

南海区域站实时气象数据质控系统研究

王立俊,江 益,程洪涛,郑虹晖

(海南省气象信息中心,海南 海口 570203)

摘 要:近年来,各类型区域站提供的气象资料对天气预报及气候研究等方面的作用越来越大。由于南海区域站位于岛礁上,数据采集易受到周围多变环境的影响,会造成数据产生偏差或错误。为保证海岛站数据的可靠性和可用性,提出了HV-Process(Horizontal and Vertical Process)算法。该算法利用双重判断和处理机制,对原有质量控制得出的疑误数据进行分析处理,降低了南海区域站气象数据的误检率。在此基础上,设计并研发了一套实时气象资料质控系统,实现了实时质控、全库质控和数据查询等功能,实时处理最新入库的疑误数据,并提供简易的实时查询。应用演练情况表明,所设计构建的实时气象资料质控系统在日常业务中提高了甄别南海区域站气象疑误数据的准确性,有效地保障了实时气象数据的质控业务。

关键词:南海区域站;气象数据;HV-Process 算法;实时质控

中图分类号:TP311

文献标识码:A

文章编号:1673-629X(2017)08-0177-05

doi:10.3969/j.issn.1673-629X.2017.08.037

Study on Quality Control System of Real Time Meteorological Data from Regional Stations on South China Sea

WANG Li-jun, JIANG Yi, CHENG Hong-tao, ZHENG Hong-hui

(Meteorological Information Center of Hainan Province, Haikou 570203, China)

Abstract: In recent years, the role of meteorological data provided by various types of regional stations in weather forecasting and climate research has becoming more and more large. Since the regional stations are located in the islands or reefs of the South China Sea, data collecting is vulnerable to be affected by the surrounding environment, which may result in the deviation or error for the data. To ensure the reliability and availability of the data from island stations, HV-Process algorithm to analyze and process the suspect data of the preceding quality control by using the mechanism of dual judging and processing is proposed, which reduces the false detection rate of meteorological data obtained from regional stations. And thus a system of quality control for real time meteorological data is developed, which implements so many functions, including quality control for real-time and whole database, and data query, real-time processing of the latest data and providing a simple real-query. The applications practice show that it has improved the accuracy of distinguishing the meteorological suspect data of regional stations and effectively guaranteed the quality control service for real-time meteorological data in the daily business.

Key words: regional stations on the South China Sea; meteorological data; HV-Process algorithm; real-time quality control

0 引言

近年来,对着中尺度、短时效和特殊行业的天气预报不断发展以及全球气候和大气环境的深入研究,对气象资料服务的需要越来越高。目前,气象部门在全国已建的自动站超过30 000个,其中很多为无人值守的区域站,又由于站点数量众多,区域站数据质量具有不确定性,所以要对区域站数据进行质控^[1-3],这样才能确保数据的代表性和准确性^[1,4]。对此,中国气象局

联合各省级气象局,研发了台站-省级-国家级的MDOS(Meteorological Data Operation System,气象资料业务系统)。通过该系统对各种类型的区域自动站数据进行质控,对质控后的疑误数据,再通过人机交互的形式进行审核。

南海是我国领土的重要组成部分,具有重要的战略军事意义,因此预报南海区域的天气,南海气象资料显得尤为重要。由于南海区域自动站分布在南海的各

收稿日期:2016-10-05

修回日期:2017-01-09

网络出版时间:2017-07-05

基金项目:国家科技支撑计划课题(2013BAK05B03);海南省气象局科技创新项目(HNQQN201405)

作者简介:王立俊(1989-),男,助理工程师,研究方向为气象数据理论和应用。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20170705.1652.078.html>

个岛礁上,其位置具有特殊性,且数据采集易受周围环境的影响,会造成数据出现偏差甚至错误,为确保自动站资料的可靠性和可用性,需要探索适用于海岛区域自动站资料的质控方法。

针对南海区域自动站数据的质控特点,在 MDOS 的质控数据基础上,提出了 HV-Process 算法。该算法利用双重判断和处理机制,再次对原有质控数据进行质控,剔除掉不满足数据对比一致性的记录,降低南海区域站数据的误检率;在此基础上,研发了一套具有实时质控、全控质控和数据查询等功能的南海区域站实时气象数据质控系统。应用演练表明,在日常业务工作中,系统提高了甄别南海区域站气象数据的准确性和值班人员的工作效率,有效地保障了南海区域站实时气象数据的质控业务。

1 相关技术

1.1 质控技术

针对不同的气象数据,国内外研制了不同的质控系统,例如:在地面观测业务中,北欧采用台站级质控、入库前实时质控、入库后非实时质控和人工质控的方式对自动站数据进行质控^[4],美国则采用台站级、州级和国家级的三级质控方式对地面观测数据进行质控^[3];与国外相比,国内是对自动站资料进行台站级、省级和国家级的三级质控^[5]。

针对不同质控系统中数据的质量检查过程,主要分为以下几种检查方法^[6-13]:要素主要变化检查、要素内部一致性检查、要素时间一致性检查和空间一致性检查。

(1)要素主要变化检查:指某个固定站点历史上出现过的最大(最小)值^[6],超出主要变化范围的数据,为历史极值,需要进一步确认,同样主要通过站点历史数据的统计分析得到。

(2)要素内部一致性检查:不同要素之间是否符合一定的物理联系^[7],例如:日最低气温数值 \leq 当日各时次的气温值 \leq 日最高气温数值;极大降水量 \geq 最大降水量。当相关数据不满足这类关系时,其中有一方的数据至少为错误数据。

(3)要素时间一致性检查:不同要素要满足随时间变化的特定规律^[8],如小时气温有明显的日变化,如果连续 24 小时无变化,数据为可疑或错误。

(4)空间一致性检查:根据要素的空间分布特点,检查待测测站的数据与邻近测站的数据是否满足一定的空间连续性,目前较常用的方法有 Madsen 方法^[10]和空间回归检查法^[11]。

1.2 SSH 框架

SSH 框架是目前主流的轻量级 J2EE 软件架构,根

据 Struts2、Spring 和 Hibernate 三者的应用特性进行整合而成,其基本架构如图 1 所示。Struts2 为表现层,是所有业务逻辑的最终展现层,充当逻辑层的客户端,实现人机交互;Spring 为业务逻辑层,负责业务对象转换传递、业务逻辑组织及事务控制等工作,由 spring framework 实现业务 beans 间的依赖关系管理和运行时的注入和事务控制等功能;Hibernate 为数据持久层,负责业务数据的持久化存储,采用 O/R Mapping 技术实现。



图 1 SSH 框架基本架构图

其中,Struts2 是基于 MVC(模型-视图-控制器)开发模式的开源 Web 层应用框架,主要采用 Servlet 和 JSP 技术实现,具有很高的可配置型。Spring 为企业提供了一种轻量级的业务解决方案,通过简单的配置,可以实现“快速装配式企业组件”,在此基础上,Spring 提供了包括多种持久化数据库,声明式事务管理和 Web Services 远程访问业务逻辑的解决方案。Hibernate 是一个优秀、轻量级对象持久化的框架,通过配置映射文件对 Java 对象和数据库建立映射关系,实现操控对象即操作数据库的目的,也提供了数据查询和恢复机制。相对于传统地使用 JDBC 和 SQL 来手工操作数据,使用 Hibernate,可以大大减少操作数据库的工作。

通过使用 SSH 这三种框架的优化组合实现系统分层解耦,一方面有利于项目组各角色的明确分工,提高开发效率,缩短系统开发周期,另一方面使系统具有良好的扩展能力和可维护性^[14-15]。

2 HV-Process 算法

在地面观测数据质控业务中,国内采用台站-省级-国家级三级质控的方式,并研发了 MDOS 系统来实时质控采集到的地面观测数据。在此基础上,针对南海区域自动站数据的质控特性,提出 HV-Process 算法,对经 MDOS 质控后得到的疑误数据进行横向-纵向双重判断和处理,剔除不满足数据对比一致性的疑误数据,约简原有疑误数据的条数,并降低质控数据的误检率。

2.1 算法思路

南海区域自动站每个岛礁都建有两个测站,每个测站通过双通道来传输采集数据。因此,算法利用双站对比和累计对比的方法过滤原有的质控数据。其中,双站对比方法是指在同一地理区域内的岛礁,遍历 MDOS 质控后的数据,根据同个岛礁的不同区站号,逐一进行双站双通道的、同要素编码的数值对比;累计对

比则是指在同一地理区域具有多个岛礁,遍历质控后的数据,根据当前区站号,逐一与要素编码相同的其他站点进行计算。

整个流程主要分为四个部分,如图2所示。

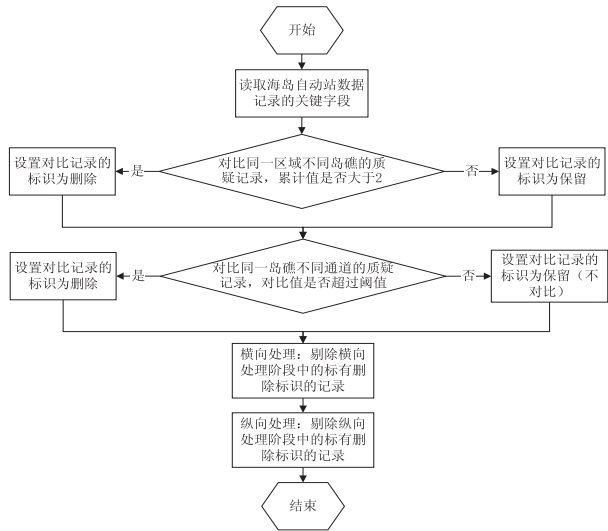


图2 HV-Process 算法流程图

(1)数据纵向判断:遍历当前时次的所有疑误记录,统计在当前时次中,出现与当前记录的要素编码相同的其他记录(同一区域不同岛礁的台站),累计值超过数值2的记录,置该记录的标志为删除标识;否则,置该记录的标志为保留标识。

(2)数据横向判断:遍历当前时次的所有疑误记录,与当前记录的要素编码相同的其他记录做数值对比(同一区域同一岛礁的其他台站),参考自动气象站测量性能指标要求^[13]和自动站与备份站之间各要素的差值范围,确定当前记录的标识,超过设定阈值,标记该记录的标志为删除标识;否则,标记为保留标识。

(3)数据横向处理:遍历上述横向判断后的所有记录,剔除掉所有标有删除标识的记录。

(4)数据纵向处理:遍历(3)阶段处理完的所有记录,剔除掉所有标有删除标识的记录。

2.2 实验

2.2.1 实验环境

使用Java编程语言来实现算法,算法的实验环境如表1所示。

表1 实验环境参数表

设备	参数
CPU	Intel Core i5-3337U CPU 主频 1.80 Hz
内存	8 G
硬盘	WDC 1T
操作系统	Windows 8.1 64 bit

2.2.2 数据集

由于MDOS质控系统是2015年7月后正式运行

的,所以,算法的实验数据集为MDOS数据库中2015年9月至2016年9月的海岛区域自动站的疑误数据。

2.2.3 实验结果及分析

根据上述的方法,将HV-Process算法拆分成横向和纵向两类操作,分别对数据集进行了以下操作:

- (a)15次横向判断-处理操作;
- (b)15次纵向判断-处理操作;
- (c)15次横纵向判断-处理操作。

实验结果见表2(均取执行15次操作后的平均值)。

表2 实验结果

操作	记录数 (前)/条	记录数 (后)/条	有效剔除率/%	耗时 /s
a	10 584	4 986	64.9	39
b	10 584	2 055	98.5	43
c	10 584	2 023	98.8	62

其中,记录数(前)为质控后的数据记录条数;记录数(后)为执行上述某操作后的数据记录条数;有效剔除率为经处理后正确删除的数据记录与处理前正确数据被检测为异常的数据记录之比;耗时为执行上述某操作的时间。

实验结果表明:在有效剔除率方面,由于数据集中存在某些岛礁只有单个区站的质控数据,执行(a)操作时,不能进行质控数据的横向对比、判断,所以有效剔除率只有64.9%,而执行(b)操作后的有效剔除率会接近99%,效果与操作(c)相近。但如果出现同个区域内,仅有3个以内的岛礁的质控数据的情形,执行(b)操作后的有效剔除率将会明显下降。在时间消耗方面,由于算法在对比-处理过程中有效地使用了数据缓存机制,执行操作(c)的时间消耗仅比执行操作(a)和(b)多一些,但能保证较高的有效剔除率。

因此,文中算法将操作(a)和(b)整合在一起,有效地降低了质控数据的误检率,且对于正确数据被检测为异常的情况,与原有质控方式相比,该算法利用双重判断-处理机制(操作(c)),有效地删除掉被检测为异常的正确数据,大大约简了质控后的数据记录,保证了南海区域站气象质控数据的正确性和可用性,提高了MDOS值班工作人员的工作效率。

3 系统的设计与实现

为了能更好、更快地对疑误数据进行校对,提高值班人员的工作效率,在上述基础上,根据MDOS值班人员的工作需求,经分析得到如图3所示的系统工作流程,并研发了一套岛礁区域自动站的,集实时质控、数据查询和全库质控于一体的实时质控系统。

3.1 系统整体架构

系统的工作流程如图 3 所示。

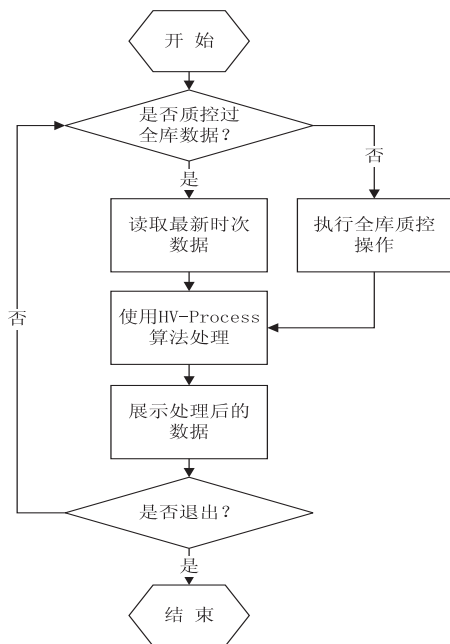


图3 系统工作流程图

首先,如果是第一次安装、部署系统,需由管理员判断数据库是否执行过全库质控操作,若没执行过,执行全库质控操作,若已执行过,跳过全库质控操作,进入实时质控环节。

主要分为两个部分：

(1)全库质控部分:导入历史数据库或者还原备份数据库后,须由管理人员执行该操作,使用 HV-Process 算法处理当前数据库中的区域站疑似数据表中的数据,接着将处理后的数据展示在客户端。

(2) 实时质控部分:系统正常运行时,会实时监控最新时次的区域站疑误数据的入库行为,若监控发现有数据的入库行为,则调用 HV-Process 算法进行处理,并在客户端展示处理后的数据。

3.2 系统功能实现

根据图3,将系统功能主要划分为三个部分:实时质控、全库质控和数据查询,它们都通过文中算法来处理原有的质控数据,并将处理后的数据展示到客户端的浏览器上。

系统主要分为应用服务器和数据库服务器两个组成部分,前者部署业务应用系统,后者部署各种数据库管理工具。应用服务器与数据库服务器是分离的。该业务系统的运行环境为已安装、配置 JDK 1.7 的 Windows 7 以上操作系统,中间件服务选用 Tomcat 6.0;数据库服务器的运行环境为 Windows Server 2008 以上操作系统,数据库服务器选用 Microsoft SQL Server 2008 R2,采用集中式数据管理。

系统遵循 MVC 开发原则, 基于多层组件式 B/S 架

构和 SSH 框架,采用 Java 编程语言,整个系统操作简单且人性化,具有良好的可维护性和可拓展性,为 MDOS 业务值班提供了极大的支持。

在日常值班业务中,值班人员只需将系统的后台服务开启,系统会自动监控、处理最新入库的疑误数据,无需额外的人工操作。

系统的主界面如图 4 所示。当用户登录到主界面时,系统后台会自动启动实时质控功能,监控 MDOS 数据库的入库行为,处理当前最新时次的质控数据,并将处理后的结果显示在用户的浏览器上。



图4 系统主页面

图5为全库质控功能界面。当用户导入历史数据库或还原备份数据库后,可点击页面的“处理全部记录”按钮,系统后台会调用 HV-Process 算法,处理当前 MDOS 数据库的区域站疑误数据表中的数据,完成全库质控的操作后,将处理完后的数据展示在前台页面。



图5 全库质控执行页面

根据工作人员的使用习惯,系统还提供了特定时间段的数据查询页面,用户可自定义查询时间段和搜索字段,后台根据用户提交的搜索信息,在不刷新当前页面的前提下,即可获取搜索结果,如图6所示。

4 结束语

为提高南海气象数据的准确性,在分析研究南海



图 6 数据查询页面

区域自动站数据的质控特点及 MDOS 数据质量控制的基础上,提出了 HV-Process 算法。该算法利用双重判断-处理机制来约简质控后的数据,降低了南海区域自动站质控数据的误检率。在此基础上,设计并研发了一套集实时质控、全库质控和数据查询等功能的质控系统。该系统构建完成后,经过了应用演练的检验,各个功能模块运行稳定,有效地提高了值班人员的工作效率。

参考文献:

[1] World Meteorological Organization. Guide to meteorological instruments and methods of observation[M]. [s. l.]:Secretariat of the World Meteorological Organization,1983.

[2] Shafer M A, Fiebrich C A, Arndt D S. Quality assurance procedures in the Oklahoma Mesonet[J]. Journal of Atmospheric and Oceanic Technology,2000,17(4):474-494.

[3] Yan T, Kumar V, Ganesan D. Crowdsearch: exploiting crowds for accurate real-time image search on mobile phones[C]//Proceedings of the 8th international conference on mobile systems, applications, and services. [s. l.]:ACM,2010:77-90.

[4] Alonso O, Rose D E, Stewart B. Crowdsourcing for relevance evaluation[J]. ACM SIGIR Forum,2008,42(2):9-15.

[5] Callison-Burch C. Fast, cheap, and creative: evaluating translation quality using Amazon's Mechanical Turk[C]//Proceedings of the 2009 conference on empirical methods in natural language processing: volume. [s. l.]:Association for Computational Linguistics,2009:286-295.

[6] Bigham J P, Jayant C, Ji H, et al. VizWiz: nearly real-time answers to visual questions[C]//Proceedings of the 23rd annual ACM symposium on user interface software and technology. [s. l.]:ACM,2010:333-342.

[3] Jiménez P A, González-Rouco J F, Navarro J, et al. Quality assurance of surface wind observations from automated weather stations[J]. Journal of Atmospheric and Oceanic Technology, 2010,27(7):1101-1122.

[4] 任芝花,熊安元.地面自动站观测资料三级质量控制业务系统的研制[J].气象,2007,33(1):19-24.

[5] 熊安元.北欧气象观测资料的质量控制[J].气象科技,2003,31(5):314-320.

[6] 周林,李湘.关于美国自动地面观测系统(ASOS)的考察报告[J].气象科技合作动态,2002(4):18.

[7] 王新华,罗四维,刘小宁,等.国家级地面自动站 A 文件质量控制方法及软件开发[J].气象,2006,32(3):107-112.

[8] 任芝花,赵平,张强,等.适用于全国自动站小时降水资料的质量控制方法[J].气象,2010,36(7):123-132.

[9] 赵煜飞,任芝花,张强.适用于全国气象自动站正点相对湿度资料的质量控制方法[J].气象科学,2011,31(6):687-693.

[10] 鞠晓慧,曹丽娟,朱建华.地面自动站气压的台站极值检查方法研究[J].气象与环境学报,2010,26(3):48-52.

[11] Lanzante J R. Resistant, robust and non-parametric techniques for the analysis of climate data: theory and examples, including applications to historical radiosonde station data[J]. International Journal of Climatology, 1996,16(11):1197-1226.

[12] 王海军,刘莹.综合一致性质量控制方法及其在气温中的应用[J].应用气象学报,2012,23(1):69-76.

[13] 何志军,封秀燕,何利德,等.气象观测资料的四方空间一致性检验[J].气象,2010,36(5):118-122.

[14] 黄美林,马建华,李东.基于 SSH 框架与泛型的通用分页方法设计与实现[J].计算机技术与发展,2012,22(1):67-71.

[15] 付更丽,曹宝香. SOA-SSH 分层架构的设计与应用[J].计算机技术与发展,2010,20(1):74-77.

[9] Franklin M J, Kossmann D, Kraska T, et al. CrowdDB: answering queries with crowdsourcing[C]//Proceedings of the 2011 ACM SIGMOD international conference on management of data. [s. l.]:ACM,2011:61-72.

[10] 冯剑红,李国良,冯建华.众包技术研究综述[J].计算机学报,2015,38(9):1713-1726.

[11] 高梦超,胡庆宝,程耀东,等.基于众包的社交网络数据采集模型设计与实现[J].计算机工程,2015,41(4):36-40.

[12] 黄冬梅,许坤,张明华. Entropy-Beta: 用于模式匹配众包方法中的发包策略[J]. 计算机科学与探索,2015,9(7):887-896.

[13] 阮闪闪,王小平,薛小平.基于证据理论信任模型的众包质量监控[J].计算机应用,2015,35(8):2380-2385.

[14] 张志强,逢居升,谢晓芹,等.众包质量控制策略及评估算法研究[J].计算机学报,2013,36(8):1636-1649.