

# 基于时间序列的时空插值算法改进研究

李 彦 王丽娜

(重庆理工大学计算机科学与工程学院 重庆 400054)

**摘 要** 为了提高气象预报的精细化程度,针对空间插值算法进行了改进研究,加入了时间序列后的时空混合插值更加符合气象要素的时空特性,从源头上解决了由于各种不可抗力因素(如仪器故障、传输线路故障等)造成的气象站点实时数据缺失。利用重庆市 900 多个气象站近一年的观测数据进行了插值方法的对比研究,证实了该算法的可行性。

**关键词** 时间序列,时空插值,气象要素,实时预测

中图法分类号 TP301.6 文献标识码 A

## Research of Spatio-temporal Interpolation Algorithm Based on Time Series

LI Yan WANG Li-na

(School of Computer Science and Engineering, Chongqing University of Technology, Chongqing 400054, China)

**Abstract** In order to enhance the degree of the precision of weather forecast, the research carry on the spatial interpolation algorithm. Joined the time series in the spatio-temporal interpolation, it was more accordance with the spatial and temporal characteristics of meteorological elements, and solved the lack of real-time data caused by a variety of force majeure from the source (such as instrument meteorological site failure, transmission line failure, etc.). Based on the observation data of over 900 weather stations in Chongqing of nearly a year, it conducted a comparative study of interpolation methods, and confirmed the feasibility of the algorithm.

**Keywords** Time series, Spatio-temporal interpolation, Meteorological elements, Real-time prediction

## 1 引言

随着科技和社会的进步,气象科技在生产生活决策中的作用越来越受人关注。很多与气象息息相关的行业更是提出了气象精细化服务,在气象精细化服务中关键是要解决“精”和“细”的问题<sup>[1]</sup>。通常我们在提到气象精细化时,更多强调的是气象要素细化、时空分辨率提高等问题,而忽略了气象要素“精”度问题。为了提高气象预报的精细化程度,针对空间插值算法进行了改进研究。由于气温、降雨量等气象要素观测值是典型的时空数据,它们有明显的时间序列特征和很强的空间地域性<sup>[2-4]</sup>。现有的气象要素插值方法往往将时空分开来进行研究,单一角度造成大量有价值的信息丢失,导致气象数据的预测精度不够高。因此我们需要从时空混合的角度对气象数据进行分析。时空插值主要应用于以下两个方面:一个是对不规则数据集插值,这样解决了气象数据时间序列的不完整对气象统计造成的不便<sup>[2]</sup>;另一个是对气象站点缺失数据进行修补,这样可以解决由于样本点减少引起的插值精度降低的问题<sup>[5]</sup>。

传统的时空插值主要是针对历史气象数据的填补,实现了气象统计的功能。但是很多时候,我们并不仅仅关心站点

的历史数据,精确的实时数据才是我们的关注要点。实时信息关系到气象部门的各种决策的执行,还有许多对气象敏感的部门,比如水利、电力部门都需要将各气象站点的实时数据作为日常调度的一个重要参考。但是由于一些不可拒因素的影响,比如极端天气、传输故障、观测机器故障等原因,都会造成有的站点数据丢失,这时就需要用一定的预测手段由站点的历史记录来推断缺失站点的实时气象要素值。本文在此需求的基础上对时空插值算法进行改进,这样就可以实现在异常情况下气象站点实时气象要素的预测。以下的章节给出了现有时空插值面临的问题和改进后算法的描述和实现。最后用图示的方法展示了现有的插值方法和改进后的插值方法在插值精度上的对比。

## 2 数据来源与现有研究方法概述

### 2.1 数据来源

本文选用重庆市气象局一年来(20120801~20130730) 900 多个气象观测站的日平均温度、日平均降水、日平均相对湿度等气象资料,并获取气象站的位置坐标数据,其中 911 个气象的站分布状况如图 1 所示,其中紫色站点为区域站,红色为自动站。

本文受重庆市气象服务中心科技项目(2012Q86)资助。

李彦(1976-),男,教授,主要研究方向为计算机网络与软件;王丽娜(1988-),女,硕士生,主要研究方向为计算机网络与软件, E-mail: wanglina164@126.com。

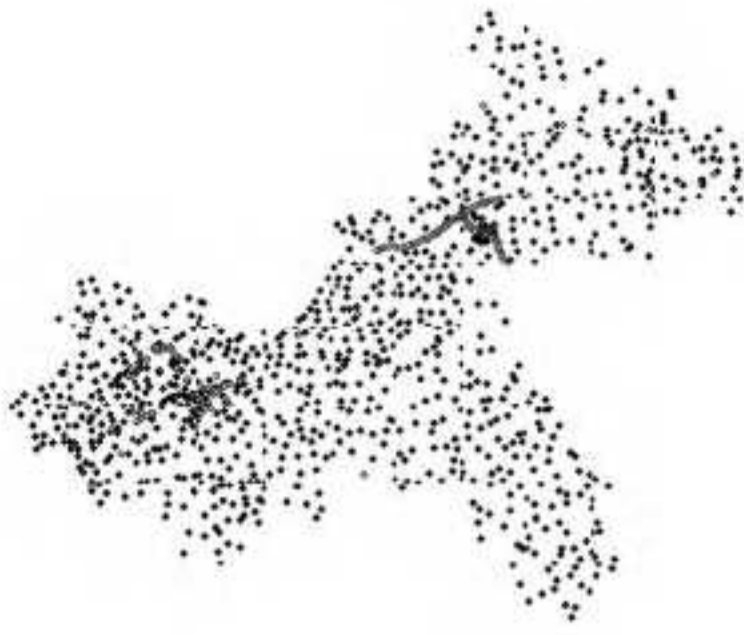


图1 气象站分布图

## 2.2 现有研究方法概述

已有的插值研究中,空间插值和时间插值的整合研究相对较少。在现有的空间插值方法的基础上引入时间插值,目前常用的方法为约减法和扩展法,这两种方法都可以用来进行不规则数据集的时空插值,但是也存在着不可忽视的漏洞。

### 2.2.1 约减法

约减法是将时间作为独立维进行考虑,主要用于气象样本站点在某些时刻有缺失时的数据弥补,所以通常先利用时间插值对缺失数据的气象样本站点进行数据补漏,再在此基础上进行空间插值。这里以反距离加权法为例,假设所有气象站点在 $t$ 时刻的气象数据丢失, $t$ 时刻前后数据存在,要求 $q$ 站点 $t$ 时刻的气象值,约减法时空插值的计算公式如式(1)和式(2)所示:

$$Z_i(t) = \frac{t_{i2} - t}{t_{i2} - t_{i1}} * Z_{i1} + \frac{t - t_{i1}}{t_{i2} - t_{i1}} * Z_{i2} \quad (1)$$

$$Z(t) = \frac{\sum_{i=1}^m Z_i(t)}{\sum_{i=1}^m \frac{1}{d_i^n}} \quad (2)$$

其中, $Z_{i1}(t)$ 、 $Z_{i2}(t)$ 为第 $i$ 个气象样本点在 $t_{i1}$ 、 $t_{i2}$ 时刻的实测值, $Z_i(t)$ 为第 $i$ 个气象样本点在时刻 $t$ 的估算值; $d_i$ 为第 $i$ 个气象站点与待插点之间的空间距离, $Z(t)$ 为待插站点在时刻 $t$ 的气象要素估算值; $m$ 为参与插值计算的气象样本点个数, $n$ 为幂指数。

### 2.2.2 扩展法

扩展法是把时间作为与空间相同的第三维来考虑的,同样以反距离加权法为例,扩展法的计算如式(3)、式(4)所示,选取与待插点最近“距离”的气象样本点作为样本值,这里的距离采用时空距离。

$$Z(t) = \frac{\sum_{i=1}^m Z_i(t_i)}{\sum_{i=1}^m \frac{1}{d_i^n}} \quad (3)$$

$$d_i = \sqrt{(x_i - x)^2 + (y_i - y)^2 + (t_i - t)^2} \quad (4)$$

其中, $Z_i(t_i)$ 为第 $i$ 个气象样本点在 $t_i$ 时刻的实测值, $t_i$ 为 $t$ 临近时刻, $d_i$ 为第 $i$ 个气象样本点与待插点之间的时空距离。

### 2.2.3 存在的问题

时空插值的约减法与扩展法开创了时空混合插值的先河,具有很大的参考再研究价值<sup>[6]</sup>,但这两种方法存在的问题也不容忽视。约减法中待插值时刻气象站点的缺失值需要从该时刻前后各选出一个最近时刻的实测值进行估算,它只能用于站点历史缺失值的填补,这样只能用于气象统计,但是对气象信息的实时监测的要求就不能胜任了。而扩展法中时空单位的不确定性会导致插值结果存在很大的差异性。因为空间单位的米或者千米和时间单位的年、月、日、时、分或秒做不

同组合时,得到的插值结果会有很大的差别,所以扩展法存在不稳定性。本文将以约减法为基础,加入时间序列模型,为气象站点的实时数据缺失提供有效的解决方法。以下将详细介绍该算法的思路、实现步骤及实验结果讨论。

## 3 基于时间序列的气象插值算法

时间序列是一种将具有统计指标的数值按时间顺序排列形成的数列<sup>[7]</sup>。时间序列气象插值算法就是通过编制和统计分析气象站点时间序列,根据序列反映出来的统计规律或线性趋势预测下一时刻的站点气象数值。因为在气象数据插值研究中我们不仅关心序列形态,很多与气象相关的行业更关心气象实时数据,也即序列某时刻值的大小。

### 3.1 改进的时空插值算法

$X = \langle x_0, x_1, \dots, x_i, \dots, x_n \rangle$ ,  $x_0$ 至 $x_n$ 是一个采样间隔为 $\Delta t = x_i - x_{i-1}$ 的时间序列,其中 $x_i$ 表示采样时刻 $i$ 的气象数据。首先对原始序列 $X$ 进行处理,找出区间隔离特征点,把相邻区间隔离特征点相连组成一个个子区间,然后提取出每个子区间的斜率、均值及区间长度。区间特征点有3种情况:第1种为时间序列的起点和终点;第2种为时间序列的极值点;第3种为某点前后线段的斜率差大于一定阈值。以下是该算法的重要点的确定方法及插值原理:

1)极值点的确定:对于时间序列 $X$ ,如果 $X$ 满足条件 $x_q \leq x_{q+1} \leq \dots \leq x_i$ ,且 $x_i \geq x_{i+1} \geq \dots \geq x_r$  ( $1 \leq q \leq i \leq r \leq n$ ),或者满足 $x_q \geq x_{q+1} \geq \dots \geq x_i$ 且 $x_i \leq x_{i+1} \leq \dots \leq x_r$  ( $1 \leq q \leq i \leq r \leq n$ ),即时间序列 $X$ 的单调性在时刻 $i$ 发生变化,则 $x_i$ 就被认为是序列 $X$ 的极值点。

2)斜率差 $\Delta k$ 的确定:当时间序列没有明显趋势的变化时,可以用斜率差 $\Delta k$ 来反映序列的转折情况。如图2所示,当两条线段相交时, $\Delta k$ 越大这两条相邻线段的趋势相差越大,其中①线段表示当 $\Delta k = 0$ 时,两条线段 $L_1, L_2$ 在一条直线上;②③表明在 $\Delta k \neq 0$ 且单调性不变时线段中间有弯折。此时的拐点由斜率阈值来确定是否为区间特征点。因为是用气温进行验证,本文中选取斜率阈值为 $\Delta k = 0.5$ 。



图2 斜率差示意图

3) $t$ 时刻缺失数据站点气象要素预测:时间序列插值法通过对序列 $X$ 的一次顺序扫描可以求出序列特征点,然后在此基础上计算出区间中各段的斜率、区间长度、均值等特征,最后可以拟合出该线段的线性函数 $Z(t)$ ,则时刻 $t$ 的气象要素值就可以通过线性函数预测出来,即为 $Z(t)$ 。

$$Z(t) = \begin{cases} k - tz * \Delta x + X(t), & \Delta k > 0.5 \\ k(t) * \Delta x + X(t), & \Delta k \leq 0.5 \end{cases} \quad (5)$$

其中, $k(t)$ 为时刻 $t$ 的斜率数列, $\Delta k$ 为斜率差, $k - tz = \frac{1}{m} \sum_{i=1}^m k$ , $m$ 为 $t$ 时刻所在区间长度。

4)加入空间插值模型:单独的空间插值方法和单独的时间序列方法都没能周全地考虑气象观测值的时空特征,在修补不规则数据集和填补实时缺失数据时精确度还不够。这里

将时间序列模型加入到空间插值的公式如式(6)所示:

$$Z = w * Z_p + (1 - w) * Z_t + q \quad (6)$$

式中,  $q$  为常数项,  $w$  为空间插值估算值的系数, 将空间插值估算值  $Z_p$  和时间插值估算值  $Z_t$  代入到式(6), 采用最小二乘法解出最优的系数  $w$ 。

### 3.2 时间序列预测算法步骤

为了提高插值精度, 本文将空间插值和时间插值结合于一体, 改进后的插值算法流程如图 3 所示。

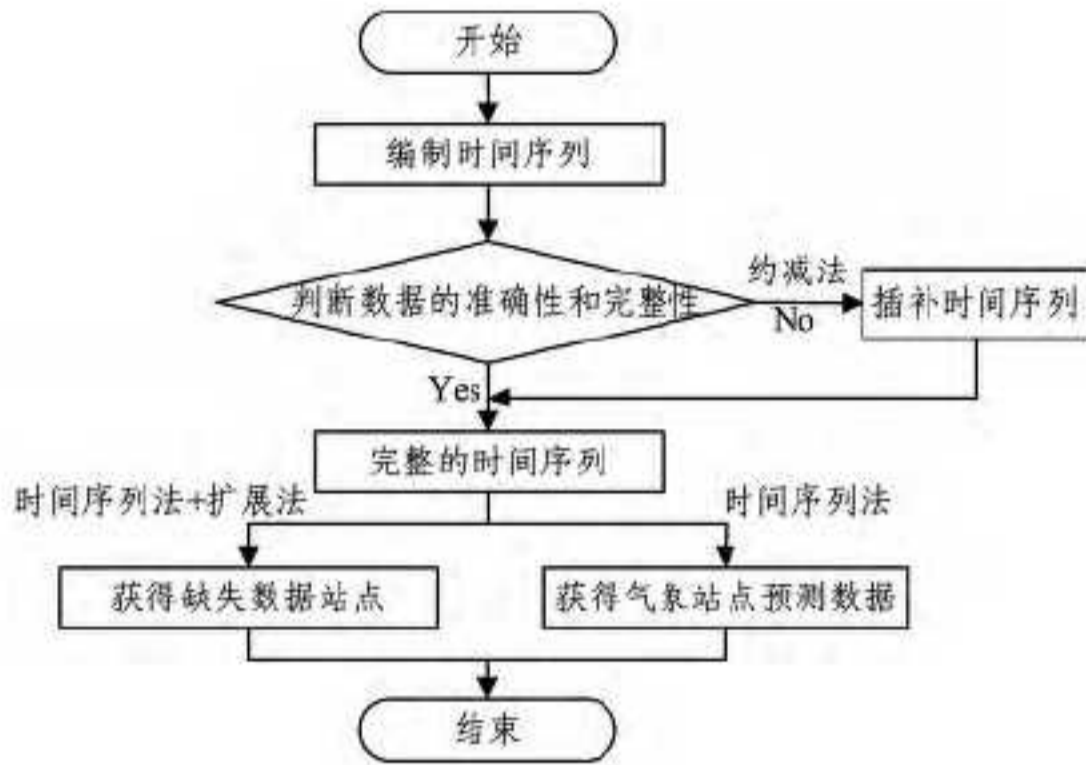


图 3 时空插值流程

#### Step 1 编制时间序列

收集气象资料, 为每一个气象站点编制一个时长为一年的时间序列。共计 914 个时间序列可供分析验证。

#### Step 2 插补时间序列

由于本文的研究还处于起始阶段, 会遇到原始数据缺失的问题, 这里选用临近点插值的方式进行初步解决, 即把空缺值或奇异值用空间插值的方式, 利用临近站点的气象值进行插补, 使之成为一个完整的时间序列。

#### Step 3 建立时间序列插值模型

为解决传统时空插值只能用于过去缺失数据的填补, 不能针对实时数据缺失进行预测的问题, 本文建立时间序列插值模型。

$$Z_p = \sum_{i=1}^m \frac{Z_i}{d_i^2} / \sum_{i=1}^m \frac{1}{d_i^2} \quad (7)$$

$$Z_t = Z(t) \quad (8)$$

$$Z = w * Z_p + (1 - w) * Z_t + q \quad (9)$$

其中,  $Z_p$  为  $t$  时刻空间插值结果,  $Z_t$  为  $t$  时刻缺失数据站点时间序列法插值结果。  $Z$  为  $t$  时刻基于时间序列的时空混合插值结果。

#### Step 4 模型预测

根据实际需求利用时间序列插值模型进行气象站点的实时数据修补或站点的气象预测, 为许多对实时气象数据敏感的部门提供更为精确的实时插值数据, 其次也为气象站点提供完备的时间序列数据, 为后来的气象统计和数据建模省去了数据预处理步骤。

## 4 结果分析

为了提高算法检验的效率, 本文选用最常用的空间插值方法即反距离加权法 (IDW), 建立时间序列预测模型, 时空插值选用加权平均法, 其中权重  $w$  的选择依赖于经验值, 实验时选为 0.5。

### 4.1 空间插值方法结果

为了提高检验准确率和可信度, 本文对 911 个气象区域站进行回归迭代检验, 即: 假设每一个站点为数据缺失站点, 然后用反距离加权法插值进行数据插补, 最后将所得估值数据和原始观测数据进行对比, 这里采用平均绝对误差 MAE (Mean Absolute Error) 和平方根误差 RMSIE (Root Mean Squared Interpolation Error) 计算所有站点的实测值与估算值的误差来评估空间插值方法的精度。结果如图 4 所示。

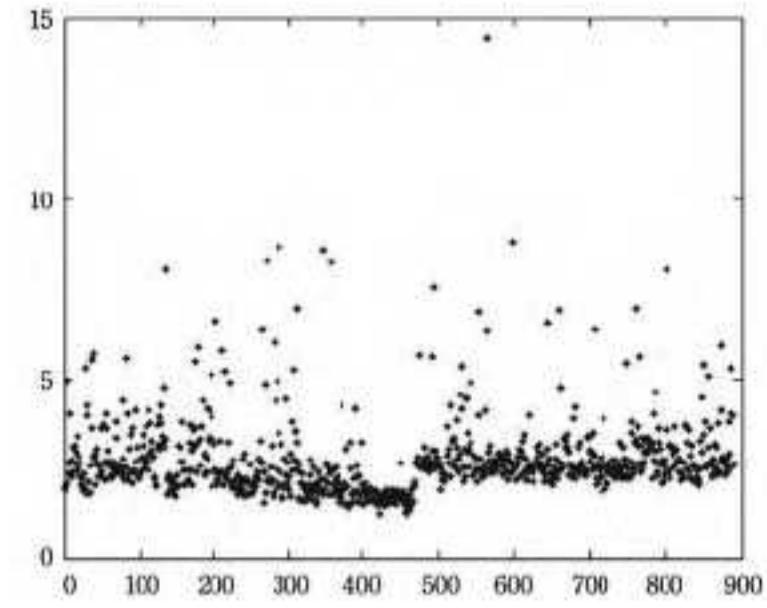


图 4 空间插值误差

由图 4 看出, 误差主要集中在  $2^{\circ}\text{C}$  上下, 有些误差比较大的站点数据本身就有异常的可能。

### 4.2 时间序列插值结果分析

为了进一步细化检验结果, 这里首先对时间序列模型的插值结果与空间插值结果进行对比, 如图 5 所示。

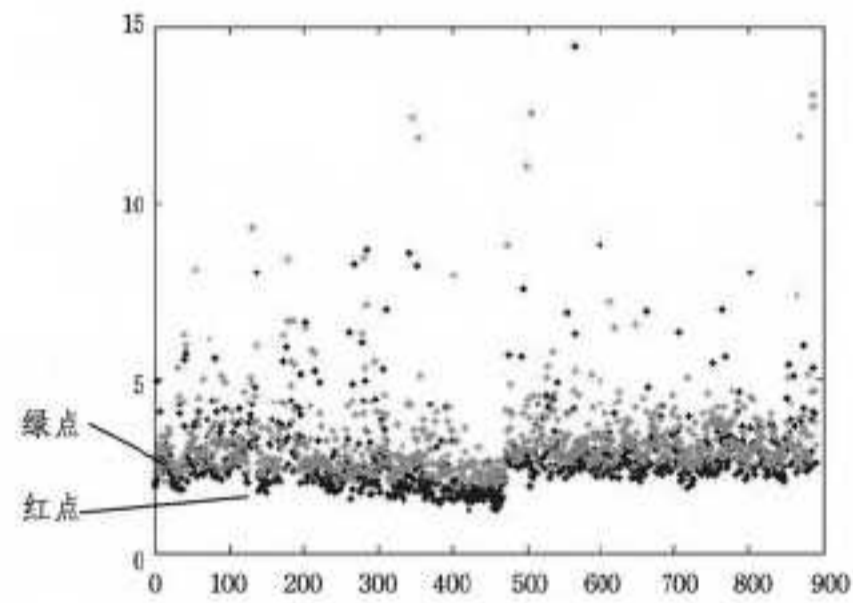


图 5 时间序列插值

图 5 中, 绿色点为单独的时间序列插值误差。红色点为空间插值 (idw) 的估值误差。通过对比我们可以发现这两者的误差大小大致相同, 都能在精度要求不高的情况下满足需求。

### 4.3 时空混合插值结果对比

如上所述, 气象要素具有很大的时空特性。为了提高插值精度, 本文在空间插值的基础上加入时间序列模型, 时空混合插值的结果如图 6 所示。

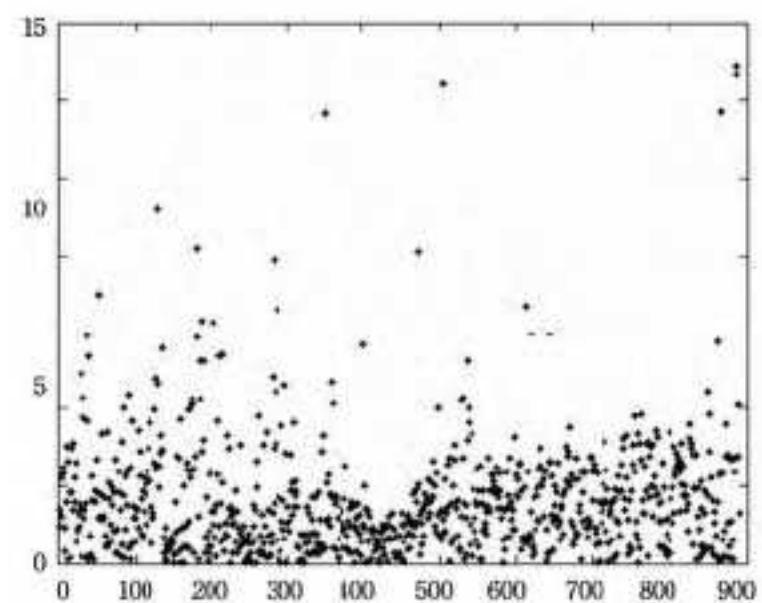


图 6 加入时间序列的插值误差图

(下转第 424 页)

的信息十分关注,技术人员对工艺流程、理化检测报告、精密检测报告等技术类信息的需求强烈,但两者均对不同车间、不同班组、不同工种之间的工作负荷对比以及变化趋势等较为宏观的信息没有兴趣,而这些信息却正是管理人员非常关心的。

表 4 三大类用户群的强识别规则集

编号	大类识别规则	置信度	覆盖度
R <sub>1</sub>	$a_4=5$ and $a_5=1 \rightarrow D=I$	9/9=1	9/9=1
R <sub>2</sub>	$a_2=5$ and $a_5=1 \rightarrow D=II$	6/6=1	6/6=1
R <sub>3</sub>	$a_5=5 \rightarrow D=III$	8/8=1	8/9=0.889

#### 4.2 发现子类用户群信息需求的差异性

采用同样的计算步骤,可以在大类用户群识别规则的前提下,从细粒度用户空间下的决策表中获取各子类用户群的强识别规则集,如表 5 所列。

表 5 各子类用户群的强识别规则集

编号	大类识别规则	子类识别规则	置信度	覆盖度
R <sub>1</sub> '	R <sub>1</sub>	$a_1=5 \rightarrow D'=I_1$	3/3=1	3/3=1
R <sub>2</sub> '		$a_1=1 \rightarrow D'=I_2$ or $D'=I_3$	6/6=1	6/6=1
R <sub>3</sub> '	R <sub>2</sub>	$a_3=5 \rightarrow D'=II_1$	3/3=1	3/3=1
R <sub>4</sub> '		$a_3=1 \rightarrow D'=II_2$	3/3=1	3/3=1
R <sub>5</sub> '	R <sub>3</sub>	$a_1=1$ and $a_4=1 \rightarrow D'=III_1$	2/2=1	2/3=0.667
R <sub>6</sub> '		$a_1=1$ and $a_4=5 \rightarrow D'=III_2$	3/3=1	3/3=1
R <sub>7</sub> '		$a_1=4$ and $a_4=1 \rightarrow D'=III_3$	3/3=1	3/3=1

从表 5 中可以发现各子类用户群信息需求的差异性,对于三类技能人员来说,其差异性集中体现在对生产任务类信息的需求上,可以看出加工操作人员非常关心车间生产任务的进展情况,迫切希望了解未来一段时间本人和所在班组将要承担的加工任务;对于两类技术人员来说,其差异性突出表现在对制造资源类信息的需求上,可以发现工艺编制人员在此方面的信息需求强烈,说明他们必须通过掌握加工设备的工艺参数与运行状态以及车间刀具、工装等库存信息来确定合理可行的工艺路线,同时也需要深入了解企业器材材料的库存情况,以便准确地编制工艺流程的材料清单;三类管理人员的信息需求相比较而言,企业生产管理人员作为整个企业生产任务的全局性计划和协调人员,比其他两者更关注生产任务类信息,而车间计划调度人员却更关心车间个人工作量

(上接第 416 页)

如图 6,因为时间序列和空间插值的结合,使得插值误差有了很大的降低,平均在 1℃ 左右,这对插值精度的提升具有很大的意义。

结束语 基于时间序列模型时空插值算法,加入时间序列模型的插值算法因为可以既考虑空间因素的影响又同时考虑时间因素对气象数据的影响,避免了空间插值的单一性。实验表明,它比目前经常使用的数据修补方法更加精确,不仅实现了已经残缺的气象观测序列的修补,其最大的贡献在于解决了极端天气下线路故障或气象设备损坏而收不到气象观测站的实时气象数据的问题,方便气象敏感部门(如电力部门)根据实时的情况做出相应的措施。但是由于数据的变化不一定是严格按某一规律进行的,因此该方法不能保证对所有的时间序列都适合,尤其对于数据突变明显的时间序列。

方面的信息。

结束语 满足用户的多样化信息需求是企业级信息系统实施成功与否的关键要素之一。基于粗糙集和商空间理论的思想,本文提出了一种通过企业级信息系统的日志挖掘,从不同的用户粒度空间分析企业用户群信息需求一致性和差异性的技术方法,并通过一个运行多年的企业级信息系统的实例分析,验证了方法的可行性和有效性。如何从单个功能模块的使用中发现更深层次的用户信息需求,以及如何利用信息推送技术将获取的用户客观信息需求转化为信息系统细致入微的主动信息服务,将是下一步研究工作的重点。

#### 参考文献

- [1] 付博,赵世奇,刘挺. Web 查询日志研究综述[J]. 电子学报, 2013,40(9):1800-1808
- [2] 何跃,马丽霞,腾格尔. 基于用户访问兴趣的 Web 日志挖掘[J]. 系统工程理论与实践,2012,32(6):1353-1361
- [3] 王国胤,张清华,胡军. 粒计算研究综述[J]. 智能系统学报, 2007,2(6):8-26
- [4] 苗夺谦,王国胤,刘清,等. 粒计算:过去、现在与展望[M]. 北京:科学出版社,2007
- [5] 张钊,张铃. 粒计算未来发展方向探讨[J]. 重庆邮电大学学报:自然科学版,2010,22(5):538-540
- [6] 薛志远,张清华. 复合粒计算模型研究进展[J]. 重庆邮电大学学报:自然科学版,2010,22(5):631-640
- [7] 王国胤,姚一豫,于洪. 粗糙集理论与应用研究综述[J]. 计算机学报,2009,32(7):1229-1246
- [8] 苗夺谦,李道国. 粗糙集理论、算法与应用[M]. 北京:清华大学出版社,2008
- [9] 吴伟志,杨玉芳. 空间遥感数据的多粒度标记分类[J]. 计算机科学,2012,39(4):23-27
- [10] 张铃,张钊. 问题求解理论及应用\_商空间粒度计算理论及应用[M]. 北京:清华大学出版社,2007
- [11] 周丹晨,尚黎,周战强. 面向产品制造全过程的企业信息集成平台研究[J]. 计算机应用与软件,2008,25(7):149-151
- [12] 王国胤,于洪,杨大春. 基于条件信息熵的决策表约简[J]. 计算机学报,2002,25(7):759-766

#### 参考文献

- [1] 矫梅燕. 在改革中探索精细化预报业务发展[N]. 中国气象报, 2006,10(26)
- [2] 刘晓晓. 气象监测数据的时空特征分析与建模研究[D]. 开封:河南大学,2009
- [3] 李莎,舒红,等. 利用时空 kriging 进行气温插值研究[J]. 武汉大学学报,2012(2):237-240
- [4] 张建龙,王玲. 重庆岩溶区的气候时空变化特征分析及趋势预测[J]. 热带气象学报,2008(6):239-248
- [5] 徐爱萍,胡力,等. 空间克里金插值的时空扩展与实现[J]. 计算机应用,2011(31):273-276
- [6] 彭思岭. 气象要素时空插值方法研究[D]. 长沙:中南大学,2010
- [7] 谢福鼎,王赫楠,等. 基于函数的时间序列分段线性表示方法[J]. 计算机科学,2011,38(11):153-155