



## (12)发明专利申请

(10)申请公布号 CN 105844346 A

(43)申请公布日 2016.08.10

(21)申请号 201610152208.4

(22)申请日 2016.03.17

(71)申请人 福州大学

地址 350108 福建省福州市闽侯县上街镇  
大学城学园路2号福州大学新区

(72)发明人 郑相涵 叶慧娟 郭文忠

(74)专利代理机构 福州元创专利商标代理有限公司 35100

代理人 蔡学俊

(51)Int.Cl.

G06Q 10/04(2012.01)

G06Q 50/30(2012.01)

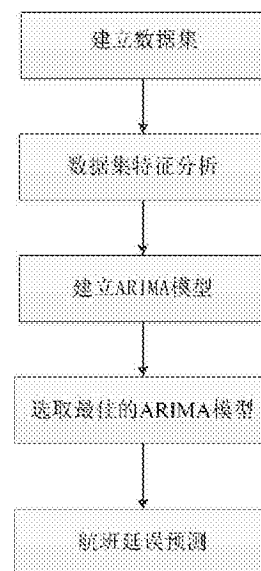
权利要求书2页 说明书4页 附图1页

### (54)发明名称

基于ARIMA模型的航班延误预测方法

### (57)摘要

本发明涉及一种基于ARIMA模型的航班延误预测方法,具体包括以下步骤:步骤S1:建立数据集:采集航班数据,所述航班数包括航班信息、时间信息、机场信息以及延误时间信息;步骤S2:数据集特征分析:分析各个机场与航空公司间的差异以及天气对航班的影响;步骤S3:建立ARIMA模型:ARIMA模型的建立包括移动平均过程、自回归过程、自回归移动平均过程以及ARIMA过程;步骤S4:选取最佳的ARIMA模型:对建立的ARIMA模型进行验证并选取最佳参数;步骤S5:航班延误预测:获取任意一航班数据后,选取最佳的ARIMA函数作为模型延误函数,选取多元线性回归函数作为天气延误函数,相加后得到最后的预测结果。本发明能实现更高的预测精度,缩短使用时间,有效预测航班延误。



1. 一种基于ARIMA模型的航班延误预测方法,其特征在于:具体包括以下步骤:

步骤S1:建立数据集:采集航班数据,所述航班数包括航班信息、时间信息、机场信息以及延误时间信息;

步骤S2:数据集特征分析:分析各个机场与航空公司间的差异以及天气对航班的影响;

步骤S3:建立ARIMA模型:ARIMA模型的建立包括移动平均过程MA、自回归过程AR、自回归移动平均过程ARMA以及ARIMA过程;

步骤S4:选取最佳的ARIMA模型:对建立的ARIMA模型进行验证并选取最佳参数;

步骤S5:航班延误预测:获取任意一航班数据后,选取最佳的ARIMA函数作为模型延误函数,选取多元线性回归函数作为天气延误函数,并将所述模型延误函数与天气延误函数相加,得到最后的预测结果。

2. 根据权利要求1所述的一种基于ARIMA模型的航班延误预测方法,其特征在于:所述步骤S1中,所述航班信息包括航空公司、航班号以及飞机尾号;所述时间信息包括日期、预计出发时间、预计到达时间、预计飞行时间、实际飞行时间、延误出发时间以及延误到达时间;所述机场信息包括出发机场、到达机场、税率期限、起飞时间以及降落时间;延误时间信息包括极端天气情况、航空公司的飞机控制情况、国家航空系统的飞机控制情况以及安全问题。

3. 根据权利要求2所述的一种基于ARIMA模型的航班延误预测方法,其特征在于:所述极端天气情况包括龙卷风与暴风雪;所述航空公司的飞机控制情况包括飞机维修或清洁与行李装载;所述国家航空系统的飞机控制情况包括机场运营与交通堵塞下取消航线;所述安全问题包括飞机的安全漏洞。

4. 根据权利要求1所述的一种基于ARIMA模型的航班延误预测方法,其特征在于:所述步骤S3中,所述ARIMA模型根据原序列是否平稳以及回归中所含部分的不同,包括移动平均过程MA、自回归过程AR、自回归移动平均过程ARMA以及ARIMA过程;将非周期性的ARIMA模型被划分为ARIMA(p,d,q)模型,其中p为自回归项,q为移动平均项数,d为时间序列成为平稳时所做的差分次数,则所述ARIMA模型如下:

$$\begin{cases} \Phi(L)\Delta^d y_t = c + \Theta(L)\varepsilon_t \\ E(\varepsilon_t) = 0, \text{Var}(\varepsilon_t) = \sigma^2, E(\varepsilon_t, \varepsilon_s) = 0, s \neq t(1) \\ E y_s \varepsilon_t = 0, \forall s < t \end{cases}$$

其中  $\Delta^d = (1-L)^d$ ;  $\Phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p$  为自回归(AR)系数多项式;  $\Theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q$  为移动平均(MA)系数多项式;  $\{\varepsilon_t\}$  为零均值白噪声序列。

5. 根据权利要求1所述的一种基于ARIMA模型的航班延误预测方法,其特征在于:所述步骤S4具体包括以下步骤:

步骤S41:进行数据平稳性检验:绘制时间序列并对图形进行评价或运用ADF单位根进行检验;

步骤S42:拟合平稳时间序列ARMA:使用 $\{y\}$ 表示进行差分之后的平稳时间序列,并计算

自相关函数ACF和偏自相关函数PACF,从ACF和PACF测试模式中选择合适的p和q的值;

步骤S43:选择参数:测试p和q的不同组合,应用AIC和SC准则选择最佳的模型参数;

步骤S44:模型验证:通过检查残差是否是一个白噪声序列来验证该模型是否准确;

步骤S45:模型预测:根据被选定的模型,用于对航班模型延误的未来值作出预测。

6.根据权利要求1所述的一种基于ARIMA模型的航班延误预测方法,其特征在于:所述步骤S5具体包括以下步骤:

步骤S51:选取模型延误函数:选择ARIMA函数作为模型延误函数;

步骤S52:选取天气延误:通过主成分分析以及因素分析方法获取影响航班延误的三个主要因素,所述三个主要因素包括风、可见度和空气条件,运用多元线性回归函数作为天气延误函数:

$$y_i = \beta_1 w_i + \beta_2 v_i + \beta_3 s_i + \varepsilon_i \quad i = 1, \dots, n$$

其中  $w_i$  指的是风速,  $v_i$  指的是可见度,  $s_i$  指的是空气条件和  $i$  代表航班号;

步骤S53:应用所有历史天气延误数据集线性回归和最小化可能值,获得合理的,其次建立适用于天气延误的回归方程,再分析函数的优势,检验回归函数的正确性、天气结果的随机性以及特殊的回归协同因素的正确性,最后对航班天气延误的未来值作出预测。

## 基于ARIMA模型的航班延误预测方法

### 技术领域

[0001] 本发明涉及航班信息分析技术领域,具体涉及一种基于ARIMA模型的航班延误预测方法。

### 背景技术

[0002] 有效的数据挖掘算法和分析策略,可以通过获取先验知识为公司或个人提取有价值的信息,并帮助他们作出进一步的决定。在那些涉及大数据的领域里,航班延误预测已引起了广泛的关注。近年来,航班延误的风险管理的重要性显而易见。比如,骇人听闻的MH370飞行事故的发生,以及国内航班延误导致的巨大经济损失给航空公司和乘客带来的巨大不满,将飞行风险管理推到了一个非常紧急的位置。所以无论从安全系数还是经济方面来说,更有效的航班延误预测模型是十分必要的。

[0003] 然而,鉴于航班数据的特征(如数据量大、多样性等),高精度地预测航班延误,同时保证计算复杂性和延迟在可接受的范围内是很困难的。此外,影响航班延误的特征中,天气等因素可能会动态改变。因此,提出基于ARIMA模型的航班延误预测方法,它能够有效地预测航班延误。

### 发明内容

[0004] 有鉴于此,本发明的目的是提供一种基于ARIMA模型的航班延误预测方法,以实现更高的预测精度,缩短使用时间,有效预测航班延误。

[0005] 本发明采用以下方案实现:一种基于ARIMA模型的航班延误预测方法,具体包括以下步骤:

步骤S1:建立数据集:采集航班数据,所述航班数包括航班信息、时间信息、机场信息以及延误时间信息;

步骤S2:数据集特征分析:分析各个机场与航空公司间的差异以及天气对航班的影响;

步骤S3:建立ARIMA模型:ARIMA模型的建立包括移动平均过程MA、自回归过程AR、自回归移动平均过程ARMA以及ARIMA过程;

步骤S4:选取最佳的ARIMA模型:对建立的ARIMA模型进行验证并选取最佳参数;

步骤S5:航班延误预测:获取任意一航班数据后,选取最佳的ARIMA函数作为模型延误函数,选取多元线性回归函数作为天气延误函数,并将所述模型延误函数与天气延误函数相加,得到最后的预测结果。

[0006] 进一步地,所述步骤S1中,所述航班信息包括航空公司、航班号以及飞机尾号;所述时间信息包括日期、预计出发时间、预计到达时间、预计飞行时间、实际飞行时间、延误出发时间以及延误到达时间;所述机场信息包括出发机场、到达机场、税率期限、起飞时间以及降落时间;延误时间信息包括极端天气情况、航空公司的飞机控制情况、国家航空系统的飞机控制情况以及安全问题。

[0007] 进一步地,所述极端天气情况包括龙卷风与暴风雪;所述航空公司的飞机控制情

况包括飞机维修或清洁与行李装载;所述国家航空系统的飞机控制情况包括机场运营与交通堵塞下取消航线;所述安全问题包括飞机的安全漏洞。

[0008] 进一步地,所述步骤S3中,所述ARIMA模型根据原序列是否平稳以及回归中所含部分的不同,包括移动平均过程MA、自回归过程AR、自回归移动平均过程ARMA以及ARIMA过程;将非周期性的ARIMA模型被划分为ARIMA(p,d,q)模型,其中p为自回归项,q为移动平均项数,d为时间序列成为平稳时所做的差分次数,则所述ARIMA模型如下:

$$\begin{cases} \Phi(L)\Delta^d y_t = c + \Theta(L)\varepsilon_t \\ E(\varepsilon_t) = 0, \text{Var}(\varepsilon_t) = \sigma^2, E(\varepsilon_t, \varepsilon_s) = 0, s \neq t(1) \\ Ey_s \varepsilon_t = 0, \forall s < t \end{cases}$$

其中  $\Delta^d = (1-L)^d$ ;  $\Phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p$  为自回归(AR)系数多项式;  $\Theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q$  为移动平均(MA)系数多项式;  $\{\varepsilon_t\}$  为零均值白噪声序列。

[0009] 进一步地,所述步骤S4具体包括以下步骤:

步骤S41:进行数据平稳性检验:绘制时间序列并对图形进行评价或运用ADF单位根进行检验;

步骤S42:拟合平稳时间序列ARMA:使用{y}表示进行差分之后的平稳时间序列,并计算自相关函数ACF和偏自相关函数PACF,从ACF和PACF测试模式中选择合适的p和q的值;

步骤S43:选择参数:测试p和q的不同组合,应用AIC和SC准则选择最佳的模型参数;

步骤S44:模型验证:通过检查残差是否是一个白噪声序列来验证该模型是否准确;

步骤S45:模型预测:根据被选定的模型,用于对航班模型延误的未来值作出预测。

[0010] 进一步地,所述步骤S5具体包括以下步骤:

步骤S51:选取模型延误函数:选择ARIMA函数作为模型延误函数;

步骤S52:选取天气延误:通过主成分分析以及因素分析方法获取影响航班延误的三个主要因素,所述三个主要因素包括风、可见度和空气条件,运用多元线性回归函数作为天气延误函数:

$$y_i = \beta_1 w_i + \beta_2 v_i + \beta_3 s_i + \varepsilon_i \quad i = 1, \dots, n,$$

其中  $w_i$  指的是风速,  $v_i$  指的是可见度,  $s_i$  指的是空气条件和  $i$  代表航班号。

[0011] 步骤S53:应用所有历史天气延误数据集线性回归和最小化可能值,获得合理的,其次建立适用于天气延误的回归方程,再分析函数的优势,检验回归函数的正确性、天气结果的随机性以及特殊的回归协同因素的正确性,最后对航班天气延误的未来值作出预测。

[0012] 与现有技术相比,本发明的基于ARIMA(差分自回归移动平均)模型的航班延误预测方法,实现更高的预测精度,缩短使用时间,有效预测航班延误。

## 附图说明

[0013] 图1 为本发明的流程图示意图。

## 具体实施方式

[0014] 下面结合附图及实施例对本发明做进一步说明。

[0015] 本实施提供一种基于ARIMA模型的航班延误预测方法,如图1所示,具体包括以下步骤:

步骤S1:建立数据集:采集航班数据,所述航班数包括航班信息、时间信息、机场信息以及延误时间信息;

步骤S2:数据集特征分析:分析各个机场与航空公司间的差异以及天气对航班的影响;

步骤S3:建立ARIMA模型:ARIMA模型的建立包括移动平均过程MA、自回归过程AR、自回归移动平均过程ARMA以及ARIMA过程;

步骤S4:选取最佳的ARIMA模型:对建立的ARIMA模型进行验证并选取最佳参数;

步骤S5:航班延误预测:获取任意一航班数据后,选取最佳的ARIMA函数作为模型延误函数,选取多元线性回归函数作为天气延误函数,并将所述模型延误函数与天气延误函数相加,得到最后的预测结果。

[0016] 在本实施例中,所述步骤S1中,所述航班信息包括航空公司、航班号以及飞机尾号;所述时间信息包括日期、预计出发时间、预计到达时间、预计飞行时间、实际飞行时间、延误出发时间以及延误到达时间;所述机场信息包括出发机场、到达机场、税率期限、起飞时间以及降落时间;延误时间信息包括极端天气情况、航空公司的飞机控制情况、国家航空系统的飞机控制情况以及安全问题。

[0017] 在本实施例中,所述极端天气情况包括龙卷风与暴风雪;所述航空公司的飞机控制情况包括飞机维修或清洁与行李装载;所述国家航空系统的飞机控制情况包括机场运营与交通堵塞下取消航线;所述安全问题包括飞机的安全漏洞。

[0018] 在本实施例中,所述步骤S2通过对一些机场与航空公司的分析,可得出如下结论:1、机场模式:各机场的航班数量、平均延误时间、航空公司数量和取消率存在差异。2、航空模式:美国的航空公司会比西南航空公司有一个更高的出发/到达延误率和更长的延误时间。3、天气影响:据统计,百分之四的航班延误是因为极端天气的影响,而非极端天气也会导致飞机晚点,天气为影响航班延误的一个重要因素。

[0019] 在本实施例中,ARMA模型包括自回归(AR)和移动平均(MA)。ARMA与ARIMA的区别为:ARMA模型用于预测平稳时间序列,而ARIMA用于预测非平稳时间序列。ARIMA模型是差分与ARMA的组合,称为差分自回归移动平均模型,是指将非平稳时间序列转化为平稳时间序列,然后将因变量仅对它的滞后值以及随机误差项的现值和滞后值进行回归所建立的模型。

[0020] 则所述步骤S3中,所述ARIMA模型根据原序列是否平稳以及回归中所含部分的不同,包括移动平均过程MA、自回归过程AR、自回归移动平均过程ARMA以及ARIMA过程;将非周期性的ARIMA模型被划分为ARIMA(p,d,q)模型

,其中p为自回归项,q为移动平均项数,d为时间序列成为平稳时所做的差分次数,则所述ARIMA模型如下:

$$\begin{cases} \Phi(L)\Delta^d y_t = c + \Theta(L)\varepsilon_t \\ E(\varepsilon_t) = 0, \text{Var}(\varepsilon_t) = \sigma^2, E(\varepsilon_t, \varepsilon_s) = 0, s \neq t(1) \\ E y_s \varepsilon_t = 0, \forall s < t \end{cases}$$

其中  $\Delta^d = (1-L)^d$ ;  $\Phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p$  为自回归(AR)系数多项式;  $\Theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q$  为移动平均(MA)系数多项式;  $\{\varepsilon_t\}$  为零均值白噪声序列。

[0021] 在本实施例中,所述步骤S4具体包括以下步骤:

步骤S41:进行数据平稳性检验:一个简单的方法就是绘制时间序列,并对图形进行评价,更为精确的方法就是运用ADF单位根进行检验;

步骤S42:拟合平稳时间序列ARMA:使用 $\{y\}$ 表示进行差分之后的平稳时间序列,并计算自相关函数ACF和偏自相关函数PACF,从ACF和PACF测试模式中选择合适的p和q的值;

步骤S43:选择参数:测试p和q的不同组合,应用AIC和SC准则选择最佳的模型参数;

步骤S44:模型验证:通过检查残差是否是一个白噪声序列来验证该模型是否准确;

步骤S45:模型预测:根据被选定的模型,用于对航班模型延误的未来值作出预测。

[0022] 在本实施例中,所述步骤S5中,获取航班数据后,提取两个主要导致航班延误的特征向量,进行预测,具体包括以下步骤:

步骤S51:选取模型延误函数:选择ARIMA函数作为模型延误函数;其中所述ARIMA函数的选取即根据上述方法获得:首先绘制航班模型延误的数据分布图,进行数据平稳性检验,判断是否为平稳时间序列;如若不是,则进行时间序列差分,直到得到一个平稳的时间序列;其次选择合适的ARIMA模型,使用R中的“acf”和“pacf”函数来分别(自)相关图和偏相关图,并在“acf”和“pacf”设定“plot=FALSE”来得到自相关和偏相关的真实值。再则使用R中的“arima()”函数来估计ARIMA(p,d,q)模型中的参数,选择最佳模型参数,使用残差验证模型正确性,最后使用它们做出预测模型,对航班模型延误的未来值作出预测;

步骤S52:选取天气延误:通过主成分分析以及因素分析方法获取影响航班延误的三个主要因素,所述三个主要因素包括风、可见度和空气条件,运用多元线性回归函数作为天气延误函数:

$$y_i = \beta_1 w_i + \beta_2 v_i + \beta_3 s_i + \varepsilon_i \quad i = 1, \dots, n,$$

其中  $w_i$  指的是风速,  $v_i$  指的是可见度,  $s_i$  指的是空气条件和  $i$  代表航班号。

[0023] 步骤S53:应用所有历史天气延误数据集线性回归和最小化可能值,获得合理的,其次建立适用于天气延误的回归方程,再分析函数的优势,检验回归函数的正确性、天气结果的随机性以及特殊的回归协同因素的正确性,最后对航班天气延误的未来值作出预测。

[0024] 以上所述仅为本发明的较佳实施例,凡依本发明申请专利范围所做的均等变化与修饰,皆应属本发明的涵盖范围。

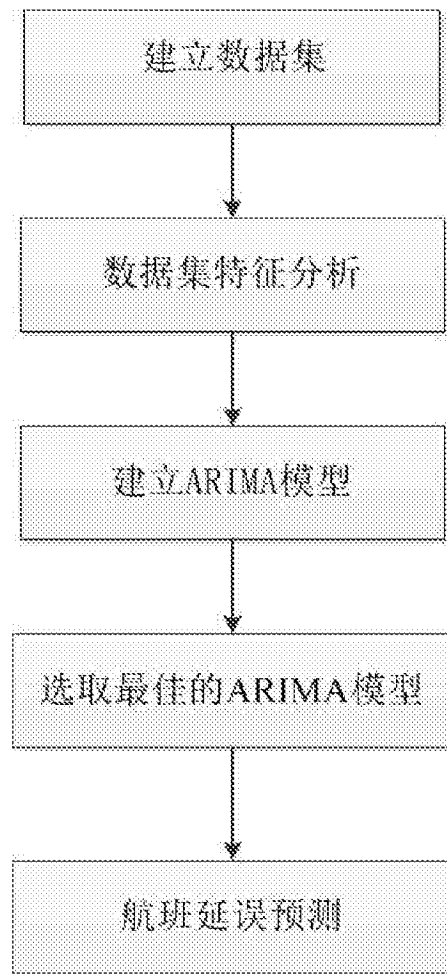


图1