

微网光伏微源功率的超短期预测

为提供微网经济运行与优化控制所需的光伏微源功率信息，减少光伏微源的波动对电网的影响，文中建立了数据挖掘技术和支持向量机相结合的光伏微源功率预测模型。根据光伏微源出力与太阳辐照强度、环境温度和云量等因素密切相关的特点，引入数据挖掘技术，通过日类型模糊分类器与灰色相关度分析法对大量历史数据进行处理，消除冗余信息；建立支持向量机预测模型，实现了与数据挖掘信息的对接。基于某地小型光伏电站实测数据进行功率预测，实例表明该预测模型预测速度快，预测精度较高。

焦然 马龙飞 宫成 张宝群 迟忠君 杨冰 时锐 刘慧珍 曾爽/国网北京市电力公司



焦然/工程师

关键词/Keywords

微网 ·
光伏 ·
功率预测 ·
数据挖掘 ·
支持向量机 ·

0 引言

太阳能是无污染、可再生的清洁能源，安装方便并可普遍推广，因此得到国内外大力推崇^[1]。但光伏发电具有波动性和间歇性的缺点，大规模光伏电站并网会影响电力系统的安全稳定，为了减少对大电网的冲击，系统往往采取限制与隔离的方式来处置光伏站点，如：IEEE1547规定：当电力系统发生故障时DG（分布式能源主要指光伏和风电）必须马上退出运行，大大限制了光伏发电系统效能的充分发挥^[2,3]。在21世纪初，众多学者提出了微网（microgrid）的概念，微电网将发电、负荷、储能装置及控制装置等结合，形成一个单一可控的单元，同时向用户供给电能和热能^[4]。微网中发电装置包括光伏微源在内的分布式能源，这样不仅解决了DG高渗透率

并网问题，还充分发挥了DG为用户和电网带来的潜在价值和经济效益。微网作为大电网的有效补充与DG的有效利用形式，受到世界各国越来越多的关注和支持，微网的发展潜力无疑是巨大的。

功率预测从时间尺度上可以分为超短期功率预测和短期功率预测。超短期功率预测的时间尺度为30 min~6 h，短期功率预测的时间尺度一般为1~2 d^[8]，由于光伏发电时段的特殊性，根据本文实例将预测时段选为9:00~18:00，预测时间尺度30 min，一天共计19个预测点。

目前微网研究热点集中于：微网控制策略^[5]、微网经济运行^[6]和微网孤网判据^[7]等方面。目前对微网内不可控微源的功率预测研究较少，但其重要作用是不能被忽视的。微网中光伏微源容量小，对预测准确度提出了更高要求；同时超短期预测，对预测速度有较高的要求。为解决上述问题本文提出了数据挖掘技术与支持向量机（SVM）结合的方法对微网中光伏功率进行超短期预测。

1 微网中光伏发电系统

1.1 光伏微源的基本结构

微网是由负荷和多种微型电源共同组成的系统，它可同时提供电能和热量；内部微源主要由电力电子器件负责能量的转换，并提供必需的控制；微电网相对于外部大电网表现为单一的受控单元，并可同时满足用户对电能质量和供电安全等的要求。光伏微源是微网中一种间歇性微源，光伏微源基本结构如图1所示。

光伏微源与并网光伏电站不同：光伏微源容量一般较小，一般为几兆瓦

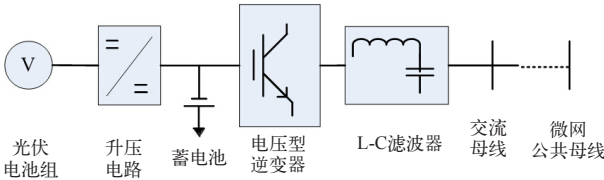


图1 光伏微源基本结构

至几十兆瓦；光伏微源出力受到微网控制策略的约束，为了达到微网的经济运行，光伏微源在微网中一般都满发状态；微网中光伏微源与燃料电池、超级电容等储能装置结合紧密，并网运行时能量高效平稳，孤岛运行时充分利用储能系统的快速充放电功能，可持续可靠地对微网内重要用户供电^[9]。

1.2 光伏微源的环境影响因素

影响光伏微源功率的环境因素很多，如微网所在经纬度、海拔高度、当地太阳高度角、大气透明度、大气压、温度、湿度和日照时间等。对某地区光伏微源而言，可归纳总结为三个主要影响因素。

(1) 光照强度对光伏微源的影响

光伏微源功率与光照强度基本成正比。研究表明在温度固定的条件下，光强在100~1 000 W/m²范围内，光电流始终随光强的增长而线性增长，光照强度与光伏组件的光电流成正比关系；当光照强度在400~1 000 W/m²范围内变化，光伏组件的开路电压基本保持恒定，光照强度对光电压的影响很小。

(2) 云量对光伏微源的影响

云量对太阳辐射具有较强的削弱作用，云量较大时将减少光伏微源的输出功率。云量是表征天空遮蔽程度的气象因子，云量值越大，表示大气的透明度越低，太阳辐射被削弱的越多；云量很小，表示大气透明度高，到达地面的太阳辐射就多。天空中风沙、灰尘和云雾都是影响云量的重要指标，对于干燥地区，削弱太阳直接辐射的主要因子是悬浮在大气中的固体微粒；而对于湿润地区，直接辐射的削弱主要与大气中的水汽和气溶胶中的液态微粒有关。

(3) 温度对光伏微源的影响

当温度变化时，光伏微源的输出功率将发生变化。对一般的晶体硅光伏电池来说，随着温度的升高，短路电流会略有上升，而开路电压要下

降且下降较多，因此总的输出功率要下降。在规定的试验条件下，温度每变化1℃，光伏电池输出功率的变化值称为功率温度系数，通常用 γ 表示。对于一般的晶体硅光伏电池，温度每上升1℃， γ 下降0.35~0.5个百分点。

2 数据挖掘与支持向量机预测

光伏功率预测方法一般分为物理方法和统计方法。物理方法的不需要历史运行数据，光伏电站投产就可以进行预测，但是需要光伏电站详细的地形图、光伏电站布置坐标和光伏电站功率曲线等数据。统计方法基于光伏电站历史运行数据，对光伏电站具体布置资料没有要求，并且程序简明。本文数据挖掘与支持向量机结合预测方法属于统计学方法。

2.1 数据挖掘处理模型

数据挖掘技术指使用模式识别技术、统计和数学技术，在大量的数据中发现有意义的新关系、模式和趋势的过程，也就是从海量数据中挖掘出可能有潜在价值的信息技术^[10]。充分考虑光伏微源功率影响因素和近邻相关度特点，对已有的数据样本进行数据挖掘。

(1) 日类型模糊分类器

光伏功率受光照强度、云量和温度主要因素影响，设日类型用 Y 表示，其包含指标：最高光照强度、平均光照强度、最高气温、平均气温和云量。用 X_1 、 X_2 、 X_3 、 X_4 及 X_5 分别代表最高光照强度、平均光照强度、最高气温、平均气温和云量，得 $Y=\{X_1, X_2, X_3, X_4, X_5\}$ 。由于各地光照强度、气温和云量的低、中、高标准不同，因此可按本地区的实际情况制定一个模糊化标准，将指标的低、中、高，分别取值1、2、3。各类指标模糊值选定后，当日的日类型也初步确定。如某日 $Y_n=\{2, 2, 2, 1, 3\}$ 代表该日最高光照强度，平均光照强度和最高气温为该地区中等水平，平均气温相对较低，云量较大。

将每日光伏微源功率曲线和当日的日类型输入数据库，形成历史数据。根据气象预报得到预测日日类型，如该日日类型为 $Y=\{2, 2, 2, 1, 3\}$ ，则从历史数据库中反向抽取该日类型的所有功率曲线，形成一个同日类型模糊分类库，这个

库中的所有日类型都为{2、2、2、1、3}。

(2) 灰色相关度分析法精选样本数据

灰色相关分析法是一种分析系统中各因素关联程度的方法，其基本思想是根据曲线间相似程度来判断相关程度。采用该方法在已形成的同日类型模糊分类库中，选取与预测日最相关的若干日期，以这些日期的功率信息作为SVM模型的输入数据。灰色相关度分析法步骤如下。

1) 构造序列矩阵。利用预测日的同日类型模糊分类库中每天的气象特征构造序列矩阵。预测日气象预报相关指标准确值为：最高光照强度1 031 W/m²；平均光照强度341 W/m²；最高气温34 ℃；平均气温19 ℃；云量0.5，预测日气象特征（用Q₀表示）为Q₀=[1 031、341、34、19、0.5]^T。同理，同日类型模糊分类库中每日的气象特征为Q₁、Q₂、Q₃…Q_n。这n+1个序列构成序列矩阵表示如下

$$[Q_0, Q_1, Q_2, Q_3, \dots, Q_n] = \begin{bmatrix} Q_{01} & Q_{11} & Q_{21} & \dots & Q_{n1} \\ Q_{02} & Q_{12} & Q_{22} & \dots & Q_{n2} \\ Q_{03} & Q_{13} & Q_{23} & \dots & Q_{n3} \\ Q_{04} & Q_{14} & Q_{24} & \dots & Q_{n4} \\ Q_{05} & Q_{15} & Q_{25} & \dots & Q_{n5} \end{bmatrix} \quad (1)$$

2) 无量纲化处理。由于各因素有不同的计量单位，因而原始数据存在量纲和数量级上的差异，不同的量纲和数量级不便于比较。在计算关联度之前，通常要对原始数据进行无量纲化处理。考虑指标数值差距较大，本文采用极值处理法进行无量纲化

$$\begin{cases} Q_{ij}^* = \frac{Q_{ij} - m_i}{M_i - m_i} \\ M_i = \max\{Q_{0j}, Q_{1j}, Q_{2j}, \dots, Q_{nj}\} \\ m_i = \min\{Q_{0j}, Q_{1j}, Q_{2j}, \dots, Q_{nj}\} \end{cases} \quad (2)$$

式中，i=0, 1, 2, …, n；j=1, 2, 3, 4, 5。无量纲化处理后矩阵如下

$$\begin{bmatrix} Q_{01}^* & Q_{11}^* & Q_{21}^* & \dots & Q_{n1}^* \\ Q_{02}^* & Q_{12}^* & Q_{22}^* & \dots & Q_{n2}^* \\ Q_{03}^* & Q_{13}^* & Q_{23}^* & \dots & Q_{n3}^* \\ Q_{04}^* & Q_{14}^* & Q_{24}^* & \dots & Q_{n4}^* \\ Q_{05}^* & Q_{15}^* & Q_{25}^* & \dots & Q_{n5}^* \end{bmatrix} \quad (3)$$

3) 相关度计算。计算公式(3)中Q_{0j}^{*}与Q_{ij}^{*}相似系数，计算公式如下

$$\varepsilon_{ij} = \frac{\min_i \min_j |Q_{0j}^* - Q_{ij}^*| + \rho \max_i \max_j |Q_{0j}^* - Q_{ij}^*|}{|Q_{0j}^* - Q_{ij}^*| + \rho \max_i \max_j |Q_{0j}^* - Q_{ij}^*|} \quad (4)$$

式中，i=1, 2, …, n；j=1, 2, 3, 4, 5。ρ是分辨系数，ρ∈(0, 1)，通常ρ取0.5。得到关联系数矩阵如下

$$\begin{bmatrix} \varepsilon_{11} & \varepsilon_{21} & \dots & \varepsilon_{n1} \\ \varepsilon_{12} & \varepsilon_{22} & \dots & \varepsilon_{n2} \\ \varepsilon_{13} & \varepsilon_{23} & \dots & \varepsilon_{n3} \\ \varepsilon_{14} & \varepsilon_{24} & \dots & \varepsilon_{n4} \\ \varepsilon_{15} & \varepsilon_{25} & \dots & \varepsilon_{n5} \end{bmatrix} \quad (5)$$

由公式(5)计算Q₀与Q_i的相关度，公式如下

$$\gamma_i = \frac{1}{5} \sum_{j=1}^5 \varepsilon_{ij} (i=1, 2, 3, \dots, n) \quad (6)$$

4) 确定微源功率预测所需历史序列

根据预测日气象特征Q₀与同日类型模糊分类库每日气象特征序列[Q_i]相比较，得出相关度γ_i，按照相关度从大到小的顺序排列，取前6组序列，并将这6组序列按时间顺序排列作为SVM预测的样本数据。

2.2 支持向量回归预测模型

支持向量机(SVM)方法基于结构风险最小化原则，可提高学习机的泛化能力；此外SVM算法是一个凸二次优化问题，能够保证找到的极值解也就是全局最优解。它在解决小样本、非线性及高维模式识别中表现出许多特有的优势^[11]。

用SVM处理回归问题，基本思想是通过一个非线性映射φ，把输入空间的数据x映射到一个高维特征空间中去，然后在这一高维空间中作线性回归，也称之为SVR(Support Vector Regression)。如给定样本容量{x_i, y_i} (i=1, 2, …, n), n为样本容量；x_i为输入向量；y_i为目标函数输出数据。高维特征空间内函数f

$$y = f(x, w) = w \phi(x) + b \quad (7)$$

式中，w为权值向量；b为常数。对每一个样本f(x_i, w)和y_i的差值都很小，则f(x, w)能从x准确地预测y。寻找最优函数需要，满足下面公式

$$\begin{cases} \min \frac{1}{2} \|w\|^2 \\ |y - f(x, w)| \leq \varepsilon \end{cases} \quad (8)$$

其中， $\varepsilon \geq 0$ ，用来表示SVM预测值与实际值最大的差距。由于噪声、误差等原因，在 ε 选定合理状态下，有一些样本点会落到 ε 之外，需要引入松弛变量 ζ_i 和 ζ_i^* ，惩罚因子 C ，处理该问题，公式 (8) 做如下修改

$$\begin{cases} \min \frac{1}{2} \|w\|^2 + c \sum_{i=1}^l (\zeta_i + \zeta_i^*) \\ |y - f(x, w)| \leq \varepsilon + \zeta_i \\ f(x, w) - y \leq \varepsilon + \zeta_i^* \\ \zeta_i, \zeta_i^* \geq 0 \end{cases} \quad (9)$$

上面公式实质是一个凸集二次规划问题。引入拉格朗日乘子 α_i 、 α_i^* ，并引入核函数，公式 (7) 改写成系数为 α_i 、 α_i^* 和 b 的函数

$$\begin{aligned} y &= f(x, \alpha_i, \alpha_i^*) \\ &= \sum_{i=1}^n (\alpha_i, \alpha_i^*) K(x, x_i) + b \end{aligned} \quad (10)$$

式中， α_i 、 $\alpha_i^* \in [0, C]$ 。

$$K(x, x_i) = \exp[-g(x - x_i)^2] \quad (11)$$

$K(x, x_i)$ 为核函数，一般有多项式核函数、径向基 (RBF) 核函数等，式 (11) 为本文所选核函数，RBF核函数。

3 数据挖掘与SVM结合预测步骤

数据挖掘技术通过发现数据间的关系，去掉数据中的冗余信息，从而大大简化输入信息的数据空间位数；SVM具有较好的噪声干扰抑制能力，且具有良好的泛化能力。两者结合可以提高运算速度和准确性。预测流程如下。

- 1) 整理历史数据，得出样本数据。对统计的历史数据通过本文提到的数据挖掘技术，形成具有高相似性气象特征的样本数据。
- 2) 样本数据处理，得到输入格式。将样本数据中与预测时刻点相同的数据作为SVM中的 y_i 值， x_i 值为对应时刻点前4个时段的功率值，并作归一化处理，形成 (y_i, x_i) 格式。
- 3) 对SVM计算模型参数初始化。将拉格朗日乘子 α_i 、 α_i^* 和阈值 b 赋予初始值。

- 4) 输入 $[y_i, x_i]$ 样本数据，采用SVM计算工具优化 α_i 、 α_i^* 和 b 。
- 5) 将预测数据整理为 $[y^*, x]$ 格式，输入SVM模型中，得到预测值 y 。

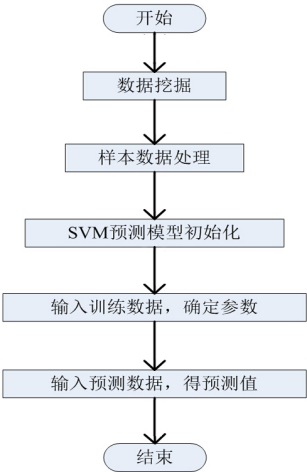


图2 光伏微源功率预测流程图

4 实例分析

根据微网中光伏微源的特点，选择与其特性相似的某小型光伏电站，采用该站实测数据进行分析。以2012年1月共计31天数据作为历史数据，对2012年2月1号全天19个预测点进行功率预测。

根据2月1号的气象预报，该气象日类型为 $y=[3, 3, 1, 1, 1]$ ，形成同日类型模糊分类库，采用灰色相关度分析法确定样本数据日期及整点功率值如表1所示。

表1 样本数据日期及整点功率值 (单位: MW)

	9/h	10/h	11/h	12/h	13/h
5日	0.011	1.407	2.840	3.240	4.294
11日	0.011	1.572	3.229	3.916	4.156
22日	0.032	1.785	3.069	3.959	4.332
23日	0.043	1.433	2.755	3.309	4.167
25日	0.059	2.041	3.298	4.076	4.460
31日	0.107	1.987	3.287	4.033	4.353
	14/h	15/h	16/h	17/h	18/h
5日	4.289	3.996	3.506	1.156	0.133
11日	4.220	4.103	3.437	2.531	0.352
22日	3.474	3.261	2.819	2.563	0.741
23日	3.863	3.644	3.479	1.545	0.293
25日	4.524	3.778	3.399	2.925	1.263
31日	4.396	4.183	3.714	2.867	0.591

整理样本数据，利用SVM预测模型预测2月1日功率。预测值与真实值19个预测点拟合曲线如图3所示。预测值与真实值整点功率及相对误差见表2。

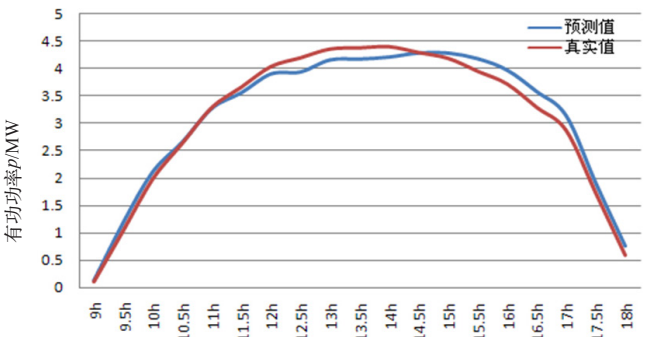


图3 预测值与真实值功率拟合曲线

表2 预测值与真实值整点功率及相对误差 P (单位: MW)

/h	真实值	预测值	相对误差(%)
9	0.106 6	0.129 1	-21.11
10	1.987 4	2.112 4	-6.29
11	3.287 4	3.267 2	0.61
12	4.033 4	3.892 07	3.50
13	4.353 0	4.147 4	4.72
14	4.395 7	4.198 77	4.48
15	4.182 5	4.271 81	-2.14
16	3.713 7	3.960 59	-6.65
17	2.866 5	3.130 38	-9.21
18	0.591 4	0.758 72	-28.29

预测点14~30相对误差最小为0.37%，预测点18~00相对误差最大为-28.29%，预测值与真实值19个预测点平均绝对误差为7.39%。

从相对误差来看，9 h和18 h误差较大，主要原因光伏微源功率输出很小，预测稍有偏差就会导致较大的误差；其余误差绝对值都在10%以内，中间大部分时间点误差能保持在5%以内。

5 结束语

本文充分考虑了气象因素对光伏微源出力

的影响，提出了日类型模糊分类和灰色相关度分析的数据挖掘技术，对训练样本进行了有效的约简，大大提高了训练速度；采用SVM模型进行预测，在光伏微源容量很小，预测相对困难的情况下，绝对误差平均值为7.39%，较为准确，对微网的运行控制与经济运行有一定的指导作用。

参考文献

[1] 张抒阳,张沛,刘珊珊.太阳能技术及其并网特性综述[J].南方电网技术,2009,3(4):64-67

[2] 郑志杰,李磊,王葵.大规模光伏并网电站接入系统若干问题的探讨[J].电网与清洁能源,2010,26(2):74-76

[3] 张兴科.光伏并网发电功率波动与对策[J].电网与清洁能源,2011,27(6):55-60

[4] 盛鸥,孔力,齐智平等.新型电网-微电网 (Microgrid) 研究综述[J].继电器,2007,35(12):75-80

[5] 宋斌.基于Agent的微网控制系统研究[J].电力科学与工程,2009,25(9):22-26

[6] 郭佳欢.微网经济运行优化的研究[D].华北电力大学硕士学位论文,2010

[7] 张岚,殷桂梁,韩郁等.基于无功微扰/频率偏差相关度的光伏发电系统孤岛检测方法[J].南方电网技术, 2010,4(1):130-134

[8] 卢静,翟海青等.光伏发电功率预测统计方法研究[J].华东电力, 2010,38(4)

[9] 雷金勇,郝木凯,张兴.微网中储能系统功率控制策略及其动模实验[J].南方电网技术, 2012,6(3):57-61

[10] Chen L D, Toru S. Data mining methods, applications, and tools[J]. Information system management, 2000, 17(1): 65-70.

[11] 朱永强,田军.最小二乘支持向量机在光伏功率预测中的应用[J].电网技术,2011,35(7):54-59.

作者简介

焦然 (1987-), 男, 本科, 工程师, 主要从事新能源与电能质量工作。

(收稿日期: 2015-12-10) **EA**

(编辑: 张万英)