附件给出了在线监测装置 OTLM1 与 OTLM2 设备采集的数据，记录了2019年7月至2020年7月一年内等时段间隔的电流负载状态。通过对电流的监控，进而可以获知输电效率与时间和耗电之间的关系。随着各行业耗电量的逐步增加，输电线路必须不断完善。为了满足一天之内所有行业耗电需求，特别是高峰时段用电需求，现需结合实际情况与所给信息建立数学模型，分析以下问题：

问题一：结合相关文献，依据 OTLM1 设备的采样数据，整理相关数据，建立数学模型预测 2020 年 8 月的动态载流量需求，并分析不同时间间隔载流量的变化情况，分析电流的周期性特征。

问题二：综合考虑 OTLM1 和 OTLM2 采集的数据，分析温度、弧垂角度对载流量的定量影响。

问题三：附件中 OTLM 设备的输出数据记录了具有 L 型切面四脚钢结构的架空输电线路温度以及振动数据。在 L 型切面脚钢结构中，应力反映了导线内张力的变化，而张力与导线温度有关，张力与 OTLM1 与 OTLM2 设备监测到的弧垂角度呈线性关系。结合图5和图6的相关信息，建立数学模型来讨论导线电流变化与钢架腿部每侧应力变化之间的关系。

问题四：已知 2020 年 8 月 15 日用电量有大幅度增加，整合分析附件中的相关数据，给出相应的增容方案，并给出最大增容量。

利用预测的数据，利用BP神经网络算法，预测2020年8月的动态载流量需求。

利用SPSS软件，绘制OTLM1与OTLM2采集的各组数据之间的矩阵散点图。通过矩阵散点图可初步判断，各组变量之间有一定的线性相关性，其中Line Temperature [°C]与Ambient Temperature OTLM [°C]呈现出较强的线性相关性。为更精确地描述各组数据之间的关系，先对各组数据进行正态分布检验，以判断其是否满足皮尔逊相关系数假设检验的条件。

由表3和表4可知：

1. OTLM采集的Line Temperature [°C]与Ambient Temperature OTLM [°C]之间相关系数较高，呈正相关。OTLM1与OTLM2采集的Line Temperature [°C]与Ambient Temperature OTLM [°C]的Spearman相关系数分别为0.9847与0.9838，表明它们之间存在较强的正相关关系。
2. Humidity [%] 与Line Temperature [°C]，Ambient Temperature OTLM [°C]的Spearman相关系数均小于0，表明它们之间存在负相关关系。但OTLM1与OTLM2采集的Humidity [%]与Line Current [A]的相关系数分别为0.0074与0.0588，相关性近似于0。
3. OTLM1采集的Angle [°]与Line Current [A]，Ambient Temperature OTLM [°C]的Spearman相关系数分别为-0.7575与-0.7714，表明它们之间存在一定的负相关关系。相反，OTLM2采集的Angle [°]与Line Current [A]，Ambient Temperature OTLM [°C]的Spearman相关系数分别为0.8982与0.8906，它们之间存在较强的正相关关系。

附件中数据给出了钢架四条腿部每侧应力的相关数据。首先使用因子分析，通过研究变量间的相关系数矩阵，将钢架每侧的应力变量反映成一个综合变量，这一过程又可称为降维。然后建立典型相关分析，从Line Temperature [°C]、Ambient Temperature OTLM [°C]、Humidity [%]、Angle [°] 、Altitude [m]和Vbat [V]中选取若干典型变量，其目的是分析电流与应力之间存在的关系。最后建立多元线性回归方程，更精确地定量描述应力与弧垂角度之间的关系。

利用因子分析进行应力数据的处理分析

因子分析是主成分分析的推广，它也是一种把多个变量化为少数几个综合变量的多元分析方法，其目的是用有限个不可观测的隐变量来解释原始变量之间的相关关系。

（清风，因子分析的原理一页）

因子分析模型的假设：（清风，假设一页）

KMO检验和巴特利特球形检验

**KMO检验**

KMO检验是 Kaiser, Meyer和 Olkin提出的，该检验是对原始变量之间的简单相关系数和偏相关系 数的相对大小进行检验，主要应用于多元统计的因子分析。KMO统计量是取值在0和1之间，当所有变量间的简单相关系数平方和远远大于偏相关系数平方和 时，KMO值越接近于1，意味着变量间的相关性越强，原有变量越适合作因子分析;当所有变量 间的简单相关系数平方和接近0时，KMO值越接近于0,意味着变量间的相关性越弱，原有变量越不适合作因子分析。

**巴特利特球形检验**

巴特利特球形检验是一种检验各个变量之间相关性程度的检验方法。巴特利特球形检验是以变量的相关系数矩阵为出发点的。它的原假设是相关系数矩阵是一个单位阵(不适合做因子分析，指标之间的相关性很差，不适合降维)，即相关系数矩阵对角线上的所有元素都是1，所有非对角线上的元素都为0。巴特利特球形检验的统计量是根据相关系数矩阵的行列式得到的。如果该值较大，且其对应的p值小于用户心中的显著性水平(一般为0.05)，那么应该拒绝原假设，认为相关系数不可能是单位阵，即原始变量之间存在相关性，适合于作因子分析。相反不适合作因子分析。

通过对附件中的数据建立预处理、特征工程、特征选择，变量从原来的8个变成了2个。而特征维度太高对主观法与客观法建立评价体系来说都很复杂，因此我们利用因子分析降维，将8个变量降成了2维。这里展示重要公共因子：

（一张表，因子分析的各个变量，详见华为杯论文）

处理一下下面的表

|  |  |  |
| --- | --- | --- |
| **KMO 和巴特利特检验（1侧）** | | |
| KMO 取样适切性量数。 | | .537 |
| 巴特利特球形度检验 | 近似卡方 | 106914.762 |
| 自由度 | 6 |
| 显著性 | <.001 |
|  |  |  |

|  |  |  |
| --- | --- | --- |
| **KMO 和巴特利特检验（2侧）** | | |
| KMO 取样适切性量数。 | | .549 |
| 巴特利特球形度检验 | 近似卡方 | 178172.216 |
| 自由度 | 6 |
| 显著性 | <.001 |

原假设：数据不适合进行因子分析

备择假设：数据适合进行因子分析

对因子进行KMO 和巴特利特检验，由KMO 取样适切性量数和巴特利特检验的显著性水平可知，在99%的置信水平下，我们拒绝原假设，即认为原数据适合因子分析，通过了KMO 和巴特利特检验。

|  |  |  |
| --- | --- | --- |
| **公因子方差1侧** | | |
|  | 初始 | 提取 |
| S1 Ch1 Avg [MPa] | 1.000 | .590 |
| S2 Ch1 Avg [MPa] | 1.000 | .924 |
| S3 Ch1 Avg [MPa] | 1.000 | .799 |
| S4 Ch1 Avg [MPa] | 1.000 | .081 |
| 提取方法：主成分分析法。 | | |

|  |  |  |
| --- | --- | --- |
| **公因子方差2侧** | | |
|  | 初始 | 提取 |
| S1 Ch2 Avg [MPa] | 1.000 | .669 |
| S2 Ch2 Avg [MPa] | 1.000 | .674 |
| S3 Ch2 Avg [MPa] | 1.000 | .619 |
| S4 Ch2 Avg [MPa] | 1.000 | .797 |
| 提取方法：主成分分析法。 | | |

（写共因子数据结论）多少占多少

通过共因子方差可以看出，提取的公因子对各个变量的贡献度均较大，这表明了这两个公因子可以较好地反应各个变量的信息。

|  |  |
| --- | --- |
| **1侧成分矩阵a** | |
|  | 成分 |
| 1 |
| S1 Ch1 Avg [MPa] | .768 |
| S2 Ch1 Avg [MPa] | .961 |
| S3 Ch1 Avg [MPa] | -.894 |
| S4 Ch1 Avg [MPa] | -.285 |
| 提取方法：主成分分析法。 | |
| a. 提取了 1 个成分。 | |

|  |  |
| --- | --- |
| **2侧成分矩阵a** | |
|  | 成分 |
| 1 |
| S1 Ch2 Avg [MPa] | .818 |
| S2 Ch2 Avg [MPa] | -.821 |
| S3 Ch2 Avg [MPa] | .787 |
| S4 Ch2 Avg [MPa] | .893 |
| 提取方法：主成分分析法。 | |
| a. 提取了 1 个成分。 | |

降维后将得到因子载荷矩阵，因子也具有解释性：比较电流、应力与其余各变量的Spearman相关系数，因子1可看作是电流显著性因子，因子2可看作是应力显著性因子。

**利用典型相关分析分析电流与应力变化的关系**

为了研究电流与应力变化之间的关系，令电流为输入变量，应力变化为输出变量，采用典型相关分析法。

典型相关分析由Hotelling提出，其基本思想是利用主成分思想，分别找出输入变量与输出变量的线性组合，然后讨论线性组合之间的相关关系。。

首先在每组变量中找出变量的线性组合，使得两组的 线性组合之间具有最大的相关系数;

然后选取和最初挑选的这对线性组合不相关的线性组合，使其配对，并选取相关系数最大的一对;

如此继续下去，直到两组变量之间的相关性被提取完毕为止。

被选出的线性组合配对称为典型变量，它们的相关系 数称为典型相关系数。典型相关系数度量了这两组变量之间联系的强度。

假设电流与应力数据服从联合正态分布。