

Short-Term Wind Power Prediction Based on K-Means Clustering Algorithm

Xiaojuan Yan, Renxi Gong

School of Electrical Engineering, Guangxi University, Nanning
Email: 617956088@qq.com, rxgong@gxu.edu.cn

Received: Apr. 18th, 2012; revised: Apr. 25th, 2012; accepted: Apr. 27th, 2012

Abstract: Forecasting the output power of wind farm play a vital role in the reducing of the running cost of wind power plants and the reasonable arrangements for the dispatch of power systems. Improving the prediction accuracy of wind power can contribute to lower the detrimental impact of wind power plants on power grid as well as improve the competitiveness of wind power plants against others in electricity markets. As in the establishment of short-term wind power forecasting model, the sample selection has a greater impact on prediction accuracy, so the study of sample selection method is very important. In this paper, a new method is proposed in which the history wind power data is clustered by K-means algorithm, data is classified through the LVQ net and the prediction model of wind power is established with the least-squares method. The practical application shows that the method can be utilized to predict the wind power effectively and precisely, and it is quite significant for the regulation of wind power.

Keywords: K-Means Clustering; LVQ; Short-Term Prediction of Wind Power

基于 K-均值聚类的风能短期功率预测

颜晓娟, 龚仁喜

广西大学电气工程学院, 南宁
Email: 617956088@qq.com, rxgong@gxu.edu.cn

收稿日期: 2012 年 4 月 18 日; 修回日期: 2012 年 4 月 25 日; 录用日期: 2012 年 4 月 27 日

摘要: 风电场输出功率的预测对降低风电场运行成本及合理安排电力系统调度具有重要意义。提高风能功率的预测精度, 可以有效地减轻风电对电网的影响, 同时提高风电场在电力市场中的竞争能力。在建立风能短期功率预测模型时, 由于样本的选取对预测的精度有较大的影响, 因此研究样本的选取方法具有重要意义。本文提出了一种利用 K-均值算法对历史功率数据样本进行聚类, 根据得到的聚类结果训练学习向量量化(LVQ)神经网络, 利用训练好的神经网络对待预测数据进行自动分类, 最后使用最小二乘法建立风能功率预测模型的方法。通过实验验证了该方法的有效性和可行性, 对于风电调度具有一定的参考意义。

关键词: K-均值聚类; LVQ; 风能短期功率预测

1. 引言

随着全球能源需求的不断扩大和环境污染的日益严重, 可再生能源的开发与利用成为能源产业研究的热点和重要课题, 风能作为一种清洁、开发潜力巨大的可再生能源更是成为各国政府发展能源产业的优先主题而倍受重视, 我国作为一个能源需求和消费大国,

大力发展并网型风力发电成为我国解决能源紧张, 减轻环境污染和施行可持续发展战略的重要举措。

由于风能存在随机性, 波动性和间歇性的特点, 使得大规模风电的并网成为影响电力系统稳定运行和电能质量的重要因素, 对风能功率的准确预测有利于了解电力系统在一定时间段内的运行情况, 从而采

取有效措施确保电力系统的功率平衡和经济调度,对电力系统稳定可靠运行具有十分重要的意义^[1-3]。风能功率的预测可根据不同的方式进行分类。根据预测时所依据的物理量的不同可分为:根据温度、湿度、风力、降雨量等众多的天气因素预测风电场输出功率和根据风电场单一的风速或输出功率直接预测下一时段的功率两类;也可根据预测时选用的方法来分:分为人工神经网络法(ANN)、时间序列法、模糊预测法(ANFIS)、卡尔曼滤波法(Kalman-filter)、最小二乘法以及支持向量机(SVM)预测;还可根据预测的时间长短来分,分为超短期预测和短期预测。一般将不超过 30 min 的预测称为超短期预测,短期预测一般可认为是 30 min~72 h 的预测。目前对于更长时间尺度的预测,还存在较大的困难^[4-7]。

本文首先利用 K-均值算法对某风电场的历史功率数据进行数据聚类分析,然后根据 K-均值算法的聚类结果对 LVQ 网络进行分类训练,当达到分类识别要求后保存其分类号,最后利用最小二乘法根据分类号调用相应的拟合函数组合方法实现对功率的预测。通过具体模拟实验,本文不仅验证了利用这种模型进行预测的可行性,而且也显示得到的预测精度比相关的运用统计聚类分析风速,再由风速导出功率的预测精度更高。

2. K-均值聚类

聚类即是把一组个体按照相似性归成若干类别即“物以类聚”,它的目的是使得属于同一类别的个体之间的距离尽可能的小而不同类别上的个体间的距离尽可能的大,通过对单一的历史功率值聚类后可以发现和归纳出规律,如风速的大小和强度等,由此可以通过对历史数据的聚类结果来对未来的变化趋势进行外推,在这种情况下,对历史数据进行聚类是非常必要的。

聚类的方法有很多种,但应用最广泛的仍然是 K-均值聚类。其基本思想是:对于给定的 N 个对象的数据集,将构造数据划分 K 类,每个类代表一个簇, $K \leq N$ 。也就是说, K-均值算法将数据划分为 K 个簇,这 K 个簇满足 2 个条件: 1) 每个簇中至少包含一个对象; 2) 每一个对象属于且仅属于一个簇。对于给定的 K, 算法首先给出一个初始的划分方法,以后通过

反复迭代的方法改变划分,使得每一次改进之后的划分方案都比前一次更好,其中类内差异 $W(c)$ 函数和类间差异 $b(c)$ 函数是聚类结果的评价函数(类内差异函数衡量类内的紧凑性,类间差异函数衡量不同类之间的距离),而 K-均值算法的准则函数则保证生成的结果簇尽可能的紧凑和独立^[8]。

类内差异衡量函数如式(1)所示:

$$W(c) = \sum_{i=1}^K W(ci) = \sum_{i=1}^K \sum_{x \in ci} d(x, \bar{x}_i)^2 \quad (1)$$

类间差异衡量函数如式(2)所示:

$$b(c) = \sum_{1 \leq j(i \leq K)} d(\bar{x}_j, \bar{x}_i)^2 \quad (2)$$

K-均值算法的准则函数如式(3)所示:

$$E = \sum_{i=1}^K \sum_{x \in ci} |x - \bar{x}_i|^2 \quad (3)$$

其中 K 表示类, x 表示给定的点, x_i 为簇 ci 的平均值。

3. LVQ 网络分类

LVQ 神经网络由芬兰学者 Teuvo kohonen 提出,它将竞争学习思想和有监督学习算法相结合,在网络学习过程中,通过教师信号对输入样本的分配类别进行规定,从而克服了自组织网络采用无监督学习算法带来的缺乏分类信息的弱点。LVQ 神经网络在模式识别和优化领域有着广泛的应用。

LVQ 网络的算法是一种监督类型的分类方法,该算法与自组织特征映射神经网络(SOM)算法最大的区别在于提供给 LVQ 网络的每个训练例都有一个“标记”(label),该“标记”用于指导每个训练例所属的类别,在网络的训练过程中起到一定的监督作用。LVQ 网络由输入层、隐含层和输出层三层组成。输入层与隐含层之间完全连接,每个输出层神经元与隐含层神经元的不同组相连接。隐含层和输出层神经元之间的连接权值固定为 1。在网络的训练过程中,输入层和输出层神经元之间的权值被修改。当某个输入模式被送至网络时,最接近输入模式的隐含神经元因获得激发而赢得竞争,从而容许其输出为 1,而其他输出均为 0^[9]。

LVQ 网络学习算法的步骤如下:

1) 初始化。竞争层神经元权值向量 $W_{j1}(0)$, $j = 1, 2, \dots, M$ 赋小随机数, 确定初始学习速率 $\eta(0)$ 和训练次数 K 。

2) 输入样本向量 X 。将输入向量 $X = [x_1, x_2, \dots, x_n]$ 送入输入层。

3) 计算隐含层权值向量与输入向量之间的距离, 计算公式如式(4)所示:

$$d_j = \sqrt{\sum_{i=1}^n (x_i - w_{ij})^2} \quad (4)$$

4) 选择与权值向量的距离最小的神经元, 并把其称为胜出神经元, 记为 j^* 。

5) 更新连接权值。如果胜出神经元和预先指定的分类一致, 则为正确分类, 否则称为不正确分类。权值的调整量如式(5)所示:

$$\Delta w_{ij} = \pm \eta (x_i - w_{ij}) \quad (5)$$

其中, 正确调整符号为正, 不正确调整符号为负。

6) 判断算法是否结束。如果迭代次数大于预先设定的次数, 算法结束, 否则返回第二步, 进入下一轮学习。

LVQ 网络工作流程图如图 1 所示。

4. 功率预测模型

在本文中具体功率预测模型的函数通过拟合回归函数完成。虽然有很多不同的方法来求样本回归函数, 但广泛使用的是最小二乘法^[10]。最小二乘法根据一组给定的数据 (x_i, y_i) ($i = 0, 1, \dots, m$), 在函数空间 $\Phi = \text{span}\{\varphi_0, \varphi_1, \dots, \varphi_n\}$ 中找一个拟合函数 $y = S(x)$, 使误差平方和最小。

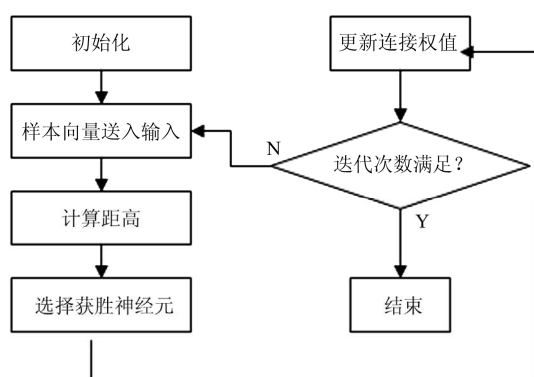


Figure 1. Work flow chart of LVQ net
图 1. LVQ 网络工作流程图

拟合函数形式如式(6)所示:

$$S(x) = a_0 \varphi_0(x) + a_1 \varphi_1(x) + \dots + a_n \varphi_n(x) \quad (n < m) \quad (6)$$

误差平方和公式如式(7)所示:

$$\|\delta\|_2^2 = \sum_{i=0}^m \delta_i^2 = \sum_{i=0}^m [S^*(xi) - yi]^2 = \min_{S(x) \in \phi} \sum_{i=1}^m [S(xi) - yi]^2 \quad (7)$$

5. 实例分析

以某风电场的 112 天的历史功率数据为例进行分析, 每天 96 点的历史功率值, 由这些数据创建短期功率预测模型为例。首先进行 K-均值聚类, 根据问题的性质, 本文把每 4 个小时(16 个数据)作为一个时段, 对前面 3 个小时(12 个数据)的数据进行聚类, 后面相应的 1 个小时(4 个数据)作为预测数据。一共划分为 672 条数据, 聚为 12 类。在聚类这方面已经有很多成熟的统计软件如 SPSS(Statistical Product and Service Solutions), SAS 等。由于 SPSS 统计软件简单易学和界面友好等特点, 选用 SPSS 来进行聚类工作。利用 SPSS 软件对国家能源局给的 4 个表中的数据进行聚类运算, 聚类中心和类别如表 1。

由于数据有限, 利用 LVQ 网络对历史功率数据的聚类结果进行分类训练时, 如果网络训练不充分就会导致分类错误。为了充分训练网络, 选择其中的 662 条数据作为网络的训练样本, 其余 10 条数据的作为测试样本。因此对应的 LVQ 网络的输入层单元数为 12, 输出层单元数为 12。隐层单元数根据经验公式(8)选择。

Tab 1. Cluster centers
表 1. 聚类中心

类号	类中心值											
第 1 类	321	285	314	352	387	443	471	535	518	550	498	485
第 2 类	622	586	649	692	678	545	667	581	523	535	476	448
第 3 类	570	665	636	664	662	613	550	398	364	454	425	445
第 4 类	278	276	262	222	204	178	183	179	177	166	167	159
第 5 类	704	756	746	734	745	772	744	737	741	755	755	752
第 6 类	245	242	396	361	279	141	111	177	130	426	716	763
第 7 类	63	53	55	49	56	51	57	54	56	49	56	63
第 8 类	400	348	391	438	474	529	594	631	686	728	671	665
第 9 类	845	807	477	602	605	289	208	207	847	851	829	843
第 10 类	127	124	160	166	150	181	224	258	314	359	394	394
第 11 类	852	846	846	849	0	850	849	854	0	850	846	848
第 12 类	308	368	381	411	485	510	462	386	335	274	281	264

$$H = (I + J) * 0.5 + M \quad (8)$$

其中 H 为隐层单元数, I 为输入单元数, J 为输出单元数, M 为 1~10 之间的整常数。

通过实验我们发现 LVQ 网络的分类正确率在 50%~90%之间, BP 网络的分类正确率在 30%~70%之间, 由于 LVQ 网络将竞争学习思想和有监督学习算法相结合, 在分类这方面更优于 BP 网络。当 LVQ 网络达到训练要求时保存网络和权值, 以便判断分类时直接调用网络。

在建立预测模型的函数时, 以每个类的聚类中心和后面对应的功率值作为预测值, 建立相应的最小二乘拟合函数, 预测时根据 LVQ 网络判断出待预测数据的分类号调用相应的最小二乘拟合函数进行功率预测, 完成预测工作。某一天相应时间的预测值和实测值如表 2。

实验结果数据表明, 依照上述原理建立的模型, 误差小于 21%。参考文献[11]运用统计聚类分析风速, 再由风速导出功率^[11]的方法的误差一般在 22%~24%之间, 相比之下, 聚类分析历史功率值, 再根据数据的分类来预测的方法具有更好的预测精度。

Table 2. Forecasting results
表 2. 预测结果

时刻	4:00	4:15	4:30	4:45	5:00	5:15
实际值	249	396	389	494	508	462
预测值	300	463	469	582	613	548
误差	20.5%	16.9%	13.9%	17.8%	20.7%	18.7%
时刻	5:30	5:45	6:00	6:15	6:30	6:45
实际值	357	327	221	233	231	249
预测值	431	392	263	276	278	201
误差	20.8%	19.9%	19.0%	18.5%	20.5%	19.2%

6. 结论

风电场的规模日益扩大, 对电网的稳定运行的影响也日益深远。对风电场的输出功率进行预测就显得尤为重要。考虑到只有单一的历史功率值数据, 本文尝试了借鉴负荷预测的经验, 建立了基于统计聚类, 利用 LVQ 网络分类, 最小二乘法拟合来进行风电功率的预测的模型。通过具体的实例验证了这种方法的可行性。此外, 运用数据挖掘技术, 深入挖掘数据序列本身的相似性, 调用数据库动态选择训练数据以及找到更好的拟合方程是以后要做的工作。

参考文献 (References)

- [1] 都志杰, 马丽娜. 风力发电[M]. 北京: 化学工业出版社, 2009: 29-35.
- [2] 顾为东. 中国风电产业发展新战略与风电并网理论[M]. 北京: 化学工业出版社, 2006: 5-8.
- [3] 王丽婕, 廖晓钟, 高阳等. 风电场发电功率的建模和预测研究综述[J]. 电力系统保护与控制, 2009, 37(13): 118-121.
- [4] 洪翠, 林维明. 风电场风速及风电功率预测方法研究综述[J]. 电网与清洁能源, 2011, 27(1): 60-66.
- [5] 谷兴凯, 范高峰, 晓蓉等. 风电场功率预测技术综述[J]. 电网技术, 2007, 31(S2): 335-338.
- [6] 吴兴华, 周晖, 黄梅. 基于模式识别的风电场风速和发电功率预测[J]. 继电器, 2008, 36(1): 27-32.
- [7] 杨秀媛, 肖洋, 陈树勇. 风电场风速和发电功率预测研究[J]. 中国电机工程学报, 2005, 25(11): 1-5.
- [8] 毛国君, 段立娟. 数据挖掘原理与算法[M]. 北京: 清华大学出版社, 2007: 166-175.
- [9] 韩立群. 人工神经网络理论、设计及应用[M]. 北京: 化学工业出版社, 2007: 107-110.
- [10] 李庆阳, 王能超, 易大义. 数值分析[M]. 武汉: 华中科技大学出版社, 2008: 64-65.
- [11] 方江晓, 周晖, 黄梅等. 基于统计聚类分析的短期风电功率预测[J]. 电力系统保护与控制, 2011, 11(39): 69-73.