

doi: 10.3969/j.issn.1671-7775.2016.02.014

基于 Mycielski 改进的 Markov 链超短期 风电功率预测方法

尹天骄, 沈艳霞, 陆 欣

(江南大学 电气自动化研究所, 江苏 无锡 214122)

摘要: 针对在 Markov 链超短期风电功率预测过程中未考虑风电功率变化趋势, 在不同风电功率变化区间均采用同一状态转移概率矩阵, 导致预测精度欠佳的问题, 提出了一种基于 Mycielski 方法改进的 Markov 链预测方法. 首先利用 Mycielski 方法在风电功率历史序列中寻找最长长度的重复序列, 然后将计算每一时刻 Markov 链状态转移概率矩阵的序列重新定义, 最后利用每一时刻的 Markov 链状态转移概率矩阵进行风电功率预测. 对我国某风电场超短期风电功率进行总体预测. 结果表明, 在均方根误差的对比上, 基于 Mycielski 方法改进的 Markov 链预测方法能够提高 14.15% 的预测精度, 具有一定的实用价值.

关键词: 风电功率; Markov; Mycielski; 预测; 转移概率矩阵

中图分类号: TM743 **文献标志码:** A **文章编号:** 1671-7775(2016)02-0209-05

引文格式: 尹天骄, 沈艳霞, 陆 欣. 基于 Mycielski 改进的 Markov 链超短期风电功率预测方法[J]. 江苏大学学报(自然科学版), 2016, 37(2): 209-213.

Improved Markov chain for ultra-short-term wind power prediction method based on Mycielski algorithm

YIN Tianjiao, SHEN Yanxia, LU Xin

(Institute of Electrical Automation, Jiangnan University, Wuxi, Jiangsu 214122, China)

Abstract: During the ultra-short-term wind power prediction using Markov chains, to solve the poor prediction accuracy due to the using same-state transition probability matrix in different variation scopes of wind power and without considering wind power real-time trends, an improved prediction method using Markov chains was proposed based on Mycielski algorithm. The Mycielski approach was employed to find the longest repetitive sequence from the historical sequence of wind power. The sequence in state transition probability matrix of Markov chain at every moment was redefined. The ultra-short-term wind power prediction was realized by the state transition probability matrix of the Markov chain at every moment. The ultra-short-term wind power prediction was also implemented in a domestic wind farm. The results show that for root-mean-square error, the modified prediction using Markov chains based on Mycielski algorithm can effectively increase prediction accuracy by 14.15% with high practical value.

Key words: wind power; Markov; Mycielski; prediction; transition probability matrix

收稿日期: 2015-09-07

基金项目: 国家自然科学基金资助项目(61104183); 高等学校博士学科点专项科研基金资助项目(20130093110011); 江苏省自然科学基金资助项目(BK20141114)

作者简介: 尹天骄(1993—), 女, 内蒙古赤峰人, 硕士研究生(yintianjiao2014@163.com), 主要从事风速风电功率预测研究.

沈艳霞(1973—), 女, 山东淄博人, 教授, 博士生导师(shenyx@jiangnan.edu.cn), 主要从事风电系统优化研究.

风能是一种清洁、无污染的可再生能源,但风具有随机性及不可控性,不稳定的风能在并网后会对电网产生巨大冲击,影响电网系统的安全稳定运行.风电功率预测是帮助实现风电并网稳定运行的重要手段^[1].

风电功率预测方法一般可分为2类:①利用数值天气预报建立预测模型,将天气数据转换成风电功率输出.该方法不受历史数据的限制,可用于新建风电场的功率预测.丹麦、德国、西班牙等国均有基于数值天气预报的风电功率预测系统投入运行,如Pridiktor,Previento等^[2].②基于历史数据进行预测,分为学习方法(如人工神经网络^[3]、小波分析法等^[4])和统计方法.其中统计方法如卡尔曼滤波法^[5]、Markov链方法^[6-8].根据风电场风速的历史数据与实时数据间的关系建立预测模型,进行风速预测,具有预测速度快、模型简单的特点,适用于短期、超短期的风电功率预测.

Markov链预测法因其算法简单、可操作性强在电力系统中得到应用.文献[6]中建立了基于Markov链的光伏并网发电量预测的数学模型,并利用该模型完成日发电量、月发电量预测.文献[7-8]提出了Markov链在风电功率预测和风向建模中的应用,均得到了较好的预测效果.但是,在单一的Markov链预测方法中,由于未考虑不同区间风电功率的变化趋势,在风电功率预测过程中全部采用相同的状态转移概率矩阵,导致预测精度欠佳.

文中基于文献[6-8],拟提出基于Mycielski改进的Markov链超短期风电率预测方法,并利用实测数据验证该改进预测方法能否有效提高单一的Markov链风电功率预测精度.

1 Markov链预测模型

Markov链为状态空间中经过从一个状态到另一个状态的转换的随机过程.该过程具有无后效性,即下一时刻的状态只由当前状态决定,而与它之前的状态无关.用公式表示如下:

$$X(x_{n+1}=y_{n+1}|x_0=y_0, x_1=y_1, \dots, x_n=y_n) = X(x_{n+1}=y_{n+1}|x_n=y_n), \quad (1)$$

式中: $X(\cdot)$ 为概率函数; x_1, x_2, \dots, x_n 为一个时间序列第 n 个随机变量; y_1, y_2, \dots, y_n 为第 n 个随机变量所对应的状态值.第 x_{n+1} 的状态值 y_{n+1} 只与 x_n 的状态值 y_n 有关,而与其他的状态值无关.

Markov链的所有状态转移概率可以组成一个

矩阵 \mathbf{P} ,称为状态转移概率矩阵,即为

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{bmatrix}, \quad (2)$$

式中: m 为状态数; \mathbf{P} 内部元素 p_{ij} 表示该随机过程由状态 i 转移至状态 j 的状态转移概率,即为

$$p_{ij} = \frac{N_{ij}}{\sum_{j=1}^m N_{ij}}, \quad (3)$$

式中 N_{ij} 为状态 i 转移至状态 j 的转移频数, $i, j \in [0, m]$.通过状态转移概率 p_{ij} 计算累计转移概率,即为

$$p_{ij'} = \sum_{j=1}^{j'} p_{ij}, \quad (4)$$

式中 $p_{ij'}$ 为在 i 行中 j' 状态的累计转移概率.所有累计转移概率可以组成累计转移概率矩阵 \mathbf{P}_C ,即为

$$\mathbf{P}_C = \begin{bmatrix} p_{11} & \cdots & \sum_{j=1}^{j'} p_{1j} & \cdots & \sum_{j=1}^m p_{1j} \\ p_{21} & \cdots & \sum_{j=1}^{j'} p_{2j} & \cdots & \sum_{j=1}^m p_{2j} \\ \vdots & & \vdots & & \vdots \\ p_{m1} & \cdots & \sum_{j=1}^{j'} p_{mj} & \cdots & \sum_{j=1}^m p_{mj} \end{bmatrix}. \quad (5)$$

得到累计转移概率矩阵 \mathbf{P}_C ,即可由初始状态以及计算机随机产生的介于0到1之间的随机数,通过与 \mathbf{P}_C 矩阵中的累积概率对比得到该时刻的风电功率预测值.

2 基于Mycielski改进Markov链方法

2.1 Mycielski方法

Mycielski方法由A. EHRENFUCHT和J. MYCIELSKI^[9]提出,是一种基于模式匹配的预测算法.该方法在信号预测编码以及蛋白质的DNA序列的预测等领域都有应用.近年来C. CROONEN-BROECK等^[10]又通过Mycielski方法进行风电功率预测.该方法主要思想是将历史数据作为一个样本序列,通过对这个序列从后向前搜索,找出在该样本序列中存在的与该序列尾部子序列相同的最长样本子序列,然后将该样本子序列的下一个或几个样本值作为预测值.其数学表达式为

$$\hat{a}(t+1) = f_{t+1}(a(1), a(2), \dots, a(t)), \quad (6)$$

式中: f_{t+1} 为匹配函数; $\hat{a}(t+1)$ 为预测值; $a(1)$,

$a(2), \dots, a(t)$ 为样本序列. 即取全部的历史数据为样本序列 $\{a(1), a(2), \dots, a(t)\}$, 从该序列的尾部 $\{a(t)\}$ 开始在样本序列 S 中进行匹配, 当序列尾部元素匹配成功后便增加尾部子序列的长度, 即再对 $\{a(t), a(t-1)\}$ 进行匹配, 直至找出在样本序列中与其相同的最长样本子序列, 该样本子序列的下一个元素就是下一时刻的预测值, 该方法在每一次预测中都要搜索一遍整个历史数据序列. 其具体匹配过程如式(7) - (8)所示:

$$\begin{aligned} \lambda &= \operatorname{argmax}_k \{a(k) = a(t), a(k-1) = \\ &a(t-1), \dots, a(k-L) = a(t-L)\}, \quad (7) \\ \hat{a}(t+1) &= f_{t+1}(a(1), a(2), \dots, a(t)) = a(\lambda+1). \quad (8) \end{aligned}$$

可以看出, λ 即为当 L 取得最大值时的 k 值, 最后的预测结果 $\hat{a}(t+1)$ 就等于 $a(\lambda+1)$.

2.2 基于 Mycielski 改进的 Markov 链风电功率预测方法

对风电功率进行预测时, 单一的 Markov 链风电功率预测方法在预测过程中使用同一个状态转移概率矩阵, 即由全部的历史数据求得的状态转移概率矩阵, 未考虑不同区间风电功率变化趋势, 预测精度有限. 采用 Mycielski 方法改进 Markov 链预测方法, 考虑了风电功率的变化特点, 对每一预测时刻的 Markov 链状态转移概率矩阵通过 Mycielski 方法重新匹配样本集.

改进方法预测步骤如下:

1) 对历史风电功率数据归一化处理, 得到序列 S , 把 S 中数据平均划分为 ξ 个状态, 得到序列 A .

2) 将序列 A 中每 δ 个数据取几何平均值后得到序列 A_{geo} . 用 Mycielski 方法由 A_{geo} 的序列尾部向首端匹配相同的数据段. 设预测 t 时刻的风电功率值, 则所搜索的数据区间为 $1, 2, \dots, (t-1)$ 时刻的数据; 若预测第 $(t+1)$ 时刻的风电功率值, 则搜索的数据区间为 $1, 2, \dots, t$ 时刻的数据. 在序列 A_{geo} 中匹配出最长子序列后, 记录该匹配序列以及其下一个数据在序列 S 中的对应数据.

3) 设由步骤2)得到 S 序列中的子序列数据为 A_i . 取 A_i 内数据化分状态值后作为 Markov 链状态转移概率矩阵样本集.

4) 通过选取的样本集求取对应 t 时刻的 P 和 P_c .

5) 选择初始状态 η , 生成一个 0 到 1 之间的随机数 μ , 将 μ 与累计转移概率矩阵 P_c 中第 η 行的元素进行比较, 若 μ 在 2 个状态值之间, 则取后一状态值为预测状态, 并在该状态中随机取值 ε 作为最终

预测结果.

6) 通过 Markov 链模型对 t 时刻的风电功率进行预测, 每预测一个时刻的风电功率值即重复一次以上步骤, 直至预测结束.

3 算例分析

算例数据集为江苏某风电场风电功率数据, 共 4 800 组, 前 3 840 组数据为样本集, 后 960 组数据为测试集. 分辨率为 15 min.

3.1 样本序列马氏性检验

检验一个随机过程是否具有马尔科夫性是应用马尔科夫链预测模型的必要前提. 通常采用 χ^2 统计量检验离散序列的马尔科夫链^[11].

$$\chi^2 = 2 \sum_{i=1}^n \sum_{j=1}^n N_{ij} \left| \ln \left(\frac{p_{ij}}{p_j} \right) \right|, \quad (9)$$

式中 p_j 为边际概率, 即为

$$p_j = \frac{\sum_{i=1}^n N_{ij}}{\sum_{i=1}^n \sum_{j=1}^n N_{ij}}. \quad (10)$$

统计量 χ^2 服从自由度为 $(n-1)^2$ 的 χ^2 分布. 给定显著性水平 α , 查表可得 $\chi_{\alpha}^2((n-1)^2)$ 的值, 若 $\chi^2 > \chi_{\alpha}^2((n-1)^2)$, 则认为该序列符合马尔科夫性, 可应用马尔科夫链预测模型.

对文中所使用的历史数据进行马尔科夫性检验, 当 $\alpha = 0.05$ ($\mu = 0.95$)、自由度 $(m-1) = 19$ 时, $\chi_{0.05}^2((20-1)^2) = 10.85$, $\chi^2 = 226.756$, 故有 $\chi^2 > \chi_{0.05}^2((20-1)^2)$, 证明该风电功率序列具有马尔科夫性.

3.2 算法参数设置及预测结果分析

先利用 Mycielski 方法匹配历史风电功率数据中符合规则的子序列, 并对匹配出的子序列进行状态划分, 再通过建立 Markov 链预测模型, 计算状态转移概率, 得到预测结果.

3.2.1 Mycielski 匹配子序列参数设置

在用 Mycielski 匹配子序列过程中, 参考文献 [12-13] 后, 通过多次试验, 确定以下参数设置.

把经过归一化的历史数据 S 等分为 5 个状态值, 即 $\xi = 5$, 得到序列 A . 其中归一化后风电功率在 $[0, 0.2)$, $[0.2, 0.4)$, $[0.4, 0.6)$, $[0.6, 0.8)$, $[0.8, 1.0]$ 区间内的数据分别对应的状态值为 0, 1, 2, 3, 4. 为避免高频分量对原始数据产生不良影响, 将序列 A 中每 24 个数据求得一个几何平均数, 按照规则匹配出序列 A_{geo} .

3.2.2 Markov 链预测参数设置

根据序列 A_{geo} , 提取历史数据 S 中的对应数据段 A_i . 利用数据段 A_i 进行 Markov 链预测, 参考文献 [6-8] 后, 通过多次试验, 决定将数据平均分为 20 个状态值, 即 $v=20$. A_i 数据区间与状态值对应关系如下: $[0, 0.05)$, 0; $[0.05, 0.10)$, 1; $[0.10, 0.15)$, 2; \cdots ; $[0.90, 0.95)$, 18; $[0.95, 1.00]$, 19. 完成数据的预处理后, 即可进行风电功率预测.

3.3 结果分析

依照文中所述预测步骤, 使用 Markov 链预测方法和基于 Mycielski 改进的 Markov 链方法分别进行

预测. 根据超短期风电功率预测时间分辨率的要求, 文中采用 15 min 的预测分辨率. 经过多次预测后, 2 种方法的预测曲线以及绝对误差曲线如图 1 所示, 2 种预测模型的绝对误差统计结果如表 1 所示, 预测结果的分段均方根误差 ($RMSE$) 如表 2 所示. 其中均方根误差 $RMSE$ 公式为

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (X_{obs,i} - X_{model,i})^2}{N}}, \quad (11)$$

式中: $X_{obs,i}$, $X_{model,i}$ 分别为实际测量值和预测值; N 为预测个数.

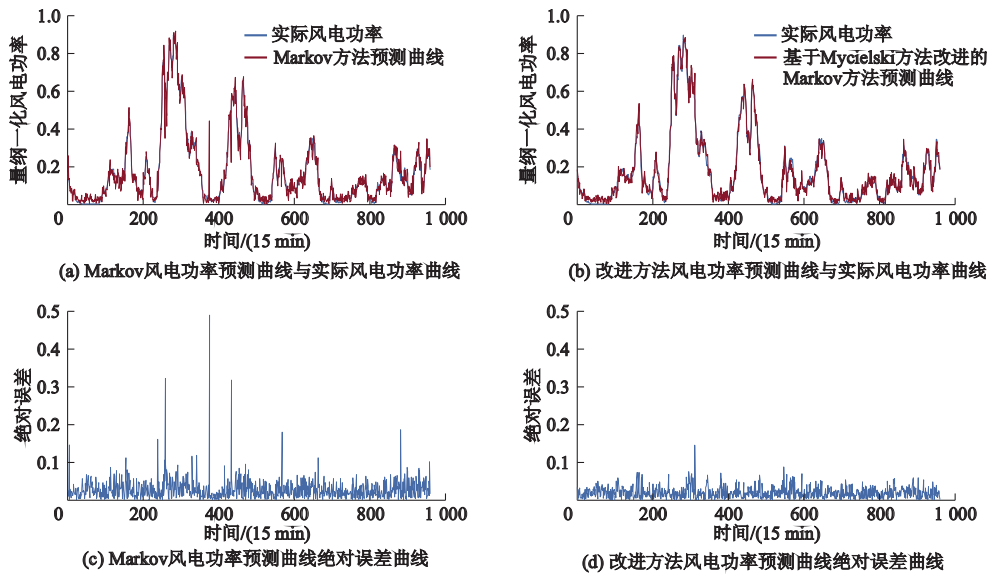


图 1 2 种预测方法的预测结果及绝对误差

表 1 2 种方法绝对误差统计

参数	最大值	均值
Markov 链预测绝对误差	0.407 6	0.022 5
改进方法预测绝对误差	0.145 0	0.019 3

表 2 2 种预测方法 $RMSE$ 分段预测结果

方法	$RMSE$					
	3 841 - 4 041 *	4 041 - 4 241	4 241 - 4 441	4 441 - 4 641	4 641 - 4 800	3 841 - 4 800
Markov 链方法	0.545 458	0.659 250	0.640 756	0.561 945	0.516 580	0.608 434
改进方法	0.508 105	0.597 876	0.527 920	0.484 235	0.476 519	0.522 285

注: * 行为时间分段, 单位为 15 min.

由图 1a,b 曲线可见, 2 种预测模型都能够近似预测风电功率, 但单一的 Markov 链方法在一些点处的预测结果会出现较大的误差; 由图 1c,d 的对比可见, 改进方法的预测精度整体误差曲线较为平稳, 结合表 1 中对 2 种方法绝对误差的统计结果, 可以看

出改进方法能够减少误差较大预测点. 同时比较表 2 中 2 种预测方法的 $RMSE$, 在不同的数据段, 改进方法的误差均小于单一的 Markov 链方法, 其中 4 241 - 4 441 (单位为 15 min) 段数据, 改进方法的 $RMSE$ 结果提高了 17.61% 的预测精度, 而在总体预

测精度上也提高了 14.15%。由试验结果看出, 基于 Mycielski 改进的 Markov 链方法在分段及总体的预测精度上都优于单一的 Markov 链预测方法, 体现了改进方法的优势。

4 结 论

1) 文中针对 Markov 链风电功率预测方法的不足, 提出了基于 Mycielski 改进的 Markov 链超短期风电功率预测方法, 并阐释了该方法的预测步骤。

2) 利用 Mycielski 方法为每一预测时刻 Markov 链状态转移概率矩阵的样本序列进行重新定义, 改进了 Markov 链预测方法。

3) 通过试验和结果分析验证了改进方法的预测结果较单一的 Markov 链预测方法在 RMSE 及绝对误差上有更好的预测精度。

参考文献 (References)

- [1] LIU Y L, ZHANG Y B, CHEN Z Y. Wind power prediction investigation [J]. Research Journal of Applied Sciences, Engineering and Technology, 2013, 5 (5): 1762 – 1768.
- [2] 冯双磊, 王伟胜, 刘纯, 等. 风电场功率预测物理方法研究[J]. 中国电机工程学报, 2010, 30 (2): 1 – 6. FENG S L, WANG W S, LIU C, et al. Study on the physical approach to wind power prediction [J]. Proceedings of the CSEE, 2010, 30 (2): 1 – 6. (in Chinese)
- [3] HERVÁS-MARTÍNEZ C, SALCEDO-SANZ S, GUTIÉRREZ P A, et al. Evolutionary product unit neural networks for short-term wind speed forecasting in wind farms [J]. Neural Computing and Applications, 2012, 21 (5): 993 – 1005.
- [4] 刘辉, 田红旗, 李燕飞. 基于小波分析法与滚动式时间序列法的风电场风速短期预测优化算法[J]. 中南大学学报(自然科学版), 2010, 41 (1): 370 – 375. LIU H, TIAN H Q, LI Y F. Short-term forecasting optimization algorithm for wind speed from wind farms based on wavelet analysis method and rolling time series method [J]. Journal of Central South University (Science and Technology), 2010, 41 (1): 370 – 375. (in Chinese)
- [5] CASSOLA F, BURLANDO M. Wind speed and wind energy forecast through Kalman filtering of numerical weather prediction model output [J]. Applied Energy, 2012, 99 (6): 154 – 166.
- [6] 李英姿, 贺琳, 牛进苍. 基于马尔可夫链的光伏并网发电量预测[J]. 太阳能学报, 2014, 35 (4): 611 – 616. LI Y Z, HE L, NIU J C. Forecasting power generation of grid-connected solar PV system based on Markov chain [J]. Acta Energiæ Solaris Sinica, 2014, 35 (4): 611 – 616. (in Chinese)
- [7] CARPINONE A, GIORGIO M, LANGELLA R, et al. Markov chain modeling for very-short-term wind power forecasting [J]. Electric Power Systems Research, 2015, 122: 152 – 158.
- [8] 季亮, 谭洪卫, 王亮. 基于马尔可夫链的自然风风向建模[J]. 中南大学学报(自然科学版), 2012, 43 (8): 3274 – 3279. JI L, TAN H W, WANG L. Wind direction modeling using Markov chain [J]. Journal of Central South University (Science and Technology), 2012, 43 (8): 3274 – 3279. (in Chinese)
- [9] EHRENFUCHT A, MYCIELSKI J. A pseudorandom sequence-how random is it? [J]. American Mathematical Monthly, 1992, 4 (99): 374 – 375.
- [10] CROONENBROECK C, AMBACH D. A selection of time series models for short- to medium-term wind power forecasting [J]. Journal of Wind Engineering and Industrial Aerodynamics, 2015, 136: 201 – 210.
- [11] 周松林, 茆美琴, 苏建徽. 考虑风力发电随机性的微电网潮流预测[J]. 中国电机工程学报, 2013, 33 (22): 26 – 34. ZHOU S L, MAO M Q, SU J H. Power flow forecasting of micro-grid considering randomness of wind power [J]. Proceedings of the CSEE, 2013, 33 (22): 26 – 34. (in Chinese)
- [12] 邹文, 丁巧林, 杨宏, 等. 基于 Mycielski 算法的风电场风速预测[J]. 电力科学与工程, 2011, 27 (3): 1 – 4. ZOU W, DING Q L, YANG H, et al. Wind speed prediction based on Mycielski algorithm [J]. Electric Power Science and Engineering, 2011, 27 (3): 1 – 4. (in Chinese)
- [13] 甘敏, 丁明, 董学平. 基于改进 Mycielski 方法的风速预测[J]. 系统工程理论与实践, 2013, 33 (4): 1084 – 1088. GAN M, DING M, DONG X P. Improved Mycielski approach for wind speed prediction [J]. Systems Engineering-Theory & Practice, 2013, 33 (4): 1084 – 1088. (in Chinese)

(责任编辑 梁家峰)