

文章编号 : 1673-8411 (2019)02-0086-04

区域自动站降水分时统计关键技术与实现

黄志, 丘平珠, 谭斐

(广西气象信息中心, 南宁 530022)

摘要: 利用 2008-2019 年广西区域 2500 个自动站小时分钟降水数据, 采用基于多线程技术对自动站小时分钟降水数据中的无降水数据进行筛选, 通过数据预处理剔除冗余数据仅保留有降水的数据。研究表明, 处理耗时较直接通过数据库进行查询读取统计的方式减少 90%, 并且通过对比分析发现, 数据总体可用性高达 95%。由此可见, 高效的处理方式和数据的高可用性, 为广西暴雨洪涝风险普查提供了良好的数据支撑和参考依据。

关键词: 降水; 分时统计; 预处理; 多线程

中图分类号: P412.13

文献标识码: A

Research and Implementation of Precipitation Timesharing Accounting Key Technologies for Regional Automatic Station

Huang Zhi, Qiu Pingzhu, Tan Fei

(Guangxi Meteorological Information Center, Nanning 530022)

Abstract: In this paper, based on multi-threading technology, the precipitation-free data of automatic station hourly and minute precipitation data are screened, and the redundant data are eliminated through the pretreatment process, only the precipitation data are retained. The processing time is reduced by 90% compared with directly querying through the database, and the overall availability of the checked data is up to 95%. The efficient processing method and the high availability of the statistical results provide good data support and reference basis for the storm flood risk survey in Guangxi.

Keywords: precipitation; time-sharing statistics; preprocessing; multithreading

引言

广西地处亚热带季风气候区。主要特征是夏天时间长、气温较高、降水多, 冬天时间短、气候干暖, 年均降雨量 1835mm, 并且全区各地降水分布极为不均, 从 1100-1900mm 不等, 给每年汛期的暴雨洪涝防灾减灾工作带来极大挑战^[1]。

为应对暴雨洪涝对广西经济建设以及人民生命财产的影响与侵袭, 广西暴雨洪涝风险普查工作于 2013 年启动, 其中一项重点工作就是对区域自动站的有记录以来的分钟数据根据不同的统计时效, 计算统计其降水极值以及出现时间, 其中统计时效分为 10min、30min、1h、3h、6h、

12h、24h 共 7 个。计算统计的结果将为暴雨洪涝预报业务与研究提供数据支撑, 为防灾减灾提供数据参考依据, 因此此项工作意义重大。

与 91 个建站年代久远、数据完整的国家基本基准站相比, 区域自动站的建站时间较短且站点数量大, 缺少与之匹配的完整的 AJ 月审核数据报表, 原始的区域自动站小时分钟降水数据只能从自动站数据库中进行提取, 经过近 10a 的积累, 区域自动站小时数据已经达到了亿数量级, 如果采用对数据库直接进行查询读取处理并获取统计结果的方式, 会使得数据库连接通道长时间被占用以及大数据量的 IO 读写导致程序运行异常和频繁掉线, 常规处理方式给此项工作增添了极大的困难和挑战。

收稿日期: 2019-03-28

基金项目: 广西区气象局气象科研项目“基于 CIMISS 的广西天气气候监测与评价系统”(桂气科 2017M03)

作者简介: 黄志(1981-), 男, 硕士研究生, 高级工程师, 主要从事气象信息化与大数据研究与应用。


```

ssda:=strtodate(statmp)+i;
datetimetosting(sdate, ' yyyymmdd' ,ssda);
for y:=0 to 23 do begin
if y<10 then
sdate:=sdate+' 0' +inttostr(y);
if y>9 then
sdate:=sdate+inttostr(y);
sql:='select StationNum, ObservTimes, PrecipitationAmount from tabAWSTimeData where
StationNum=''' +sta_N+'''' and ObservTimes=''' +sdate+'''' ;// 拼接 sql 查询语句获取遍历小时分钟降水数据
form1.NumADOQuery.Close;
form1.NumADOQuery.SQL.Clear;
form1.NumADOQuery.SQL.Add(sql);
form1.NumADOQuery.Open;
if (form1.NumADOQuery.RecordCount>0) then
begin
.....
minirain:=form1.NumADOQuery.FieldByName('PrecipitationAmount').AsString;// 获取小时分钟
降水数据
.....
for n:=0 to 59 do begin
comtmp:=copy(minirain, n*2, 2);
comtmp1:=comtmp1+comtmp;
if (comtmp='00') or (comtmp=' //') or (comtmp=' 0/') or (comtmp=' /0') then
// 判断每分钟的降水量是否为 0
.....
TextGroup.SaveToFile(extractfilePath(application.exeName)+' data/' +sta_N+' .txt' );
// 预处理完成, 将非零小时分钟降水数据集流转为文本

```

2 算法设计与实现

完成前期的数据预处理, 为之后的数据统计精简了大量无用的冗余数据, 接下来的工作就是针对每个站点生成的小时雨量文件进行处理统计, 计算统计其降水极值以及出现时间, 并进行汇总, 以 10min 时效的降水极值统计和出现时间为例就其算法做分析。具体的流程如图 3 所示。

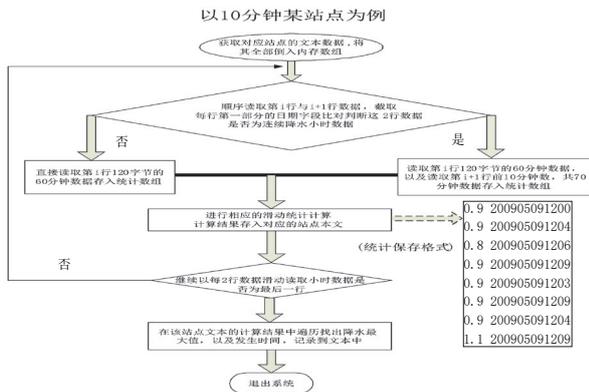


图 3 分钟降水极值处理统计流程

由图 2 可以看出每行数据的小时时间时序连续和中断参差不齐, 以 10min 时效的降水极值统计为例, 任意一行从 1-50min 的统计可以在本行数据中进行; 但是从第 51-59min, 如果下一小时为连续时序, 则需要读取下一小时前 10min 的数据参与统计, 所以此时需要判断第二行数据的起始时间是否连续, 如不连续则对 51-60min 降水则逐个累加即可; 如为连续小时降水, 则需提取第二行前 9min 数据追加至第一行数据末尾变换为 60+9=69min 的数据格式进行统计, 这些操作可在数据内存中进行, 统计完成则释放内存。

此次类推, 30min 降水时效则需一次读取 2 行数据判断是否连续, 1 个小时统计时效需读取 2 行, 从 3h 开始, 每次读取的行数 = 统计时效 * 2 倍数读取, 如 3h 统计读取 6 行, 9h 读取 18 行, 12h 读取 24 行, 24h 读取 48 行, 通过判断所读取的每行数据开头的年月日时字段是否为连续降水时序, 决定导入内存数组的行数。在统计中发现 12h 和 24h 全为降水连续时效出现概率较小。

每个站点在处理过程中都保存了每次连续时

序的统计数值, 之后通过遍历查找统计记录得出极值结果和与之对应的日期, 抽取部分数据与临近站点的数据根据相关性来检验其有效性, 并通

过与历年所属区域的国家基本站的降水最大值出现日期比对分析, 确保提取的最大值准确可靠。以下图4为各个时效的部分统计结果汇总。

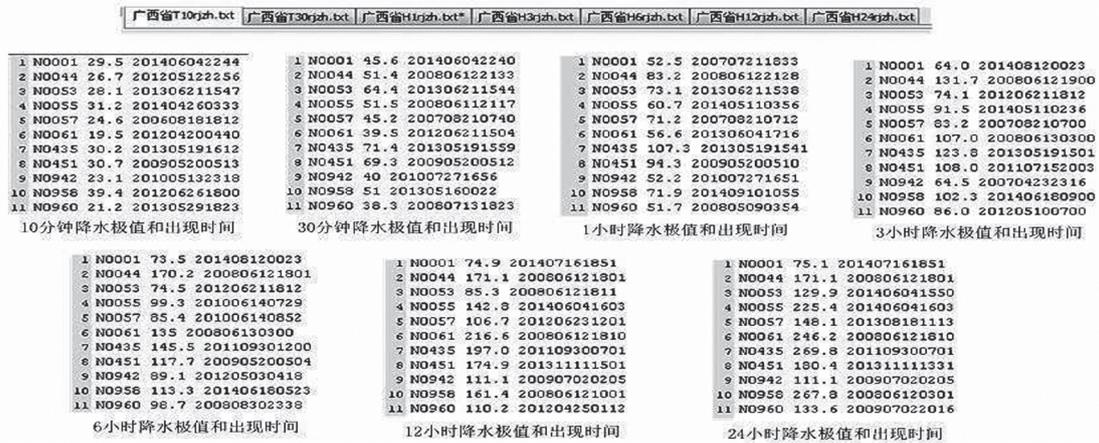


图 4 各个时效的部分统计结果汇总

3 处理效率与结果分析

在实际的处理过程中, 采用 3 路线程并发处理的耗时较直接通过数据库进行查询读取统计的方式减少 90%, 处理效率大大提高, 数据的统计处理工作可以顺利开展。经过对处理之后的统计结果进行分析比对, 所有站点各个降水时效的极值与出现时间大部分集中在 6、7、8 三个月, 少数出现在春夏和夏秋过渡期间, 这与广西总体的气候降水特征基本符合, 统计结果与实况结果的比对表明其数据总体可用性高达 95%, 同类数据较国家基本站稍微低但基本持平。

4 总结

采用数据预处理减少冗余数据, 数据处理对象转变成了 txt 文本, 无需考虑程序因数据库检索长期占用数据库资源导致的锁死和异常掉线问题, 程序运行的鲁邦性大大提高; 并且处理程序的线程数可以根据实际的站点数和数据量进行增加和减少, 模块化设计特征显著。

区域自动站小时分钟在不同统计时效下的降水极值统计工作填补了广西区域自动站分钟降水精细化统计分析的空白, 为防洪减灾和经济建设提供了有效数据支撑, 随着区域自动站的不断扩建和数据传输频次的提高, 此项工作将会与现有的 CIMISS 系统、MODS 系统以及历史 AJ 文件报表做进一步扩展和延伸。

参考文献:

- [1] 黄雪松, 周惠文, 黄梅丽, 等. 广西近 50 年来气温、降水气候变化 [J]. 气象研究与应用, 2005, 26(4): 9-11.
- [2] 黄志, 高钰杰, 廖伟平. 基于多线程的自动站数据处理系统 [J]. 成都信息工程学院学报, 2012, 4(27): 192-195.
- [3] 曾行吉, 李莹, 宋瑶. 地面气象观测数据入库多线程并行设计与实现 [J]. 气象研究与应用, 2017, 38(3): 114-116.
- [4] 黎颖智, 史彩霞, 刘世学. 关于异步多线程快速提取 CIMISS 数据入库方法的研究与应用 [J]. 气象研究与应用, 2018, 39(1): 118-120.
- [5] 张金标, 张恩红. 基于多线程流水线的光盘自动刻录技术研究 [J]. 气象研究与应用, 2018, 39(2): 94-97.