海上目标自动识别

问题一

1.1 时空位置信息误差修正模型和有效性检验

1.1.1 误差修正模型

附件一数据描述了四种信号源批号的传感器记录，内容包括时间、纬度、经度、传感器类型、记录值、信号源批号。本节主要针对四种信号源数据中的时间、经度、纬度数据进行误差评估并建立数学模型修正之，最后评价该模型的修正效果。

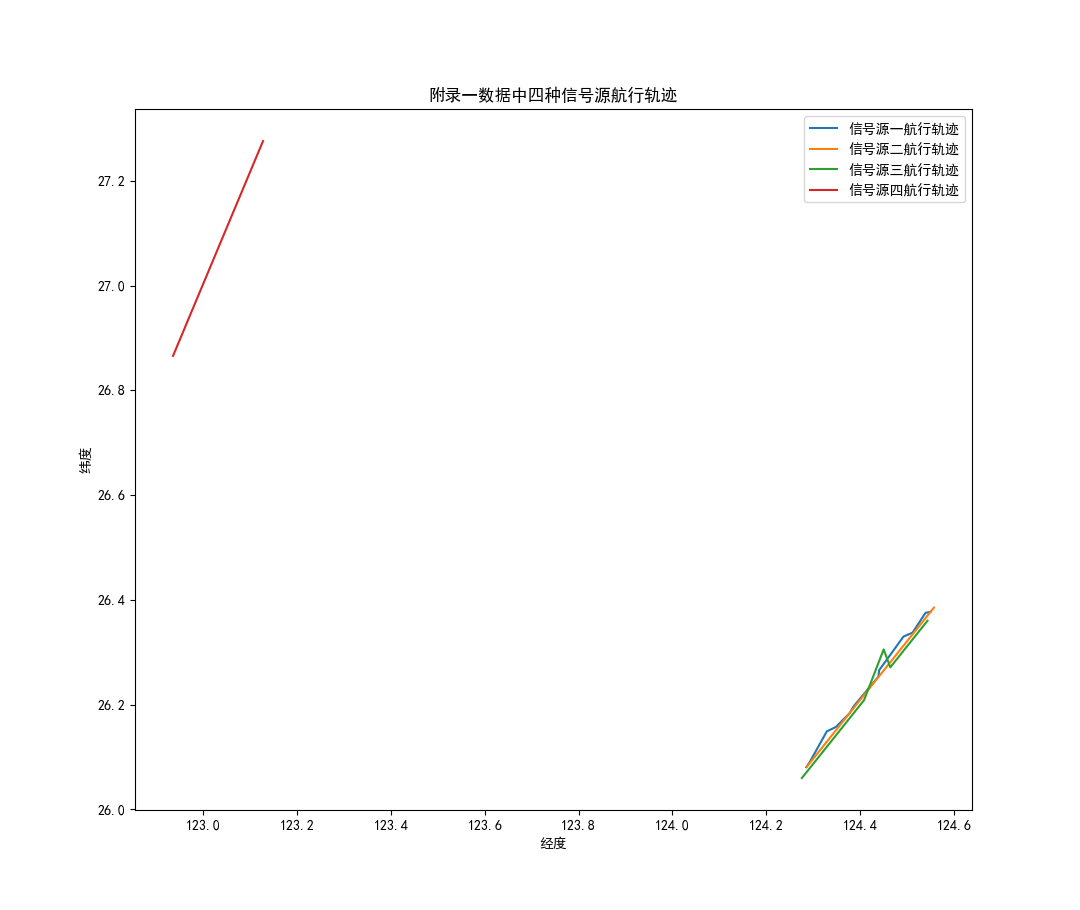


图 1.1 附件一数据中四种信号源航行轨迹对比

如图1.1所示，建立以经纬度为坐标轴的平面直角坐标系，分别绘制四种信号源航行轨迹，编号为信号源一、信号源二、信号源三、信号源四，其中编号原则为信号源批号差异，我们认为题意所指的同一信号源批号的数据是同种信号源发出的。图中信号源一和信号源三轨迹更加曲折，而信号源二和信号源四的轨迹更加平直。

我们推测时空数据的误差之一是源自电磁信号在非均匀介质传播自身造成的误差，并假定出于燃料消耗最少和航行时间最短的原则，船舶的航迹应尽可能地接近直线，用船舶航行轨迹的拟合直线代表船舶的真实航行轨迹，因此我们从图1.1可以定性地推断出信号源一和信号源三的时空数据误差大于信号源二和信号源四的误差。

建立数学模型减小其误差的首要前提是用量化的方法描述误差，基于前文的推理我们首先给出以下两条假设：

·船舶实际运动轨迹应尽量接近直线

·传感器记录的每条航行轨迹对应的拟合直线为该船舶的实际航行轨迹

我们拟采用航迹接近其拟合直线的接近程度来描述时空数据的误差大小，即航行轨迹越曲折，说明传感器采集的时空数据误差越大，轨迹越平直，时空数据越精准。

至此，我们可以抽象出此问题的数学语言描述：船舶航行轨迹为平面中的折线段，平面中折线段与其拟合直线的接近程度代表传感器的经纬度数据记录精度，现求一数学模型，传感器采集经纬度数据的同时，将每次传感器记录的经纬度数值输入到该模型，输出一个修正的经纬度数值替代原始数值并记录，以期最终的信号源航行轨迹与原始经纬度数值相比更加接近直线。

现在介绍如何评价一折线段接近直线的接近程度，在数学上常用最小二乘原理求离散点的回归直线方程，衡量回归方程优劣的方法是求离散点与回归直线的残差平方和，同样我们用残余误差平方和的均值来评价折线段接近其拟合直线的接近程度，本模型的用途即是减小轨迹点与其拟合直线的平均残差平方和。

考虑到实际应用场景，传感器在某时刻记录的经纬度数值受之前若干个记录结果和本时刻初始（未经模型修正）经纬度数值影响，且越接近本时刻的数据记录对本时刻经纬度数值的影响越大。我们用d：d0，d1，d2，…, dn表示经纬度数据的初始值，用D：D0，D1，D2，…, Dn表示经纬度数据修正值，用函数R()表示修正（Revise）方法，即我们的数学修正模型，则此问题可以看作求满足一定条件的将集合d映射到集合D的映射方法。“一定条件”即指上文修正经纬度数值绘制的航迹更加接近直线。有如下关系式：

Di = R(di,di-1,di-2,…,d0) (1.1)

式（1.1）描述了修正模型的输入输出，输入为当前时刻以及之前时刻的经纬度数据输入，输出为当前时刻修正的经纬度数据。

考虑到基于前文两点假设的船舶航向始终固定，因此用当前时刻至初始时刻的所有经纬度点修正船舶航向，用船舶前一时刻的速率代表前一时间段的航行速率，在前一时刻经纬度数值确定的情况下，便可针对本时刻的经纬度数值进行修正。我们建立的模型的修正原理如下：

1. 求初始时刻至当前时刻所有初始经纬度点的拟合直线，以此直线方向代表带修正时刻船舶航行方向；
2. 求前一时刻至当前时刻船舶位移量，对航行时间求导得到平均速率，以此速率值代表待修正时刻至前一时刻的船舶航行速率；
3. 以前一时刻船舶所在位置为起点，以1、2步求得的速度矢量为该时间段内船舶航行速度，求得本时刻船舶所在位置的预测值；
4. 将3中的位置与本时刻传感器探测到的船舶位置求和取平均，作为本时刻轮船位置经纬度数据的修正结果。

我们用Python语言编写程序实现了上述过程，并且将四种信号源修正经纬度数据与原始经纬度数据绘图比较。

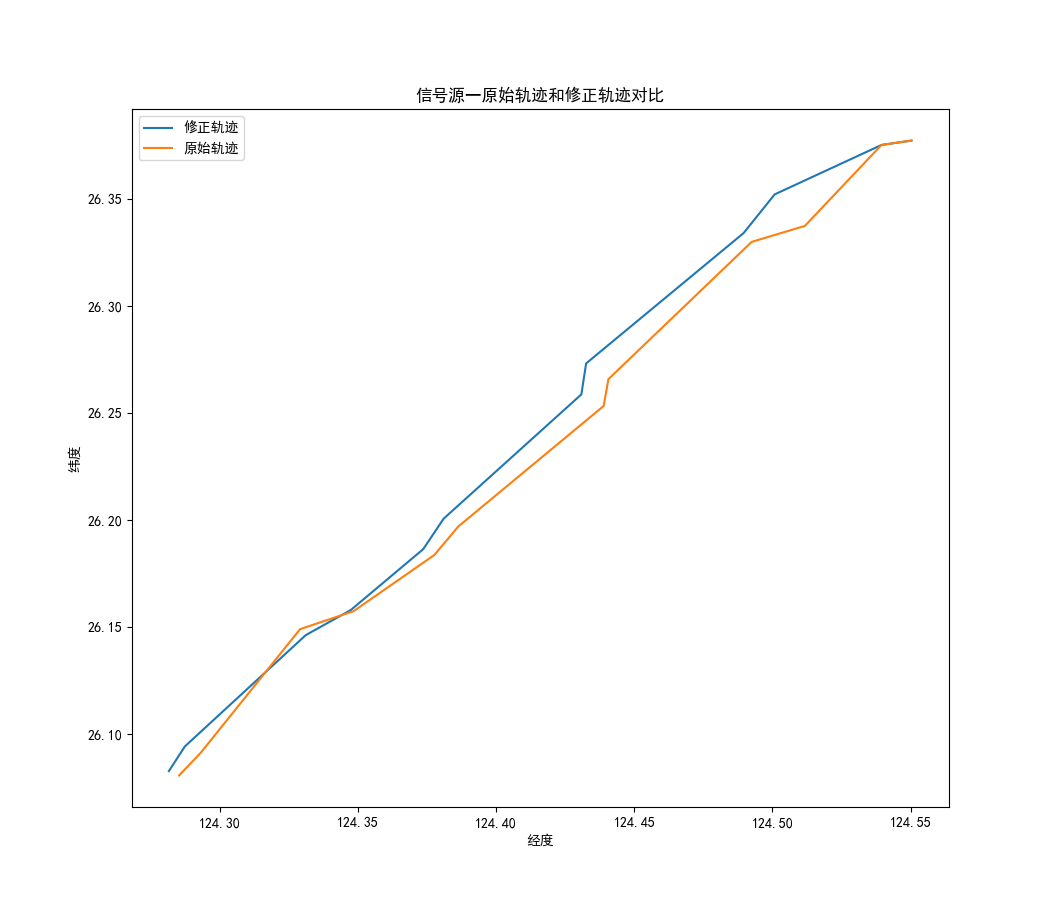


图1.2 信号源一原始轨迹和修正轨迹对比

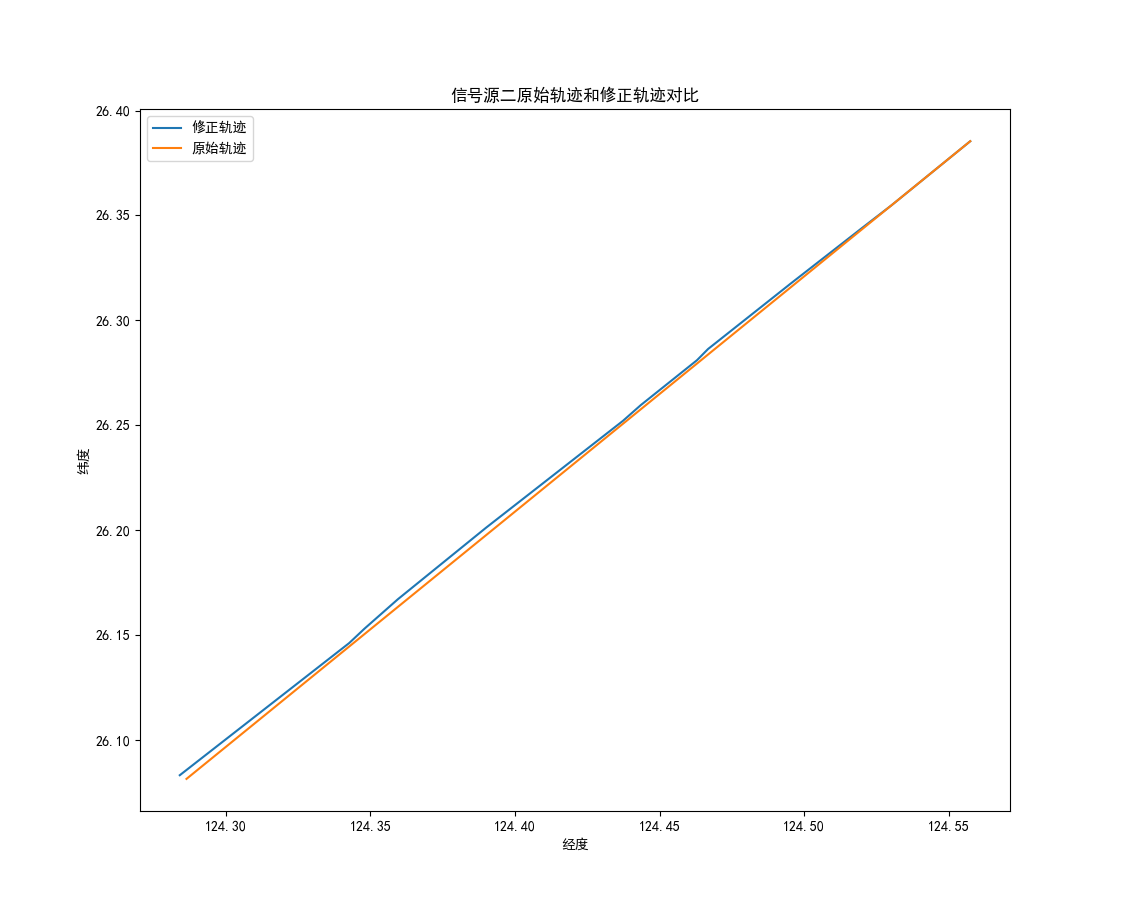


图1.3 信号源二原始轨迹和修正轨迹对比

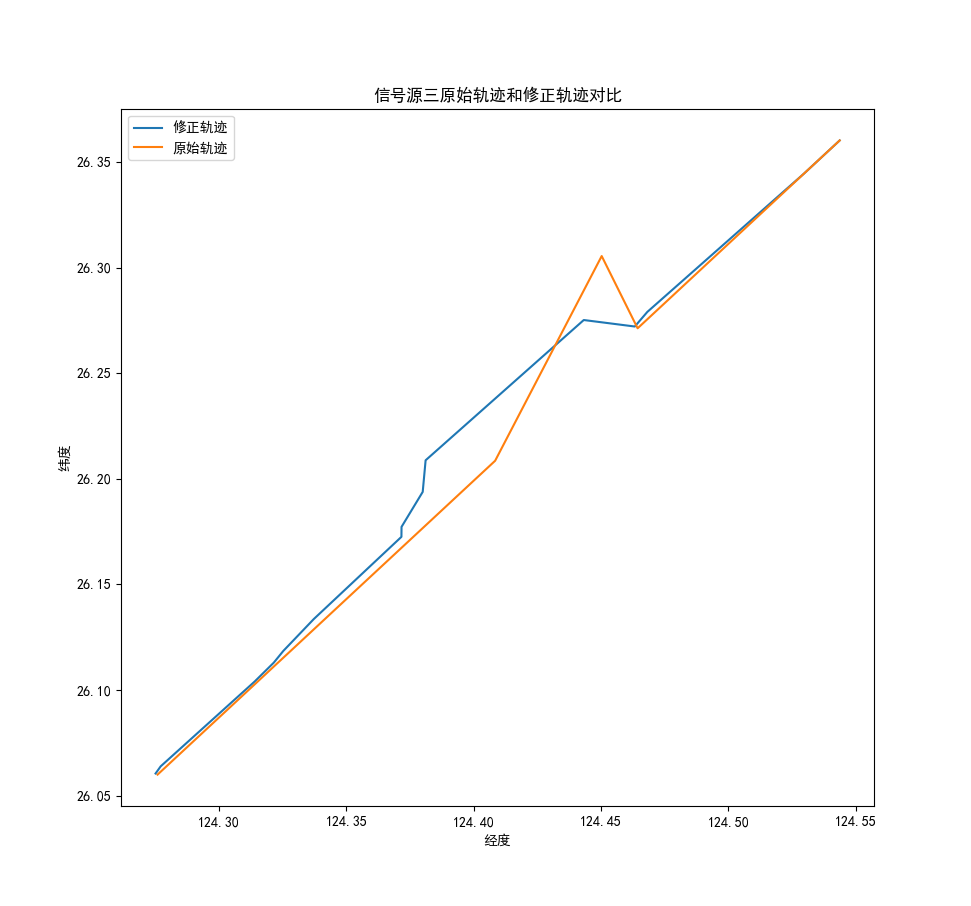


图1.4 信号源三原始轨迹和修正轨迹对比

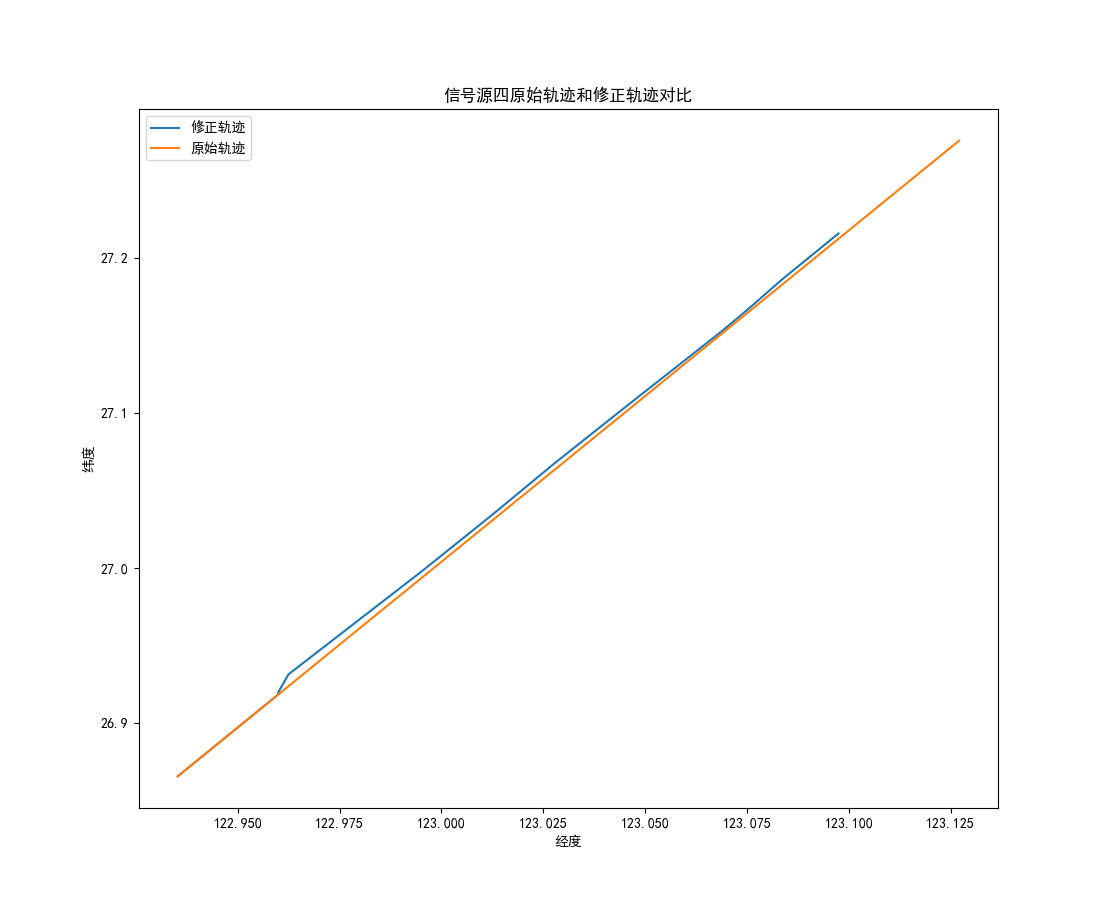


图1.5 信号源四原始轨迹和修正轨迹对比

由以上四张图片可以看出，针对信号源一和信号源三的修正效果明显，由于信号源二和信号源四本身更接近直线，因此效果稍弱，也说明信号源二和信号源四本身更接近其拟合直线，误差很小，不需要重新建立模型修正。

下节将定量判定本修正模型有效性。

1.1.2 有效性检验

如上文所述，我们拟采用平均残余误差平方和衡量折线段与其拟合直线的接近程度。有离散点xi,yi,其拟合直线为y = kx + b,平均残余误差平方和由下式计算：

Sm =  (1.2)

Sm表示平均残余平方误差平方和，数值越小，说明折线段越接近其拟合直线，也即经纬度记录数据误差越小。本节分别求出四种信号源初始经纬度数据的平均残余误差平方和和修正经纬度数据的平均残余误差平方和，通过比较二者证明本模型的优越性。

表1.1 各信号源经纬度数据初始精度与修正后经度比较

|  |  |  |
| --- | --- | --- |
| 信号源编号 | 原始经纬度数据Sm | 修正经纬度数据Sm |
| 信号源一 | 4.4858170255788915e-05 | 2.1551532159581447e-05 |
| 信号源二 | 4.188629709397054e-11 | 5.3723867165193035e-07 |
| 信号源三 | 15．423761162676271e-05 | 8.9947107772433364e-05 |
| 信号源四 | 1.8911361860828413e-10 | 23030.469118566555e-10 |

由表1.1可以看出，经过该模型修正后，信号源一和信号源三记录精度（平均残余误差平方和）度均有提高，经修正经纬度数值误差更小，航行轨迹更加平直。而信号源二和信号源四由上节轨迹图可看出其本身平均残余误差平方和极小，不需通过本模型修正。在实际应用中应首先剔除此类轨迹的经纬度数据并认为其是足够精确的，之后再根据本模型修正轨迹较为曲折的其它大误差数据。

由此我们认为此数学模型对传感器记录的原始时空数据误差减小有一定作用。

1.2 信号特征信息误差修正模型

1.2.1 误差修正模型

从附件一所给的数据中可以观察到，不同于时空位置信息在船舶航行过程中随时间连续变化的特点，信号特征信息理想情况下为常量，同一批号的信号源发出的特征信号信息基本维持不变，因此较于针对时空位置信息的数学修正模型，本节建立的模型更为简单。附件一给出了四种信号源批号的三种特征信号信息，本节将提出针对这十二组数据提出修正模型并验证有效性。

同理首先建立量化误差方法，在数学中对于一维离散数据常常用均方误差衡量数据偏离均值的波动情况，方差越大数据波动越大，对应地，同一信号源批号的同一信号特征信息波动越大说明其误差越大，因此理想情况下传感器接收到的特征信号信息应恒为常数。本节也将用离散数据的方差作为衡量数据精度的指标，方差越小，说明数据越精确。根据以上推理，可以得出本节建立数学模型的作用是最小化特定信号源批号的特定特征信号信息的均方误差。

每一时刻的特征信号信息同样受之前所有时刻的特征信号信息影响，因此本模型的输入为当前时刻和之前所有时刻的传感器接收到的特征信号信息，输出为修正的特征信号信息。考虑到应尽量使本时刻的修正值更加接近于特征信号数值的均值，我们提出的模型如下：

1. 取特定信号源批号、特定特征信息时刻之前所有时刻的记录数值平均值；
2. 求当前时刻初始值与之前所有时刻的平均值的平均值作为本时刻的修正值记录。

针对附件中的十二组信号，我们分别进行了数值修订，并计算得出其对应的均方误差，用以比较该模型的误差修正效果。

1.2.2 有效性检验

我们将每组信号分别输入上节模型修正，得出了修正结果和原始数据方差，如下表所示：

表1.2 附件一中12种特征信号修正前后精度对比

|  |  |  |
| --- | --- | --- |
| 特征信号信息编号 | 初始值均方误差 | 修正值均方误差 |
| 1 | 0.00501094576389 | 0.00175540654605 |
| 2 | 0.00760188213018 | 0.00632761817811 |
| 3 | 0.00810335246094 | 0.0066553737075 |
| 4 | 0.0609384823669 | 0.0507236816052 |
| 5 | 0.0106928413889 | 0.00217610844163 |
| 6 | 0.103516824379 | 0.0656231223147 |
| 7 | 0.0132924521484 | 0.0023570911609 |
| 8 | 0.0471298436686 | 0.0298931953109 |
| 9 | 0.0274305555556 | 0.0111989825752 |
| 10 | 0.0 | 0.0 |
| 11 | 0.0246625 | 0.00298406956202 |
| 12 | 0.0253467455621 | 0.00271762556229 |

表1.2中第一列为12组特征信号编号，第二栏是初始数值计算的均方误差，第三栏是经模型修正后计算的均方误差，可以看出每一组数据经修正后方差均小于修正前，即修正后的数值更趋于稳定，误差更小。由此认为上节建立的误差修正模型是有效的。

问题三

3.1 问题分析

由附件三数据可知，同一船舶可携带多种型号信号源；由附件四数据可知，同一信号源可发出多种信号数据，值得注意的是，A类信号源发出的信号都相同，因此可用MMSI编号作为识别A类信号源的唯一标识。假设一艘船舶上有m种信号源，每种信号源平均发出n种信号，那么传感器对其可能捕获的信号就有m \* n种之多，按附件三数据给出的某船舶至多有6中传感器，附件四数据某信号源至多发出8种信号计算，那么此船舶可能会发出6 \* 8 = 48种信号之多。附件五中数据最复杂的一项是目标编号为101的数据，这个目标共发出了17种信号，假设按数据表索引，用目标发出的这17种信号去分别匹配数据库中的48种信号，这种穷举法运算量巨大，显然不适合用来解决此问题。

本题所给的信号源数据库种每种信号都有多种数据表现形式，因此每种信号源都可看作一个由高维张量描述的点，比如下列R类信号：

表3.1 某R类信号源数据示例

|  |  |
| --- | --- |
| **信号源名称** | **DRZ-60** |
| 传感器类型 | R |
| RA\_1\_1 | 3 |
| RA\_1\_2 | 2011, 2105.6 |
| RA\_2\_1 | 2 |
| RA\_2\_2 | 15.200, 20.100, 30, 40.400 |
| RA\_3\_1 | 1 |
| RA\_3\_2 | 1001.6, 1105.3, 1201.7 |
| RA\_4\_1 | 2 |
| RA\_4\_2 | 0.78 |

显然用一个8维的高维张量描述一个R类信号源有极大的冗余，因此理论上一定可以将各信号源的描述方式由高维张量降低为低维张量。降维的通常方法是投影，但本节建立模型的降维过程不采用计算量太大的投影法。但明确一点，降维不可过度，因为过度降维会模糊各信号源的边界，导致对给定信号信息无法准确分类，因此降维的程度大小取决于一下两点：

·数据库中的信号源数据保证足够的区分程度以满足信号信息分类；

·尽量降到计算开销最少、复杂度最小的更低维度。

这里引出了一个问题，如何判定降维后的数据是边界足够清晰的，能满足信号信息分类需求？下节将对此进行讨论。信号源的描述方式（高维张量）可以称为该信号源的特征（feature），高维张量描述的信号源特征（表3.1）成为复杂特征，低维张量描述的信号源特征成为简单特征。

前文已经说明船舶-信号源-信号数据之间的关系是树状的，从数据结构的角度来说，船舶是根节点，信号源是船舶的分支，信号数据是信号源的分支，为该结构的子节点。附件五数据给出了多个目标的信号数据信息，题目要求反推出这几类目标分别是哪几种船舶。解决思路就是从该数据结构的子节点依次向上遍历个节点，最终找到根节点的船舶编号或名称信息，至此分类完成。

附件五中数据的信号特征信息也是高维张量描述的，如果不加以转换，遵循上述思路为其分类时将会出现用高维张量匹配高维张量的局面，不利于为题求解，因此出于简化计算、加快匹配速度的考虑，对目标物信号特征信息也采用降维处理，使其能够匹配数据库中降维处理之后的信号源数据信息，从而实现快速匹配。

单条目标物信号数据信息可能匹配到多条信号源信息，多条信号源信息可能匹配到多条船舶信息，因此对单条信号信息的溯源结果可能出现多条船舶，而题目所说的同一“目标编号”的信号源可视为来自同一船舶目标，附件五中数据船舶存在多条信号信息记录，由于每条记录均可匹配到多条传播，最终的匹配结果则可用一种概率密度函数来描述，即认为匹配结果出现次数较多的船舶为正确结果的可能性较大，并且对于所有匹配到的结果，其可能正确的概率之和为1.

Ri为同一信号源特征信息多次匹配结果,P(Ri)为其正确的概率，这样就可以构建题目要求的可信程度评价方式。

3.2 模型构建

3.2.1 信号源特征降维

R类信号源由8种参数描述，L1类信号源、L2类信号源、A类信号源由3种参数描述。现将其转化为由单一参数描述的低维特征，并论证此种做法的可行性。

观察数据库中各信号源特征数据，可得到以下特点：

·R类信号源数据描述最复杂，多达8个参数，某些参数又由多个数值描述；

·各数据取值范围不尽相同，差异大；

·某些数据项数值固定，始终不变，此类在分类时数据无参考价值；

·A信号源数据的MMSI编码可作为其唯一识别依据；

·传感器记录值虽有起伏，在数量较大的情况下可取平均值近似处理。

附件五中的传感器记录数据如题所述存在误差，可用题目一中给定方法修正，以提高分类可信度。

本模型针对高维信息特征的降维处理方法为对应值简单相加，如此可将8维R类信号数据、3维L1类信号数据、3维L2类信号数据、3维A类信号数据处理为统一的一维格式。如此一来，数据库中的每种每种信号源皆可用一单一数值来描述，此数值即信号源新的低维特征。

现论述此种降维方式和信号特征描述方式的可行性。

从计算复杂度上讲，一维特征数据之间的相互匹配远远小于于8维或3维特征数据的匹配难度，此种处理方式满足上节所述的“尽量降到计算开销最少、复杂度最小的更低维度”。后续针对“数据库中的信号源数据保证足够的区分程度以满足信号信息分类”这一要求进行可行性论证。

在对采集信号数据进行分类时，同样先将其转化为一维特征描述数据，与信号源数据库中经降维处理的特征进行对比，归类即可得到结果。现在需要考虑的是数据库中新的一维特征数据差异是否足够大、能否满足对采集信号数据的分类要求。

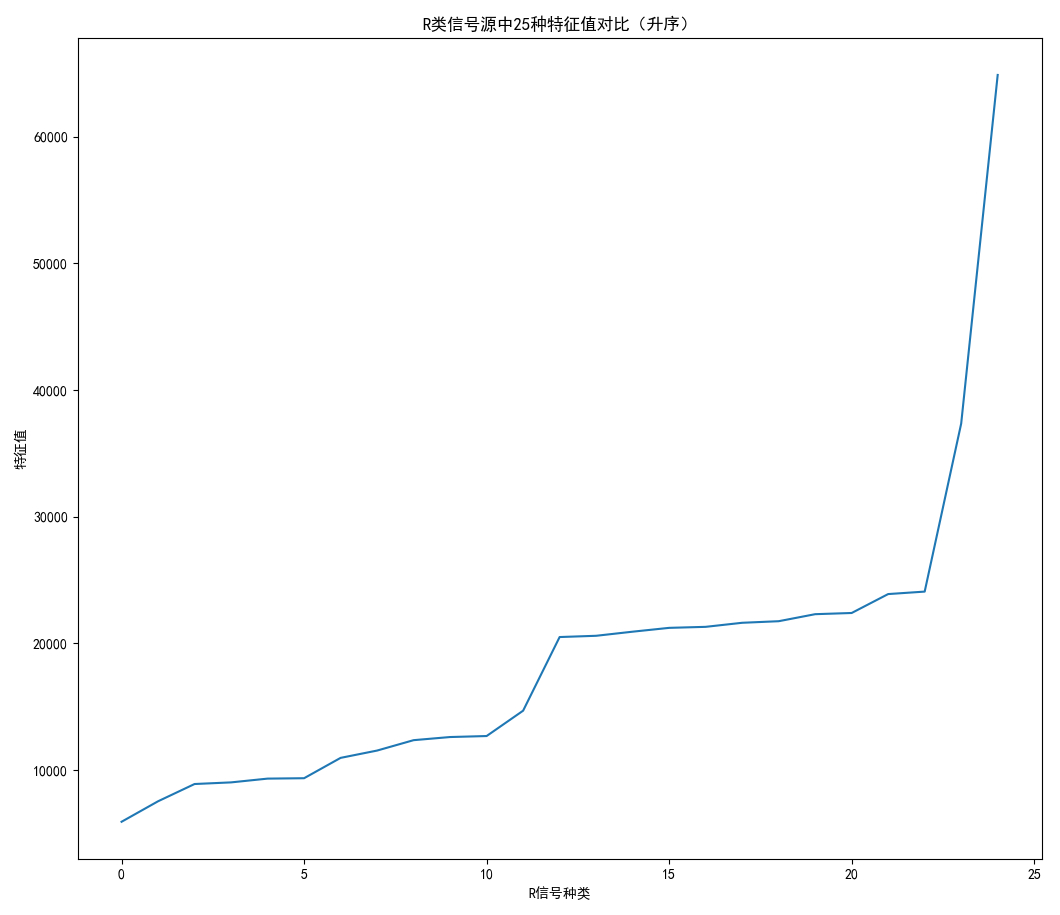


图3.1 数据库中R类信号源降维后特征值对比

图3.1反映了R类信号源降维单一特征值描述数据后的分布状况，图中按照升序将其排列，其最小值在500以下，最大值超过6000.

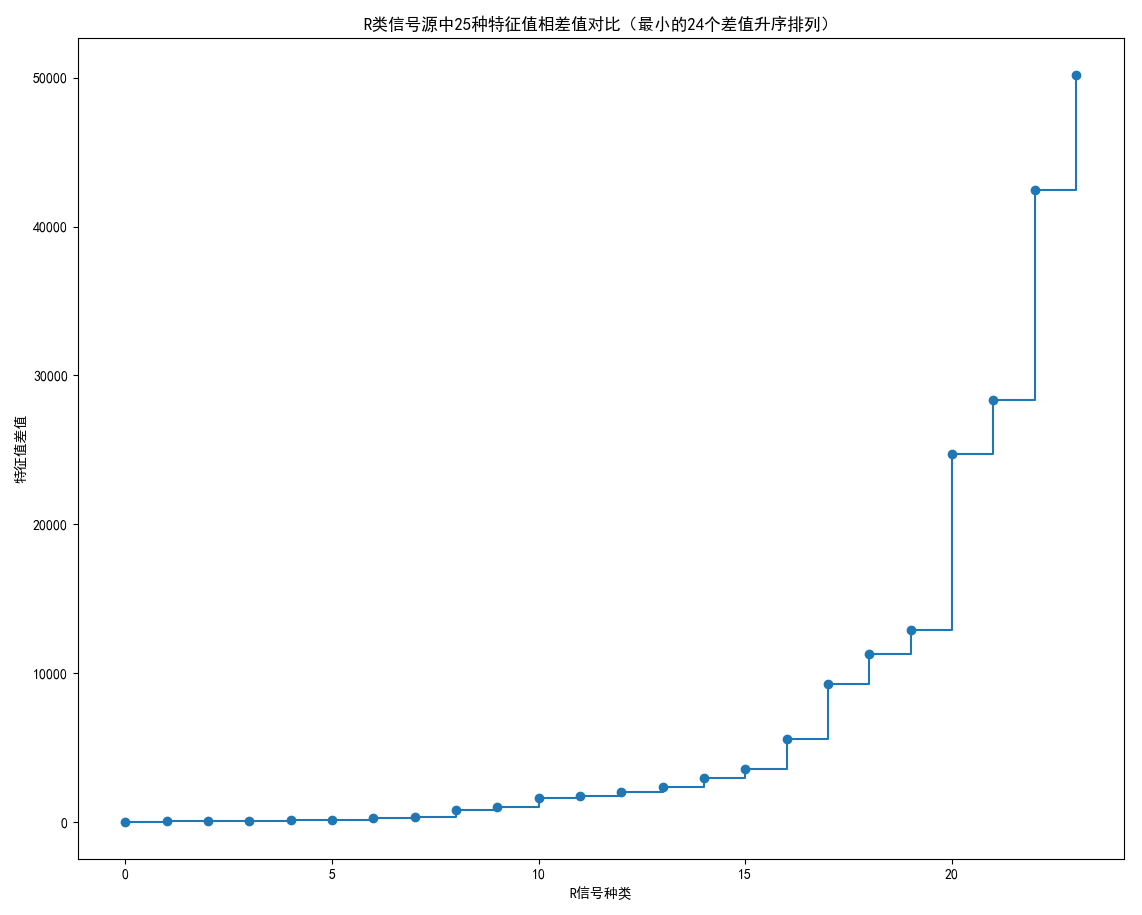


图3.2 R类信号降为一维后其相互间差值变化

如图3.2所示，我们将数据库中25种R类信号源按简单相加法降维以为特征后按照升序排序，并将相邻数据作差取绝对值，将这些绝对值按升序排列，得到了上图折线段，这张图反映了25种R类信号差异情况，其最小差值在30左右，最大差值大于5000，由此可知，只要附件五中的传感器采集到的R类数据经过相同降维方式处理之后插值能够超过30，那么即可认为以数据库中的一维特征数据为分类依据，样本理论上是可分的。下表列出了24个详细的差值数据。

表3.2 R类信号源特征数据降维后差异情况

|  |  |
| --- | --- |
| **序号** | **差值** |
| 1 | 31.2 |
| 2 | 80 |
| 3 | 94.5 |
| 4 | 98.9 |
| 5 | 123.96 |
| 6 | 128.8 |
| 7 | 302.149 |
| 8 | 323.687 |
| 9 | 820.31 |
| 10 | 999.194 |
| 11 | 1605.54 |
| 12 | 1788.48 |
| 13 | 2000.5 |
| 14 | 2337.44 |
| 15 | 2979.594 |
| 16 | 3586.59 |
| 17 | 5620.374 |
| 18 | 9265.25 |
| 19 | 11284.64 |
| 20 | 12929.79 |
| 21 | 24743.52 |
| 22 | 28323.22 |
| 23 | 42468.62 |
| 24 | 50180.4 |

由上表可看出，此种降维处理方式保留了原始信号源差异信息，能够作为后续对附件五中数据进行信号源类别分类的依据。

同理可分别论证L1类信号源、L2类信号源、A类信号源按此种方式处理的合理性，在此不再赘述论证方法，仅给出参考图表。

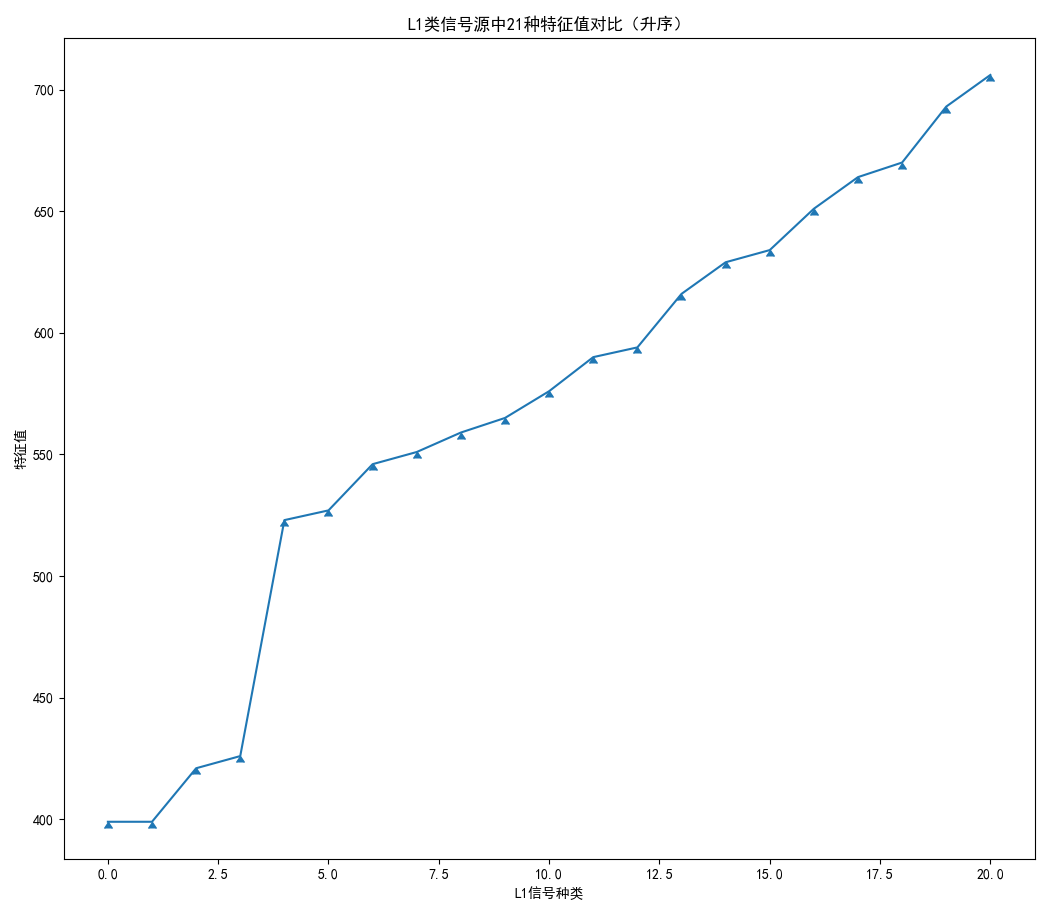


图3.3 L1类信号源降维后数值分布情况

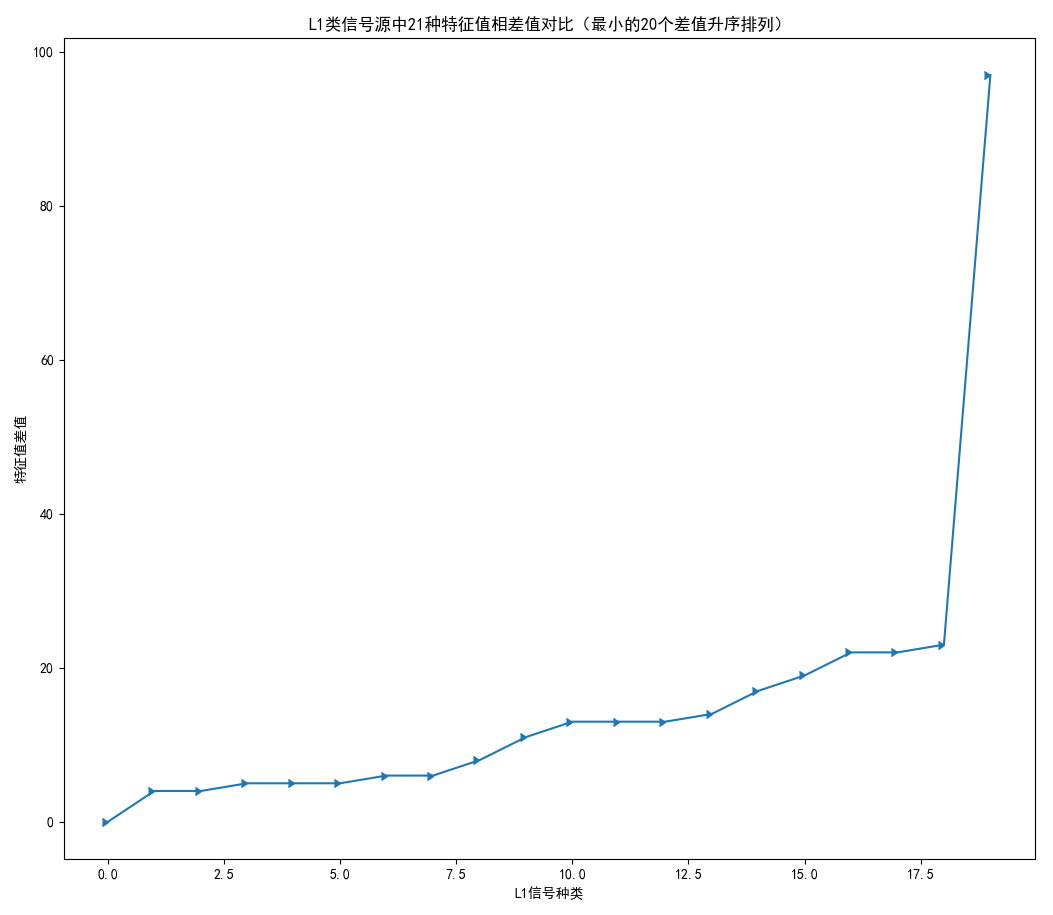


图3.4 L1类信号源降维后差异情况

表3.3 L1类信号源降维处理后详细差异数值

|  |  |
| --- | --- |
| **序号** | **差值** |
| 1 | 0 |
| 2 | 4 |
| 3 | 4 |
| 4 | 5 |
| 5 | 5 |
| 6 | 5 |
| 7 | 6 |
| 8 | 6 |
| 9 | 8 |
| 10 | 11 |
| 11 | 13 |
| 12 | 13 |
| 13 | 13 |
| 14 | 14 |
| 15 | 17 |
| 16 | 19 |
| 17 | 22 |
| 18 | 22 |
| 19 | 23 |
| 20 | 97 |

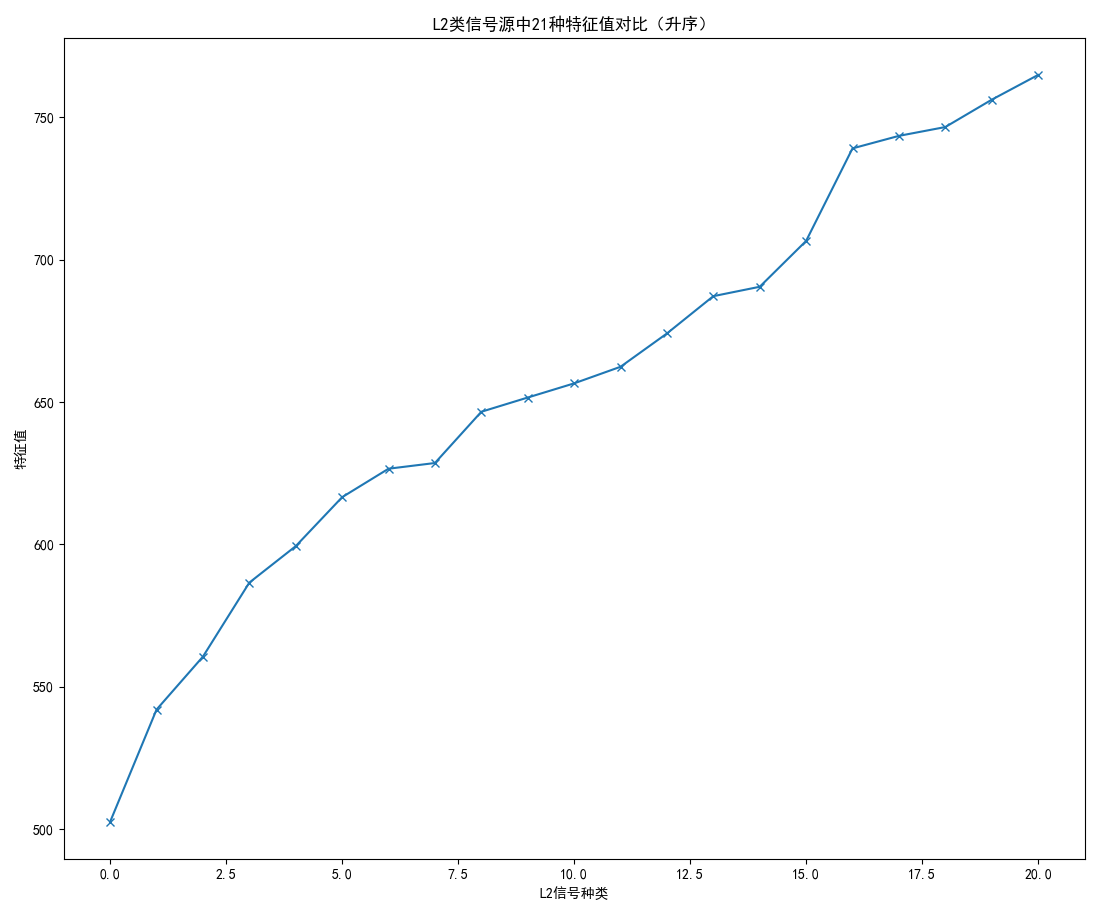


图3.5 L2类信号源降维后数值分布情况

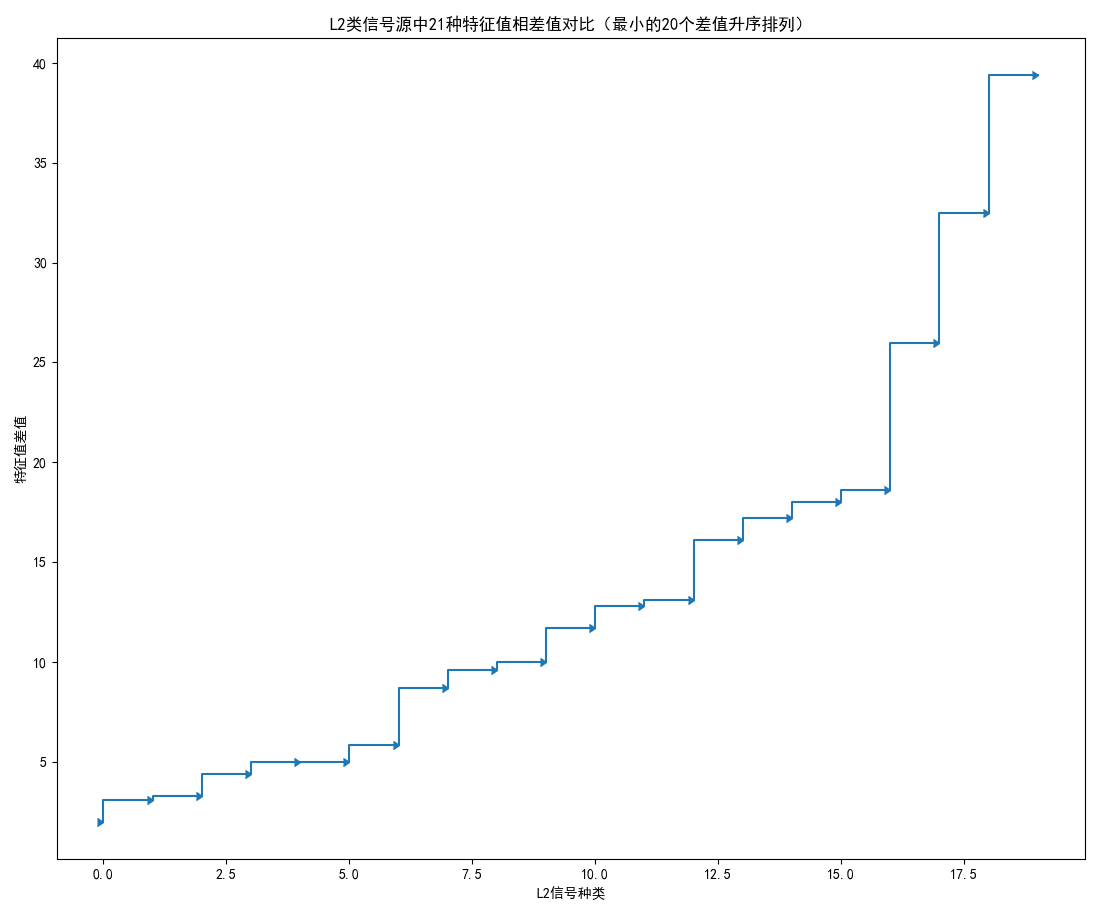


图3.6 L2类信号源降维后差异情况

表3.4 L2类信号源降维处理后详细差异数值

|  |  |
| --- | --- |
| **序号** | **差值** |
| 1 | 2 |
| 2 | 3.1 |
| 3 | 3.3 |
| 4 | 4.4 |
| 5 | 4.975 |
| 6 | 5 |
| 7 | 5.825 |
| 8 | 8.7 |
| 9 | 9.6 |
| 10 | 10 |
| 11 | 11.7 |
| 12 | 12.8 |
| 13 | 13.1 |
| 14 | 16.1 |
| 15 | 17.2 |
| 16 | 18 |
| 17 | 18.6 |
| 18 | 26 |
| 19 | 32.5 |
| 20 | 39.4 |

对于A类信号，如前文所述，将其MMSI编号作为唯一识别依据，不再做降维处理。

本节论证了简单相加法对高维数据降维处理在本题中的有效性，其最小差异值将作为判定附件五降维处理的各类别信号的分类依据，将目标信号源特征值与数据库信号源特征值差值取绝对值，该值在误差允许范围内落在哪个信号源类别区间，即可认为该目标信号来自该信号源。

3.2.2 模型建立——针对本题的具体目标识别方案

上节处理了高维信号特征数据，将各类信号源特征转化为单一参数描述数据，本节提出具体的目标物识别方法。

先抽象问题：集合A为传感器采集目标物信号特征，集合B为信号源名称与其特征值对应关系，集合C为信号源名称与船舶名称或编号对应关系，以上三个集合拥有自顶向下的两两相关的树状关系，求集合A直接到集合C 的映射方案。

待求数学模型输入为目标物信号特征数据，输出为个目标物船舶名称或船舶编号，船舶-信号源种类、信号源种类-信号源特征值为已知对应关系。现提出以下解决方案：

1. 误差修正。将传感器采集到的信号源特征数据按照题目一所给方法修正处理。
2. 特征张量降维。将多维特征参数用对应项简单相加法降低到低维，作为各信号源种类的新特征值，分别对目标物信号源特征数据和数据库信号源特征数据按此方式处理。
3. 信号源种类匹配。依次分别将附件五中的信号源特征值与数据库中信号源特征值对比，以上节中最小特征值差异作为衡量标准，对各信号源分类，得到一对多的信号源特征-信号源名称组合。
4. 船舶名称（或编号）匹配。依次分别将3中匹配到的信号源名称与数据库中船舶-信号源数据对比匹配，得到一对多的信号源名称-船舶名称组合。
5. 结果订正。由4可得到信号源特征-传播名称一对多的匹配组合，按照目标物编号和信号源批号进一步缩小结果空间，同一目标物有多批信号源批号的数据，说明该船舶一定同时具有所有信号源批号的信号源类别，对所有同目标编号的匹配结果取交集，可按此方法进一步剔除无效船舶名称。
6. 可信度评价。5可得到单一目标物的多种船舶分类结果，统计结果中同种船舶出现的次数，按出现次数为比例平均分配各结果正确概率，作为可信度评价。

3.3 实验验证

我们按照上节所述数学模型，借助Python语言编写了针对附件五数据的船舶分类程序，并对所有信号特征数据进行归类，得到了各目标物编号的传感器信息所来自的船舶编号和船舶名称。在本节将展示主要的程序运行结果。

表3.5 R类信号源降维后特征值

|  |  |
| --- | --- |
| **信号源批号** | **一维特征值** |
| 1011 | 7529.9655999999995 |
| 1012 | 20509.303899999999 |
| 1013 | 14698.828800000001 |
| 1041 | 9328.6274000000012 |
| 1042 | 23923.128199999999 |
| 1043 | 20489.708099999996 |
| 1071 | 12379.063 |
| 1072 | 21743.856199999998 |
| 1073 | 12688.042600000001 |
| 1101 | 7520.7380000000003 |
| 1102 | 37348.457829999999 |
| 1103 | 21193.968499999999 |
| 1131 | 21237.280299999999 |

表3.6 L1类信号源降维后特征值

|  |  |
| --- | --- |
| **信号源批号** | **一维特征值** |
| 1014 | 576.00099166999996 |
| 1044 | 670.09023330000002 |
| 1074 | 593.986314257 |
| 1104 | 692.48532496699988 |

表3.7 L2类信号源降维后特征值

|  |  |
| --- | --- |
| **信号源批号** | **一维特征值** |
| 1015 | 744.01959999999997 |

表3.8 A类信号源降维后特征值

|  |  |
| --- | --- |
| **信号源批号** | **一维特征值（MMSI编码）** |
| 1106 | 440999674 |

表3.9 R类信号源数据特征匹配到的船舶编号

|  |  |
| --- | --- |
| **信号源批号** | **匹配到的船舶编号** |
| 1011 | 431009, 440001, 440002, 366009, 366001, 366004, 366007, 431001 |
| 1012 | 431009, 431010, 366001, 366003, 431001, 431007 |
| 1013 | 431001, 431006 |
| 1041 | 431010, 440003, 440004, 366002, 366010, 366014, 431007 |
| 1042 | 440002, 366001, 431006, 366005, 366006, 366008, 431002, 431003, 366012, 431007, 366014, 366015 |
| 1043 | 431009, 431010, 366001, 366003, 431001, 431007 |
| 1071 | 431003, 431012, 431014 |
| 1072 | 431004, 431014 |
| 1073 | 431004, 431005, 431014 |
| 1101 | 431009, 440001, 440002, 366009, 366001, 366004, 366007, 431001 |
| 1102 | 440001, 366004, 366010, 366013, 366014 |
| 1103 | 无匹配 |
| 1131 | 440001, 440002, 412001, 440006, 431016, 440009, 431006 |

表3.10 L1类信号源数据特征匹配到的船舶编号

|  |  |
| --- | --- |
| **信号源批号** | **匹配到的船舶编号** |
| 1014 | 431001, 431010 |
| 1044 | 431012, 431006, 431007 |
| 1074 | 431014 |
| 1104 | 440001, 440002 |

表3.11 L2类信号源数据特征匹配到的船舶编号

|  |  |
| --- | --- |
| **信号源批号** | **匹配到的船舶编号** |
| 1015 | 431001 |

表3.12 A类信号源数据特征匹配到的船舶编号

|  |  |
| --- | --- |
| **信号源批号** | **匹配到的船舶编号** |
| 1106 | 440001 |

表3.13 最终分类结果

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **目标编号** | **信号源批号** | **船舶编号** | **船舶名称** | **可信度** |
| 101 | 1011, 1012, 1013, 1014, 1015 | 431001 | DBE-AY01 | 100% |
| 104 | 1041, 1042, 1043, 1044 | 431007 | DBH-SS01 | 100% |
| 107 | 1071, 1072, 1073, 1074 | 431014 | DBG-WD | 100% |
| 110 | 1101, 1102, 1103, 1104, 1106 | 440001 | DBE-KDX1-GKTDW | 100% |
| 113 | 1131 | 440001 | DBE-KDX1-GKTDW | 14% |
| 440002 | DBE-KDX1-YZWD | 14% |
| 412001 | DBC-HAIYANGDAO | 14% |
| 440006 | DBC-SE01 | 14% |
| 431016 | DBC-JD | 14% |
| 440009 | DBS-CY | 14% |
| 431006 | DBE-JG04 | 14% |

题目四

4.1 问题分析

通过分析附6数据并与附件5对比，我们发现附件6中的仿真场景具有每一种信号的各时刻地理位置信息，而题目二已经解决了通过信号源轨迹为各信号源分类的问题，题目三解决了通过信号源种类为目标分类问题，因此只需将附件6中的数据先按照轨迹信息聚类，即可沿用题目三的数学模型为整个仿真场景中的信号源分类，确定其中的目标物种类和目标物名称。

4.2 模型建立

解决此问题的思路是首先将附录6中数据通过轨迹为各信号源批号的数据分配目标物标签，确定其目标物个数，这样可以得到与附件5数据相同格式的信号源数据，再套用题目三中的模型即可得到结果。

本题解决步骤如下：

1. 数据修正。按题目一方法分别修正经纬度、信号特征数值数据，减小后面的分类误差；
2. 目标分类。处理经纬度数据，得到各信号源批号数据轨迹特征，将各信号源批号的数据按题目二方式聚类，分成若干目标物，将个目标物按序编号，作为信号源批号对应地标签。
3. 数据重构。将附件6中添加目标物编号标签的数据重新整理，得到形如附件5数据格式的信号数据。
4. 信号源分类。将3中重构数据输入题目三建立的模型，得到最终各目标识别结果，包括船舶编号和船舶名称。
5. 可信度评价。原理同题目三。

4.3 实验验证

本节以附件6数据为输入，以目标种类、目标编号、目标识别结果为输出，按照上节数学模型设计程序验证海上目标识别效果。在此给出关键的几个程序运行结果。

表4.1 信号源批号聚类结果

|  |  |
| --- | --- |
| **目标物编号** | **信号源批号** |
| 1 | 1225, 1326, 1391, 327 |
| 2 | 1031, 1113, 1169, 365 |
| 3 | 1373, 1405, 1416, 319 |
| 4 | 314 |
| 5 | 124, 212, 225 |
| 6 | 317 |
| 7 | 323 |
| 8 | 330 |
| 9 | 184 |
| 10 | 438 |
| 11 | 112, 146, 158 |
| 12 | 129 |
| 13 | 109, 139, 178 |
| 14 | 1425 |
| 15 | 292 |
| 16 | 278, 304 |
| 17 | 1476 |
| 18 | 282, 1297 |
| 19 | 345 |
| 20 | 360, 736 |
| 21 | 239, 256, 289 |
| 22 | 251, 274 |
| 23 | 1501 |
| 24 | 1503 |
| 25 | 1584 |

上表给出了按题目二方法的轨迹聚类方式将附件6仿真场景中的数据分为25类目标以及每种目标物对应地信号源批号。

表4.2 针对各类信号源进行目标识别的结果

|  |  |
| --- | --- |
| **信号源批号** | **识别结果：船舶编号** |
| 109 | 366008, 366009, 366002, 366003 |
| 112 | 431009, 440001, 440002, 366009, 366001, 366004, 366007, 431001 |
| 124 | 366009, 366010, 366011, 366012, 366013 |
| 129 | 431009, 431010, 366001, 366003, 431001, 431007 |
| 139 | 440002, 366001, 431006, 366005, 366006, 366008, 431002, 431003, 366012, 431007, 366014, 366015 |
| 146 | 431009, 440001, 440002, 366009, 366001, 366004, 366007, 431001 |
| 158 | 无结果 |
| 178 | 440003, 431013, 366005, 366007, 431004, 366013, 366015 |
| 184 | 366009, 366010, 366011, 366012, 366013 |
| 212 | 440001, 366004, 366010, 366013, 366014 |
| 225 | 366011, 366006 |
| 239 | 431009, 440001, 440002, 366009, 366001, 366004, 366007, 431001 |
| 251 | 431001, 431006 |
| 256 | 无结果 |
| 274 | 431010, 440003, 440004, 366002, 366010, 366014, 431007 |
| 278 | 440002, 366001, 431006, 366005, 366006, 366008, 431002, 431003, 366012, 431007, 366014, 366015 |
| 282 | 431009, 431010, 366001, 366003, 431001, 431007 |
| 289 | 431003, 431012, 431014 |
| 292 | 431004, 431014 |
| 1031 | 440002, 366001, 431006, 366005, 366006, 366008, 431002, 431003, 366012, 431007, 366014, 366015 |
| 1113 | 440008, 431018, 431003, 440005 |
| 1169 | 431003, 431012, 431014 |
| 1225 | 431003, 431012, 431014 |
| 1297 | 431004, 431005, 431014 |
| 1326 | 431008, 431002, 431011, 431012 |
| 1373 | 431010, 440003, 440004, 366002, 366010, 366014, 431007 |
| 1391 | 431010, 431002, 431012 |
| 1405 | 431009, 431010, 366001, 366003, 431001, 431007 |
| 1416 | 无结果 |
| 1425 | 440008, 431018, 431003, 440005 |
| 1476 | 412002, 431013, 440007, 431017, 440010, 366005 |
| 304 | 431014 |
| 314 | 366009, 366002, 366004 |
| 317 | 366001, 366007 |
| 319 | 431001, 431010 |
| 323 | 366008, 366003, 366005, 366006 |
| 327 | 431012, 431006, 431007 |
| 330 | 366011, 366013, 366014 |
| 345 | 431001, 431010 |
| 360 | 431012, 431006, 431007 |
| 365 | 431009, 431003, 431005, 431011 |
| 438 | 366009, 366012 |
| 736 | 431001 |
| 1501 | 366009 |
| 1503 | 431014 |
| 1584 | 440005 |

表4.3 最终识别结果和可信度

|  |  |  |  |
| --- | --- | --- | --- |
| 目标物编号 | 船舶编号 | 船舶名称 | 可信度 |
| 1 | 431012 | DBH-AWW01 | 100% |
| 2 | 431003 | DBE-JG01 | 100% |
| 3 | 431001 | DBE-AY01 | 50% |
| 431010 | DBH-XZ02 | 50% |
| 4 | 366009 | DBM-HSD01 | 33.3% |
| 366002 | DBE-KEWS02 | 33.3% |
| 366004 | DBE-NMZ02 | 33.3% |
| 5 | 无法识别 | 无法识别 | 100% |
| 6 | 366001 | DBE-HSD02 | 50% |
| 366007 | DBH-LSF03 | 50% |
| 7 | 366008 | DBH-NMZ03 | 25% |
| 366003 | DBE-LSF02 | 25% |
| 366005 | DBE-BS02 | 25% |
| 366006 | DBH-HSD03 | 25% |
| 8 | 366011 | DBM-LSF01 | 33.3% |
| 366013 | DBX-TKDLJ01 | 33.3% |
| 366014 | DBX-TKDLJ02 | 33.3% |
| 9 | 366009 | DBM-HSD01 | 20% |
| 366010 | DBM-KEWS01 | 20% |
| 366011 | DBM-LSF01 | 20% |
| 366012 | DBM-NMZ01 | 20% |
| 366013 | DBX-TKDLJ01 | 20% |
| 10 | 366009 | DBM-HSD01 | 50% |
| 366012 | DBM-NMZ01 | 50% |
| 11 | 431009 | DBH-XZ01 | 12.5% |
| 440002 | DBE-KDX1-YZWD | 12.5% |
| 440001 | DBE-KDX1-GKTDW | 12.5% |
| 366001 | DBE-HSD02 | 12.5% |
| 366004 | DBE-NMZ02 | 12.5% |
| 366007 | DBH-LSF03 | 12.5% |
| 366009 | DBM-HSD01 | 12.5% |
| 431001 | DBE-AY01 | 12.5% |
| 12 | 431009 | DBH-XZ01 | 16.7% |
| 431010 | DBH-XZ02 | 16.7% |
| 366001 | DBE-HSD02 | 16.7% |
| 366003 | DBE-LSF02 | 16.7% |
| 431001 | DBE-AY01 | 16.7% |
| 431007 | DBH-SS01 | 16.7% |
| 13 | 366005 | DBE-BS02 | 50% |
| 366015 | DBX-TKDLJ03 | 50% |
| 14 | 440008 | DBS-DZR | 25% |
| 431018 | DBS-SG | 25% |
| 431003 | DBE-JG01 | 25% |
| 440005 | DBC-JZD | 25% |
| 15 | 431004 | DBE-JG02 | 50% |
| 431014 | DBG-WD | 50% |
| 16 | 无法识别 | 无法识别 | 100% |
| 17 | 412002 | DBC-HULUDAO | 16.7% |
| 431013 | DBY-YH | 16.7% |
| 440007 | DBC-SE02 | 16.7% |
| 431017 | DBS-DB | 16.7% |
| 440010 | DBS-YZLD | 16.7% |
| 366005 | DBE-BS02 | 16.7% |
| 18 | 无法识别 | 无法识别 | 100% |
| 19 | 431001 | DBE-AY01 | 50% |
| 431010 | DBH-XZ02 | 50% |
| 20 | 无法识别 | 无法识别 | 100% |
| 21 | 无法识别 | 无法识别 | 100% |
| 22 | 无法识别 | 无法识别 | 100% |
| 23 | 366009 | DBM-HSD01 | 100% |
| 24 | 431014 | DBG-WD | 100% |
| 25 | 440005 | DBC-JZD | 100% |