

文章编号:1006-7639(2005)-01-0061-04

西北干旱区沙尘暴数据库系统介绍

陈旭辉¹,陆登荣¹,冯建英²

(1.兰州中心气象台,甘肃 兰州 730020;2.中国气象局兰州干旱气象研究所,甘肃 兰州 730020)

摘 要:本系统应用微软 access 数据库系统建立了西北沙尘天气数据库,采用 C/S 结构在数据库环境下编程,设计了数据挖掘系统(数据挖掘机)、资料维护系统、数据接口系统。数据挖掘机实现了一般用户在不了解数据格式的情况下,通过可视化的界面按任意条件对资料进行查询、统计,并通过各种算法对资料进行处理,生成用户需要的资料序列;数据接口系统实现将查询结果以图形、文本文件、绘图格式文件等多种格式输出;资料维护系统实现资料的追加及维护等功能。

关键词:沙尘暴;数据库;客户服务器;数据挖掘

中图分类号:TP311.132.4

文献标识码:A

前 言

长期以来气象地面观测资料都是以文件的方式管理,如 A 格式、D 格式、W 格式等,这些资料均为特定格式的文件,只有了解资料存储格式和观测规范的专业人员才能对资料进行处理,所以资料处理成为沙尘暴研究中的一大障碍,使科研人员的许多“灵感”因不能获取某种资料而不能实现。本系统采用数据库技术,对沙尘资料进行存储和管理,并采用数据挖掘技术,为用户提供了一个可视化的分析、查询平台,使用户在不了解资料存储细节的情况下,定制自己所需的资料。

1 沙尘数据及本系统中几点约定

本系统使用的气象资料均来自地面月报表 A0 文件。对于沙尘天气,本系统有以下几个约定。

(1) 沙尘天气定义及标准按中国气象局 2002 [13]号文件执行。

(2) 区域性沙尘暴标准。分别以出现沙尘天气的 n 个站为圆心,画 n 个半径为 RA 的圆,若这 n 个圆中,只要有一个圆内能包含 m ($m \leq n$) 个站,则认为本次沙尘天气为区域性沙尘天气(图 1)。其中 RA 表示区域的范围大小,1.9 m 为区域站数标准(中国气象局规定 m 为 3,且为基本基准站)

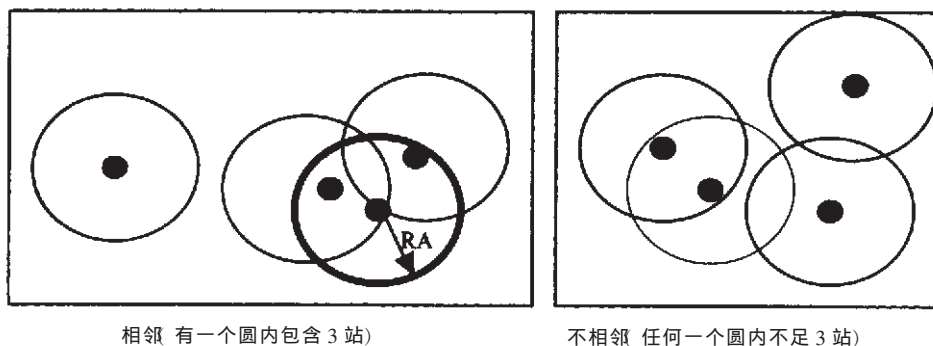


图 1 3 站相邻示意图

Fig.1 The sketch map of regional sandstorm

收稿日期:2004-12-01;改回日期:2005-02-25

基金项目:中国科技部项目 西北干旱区沙尘暴预警、服务系统研究”01 子专题 2000DLA1003)资助

作者简介:陈旭辉(1968-),男,甘肃渭源人,工程师,从事管理工作。

(3) 时间连续标准:对于沙尘天气过程,若连续几天出现,中间没有中断,则认为是时间连续的。

2 数据库设计

建立数据库主要问题是如何有效地规划数据库,使其既符合关系型数据库的模型,又能减小数据冗余、提高执行效率。本系统对沙尘数据进行分析,根据其使用频度、信息量、数据间关系等将沙尘数据分为沙尘个例库、要素库、图片库、距离矩阵库、工作库等,具体规划如下。

2.1 沙尘个例库

沙尘个例库保存使用频率最高、信息量较小的信息,主要由个列表、文献表、灾情表等 4 个数据库表组成。

2.1.1 台站信息表

台站信息是使用气象资料时必不可少的,但它又不同于气象资料,具有稳定、变化少等特点。为了使用方便,将台站信息收录在一个表中,本表包含了西北 241 个站的省名、站名、站号、经度、纬度、海拔、站类型、建站时间等信息,并通过站号和其它表建立关系。

2.1.2 沙尘天气个列表

沙尘个例是研究沙尘天气时使用频率最高的资料,为了提高查询速度,将沙尘天气个例资料放在一个表中,由日期、站号、等级、开始时间、结束时间、持续时间、最小能见度、最大风速、最大风速的风向、大风等 16 个字段组成,这个表包含了研究沙尘天气常用的要素。

2.1.3 文献表

文献表保存了沙尘暴方面的一些文献,由刊物、文章名称、卷号、期号、发行时间、关键词、摘要、作者、作者信息、页码等 10 个字段。

2.1.4 灾情表

灾情表收集了西北区历史上重大沙尘暴天气造成的灾情情况,由开始日期、结束日期、地名、灾害类型、灾情等 5 个字段组成。

2.2 要素库

此数据库只有一个表,保存了分析沙尘天气常用的温度、降水等 10 个要素的逐日资料。单独建库主要是为了克服 access 数据库容量不能超过 1G 的限制。

2.3 图片库

由历史天气图扫描得到的电子化天气图及典型

个例的分析图组成,共计 200 多张。

2.4 距离矩阵库

在区域分析时,要用到任意两站之间的距离,为了提高计算速度,将 241 个站相互之间的距离提前计算后生成一个距离矩阵(241×241),该矩阵是一个三角矩阵,实际有用的信息有(241×240/2)个,为方便读取,保存为二进制文件。

2.5 工作库

目前原始资料库已达 700 MB,在统计中会产生大量的临时数据,为了防止在统计过程中造成数据库容量超过 1G,系统专门设计了一个工作数据库,用来保存统计中产生的各种中间结果,其内容在系统运行中随时都在改变。

3 系统平台设计

为了方便用户维护、使用数据库中资料,本系统设计了一个集成的应用平台,该平台具有资料转换、数据挖掘、区域分析、灾情及文献查阅、图片显示、结果输出等功能,其中数据挖掘是本系统的核心,通过它用户可以建立任意条件下沙尘天气的序列。

3.1 资料转换

资料转换用于将特定格式的文件转换为数据库格式,用于建库初期导入资料和后期追加资料。

3.2 数据挖掘

数据挖掘是指从海量的信息中提取有价值的信息,本系统提供的数据挖掘机可以使用户获取任意范围、任意强度和任意时间尺度的沙尘资料。使用时,用户只要通过查询条件生成器输入条件,即可获取所需资料见图 2。系统提供的主要查询、统计功能有:

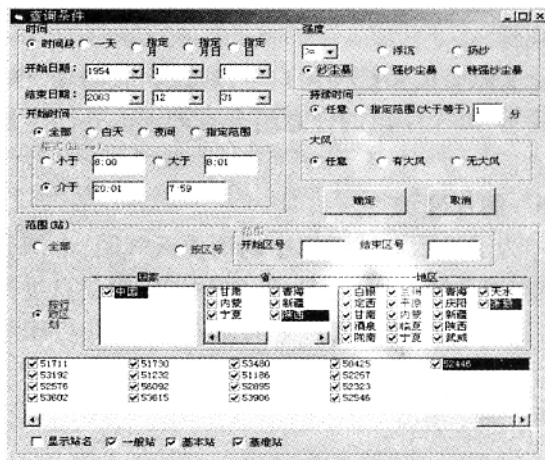


图 2 资料挖掘界面

Fig.2 The interface of date mining

按站统计:可重建各站逐日、逐月、逐年、月合计、年合计、任意时段同期历年值、任意时段同期历年合计值等序列。

按站次统计:可以统计某一范围内,1 d 出现沙尘天气的站数,包括逐日、逐月、逐年、月合计、年合计、任意时段同期历年值、任意时段同期历年合计值等。

按次数统计:可统计某一范围内沙尘日的逐日、逐月、逐年、月合计、年合计等值。沙尘日指某一范围内,某日有 m 个站出现沙尘天气,则为一个沙尘日, m 由用户设置,本系统缺省为 1 站。

按区域统计:对区域性沙尘天气按“地域不连续、时间不连续”、“地域不连续、时间连续”、“地域连续、时间不连续”、“地域连续、时间连续”4 种情况分别按日、月、年、月合计、年合计等进行统计。

3.3 区域分析

对区域性沙尘暴,中国气象局旧的标准中有“相邻 3 站”的要求,但相邻只是一个定性的概念,没有量化,难以在业务中使用。本系统为了解决这个问题,首先提出空间“相邻”的量化标准,并根据业务实际,提出时间连续的标准,实现了按任意的空间尺度标准、时间标准、强度标准进行区域分析,建立区域性沙尘天气序列的功能。为了方便不同用户使用,本系统对任意条件下的沙尘天气可同时进行“地域不连续、时间不连续”、“地域不连续、时间连续”、“地域连续、时间不连续”、“地域连续、时间连续”4 种分析,为业务、科研人员研究沙尘天气的范围、时间等特性提供了选择。

3.4 灾情及文献查阅

灾情查阅系统使用户可以按日期、范围查询西北区的沙尘、大风灾情,结果可直接显示,也可保存为文件。

文献查阅系统使用户可以按期刊、期号、作者、摘要等 10 个字段的任意组合查询国内外沙尘暴文献。

3.5 图片显示

为了显示本数据库中的历史天气图、个例分析图等,本系统采用 ole 技术,对数据库中信息进行还原,实现对图片信息的显示、修改、另存、打印等功能。

3.6 结果输出

对于查询、统计得到的结果本系统提供了表格、曲线等直接显示方式,同时为了方便用户使用,系统可将数据保存为文本文件或 EXCEL 文件、surfer 绘图格式文件等,为用户二次处理资料提供了接口。

4 系统解决的几个关键问题

(1) 可扩展性。系统具有任意范围、任意时段、任意强度沙尘天气的统计功能,故只要有资料,程序不需要做任何修改,就可实现对全国,甚至全世界范围内沙尘天气的分析。

(2) 区域标准划分的微机实现。早期区域性沙尘暴的定义中有“相邻 3 站”的标准,但没有对“相邻”做具体规定,为了便于计算机处理,我们定义:分别以出现沙尘天气的 n 个站为圆心,画 n 个半径为 RA 的圆,若这 n 个圆中任意一个圆内包含 n_1 ($m \leq n_1 \leq n$) 个站,则认为本次沙尘天气为区域性沙尘天气。其中 RA 表示区域的面积大小, m 为区域站数(中国气象局规定为 3)并将这个算法在计算机上实现。

(3) 任意时段、任意范围、任意强度沙尘资料序列的重建。

5 小 结

(1) 采用数据库系统对资料进行管理,可提高数据的安全性,增强资料的可读性、减少编程的工作量,也便于实现资料共享。

(2) 强大的数据挖掘功能是本系统的特点,用户通过本系统提供的功能,可以从库中提取任何有价值的信息。

(3) 本系统具有灵活的资料接口,为用户提供了常用格式文件的输出功能。

(4) 开放性是本系统的一个特点。系统虽然为西北沙尘暴数据库系统,但设计为一个开放的系统,可以用于任何范围,用户只需通过维护台站信息就可实现范围的增减;此外,本系统所有参数如区域性沙尘暴的相邻标准、站数标准等都是可调的,适用于任何地方使用。

(5) 本系统首次提出采用两站之间距离作为判断“相邻”的标准,并通过采用一定的算法建立了面向“区域性沙尘暴”这个主题的数据挖掘模型。

The Sandstorm Database System Introduction in Arid Region in Northwest China

CHEN Xu-hui¹, LU Deng-rong¹, FENG Jian-ying²

(1. Lanzhou Central Meteorology Observation, Lanzhou 730020, China;

2. Institute of Arid Meteorology, CMA, Lanzhou 730020, China)

Abstract: Using the microsoft access database system, a database of sandstorm in arid region in Northwest China has been created. We designed system of data mining (data mining machine), data maintaining, data querying and data interface by C/S programming structure under database environment. When normal user doesn't know data format, the data mining machine can achieve data querying and statistics according to any condition by visual interface, and process data by all kinds of algorithms and come up with user needful data sequence. This system can output querying results by graph, text file, mapping file and so on. The new data can be appended by the data maintaining system easily.

Key words: sandstorm; database; client server; data mining

(上接第 53 页)

Identification Geosynchronous Satellite Infrared Image and Detection Violent Convective Weather System in Flood Season by Textural Analysis Technique

YUE Ping^{1,2}, LIU Xiao-yun², GUO Liang-cai², LU Chang-hong²

(1. Nanjing University of Information Science & Technology KLME, Nanjing 210044, China;

2. Jiuquan Meteorological Bureau, Jiuquan 735000, Gansu, China)

Abstract: Based on the textural analytic theory of gray scale co-occurrence matrix, the geosynchronous satellite IR image analytic system was made. Fifty pieces of FY-2 IR image, which attended by violent convective weather in the west areas of Gansu province in flood season from 2003 to 2004, were selected to compute the inertia, entropy and energy quadrature. A hail case, which happened on 29th June, 2004, was selected to validate the software system. The results show that the geosynchronous satellite IR image analytic system is good for monitoring violent convective weather system, and for the convective weather system, the getparms value of the energy is closed to the value one. On the contrary, the values of the inertia and entropy are closed to zero, the getparms value of the energy is two scalar higher than those of entropy and entropy.

Key words: texture; gray level co-occurrence matrix; IR image; convection; weather system