

短期负荷预测中气象因素处理的费歇信息方法

蔡舒平¹, 张保会², 汤大海³, 陈燕³, 郝治国²

(1. 江苏大学 电气信息工程学院, 江苏 镇江 212013; 2. 西安交通大学 电气工程学院, 陕西 西安 710049;
3. 国网江苏省电力有限公司镇江供电分公司, 江苏 镇江 212001)

摘要:在能提供大量实时负荷数据和气象数据的智能电网大数据环境下,挖掘合适的气象因素处理方法对提高短期负荷预测精度尤为重要。针对一个或多个气象变量,解决一维或多维费歇信息计算问题。在此基础上,提出基于费歇信息的气象因素建模方法及新预测模型。实际测试结果表明:采用所提模型可以获得更精确的预测结果,解决了短期负荷预测中对气象因素处理的主观随意性问题。

关键词:智能电网;短期负荷预测;气象因素;费歇信息;累积效应;模型

中图分类号:TM 714

文献标志码:A

DOI:10.16081/j.epae.202002027

0 引言

短期负荷预测是电力系统安全经济运行的基础,也是智能电网能源管理系统的重要组成部分,其预测精度受到多种相互依存因素的影响,其中气象因素对其影响巨大。正确处理气象因素对提高短期负荷预测精度意义重大,而预测精度上很小的改进都意味着每年可以节省数百万美元,且对保护环境也有很大作用。与此同时,全球范围内智能电网技术的大规模实施,给负荷预测既带来了机遇也增加了挑战。不同于传统电网中气象数据只能由气象部门提供,智能电网用户端大量引入气象数据采集仪表,可随时获取更加丰富的气象信息,使得短期负荷预测中考虑实时气象信息成为可能。因此,面对智能电网负荷预测中大量的实时数据和丰富的气象信息,研究不同于传统方法的在线预测技术以及挖掘实时气象因素处理方法是提高智能电网负荷预测准确性和快速性的关键。

研究发现,温度是影响短期负荷预测最主要的气象因素。一般而言,温度和气象敏感负荷需求间存在非线性关系。其他气象因素,如湿度、风速,只有在和温度组合时才加以考虑。另外,降雨对负荷需求的影响是通过温度、湿度等体现的。短期负荷预测中既可以直接引入这些气象因素,也可以通过组合其中若干因素以一种气象综合指数的形式间接引入,如温湿指数 THI(Temperature and Humidity Index)^[1]、人体舒适度指数^[2]等,这些气象综合指数

极大地丰富了短期负荷预测中气象因素的处理方法。但上述方法均未考虑温度、湿度等气象变量的累积效应,显然,持续多日高温和单独一日高温对当日负荷需求的影响截然不同^[1]。文献[3]给出2种考虑温度累积效应的修正模型,但需要动态调整累积效应系数。文献[4]采用曲线拟合的方法得到积温效应的量化公式,但持续高温天数与日最高气温对积温效应贡献的比重系数因不同地域而异,并不具有普遍适用性。此外,上述所有综合气象指标都是基于日特征气象因素,如日最高温度、日最低温度、日平均温度等,这些指标在传统电网负荷预测中发挥了重要作用,但在能够提供实时气象信息的智能电网环境下,仅考虑日特征气象因素显然无法得到更加精细化的预测结果。文献[5]给出一种考虑实时气象因素的神经网络模型,但需要增加大量的输入层变量,这增加了网络的训练时间和预测时间。同时,这些气象综合指数在引入短期预测模型时也较少考虑气象因素的累积效应。文献[1]给出一种考虑温度和湿度累积效应的加权温湿指数(WTHI),但需要根据不同地区的负荷预测重新确定加权系数。因此,探寻智能电网环境下短期负荷预测中实时气象因素的建模方法愈显重要和迫切。

近年来,信息理论在短期负荷预测中的应用日渐广泛,这是因为本质上任何类型的模型和数据都是对客观事物的一种信息表征而与它们的学科来源无关^[6-7]。另外,智能电网中的气象数据采集仪表所提供的丰富的气象数据信息,为深入探究气象因素对短期负荷预测的影响提供了极大的便利。因此,问题的关键是如何利用这些气象数据信息,寻找一种新的途径来更好地描述气象因素对电力系统短期负荷预测的影响。

费歇信息 FI(Fisher Information)提供了表征气象因素这类特征变量的可能性。不同于系统信息的其他度量方法,费歇信息通过监测描述系统的数据

收稿日期:2019-04-26;修回日期:2019-12-29

基金项目:国家自然科学基金资助项目(61301138,51975260);
江苏高校优势学科建设工程资助课题(批准号:苏政办发(2011)6)

Project supported by the National Natural Science Foundation of China(61301138, 51975260) and the Priority Academic Program Development of Jiangsu Higher Education Institutions((2011)6)

监测系统变化过程^[8]。本文采用费歇信息理论解决智能电网短期负荷预测中气象因素的建模问题,并结合实际气象数据和负荷数据验证其有效性。

1 费歇信息理论及其计算

1.1 费歇信息

系统状态变量的行为决定了系统稳定性^[6]。因此,对于一个完全稳定的系统,在容许的测量误差范围内,一个状态变量随时间重复观测会产生相同的值,概率密度函数(PDF)对该变量急剧倾斜,其对该变量的导数接近无穷大;对于一个相当不稳定的系统,则情况恰好相反,对变量的所有观测会产生完全不同和不相关的值,观测到任何一个值的概率均等,概率密度函数是扁平的,其对该状态变量的导数接近于0^[9]。据此,统计学家罗纳德·费歇推导出另一种基于概率密度函数导数的度量稳定性的方法——费歇信息,其定义式如下^[10]:

$$I = \int \frac{ds}{P(s)} \left(\frac{dP(s)}{ds} \right)^2 \quad (1)$$

其中, s 为一系统状态变量; $P(s)$ 为该变量的一种概率密度分布函数。

稳定性的概念在许多不同学科中均有涉及,不同学科有不同的表述和衡量标准。费歇信息是从系统信息的角度给出一种定量度量系统稳定性的方法。某事件越稳定,所提供的关于该事件的信息量越大,费歇信息值就越大。由于实际中绝大多数系统常位于非常稳定或非常不稳定(分别对应于费歇信息值为最大和0)这2种极端情况之间。因此,理论上费歇信息是对系统稳定性的一种很好的测量。

为计算方便,进行简单的代换,即令 $q^2(s)=P(s)$,则式(1)变为:

$$I = 4 \int ds \left(\frac{dq(s)}{ds} \right)^2 \quad (2)$$

实际中,绝大多数情况下给出的是系统的观测数据序列,因此式(2)中的微分以差分代替,而积分以求和代替,即分别用 $\Delta s = s_i - s_{i+1}$ 和 $\Delta q = q_i - q_{i+1}$ 代替 ds 和 dq ,其中 s_i 为系统状态标号,表示第 i 个状态, q_i 为该状态处的一个过渡变量。因此相邻2个状态标号之差等于1,即 $s_i - s_{i+1} = 1$ 。从而费歇信息近似为:

$$I \approx 4 \sum_{i=1}^n \left(\frac{q_i - q_{i+1}}{s_i - s_{i+1}} \right)^2 (s_i - s_{i+1}) \quad (3)$$

其中, n 为状态数。

则计算费歇信息的最终表达式为^[11]:

$$I \approx 4 \sum_{i=1}^n (q_i - q_{i+1})^2 \quad (4)$$

1.2 单变量费歇信息计算

单变量费歇信息的计算采用时间窗方法,即在

所给单变量数据集上建立一个滑动的数据窗口,窗口大小取决于整个数据集的大小。按照经验,窗口中所含数据的数量不应少于8个,以防止窗口中的各点过度影响整个计算^[11]。窗口滑动因子的取值应不超过窗口宽度,以使相邻窗口间产生重叠,这样做是为检测到可能超出该窗口边界外的系统稳定情况。

假设任一观察数据序列集为 $D = \{d(k), k = 1, 2, \dots, N\}$,其中 N 为序列总长度,则在该序列集上定义滑动窗 W 为:

$$W(m, w, \delta) = \{d(k), k = 1 + m\delta, \dots, w + m\delta\} \quad (5)$$

其中, w 为窗宽, δ 为滑动因子,均为不大于 N 的正整数; $m = 1, 2, \dots, M, M = (N-w)/\delta$ 为窗口数。滑动窗口示意图见图1,图中 w_M 表示窗宽为 w 的第 M 个窗口。

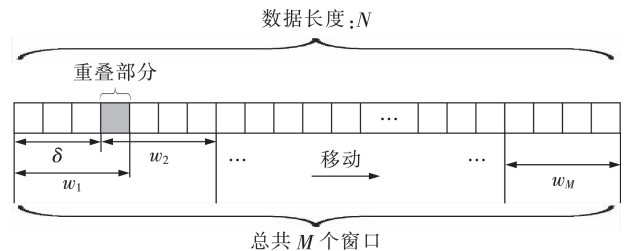


图1 滑动窗口示意图

Fig.1 Schematic diagram of sliding window

将滑动窗 W 划分为 I 个区间:

$$W(m, w, \delta) = \bigcup_{i=1}^I z_i \quad (6)$$

其中, $z_i = \{(s_{i-1}, s_i), i = 1, 2, \dots, I\}$,区间互不相交。

$$s_0 < s_1 < s_2 < \dots < s_I \quad (7)$$

$$s_0 = \min W(m, w, \delta) =$$

$$\min \{d(k), k = 1 + m\delta, \dots, w + m\delta\} \quad (8)$$

$$s_I = \max W(m, w, \delta) =$$

$$\max \{d(k), k = 1 + m\delta, \dots, w + m\delta\} \quad (9)$$

式(8)、(9)分别用于求取滑动数据窗 W 中所含数据的最小值和最大值。

因此, $d(k) \in W(m, w, \delta)$ 位于区间 z_i 的概率 $P(z_i)$ 为 $d(k) \in W(m, w, \delta)$ 落入区间 z_i 的数量与 $W(m, w, \delta)$ 中总数据个数 w 的比值,即:

$$P(z_i) = \frac{\text{落入区间} z_i \text{的数据个数}}{\text{时间窗口中的总数据个数}} \quad (10)$$

利用式(10)即可计算对应各状态的 q_i ,从而可采用式(4)为各时间窗口估计出费歇信息,并将估计出的费歇信息值分配给各时间窗口的中间位置。从上述计算过程可看出,数据分组方法采用组距分组法,即将用来计算费歇信息的一个窗口的数据划分为若干个区间。按照经验,一组数据所分组数范围为 $5 < I < 15$,本文取值为9。以此为基础即可确定组距,即组距=全部数据的最大值和最小值之差/组数。

计算单变量费歇信息的流程见附录A中图A1。

1.3 多变量费歇信息计算

设由 M 个变量定义的 M 维状态空间中有一长度为 N 的时间序列数据集 $\{(X(t_1), X(t_2), \dots, X(t_i), \dots, X(t_N))\}$, 其中 $X(t_i) = (x_1(t_i), x_2(t_i), \dots, x_M(t_i))$ 。

类似上述单变量数据集滑动窗的定义方法, 在多维数据集上定义一系列相互重叠的窗口序列。假设各数据窗描述 M 维状态空间中系统的若干个状态。并且若 $|X(t_i) - X(t_j)| \leq \Delta X$ (ΔX 为容许的测量误差向量), 则认为 $X(t_i)$ 和 $X(t_j)$ 属于同一个状态(同一值的2次观测)。 ΔX 由以下Chebyshev定理确定。

Chebyshev定理: 设随机变量 X 具有数学期望 $E(X) = \mu$ 、方差 $D(X) = \sigma^2$, 则对任意正数 ε , 不等式 $P\{|X - \mu| \geq \varepsilon\} \leq \frac{\sigma^2}{\varepsilon^2}$ 或 $P\{|X - \mu| < \varepsilon\} \geq \frac{\sigma^2}{\varepsilon^2}$ 成立。

该定理表明, 任意一个数据集中, 位于其平均数 g (g 为任意大于1的正整数)倍标准差范围内的比例(或部分)至少为 $1/g^2 \sim 1$ 。由于本文所研究的气象变量均满足 $E(X)$ 和 $D(X)$ 存在且有限这一条件, 故若取 $\Delta X = 2\rho$ (ρ 为上述数据窗的标准差), 则按照该定理, 上述数据窗中至少有75%的数据位于平均数2倍标准差范围内(即落在2倍标准差范围内的观测比例至少是75%)而不管其概率分布形式如何^[12]。

考虑到上述数据窗中每一个数据点为一个 M 维向量, 因此以 $2\rho_1, 2\rho_2, \dots, 2\rho_M$ ($\rho_1, \rho_2, \dots, \rho_M$ 为各维数据的标准差)为 M 个边长来构建 M 维超矩形。具体而言, 在第一个数据窗内, 按时间顺序选取第一个点作为第一个状态的中心, 围绕该点构建一个状态超矩形。清点落入超矩形中所有点的数量 Z_1 , 即:

$$Z_1 = \left\{ Z_1: \left(|x_1(t_1) - x_1(t_j)| \leq 2\rho_1 \right) \& \right. \\ \left. \left(|x_2(t_1) - x_2(t_j)| \leq 2\rho_2 \right) \& \dots \& \right. \\ \left. \left(|x_M(t_1) - x_M(t_j)| \leq 2\rho_M \right) \right\} \quad j = 2, 3, \dots, N$$

得到由 Z_1 个数据组成的第一个状态。在该数据窗内选择下一个未被计入的点, 围绕该点构建另一个 M 维超矩形, 同理清点落入该超矩形中数据的数量 Z_2 , 得到由 Z_2 个数据组成的第二个状态。依此类推, 在数据窗内重复该过程直到将所有点计入不同的状态。由于下一个 M 维超矩形数据清点计入过程是在前面所有 M 维超矩形清点装入完剩下的数据基础上进行的, 这就保证了每个超矩形中所包含数据的唯一性。在上述窗口中的所有数据完成装箱后, 该过程移入下一个数据窗。

设数据窗中所有元素可归为 L 个状态, 则有:

$$N_w = \sum_{h=1}^L Z_h \quad (11)$$

其中, N_w 为数据窗中元素总数, 其等于各个状态中元素数量之和。

由此可得计算多变量费歇信息的基本流程如附录A中图A2所示。

2 预测模型中气象因素的费歇信息建模方法

在所有影响用电负荷的气象因素中, 温度起着最为关键的作用。盛夏高温天气时, 用电负荷随着温度的升高而急剧升高, 尤其是持续多日高温天气还会对电力负荷产生累积效应, 而其他气象因素也有类似现象。气象因素的累积效应会严重影响负荷预测的精度, 因此, 进行负荷预测时要考虑累积效应的影响^[4]。费歇信息是对系统稳定性的一种度量, 其能定量地反映之前若干时段内气象因素变化的稳定程度, 用它对当前气象因素进行加权(即相乘)就能很好地刻画这种累积效应, 这与人们对气象变化的日常感受相一致。以下就2种情况分别展开讨论。

2.1 单一气象因素的费歇信息建模

文献[3]考虑气象敏感负荷与各时段实时温度因素, 以午高峰为例研究了温度累积效应对电力负荷的影响机理。研究发现, 温度对负荷需求所产生的累积效应由当日内累积效应和多日累积效应两部分组成^[3]。当日内累积效应指的是当日内以前各时段的温度对当日当前负荷产生的影响, 这种影响主要来源于前一时段、前两时段的温度, 而前三时段的温度对此影响甚微^[3](在这些研究中1个时段为1h)。将上述以1h为1个时段的当前时段、前一时段、前两时段换算成本文中以15min为1个时段来计算, 则总共有 $4 \times 3 = 12$ 个时段, 取这12个时段中的气象变量数据组成计算积温效应的费歇信息数据窗中当日累积效应的第一部分数据。多日温度累积效应指的是负荷出现一定程度的异常增长与持续多日低温或高温天气状况有关^[3]。笔者以某地夏季持续高温日期间的几个午高峰负荷为例进行研究发现, 某日午高峰负荷的持续上升与前两日午高峰温度的稳定增加有关, 而与再往前日期温度的变化情况相关性不大^[3], 其他气象因素对负荷的影响具有类似的情况。考虑到负荷预测中应对距离预测点较近的影响因素给予更多的关注, 即“近大远小”取数原则, 因此, 仅考虑预测前一日同一预测时段处当前时段、前一时段的温度影响, 以及预测前两日同一预测时段处当前时段的温度影响。采用与上述类似的时段换算方法即可分别得到计算积温效应的费歇信息数据窗中前一日累积效应的第二部分8个数据和前两日累积效应的第三部分4个数据。这样可用以下3

个部分的数据组合来构成计算单变量气象因素的单次费歇信息,并用其对当前气象因素进行加权即可充分反映气象变量的当日累积效应和多日累积效应(以采样间隔取15 min为例):

$$\left\{ \begin{aligned} &(X_{d,t} \ X_{d,t-1} \ \cdots \ X_{d,t-11}), \\ &(X_{d-1,t} \ X_{d-1,t-1} \ \cdots \ X_{d-1,t-7}), \\ &(X_{d-2,t} \ X_{d-2,t-1} \ \cdots \ X_{d-2,t-3}) \end{aligned} \right\}$$

其中, d 表示预测当日; t 为当前预测时段。式中第一部分数据由预测当日当前时段3 h内的12个气象数据采集点组成,第二部分数据由预测前一日当前时段2 h内的8个气象数据采集点组成,第三部分数据由预测前二日当前时段1 h内的4个气象数据采集点组成。即计算一次费歇信息的数据窗由总共24个数据点组成,符合负荷预测数据“近大远小”的取数原则。由上述可知,该数据选取方法在现有研究成果^[3-4]的基础上,兼顾了负荷预测数据的取数原则,使得气象因素对负荷的累积效应得以充分体现。

为进一步检验上述计算费歇信息气象因素累积效应数据窗组成方法的有效性,附录A中图A3给出了2016年7月1日至8月31日某地区日最大负荷对日最高温度按上述方法进行加权处理前后的散点图,可以看出,处理后两者的关系更紧密,接近于线性变化,相关系数由原来的0.8525上升至0.9243。

2.2 气象综合指数的费歇信息建模

研究表明,相较于影响负荷需求的温度、湿度等单一因素,气象综合指数能更有效地跟踪和描述负荷的变化^[1,13],这是由于人体的真实感觉无法通过单个气象指标进行准确描述,其是在温度、湿度及风速等多种气象因素的相互耦合作用下形成的,因此学者提出多种考虑多个气象因素的综合气象指数模型,如实感温度、温湿指数、风冷力指数、人体舒适度等,但这些综合气象指数均是基于日特征气象因素,对负荷的影响并不具有实时的效果。

气象综合指数通常定义为多个单一气象变量的函数以反映其耦合效果,该效果与各气象变量在多维空间中变化的稳定性有着必然的联系,而这种稳定性可以通过多变量费歇信息来计算。此外,类似于单一气象因素,气象综合指数同样具有累积效应,这使得人体的反应往往滞后于外部天气的变化。因此,在预测模型中引入气象综合指数时必须考虑这种效应,而该效应可通过多变量费歇信息对综合指数的加权来实现。由此可见,气象综合指数的费歇信息处理方法很好地体现了气象综合指数的耦合效应和累积效应。类似于计算单一气象因素费歇信息中数据窗数据的构成方法,计算多变量费歇信息的数据窗数据组成如下(取采样间隔为15 min):

$$\left\{ \begin{aligned} &(X_{d,t} \ X_{d,t-1} \ \cdots \ X_{d,t-11}), \\ &(X_{d-1,t} \ X_{d-1,t-1} \ \cdots \ X_{d-1,t-7}), \\ &(X_{d-2,t} \ X_{d-2,t-1} \ \cdots \ X_{d-2,t-3}) \end{aligned} \right\}$$

其中, $X_{d,t}=(x_{d,t}(1), x_{d,t}(2), \dots, x_{d,t}(M))$ 。

3 气象因素的费歇信息建模方法及实例分析

相比于短期负荷预测中的概率模型^[14],本文采用具有更好泛化性能和较高预测精度的支持向量机(SVM)模型来验证气象因素的费歇信息建模方法在智能电网短期负荷预测中的有效性。以下就3种情况分别展开讨论。

3.1 SVM预测模型中气象因素直接输入法

研究发现,当气象因素发生剧烈变化时,对应时段的负荷也会相应改变,即气象因素对负荷的影响具有实时效应^[5]。另一方面,实时气象因素与负荷的变化之间存在“时差”,即气象因素变化在前,负荷变化在后。结合负荷变化固有的周期性特点,一个气象变量未经费歇信息加权处理而直接输入的每日96点SVM预测模型结构由三部分组成:1个隐含层;只有1个神经元,即待预测日某时刻负荷值的输出层;由19个神经元组成的输入层,其输入变量组成见附录B中表B1。附录B表B1中输入变量选取前一、四时段而忽略前二、三时段,这是基于已有研究成果,并结合大量的仿真实验验证而确定的。研究发现,当取前一、四时段的气象因素和相应时段中的负荷值时,能够取得更加精确的预测结果,也能更准确地反映气象因素对预测负荷的影响^[1,3,5]。

将附录B表B1中所列出的19个输入神经元作为SVM预测模型的输入变量来考察实时气象因素对负荷预测效果的影响,本文将其记为模型1,并与下述基于费歇信息的气象因素建模方法进行对比。

3.2 SVM预测模型中气象因素输入的费歇信息法

在上述模型的基础上,对其中的温度和湿度气象变量采用基于费歇信息的处理策略,则输入层变量减为11个,见附录B中表B2。本文将以这种方法建立的模型记为模型2。模型中,负荷变化对气象因素的累积效应和滞后效应通过费歇信息对当日气象因素加权得以体现,同时减少了输入神经元的数量,简化了模型,缩短了训练时间,加快了预测速度。

3.3 SVM预测模型中综合气象指数输入的费歇信息方法

进一步,在模型2的基础上引入温湿指数^[1]:

$$I_{TH} = T_F - (0.55 - 0.55H)(T_F - 58) \quad (12)$$

其中, T_F 为时段华氏温度, H 为时段湿度,对它们进行费歇信息加权处理,则输入层的变量减为9个,如

附录B中表B3所示。本文将对应这种情况所建立的模型记为模型3。在该模型中,通过费歇信息对当日综合气象指数加权不但很好地体现了它们对负荷产生的累积效应,还抵消了综合气象指数与负荷变化之间存在的“时差”,使得负荷实时跟随综合气象指数的变化而变化,此外模型中输入神经元的数量进一步减少,缩减了训练时间,加快了预测速度。

3.4 实例分析

本文选取江苏省某市2016年7月1日至8月31日共计62 d的日负荷数据和气象数据作为SVM的训练样本,训练样本和测试样本按3:1的比例进行划分,兼顾预测模型的性能退化问题,实际应用中以3 d的数据进行训练,以接下来1 d的数据进行预测,如此循环向前推进,这就出现了测试样本包含在训练样本中的情况。本文选取8月12日负荷数据为测试样本。所收集到的负荷数据集和气象数据集的采样间隔分别为每15 min一个点和每1 h一个点。为了和负荷数据集保持同样的采样间隔,将气象数据集(含温度、湿度和风速)进行插值,对插值处理后的气象数据集再按上述计算单变量气象因素费歇信息数据窗组成方法形成窗口数据。为了更有效地捕捉单一气象因素的变化,取滑动因子 $\delta=1$ 。对2个气象因素(此处为温度、湿度),窗口内的每个数据点为由温度和湿度所组成的二维数据,采用多变量费歇信息的计算方法构造一个二维的长方形,其2个边长分别为二维窗口数据内温度、湿度分量标准差的2倍,据此可算出二维变量的温湿指数费歇信息。

本文实例研究所用计算机主频为2.5 GHz,内存为12 GB。依次对模型1—3进行测试,结果见附录A中图A4。图中,曲线的两端(分别对应横轴1~20, 70~96)分别对应于一天中的凌晨和傍晚时段,该时段负荷的积温效应较明显,3个模型对负荷的变化均表现出较好的拟合效果。曲线的中间(对应横轴30~60)对应上午和午高峰时段,该时段温度迅速攀升,气象敏感负荷增加较快,相较于模型1和2,模型3更好地展现出了对高峰负荷的实时跟踪效果。

为了评价各个模型的预测效果,本文采用平均绝对百分比误差MAPE(Mean Absolute Percentage Error)、平均误差ME(Mean Error)、均方根误差RMSE(Root Mean Square Error)作为评判的依据,分别记为 e_{MAPE} 、 e_{ME} 、 e_{RMSE} 。从附录A中图A4中各个模型每15 min的相对误差及相应的96点 e_{MAPE} 来看,模型3最好,模型2次之,模型1稍差。

表1给出了3个模型在该天整点时刻的预测误差及单次平均预测时间 $\bar{t}$ 。

由表1结果知,模型2的 e_{MAPE} 、 e_{ME} 、 e_{RMSE} 比模型1分别降低了0.76%、34.78 MW和1.36%,这表明相较于模型1,模型2对气象敏感负荷的预测有一定的优

表1 3个模型预测结果

Table 1 Forecasting results of three models

模型	$e_{\text{MAPE}} / \%$	$e_{\text{ME}} / \text{MW}$	$e_{\text{RMSE}} / \%$	$\bar{t} / \text{s}$
1	2.29	78.82	2.11	56.94
2	1.53	44.04	0.75	42.09
3	1.22	34.95	0.13	34.37

势,这是由于模型2中采用费歇信息建模方法来处理气象因素的累积效应,避免了模型1中气象因素的直接输入而导致对该类问题处理的不足。同时,模型2还解决了负荷变化和气象因素间的“时差”问题,使负荷随气象因素的变化表现出实时的效果,这从附录A中图A4(b)可看出。此外,由于气象因素经费歇信息处理后输入层变量减少,这使得SVM训练时间缩短,单次平均预测时间比模型1缩短了14.85 s。

模型3的 e_{MAPE} 、 e_{ME} 、 e_{RMSE} 在模型2的基础上又分别降低了0.31%、9.09 MW和0.62%,这说明模型3的预测精度得到了进一步的提高。和模型2不同的是,模型3采用气象综合指数体现气象因素对负荷的影响,由于它能更准确地描述人体的真实感觉进而影响人们的用电行为,因而能更好地跟踪负荷的变化,同样,由于使用费歇信息方法解决累积效应和滞后效应,模型3具有更好的拟合效果和预测精度。此外,模型3采用综合指数后输入层变量进一步减少,训练时间比模型2又有了进一步的缩短,单次平均预测时间比模型1和2分别缩短了22.57、7.72 s。

4 结论

鉴于传统气象因素处理方法在短期负荷预测中的不足,并考虑到智能电网实时数据信息的特点,本文提出基于费歇信息的气象因素建模策略,并将其应用于实际的智能电网短期负荷预测。

预测结果表明,无论是对原始气象变量还是对气象综合指数,该方法不但很好地解决了气象因素的累积效应,提高了负荷预测精度,还消除了负荷随气象因素变化的滞后效应。更重要的是,采用气象因素的费歇信息建模方法大幅减少了智能预测模型输入层变量,缩短了训练时间,提高了预测速度,即本文仅通过对预测模型中气象输入变量进行优化处理就以最少的输入资源获得了最大的预测性价比。

附录见本刊网络版(<http://www.epae.cn>)。

参考文献:

- [1] 康重庆,夏清,刘梅. 电力系统负荷预测[M]. 北京:中国电力出版社,2017:234-238,241-244.
- [2] 张贲,邵常宁,赵燃. 基于误差修正的短期负荷预测方法[J]. 电力自动化设备,2015,35(11):152-157.
ZHANG Ben, SHAO Changning, ZHAO Ran. Short-term load forecasting based on deviation correction[J]. Electric Power

- Automation Equipment, 2015, 35(11):152-157.
- [3] 方鹏飞, 胡长洪, 郑奕辉, 等. 考虑夏季气象因素的短期负荷预测方法研究[J]. 电力系统保护与控制, 2010, 38(22):100-104. FANG Gefei, HU Changhong, ZHENG Yihui, et al. Study on the method of short-term load forecasting considering summer weather factors[J]. Power System Protection and Control, 2010, 38(22):100-104.
- [4] 蒋建东, 程志豪, 朱明嘉. 考虑积温效应的短期负荷组合预测方法[J]. 电力自动化设备, 2011, 31(10):28-31. JIANG Jiandong, CHENG Zhihao, ZHU Mingjia. Combined short-term load forecast with accumulated temperature effect[J]. Electric Power Automation Equipment, 2011, 31(10):28-31.
- [5] 康重庆, 周安石, 王鹏, 等. 短期负荷预测中实时气象因素的影响分析及其处理策略[J]. 电网技术, 2006, 30(7):5-10. KANG Chongqing, ZHOU Anshi, WANG Peng, et al. Impact analysis of hourly weather factors in short-term load forecasting and its processing strategy[J]. Power System Technology, 2006, 30(7):5-10.
- [6] CABEZAS H, FATH B D. Towards a theory of sustainable systems[J]. Fluid Phase Equilibria, 2002(197):3-14.
- [7] 陈丽丹, 张尧, Antonio Figueiredo. 融合多源信息的电动汽车充电负荷预测及其对配电网的影响[J]. 电力自动化设备, 2018, 38(12):1-10. CHEN Lidan, ZHANG Yao, ANTONIO Figueiredo. Charging load forecasting of electric vehicles based on multi-source information fusion and its influence on distribution network[J]. Electric Power Automation Equipment, 2018, 38(12):1-10.
- [8] FATH B D, CABEZAS H, PAWLOWSKI C W. Regime changes in ecological systems: an information theory approach[J]. Journal of Theoretical Biology, 2003, 222(4):517-530.
- [9] KARUNANITHI A T, CABEZAS H, FRIEDEN B R, et al. Detection and assessment of ecosystem regime shifts from Fisher information[J]. Ecology and Society, 2008, 13(1):1-15.
- [10] FISHER R A. On the mathematical foundations of theoretical statistics[J]. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 1922, 222(594):309-368.
- [11] EASON T, CABEZAS H. Evaluating the sustainability of a regional system using Fisher information in the San Luis Basin, Colorado[J]. Journal of Environmental Management, 2012, 94(1):41-49.
- [12] LAPIN L. Statistics: meaning and method[M]. New York, USA: Harcourt Brace Jovanovich, 1975:11-37.
- [13] 王雁凌, 吴梦凯. 经济新常态下基于偏最小二乘回归的中长期负荷预测模型[J]. 电力自动化设备, 2018, 38(3):133-139. WANG Yanling, WU Mengkai. Medium and long term load forecasting model based on partial least-square regression under new normal economy[J]. Electric Power Automation Equipment, 2018, 38(3):133-139.
- [14] 王浩林, 张勇军, 毛海鹏. 基于时刻充电概率的电动汽车充电负荷预测方法[J]. 电力自动化设备, 2019, 39(3):207-213. WANG Haolin, ZHANG Yongjun, MAO Haipeng. Charging load forecasting method based on instantaneous charging probability for electric vehicles[J]. Electric Power Automation Equipment, 2019, 39(3):207-213.
- [15] MOGHRAH I, RAHMAN S. Analysis and evaluation of five short-term load forecasting techniques[J]. IEEE Transactions on Power Systems, 1989, 4(4):1484-1491.

作者简介:



蔡舒平

蔡舒平(1963—),男,甘肃天水人,副教授,主要研究方向为电力负荷预测、智能计算在电力系统中的应用等(E-mail: spcai@ujs.edu.cn);

张保会(1953—),男,河北魏县人,教授,博士研究生导师,博士,研究方向为电力系统安全稳定控制系统理论及装置、电力系统新型继电保护以及电力系统通信等(E-mail: bhzhang@mail.xjtu.edu.cn);

汤大海(1963—),男,江苏镇江人,研究员级高级工程师,长期从事电网运行、继电保护等方面的工作(E-mail: 13921598622@163.com)。

(编辑 王锦秀)

Fisher information method for processing weather factors in short-term load forecasting

CAI Shuping¹, ZHANG Baohui², TANG Dahai³, CHEN Yan³, HAO Zhiguo²

(1. College of Electrical and Information Engineering, Jiangsu University, Zhenjiang 212013, China;

2. School of Electrical Engineering, Xi'an Jiaotong University, Xi'an 710049, China;

3. Zhenjiang Power Supply Company of State Grid Jiangsu Electric Power Co., Ltd., Zhenjiang 212001, China)

Abstract: In the big data environment of smart grid that can afford massive real-time load data and weather data, it is especially important to find appropriate weather factor processing method for improving the accuracy of short-term load forecasting. Aiming at one or multiple weather variables, the problems of one or multi-dimension Fisher information calculation are solved. On this basis, the modeling method of weather factors based on Fisher information and the new forecasting model are proposed. The practical test results show that the proposed model can obtain more accurate forecasting results, which solves the subjective randomness problem in dealing with weather factors of short-term load forecasting.

Key words: smart grid; short-term load forecasting; weather factor; Fisher information; accumulated effect; models

附录 A:

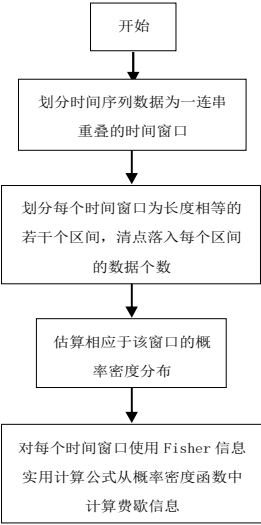


图 A1 单变量费歇信息计算流程图
Fig.A1 FI calculation flowchart of single variable

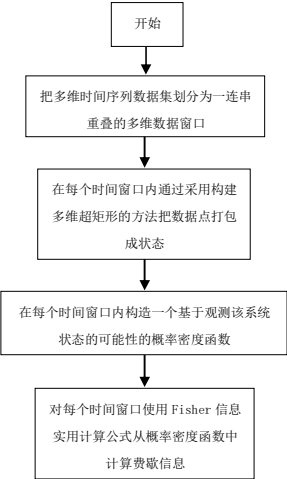
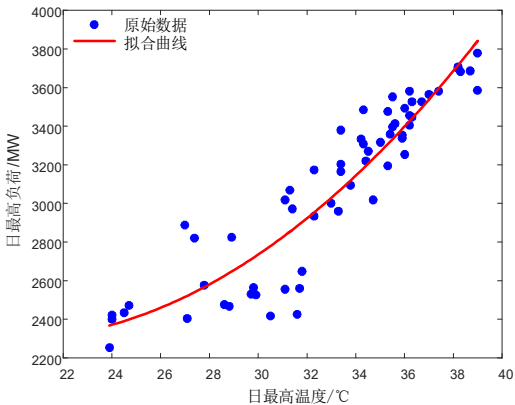
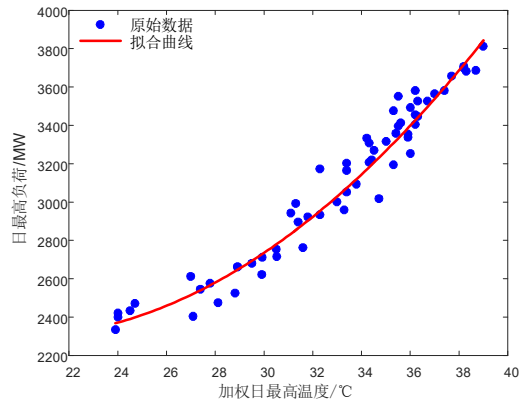


图 A2 多变量费歇信息计算流程图
Fig.A2 FI calculation flowchart of multiple variables



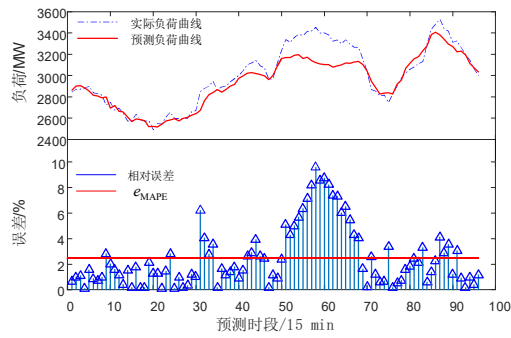
(a) 日最高负荷对日最高温度散点图



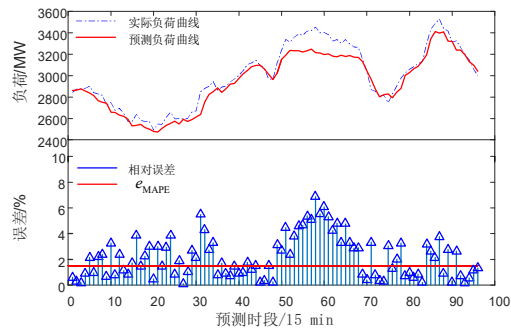
(b) 日最高负荷对加权日最高温度散点图

图 A3 日最高负荷对日最高温度处理前后散点图

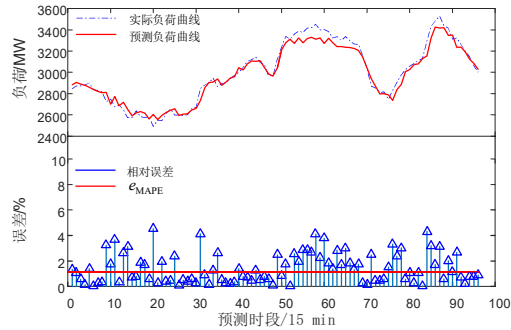
Fig.A3 Scatter plot for daily maximum load vs. daily maximum temperature before and after dealing



(a)模型1



(b)模型 2



(c)模型 3

图 A4 负荷预测结果及其误差曲线

Fig.A4 Forecasting results ofload and corresponding error curves

附录 B:

表 B1 气象因素直接输入时预测模型输入量组成表

Table B1 Direct input climate variables of foresting model

输入量	描述
1、2	对应预测日星期类型及当前预测时刻
3、4	对应预测日预测时刻的前一时刻、 前四时刻的负荷值
5—10	对应预测日预测时刻、前一时刻、 前四时刻的温度、湿度
11—13	对应上周同类型日预测时刻、 前一时刻、前四时刻的负荷值
14—19	对应上周同类型日预测时刻、前一时刻、 前四时刻的温度、湿度

表 B2 气象因素经费歇信息处理后预测模型输入量组成

Table B2 Input climate variables of foresting model after dealing by FI

输入量	描述
1、2	对应预测日星期类型及当前预测时刻
3、4	对应预测日当前预测时刻的前一时刻、 前四时刻的负荷值
5、6	对应预测日当前预测时刻温度、 湿度的费歇信息加权值
7—9	对应上周同类型日当前预测时刻、 前一时刻、前四时刻的负荷值
10、11	对应上周同类型日当前预测时刻温度、 湿度的费歇信息加权值

表 B3 温湿指数经费歇信息处理后预测模型输入量组成

Table B3 Input THIo of foresting model after dealing by FI

输入量	描述
1、2	对应预测日星期类型及当前预测时刻
3、4	对应预测日当前预测时刻的前一时刻、 前四时刻的负荷值
5	对应预测日当前预测时刻 温湿指数的费歇信息加权值
6—8	对应上周同类型日当前预测时刻、 前一时刻、前四时刻的负荷值
9	对应上周同类型日当前预测时刻 温湿指数的费歇信息加权值