

本节核心是建立基于合作博弈的利益分配模型, 解决待预测风电场与多个数据提供方之间的数据定价问题。

2.3.1 联盟合作博弈模型

博弈(game)是指所有成员在一定的游戏规则下根据自身所掌握的信息, 以实现自身利益最大化或风险最小化为目标, 考虑采取最优策略进行行动的过程。博弈模型通常分为合作博弈(cooperative game)与非合作博弈(non-cooperative game), 其中合作博弈主要用于考虑合作联盟中各成员内部利益的分配问题^[30]。

风电场使用其他场站数据进行合作预测是一种典型的合作博弈, 可以采用可转移支付联盟博弈的方式, 使用相同度量方式作为联盟价值函数, 衡量各参与成员的贡献, 建立联盟合作博弈模型。

首先定义待预测风电场为成员 0, 并称其为“目标风电场”。假设一共有 M 个风电场协助目标风电场进行合作预测, 定义这些协助风电场依次为成员 1、成员 2、成员 3、...、成员 M , 则所有参与合作预测的风电场就组成了合作博弈联盟 $\mathcal{M} = \{\text{成员 } 0, \text{成员 } 1, \dots, \text{成员 } M\}$ 。

联盟价值特征函数的自变量取值集合为参与合作博弈的所有子联盟 $\mathcal{S} (\mathcal{S} \subseteq \mathcal{M})$, 其函数值定义为“目标风电场参与该子联盟合作获得的期望利润”与“不参与合作仅使用自身数据获得的期望利润”

之间的差值,即前面提及的“超额利润”,联盟价值特征函数 $v(S)$ 可表示为

$$v(S) = \begin{cases} 0, & \text{if } 0 \notin S \\ \mathbb{E}_{Q_M}[H(q_S^*, Q_M)] - \mathbb{E}_{Q_M}[H(q_{\{0\}}^*, Q_M)], & \text{if } 0 \in S \end{cases} \quad (20)$$

式中: Q_M 为目标风电场参与总联盟 \mathcal{M} 合作,使用VAR模型所得到的风电功率概率分布; $\mathbb{E}_{Q_M}[H(q, Q_M)]$ 为风电场投标电量为 q 、实际发电功率服从概率分布 Q_M 时的期望利润函数; q_S^* 为目标风电场参与子联盟 S 合作后,使用其他场站数据进行VAR预测所确定的最佳出清电量; $q_{\{0\}}^*$ 为目标风电场仅使用自身数据进行AR预测所确定的最佳出清电量。

2.3.2 联盟价值分配法则

在风电数据交易后,需要对使用数据产生的超额利润,即联盟价值特征函数进行合理分配,将其作为数据购买价格支付给数据售卖方。本节将介绍两种分配方法:Shapley值法、核仁法。

首先定义一般意义上的利润分配模型,考虑在合作博弈模型中,成员 i 所分配到的利润为 c_i ,则应保证“总联盟价值原则”,即所有成员分配到的利润之和等于总联盟价值,如式(21)所示。

$$\sum_{i \in \mathcal{M}} c_i = v(\mathcal{M}) \quad (21)$$

式中, $v(\mathcal{M})$ 为 S 取为纵联盟 \mathcal{M} 时的联盟价值特征函数值。

同时,还应保证“成员价值原则”,即每个成员所分配到的利润不小于独自参与博弈时获得的价值。在本研究中,单个成员参与合作博弈不会产生超额利润,因此“成员价值原则”要求每个成员分配到的利润不小于0,如式(22)所示。

$$c_i \geq v(\{i\}) = 0, \quad \forall i \in \mathcal{M} \quad (22)$$

Shapley值法^[31]通过比较成员 i 在所有子联盟 S 中参与和不参与博弈时产生的联盟价值函数差值,并考虑子联盟 S 中成员个数作为权重系数,最后进行加权平均计算出Shapley值 c_i ,如式(23)所示。

$$c_i = \sum_{S \subseteq \mathcal{M}} \frac{(|S|-1)! (|\mathcal{M}|-|S|)!}{|\mathcal{M}|!} [v(S) - v(S \setminus \{i\})] \quad (23)$$

式中: $|S|$ 表示子联盟 S 中参与成员的总数; $S \setminus \{i\}$ 表示子联盟 S 去除成员 i 后的成员集合。

Shapley值法按Shapley值进行价值分配,对所有成员的Shapley值求和可得 $\sum_{i \in \mathcal{M}} c_i = v(\mathcal{M})$,即满足

总联盟价值原则。然而它不能保证成员价值原则,满足成员价值原则的一个必要条件是联盟价值函数是广义单调增函数,即满足式(24)。

$$T_1 \subseteq T_2 \Rightarrow v(T_1) \leq v(T_2) \quad (24)$$

式中, T_1, T_2 为总联盟 \mathcal{M} 中的任意子联盟。

本文建立的联盟价值函数 $v(S)$ 并不满足式(24)。因为对某个子联盟 \mathcal{T} 来说,增添一个风电场的数据进行合作预测,并不意味着在所有时段都能提升预测精度并产生超额利润。此外,某风场在某时段分配到的利润为负值,也不意味着需要把该风场从合作对象中剔除。一般推荐在尽量长的时段内,综合考虑该风场数据是否带来超额利润,再判断是否值得将其选为合作对象。

核仁法通过定义每个子联盟 S 在合作博弈时的不满意度,并最小化每个成员对分配利润的不满意度确定分配价值结果^[32]。首先定义子联盟 S 在合作博弈时的不满意度 $o(c, S)$,用式(25)表示。

$$o(c, S) = v(S) - \sum_{i \in S} c_i \quad (25)$$

式中, c 表示全部成员的利润分配向量。式(25)所定义的不满意度可以解释为:联盟 S 的价值函数与联盟内成员分配利润和的差值。

然后建立一个最小-最大优化问题,其目标是 minimized 所有子联盟中最大不满意度,决策变量是成员利润分配向量 c ,该优化问题可表述为式(26)。

$$\begin{cases} \min_x \max_{S \subseteq \mathcal{M}} o(c, S) \\ \text{s.t. } o(c, \mathcal{M}) = 0 \\ c_i \geq 0, \forall i \in \mathcal{M} \end{cases} \quad (26)$$

上述优化问题的两个约束条件分别表示总联盟价值原则与成员价值原则。经证明该优化问题是凸优化问题^[33],因此在使用核仁法进行分配时,结果同样存在且唯一。

通过上述建立的合作博弈模型,就能将第2.2节所衡量的进行风电数据交易产生的超额利润,以合理公平的方式分配给各售卖方,这体现了对每个售卖方提供的风电数据价值的定量估计,最终实现对风电数据的定价。

3 考虑风电时空特性的数据定价方法

在设计出数据定价方法后,为了充分挖掘邻近场站的数据价值,需要选择合适的数据交易对象。考虑到风力资源的时空相关性,只有选择合适的数据交易对象,才能实现风电预测精度的提高和风场利润的最大化。因此,在第2节设计的风电数据定价方法基础上,进一步对潜在的数据交易对象进行筛选是至关重要的,这也能进一步确保数据定价的合理性和公平性。

为降低后续筛选的计算负担,首先对潜在的交易对象进行初步筛选,基本方法是通过先验指标评

估潜在合作场站数据与待预测场站数据的相似程度。一般而言, 相似程度越高的数据越有利于提升预测精度并获得更多超额利润。本文所采用的离散-连续混合概率分布中连续部分为 logit-normal 概率分布, 风电功率经 logit 变换后可以认为服从正态分布, 因此可以使用 Pearson 相关系数评估两两场站数据之间的相似程度。同时, 对于本文采用的向量自回归这一线性预测模型, Pearson 相关系数可以筛选出线性相关程度更高的风电场群, 以期实现预测精度的提高。

初步筛选主要考虑数据的相似程度, 而在选择交易对象时, 最终目标是为了获取更高的期望利润。如果合作对象过多, 可能导致单个合作对象分配到的利润减少, 从而减弱其参与交易的动力。因此, 作为数据购买方的风电场需要着重考虑交易后的净利润最大化来确定最佳的合作对象。同时, 作为数据提供方的风电场可以选择第 2.3 节中对协助风电场有利的分配方法, 尽可能提升其交易意愿。

将经过初步筛选后的风场联盟记为 \mathcal{D} , 经过最终筛选后的风场联盟记为 \mathcal{N} , 可知 \mathcal{N} 是 \mathcal{D} 的一个子联盟。为此需要度量不同合作对象联盟 \mathcal{N} 的优劣程度, 故定义待预测场站的剩余利润 $W(\mathcal{N})$, 其计算公式如式(27)所示。

$$W(\mathcal{N}) = \mathbb{E}_{Q_{\mathcal{N}}} [H(q_{\{0\}}^*, Q_{\mathcal{N}})] + c_{0,\mathcal{N}} = \mathbb{E}_{Q_{\mathcal{N}}} [H(q_{\mathcal{N}}^*, Q_{\mathcal{N}})] - \sum_{j \in \mathcal{N}, j \neq 0} c_{j,\mathcal{N}} \quad (27)$$

式中: $c_{0,\mathcal{N}}$ 为待预测场站在子联盟 \mathcal{N} 中分配到的利润; $c_{j,\mathcal{N}}$ 为合作场站 j 在子联盟 \mathcal{N} 中分配到的利润。待预测场站的剩余利润可表示为, 仅使用自身数据进行预测获得的期望利润 $\mathbb{E}_{Q_{\mathcal{N}}} [H(q_{\{0\}}^*, Q_{\mathcal{N}})]$, 加上在子联盟 \mathcal{N} 中合作预测后分配到的利润 $c_{0,\mathcal{N}}$; 同时还可以表示为, 使用子联盟 \mathcal{N} 中其他场站数据进行预测获得的期望利润 $\mathbb{E}_{Q_{\mathcal{N}}} [H(q_{\mathcal{N}}^*, Q_{\mathcal{N}})]$, 减去全部合作场站分配到的总利润 $\sum_{j \in \mathcal{N}, j \neq 0} c_{j,\mathcal{N}}$ 。

由于风力资源时空相关性存在季节性变化, 为提高数据交易的稳定性, 同时充分利用时空相关性, 可设置合适的风电数据交易周期, 将交易周期内待预测场站的总剩余利润作为优化目标, 来确定最佳合作对象, 由此建立优化问题如式(28)所示。

$$\max_{\mathcal{N}} \sum_{t=0}^{T_c} W_t(\mathcal{N}), \mathcal{N} \subseteq \mathcal{D} \quad (28)$$

式中: T_c 为风电场数据交易周期; W_t 为待预测场站在 t 时刻的剩余利润。在求解上述优化时, 采用枚举法统计出交易周期内所有可能的子联盟 $\mathcal{N} \subseteq \mathcal{D}$

的总剩余利润, 进而确定出最大总剩余利润的风电场联盟 \mathcal{N} 。

4 算例分析

4.1 算例设定

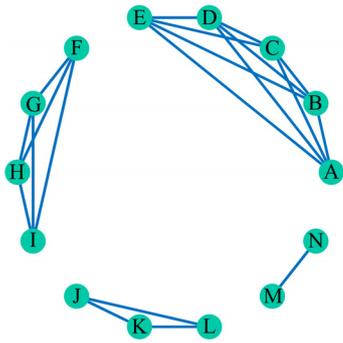
本文所使用数据集包括两部分: 1) 开放数据集的 22 个风电场的归一化发电功率数据^[23], 共计 731 天, 原始时间分辨率为 5 min, 采用“窗口平均法”将时间分辨率转化为 1 h; 2) 北欧电力市场出清价格与不平衡价格数据, 共计 365 天, 时间分辨率为 1 h。电价单位为丹麦克朗 DKK, 因此算例分析中的利润单位为 DKK/MW。

首先按照第 3 节中介绍的数据交易对象筛选方法, 初步筛选出多个线性相关程度较高的风电场群。将 22 个风电场依次命名为 A、B、...、V, 并计算两两之间的 Pearson 相关系数。图 2(a)展示了它们之间相关系数的可视化结果, 可以看到部分风电场群具有较高的空间相关性。图 2(b)展示了相关系数大于 0.6 的风电场群拓扑关系, 因此通过相关系数可以初步筛选出 4 个风电场群, 分别由 5 个、4 个、3 个、2 个风电场组成。本文将选取 A、B、D、E、F 这 5 个风电场组成风电场群进行数据交易算例分析。

针对该风电场群, 首先任选一个目标风电场作为数据购买方, 其余 4 个为数据售卖方。按照第 2 节的数据定价方法, 首先针对目标风电场的预测问题, 划分训练集和测试集, 在训练集上通过时序交叉验证确定 AR 和 VAR 模型阶数, 在测试集上预测各时刻的风电功率概率分布; 然后根据市场价格 π_s 、 π_u 和 π_d 计算上调超出价格 $\bar{\lambda}_u$ 与下调降低价格

| | | | | | | | | | | | | | | | | | | | | | | | |
|---|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| A | 1.0 | 0.8 | 0.9 | 0.9 | 0.7 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.2 | 0.3 | 0.2 | 0.6 | 0.5 | 0.4 | 0.4 | 0.4 | 0.3 | 0.1 | 0.4 | 0.2 | 0.3 |
| B | 0.8 | 1.0 | 0.8 | 0.8 | 0.7 | 0.1 | 0.1 | 0.1 | 0.2 | 0.3 | 0.4 | 0.2 | 0.5 | 0.5 | 0.4 | 0.4 | 0.4 | 0.3 | 0.3 | 0.1 | 0.5 | 0.3 | 0.3 |
| C | 0.9 | 0.8 | 1.0 | 0.8 | 0.8 | 0.1 | 0.1 | 0.1 | 0.1 | 0.3 | 0.3 | 0.2 | 0.6 | 0.5 | 0.4 | 0.4 | 0.4 | 0.3 | 0.1 | 0.4 | 0.2 | 0.2 | 0.3 |
| D | 0.9 | 0.8 | 0.8 | 1.0 | 0.7 | 0.1 | 0.1 | 0.1 | 0.1 | 0.2 | 0.3 | 0.1 | 0.6 | 0.6 | 0.4 | 0.4 | 0.4 | 0.3 | 0.0 | 0.4 | 0.2 | 0.2 | 0.2 |
| E | 0.7 | 0.7 | 0.8 | 0.7 | 1.0 | 0.1 | 0.1 | 0.1 | 0.1 | 0.3 | 0.3 | 0.2 | 0.5 | 0.5 | 0.4 | 0.4 | 0.4 | 0.3 | 0.1 | 0.4 | 0.2 | 0.2 | 0.3 |
| F | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 1.0 | 0.6 | 0.8 | 0.6 | 0.2 | 0.2 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.2 | 0.1 | 0.1 | 0.2 | 0.2 | 0.2 |
| G | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.6 | 1.0 | 0.7 | 0.7 | 0.2 | 0.2 | 0.2 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.2 | 0.1 | 0.1 | 0.2 | 0.2 | 0.3 |
| H | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.8 | 0.7 | 1.0 | 0.6 | 0.1 | 0.2 | 0.1 | 0.1 | 0.1 | 0.1 | 0.0 | 0.1 | 0.2 | 0.1 | 0.1 | 0.2 | 0.2 | 0.2 |
| I | 0.1 | 0.2 | 0.1 | 0.1 | 0.1 | 0.6 | 0.7 | 0.6 | 1.0 | 0.2 | 0.2 | 0.2 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.2 | 0.1 | 0.1 | 0.2 | 0.1 | 0.2 |
| J | 0.2 | 0.3 | 0.3 | 0.2 | 0.3 | 0.2 | 0.2 | 0.2 | 0.1 | 1.0 | 0.9 | 0.8 | 0.2 | 0.2 | 0.3 | 0.2 | 0.3 | 0.2 | 0.4 | 0.6 | 0.5 | 0.5 | 0.5 |
| K | 0.3 | 0.4 | 0.3 | 0.3 | 0.3 | 0.2 | 0.2 | 0.2 | 0.2 | 0.9 | 1.0 | 0.7 | 0.2 | 0.2 | 0.3 | 0.3 | 0.4 | 0.3 | 0.2 | 0.4 | 0.6 | 0.5 | 0.5 |
| L | 0.2 | 0.2 | 0.2 | 0.1 | 0.2 | 0.1 | 0.2 | 0.1 | 0.2 | 0.8 | 0.7 | 1.0 | 0.2 | 0.1 | 0.2 | 0.2 | 0.3 | 0.2 | 0.2 | 0.3 | 0.5 | 0.4 | 0.4 |
| M | 0.6 | 0.5 | 0.6 | 0.6 | 0.5 | 0.1 | 0.1 | 0.1 | 0.1 | 0.2 | 0.2 | 0.2 | 1.0 | 0.7 | 0.3 | 0.4 | 0.4 | 0.3 | 0.0 | 0.4 | 0.2 | 0.2 | 0.2 |
| N | 0.5 | 0.5 | 0.5 | 0.6 | 0.5 | 0.1 | 0.1 | 0.1 | 0.1 | 0.2 | 0.2 | 0.1 | 0.7 | 1.0 | 0.3 | 0.4 | 0.3 | 0.3 | 0.0 | 0.4 | 0.1 | 0.2 | 0.2 |
| O | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.1 | 0.1 | 0.1 | 0.1 | 0.3 | 0.3 | 0.2 | 0.3 | 0.3 | 1.0 | 0.4 | 0.4 | 0.2 | 0.0 | 0.5 | 0.2 | 0.2 | 0.2 |
| P | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.1 | 0.1 | 0.1 | 0.1 | 0.2 | 0.3 | 0.2 | 0.4 | 0.4 | 0.4 | 1.0 | 0.3 | 0.2 | 0.0 | 0.4 | 0.2 | 0.2 | 0.2 |
| Q | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.1 | 0.1 | 0.1 | 0.1 | 0.3 | 0.4 | 0.3 | 0.4 | 0.3 | 0.4 | 0.3 | 1.0 | 0.3 | 0.1 | 0.6 | 0.3 | 0.3 | 0.3 |
| R | 0.3 | 0.3 | 0.3 | 0.3 | 0.3 | 0.2 | 0.2 | 0.2 | 0.2 | 0.3 | 0.3 | 0.2 | 0.3 | 0.3 | 0.2 | 0.3 | 0.3 | 1.0 | 0.3 | 0.3 | 0.4 | 0.5 | 0.5 |
| S | 0.1 | 0.1 | 0.1 | 0.0 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.2 | 0.2 | 0.2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.3 | 1.0 | 0.2 | 0.2 | 0.2 | 0.3 |
| T | 0.4 | 0.5 | 0.4 | 0.4 | 0.4 | 0.1 | 0.1 | 0.1 | 0.1 | 0.4 | 0.4 | 0.3 | 0.4 | 0.4 | 0.5 | 0.4 | 0.6 | 0.3 | 0.2 | 1.0 | 0.3 | 0.4 | 0.4 |
| U | 0.2 | 0.3 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.6 | 0.6 | 0.5 | 0.2 | 0.1 | 0.2 | 0.2 | 0.3 | 0.4 | 0.2 | 0.3 | 1.0 | 0.5 | 0.5 |
| V | 0.3 | 0.3 | 0.3 | 0.2 | 0.3 | 0.2 | 0.3 | 0.2 | 0.3 | 0.5 | 0.5 | 0.4 | 0.2 | 0.2 | 0.2 | 0.2 | 0.3 | 0.5 | 0.3 | 0.4 | 0.5 | 1.0 | 1.0 |

(a) 22 个风电场的相关系数热力图



(b) 具有强相关性的风电场聚类结果

图 2 风电场群初步筛选结果

Fig. 2 Preliminary selecting result of wind farm clusters

$\bar{\lambda}_d$ 的历史平均值, 并根据式(18)与式(13)分别计算目标风电场每小时的最佳投标发电量 q^* 和期望利润 $\mathbb{E}_Q[H(q, Q)]$; 据此构建合作博弈模型, 得到目标风电场和协助风电场的利润分配结果 c , 实现风电数据的合理定价。

4.2 风电数据定价结果

由于风力资源的时空相关性具有明显的季节性特点, 所选取的交易周期不应过长, 同时考虑数据交易的稳定性要求, 本文设置风电数据交易周期为一个月, 即交易一个月的数据并对此定价。在本节中, 选取风电场 A 作为目标风电场、风电场 B、C、D、E 为协助风电场提供相关数据, 并对 2013 年 6 月的数据定价进行算例分析。

4.2.1 风电功率预测精度的提升效果

在采用 AR 和 VAR 模型对风电场 A 的发电功率进行预测时, AR 模型仅利用了风电场 A 自身的数据, 而 VAR 模型利用了 5 个风电场的全部数据。通过时序交叉验证, 确定了两种模型的最佳阶数均为 2, 并设置了 700 h 的动态窗口宽度。在本节选取的数据交易周期内, AR 预测模型的均方根误差(root mean square error, RMSE)为 11.89, VAR 预测模型 RMSE 为 10.82, 预测精度提高了 9.00%。这一结果表明, VAR 模型通过引入邻近场站数据, 能够显著提升预测精度, 为风电场提供更准确的预测信息。

4.2.2 电力市场交易的超额利润结果

首先根据历史数据的平均值确定不平衡结算价格, 上调超出部分修正价格 $\lambda'_u = 8.610$ DKK/MW, 下调降低部分修正价格 $\lambda'_d = 8.424$ DKK/MW。进而计算出不平衡价格修正费用比 $\gamma' = 0.495$ 。继续根据式(18)计算风电最优投标电量 q^* 。最后根据式(13)计算出风电场 A 参与不同子联盟 S 的期望利润 $\mathbb{E}_Q[H(q, Q)]$ 。

图 3 展示了风电场 A 进行数据交易前后所产生

的超额利润, 即根据 VAR 预测确定最优投标发电量计算出的 $\mathbb{E}_{Q_M}[H(q_M^*, Q_M)]$ 与根据 AR 预测计算出的 $\mathbb{E}_{Q_M}[H(q_{(0)}^*, Q_M)]$ 之间的差值。图 3 结果同时说明了风电场 A 通过参与电力市场交易能够实现获得更高利润(即“超额利润”), 为下一环节的数据定价提供了保证, 使数据交易成为可能。

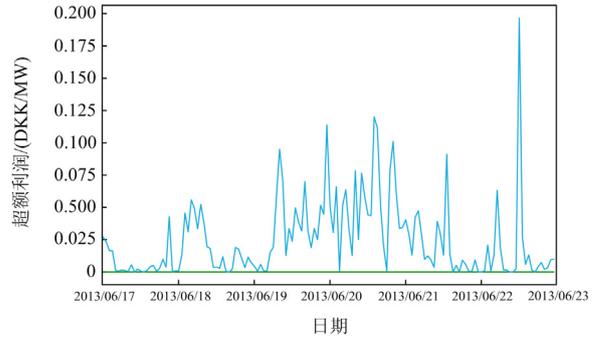
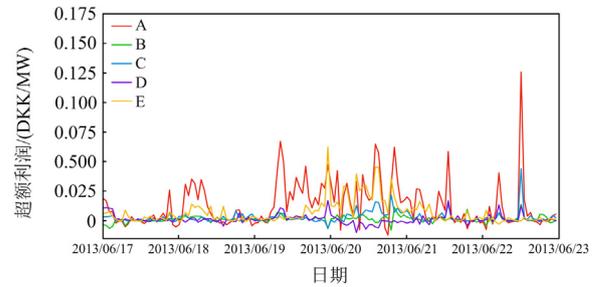


图 3 经数据交易后产生的超额利润

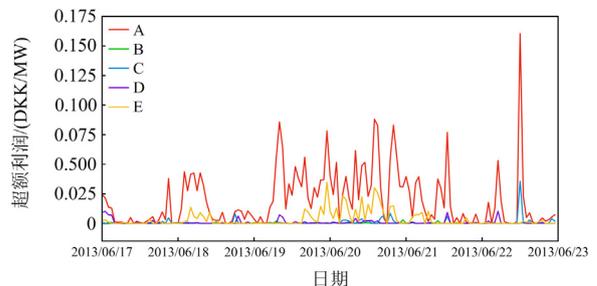
Fig. 3 Excess profit generated by data trading

4.2.3 基于利润分配的数据定价结果

根据 4.2.2 节计算出的每个子联盟 S 下目标风电场参与市场交易获得的期望利润 $\mathbb{E}_{Q_M}[H(q_S^*, Q_M)]$, 按照式(20)构建合作博弈联盟的特征函数, 并采用 Shapley 值法、核仁法分配超额利润实现数据定价, 风电场站 A、B、C、D、E 在 6 天内各时刻分配到的利润曲线如图 4 所示。



(a) Shapley 值法分配结果



(b) 核仁法分配结果

图 4 风电数据定价中各方利润分配情况

Fig. 4 Profit allocation results of each participant taking part in wind power data pricing

图 4(a)展示了 Shapley 值法的利润分配情况, 可以看出在部分时间节点数据交易会为目标风电场(A)和协助风电场(B、C、D、E)带来负利润。这是因为在这些时间节点, 联盟特征函数不满足广义单调增函数这一性质(即式(24)), 因此其分配结果不满足“成员价值原则”。但从长期分配结果来看, 交易双方最终都能获得正利润(累计值)。图 4(b)展示了核仁法的利润分配情况, 其分配结果始终为正, 可以满足“成员价值原则”。此外图 4(a)和图 4(b)都表明各方分配到的利润变化幅度较大, 在分配利润较小时, 可能导致激励不足, 降低各方数据交易意愿, 不利于数据交易的稳定进行。因此进行长周期的数据交易, 并对此进行数据定价是极其必要且有意义的。

将数据交易周期设定为 1 个月, 在表 2 中展示了采用 Shapley 值法、核仁法对超额利润的整体分配结果。从分配结果可以看出, Shapley 值法相较于核仁法认为目标风电场 A 所做出的贡献较小, 而认为 4 个协助风电场贡献更大。因此 Shapley 值法倾向于给目标风电场 A 分配相对较少的利润, 核仁法倾向于给目标风电场 A 分配相对更多的利润。

表 2 超额利润在各场站间的分配结果

| 风电场 | DKK/MW | | | | |
|------------|--------|-------|-------|-------|-------|
| | A | B | C | D | E |
| Shapley 值法 | 7.113 | 0.924 | 1.024 | 2.277 | 3.924 |
| 核仁法 | 11.658 | 0.107 | 0.151 | 1.550 | 1.796 |

在选择这两种分配方法时, 可以参照第 3 节中的取舍原则, 即选择更有利于协助风电场的分配方式, 以便在优先保证目标风电场利润后还能最大限度地提升针对协助风电场的经济激励。因此本文推荐优先采用 Shapley 值法进行超额利润的分配, 当其在长期交易的分配结果不满足“成员价值原则”时, 再使用核仁法进行分配。

4.3 考虑风电时空特性的数据定价结果

本节分别以 A、B、C、D、E 为目标风电场, 使用第 3 节建立的最大化剩余利润模型, 筛选出各月最佳交易对象, 如表 3 所示。其中“/”表示在该月内目标风电场不进行数据交易比进行数据交易能够获得更高的交易利润。此类情况在表 3 中只出现 6 次, 占比仅为 10%。因此在普遍情况下, 本文提出的数据交易定价方法能够提高风电场利润。

从表 3 的交易对象筛选结果可以看出, 数据交易对象随时间变化有明显的季节性特征, 据此可将其分为三类: 长期交易对象、季节性交易对象、短期交易对象。例如, E 场站倾向于长期购买 A 场站风电数据, B 场站倾向于在下半年购买 A 场站风电

表 3 各风场数据交易对象筛选结果

Table 3 Selecting results of trading objects of each wind farm data

| 月份 | A | B | C | D | E |
|----|---------|-------|---------|---------|---------|
| 1 | B,C,E | C,D | B | B,C,E | A,B |
| 2 | B,D | D | D | / | D |
| 3 | B,C,E | C,E | E | A,B,E | A,B,C,D |
| 4 | C,D | / | / | / | C,D |
| 5 | B,D | D | A,B,D | A,B | A,B,C,D |
| 6 | B,C,D,E | C,E | / | B,C,E | A,B,D |
| 7 | B,C,D,E | D | A,B,D | A,B,C,E | A,B,C,D |
| 8 | B,C | A,C,E | B | A,B,C | A |
| 9 | C,D,E | D,E | D,E | A,C,E | D |
| 10 | B,C,D,E | A,D | A,B,D,E | A,B,E | A,B,C,D |
| 11 | B,D | A,D | A,B,D,E | A,B,E | A,B,C,D |
| 12 | B,C | / | B | A,B | B,C |

数据, C 场站仅在部分月内购买 E 场站数据。这也反映出风力资源时空相关性的季节性特征, 因此在不同时期改变数据交易对象, 可以更好地提高预测精度, 进而获得更高利润。

进一步按照各个风电场筛选出的数据交易对象(如表 3 所示)计算出对应的市场利润变化情况, 结果如图 5 所示, 图中“基础利润”表示未进行数据交易的市场利润, “提高利润”表示进行数据交易后提高的市场利润。

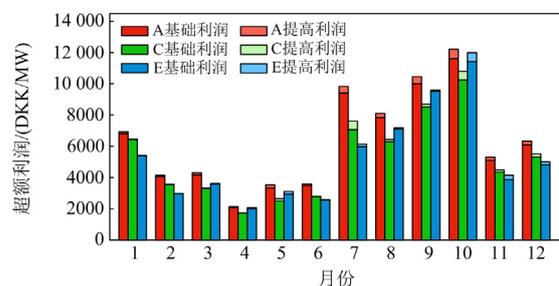


图 5 部分风电场参与数据交易的市场利润变化

Fig. 5 Market profit changes of partial wind farm involving data transaction

从图 5 可以看出, 在 7 月至 10 月, 风电数据交易后能够为绝大多数风电场提升更多的市场利润。进一步分析上述 4 个月的计算结果, 在这一时期内, 该风电场群附近的风力资源较为充沛。因此在这一阶段开展数据交易就显得更有意义, 目标风电场也更愿意为此购买更多风电数据以及支付更高数据成本, 从而获得更精确的预测结果, 最大化参与电力市场的交易利润。结合表 3 的计算结果, 4 月是该风电场群风力资源最差的月份, 因此风电场 B、C、D 在 4 月选择不进行数据交易。

为了说明在数据定价时考虑风电时空特性的必

要性,图6对比了风电场A直接与B、C、D、E进行数据交易与考虑风电时空特性进行合作筛选的期望利润提升百分数。在6月、7月和10月中,是否进行合作对象筛选不影响风电场A的获利情况,这与表3的结果相符。但在剩余的9个月中,筛选合作对象平均能使风电场A的市场利润提升1.35%,甚至在12月提升了2.93%。这说明在进行风电数据定价时,需要考虑风电时空特性进行交易对象筛选,才能实现数据的合理定价。同时也说明更多的风电数据可能会导致信息冗余,预测精度提升所带来的效益反而少于所支出的数据购买成本。因此由于这种边际效应,使用这部分场站数据反而会风电场带来负效益。

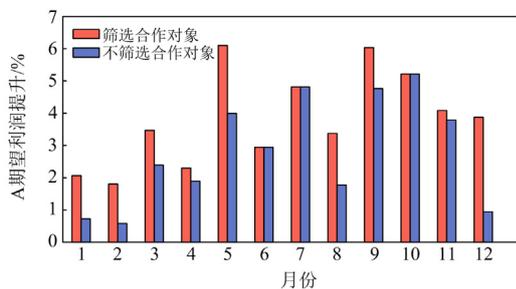


图6 风电场A筛选数据交易对象的市场利润变化

Fig. 6 Market profit change of trading objects of selecting data of wind farm A

5 结语

在电力大数据背景下,本文聚焦于新能源数据交易问题,针对风电数据定价问题展开详细研究,提出了一套完整的风电数据交易流程以及针对风电功率预测任务的数据定价方法。算例分析结果验证了本文所提方法的有效性,所得主要结论如下:

1) 与仅利用自身场站数据相比,通过与邻近场站进行风电数据交易,可以有效利用风电资源的时空相关性,提升风电预测精度,改进风电场参与电力市场交易的投标策略,最终实现市场交易利润的提升。

2) 利用联盟合作博弈理论,可以合理分配由风电数据交易带来的超额利润,向数据售卖方支付数据使用费。长期来看,优先采用 Shapley 值法,并将核仁法作为参考,有助于实现数据定价的公平性和合理性。

3) 在进行风电数据定价时,应考虑风力资源的时空特性。由于不同季节的时空特性各异,设置适当的数据交易周期并在不同时期调整交易对象,是充分挖掘数据价值、避免购买冗余信息、实现高效数据定价的关键。

本文初步探索了数据交易应用于风电预测任务的可能性,并提出了一种基于合作博弈模型的合理公平的数据定价方法。这使得具有竞争关系的不同风电企业之间合作共赢成为可能,进而提供更加准确和可靠的风电功率预测结果。

参考文献

- [1] 李婷, 胥威汀, 刘向龙, 等. 含高比例可再生能源的交直流混联电网规划技术研究综述[J]. 电力系统保护与控制, 2019, 47(12): 177-187.
LI Ting, XU Weiting, LIU Xianglong, et al. Review on planning technology of AC/DC hybrid system with high proportion of renewable energy[J]. Power System Protection and Control, 2019, 47(12): 177-187.
- [2] GONCALVES C, BESSA R J, PINSON P. A critical overview of privacy-preserving approaches for collaborative forecasting[J]. International Journal of Forecasting, 2021, 37(1): 322-342.
- [3] 赵永宁, 李卓, 叶林, 等. 基于时空相关性的风电功率超短期自适应预测方法[J]. 电力系统保护与控制, 2023, 51(6): 94-105.
ZHAO Yongning, LI Zhuo, YE Lin, et al. A very short-term adaptive wind power forecasting method based on spatio-temporal correlation[J]. Power System Protection and Control, 2023, 51(6): 94-105.
- [4] 赵寒亭, 张耀, 霍巍, 等. 基于纵向联邦学习的短期风电功率协同预测方法[J]. 电力系统自动化, 2023, 47(16): 44-53.
ZHAO Hanting, ZHANG Yao, HUO Wei, et al. Collaborative forecasting method for short-term wind power based on vertical federated learning[J]. Automation of Electric Power Systems, 2023, 47(16): 44-53.
- [5] 熊巧琴, 汤珂. 数据要素的界权、交易和定价研究进展[J]. 经济动态, 2021(2): 143-158.
XIONG Qiaoqin, TANG Ke. Research progress on the right delimitation, exchange and pricing of data[J]. Economic Perspectives, 2021(2): 143-158.
- [6] 蔡莉, 黄振弘, 梁宇, 等. 数据定价研究综述[J]. 计算机科学与探索, 2021, 15(9): 1595-1606.
CAI Li, HUANG Zhenhong, LIANG Yu, et al. Survey of data pricing[J]. Journal of Frontiers of Computer Science and Technology, 2021, 15(9): 1595-1606.
- [7] 江东, 袁野, 张小伟, 等. 数据定价与交易研究综述[J]. 软件学报, 2023, 34(3): 1396-1424.
JIANG Dong, YUAN Ye, ZHANG Xiaowei, et al. Survey on data pricing and trading research[J]. Journal of Software, 2023, 34(3): 1396-1424.
- [8] 郑臻哲. 移动互联网中博弈模型与机制设计研究[D]. 上海: 上海交通大学, 2018.
ZHENG Zhenzhe. Game models and mechanism design for mobile internet[D]. Shanghai: Shanghai Jiao Tong University, 2018.
- [9] MCAFEE R P. A dominant strategy double auction[J]. Journal of Economic Theory, 1992, 56(2): 434-450.
- [10] GHOSH A, ROTH A. Selling privacy at auction[J]. Games and Economic Behavior, 2015, 91: 334-346.

- [11] ZHANG Mengxiao, BELTRAN F, LIU Jiamou. Selling data at an auction under privacy constraints[C] // 36th Conference on Uncertainty in Artificial Intelligence, August 3-6, 2020, Toronto, Canada: 669-678.
- [12] SHANNON C E. A mathematical theory of communication[J]. Bell System Technical Journal, 1948, 27(3): 379-423.
- [13] YANG Jian, ZHAO Chongchong, XING Chunxiao. Big data market optimization pricing model based on data quality[J]. Complexity, 2019: 5964068.
- [14] STAHL F, VOSSEN G. Fair knapsack pricing for data marketplaces[C] // Advances in Databases and Information Systems, August 28-31, 2016, Prague, Czech Republic: 46-59.
- [15] YU Haifei, ZHANG Mengxiao. Data pricing strategy based on data quality[J]. Computers & Industrial Engineering, 2017, 112: 1-10.
- [16] KOUTRIS P, UPADHYAYA P, BALAZINSKA M, et al. Query-based data pricing[J]. Journal of the ACM, 2015, 62(5): 1-44.
- [17] CHEN Lei, OZCAN F, QUAMAR A, et al. Efficient task-specific data valuation for nearest neighbor algorithms[J]. Proceedings of the VLDB Endowment, 2019, 12(11): 1610-1623.
- [18] CHEN Lingjiao, KOUTRIS P, KUMAR A. Towards model-based pricing for machine learning in a data marketplace[C] // Proceedings of the 2019 International Conference on Management of Data, June 30-July 5, 2019, New York, USA: 1535-1552.
- [19] CHEN Lingjiao, KOUTRIS P, KUMAR A. Model-based pricing: do not pay for more than what you learn![C] // Proceedings of the 1st Workshop on Data Management for End-to-End Machine Learning, May 14-19, 2017, New York, USA: 1-4.
- [20] AGARWAL A, DAHLEH M, SARKAR T. A marketplace for data: an algorithmic solution[C] // Proceedings of the 2019 ACM Conference on Economics and Computation, June 24-28, 2019, Phoenix, AZ, USA: 701-726.
- [21] 程杉, 傅桐, 李洋洋, 等. 含高渗透可再生能源的配电网灵活性供需协同规划[J]. 电力系统保护与控制, 2023, 51(22): 1-12.
CHENG Shan, FU Tong, LI Fengyang, et al. Flexible supply demand collaborative planning for distribution networks with high penetration of renewable energy[J]. Power System Protection and Control, 2023, 51(22): 1-12.
- [22] HAN L, KAZEMPOUR J, PINSON P. Monetizing customer load data for an energy retailer: a cooperative game approach[C] // 2021 IEEE Madrid PowerTech, June 28-July 2, 2021, Madrid, Spain: 1-6.
- [23] DOWELL J, PINSON P. Very-short-term probabilistic wind power forecasts by sparse vector autoregression[J]. IEEE Transactions on Smart Grid, 2016, 7(2): 763-770.
- [24] JIAN P. A survey on data pricing: from economics to data science[J]. IEEE Transactions on Knowledge and Data Engineering, 2020, 34(10): 4586-4608.
- [25] 孙可, 兰洲, 林振智, 等. 国际典型电力市场阻塞管理机制及其对中国的启示研究[J]. 电力系统保护与控制, 2020, 48(12): 170-178.
SUN Ke, LAN Zhou, LIN Zhenzhi, et al. Transmission congestion management mechanism of typical international power markets and possible guidance for China's power market[J]. Power System Protection and Control, 2020, 48(12): 170-178.
- [26] TSDTU J, PINSON P, TROMBE P J, et al. Probabilistic forecasts of wind power generation accounting for geographically dispersed information[J]. IEEE Transactions on Smart Grid, 2013, 5(1): 480-489.
- [27] SOMMER B, PINSON P, MESSNER J W, et al. Online distributed learning in wind power forecasting[J]. International Journal of Forecasting, 2021, 37(1): 205-223.
- [28] 徐扬, 张耀, 陈宇轩, 等. 基于 Bagging 混合策略的多风电场稀疏向量自回归概率预测[J]. 电力系统保护与控制, 2023, 51(7): 95-106.
XU Yang, ZHANG Yao, CHEN Yuxuan, et al. Bagging ensemble method of probabilistic forecasting for multiple wind farms by sparse vector autoregression[J]. Power System Protection and Control, 2023, 51(7): 95-106.
- [29] 程建东, 赵浩然, 韩明哲. 市场机制下推动风电参与电力市场的实践总结与启示[J]. 电网技术, 2022, 46(7): 2619-2631.
CHENG Jiandong, ZHAO Haoran, HAN Mingzhe. Practice summary and enlightenment of promoting wind power to participate in power market under market mechanism[J]. Power System Technology, 2022, 46(7): 2619-2631.
- [30] MONTERO-MANSO P, ATHANASOPOULOS G, HYNDMAN R J, et al. FFORMA: feature-based forecast model averaging[J]. International Journal of Forecasting, 2020, 36(1): 86-92.
- [31] 杨晓辉, 袁志鑫, 肖锦扬, 等. 考虑电池寿命的混合储能微电网优化配置[J]. 电力系统保护与控制, 2023, 51(4): 22-31.
YANG Xiaohui, YUAN Zhixin, XIAO Jinyang, et al. Optimal configuration of hybrid energy storage microgrid considering battery life[J]. Power System Protection and Control, 2023, 51(4): 22-31.
- [32] KARACA O, DELIKARAOGLU S, HUG G, et al. Enabling inter-area reserve exchange through stable benefit allocation mechanisms[J]. Omega-International Journal of Management Science, 2022, 113: 102711.
- [33] BENEDEK M, FLIEGE J, NGUYEN T D. Finding and verifying the nucleolus of cooperative games[J]. Mathematical Programming, 2021, 190(1-2): 135-170.

收稿日期: 2023-12-14; 修回日期: 2024-04-03

作者简介:

霍巍(2000—), 男, 博士研究生, 研究方向为电力大数据与人工智能; E-mail: liuxi26@stu.xjtu.edu.cn

张耀(1988—), 男, 通信作者, 博士, 副教授, 博士生导师, 研究方向为电力系统运行与规划、可再生能源预测等。E-mail: yaozhang_ee@xjtu.edu.cn

(编辑 周金梅)