

天气预报文本自动生成系统的天气图编辑模块的设计与改进

曾庆辉¹⁾ 姚天¹⁾ 雷小途²⁾ 姚志展²⁾

(1) 上海交通大学计算机科学与工程系, 上海 200030; 2) 上海市气象局, 上海 200030)

摘要: 天气预报文本自动生成是天气预报科学化、客观化、高效化的重要途径, 借助于天气图编辑器, 可使预报准确性得到提高。本文介绍了天气图编辑软件的设计思想和方法, 以及在本系统原框架基础上, 对其主要功能模块实现方法的改进和完善, 并对图形编辑器与文本生成器之间的数据接口作了介绍。

关键词: 天气图; 图形编辑; 等值线; 等值圈

中图分类号: TP391.4 **文献标识码:** A

近年来, 随着计算机科学的发展及其在天气预报上的应用, 数值天气预报业务化逐渐成为现实。在未来几年, 预报员对数值预报产品的理解、分析与修改所形成的人机交互天气分析预报技术必将得到迅速发展与应用。传统的天气图表在纸面、书面上表达形式将逐渐被计算机图形(像)多功能显示及多功能复制再现方式所替代。但是到目前为止, 以自然语言理解和文本生成技术为基础的天气预报文本自动生成技术在我国仍未得到应用。

笔者所在课题组研究、设计了一个天气预报文本自动生成系统^{1,2)}, 借鉴加拿大的 FoG (Forecast Generator) 系统³⁾。此系统针对天气预报客观化、自动化和预报实时性、文本多语种性、文本形式的灵活性、以及用户使用的方便性等要求进行设计。针对目前气象数据量大、复杂, 在人机交互条件下对气象信息的显示、调整和修改比较困难的特点, 本系统设计了天气图处理模块, 用于将原始天气数据以等值线图方式显示, 并提供工具使预报员能够直观地调整图形, 并修改数据。使数值预报和预报员的经验有机地相结合, 进一步提高天气预报的准确性。当图形编辑完成后, 预报员可以对图形所对应地理区域的任何地点进行预报, 实际操作是向数据库添加记录, 供文本生成模块读取相应天气数据。

1 天气图编辑器的框架和流程

本系统是以图形化方式帮助预报员理解、分析与修改等值线天气图的人机交互气象预报天气图编辑软件。系统在原来能较好进行中心点和等值线编辑的单要素框架¹⁾上改进和完善而成。天气图编辑器的框架如图 1 所示。

收稿日期: 1998-10-16; 改回日期: 1998-12-08

基金项目: 国家自然科学基金(69673008)、上海市科委科技发展基金(962907002)和德国大众(Volkswagen-Stiftung)资助项目

第一作者简介: 曾庆辉, 男, 1975 年 12 月生, 硕士生

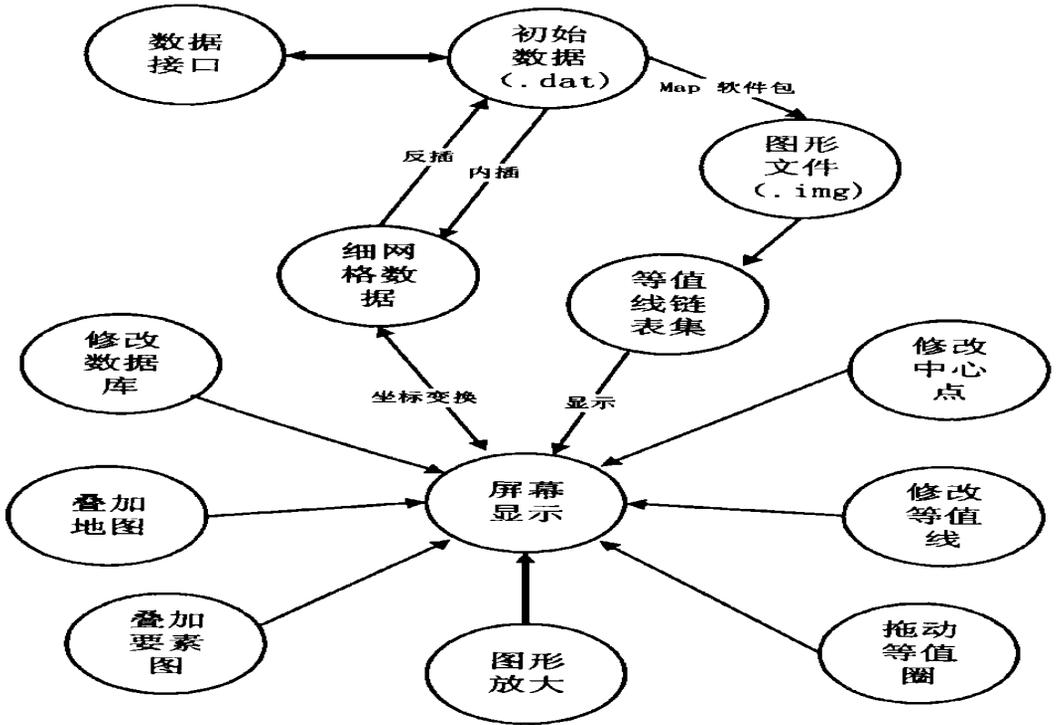


图1 天气图编辑器框架图

Fig. 1 Framework of weather chart editor

系统流程为:

(1) 天气图形的生成和显示: 指系统从文件中读取原始网格数据, 利用专门的图形文件生成器(Map)生成要显示的图形文件, 根据图形文件显示在用户屏幕上(图2)。



图2 天气图形的生成和显示

Fig. 2 Formation and display of weather chart

其中Map是在NCAR软件包的基础上修改而成的图形生成器, 图形文件的存取方式为直接存取, 以压缩形式存放图形数据记录长为200字节, 每4字节(即32位二进制)表示一个坐标点、颜色抬落笔信息, 每一个记录包含50个坐标。同一等值线的坐标用直线(相同的颜色)互相连接起来, 通过抬落笔信息区分是否属于同一等值线。所有等值线整个图形文件的等值线集。记录形式为:

15 12 11 0 15 12 11 0
 笔 X 坐标 颜色 Y 坐标

屏幕图形的内部数据结构和生成算法见文献[1]。该部分对应框架图中‘初始数据’、‘图形文件’、‘细网格数据’、‘等值线链表集’四个模块。

(2) 天气图编辑: 图形生成后, 预报员利用编辑工具对图形进行编辑。为此, 首先要将数据文件的粗网格数据转化为与屏幕坐标相对应的细网格数据, 采用的内插公式是唐纳德公式(Donald M Monro)^[4]; 修改中心点、修改等值线及拖动等值圈是图形编辑的主要方法, 并且能

在放大情况下进行, 地图和要素图的叠加用于让预报员以当前正编辑的天气图(某种要素图, 如温度图)相对应的地图及其他相关要素图(如气压图、湿度图等)作为参照来编辑(图 3)。

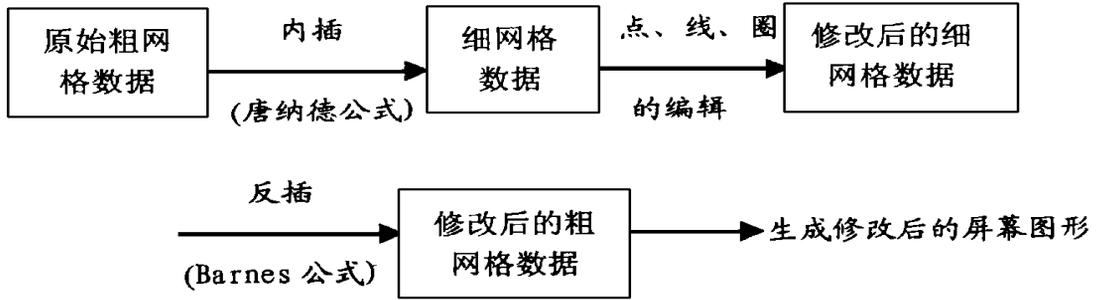


图 3 流程中数据格式的变化

Fig. 3 Data format shift in the data flow

以下是对唐纳德公式利用三次曲线拟合函数, 进行网格内插的方法:

$$f(x) = a + bx + cx^2 + dx^3 \quad 0 < x < 1 \quad (1)$$

指定区间 $[j, j+1]$, 往该区间内插, 若给定 $f(j)$ 及 $f(j+1)$, 并给出此处的导数 $f'(j)$ 和 $f'(j+1)$, 则上式各系数可由下式得出:

$$\begin{aligned} a &= f(j) & b &= f'(j) \\ c &= 3q - f'(j+1) - 2b & d &= b + f'(j+1) - 2q \end{aligned}$$

其中: $q = f'(j+1) - f'(j), 0 < j < 71$

通过公式 (1) 可以得到 j 与 $j+1$ 之间的细网格点的值 $f(j+x)$:

$$f(j+x) = a + bx + cx^2 + dx^3 \quad 0 < j < 71, 0 < x < 1$$

由细网格数据向粗网格数据反插是利用 Barnes 公式^[5], 由图 4 所示, A 点是粗网格上的点, 它的值是由它周围的细网格点 1~24 通过 Barnes 公式而来。这样就可以把 480×480 的细网格数据返回到 71×71 的粗网格数据。

$$A \text{ 点的值} = \left[\sum_{n=1}^{24} \frac{S_n}{d_n^n} \right] / \sum_{n=1}^{24} \frac{1}{d_n^n}$$

S_n 为 1~24 各点的值。 d_n 为 1~24 各点到 A 点的距离

(3) 修改数据库: 天气图编辑完成, 预报员存盘后往数据库添加记录, 才能使编辑有效。添加的记录包括预报地点、发布时间、当前时间, 以及温度、气压、降雨等要素的数据信息。此后文本生成模块就可以读取数据库中的数据, 根据规则库中的转换规则将它转化为预报文本。数据库采用 Microsoft Access 97 构造, 它以 Microsoft Jet 为数据库引擎, 是在 Windows 环境下运行的功能强大的关系数据库, 支持 ODBC (Open DataBase Connectivity 开放数据库互联) 和 DAO (Database Access Object 数据访问对象)¹⁾ 技术。利用 Visual C++ 面向对象的程序设计技术, 以 DAO 作为程序与 Access 数据库的接口操作对象, 控制数据库的访问和输入/输出。

2 功能模块的设计和实现

(1) 天气图的生成及显示模块。当用户做

1) Visual C++ 5.0 programming 联机帮助

File/Open 时,选择文件(.dat)读取原始数据,其格式为每个文件有 44 层,每层 71×71 的网格数据。其中包含了温度、气压、湿度、风力、风向等多种气象数据。根据各要素数据在文件中的偏移量读出待编辑的要素数据后,调用 Map 软件包生成该天气要素的等值线图形文件(.img),然后在屏幕上显示出来。

(2) 数据的转换。根据原始初值数据(71×71 的粗网格数据),通过唐纳德三次曲线拟合公式生成 480×480 的细网格数据,其方法是:首先把 71×71 内插成 71×480 (步骤一),然后再把 71×480 内插成 480×480 (步骤二)。此网格数据通过一定的坐标变换与屏幕坐标相对应。与此相反,细网格数据通过 Barnes 公式再生成粗网格数据。

(3) 中心点的修改。它是指对图形区域内某点的值进行修改,它的改动势必影响以该点为中心的周围某一区域点的数值。具体做法是:当用户用鼠标点取某一中心点,系统将该点的值显示到对话框中,并提供输入接口使用户输入新的值,由此得到该点原值与新值的一个差值,然后在某一影响半径内利用线性经验插值法对周围的点进行修。

(4) 等值线的修改。它是指某一段等值线的修改。等值线的修改不能跨过另一条等值线。它的改动势必也会影响周围某一区域点的数值,这些值也通过 Barnes 公式再进行修正。

(5) 等值圈的拖动。选中某一等值圈,对它及其内区域进行平移拖动。它的拖动将会引起目标位置、原位置及周围某一区域点的数值的改变,这些值也通过 Barnes 公式再进行修正。

(6) 地图的叠加。选择某一块与该天气图经纬度相对应的地图,把它叠加到天气图上,用来获得编辑天气图时的地理参照。

(7) 天气图的放大。选择某一中心位置对原有的地图进行非失真的矢量放大,而不是位图的直接放大,使放大后图形的正确率大大上升。同时保证放大时仍能进行点、线和圈的操作。

(8) 多种要素图的叠加和删除。各天气要素之间是相互关联的,把其他要素图和正编辑的天气图叠加显示,有助于预报员判断与修改。新叠加的要素图只能起参照作用,不能被编辑。

(9) 数据接口。往数据库中写入从数据文件中提取的数据,为文本生成模块提供数据。

原框架实现了(1)~(7)这七个模块。其中第一模块中天气图的生成部分由于原始数据文件的变化进行了改写。(2)~(4)这三个模块原来实现正确无误,基本上没有变化。其中中心点和等值线模块较好地考虑了效果和速度。

第(5)模块在文献[1]介绍的实现方法中存在很多问题:

1) 等值圈选择时用距离误差的方法,如误差允许太小,则不容易选中等值圈,太大则可能出现多个等值圈在误差范围内。现在则采用离单击点最近的方法来改进;2) 等值圈拖动时会擦除没选中的等值圈,现采用了异或的方法,则不会产生上述情况;3) 拖动时只把目的圈上点迁移,但圈内点没有迁移,而且原等值圈内也无变化。基于上述情况,第(5)模块重新进行了设计。

第(7)模块原来实现较差,在新系统中做了如下修改:

1) 原系统只能进行局部区域放大,采用矢量放大技术修改后,允许对全图进行放大;2) 原

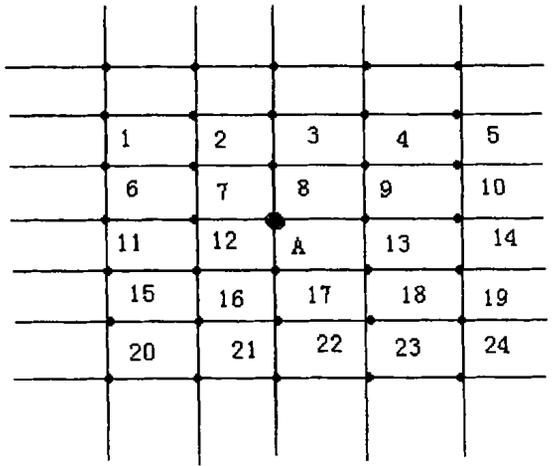


图 4 由细网格数据向粗网格数据反插

Fig. 4 Reverse interpolation from fine-grid data to coarse-grid data

系统在放大情况下不能进行操作,新系统的设计允许在放大情况下对图形进行操作(包括点、线、圈的操作);3)在具体实现中,原来采用在中心点、等值线和等值圈中分别进行分情况的方法,代码冗余烦琐,为此也进行彻底重写。

(8)、(9)模块则是为了完善功能和改善界面。多种要素的叠加显示是为了预报员更好地理解并且能对当前活动要素进行编辑。数据接口使本系统的实际应用成为可能。

下面分别介绍改进和完善的等值圈的拖动、叠加要素图及数据接口。

2.1 等值圈的拖动

等值圈的拖动在效果上要求达到整个圈迁移到目标位置,包括被选中的等值圈内的所有圈。原位置上的一系列圈则消失。该圈拖动过程中跨越的所有圈也应作出相应变化。在内部数据上则表现为被选中的等值圈内的所有数据(原区域)覆盖目的区域。所以首先要得到被选中的等值圈内的点,并将其移到目的点,同时对原来的点进行重新计算,然后计算对周围产生的影响。

(1) 判断落在圈内的点

该算法与多边形扫描算法^[6]类似。

因为等值圈链表由一系列点组成,各点间用直线连接,实际上是顶点数很多的多边形。首先讨论扫描线与多边形的关系。如图5所示,一根扫描线穿过多边形,只要决定这根扫描线中哪一段在多边形内(x 从2到4和9到13),哪一段在多边形之外(x 从0到2、4到9和13到最大 x 坐标)。对每根扫描线重复这一过程,就可以把多边形及其内部区域判断出来,这里多边形是按顶点坐标给出的。

判断扫描线上的点在多边形内可根据多边形区域的连续性避免逐点的进行判断,分为3个步骤:

1) 求出某一扫描线与多边形所有边的交点;2) 把这些点的坐标值以升序排列;3) 对每一对交点间的区域进行填充。

以图5为例,求出交点坐标,并排序 x 坐标得到的表是(2, 4, 9, 13),然后对交点2与4之间,9与13之间的所有点进行填充。

上述3步中第1步需要进行少量计算,第2步十分简单,第3步填充过程始终从奇数个交点出发到偶数个交点为止。但有一个例外,即当扫描线正好通过局部极值点坐标时,若按常规进行便会出错,这时需把这个交点两次存入排序过的交点表。对图5,当扫描线 $y=7$ 时,它与多边形的交点有3个(2, 7, 13),若按常规处理则在2到7,13到扫描线结束之间进行填充。如把顶点(7, 7)坐标两次存入则交点表为(2, 7, 7, 13),这样便从2到7,7到13之间进行填充。局部极值点的判别可通过该点的左右(或上下)两邻点落于扫描线的同侧或异侧来判断。

现在考虑交点的计算。结合等值圈的实际情况,求交点时采用对原等值圈进行插值优化的方法。插值优化后等值圈链表即记载了所有应有的交点。求出这些点后,将这些点按升序排列。一般采用将链表重组的方法。遍历整个链表, y 值相同的按 x 值升序形成一子串,各子串又按

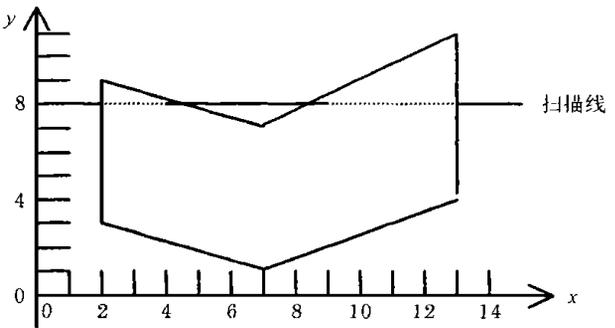


图5 多边形扫描线算法示意图

Fig. 5 Schematic diagram for algorithm of polygonal scan_line

y 值升序形成总串。在每条子链中, 从奇数个交点出发到偶数个交点, 依次求出其间的点。插值优化的算法如图 6 所示。

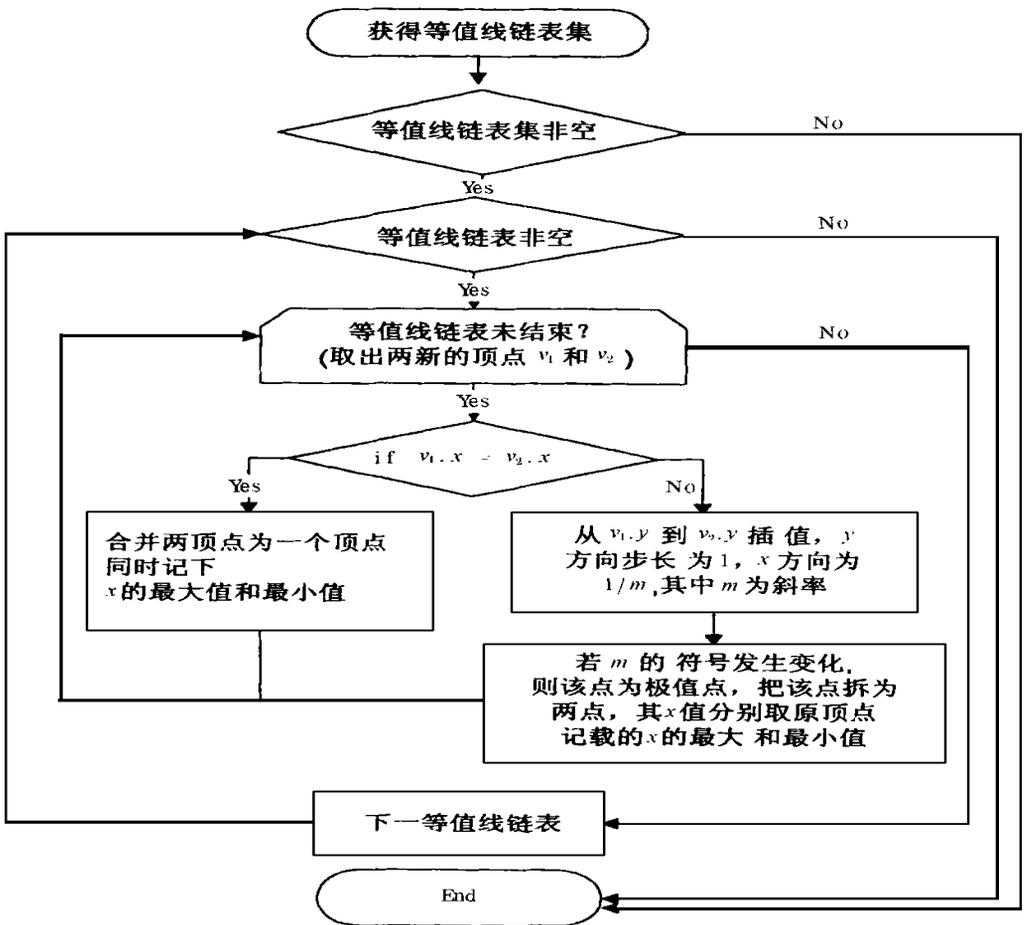


图 6 判断等值圈内的点

Fig. 6 Determining the points within isoplethic loops

(2) 圈内点的迁移

根据要求圈内点按原值搬到目的圈, 目的圈可由被选中的原来的等值圈加上偏移得到。目的圈的值很容易确定, 而原来圈内的值应根据周围环境重新进行计算, 有两种处理方法, 一种是用邻近的一固定点的固定值进行填充, 这样在移去区域可确保无新的等值圈出现; 另一种是对区域内每一点, 采用五点格式或九点格式^[7]平均方法重新计算。为简便起见, 我们采用了第一种。

(3) 对周围产生的影响

等值圈拖动时跨越的区域因为拖动而改变, 拖动范围比较大时, 其变化很难控制。考虑到实际操作时一般只是局部区域内小范围的修改, 所以以拖动幅度不是很大为前提, 这种情况可以用目的区新值与原值的差对周围的影响来近似逼近, 用 Barnes 公式计算影响值。

在实际操作过程中要注意先后次序, 我们采用的方法是先对原值进行复制, 对原区域填入新值后, 对目的区域填入复制值。此时目的区域周围点的值已发生变化, 在目的区域的影响半径内, 对所有点用 Barnes 公式计算拖动对该点的影响值。此算法可参照等值线的改动算法, 只

不过此时点集变为目的圈内的所有点。

2.2 多种要素的叠加

为了使天气预报员更好地理解天气要素图,允许多种要素叠加显示。多种要素的叠加主要是数据结构的增加,这里采用了一个数组 LineSet 来保存每种要素的各种数据结构,若该种要素没被选中,则对应指针为空。

在初始化时要求用户选择欲编辑的要素,为了使图形编辑清楚而不和其他要素图相混淆,本系统实现中允许一次只对一种要素图进行编辑。如要编辑其他要素图,需先退出本次编辑,打开欲编辑的要素天气图。用户可以任意叠加和删除非编辑的要素天气图。

主要的数据结构有:

视类中有:

```
int m- nWhich; // 表明当前选中编辑的要素
```

```
LINESET * LineSet[ 45]; // 若非空,存储了该种要素的等值线链表集
```

在视类的重画消息处理函数中,若 LineSet 中对应的指针非空,则显示该种要素。当前显示但不能编辑的要素只在 LineSet 中保存了用于显示的等值线链表结构,而当前唯一能编辑的要素则在 m- array, d- array 数组中存放了其粗网格数据和细网格数据。当用户按了叠加和删除按钮时,若该种要素对应应在 LineSet 中的指针为空,则为叠加操作,否则为删除操作。叠加的要素不能编辑。

2.3 数据接口

数据接口实际上包括两部分,一部分是来自气象台的原始数据与图形编辑模块之间的接口,另一部分是图形编辑模块与文本生成模块之间的接口。前一接口即是天气图的生成与显示(见文献[1]),以及编辑完成之后的存盘。后一接口要求往数据库中添加记录,它根据用户(预报员)指定的预报地点,从经纬度表中获得该地点的经纬度,并在天气图文件中获取相应经纬度的各天气要素(如发布时间、发布日期、降雨量、风向、风力、温度等)的数据。将这些数据加入基本表(weather. mdb)中(如表1)。在具体实现上,采用 DAO 作为程序与数据库的接对象,并且通过用户界面使得对数据库的各种操作(插入、修改、检索)均能有效进行。这里需指出的是所有地点的经纬度已预先建成一张表,每个天气要素数据文件也包含有一个相对应的地理经纬度数据文件,指出该天气要素数据所表征的地点的经纬度。

表 1 存放天气要素数据的基本表

Table 1 screenshot of the database for weather elements

ID	Region	ReleaseTime	ReleaseDate	RainFall	WindDirection
1	上海	09:00:00	1998-05-05	20.230 5	46.346
2	浙江南部	09:00:00	1998-05-05	15.653 4	60.579
3	上海	12:00:00	1998-07-31	27.106 4	220.473
4	苏州	12:00:00	1998-07-31	12.503 1	234.744
5	苏州	13:00:00	1998-07-29	12.503 1	247.145
6	上海	13:00:00	1998-07-29	27.106 4	230.473
7	上海	15:00:00	1998-08-05	27.106 4	120.023
8	苏州	15:00:00	1998-08-05	12.503 1	134.641

3 结 语

本天气图编辑器是基于国内气象预报技术发展的现状,作为天气预报文本生成系统的一

个前端模块而设计的。它的实现使数值预报的精确性与预报员的经验有机结合,为预报员提供了一个分析和修改天气图的良好的人机交互环境。使文本生成模块提供的预报数据更合理、更可靠,并使生成的预报文本与天气实况更接近。本天气预报文本生成系统曾于1998年8月在加拿大第九届自然语言生成国际研讨会上作过演示,获得与会专家的好评。另外,本天气图编辑器只需做适当修改,也可按照预报员的输入,画出天气图,进而生成预报文本。这种预报思想也可能出现在未来的天气预报中。

参 考 文 献

- [1] 姚天昉,赵敏,袁招洪,等.一种用于天气预报文本生成系统的天气图编辑方法[A].成都气象学院学报,1998,13(3):222-230
- [2] HUANG XIAORONG, YAO TIANFANG, GAO GUODONG. Generating chinese weather forecast with stylistic variations[A]. 17th international conference on computer processing of oriental language[C], Hongkong, 1997. 733-738
- [3] ELI GOLDBERG, NORBERT DRIEDGE, RICHARD KIHREDGE. Using natural language processing to produce weather forecasts[A]. IEEE experts 9, 1994. 45-53
- [4] 王继志,汤桂生.气象图形显示原理与方法[M].北京:科学出版社,1991.48-51
- [5] THOMAS W SCHLATTER. 临近预报和数值预报中气象资料客观分析方法的过去和现在[M].刘晓东译.气象科技,1990,(2):11-18
- [6] 俞勇,盛焕焯.计算机图形学——原理与实践[M].上海:上海交通大学出版社,1995.40-43
- [7] 王刚.偏微分方程数值解法——简明教程[M].呼和浩特:内蒙古大学出版社,1997.117-124

OPTIMIZATION AND IMPROVEMENT OF THE GRAPHIC EDITOR FOR WEATHER FORECAST TEXT GENERATION SYSTEM

ZENG Qing-hui YAO Tian-fang LEI Xiao-tu YAO Zhi-zhan

(Department of Computer Science and Engineering, Shanghai JiaoTong University, Shanghai 200030)

(Shanghai Meteorological Bureau, Shanghai 200030)

Abstract: The auto-generation of weather forecast text is an important way for the scientificness, objectiveness, and high-effectiveness of weather forecasts. With the aid of the graphic editor of weather charts, the auto-generation can improve the accuracy of the forecasts. This paper presents the design thought and method, the optimization and improvement of the implementation methods of the main functional modals based on the initial frame of this system, and as well as the data interface between the graphic editor and the text generator.

Keywords: weather chart; graphic editing; isopleth; isoplethic loop