

文章编号: 2095-2163(2019)03-0228-04

中图分类号: TP311.52

文献标志码: A

基于航班数据可视化系统的设计与实现

闻雅, 高志远, 王吉富, 蔡雨轩, 高晟珍, 李瑞改
(东北林业大学 信息与计算机工程学院, 哈尔滨 150040)

摘要: 随着社会的快速发展,飞机成为目前人们生活中重要的出行方式,改变了人们的生活,便利出行,但是随着高效率时代的到来,时间以及安全成为了人群衡量交通的重要因素。飞机晚点的现象成为了一个影响大家出行,制约行业发展的突出现象。此航班数据可视化系统想法来源于第七届中国软件杯大学生软件设计大赛,同时作为东北林业大学大学生创新创业训练计划项目,本系统能够查询不同地区的航班航线,以及对应航班的晚点率,开发这个系统对民航行业有个更准确清晰的目标,能够更好地针对一些晚点率较高的航班并采取相应的行动对其加以控制,同时,可以给出行群众提供参考。

关键词: 航线; 晚点率; 航班数据可视化系统

Design and implementation of flight data visualization system

WEN Ya, GAO Zhiyuan, WANG Jifu, CAI Yuxuan, GAO Shengzhen, LI Ruigai

(College of Information and Computer Engineering, Northeast Forestry University, Harbin 150040, China)

[Abstract] With the rapid development of society, aircraft has become an essential way of travel in people's life, meanwhile has changed people's life and facilitated their travel. However, nowadays people pays more attention to high efficiency, time and safety become important factors for people to measure the traffic. So plane delay is a prominent phenomenon that affects everyone's travel and restricts the development of the industry. The idea of this flight data visualization system comes from the 7th China Software Cup, the software design competition for college students. At the same time, as the students' innovative entrepreneurial training program of the Northeast Forestry University, the system is able to query in different parts of the flight routes, as well as the delay rate of the corresponding flights. The development of this system has a more accurate and clear goal for the civil aviation industry, and can better target some flights with the higher late rate and take corresponding actions to control them. At the same time, the research can provide the reference for traveling masses.

[Key words] route; delay rate; flight data visualization system

0 引言

随着中国经济的快速发展,越来越多的人选择乘坐飞机出行,航班延误问题是目前制约国内民航服务质量的焦点问题,航班正点率已成为民航业内的重中之重,同时也关系着民航业的整体运行品质^[1]。为了促进民航业的高效良性发展,进而提升用户出行体验,调研后可知将亟需对航班进行合理监控,以及开展数据对比,尽可能地降低航班取消、延误、备降等方面的概率。因此,航班数据可视化系统的研发就是至关重要的解决方式。该系统开发基本满足了对于国内航班的监控需求,尤其利于航空公司对航班进行管理控制,而对于一些晚点概率较高的航班将予以着重考察,并采取合适的整改措施,如此一来,不仅提升航空公司的运行效率,而且也更进一步地改善和优化了用户体验。

1 系统开发平台简介

PyCharm 是一种 Python IDE,带有一整套可以帮助用户在使用 Python 语言开发时提高其效率的工具,比如调试、语法高亮、Project 管理、代码跳转、智能提示、自动完成、单元测试、版本控制等。此外,该 IDE 提供了一些高级功能,可用于支持 Django 框架下的专业 Web 开发。PyCharm 是由 JetBrains 打造的一款 Python IDE,VS2010 的重构插件 Resharper 就是出自 JetBrains 的手笔。同时支持 Google App Engine,PyCharm 支持 IronPython。迄至目前,这些功能已经内嵌于先进代码分析程序中,就使得 PyCharm 成为 Python 专业开发人员和初级入门人员使用的有力工具。更重要的是,Pycharm 还配备了一个带编码补全,代码片段,支持代码折叠和分割窗口的智能、可配置的编辑器,从而可以更快、更轻

基金项目: 2018 年度东北林业大学校级大学生创新项目(201810225538)。

作者简介: 闻雅(1997-),女,本科生,主要研究方向:机器学习。

通讯作者: 闻雅 Email: 1759768433@qq.com

收稿日期: 2019-03-08

松地完成编码工作。在 PyCharm 的图形处理中, 利用 pyecharts 显示地图图形, 为了维持项目的轻量化运行, PyCharm 不再自带 pyecharts。如果需要用到世界地图, 则直接安装 echarts-countries-pypkg 包, 用到省级地图则直接安装 echarts-china-provinces-pypkg 包。有了这 2 个包, 就可获得完全显示地图效果。

2 系统算法设计

2.1 基本原理

本系统采用了基于朴素贝叶斯分类器对航班晚点进行预测, 其理论基础就是贝叶斯定理。在概率论与统计学中, 贝叶斯定理, 或称贝叶斯法则、贝叶斯规则, 描述了一个事件的可能性, 这个可能性是基于一些与此事件相关的预备知识。目前, 朴素贝叶斯算法已被广泛应用于文本分类领域。该算法基于条件独立性假设, 研究时假定给定目标对象的各个属性之间相互独立, 互不影响, 通过计算目标对象的先验概率, 利用贝叶斯定理计算出其后验概率、即该对象属于某类的概率, 然后比较后验概率的大小进行决策分类^[2]。这个假设的数学公式可表达为:

$$P(C | X) = \frac{P(C)P(X | C)}{P(X)} = \frac{P(C)}{P(X)} \prod_{i=1}^d P(X_i | C)$$

其中, $P(X)$ 表示 X 出现的概率, 属先验概率; $P(C)$ 表示不考虑相关因素时, 随机事件出现 C 情况的概率, 属先验概率; $P(C | X)$ 表示在 X 的条件下, 随机事件出现 C 情况的概率, 属后验概率; $P(X | C)$ 表示在已知事件出现 C 情况的条件下, 条件 X 出现的概率, 属后验概率。

因此, 属性条件独立性假设实际上是忽略掉了某些属性之间可能存在的关联, 假设属性的取值可能性都是独立的, 这就将 $P(x_1, x_2, x_3, \dots, x_n | c)$ 的计算简化为 $\prod_{i=1}^d P(x_i | c) = P(x_1 | c) * P(x_2 | c) * P(x_3 | c) * \dots * P(x_n | c)$ 。那么按照这种方式计算, 数据集无法覆盖所有可能的属性组合的样本的问题也得以基本解决, 此模型下不需要大量样本, 只需要保证每个属性的各个取值都能取到样本即可。

2.2 数据来源

该系统以及模型所需数据来源于 <https://zh.flightaware.com> 网站以及国外网站 <https://rp5.ru/> 中国天气_, 通过查询航班基础信息, 包括日期、出发时间、出发地点、到达时间、到达地点、飞机型号、晚

点状况等, 以及查询各个城市机场飞机起飞时的天气状况, 包括气象站水平大气压、总云量、机场附近特殊天气、水平能见度等多种因素, 最终对数据进行筛选整合, 用建立的朴素贝叶斯模型预测各条航线的晚点率, 显示各类航班的基础信息以及航班起飞时的天气状况。航班延误与人为因素^[3], 可能与管理不当, 工作人员自身疏忽^[4] 都有一定关系, 由于数据搜集困难, 该模型不考虑此因素。

2.3 实验结果分析

该模型的实验环境为 Oracle Linux 7.4 以及 Python 3, 硬件环境是基于 x64 的处理器, 处理器为 Intel® Core™ i7-5500U CPU @ 2.40 GHz 2.39 GHz, 内存为 8.00 GB, 64 位操作系统的 Windows10 企业版。实验运行步骤可描述为:

新建 python 项目对搜集的数据进行划分, 对划分的训练集以及测试集建立朴素贝叶斯模型。再对测试集进行预测, 同时绘制 ROC 曲线, 计算概率阈值。使用阈值重新预测飞机晚点概率。在这个模型中, 考虑了诸多因素, 例如出发到达时间、机型、航线、压强以及飞机起飞时相近时间点的天气因素等。由此得到的该航线晚点的混淆矩阵见表 1, 预测晚点概率阈值如图 1 所示。

表 1 混淆矩阵结果

Tab. 1 Confusion matrix results

		预测值	
		晚点(延误、备降、取消)	正常
真实值	晚点(延误、备降、取消)	TPR = 0.657 9	FPR = 0.338 2
	正常	FNR = 0.342 1	TNR = 0.661 8

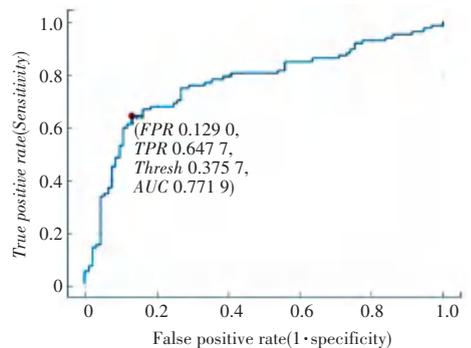


图 1 模型阈值结果预测

Fig. 1 Model threshold result prediction

3 系统功能设计

3.1 系统开发目的

本系统的设计是旨在开发一款航班信息可视化软件, 实现航班数据的显示与集成查询功能, 并以地

图的形式展示大量的航班信息,可以点击查看航班航线信息,而且可以融合一些机场地理位置信息做出具体数据分析。该系统主要根据航班的实时信息,提供航班航线可视化,使乘客更直观地了解到航班的基本信息,采集连续3个月的航班数据,结合机场的地理信息,方便航空公司监控以及查看航班的延误、取消,再据此分析航线繁忙程度、航班准点率等。故而,可供乘客选择航班的方式更加直观,航空公司也方便对那些航班经常延误以及取消的航线实施有针对性管理。

3.2 功能模块设计

该系统共有4个模块,分别是:航班信息可视化显示模块、航班集成查询模块、航班准点率预测模块和航班推荐模块。每个模块对应不同的功能。功能模块设计框架如图2所示。这里,针对各模块的设计阐释可分述如下。

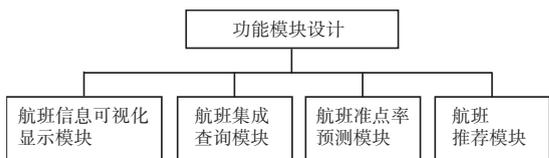


图2 功能模块设计图

Fig. 2 Function module design drawing

3.2.1 航班信息可视化显示模块

该系统使用python语言进行开发。Python具有高级的数据结构,比如列表、字典等都内置于语言本身,这样极大地提升了开发效率。现代的许多编程语言都支持面向对象,在实际的业务面前,面向对象的思维方式可以把业务抽象成对象来编程实现,好的面向对象设计可以把整个项目分成不同的模块,降低耦合,方便程序的后期调试^[5]。该系统界面友好,使用方便,易于用户操作。该系统运用airport_code文件中的2个字典,用来存储城市名称-机场代码键值对、机场代码-城市名称键值对,为了显示所有航线,这2个字典中存储了所有的城市名称-机场代码键值对、机场代码-城市名称键值对。显示结果如图3所示。

3.2.2 航班集成查询模块

该系统调用xlrd模块处理Excel表格,同时写了searchdata()函数,用来搜索特定列的特定元素,一行一行查找,查找一行输出一行。按照不同的查找方式,修改特定列的参数,实现转化查找功能。系统的查询模块共有5种方式,如图4所示。对此内容可详述如下。



图3 航线结果示意图

Fig. 3 Schematic diagram of route results

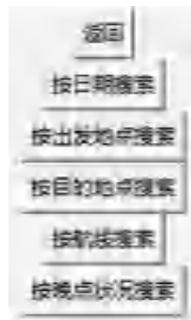


图4 查询方式

Fig. 4 Query mode

- (1)按照日期查询航班信息,输入日期,显示对应的航班数据。
- (2)按照出发地点查询航班信息,输入出发地点,显示对应的航班数据。
- (3)按照目的地点查询航班信息,输入目的地点,显示对应的航班数据。
- (4)按照航班航线查询航班信息,输入航线,显示对应的航班数据。
- (5)按照晚点情况查询航班信息输入是否晚点,显示对应的航班数据。查询样式多元化,给用户提供更多选择,研究得到了查询结果的设计显示见图5。



图5 查询结果

Fig. 5 Query results

3.2.3 航班准点率预测模块

系统使用 python 语言的 Tkinter 库并引用工程文件夹下的 flight_path 文件,完成了界面功能的显示。为了调用数据,该系统选用了 xlrd 库的 open_workbook() 函数来打开存储航班数据的 Excel 文件,并按照列循环搜索数据文件中出发地和目的地的两列。数据中出发地和目的地的数据格式为[汉语拼音书写的机场名称-机场三字代码-机场四字代码]。search() 函数按照字符位数,搜索机场的三字代码,当用户输入出发地和目的地后,接收 airport_code() 函数转化的机场三字代码作为参数。研究中,先是搜索出发地中是否存在用户意欲寻求的出发地,若存在,继续在目的地中搜索有无用户希望的目的地,若也存在,则返回一个字典 $d[T, rate]$ 。其中, T 表示存在航线, $rate$ 表示预测延误概率,可由 predicted() 函数计算出并存入数据文件,如图 6 所示,若任意一步不存在则退出。为了将用户输入和城市名称转化成一致,系统调用 flight_path 文件中的 path() 函数的 try except 结构,在 try 语句内首先调用 airport_code() 函数将界面传入的用户输入转化为机场三字代码,再以处理过的 2 个机场代码为参数调用 search() 函数,判断返回值,若 $d[0]$ 不等于 T ,则抛出错误,若等于 T 则继续。调用 search() 并得到正确的返回值后,调用 airport_ch() 函数,将用户输入转化为城市名称,使用 data[] 字典存储出发地和目的地名称。该系统航线示意图如图 7 所示。

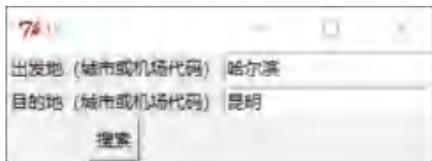


图 6 搜索框

Fig. 6 Search box

3.2.4 航班推荐模块

此模块是给乘客推荐选择符合乘客需求的最佳航班。如果乘客追求价格便宜,则推荐价格相对便

宜的航班;如果乘客关注是否准时,则推荐晚点率最低的航班;如果乘客在意航班是否是直航航线,则筛选去中途停站的航班。此外,当乘客需要服务质量相对较好的航班时,也可以依次列出各类航班服务细致程度,给用户更为上乘的体验感受。综上所述可知,此模块可根据用户的需求,灵活筛选出合适的航班。



图 7 航线预测结果显示图

Fig. 7 Route prediction result display figure

4 结束语

本论文主要设计并开发了一款航班可视化系统,具有重要的现实意义,一方面有利于民航行业对航班加强监控管理,以提高民航行业的效率;另一方面,也可以给航班用户搜罗更多的信息,对于其出行提供了更多的参考,而且可以提升乘客的用户体验。还需指出,该系统的某些功能也仍有不足之处,有待后续的研发完善,使其在未来的投入使用中能够获得更好的表现。

参考文献

- [1] 冯思轶. 大数据在机场航班延误管理中的应用[J]. 电子技术与软件工程, 2018(4): 185-187.
- [2] 赵文涛, 孟令军, 赵好好, 等. 朴素贝叶斯算法的改进与应用[J]. 测控技术, 2016, 35(2): 143-147.
- [3] 张春雷. 民航旅客运输延误问题探究[J]. 中国高新区, 2018(12): 269.
- [4] 杜洋, 孙伟社, 李国正. 我国民航航班延误的原因、损失及对策分析[J]. 民航管理, 2017(3): 9-13.
- [5] 王赟斌. 航空维修信息系统的设计与实现[D]. 郑州: 郑州大学, 2015.

(上接第 227 页)

参考文献

- [1] 李静宇, 周明, 李庚银. 基于交错 Boost 变流器和 LLC 谐振变流器的高效率车载充电机[J]. 现代电子技术, 2019, 42(1): 119-123.
- [2] 雷新颖, 王成, 苏力. 基于 STM32 的能量反馈型电子负载设计[J]. 电子设计工程, 2017, 25(21): 119-123.

- [3] 秦浩钧. 直流稳压电源的设计与测试[J]. 甘肃科技, 2018, 34(21): 15-18.
- [4] 毕磊, 张彦. 载波在 SPWM 逆变电路中的应用分析及仿真[J]. 舰船电子对抗, 2018, 41(5): 41-45.
- [5] 赵巧妮, 夏益辉. 单相 PWM 整流器比例谐振与阻尼控制研究[J]. 电气传动, 2018, 48(11): 30-33.
- [6] 余坚铨, 张超杰, 吴杰长. 模拟电路动态电源电流信号测量方法综述[J]. 船电技术, 2018, 38(11): 59-65.