计及气象因素和日期属性的短期负荷预测研究

气象因素和日期属性作为影响电力负荷的主要因素，近年来已成为负荷预测技术研究的焦点。为了既简洁又全面地表征各气象因素对负荷的影响，利用主成分分析法对多变量气象因素进行降维处理，得到综合气象指标。同时对日期属性进行量化评估，并利用思维进化算法优化Elman神经网络的权值阈值，提出MEA－Elman预测模型。最后以甘肃某地区实际负荷为例，表明了该方法能够有效提高配电网短期负荷的预测准确度。

包广清　林麒麟／兰州理工大学电气工程与信息工程学院
吴国栋／国网甘肃省电力公司调度控制中心

包广清／教授

关键词／Keywords
负荷预测·
气象因素·
日期属性·
主成分分析·
MEA－Elman预测模型·

0　引言

电力负荷预测是电力系统规划、营销、市场交易和调度等部门工作的重要依据，科学的预测是正确决策的依据和保证[1]。根据预测时间长短，通常可以将电力系统负荷预测分为长期、中期、短期以及超短期预测等几种类型[2]。以24点日负荷预测为研究对象，属于短期负荷预测研究范畴。在当前的电力市场条件下，提高短期负荷预测准确度，对安排开停机计划、机组最优组合和发电容量的合理规划以及电力市场交易等有着重要的实际意义[3]。

目前，国内外专家学者提出了各种预测方法，如时间序列预测法、回归分析预测法等[4-5]。这些预测算法是以历史负荷作为预测的主要依据，没有考虑气象、日期、经济、政治和季节等干扰因素对负荷的影响，进而影响预测准确度。而气象因素和日期因素作为影响负荷预测的主要因素，近年来已成为负荷预测的焦点[6]。利用主成分分析法对多个气象因子进行分析，得到综合气象指标，避免各气象因子之间的信息冗余，减少计算复杂度，进而提高预测速度和效率。然后对日负荷曲线进行分析，并对日期因素进行量化评估。最后针对BP神经网络动态性能的不足，建立Elman动态神经网络的预测模型，并利用思维进化算法全局寻优、并行搜索的性能优化Elman神经网络，提出基于MEA－Elman网络预测模型，以甘肃某地区真实负荷数据作为数据样本，验证所提方法的有效性。

国家自然科学基金项目（51267011）。

1　气象因素及其分析

用电负荷受众多因素影响，其中气象因子的影响尤为显著。气象因子主要包括温度、相对湿度、气压、风速和降雨量等，不同的因素会有不同程度的影响，其中温度、湿度对负荷的影响最大。炎热的夏季、寒冷的冬季都会使负荷发生急剧变化。只考虑温湿度，而忽略其他气象因素，对负荷预测的准确性亦存在不良影响。但各气象因子的直接作用会增加计算维数和信息冗余度，进而使得负荷预测变得复杂，影响预测准确度和效率。主成分分析是解决此类问题强有力的工具[7-8]，利用降维的思想，把多个指标转化为较少几个综合指标，而这些综合指标既能尽量多地反映原来的信息，同时又保持各自相互独立。

假设有n个样本，每个样本共有m

个变量，构成一个$n\times m$阶的数据矩阵，即

$$\boldsymbol{X}=\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{pmatrix} \tag{1}$$

主成分分析具体步骤如下。

步骤1：将原数据样本标准化处理，并计算各变量之间的相关系数矩阵，即

$$\boldsymbol{R}=\begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1m} \\ r_{21} & r_{22} & \cdots & r_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ r_{m1} & r_{m2} & \cdots & r_{mm} \end{pmatrix} \tag{2}$$

式中，r_{ij}为原变量的X_i与X_j之间的相关系数，其表达式为

$$r_{ij}=\frac{\sum_{k=1}^{n}(x_{ki}-\overline{x_i})(x_{kj}-\overline{x_j})}{\sqrt{\sum_{k=1}^{n}(x_{kj}-\overline{x_i})^2\sum_{k=1}^{n}(x_{kj}-\overline{x_j})^2}} \tag{3}$$

步骤2：解特征方程$|\lambda I-R|=0$，求出特征值λ_i（$i=1,2,\cdots,m$），并使其按值的大小顺序排列。

步骤3：计算各特征值对应的主成分贡献率及累计贡献率。主成分z_i的贡献率为

$$\frac{\lambda_i}{\sum_{k=1}^{m}\lambda_k}\quad i=1,2,\cdots,m \tag{4}$$

累计贡献率为

$$\frac{\sum_{k=1}^{i}\lambda_k}{\sum_{k=1}^{m}\lambda_k}\quad i=1,2,\cdots,m \tag{5}$$

通常取累计贡献率为85%～95%的特征值λ_1，λ_2，…，λ_p。

步骤4：计算特征值λ_1，λ_2，…，λ_p所对应的特征向量，则其p个主成分可以表示为

$$\begin{cases} z_1=a_{11}x_1+a_{12}x_2+\cdots+a_{1m}x_m \\ z_2=a_{21}x_1+a_{22}x_2+\cdots+a_{2m}x_m \\ \vdots \\ z_p=a_{p1}x_1+a_{p2}x_2+\cdots+a_{pm}x_m \end{cases} \tag{6}$$

其中$p\leqslant m$，进而实现降维的目的，即将m个气象因素减少为p个综合指标。

以120组气象样本数据进行主成分分析，每个样本包括温度x_1、降雨量x_2、风速x_3、云量x_4、相对湿度x_5和光强x_6等6个气象因素，根据步骤1可得其相关系数矩阵为

$$\boldsymbol{R}=\begin{pmatrix} 1.000 & 0.382\,5 & 0.069 & -0.046 & -0.031 & 0.072\,6 \\ 0.382\,5 & 1.000 & 0.734\,2 & 0.672\,1 & 0.098\,1 & 0.746\,5 \\ 0.069 & 0.734\,2 & 1.000 & 0.657\,6 & 0.221\,7 & 0.706\,7 \\ -0.046 & 0.672\,1 & 0.657\,6 & 1.000 & -0.029\,6 & 0.890\,4 \\ -0.031 & 0.098\,1 & 0.221\,7 & -0.029\,6 & 1.000 & 0.289\,5 \\ 0.072\,6 & 0.746\,5 & 0.706\,7 & 0.890\,4 & 0.289\,5 & 1.000 \end{pmatrix} \tag{7}$$

由步骤2、步骤3可得其特征值、贡献率以及累计贡献率如表1所示。

表1 特征值及各主成分贡献率

主成分	特征值	贡献率(%)	累计贡献率(%)
1	3.268 5	54.48	54.48
2	1.129 4	18.82	73.30
3	1.009 9	16.83	90.13
4	0.376 1	6.27	96.40
5	0.171 4	2.86	99.26
6	0.044 7	0.74	100.00

根据累计贡献率需达到85%以上的原则，可以取前3个特征值所对应的特征向量作为新的综合气象指标，表达式为

$$\begin{cases} z_1=-0.103x_1-0.492x_2-0.477x_3-0.484x_4-0.127x_5-0.520x_6 \\ z_2=0.844x_1+0.278x_2-0.079x_3-0.126x_4-0.411x_5-0.140x_6 \\ z_3=0.358x_1+0.031x_2+0.030x_3-0.359x_4+0.861x_5-0.003x_6 \end{cases} \tag{8}$$

式中，z_1、z_2和z_3为新的气象指标；各变量系数代表不同气象因子的比例权重。

通过主成分分析将原来6个气象因子减少为3个综合气象指标，降低了计算复杂程度。

2 负荷特性分析

采用甘肃某地区的负荷数据作为研究对象，图1为甘肃某地区的日负荷曲线。通常日负荷曲线有两个高峰，即早高峰和晚高峰，早高峰主要是由企业用电而引起的，而晚高峰是由于家庭用电

而引起的。从图1中可以看出，该地区从凌晨1点开始用电量减少，负荷量呈下降趋势；早上8点～11点用电量逐步上升；中午12点～晚上20点，负荷曲线维持在一定的范围内，存在一定的波动；晚间20点～24点，负荷量逐步下降。图2为该地区一周的负荷曲线，采样周期为1 h，共168点的负荷数据，反映了从周一～周日的7天的负荷变化情况。其中周一～周五（0～120 h）为正常工作日，周六、周日（120～168 h）为休息日，工作日内人们处于生产生活状态，产生的负荷较大，而周六、周日则处于休息状态，负荷略低于正常工作日负荷。

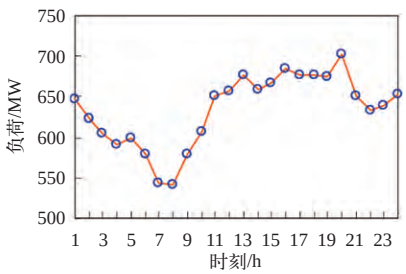

图1　甘肃某地区日负荷曲线

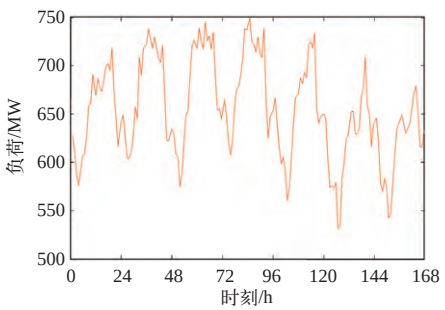

图2　甘肃某地区一周负荷曲线

同时为了考虑各日期类型之间的相关程度，采用如下量化规则，如表1所示。

表1　日期属性量化

日期属性	映射前值	映射后值
周一	1	0.1
周二	2	0.2
周三	3	0.3
周四	4	0.4
周五	5	0.5
周六	6	0.6
周日	7	0.7
节假日	6	0.6

采用下式来计算日期因素的相似系数为

$$c_i = 1 - \left| f(x_i) - f(x_0) \right| \tag{9}$$

式中，c_i为相似系数；x_i、x_0为第i个历史日和预测日的星期类型；$f(x_i)$、$f(x_0)$为x_i和x_0映射后的值。

3　MEA－Elman预测模型

3.1　MEA算法

MEA算法由孙承意等人提出，解决了进化算法中早熟、收敛速度慢等问题[9]。其与遗传算法相比，具有的特点为：可以记忆不止一代的进化信息；结构上固有的并行性；用趋同和异化代替了遗传算法中的交叉和变异，提高了进化效率[10]。趋同是在子群体内进行，通过竞争获得适应度值较高的优胜个体；异化是在整个解空间进行，将临时子群体的适应度值与优胜子群体的适应度值比较，若高于优胜子群体适应度值，则替代该优胜子群体，反之，则遗弃。

思维进化算法基本思路如下。

Step1：参数设置，设置种群大小POP_{size}，优胜子群体数目$Best_{size}$，临时子群体数目Tem_{size}，子群体大小S_G。其中S_G的表达式为

$$S_G = \frac{POP_{size}}{Best_{size} + Tem_{size}} \tag{10}$$

Step2：随机产生初始种群，然后计算子群体适应度值，并根据适应度值，产生优胜子群体和临时子群体。

Step3：将优胜子群体和临时子群体进行趋同操作，然后再进行异化操作。

Step4：解码最优个体，并判断其是否满足最优条件，若满足，输出最优个体，若不满足则返回趋同异化操作。

3.2　Elman神经网络

Elman神经网络是一种动态反馈网络，包括输入层、隐含层、承接层和输出层[11]。其中承接层是一个特别的隐含层，该层从隐含层接收反馈信号，通过联接记忆将上一个时刻的隐层状态连同当前时刻的网络输入一起作为隐层的输入，相当于状态反馈，网络结构如图3所示。

假设输入为$u_{(k-1)}$，承接层输出为$y_c(y)$，隐含

层输出为$y_h(k)$，输出层为$y_o(k)$，隐含层传递函数为$G(k)$，输出层传递函数为$F(k)$，则有该网络的状态空间表达式为

$$y_h(k)=G[w_1u(k-1)+w_2y_c(k)+b_1] \quad (11)$$

$$y_c(k)=y_h(k-1) \quad (12)$$

$$y_o(k)=F[w_3y_h(k)+b_2] \quad (13)$$

式中，w_1、w_2和w_3分别为输入层到隐含层、承接层到隐含层和隐含层到输出层的连接权值；b_1、b_2分别为隐含层阈值和输出层阈值。

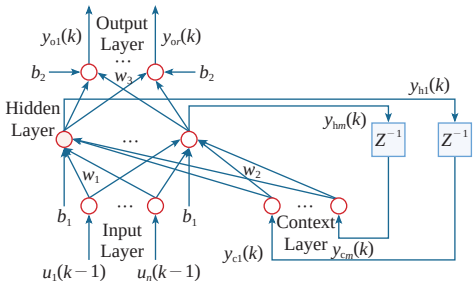

图3 Elman神经网络结构图

3.3 优化算法流程

采用思维进化算法对Elman神经网络的权值和阈值进行优化，建立MEA－Elman网络模型，具体步骤如下。

首先确定网络拓扑结构并设置MEA算法参数。根据网络结构，确定思维进化算法的编码长度，记为S，表达式为

$$S=S_1S_2+S_2S_2+S_2S_3+S_2+S_3 \quad (14)$$

式中，S_1S_2为输入层到隐含层的连接权值的个数；S_2S_2为承接层到隐含层的连接权值个数；S_2S_3为隐含层到输出层的连接权值个数；S_2为隐含层阈值个数；S_3为输出层阈值个数。

MEA算法参数设置：根据所需优化的权值与阈值个数，设置$POP_{size}=100$，$Best_{size}=1$，$Tem_{size}=4$，$S_G=20$。

其次，选取训练集的均方误差的倒数作为各个体与种群的适应度值函数，记为Fitness，其表达式为

$$Fitness=1\Big/\frac{(T_{sim}-T_{test})^2}{n} \quad (15)$$

式中，T_{sim}为预测值；T_{test}为测试集真实值；n为测试集元素的个数。

再次，产生初始种群、优胜子种群和临时子种群，然后进行趋同异化操作，根据适应度值大小，解码最优个体。

最后，以此最优个体作为初始权值和阈值，训练Elman神经网络。

4 算例分析

4.1 数据归一化处理

本例以甘肃某地区电力数据为样本，采用MEA－Elman网络模型对某日当天的负荷进行预测。为了使某些较大的输入能够限定在一定范围内，本文采用式（16）对历史负荷数据进行归一化处理。在完成预测后，采用式（17）去归一化，返回其真实值。

$$y=\frac{(y_{max}-y_{min})\times(x_{max}-x_{min})}{x_{max}-x_{min}}+y_{min} \quad (16)$$

$$x=\frac{(x_{max}-x_{min})\times(y_{max}-y_{min})}{y_{max}-y_{min}}+x_{min} \quad (17)$$

式中，y_{min}、y_{max}为归一化参数，可以自行设定，本文取$y_{min}=0$，$y_{max}=1$；x_{min}、x_{max}分别为样本中的最小值和最大值。

4.2 输入变量的选择

通过主成分分析得到综合气象指标z_1、z_2和z_3，并连同日期属性、历史负荷一起作为MEA－Elman网络模型的输入，预测了甘肃某地区某日24点的负荷，具体输入如图4所示。

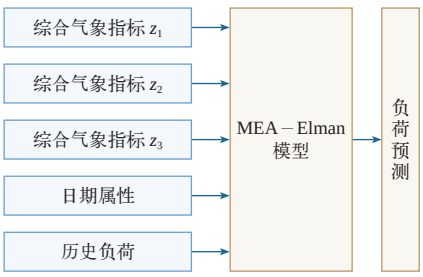

图4 预测模型的输入

4.3 结果分析

以甘肃东部某地区电力数据为样本，利用MEA－Elman网络模型、Elman网络分别对某日当天的负荷进行预测，具体预测结果如表2所示，

表2 预测结果分析

时刻	实际值/MW	MEA－Elman（考虑气象与日期属性）		Elman（考虑气象与日期属性）		MEA－Elman（不考虑气象与日期属性）	
		预测值/MW	APE(%)	预测值/MW	APE(%)	预测值/MW	APE(%)
1：00	714	710.362	0.51	705.655	1.17	702.351	1.63
2：00	686	683.843	0.31	673.735	1.79	693.452	1.09
3：00	686	690.428	0.65	675.92	1.47	670.587	2.25
4：00	658	653.283	0.72	661.843	0.58	634.538	3.57
5：00	645	649.421	0.69	652.872	1.22	637.534	1.16
6：00	640	639.286	0.11	651.396	1.78	621.821	2.84
7：00	604	610.32	1.05	621.953	2.97	619.349	2.54
8：30	628	637.542	1.52	613.841	2.25	612.328	2.50
9：00	660	650.253	1.48	646.452	2.05	642.783	2.61
10：00	696	689.421	0.95	670.534	3.66	675.563	2.94
11：00	732	720.317	1.60	720.186	1.61	710.396	2.95
12：00	717	721.122	0.57	731.552	2.03	720.634	0.51
13：00	753	740.366	1.68	742.635	1.38	730.975	2.92
14：00	734	720.249	1.87	722.846	1.52	729.564	0.60
15：00	718	720.643	0.37	710.472	1.05	720.283	0.32
16：00	708	710.000	0.28	700.548	1.05	710.354	0.33
17：00	727	721.396	0.77	713.582	1.85	707.892	2.63
18：00	747	729.321	2.37	726.000	2.81	726.516	2.74
19：00	740	730.543	1.28	729.473	1.42	725.623	1.94
20：00	749	738.452	1.41	731.862	2.29	723.529	3.40
21：00	714	705.694	1.16	723.941	1.39	712.639	0.19
22：00	701	687.513	1.92	710.243	1.32	692.548	1.21
23：00	697	694.276	0.39	690.473	0.94	673.713	3.34
24：00	703	706.428	0.49	694.716	1.18	689.234	1.96
MAPE(%)		1.01		1.70		2.01	
RMSE		8.359		12.747		15.813	

对比如图5、图6所示。在误差分析过程中，采用在电力系统负荷预测分析中使用较为广泛的APE（Absolute Percentage Error）、MAPE（Mean Absolute Percentage Error）和RMSE（Root Mean Square Error）误差分析方法[12]。其中，APE指绝对百分比误差；MAPE指平均绝对百分误差；RMSE为均方根误差。

$$\mathrm{APE}=\left|\frac{R_e-F}{R_e}\right|\times 100\% \tag{18}$$

$$\mathrm{MAPE}=\frac{1}{N}\left(\sum_{i=1}^{N}\mathrm{APE}_i\right) \tag{19}$$

$$\mathrm{RMSE}=\sqrt{\frac{\sum_{i=1}^{N}(F-R_e)^2}{N}} \tag{20}$$

式中，R_e为真实值；F为预测值；N为预测值个数。

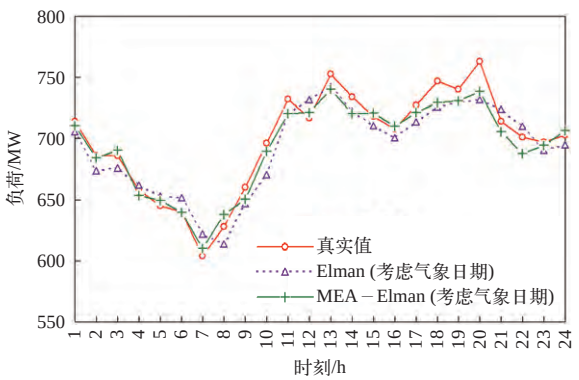

图5 MEA－Elman和Elman模型预测曲线对比图

由表2可知，在考虑气象因素和日期属性时，MEA－Elman模型的MAPE值和RMSE值分别为1.01%和8.359，而Elman模型为1.70%和12.747，

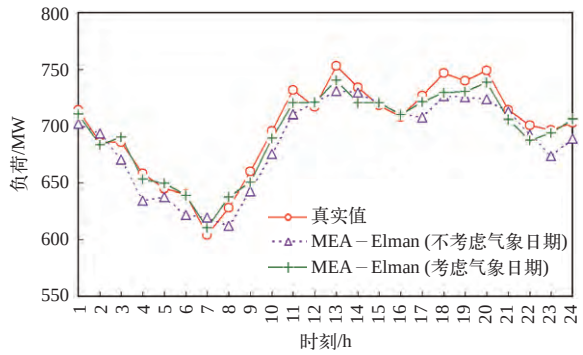

图6 气象因素和日期属性对负荷预测的影响

说明所提优化方法是有效的。而MEA－Elman模型在不考虑气象因素和日期属性时，MAPE值和RMSE值为2.01%和15.813，MAPE误差和RMSE误差增加了1.00%和7.454，表明把气象因素和日期属性加入负荷预测模型中是有必要的。图5、图6进一步表明在考虑气象因素时，MEA－Elman模型预测准确度更高，预测效果更好。

5 结束语

提出了一种计及气象因素和日期属性的短期负荷预测方法，首先利用主成分分析对6种气象因素进行分析得到3种综合气象指标，然后根据各日期类型的相关程度，对日期类型进行量化，最后提出一种基于MEA－Elman网络的短期负荷预测模型，并以甘肃某地区电力负荷数据为样本，对某日的负荷进行了预测。结果表明，所提方法提高了预测准确度，对实现发电容量的合理调度、经济调度及电力市场交易有着重要的实际意义。

参考文献

[1] 何耀耀，许启发，杨善林，等．基于RBF神经网络分位数回归的电力负荷概率密度预测方法 [J]．中国电机工程学报，2013，31 (1)：93-98．

[2] 廖旎焕，胡智宏，马莹莹，等．电力系统短期负荷预测方法综述 [J]．电力系统保护与控制，2011，39 (1)：147-152．

[3] 牛东晓，曹树华，卢建昌，等．电力负荷预测技术及其应用 [M]．北京：中国电力出版社，2009．

[4] 魏少岩，吴俊勇．基于灰色模型和Kalman平滑器的多母线短期负荷预 [J]．电工技术学报，2010，25 (2)：158-162．

[5] 万昆，柳瑞禹．区间时间序列向量自回归模型在短期电力负荷预测中的应用 [J]．电网技术，2012，36 (11)：77-81．

[6] 方鸽飞，胡长洪，郑奕辉，等．考虑夏季气象因素的短期负荷预测研究方法 [J]．电力系统保护与控制，2010，38 (22)：100-104．

[7] Jae Young Choi，Jang Myung Lee．Dynamic PCA algorithm for detecting types of electric poles [J]．Transactions of the Korean Institute of Electrical Engineers，2010，59 (3)：651-656．

[8] 唐勇波，桂卫华，彭涛，等．PCA和KICA特征提取的变压器故障诊断模型 [J]．高电压技术，2014，40 (2)：557-563．

[9] 李根，李文辉．基于思维进化算法的人脸特征点跟踪 [J]．吉林大学学报：工学版，2015，45 (2)：606-612．

[10] 郭红戈．思维进化算法的转移概率分析及几乎处处收敛性证明 [J]．控制与决策，2014，39 (12)：2201-2206．

[11] 李鹏华，柴毅，熊庆宇．量子门Elman神经网络及其梯度扩展的量子反向传播学习算法 [J]．自动化学报，2013，39 (9)：1511-1522．

[12] Hasan Hüseyin Çevik，Mehmet Çunkaş．Short-term load forecasting using fuzzy logic and ANFIS [J]．Neural Computing and Applications，2015，26 (6)：1355-1367．

（收稿日期：2016－05－15）EA

欢迎订阅 欢迎投稿 欢迎刊登广告