

一种新的天然气管网负荷预测方法

陈学锋¹ 袁宗明² 马永驰²

(1. 中国石油西南油气田分公司长输管道检测评价中心 2. 西南石油大学)

陈学锋等. 一种新的天然气管网负荷预测方法. 天然气工业, 2006, 26(3): 114-116.

摘 要 对管网负荷的准确预测是保证管网经济安全运行的前提, 在总结前人大量的管网负荷预测方法的基础上, 提出了基于新鲜度函数的模糊变权重组合预测管网负荷的方法。该方法是单模型预测技术的自然推广, 通过将新鲜度函数引入组合预测, 可有效地刻画时间序列的变化趋势, 从而实现了从模糊自适应变权重组合预测方法的有效改进。改进后的组合预测方法不仅综合利用了各种预测方法的有用信息, 而且该最佳组合预测的权系数也是时变的, 是一种进一步提高预测精度的方法, 具有预测精度高、计算简便的特点。

主题词 天然气 管网 负荷 组合 预测 变权重 方法 精度

管网负荷预测作为天然气日常调度的主要依据, 其精确程度将直接关系到能否保证用户天然气设备所需的用气压力、天然气运营企业储气设施的供气量与储气量的平衡以及高峰供气和低谷进气调节措施的实现, 从而直接影响着管网运行的费用和企业经济效益。目前存在的管网负荷组合预测方法中, 绝大多数都局限于定权重组合预测范畴, 即在组合模型中通常是将两种方法并联, 取简单的加权平均。这对于管网负荷这类复杂的动态系统的预测而言, 定权重组合预测所能提高的预测精度是很有限的。鉴于此, 笔者在对国内某燃气公司的管网负荷预测中, 重点研究了变权重的优化组合预测方法, 并首次将新鲜度函数引入到组合预测中, 提出了新的变权重组合预测算法, 并应用于该燃气公司的实际预测中, 取得了很好的效果。

一、组合预测及其国内发展现状

优化组合预测方法是当前预测科学研究最热门的课题之一。组合预测就是综合利用各种预测方法所提供的信息, 以适当的加权平均形式得出组合预测模型。组合预测方法因能有效地提高预测精度, 而受到国内外预测工作者的重视。组合预测中最关键的问题就是如何求出加权平均系数, 使得组合预测模型更加有效地提高预测精度。

对组合预测权系数确定的方法已有不少研究成果, 但在变权重组合预测方面则进展缓慢。显然, 变权重的方法比不变权重的方法更科学。因为对每一种单项预测方法而言, 它总是表现出“时好时坏”性, 但是, 由于变权重的组合预测方法比较复杂, 所以目前关于变权重的组合预测方法并不多见。文献[1]利用多项式来逼近变权重的连续函数, 并通过广义逆矩阵的循环迭代形成收敛的权系数来进行组合预测, 这不失为一种方法, 但该方法涉及大量矩阵求逆等繁琐运算, 从而妨碍了实际应用。文献[2, 3]提出了两种模糊变权重组合预测方法, 但文献[2]所提出的变权重方法, 其当前时刻的权重仅取决于上一时刻的误差及误差的变化率, 所以这种方法有一定的局限性。文献[3]对此作了改进, 其当前时刻的权重是根据最近一段时间内预测误差绝对平均值的相对指标和滚动的有限时域长度内预测误差绝对累计值的相对指标来决定的, 所以后者比前者更科学, 其仿真结果也证明了这一点。但该方法缺陷在于没有考虑时间上的先后关系, 换句话说, 那就是没有考虑样本的变化趋势, 如果两个样本序列的绝对误差平均值相等, 一个是误差逐渐变小, 一个是误差逐渐变大, 按文献[3]的方法, 所取的权重就应该是一样的, 这显然是不符合实际的。因此, 引入新鲜度函数, 并对变权重的组合预测方法作了进一步的探索。

作者简介 陈学锋, 1974年生, 2005年毕业于原西南石油学院油气储运工程专业, 获硕士学位; 主要从事油气混输技术研究。地址: (610213) 四川省成都市华阳。电话: 13540649905。E-mail: chenxuefeng74@126.com

二、模糊自适应权重组合预测算法 (FVW法^[3])及其改进

采用变权重组合预测的关键是确定各时刻各单项预测方法的权系数。

对于同一问题,设共有 $n(n \geq 2)$ 种预测方法,设: $y(i)$ 表示 i 时刻的实际观测值 ($i = 1, 2, \dots, n$); $F_j(i)$ 表示 i 时刻第 j 种方法的预测值 ($j = 1, 2, \dots, m$); $e_j(i)$ 表示 i 时刻第 j 种方法的预测误差; $k_j(i)$ 表示 i 时刻求出的第 j 种方法的加权系数, $\sum_{j=1}^m k_j(i) = 1$, $0 \leq k_j(1) \leq 1$; $f(i+1)$ 表示 $i+1$ 时刻的组合预测值, $f(i+1) = \sum_{j=1}^m k_j(i) f_j(i+1)$ ($j = 1, 2, \dots, m; i = 1, 2, \dots, n$).

对于第 j 种方法,其与实际观测值的匹配程度(权系数) k_j 由两个参数决定:最近几个采样周期预测误差绝对平均值的相对指标 E_j 和滚动的有限时域长度内预测误差绝对累计值的相对指标 EA_j .

很显然,在此仅仅考虑采样周期内预测误差平均值的相对指标 E_j 是不够的,还必须考虑预测值误差序列的变化趋势问题,同时为简化计算,最好将误差序列的变化趋势直接包含和反映在 E_j 上,在这里,引入新鲜度函数 $F(t)$ ^[4],用以描述数据在挖掘过程中所受到的重视程度或者用来描述数据在预测未来情况时作为依据的重要程度,越是最近产生的数据,越被赋予较大权重,表明越受到重视,因为越是最近产生的数据,包含的有用信息就越多,这与信息论的观点也是一致的。因此新鲜度函数为一非递减的时间函数,具体形式视具体情况而定,常用的新鲜度函数有:

$$F(t) = t \text{ 或 } F(t) = t^2 \quad (t = 1, 2, \dots, s)$$

新鲜度函数的存在可以保证越是最近的样本,对预测起的作用越大。可以通过对预测值误差序列进行加权求平均来体现这种变化趋势,其权系数是随着时间的推移逐渐变大的,这就体现了预测数值的一个新鲜度问题,也是符合事物的发展规律的。

对文献[3]中相对指标 E_j 的改进由下面的公式给出。

(1)两个相对指标 E_j 、 EA_j 值的计算(后者为改进后的计算公式):

$$\left\{ \begin{array}{l} e_j(i) = y(i) - f_j(i) \\ a_j(i) = \sum_{p=0}^{q-1} |e_j(i-p)| / q \\ s_j(i) = \sum_{p=0}^{l-1} |e_j(i-p)| \\ E_j(i) = a_j(i) / \max_{1 \leq j \leq m} [a_j(i)] \\ EA_j(i) = s_j(i) / \max_{1 \leq j \leq m} [s_j(i)] \end{array} \right. \Rightarrow \left\{ \begin{array}{l} e_j(i) = y(i) - f_j(i) \\ a_j(i) = \sum_{p=0}^{q-1} F(i-p) |e_j(i-p)| / \sum_{p=0}^{q-1} F(i-p) \\ s_j(i) = \sum_{p=0}^{l-1} |e_j(i-p)| \\ E_j(i) = a_j(i) / \max_{1 \leq j \leq m} [a_j(i)] \\ EA_j(i) = s_j(i) / \max_{1 \leq j \leq m} [s_j(i)] \end{array} \right. \quad (1)$$

式中: i 表示当前采样时刻; q 表示取平均的周期数 ($q \geq 1$); l 表示有限时域长度 ($l > q$); $a_j(i)$ 表示采样时刻 i 的第 j 种方法的 q 个预测误差绝对平均值; $s_j(i)$ 表示采样时刻 i 第 j 种方法的 l 个预测误差绝对累计值。 $F(t)$ 表示新鲜度函数。

(2) 根据式(1), $E_j(i)$ 、 $EA_j(i)$ 的值区间为 $[0, 1]$, 将 $k_j(i)$ 的取值范围也定为该区间。设 $\tilde{E}_j(i)$ 、 $\tilde{EA}_j(i)$ 、 $\tilde{k}_j(i)$ 分别为 E_j 、 EA_j 、 $k_j(i)$ 的模糊值,论域相同。其上定义相同的模糊子集:很小(VS)、较小(MS)、小(SS)、中(MM)、大(LL)、较大(ML)、很大(VL)…,则有如下一些关于权系数 $k_j(i)$ 的模糊语言推理规则:

$$\begin{array}{l} \text{if } \tilde{E}_j(i) = VS \text{ and } \tilde{EA}_j(i) = VS \text{ then} \\ \quad k_j(i) = VL \\ \text{if } \tilde{E}_j(i) = VS \text{ and } \tilde{EA}_j(i) = VL \text{ then} \\ \quad k_j(i) = VS \\ \dots \end{array}$$

据此可以建立匹配程度的取值状态表。

若各模糊子集分别由其核元素代表,则权系数的推理可用一个带修正因子 α 的公式来表示:

$$k_j'(i) = -[\alpha E_j(i) + (1 - \alpha) EA_j(i)] \quad 0 \leq \alpha \leq 1 \quad (2)$$

式中: α 反映对最近预测情况的侧重程度。显然,式(2)完全反映了关于权系数的模糊语言推理规则,而且也避免了推理规则定义中的空档或跳变现象,所以这种带修正因子的量化描述方法是合理及可行

的,并且运算极其简便。
 $\tilde{E}_j(i)$ 、 $\tilde{EA}_j(i)$ 的论域为 $[0,1]$,所以 $\tilde{k}_j(i)$ 的论域为 $[-1,0]$,由下列映射公式:
 $k_j'(i) = \tilde{k}_j(i) + 1$ 将 $[-1,0]$ 上的值映射到 $[0,1]$ 上,然后由下式 $k_j(i) = k_j'(i) / \sum_{i=1}^m k_j'(i)$ 进行归一化处理,即可得到第 j 种方法的权系数 $k_j(i)$ 。

三、应用实例

所采用的实例是基于国内某燃气公司 2003 年 9 月~11 月份的数据进行的。课题组从功能模拟的原理出发,建立了微分模拟预测模型和 BP 神经网络预测模型,其中9月份和10月份的数据作为训练样

本,11 月份的数据作为检验样本;同时,还针对这 2 种预测模型进行了组合预测,包括改进前的 FVW 组合预测和改进后的 FVW 组合预测,实践证明改进后的效果更好,进一步提高了预测精度。由于篇幅原因,在此只列出了部分检验数据(表 1)。
模糊自适应变权重 (FVW, Fuzzy Variable Weighting)算法采用式(1)的前者进行计算,其中采样周期数为 $q=3$,滚动有限时域的长度 $l=5$ 。
改进后的 FVW 组合预测算法采用式(1)的后者进行,新鲜度函数采用: $F(t)=\sqrt{t}(t=1,2,\cdots,s)$ 来计算。具体计算中,取 $s=5$ 。各种预测的预测值见表 1。各单一预测方法以及 2 种组合预测法的误差及其误差平方和见表 2。

表 1 11 月份管网负荷的历史数据及各模型得预测值表 10⁴ m³

预测方法	1 日	2 日	3 日	4 日	5 日	6 日	7 日	8 日
实际负荷	285.071	288.609	284.568	311.690	327.099	326.329	328.955	333.801
微分模拟预测	288.312	292.621	281.634	315.907	325.053	321.275	332.171	327.554
BP 神经网络预测	290.245	283.293	288.715	314.015	325.886	330.988	331.534	330.677
FVW 组合预测	289.764	286.865	282.417	315.765	324.160	329.764	331.609	330.430
改进后的 FVW 组合法	287.163	288.274	283.827	313.769	324.946	329.295	332.378	330.625

表 2 11 月份各单一预测方法和优化组合预测方法的误差及误差平方和表 10⁴ m³

预测方法	1 日	2 日	3 日	4 日	5 日	6 日	7 日	8 日	误差平方和
微分模拟预测	3.241	4.012	-2.934	4.217	-2.046	-5.054	3.216	-6.247	132.089
BP 神经网络预测	5.174	-5.316	4.147	2.325	-1.213	4.659	2.579	-3.124	117.222
FVW 组合预测	4.694	-1.744	-2.152	4.075	-2.939	3.435	2.654	-3.370	85.145
改进后的 FVW 组合法	2.092	-0.335	-0.741	2.079	-2.153	2.966	3.423	-3.176	44.595

四、结论与建议

从表 2 给出的各单一预测和优化组合预测方法的误差及其误差平方和数据可得出以下认识。
(1) 优化组合预测法的预测结果明显要优于各单一模型的预测结果,其误差平方和达到极小值,该极小值不大于参与组合预测的各单一预测方法误差平方和的最小值。
(2) 在管网负荷的预测过程中应用变权重组合预测方法,可提高预测精度,基于新鲜度函数改进后的 FVW 组合预测方法要优于改进前的 FVW 组合预测方法。
最后,还应注意,在选取各单一模型进行组合预测时,并非越多越好,事先最好进行冗余性的判断^[5],剔除冗余的模型,在模型预测阶段尽量选取从

不同角度建立的模型,这样可在组合预测阶段既充分利用全方位的信息,又能尽量减少模型之间的相关性,从而充分发挥优化组合预测的优势。

参 考 文 献
[1] 谢开贵,周家启.变权组合预测模型研究[J].系统工程理论与实践,2000(7):36-40.
[2] 曹长修,王景,唐小我.一种模糊变权重组合预测方法——FVW 法的研究[J].预测,1996,15(5):49-50.
[3] 唐小我,王景,曹长修.一种新的模糊自适应变权重组合预测算法[J].电子科技大学学报,1997,26(3):289-291.
[4] 徐宝文,张卫丰.数据挖掘技术在 Web 预取中的应用[J].计算机学报,2001,24(4):430-436.
[5] 陈华友,侯定丕.基于预测有效度的优性组合预测模型研究[J].中国科学技术大学学报,2002,32(2):172-180.
(修改回稿日期 2005-12-02 编辑 居清缙)