数据驱动的异常检测与预警

摘要

。

**针对全图障碍信息已知的情况**，。

**针对全图障碍信息未知的情况**，。

。

**关键词： 算法 蚁群算法 动态局部最优 模拟人类视觉扫描法**

一、问题重述

**1.1 问题背景**

近年来，随着社会经济水平的提升，生产企业也在迅速地发展。但随着生产企业数量增多，其安全质量呈现了下滑趋势，生产事故频频爆发。究其原因，不仅受到了生产企业硬性的限制，如生产人员密集、生产规模及能力较弱、生产中的薄弱环节较多等原因，还受到了生产企业管理层和一线生产员工安全意识的影响。许多管理人员和一线职工防范意识较弱，导致原本安全问题就频发的生产环节产生安全事故的概率大大加强。因此保证安全、防范风险在高质量推动生产企业发展中就显得尤为重要。

目前，解决生产企业高质量发展仍然是中国制造业面临的主要问题。安全的生产环境才能保证产品的质量和生产人员的人身安全。建立风险预测和评定机制也就成为了当务之急，只有制定标准化的管理、评定机制才可以使生产企业的安全得到保障。因此，我们考虑从生产中常见的非风险性异常和对生产安全产生威胁的风险性异常入手，建立预测和风险性异常程度评价模型，用适当的数学模型对生产系统的安全性进行敏感性分析和评价。

**1.2 问题提出**

**问题一**：根据附件1给出的数据特征，具有规律性、特征性、偶发性等特点的波动属于正常波动，不会对生产安全产生影响，属于非风险性异常；具有持续性、联动性特点的波动属于异常性波动，会对生产安全造成威胁，属于风险性异常。要求我们建立合适的数学模型，给出评判这两类异常数据的方案。

**问题二**：在问题一结果的基础上，建立风险性异常数据的量化评价模型，对每个时刻的数据进行量化评价，通过建立的数学模型和附件数据信息，寻找异常分值最高的5个时刻和这5个时刻所对应的异常传感器编号，并通过模型对最终的结果进行合理的分析和评价。

**问题三：**建立风险性异常预警模型来争对在未来生产过程中可能存在的安全风险，并对当日23:00:00-23:59:59时间段进行风险性异常预测，结合问题二的数据量化评价模型，对以15分钟为划分的四个预测时间段进行量化评价，给出每个时间段内的最高异常分和对应的传感器编号。

**问题四：**结合问题二和问题三的结果，建立生产系统安全行评估模型，以30分钟为时间间隔单位对当日00:00:00-23:59:59的时间段内的整个生产安全系统进行量化评价，量化结果的最终分值在0-100区间范围内，并对量化评价后的结果进行敏感性分析和评价。

二、问题分析

**2.1 问题一的分析**

对于问题一而言，我们根据题目的条件，主要考虑通过建立筛选的模型来区分风险性异常数据与非风险性异常数据。我们的模型采用数据预处理的方式，利用离群点检测和等级性数据划分的方式，对所有的传感器的数据属性进行初步的划分和筛选；再利用白噪声检测的方法，找到只存在白噪声的时序数据，将其归纳为非风险性异常数据。我们采用夏皮洛-威尔逊检验和KS检验，通过白噪声时序数据满足正态分布或均匀分布的特点进一步筛选数据属性种类；最后我们利用皮尔逊相关系数法，对剩余未能被筛选出的传感器进行相关性矩阵计算，参考文献资料后，以0.65为划分依据，当传感器时序数据大于0.65，则认为该传感器间相关性显著，属于风险性异常数据，反之，则属于非风险性异常数据。最终可以得到各个传感器的数据的属性值。

**2.2 问题二的分析**

问题二要求我们对风险性异常数据建立一个衡量异常程度的量化指标模型。我们主要考虑采用密度聚类的方法来解决指标权重和传感器的具体数值，以此计算风险性异常数据异常程度的量化值。以皮尔逊相关系数矩阵为基础，将矩阵中数据作为相对距离，寻找一个既满足于传感器数据相关性，又满足簇基本要求的簇的分类方式。之后求解出各个簇的核心对象和最小半径，簇中的每个风险性异常数据的权重即为簇的最小半径，非风险性异常数据权重为0。每个传感器任意时刻的数值通过计算其当前时刻的自相关系数得到。最终通过归一化处理，可以得到生产系统任意时刻的总异常程度数值以及该时刻下各传感器的异常程度数值。

**2.3 问题三的分析**

问题三题目要求我们建立风险性异常预警模型，以提前发现未来生产过程中可能发生的隐患。我们考虑的是通过ARIMA(自回归综合滑动平均模型)。先利用单次或多次的差分运算，使时间序列数据能够满足ARMA模型的平稳性要求；再以AIC准则函数的解最小化为目标函数，在有限区域内枚举ARMA模型中的两个参数值和，以此改变ARMA模型中的参数向量和残差，最终改变AIC准则函数的解。枚举有限区域中的所有参数值和，进行循环运算，求出所有情况的函数最终解。最终选取AIC准则函数最小值情况下的参数向量，对传感器的未来数据进行风险性预测，给出异常程度最高的5个传感器编号。

**2.4 问题四的分析**

三、模型假设

1. 。
2. 。
3. 。

四、符号说明

|  |  |
| --- | --- |
| 符号 | 符号意义 |
| *T* | 表示移动机器人沿任意路径从起点到终点所需要的时间 |
|  | 表示坐标为的栅格的成本 |
|  | 表示栅格所接受到的所有成本的最小值 |
| *S* | 表示此栅格的总像素点个数 |
|  | 表示某个栅格内障碍物所包含的像素点个数 |
|  | 表示每个栅格的状态 |
|  | 表示从起点（S）到终点（D）的估计代价 |
|  | 表示从起点到状态n的成本 |
|  | 表示从状态n到终点（D）的估计代价 |
| *k* | 表示蚁群算法中决定信息素含量的因子 |
|  | 表示障碍物栅格的默认通过率 |
|  | 表示栅格的通过率 |
|  | 表示栅格中非障碍物所占面积 |
|  | 表示栅格的总面积 |
|  | 表示从移动机器人当前所处的栅格到终点的任意路径的成本 |
|  | 表示任意两条射线之间的夹角 |
|  | 表示整个地图的长与宽 |

注：未列出符号及重复的符号以出现处为准

五、模型的建立与求解

**5.1问题一的模型建立与求解**

**5.1.1 模型的工作原理**

根据题目的要求，我们建立的问题一模型将主要通过夏皮洛-威尔逊检验法、KS检验法和皮尔逊相关系数方法对附件1给出的波动性数据进行分析，判定其为非风险性异常数据或风险性异常数据。

1).夏皮洛-威尔逊检验

夏皮洛-威尔逊检验是一种检验数据正态性的方法，目的在于判断单一传感器是否属于正态分布。

其检验的零检验是来自一个正态分布母体的样本，统计量是：

其中表示的来自母体的样本，表示样本个数，是个样本的平均值。

常量符合以下条件：

a).是的最佳线性无偏估计

b).是样本来源的正态分布的标准差

可以得到的准确值

其中矩阵是一个协方差矩阵，属于个标准正态分布的随机变量的顺序统计量，是这些变量的期望组成的一个向量。

是通常使用的的一个无偏估计。若样本数据确实来自正态分布的母体样本，则最终的统计量的分母和分子均会趋向于的估计值；对于非正态分布的数据来说，分子和分母不会趋向于同一个常数。

我们把统计量看作是顺序排列样本值和系数之间相关系统的平方或线性回归的确定性系数，其值越高，越表示样本与正态分布相匹配。统计量的最大值是1，最小值是。

我们通过设定显著性水平(常见的是0.05)，之后获得它的临界值，如果，则拒绝原假设，反之，则接受原假设。

最终可以用夏皮洛-威尔逊检验的方法，筛选出不满足正态分布的时序数据的传感器编号。

2).皮尔逊相关系数

皮尔逊相关系数是衡量向量相似度的一种方法，它的输出范围区间在，其目的在于通过传感器的时序数据，确定两个传感器之间的相关性。

其具体的计算公式如下：

其中表示传感器1的时序数据，表示传感器1时序数据的平均值；表示传感器2的时序数据，表示传感器2时序数据的平均值。

在皮尔逊相关系数的取值范围区间内。存在以下含义：

a).相关系数为0时，两个向量不相关

b).相关系数为负数时，两个向量呈负相关关系

c).相关系数为正数时，两个向量呈正相关关系

最终我们可以通过定义的相关系数的值，来筛选出时序数据不满足相关关系情况下的传感器编号。

3).KS检验

KS检验方法是一种非参数统计检验方法，该方法常被用于单样本是否符合某个已知分布。我们采用KS检验方法的目的是判断传感器时序数据是否满足均匀分布。

定义原假设：总体的具有的分布，构造检验统计量：

其中为对于一个样本集的累计分布函数，为一个假设的理论分布。

其中累积分布函数有：

为指示函数，

若当原假设为真时，依分布收敛于假设的理论分布，即

当时，趋向于0，原假设成立，即样本集的累积分布函数满足均匀分布；反之累积分布函数不满足均匀分布。

**5.1.2 模型**

1)目标的确定

根据题目要求，我们建立合理的数学模型对存在波动的数据进行分析处理，定量化地区分具有规律性、独立性、偶发性特点的非风险性异常数据波动和具有持续性、联动性特点的风险性异常数据波动。最终给出非风险性异常数据和风险性异常数据的判定方案。

2)数据预处理

a).离群点检测

我们采用基于统计模型的离群点检测方法，来对所有传感器的时序数据进行初筛选，选出非风险性异常数据的传感器。

该检测方案采用不和谐检验识别离群点。不和谐检验需检查两个假设，即工作假设和备择假设。

工作假设指若样本点的某个统计量相对于数据分布其显著性概率充分小，认为该样本点不和谐，认为此点为离群点，工作假设拒绝，此时备择假设被采用，样本点来自另一个分布模型。若它符合备择假设，则可以认为它是符合某一备择假设分布的离群点。

例如盒图方法，我们以第一个四分位数、中位数、第三个四分位数为输入数据，计算最大和最小非离群点值，作为一维数据的离群区分点。

因为我们可以通过观察传感器时序数据的分布特征，从而了解时序数据的大致分布，并且该数据是针对单一属性的，且数据十分充分。故而选取这种建立在标准统计学原理基础上的方案能够使检验非常有效。

最终通过离群点检测筛选出来的离群点都是作为风险性异常数据的传感器，对没被该方案筛选出来的传感器继续进行筛选处理。

b).等级性数据划分

经过离群点检测筛选方案后，我们再通过等级性数据划分进行第二步数据预处理。在所有的100个传感器中，我们观察可以知道，传感器存在单值或多值，即某些传感器其时序数据只存在一个值或存在多个数据值。我们将传感器时序数据的值定义为该传感器的等级性，因此每个传感器都有与之对应的唯一等级性数据。

通过计算，我们得到了传感器等级性数据在区间内的，且其时序数据信息具有明显差异的数据，是属于风险性异常数据。因此我们可以通过等级性数据划分的方案，进一步筛选出风险性异常数据所对应的传感器，对为筛选出的传感器时序数据进行模型计算和分析。

3)方案设定

*Step1*.对100个传感器的时序数据判断，寻找只存在白噪声的时间序列的传感器编号，并将其归纳到非风险性异常传感器中。因此我们对所有传感器的时序数据单独进行夏皮洛-威尔逊检验方法和KS检验方法。若传感器的时间序列数据存在白噪声，则其应该满足正态分布或均匀分布，接受原假设。

只存在白噪声的时间序列数据传感器属于正常波动，即其数据波动属于非风险性异常，这种异常误差可能是由温度等外部环境所造成的。

我们可以通过夏皮洛-威尔逊检验方法和KS检验方法，从100个传感器中筛选出时序数据不满足正态分布或不满足均匀分布的传感器，对其进行下一步操作。

*Step2.*对通过夏皮洛-威尔逊检验方法筛选后的剩余可能风险性异常数据进行皮尔逊相关系数的计算，利用这些传感器时序数据间的皮尔逊相关系数的求解值和我们设定的值进行比较

其中表示传感器的编号。

最终，我们能够通过皮尔逊相关系数的值来确定传感器是具有风险性异常数据还是非风险性异常数据。

若皮尔逊相关系数大于0.65，则这对应的两个传感器之间相关性显著，具有较强的联动性，都属于风险性异常数据的传感器；若皮尔逊相关系数小于0.65，则对应的两个传感器之间相关性较弱，关联性较差，属于非风险性异常数据的传感器。

**5.1.4 模型的结果**

通过计算机编程求解(详见代码文件 第一题.cpp)，得到的结果传感器编号和数据属性的分类情况如下所示。

表1 问题一结果呈现表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 非风险性  异常数据 | | | | 风险性  异常数据 |
| 1 | 2 | 3 | 5 | 4 |
| 6 | 8 | 9 | 10 | 7 |
| 12 | 13 | 14 | 15 | 11 |
| 16 | 18 | 19 | 22 | 17 |
| 24 | 25 | 26 | 27 | 20 |
| 28 | 29 | 30 | 31 | 21 |
| 32 | 33 | 34 | 35 | 23 |
| 36 | 37 | 38 | 39 | 41 |
| 40 | 42 | 43 | 44 | 47 |
| 45 | 46 | 48 | 49 | 57 |
| 50 | 51 | 52 | 53 | 65 |
| 54 | 55 | 56 | 58 | 71 |
| 59 | 60 | 61 | 62 | 72 |
| 63 | 64 | 66 | 67 | 73 |
| 68 | 69 | 70 | 74 | 76 |
| 75 | 77 | 78 | 79 | 94 |
| 80 | 81 | 82 | 83 | 96 |
| 84 | 85 | 86 | 87 | / |
| 88 | 89 | 90 | 91 | / |
| 92 | 93 | 95 | 97 | / |
| 98 | 99 | 100 | / | / |

**5.1.5 模型的结果分析**

**5.2问题二的模型建立与求解**

**5.2.1 模型的工作原理**

根据题目的要求，我们以密度聚类思想为核心建立风险性异常数据异常程度的量化评价模型，采用百分制方法对每个时刻数据异常程度打分，并通过模型对异常分值最高的5个时刻进行评价与分析。

1).密度聚类思想

密度聚类的核心思想是只要样本点的密度大于阙值，便可以将该样本点添加到最近的簇中。我们考虑通过密度聚类的思想来计算风险性异常数据异常程度量化模型中的权重指标。

在数据集，参数的密度聚类模型中存在着几个基本概念：

a).-邻域：数据集中除了外的其它样本与距离小于的样本集合；

b).对象：若的-邻域中样本数量大于，则样本被叫做核心对象；

c).密度直达：若样本是的-邻域中的样本，则由为密度直达；

d).密度可达：存在样本序列，其中的核心对象是，的核心对象是，且由密度直达，则由为密度可达；

e).密度相连：，使和均由密度可达，则和密度相连。

因此，我们可以从密度可达的关系中到处最大密度相连的样本集合。应该有给定参数和簇，应该满足两个条件：

连接性:和密度相连，用数学符号表示为：，

最大性：，由为密度可达，可以推出

通过上述方法，可以确定一个簇中的核心对象点，并且以此对象核心点出发，不断向密度可达的-邻域扩张，得到一个包含核心点和边界点的最大化区域，区域中的任意两点是密度相连的。

**5.2.2 模型**

1)目标函数的确定

依据题目要求，我们要对每个时刻的数据异常程度进行打分，建立量化评价模型，确立目标函数：

其中表示第个簇指标的量化权重，表示第个簇中传感器指标的数值,为计算出的簇的数量。

2)初始方案的策划

*Step1*.在问题一模型皮尔逊相关系数的基础上，确定所有传感器之间的相关系数，并建立相关特性矩阵，以此作为传感器点在空间之间的距离；

*Step2*.明确问题要求是对风险性异常数据异常程度作量化评价模型，因此在之后的模型设计考虑过程中，需要将与核心对象传感器之间存在较强相关性，但并非属于风险性异常数据的传感器指标进行及时更改，防止指标权重信息产生误差；

*Step3*.通过第一步中建立的相关特性矩阵，我们进一步寻找彼此间存在较强相关性的传感器。利用密度聚类的思想，把所有在相关特性矩阵中反应的相对距离呈现在一维平面中，寻找出一个既能满足于传感器数据之间相关性筛选关系，又能满足簇的个数等基本要求的簇的分类方法和方式。以该种簇的分类方式将一维图像中所包含的传感器聚成一个簇。这样就可以形成几个互相之间并无显著关联特性的簇，并进一步对每个簇分析。

*Step4*.求解每个簇的最小半径。运用密度聚类思想中的密度直达、密度可达、密度相连、邻域和核心对象的概念，以满足连接性和最大性为目标条件，寻找每个簇中以任意一个传感器点为核心对象所求出的对应最小半径。

最终求出的每个簇的最小半径是所有最小半径中的最小值，核心对象是求出的最小半径所对应的核心对象。

*Step5*.以每个簇中的最小半径作为每个簇的权重，若簇中含有非风险性异常数据的传感器，则将该传感器的权重定义为0，不对其所在簇中的其它传感器产生任何影响。在最后计算任意时刻数据异常程度评价时，以每个簇的权重作为该簇中的所有含有风险性异常数据的传感器的权重。这里的每个簇的权重也就是模型的目标函数中的量化权重。

*Step6*.计算目标函数中每个传感器指标的数值，我们采取自相关系数表示每个传感器指标在任意时刻的数值，得到公式

其中常数表示滞后数，即两个同一时序数据中的时序差值。

*Step6*.最终通过密度聚类方法计算出的簇权重和每个传感器指标中的任意时刻自相关系数求解出最终量化后的风险性异常数据异常程度的量化指标，并选出异常分值最高的5个时刻及这5个时刻对应的5个异常程度最高的传感器编号。

**5.2.3 模型的求解**

根据问题二数学模型的描述，我们通过算法DBSCAN来解决密度聚类问题，通过矩阵和迭代运算来进行传感器的任意时刻自相关系数计算和风险性异常数据异常程度量化数值。

1)DBSCAN算法

算法也是以迭代为基础，我们先通过所有传感器的相关性矩阵数据，得到所有传感器在空间位置中的相对位置点。

a).DBSCAN算法中先假定一个点其所在的-邻域内存在个对象，创建一个以点为核心对象的簇；

b).其次寻找合并核心对象点直接密度可达的对象；

c).不断以核心对象点为圆心扩大簇的半径，扩充新的样本点进入簇；

d).直到样本点数量足够或簇的半径合适时，且在簇的半径略微变化时，并无新的样本点可以更新簇后，算法结束。

不断迭代各个簇，直到明确最终求出的簇中所包含的传感器编号和各个簇的的半径，即传感器异常程度数值计算的权重。

2).自相关系数计算

以每个传感器任意时刻的数据为目标点位，利用迭代方法求解以附件1中给的传感器任意时刻前的参数数据为基础的100个传感器在任意时刻的自相关系数。

传感器自相关系数作为传感器任意时刻的数值。利用目标函数的求解公式，带入传感器任意时刻的数值和各个簇的异常程度数值权重。量化求解出各时刻风险性异常数据异常程度的具体数值和5个异常程度最高时刻中5个传感器异常程度最高的编号。

**5.2.4 模型的结果**

通过计算机编程求解(详见代码文件 第一题.cpp)，得到的结果如下所示，可以看到最高异常分值的5个时刻以及其所对应的5个异常程度最高的传感器编号。

表2 问题二结果呈现表

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 第一高分 | 第二高分 | 第三高分 | 第四高分 | 第五高分 |
| 异常程度得分 | 72.81424496 | 72.76768961 | 72.76160024 | 72.75144977 | 72.63850413 |
| 异常时刻编号 | 1424 | 1423 | 1426 | 1422 | 1427 |
| 异常传感器编号 | 4 | 4 | 4 | 4 | 4 |
| 异常传感器编号 | 20 | 20 | 20 | 20 | 20 |
| 异常传感器编号 | 41 | 41 | 41 | 41 | 41 |
| 异常传感器编号 | 76 | 76 | 76 | 76 | 76 |
| 异常传感器编号 | 94 | 94 | 94 | 94 | 94 |

**5.2.5 模型的结果分析**

**5.3问题三的模型建立与求解**

**5.3.1模型的工作原理**

根据题目要求，建立风险性异常预警模型，提前发现当日23时至24时可能产生的风险性异常，并结合问题二的量化评价模型，给出四个时段的异常程度数值。我们运用ARIMA模型(自回归综合滑动平均模型)来预测和分析未来的1h内异常程度数值的变化情况。

1).ARIMA模型

ARIMA模型其实质就是差分运算和ARMA模型的结合，而ARMA模型又可以细分成AR模型和MA模型。因此我们利用差分运算和ARMA模型来实现ARIMA模型的应用。

a).差分运算

因为ARMA模型必须满足平稳性要求，即数据应该没有明显的上升或下降趋势，在一个恒定的区间内浮动。因此我们通过差分运算将具体数据转化为能够被ARMA模型所用的平稳性数据。当数据序列蕴含线性趋势时，我们可以用一阶差分实现趋势平稳；当序列蕴含曲线趋势时，我们运用二阶或三阶差分提取出原数据的趋势，从而实现平稳的特性；若序列含有固定的周期，我们能用步长为周期的长度差分，从原序列中过滤出周期的影响，实现趋势的平稳。

b).AR模型

AR模型是自回归模型，其变量当前的取值与过去的取值相关，即时间序列中前后存在相互影响的关系。其蕴含参数阶数，阶自回归模型可以表示为：

对于中心化的AR()模型而言，当时成立，即有

另外AR()模型的自相关系数具有拖尾性，且呈负指数衰减，不会在有限阶数后恒等于0；AR()模型的偏自相关系数具有阶截尾性。

c).MA模型

MA模型为滑动平均模型，其当前序列的数据变量的记忆是关于过去外部干扰的记忆，即当前时间序列可以表示成现在干扰值和过去干扰之的线性组合关系。其也存在参数阶数，阶移动平均模型可以表示为：

对于中心化的MA()模型而言，当时成立，即有

另外MA()模型的自相关系数是以阶截尾的，MA()模型的偏自相关系数具有拖尾性，其偏自相关系数不会在有限阶段后恒为零。

同时对AR()模型和MA()模型引入延迟算子，可以得到中心化AR()和中心化的MA()模型

**5.3.2模型**

1)目标函数的确定

根据题目要求和ARMA模型的需求，我们以*AIC*准则函数作为目标函数，其既考虑模型对原始数据的预测准确性，也考虑了模型中所含待定参数的个数，能够很好地适用于ARMA模型的检验。

其中是残差的方差，表示白噪声，为样本容量，和为模型的阶数。

我们所需要的目标函数是，准则函数的最小值。由其表达式可以看出，该函数受到第一项和第二项的影响。我们最终利用该方案求解出来的参数和，来确定ARMA模型中的系数向量，通过系数向量的解，计算所有预测传感器时序数据的具体数值。

当第一项参数较大时，第二项参数会较小，函数处于过拟合状态；当第二项参数较大时，第一项参数会较小，函数处于欠拟合的状态。而第一项参数能够体现模型拟合的好坏程度，随阶数增大而减；第二项参数能够体现模型本身的复杂程度，随阶数增大而增大。

2)初始方案的策划

*Step1*.观察题目附件1中所给出的函数时序数据序列值，对100个传感器的时序数据有一个大致的认识，以便于接下来能够寻找更好的方法对这一组的数据进行处理和分析；

*Step2*.平稳性检验。因为ARMA模型对传感器的时序数据存在平稳性的要求，因而我们采取差分运算的方法，去除时序数据中所含有的非平稳性因素。无论时序数据是线性的还是曲线的，或者是周期性的，都可以通过单次或多次差分运算来提取出数据的平稳性，来满足ARMA模型的需求。

*Step3*.白噪声检验。我们通过对残差进行检验来确定传感器的时序数据是不是存在白噪声，具体检验的方法是问题一模型中的夏皮洛-威尔逊检验和KS检验来进行，通过检测时序数据是否满足正态分布或均匀分布来判断。

*Step4*.若传感器时序数据参数通过白噪声检验，则可以直接通过其确认的参数和，以及系数向量来求解传感器的在未来1h内的预测数据值；反之，则需要重新改变参数数据，重新进行白噪声检验，寻找最优的满足*AIC*准则函数的最小值，从而求解传感器的预测数据量。

*Step5*.当时序数据不通过检验后，需要重新确定参数和的数据，因为参数和均为整数，因此我们选择采用有限区域枚举的方案，对参数和在的区间中任意取整数对，寻找*AIC*准则函数的最小值；

*Step6*.通过上一步确定的参数和，进一步确定ARMA模型中的系数向量，从而进一步计算残差值。我们将假设枚举的参数和的值以及以此计算出的残差的方差代入*AIC*准则函数中，计算出该参数和取值情况下的目标函数值；

*Step7*.进一步对参数和进行枚举，得到很多种情况下的系数向量和残差值，利用这些数据计算出参数和在各个取值情况下的最终*AIC*准则函数解。对所有取值情况下的最终函数值进行筛选，选取最小情况时的参数和及系统向量。并以此为依据，求解各传感器的预测数据数值。

**5.3.3模型的求解**

我们在运用数学模型求解传感器预测结果的过程中，所面临的两个问题分别是*AIC*准则函数和自回归移动平均模型中的系数求解。

1). *AIC*准则函数

*AIC*准则函数是衡量统计模型拟合性能的一种标准，我们采用的是因子分析模型，利用该准则确定公共因子的个数，使得公共因子不仅可以解释原始变量较多的信息情况，并且还不会产生由于因子较多而导致的信息冗余。

其中还存在关系式有：

其是极大似然估计的本质，等价于使K-L距离达到最小的参数估计，满足*AIC*准则函数的实质。因此我们运用ARMA模型中的准则函数最小为目标值，求取传感器预测数据解。

2).自回归移动平均模型中的系数求解

由ARMA(p,q)模型中的特征，我们可以知道中心化ARMA(p,q)模型是当时成立，与AR模型的判定条件相一致，这时，满足序列为：

可以从上述表达式看出，当前时间序列值，与过去的时间序列值乘线性关系。当假设ARMA(p,q)模型中的(p,q)解值后，可以通过附件1中的数据和ARMA模型中的(p,q)数值，进一步求解出自回归移动平均模型中的系数以及残差。并以此作为白噪声检验的依据，若满足白噪声检验，则通过该系数向量求解序列数据的预测数值；若不满足白噪声检验，则在有效区域内枚举可能的(p,q)解值对，直到最终的准则函数满足白噪声检验为止。

**5.3.4模型的结果**

利用算法求解在无法知晓栅格地图全貌信息下的最优路径规划模型(详见代码文件 第二题K.cpp)。我们通过随机生成障碍物的方式产生栅格地图模型，带入我们的算法求解的结果如下。

**5.4问题四的模型建立与求解**

**5.4.1 模型的工作原理**

**5.4.3 模型**

1)目标函数的确定

2)初始方案的策划

*Step1*.

*Step2*.

*Step3*.

**5.4.4 模型的结果**

利用算法求解在无法知晓栅格地图全貌信息下的最优路径规划模型(详见代码文件 第二题K.cpp)。我们通过随机生成障碍物的方式产生栅格地图模型，带入我们的算法求解的结果如下。

**5.4.5 模型的灵敏度分析**

六、模型的综合评价和推广

**6.1 模型的综合评价**

**6.1.1 模型的优点**

本文在

本文在。

本文在。

**6.1.2 模型的缺点分析**

本文。

。

**6.2模型的推广**

。

七、参考文献

[1]刘永建,曾国辉,黄勃,等.改进蚁群算法的机器人路径规划研究[J].电子科技,2020(1):13-18.

[2]任红格,胡鸿长,史涛.基于改进蚁群算法的移动机器人全局路径规划[J].华北理工大学学报(自然科学版),2021,43(02):102-109.

[3]杨凌耀,张爱华,张洁,宋季强.栅格地图环境下机器人速度势实时路径规划[J/OL].计算机工程与应用:1-8[2021-05- 24]. http://kns.cnki.net/kcms/detail/11.2127.tp.20210514.1007.008.html.

[4]蒋林,方东君,周和文,黄惠保.基于射线模型的改进全局路径规划算法[J/OL].电子学报:1-10[2021-05-24]. http://kns.cnki.net/kcms/detail/11.2087.TN.20210429.1538.002.html.

[5]彭湘,向凤红,毛剑琳.一种未知环境下的移动机器人路径规划方法[J].小型微型计算机系统,2021,42(05):961-966.

[6]杨光辉.基于人工智能优化算法的大型舰船紧急疏散路径规划[J].舰船科学技术,2021,43(08):52-54.