报名序号：4999

赛题题目：人工智能对大学生学习影响的评价

**基于主成分特征提取及遗传算法优化投影寻踪法的**

**人工智能对大学生学习影响程度的评价研究**

**摘 要**

人工智能的兴起是时代不可阻挡的潮流，为研究分析人工智能对大学生学习的影响程度，本文将从现有的调查问卷对其进行数据挖掘和特征提取，根据问卷中设置的问题对其影响程度、影响方面进行量化分析。通过本次对调查问卷的研究，可以为人工智能在校园的发展策略提供科学的智力支持，并为校园科技创新赋予新动能。

针对问题一，为对数据进行数值化处理和数据分析，本文去除了问题5、6共108个异常样本和问题7、8共106个异常样本，并将问题划分为学生情况问题、单项选择问题、多项选择问题三种类别，分别使用整数编码、数理统计的方法进行数值化处理。在量化处理后，本文通过对问题分类进行讨论，采用解释说明、图标可视化、数据挖掘的方法分析人工智能对大学生学习的影响程度。

针对问题二，为对数据的分析结果选取评价指标，本文采用主成分分析法将30个问题概括提取为6个主成分特征，依次为：可信度成分、目的倾向型成分、综合影响程度成分、专业导向型成分、用途受限型成分、兴趣驱动型成分。以这六个成分作为指标构建评价指标体系，并通过优先级、科学性、可操作性以及综合性原则这四个方面阐述这六个指标的合理性。

针对问题三，为评价人工智能对大学生学习的影响，本文采用基于实数编码的加速遗传算法优化的投影寻踪法对各个样本进行打分，在最后的评价得分表中提取出了前二十九名以及后三十名调查对象，使其结合问题二构建的指标评价体系进行对比分析，最终得出了人工智能对大学生的学习产生了积极影响的结论。

针对问题四，为书写一份人工智能对大学生学习影响的分析报告，本文结合对人工智能的了解、认知和判断，从学生对人工智能的依赖性、学生对人工智能的期待程度、学生是否借助人工智能进行学术造假、学生是否能从人工智能中提取自身素质这四个方面进行分析，以此为根据针对现阶段人工智能影响的利弊进行探究并给出解决建议，以及对未来人工智能的发展作出合理预判。

**关键词：数据挖掘；特征提取；主成分分析；投影寻踪；实数编码加速遗传算法**

# 问题重述

## 1.1问题背景

党的十九届五中全会明确指出，我国经济已转向高质量发展阶段。而人工智能产业为中国经济发展提供了全新的战略动能，是引领中国经济发展的重要战略抓手。

为抓住人工智能发展的机遇，让技术走进校园，教育部提出“智能化领跑教育信息化”，这一理念促进了人工智能走进大学生群体的进程，成为大学生日常学习、生活不可分割的部分。人工智能产业要实现高质量发展，培养人工智能人才是关键。因此研究分析人工智能在不同侧面对大学生学习的影响情况，是现阶段重视产业人才培养，构建“引才、留才、用才”新格局中至关重要的一个环节。

## 1.2问题要求

根据一个感兴趣的侧面，结合附件1和附件2所给出的数据，建立相应的数学模型，分析人工智能对大学生学习的影响，解决以下问题：

1. 对调查问卷的答题数据进行数据预处理，并以此为基础将字符数据转化成量化数据，最后通过转化而来的量化数据进行分析，从整体上和问题类别上探究人工智能对大学生学习的影响；
2. 从问题一的分析结果中选取指标构建评价指标体系，并从多个方面评价选取指标的合理性；
3. 建立一个数学模型，要求能够根据量化数据以及经过特征提取的评价指标，说明人工智能对大学生生活的影响程度；
4. 根据前三个问题的量化数据和结论，对现阶段人工智能发展影响作出分析，并对未来发展作出合理预判。

# 二、问题分析

## 2.1 问题一的分析

对于数据分析和数值化处理的问题，首先要对调查问卷的答案数据进行清洗去噪[1]，主要体现在异常值和缺失值两方面上。在消除了无效数据后，本文对数据划分了三个类别，并采用三种方法进行数值化处理。在将所有数据进行量化之后，本文把调查问卷的问题划分为六类，以问题为切入点，采用可视化、数理统计、解释说明的方法对人工智能对大学生学习的影响程度进行分析。其流程图如图1所示。

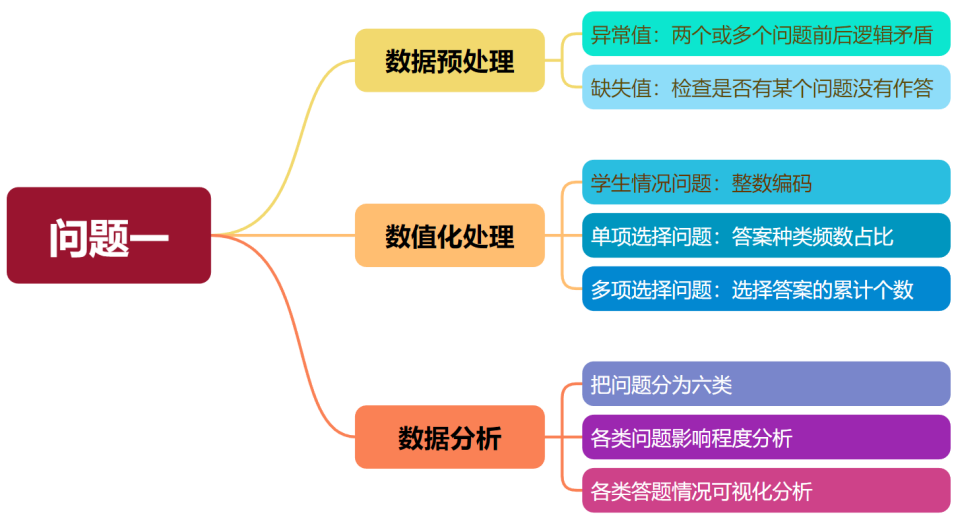


图1 问题一流程图

## 2.2 问题二的分析

对于构建评价指标体系，本文选取了前500个样本，八个有针对性的问题作为指标，使用主成分分析法对八个指标进行计算选出主成分并根据主成分构建指标评价体系，最后从多个不同的方面阐述该指标评价体系的合理性。其流程图如图2所示。

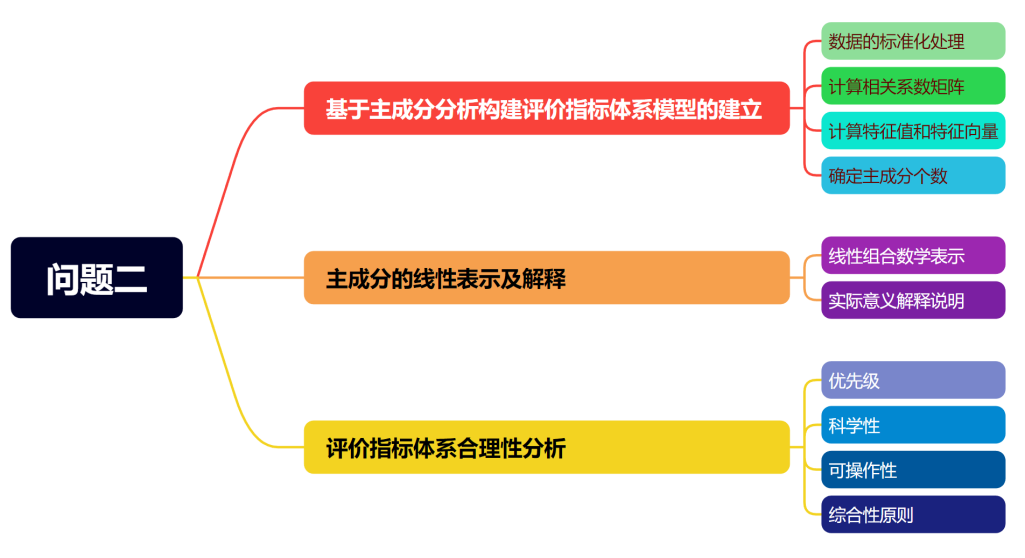


图2 问题二流程图

## 2.3 问题三的分析

对于建立数学模型，评价人工智能对大学生学习的影响的问题，根据每个样本相应的综合投影值对样本进行综合评价分析建立非线性优化模型[2]，为了优化模型求解的速率，本文采用实数编码的加速遗传算法来求解规划模型[3]，最终求得最优投影方向向量，并据此对人工智能对大学生学习的影响程度、影响方面进行打分评价。其流程图如图3所示。



图3 问题三流程图

## 2.4 问题四的分析

对于书写一份分析报告，研究人工智能对大学生学习的影响程度，本文运用了控制变量的方法，将调查问卷的问题划分为四类，即学生对人工智能的依赖性的分析、学生对人工智能的期待程度、学生是否借助人工智能进行学术造假的分析、学生是否能从人工智能中提升自身素质，据此分析现阶段人工智能对大学生学习影响的利弊，并对未来发展作出合理预判。其流程图如图4所示。

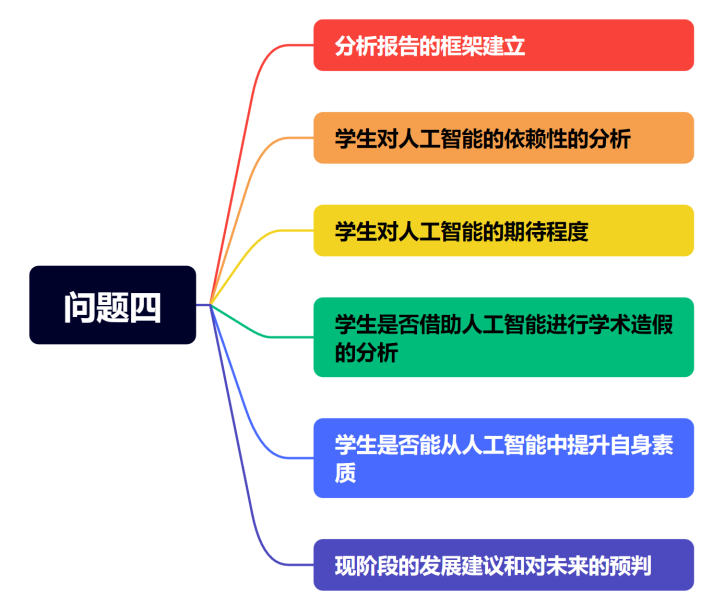


图4 问题四流程图

# 三、模型假设

1、假设剔除调查问卷中前后逻辑矛盾的答案后，不会对分析结果造成显著影响；

2、假设调查结果的可信度、有效度和区分度都处在可接受范围内；

3、假设在数据分析过程中使用控制变量法时，固定的指标不会对结果产生影响。

# 四、符号说明

|  |  |  |
| --- | --- | --- |
| 符号 | 说明 | 备注 |
|  | 第道单项选择问题第种答案的频数 |  |
|  | 第道单项选择问题第种答案的数值 |  |
|  | 第道多项选择题选择个答案的频数 |  |
|  | 第道多项选择题选择个答案的数值 |  |
|  | 投影寻踪模型中的一维投影值 | 无 |
|  | 投影寻踪模型中特征值的标准差 | 无 |
|  | 投影寻踪模型中投影特征值的局部密度 | 无 |
|  | 遗传算法中的初代投影方向 | 无 |
|  | 遗传算法中的初代编码序列 |  |
|  | 遗传算法中杂交操作后的子代个体序列 |  |
|  | 遗传算法中变异操作后的子代个体序列 | 无 |
|  | 遗传算法中的适应度 | 无 |

# 五、问题一模型的建立与求解

## 5.1 对调查结果的预处理及数值化处理

在对调查问卷答案[4]进行数值化处理之前需对其进行数据预处理，检查是否有异常值和缺失值。由于附件2并没有缺失值，因此本文只对异常值进行处理。对于字符数据异常值的定义，本文认为两个或多个问题答案出现前后逻辑矛盾的情况，即为异常值，在数值化处理的操作中这些问题的答案将不被考虑，剔除的要求分别如图5所示。

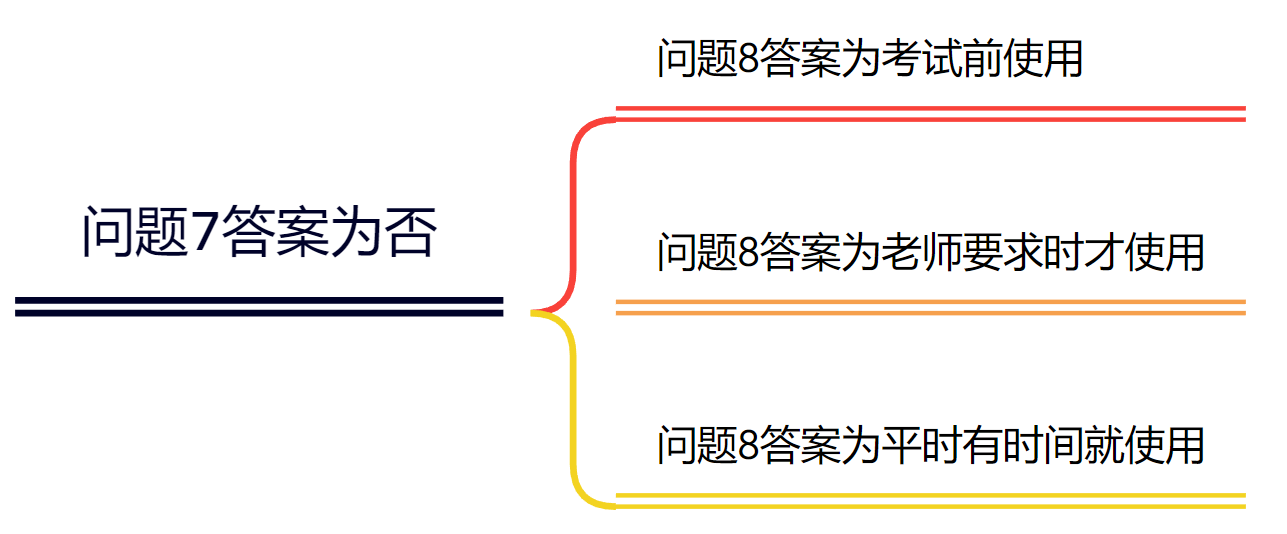
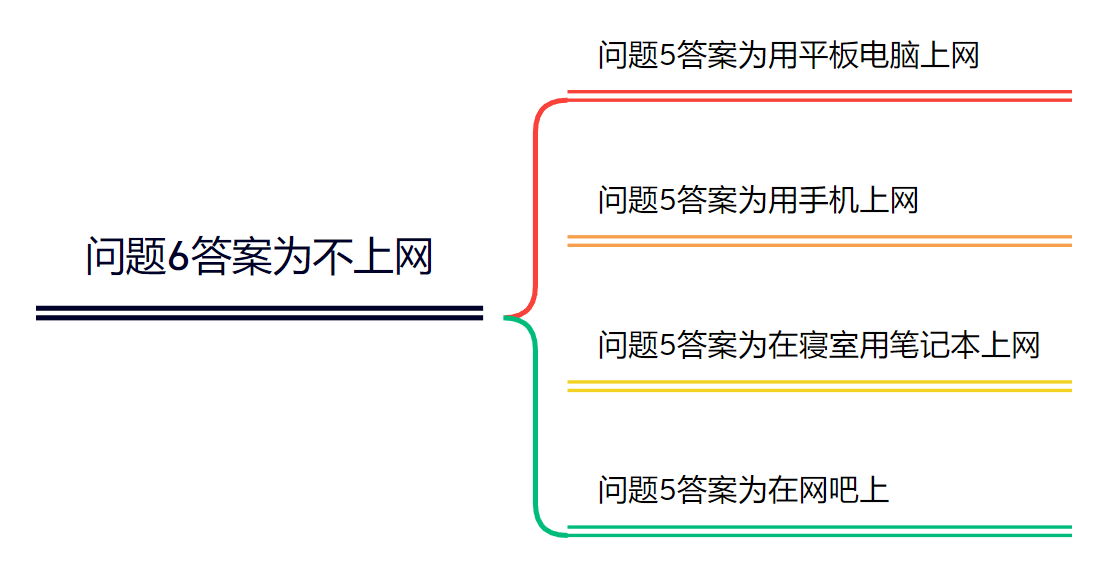


图5 问题5、6和问题7、8剔除异常值的要求

经筛选分析后发现问题5和问题6出现前后逻辑矛盾的情况，共剔除108组数据，占比2.35%；问题7和问题8也出现前后逻辑矛盾的情况，共剔除106组数据，占比2.30%。异常值处理效果如图6所示。

图6 问题5和问题8剔除异常值后效果图

对于调查问卷中设计特性，本文根据问题的性质和面向对象将其分为三大类：学生情况问题、单项选择问题以及多项选择问题，而针对不同类别的问题本文将采用不同的数值化处理方法。如表1所示。

表1 问题分类及相应的数值化处理方法

|  |  |  |
| --- | --- | --- |
| 问题类别 | 包含的问题 | 数值化处理方法 |
| 学生情况问题 | 性别、专业、年级、性格 | 按答案频数降序，将数值编码为1到的整数，其中为该问题的答案数量 |
| 单项选择问题 | 问题6到问题22 | 答案种类频数占比即为数值 |
| 多项选择问题 | 问题23到问题30 | 将每个人选择答案的个数累计起来，频数占比即为数值 |

对于学生情况问题，性别问题中男性编码为1，女性编码为0；专业、年级和性格问题按照答案频数的多少，由高到低依次编码为1到答案个数。其数值化处理结果如表2所示。

表2 学生情况问题数值化处理结果

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 序号 | 性别 | 专业 | 年级 | 性格 |
| 1号 | 0 | 4 | 2 | 3 |
| 2号 | 1 | 1 | 2 | 1 |
| 3号 | 0 | 4 | 2 | 5 |
| 4号 | 1 | 1 | 2 | 6 |
|  |  |  |  |  |
| 4604号 | 1 | 1 | 1 | 1 |
| 4605号 | 0 | 1 | 1 | 6 |

对于单项选择问题，首先将每个问题的答案个数和答题情况进行统计，某答案的数值即为某答案在其问题中的出现次数占比，计算公式如下：

|  |  |
| --- | --- |
|  | () |

如问题5中，答案“用手机上网”的出现次数为3371，占比73.2%，那么该答案可用数值0.732替代，其余同理。而对于已被剔除的答案则不再统计范围之内，并且用0替代，结果如表3所示。

表3 单项选择问题数值化处理结果

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 序号 | 问题5 | 问题6 | 问题7 | 问题8 |  | 问题22 |
| 1号 | 0.732 | 0.237 | 0.945 | 0.563 |  | 0.864 |
| 2号 | 0 | 0 | 0.031 | 0.256 |  | 0.864 |
| 3号 | 0.732 | 0.237 | 0.945 | 0.563 |  | 0.864 |
| 4号 | 0.0007 | 0.167 | 0.945 | 0.563 |  | 0.864 |
|  |  |  |  |  |  |  |
| 4604号 | 0.732 | 0.237 | 0.945 | 0.256 |  | 0.864 |
| 4605号 | 0.732 | 0.167 | 0.945 | 0.256 |  | 0.864 |

对于多项选择问题，本文通过选择答案的个数来判断大学生对该问题的认可程度，即将每个问题每个学生选择答案的个数统计出来，某答案的数值即为某答案在其问题中的出现次数占比，计算公式如下：

|  |  |
| --- | --- |
|  | () |

如问题23中，某位同学选择了7个答案，而选择7个答案的频率为18.1%，则无论选择哪7个答案，都用数值0.181替代，其结果如表4所示。

表4 多项选择问题数值化处理结果

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 序号 | 问题23 | 问题24 | 问题25 | 问题26 |  | 问题30 |
| 1号 | 0.149 | 0.425 | 0.327 | 0.258 |  | 0.384 |
| 2号 | 0.149 | 0.385 | 0.252 | 0.371 |  | 0.171 |
| 3号 | 0.181 | 0.385 | 0.252 | 0.371 |  | 0.189 |
| 4号 | 0.181 | 0.385 | 0.252 | 0.371 |  | 0.171 |
|  |  |  |  |  |  |  |
| 4604号 | 0.147 | 0.385 | 0.132 | 0.371 |  | 0.171 |
| 4605号 | 0.072 | 0.425 | 0.287 | 0.371 |  | 0.384 |

## 5.2 对调查结果的数据分析[5]

在对调查结果进行数值化处理之后，本文将从已经量化后的数据进行分析，对调查问卷设计的30个问题进行分类讨论，据此探究人工智能对大学生的影响情况。本文进行的数据分析主要有六个部分，如图7所示。

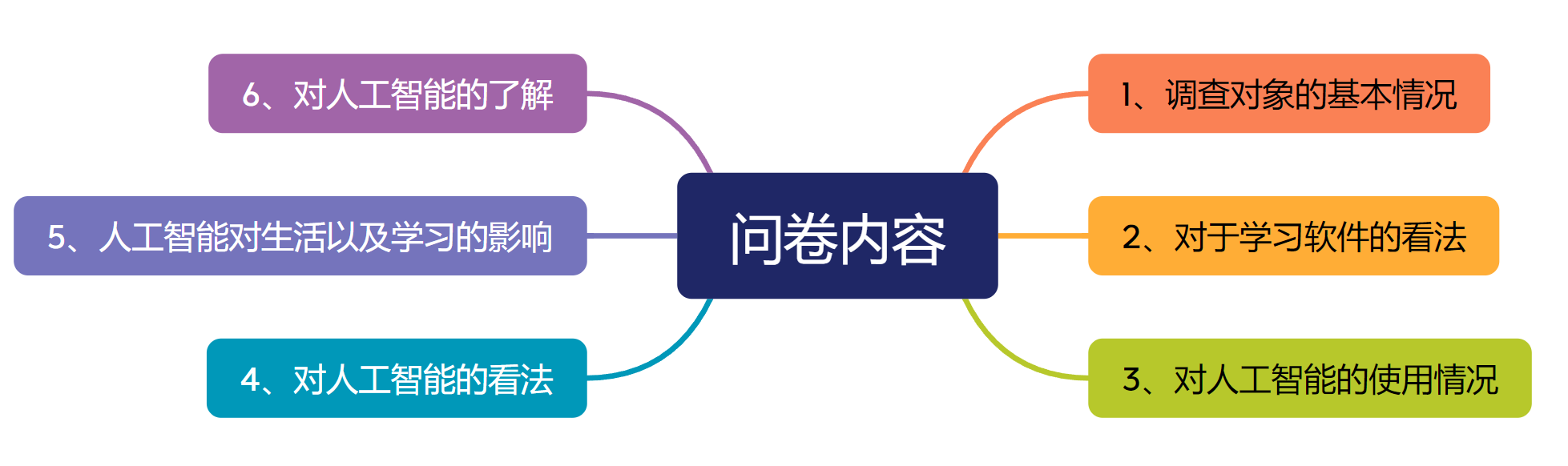
****

图7 数据分析主要内容

### 5.2.1 调查对象的基本情况

此分类包含问题1到问题6。首先从调查对象的基本情况来看，调查对象的性别、各类专业、大学年级以及性格类别的涵盖面较广也相对均衡，从而保证了结果的合理性。据问题5和问题6可知，大部分的学生使用互联网的时间较长，也从客观上反映了互联网对于大学生的日常生活的影响。可视化结果如图8所示。

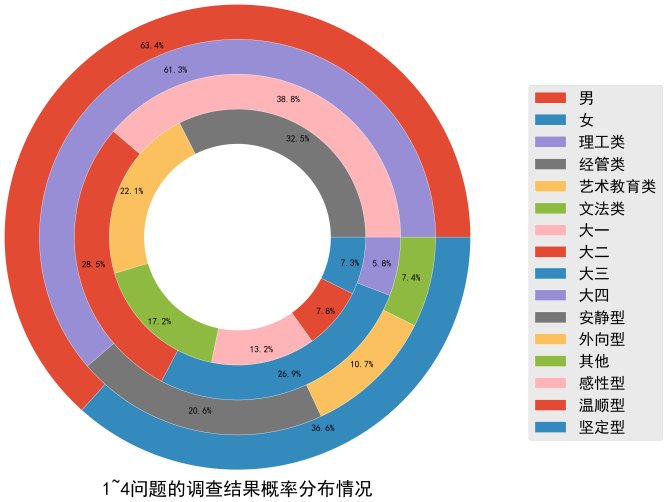
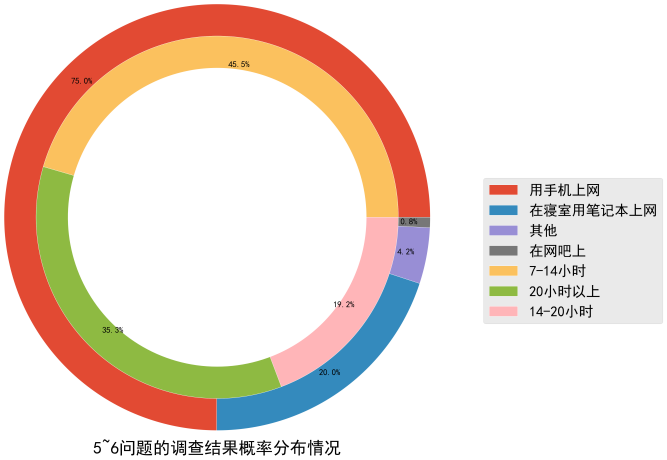
 

图8 调查对象的基本情况结果分析

### 5.2.2 对于学习软件的看法

此分类包含问题7到问题11。从图中可以看到调查对象是否使用学习软件的占比分别为97.2%、2.8%；任课老师是否推荐过学习软件占比分别为62.5%、37.5%。从以上数据我们可以看出大学生对于学习软件的使用十分普及，也能在客观上说明人工智能对于教育领域的使用范围以及效果比较客观。可视化结果如图9所示。

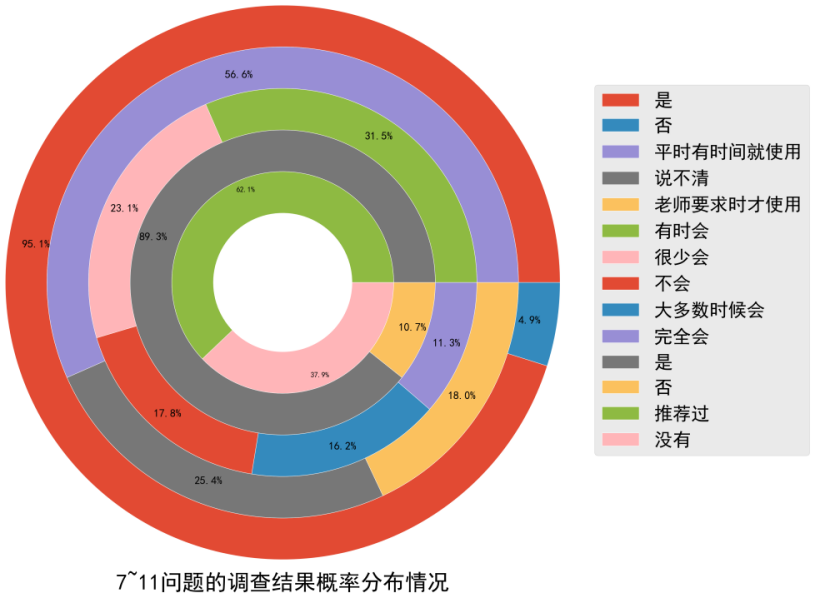


图9 对于学习软件的看法结果分析

### 5.2.3 对人工智能的使用情况

此分类包含问题12到问题18。从图中我们可以看出对于使用人工智能学习工具的占比为81.9%；其中想通过使用人工智能完成作业、完成小测验、完成论文的占比分别为42%、34%、33.2%；对于赞同大学生使用人工智能的看法的占比为59.2%。

通过以上数据大部分的大学生赞成并在人工智能在教育方面的使用较为普及、人工智能对于学生的帮助较大，但人工智能在教育领域发展并不完善，需要查询的问题与答案不一致的情况出现频繁，且绝大多数的运营商对于人工智能的部分的功能进行较高收费导致大部分的学生无法正常使用。可视化结果如图10所示。

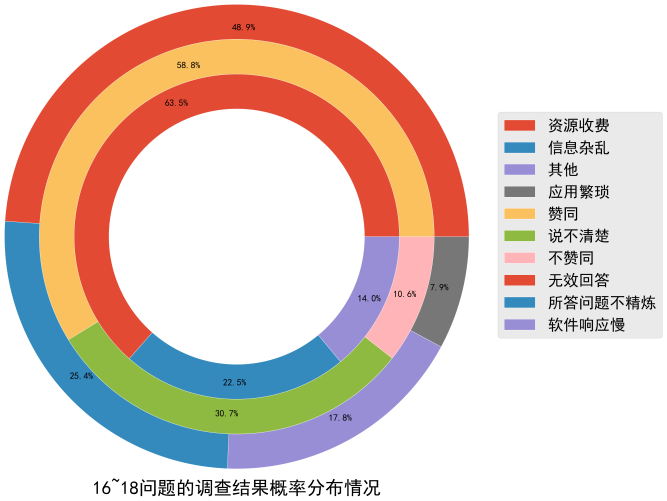
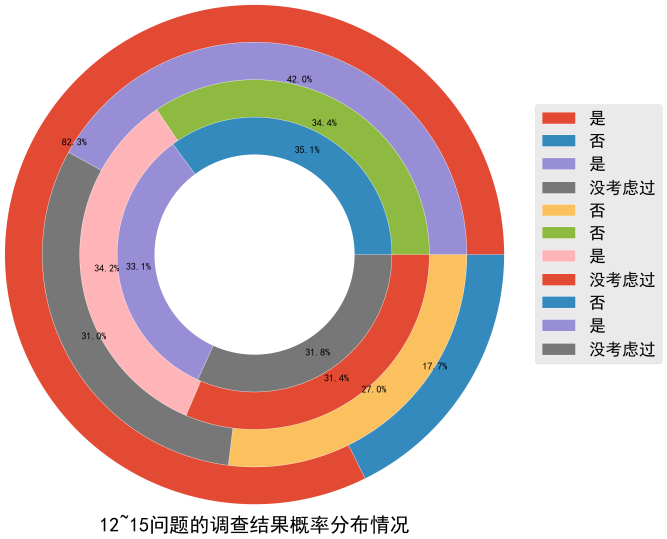


图10 对人工智能的使用情况结果分析

### 5.2.4 对人工智能的看法

此分类包含问题19到问题22。从图中我们可以看出对于人工智能学习工具的可信度中一般的占比为69.5%；在认为人工智能是否取代教师中不可能、不清楚的占比分别为48.8%、12.9%。

从图中数据可以看出人工智能在大学生看来可信度较为一般，大部分的大学生愿意接受新的方式和工具去学习并希望人工智能能够帮助提升自身的能力。从以上数据能够看出大学生愿意接受人工智能应用到教育领域但只起到辅助作用，但对其可行度持怀疑态度。可视化结果如图11所示。

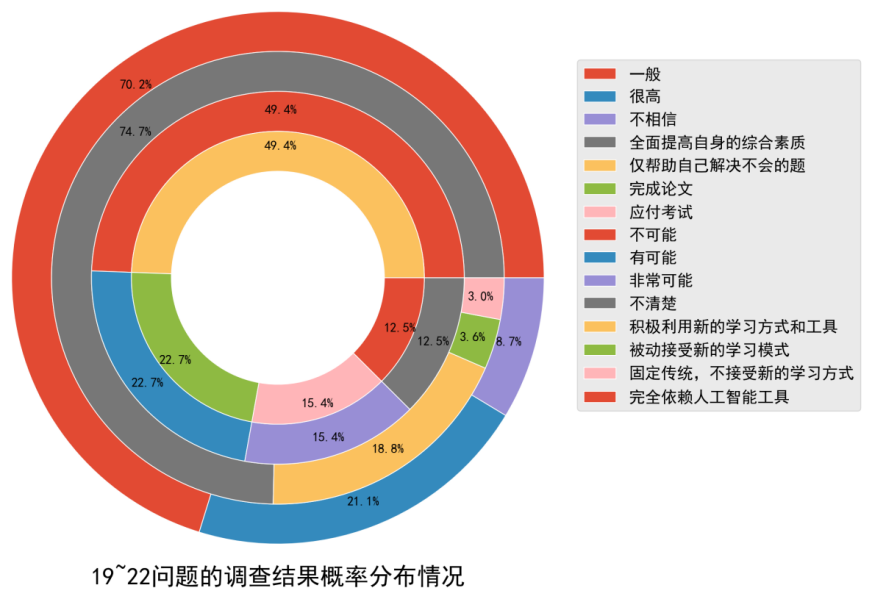


图11 对人工智能的看法结果分析

### 5.2.5 人工智能对学习的影响

此分类包含问题23到问题26。从上图我们可以看出大学生使用网络1-8项的活动的占比分别为14.9%、7.2%、11.1%、14.7%、13.5%、9.7%、18.1%、10.4%；认为学习软件比课堂有1-3处优势的占比分别为42.5%、19.0%、38.5%；认为进入大学之后在学习方面有1-4个问题的占比分别为32.7%、28.7%、13.2%、25.2%；对于学习软件的形式，关注1-4类问题的占比分别为25.8%、22.8%、14.1%、37.1%。

从图中数据我们可以看出网络对于大学生的帮助较大且相比于课堂，学习软件能够更好的帮助学生掌握和巩固所学知识；以及人工智能可以根据学生进入大学学习后产生的问题给学生提供帮助。可视化结果如图12所示。

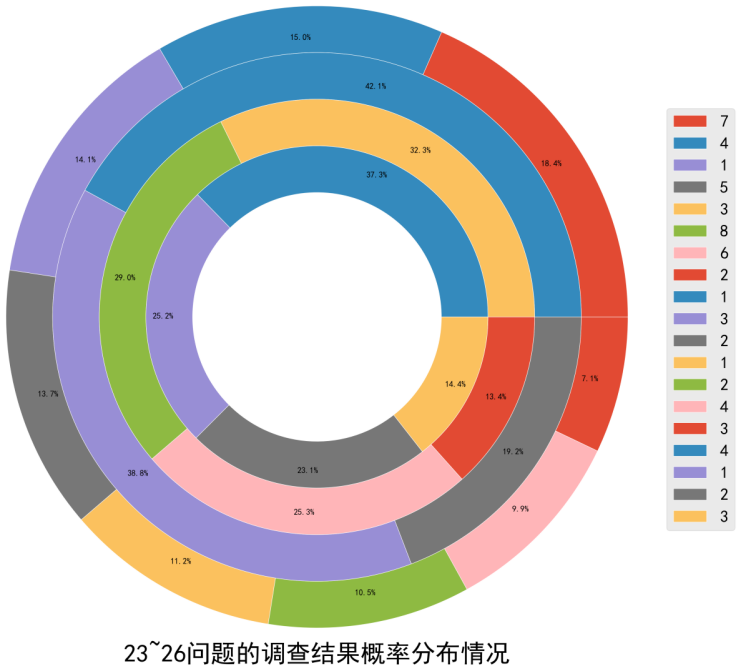


图12 人工智能对学习的影响结果分析

### 5.2.6 对人工智能的了解

此分类包含问题27到问题30。通过以上数据可以看出调查对象对于人工智能1-6类的安全性问题的占比分别为20.8%、14.6%、14.2%、8.5%、2.8%、38.8%；认为有1-3个方面对于人工智能学习工具很重要的占比分别为28.7%、19.2%、51.9%；对于人工智能需要具备的1-5种功能的占比分别为19.0%、9.7%、8.9%、10.8%、51.3%；对于人工智能应运用到1-4种学习环节的占比分别为38.4%、25.3%、18.9%、17.1%。

以上数据从客观上反应大学生群体对于人工智能许多方面的安全性问题都存在疑虑，且大部分的大学生对于人工智能的期望较高，希望人工智能具备的功能能够较为全面以及其能够融入课堂当中。可视化结果如图13所示。

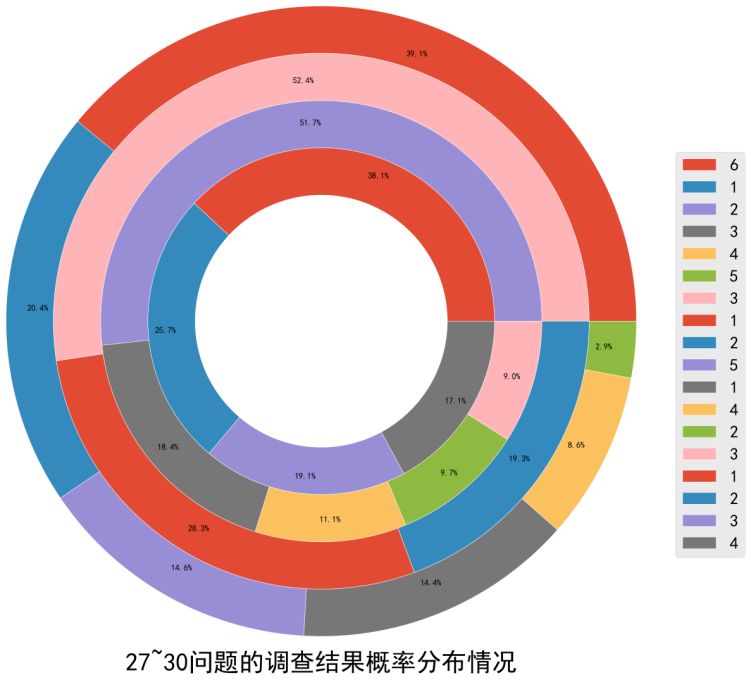


图13 对人工智能的了解结果分析

总体来看，人工智能对于大学生的学习生活的应用较为广泛且帮助较大，大学生对于人工智能的发展前景、所具备的功能、在教学领域的融入方面期望较高，但人工智能在教育方面的发展与应用仍不够完善，有资源更新、答案无效、运营商收费不合理等方面的问题发生。故人工智能在教育领域的应用技术亟待提高。

# 问题二模型的建立与求解

## 6.1 基于主成分分析[6]构建评价指标体系模型的建立

本文采用主成分分析法进行特征提取，提取出来的主成分本文即认为是评价指标。主成分分析算法是最常用的线性降维方法，它的目标是通过某种线性投影，将高维的数据映射到低维的空间中，并期望在所投影的维度上数据的信息量最大，以此使用较少的数据维度，同时保留住较多的原数据点的特性。本文将8个指标转换为用原始变量的线性组合表示的少数主成分，而这少数几个主成分即为与飞行相关的关键数据项，并通过每个指标的特征值进行重要性程度分析。

**Step 1.数据的标准化处理**

第一步需对预处理后的数据去除量纲的影响，因此先对数据进行标准化处理。按列计算的均值和标准差如下所示：

|  |  |
| --- | --- |
|  | (3) |
|  | (4) |

由于标准化的数据为，故原始样本矩阵经过标准化变为如下所示：

|  |  |
| --- | --- |
|  | (5) |

**Step 2.计算相关系数矩阵**

主成分分析在提取特征的过程中需要考虑变量之间的联系，这种联系可以用协方差来进行衡量，通过进一步的数理统计，本文将协方差整理为一个能够更高效体现变量之间相关性程度的矩阵，即为相关系数矩阵R，其计算公式如下：

|  |  |
| --- | --- |
|  | (6) |

使用MATLAB求解的相关系数矩阵R如表5所示。

表5 八个主成分之间的相关系数矩阵

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |
|  | 1 | 0.0637 | 0.0112 | 0.0485 | -0.0432 | 0.0089 | 0.0158 | 0.0370 |
|  | 0.0637 | 1 | 0.0846 | 0.1464 | 0.1034 | 0.0278 | -0.0959 | 0.0245 |
|  | 0.0112 | 0.0846 | 1 | 0.3486 | 0.2242 | 0.0875 | 0.0113 | 0.0742 |
|  | 0.0485 | 0.1464 | 0.3486 | 1 | 0.2009 | 0.0116 | -0.0799 | 0.0555 |
|  | -0.0432 | 0.1034 | 0.2242 | 0.2009 | 1 | 0.3124 | -0.0281 | 0.0880 |
|  | 0.0089 | 0.0278 | 0.0875 | 0.0116 | 0.3124 | 1 | -0.0304 | 0.0568 |
|  | 0.0158 | -0.0959 | 0.0113 | -0.0799 | -0.0281 | -0.0304 | 1 | 0.1004 |
|  | 0.0370 | 0.0245 | 0.0742 | 0.0555 | 0.0880 | 0.0568 | 0.1004 | 1 |

**Step 3.计算特征值和特征向量**

计算相关系数矩阵R的特征值及对应的标准化特征和标准化特征向量,其中,由特征向量组成8个新的指标变量如下所示：

|  |  |
| --- | --- |
|  | (7) |

其中共提取八个主成分，本文将分别计算每个主成分的贡献率，通过累计贡献率来选择最终的主成分个数，这样可以在保证原样本特征被多方位提取的同时不会造成线性组合过于复杂，主成分解释过于冗余。其中，为第1主成分；为第2主成分；；为第8主成分。

**Step 4.确定主成分个数**

首先计算特征值的贡献率和累积贡献率，分别如下所示:

|  |  |
| --- | --- |
|  | (8) |
|  | (9) |

其中，计算求得的八个主成分及其特征值、贡献率和累计贡献率如表6所示。

表6 主成分分析结果

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 指标 | 主成分1 | 主成分2 | 主成分3 | 主成分4 | 主成分5 | 主成分6 | 主成分7 | 主成分8 |
| X1 | 0 | -0.186 | 0.414 | 0.708 | 0.503 | -0.113 | -0.081 | -0.134 |
| X2 | 0.278 | -0.402 | 0 | 0.370 | -0.544 | 0.557 | 0.134 | 0.000 |
| X3 | 0.506 | 0 | 0.226 | -0.354 | 0.226 | 0 | 0.597 | -0.391 |
| X4 | 0.493 | -0.333 | 0.218 | -0.267 | 0.111 | -0.116 | -0.340 | 0.621 |
| X5 | 0.518 | 0.288 | -0.262 | 0 | 0 | 0 | -0.595 | -0.464 |
| X6 | 0.333 | 0.463 | -0.417 | 0.346 | 0.182 | 0 | 0.375 | 0.455 |
| X7 | 0 | 0.494 | 0.528 | -0.139 | 0.120 | 0.638 | -0.084 | 0.146 |
| X8 | 0.184 | 0.381 | 0.457 | 0.160 | -0.582 | -0.496 | 0.040 | -0.001 |
|  | 1.699 | 1.157 | 1.110 | 1.017 | 0.908 | 0.845 | 0.654 | 0.610 |
|  | 0.212 | 0.145 | 0.139 | 0.127 | 0.114 | 0.106 | 0.082 | 0.076 |
|  | 0.212 | 0.357 | 0.496 | 0.623 | 0.736 | 0.842 | 0.924 | 1 |

## 6.2 主成分的线性表示及解释

从表6可知，前六个指标的累计贡献率已经达到84%以上，本文认为可以较好地概括原样本特征，故选取前六个指标作为主成分。本文将对这六个主成分给出相应的线性表示及其解释。其中均为标准化后的指标，本文将这六个主成分概括为可信度成分、目的倾向成分、综合影响度成分、专业导向型成分、用途受限型成分和兴趣驱动型成分，如图14所示。

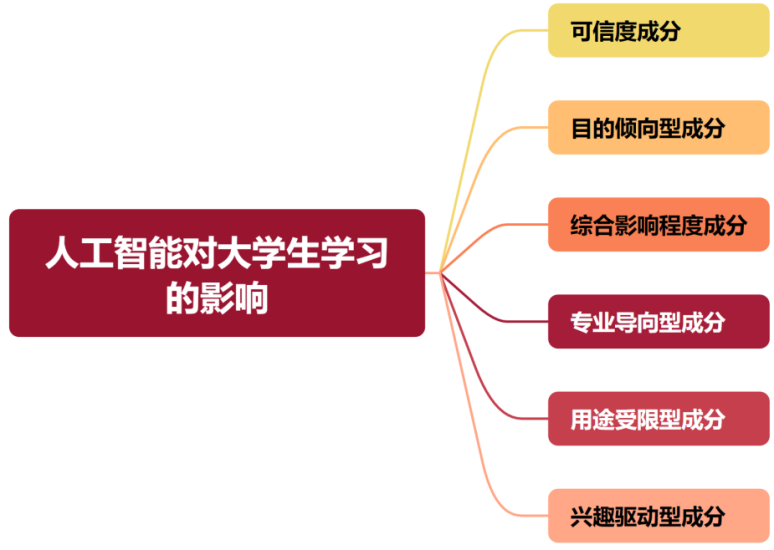


图14 六个主成分的解释

均为标准化后的指标，在线性组合表达式中已经剔除影响较小的指标，其中各个的含义如表7所示。

表7 的含义

|  |  |
| --- | --- |
| 符号 | 含义 |
|  | 专业 |
|  | 每周上网时间 |
|  | 对人工智能的选择 |
|  | 对使用人工智能的看法 |
|  | 对人工智能的适应方法 |
|  | 使用人工智能的作用 |
|  | 人工智能的优势 |
|  | 人工智能对学习的不同方面产生的结果 |

1. 从表中各个指标的系数可以看出，的正载荷占比较大，且所有指标均为正载荷，故本文将第一个主成分设定为可信度成分，其线性组合表达式为：

|  |  |
| --- | --- |
|  | (10) |

1. 从表中各个指标的系数可以看出，均为正载荷且占比较大，说明该成分偏向于对人工智能产生的效果，故本文将第二个主成分设定为目的倾向型成分，其线性组合表达式为：

|  |  |
| --- | --- |
|  | (11) |

1. 从表中各个指标的系数可以看出，均为正载荷且占比大，但均为较小的正载荷，可见该成分的各个方面较为均衡，反映的是综合性的影响程度，故本文将第三个主成分设定为综合影响程度成分，其线性组合表达式为：

|  |  |
| --- | --- |
|  | (12) |

1. 从表中各个指标的系数可以看出，均为正载荷且占比大，但均为负载荷且占比相对较大，因此该成分反映的是个人信息与功效之间的关系，故本文将第四个主成分设定为专业导向型成分，其线性组合表达式为：

|  |  |
| --- | --- |
|  | (13) |

1. 从表中各个指标的系数可以看出，为正载荷且占比大、均为正载荷但占比小，均为负载荷且占比大，因此该成分反映的是个人情况和接受程度对最后的效果产生的影响的关系，故本文将第五个主成分设定为用途受限型成分，其线性组合表达式为：

|  |  |
| --- | --- |
|  | (14) |

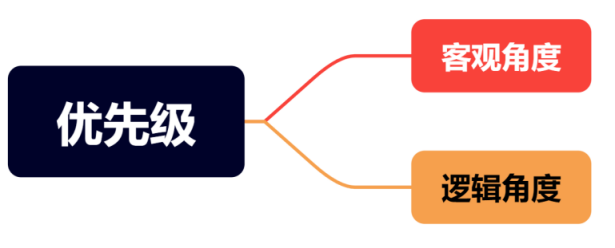
1. 从表中各个指标的系数可以看出，均为正载荷且占比大，为负载荷，因此该成分反映的是个人兴趣与结果的关系，故本文将第六个主成分为兴趣驱动型成分，其线性组合表达式为：

|  |  |
| --- | --- |
|  | (15) |

## 6.3 评价指标体系合理性分析

本文将分别从优先级、科学性、可操作性以及综合性原则这四个方面阐述六个主成分所构建的评价指标体系的合理性。

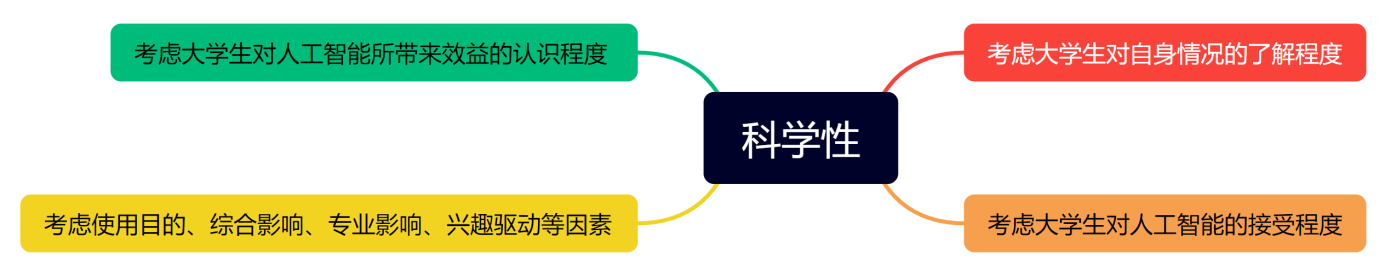
1. 优先级



从客观角度来看，由表6可知第一个主成分的贡献率相较于其他几个主成分来看占比最高，其他的几个主成分以此类推，贡献率按照从高往低的顺序排列；

从逻辑角度来看，人工智能的可信程度对于大学生是否接受使用人工智能关联性强，只有大学生开始使用人工智能的学习工具后，人工智能才能对其学习产生影响，所以第一个指标对于整体的影响较大。据此，大学生使用人工智能学习工具的目的、影响程度、其自身的专业导向、使用人工智能的用途以及兴趣驱动都是根据与人工智能对大学生学习的影响的关联性的强弱进行排序。

1. 科学性



六个主成分指标主要通过对大学生自身情况的了解、对人工智能的接受程度、以及人工智能所带来的效果三个方面中选取出来，通过评价大学生对于人工智能的可信度、使用的目的、综合影响、专业影响、使用后产生的结果、兴趣驱动六个方面，客观全面地反映出对人工智能对大学生学习产生显著影响。

1. 可操作性



在对于指标的选取中，六个主成分覆盖的范围的确定都是围绕人工智能对于大学生学习的影响来选取的，且主成分的计算方式和选取标准都是通过主成分分析法的算法步骤以及最后得出的结果，按照贡献率的高低以及累计贡献率是否超过80%进行选取。由此可见各指标涵盖的内容在整体上统筹兼顾，在局部上相互联系，具有较好的针对性。

1. 综合性原则



六个主成分指标的涵盖面广，针对大学生接受人工智能的程度、自身专业对于使用人工智能的程度、对人工智能产生的效果的重视程度、对人工智能的感兴趣程度等方面较为全面的考虑了自身、环境等多种因素，综合性较强。

# 七、问题三模型的建立与求解

## 7.1 基于投影寻踪法探究人工智能对大学生学习的影响模型的建立

本文拟建立一套能够反映人工智能对不同类别大学生影响程度的数学模型，并在此基础上确定哪些类型的大学生受人工智能影响程度的大小，最后根据建立的模型对结果进行解释说明。

## 7.2 投影寻踪评价法模型的建立

投影寻踪是处理和分析高维数据的一类统计方法，其基本思想是将高维数据投影到低维子空间上，寻找出能反映原高维数据的结构或特征的投影，达到研究分析高维数据的目的。由于投影寻踪具有稳健性、抗干扰性和准确度高等优点，因此本文认为适合此题探究人工智能对大学生学习的影响。

投影寻踪法分析评价的样本集需满足，并且数据需要进行归一化，本文采用的归一化公式对最小值和最大值作了改进，即系数分别变为0.02和0.98，目的是防止评价矩阵中含有数值0从而导致程序报错，公式如下：

|  |  |
| --- | --- |
|  | (6) |

其中表示第个样本的第个指标值，其核心思想是将n维数据综合成以为投影方向的一维投影值。其中，一维投影值如下所示：

|  |  |
| --- | --- |
|  | (17) |

本文认为可以从两个方面进行最优投影值的定义：局部投影点尽可能密集、投影点团在整体上尽可能散开。因此可以建立如下目标函数：

|  |  |
| --- | --- |
|  | (18) |

其中，为投影特征值的标准差，为投影特征值的局部密度，计算公式如下：

|  |  |
| --- | --- |
|  | (19) |
|  | (20) |

其中，为投影特征值之间的距离，为阶跃信号，为估计局部散点密度的窗宽参数，为遵循宽度内至少包含一个散点的原则，有如下关系：

|  |  |
| --- | --- |
|  | (21) |
|  | (22) |
|  | (23) |

综上所述，可得到如下非线性优化模型：

|  |  |
| --- | --- |
|  | (24) |
|  | (25) |

## 7.4 基于实数编码的加速遗传算法

遗传算法起源于对生物系统所进行的计算机模拟研究。它是模仿自然界生物进化机制发展起来的随机全局搜索和优化方法，借鉴了达尔文的进化论和孟德尔的遗传学说。其本质是一种高效、并行、全局搜索的方法，能在搜索过程中自动获取和积累有关搜索空间的知识，并自适应地控制搜索过程以求得最佳解。

由于传统的遗传算法采用需要二进制编码，但由于其频繁的编码和解码, 计算量大且只能产生有限的离散点阵, 还可能产生额外的最优点。因此本文采用基于实数编码的加速遗传算法进行求解，其建模过程如图15所示

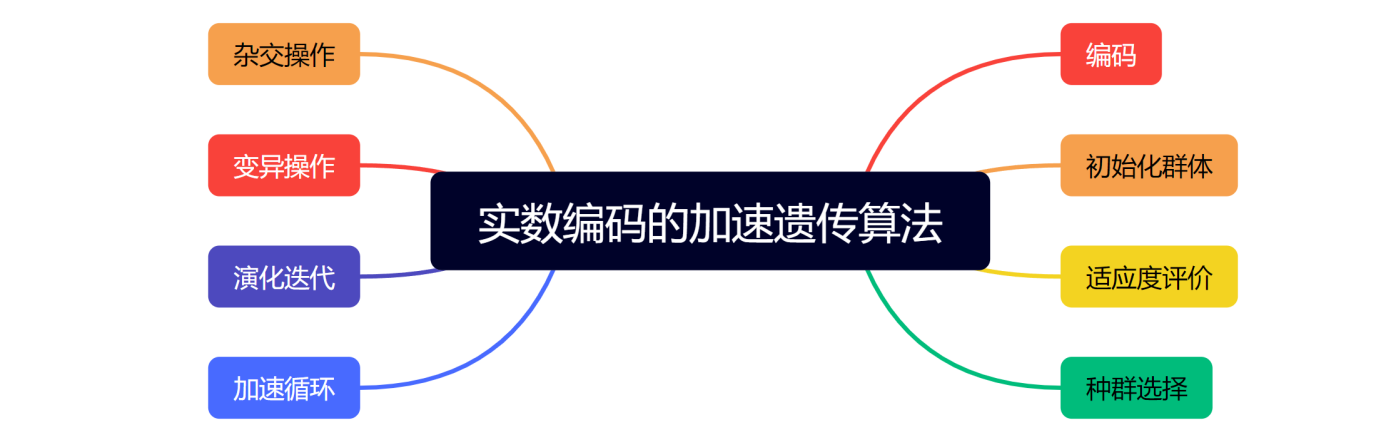


图15 实数编码的加速遗传算法

**Step 1.**编码

标准遗传算法采用二进制的编码方式，不但增加了算法执行过程中的计算量，且增加了产生额外最优解的可能性。基于实数编码的加速遗传算法采用的实数编码利用了如下的线性变换：

|  |  |
| --- | --- |
|  | (26) |

其中为决策变量，即投影方向向量的定义域，本文将该变换视为将投影方向编码为的染色体形式，且对任意的j有。

**Step 2.**初始化群体

设初始化群体个数为M，分别在定义域内的M组初代投影方向将其代入Step 1中进行实数编码后得到初代编码序列，将该序列依次代入目标函数中求取群体中

各个个体的目标函数值，依据目标函数值的大小对个体进行升序排序，选择前50个个体作为优秀个体。

其中初代投影方向和初代编码序列分别如公式27、28所示：

|  |  |
| --- | --- |
|  | (27) |
|  | (28) |

**Step 3.**适应度评价

由于目标函数值越小，表示该个体的适应度越高，再者考虑目标函数值为 0 的情况，故定义适应度函数为：

|  |  |
| --- | --- |
|  | (29) |

**Step 4.**种群选择

进行选择操作产生第 1 个子代群体，第i个个体被选择的概率为，定义，则p将区间划分为M个子区间。接着随机生成个随机数，则个体选择方式为：

|  |  |
| --- | --- |
|  | (30) |

对于剩余的个体，采用上一代中的优秀个体直接补充，即得第1个子代群体。

**Step 5.**杂交操作

进行杂交操作产生第 2 个子代群体根据Step4中得到的选择概率随机选择一对个体与进行杂交，因为编码方式的限制，故本文采用如下随机线性组合的方式进行杂交获得子代个体：

|  |  |
| --- | --- |
|  | (31) |

其中是在的随机数。

**Step 6.**变异操作

进行变异操作产生第 3 个子代群体采用n

个随机数以的概率来代替第个个体进行变异：

|  |  |
| --- | --- |
|  | (32) |

其中是在的随机数。

**Step 7.**演化迭代

将Step 4到Step 6中得到的个体依照适应度大小升序排序后，选取前M个个体构成新的种群，代入Step 3重新进行繁殖演化。

**Step 8.**加速循环

使用第一次与第二次演化迭代所产生的优秀个体对应的变量变化区间作为新的初始变化区间。算法重新进入步骤一。如此循环，直至算法达到预设的加速次数N时，结束迭代。本文设置基于实数编码的加速遗传算法迭代100次，取目标函数的最大值作为投影寻踪法最终的最优值，其对应的解为最优解，即最佳投影方向。

## 7.5 基于评价结果的结果分析

根据投影寻踪评价的评价得分，本文将对其进行分析并给出影响程度结果。由于部分问题的答题情况无显著差异，故此本文只对答题情况差异较大的问题进行分析。为使结论具有对比性和指导性，本文选取得分排名前29名的调查对象和得分排名后30名的调查对象进行对比分析。前29名和后30名的得分和排名情况如表8、表9所示。

表8 得分排名前29名的调查对象

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 排名 | 序号 | 得分 | 排名 | 序号 | 得分 |
| 1 | 4070 | 17.9750 | 16 | 857 | 17.5557 |
| 2 | 303 | 17.9484 | 17 | 1747 | 17.5339 |
| 3 | 3767 | 17.7663 | 18 | 3 | 17.5111 |
| 4 | 1362 | 17.7381 | 19 | 3648 | 17.4825 |
| 5 | 3279 | 17.6888 | 20 | 1102 | 17.4561 |
| 6 | 4029 | 17.6837 | 21 | 618 | 17.4461 |
| 7 | 235 | 17.6736 | 22 | 3111 | 17.4319 |
| 8 | 1571 | 17.6661 | 23 | 2126 | 17.4167 |
| 9 | 3673 | 17.6577 | 24 | 776 | 17.4131 |
| 10 | 2926 | 17.6509 | 25 | 1839 | 17.4008 |
| 11 | 3016 | 17.6297 | 26 | 3256 | 17.3909 |
| 12 | 2442 | 17.5756 | 27 | 788 | 17.3841 |
| 13 | 4298 | 17.5719 | 28 | 799 | 17.3841 |
| 14 | 470 | 17.5694 | 29 | 1821 | 17.3758 |
| 15 | 462 | 17.5595 |  |  |  |

表9 得分排名后30名的调查对象

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 排名 | 序号 | 得分 | 排名 | 序号 | 得分 |
| 4576 | 1763 | 7.3054 | 4591 | 137 | 7.2068 |
| 4577 | 1863 | 7.3054 | 4592 | 735 | 7.2068 |
| 4578 | 1943 | 7.3054 | 4593 | 1592 | 7.2068 |
| 4579 | 2077 | 7.3054 | 4594 | 2089 | 7.2068 |
| 4580 | 2146 | 7.3054 | 4595 | 2759 | 7.2068 |
| 4581 | 2434 | 7.3054 | 4596 | 2813 | 7.2068 |
| 4582 | 2467 | 7.3054 | 4597 | 2770 | 7.1661 |
| 4583 | 3018 | 7.3054 | 4598 | 1107 | 7.1188 |
| 4584 | 3301 | 7.3054 | 4599 | 4135 | 6.9895 |
| 4585 | 3505 | 7.3054 | 4600 | 3807 | 6.9330 |
| 4586 | 4010 | 7.3054 | 4601 | 194 | 6.9043 |
| 4587 | 4202 | 7.3054 | 4602 | 2682 | 6.8583 |
| 4588 | 524 | 7.2933 | 4603 | 2073 | 6.8012 |
| 4589 | 2037 | 7.2933 | 4604 | 2816 | 6.7152 |
| 4590 | 3460 | 7.2132 | 4605 | 3316 | 6.0942 |

本文把上述得分作为评价依据，将问题一的数据分析和问题二的评价指标体系结合起来，选取了部分有代表性的问题使得评价指标体系具体化，其中具体化情况如表10所示。

表10 评价指标体系具体化情况

|  |  |
| --- | --- |
| 指标 | 问题序号 |
| 可信度指标 | 7、9、12、17、19 |
| 目的倾向型指标 | 11、13、14、15、22 |
| 综合影响程度指标 | 24、26、29、30 |
| 专业导向型指标 | 10 |
| 用途受限型指标 | 23、28 |
| 兴趣驱动型指标 | 6、8、20 |

根据相关数据和评价指标体系的综合分析，本文认为人工智能对于前二十九名调查对象的影响较大，故对前二十九名的调查对象的选项内容进行分析，可以明显看出，人工智能对于大学生的学习产生了良好的影响。

由表7和表8可知，受人工智能影响程度得分的前二十九名调查对象与后三十名的评价得分相差了10-11分。其中对于前二十九名调查对象选项情况进行分析，发现其可信度指标的数值较大，对于使用人工智能的看法以及对其的接受程度较好；目的倾向型指标的数值较大，对于人工智能的使用目的较为明确以及适应人工智能与教育的结合的适应程度较好；综合影响程度的数值较大，在使用人工智能学习工具后得到的结果较好；兴趣驱动型指标的数值较大，人工智能的使用时间较长、学习软件的使用频率较高以及对人工智能的期望较大；用途受限型指标的数值较大，对于人工智能的使用情况以及在对人工智能学习工具的看重程度都较为良好。从上述几个指标可以清晰地看出，前二十九名在指标评价体系中各个指标的数值均较大，从而得到人工智能对其影响较大的评价。

后三十名调查对象对使用人工智能的看法以及对其接受程度相对较差，相比于前二十九名调查对象的可信度指标、目的倾向型指标、综合影响程度指标、用途受限型指标以及兴趣驱动型指标的数值均相对较低，故得出人工智能对于其影响较小的评价。

大学生对人工智能学习工具的滥用以及使用方法的不当对大学生的学习产生了一定的消极意义，但大学生使用人工智能学习工具对课堂的知识进行巩固和延展以及对课外知识的了解和学习，对其学习产生了更多的积极的意义，故从整体来看，人工智能对大学生学习产生的仍是积极的影响。

# 问题四模型的建立与求解

## 8.1 分析报告[7]的框架建立

本文在对调查问卷的答案进行数据分析后可知，问卷的内容主要可概括为六个特征，而人工智能对学生学习的影响同样也可以概括为六个特征，如表11所示。

表11 问卷特征和影响特征

|  |  |  |  |
| --- | --- | --- | --- |
| 序号 | 问卷特征 | 序号 | 影响特征 |
| 1 | 调查对象的基本情况 | 1 | 可信度成分 |
| 2 | 对于学习软件的看法 | 2 | 目的倾向型成分 |
| 3 | 对人工智能的使用情况 | 3 | 综合影响程度成分 |
| 4 | 对人工智能的看法 | 4 | 专业导向型成分 |
| 5 | 人工智能对生活以及学习的影响 | 5 | 用途受限型成分 |
| 6 | 对人工智能的了解 | 6 | 兴趣驱动型成分 |

由这些特征可以说明人工智能对大学生学习的影响主要与个人的专业和年级、老师的推荐与帮助、人工智能的生产效益与适用范围和是否对自身发展有益这四方面相关性较强。本文将根据已有的量化数据和实际生活的客观情况，作关于对人工智能了解、认知和判断，以及对未来发展的展望的技术性分析报告。

## 8.2 学生对人工智能的依赖性的分析

本文将根据问题11和问题21的量化数据进行分析，探究学生对人工智能的依赖性是否会因为老师对学习软件工具的推荐的改变。当控制变量老师没有给学生推荐学习软件工具时，学生同意未来人工智能会取代老师的人数占比为33.1%；而当控制变量老师有给学生推荐学习软件工具时，学生同意未来人工智能会取代老师的人数占比为41.3%，相比较增加了8.2%。因此本文认为老师的推荐与否以及学校对人工智能的重视与否会改变学生的人工智能的依赖程度，且二者呈正相关。可视化如图16所示。

图16 学生对人工智能的依赖性的分析

## 8.3 学生是否借助人工智能进行学术造假的分析

本文将根据问题13、问题14和问题15的量化数据进行分析，探究学生是否借助人工智能进行学术造假。首先将问题13、问题14和问题15综合成一个分析学生有无意向进行学术造假的问题，改问题认为的学术造假包括完成学校布置的作业、学校组织的测验以及需独立完成的论文。其中有意向利用人工智能进行学术造假的个数占比达到了37%，如图17所示。这说明人工智能对大学生学习的影响不仅仅是在资源丰富性和获取便捷性上，更有可能对学生学习发展造成不良影响。

图17 学生是否借助人工智能进行学术造假的分析

## 8.4 学生对人工智能的期待程度

本文将根据问题16、问题17和问题18的量化数据进行分析，探究学生对人工智能的期待程度。当控制变量大学生赞同使用人工智能学习工具时，对问题16进行统计分析可知，分别有55%和28%的学生认为资源收费和信息杂乱是影响他们使用人工智能的原因。

当控制变量大学生赞同使用人工智能学习工具时，对问题18进行统计分析可知，有57%的学生认为无效回答是他们使用人工智能时最不希望发生的事情，而选择软件响应慢的学生只有20%，这说明人工智能在学生群体的普及面较为广泛，但是在实际的生产效益上，如提供的回答与信息的质量、以及资源获取的渠道与方式，都存在推广的障碍，这也说明了学生有强大的意愿和需求使用人工智能，却受限于目前的科技水平和经济能力，如图18所示。

图18 学生对人工智能的期待程度

## 8.5 学生是否能从人工智能中提升自身素质

本文将根据问题25、问题26和问题29的量化数据进行分析，探究学生是否能从人工智能中提升自身素质的问题。对于问题25，本文认为选择答案的个数越多，便能反映学生对使用人工智能的困惑程度就越高，反之则越低；对于问题26和问题29，本文将两个问题出现答案的频数相加，其中数量定类的重视程度标准如表12所示，若总数越多，则认为学生对使用人工智能的重视程度就越高，反之则越低，如图23所示。

表12 数量定类的重视程度标准

|  |  |
| --- | --- |
| 总数数量 | 重视程度 |
| 2、3 | 较低 |
| 4、5 | 一般 |
| 6、7 | 较高 |
| 8、9 | 很高 |

图23 学生对使用人工智能的态度

由图23可知，学生对使用人工智能的重视程度和困惑程度呈正比，因此可以说明现阶段学生对人工智能的运用处于一个掌握和普及不匹配的阶段，具体体现在普及率高、覆盖率广，但是学生的使用能力仍有较大的提升空间。

## 8.6 现阶段的发展建议和对未来的预判[8]

自2012年数据的爆发式增长使人工智能发展得到突破以来，现阶段正在以多模态、新技术、广覆盖的趋势高速发展，但是由于人工智能背后需要的理论支持、资金支持和政策支持，目前仍存在信息冗杂、质量参差以及资源收费等问题，本文将提出如下三点建议：

1. 促进更多人力物力的投入，优化资源配置，提高产出信息的质量；
2. 学校开通专业板块为学生提供专门的渠道获取资源，鼓励学生合理正确使用人工智能，将其运用在自身能力提升之上；
3. 从事相关行业的人士承担起领头人的作用，带领学生群体利用人工智能为生活实践提供实质性帮助，同时加大力度培养新人，为我国高新技术突破赋予新动能，促进我国人工智能方向高质量发展。

本文结合生活实际认为未来人工智能在应用水平上，会以第五代移动通讯技术的突破为起点，建立起物体与物体之间的联系，催生出如自动驾驶、VR等更多日用产品，同时也为多项世界尖端科技的迭代扫清障碍。

# 模型的评价

## 9.1 模型的优点

1、在构建评价指标体系的时候，本文采用的主成分分析能够在保留住较多的原数据点的特性的同时提取出关键指标，通过对提取出来的主成分进行解释说明，能够较好地提升评价指标的科学性和可操作性；

2、在建立人工智能对大学生学习的影响的评价模型时，本文采用的是投影寻踪评价法，此方法具有稳健性、抗干扰性和准确度高等优点，由于传统的遗传算法求解投影寻踪具有繁琐且生成解有限的弊端，本文采用的实数编码优化的加速遗传可以使寻优过程更接近全局最优解。

## 9.2 模型的缺点

1、使用实数编码加速的遗传算法对投影寻踪进行求解时，寻优时间较长，不利于大规模的数据及模型推广；

2、主成分分析法得出的主成分模糊了原有指标表达的含义，减少了样本中的信息量。

## 9.3 模型的推广

本文对人工智能对于大学生学习的影响做了较为完整的分析，综合考虑了人工智能对大学生学习产生影响的多种因素，包括可信度、兴趣导向等因素，因而对人工智能在教育领域的发展方向有一定的借鉴意义。在求解问题三中，所用到的基于实数编码加速遗传算法优化的投影寻踪算法可以运用到各种问卷调查中对于调查对象进行评分。

# 参考文献

1. 申莎莎. 基于小波变换与傅里叶变换对比分析及其在信号去噪中的应用[J]. 山西师范大学学报(自然科学版),2018,32(03):27-32.
2. 赵小勇. 投影寻踪模型及其在水土资源中的应用[D].东北农业大学,2006.
3. 杨淼. 改进的遗传算法在多元函数优化中的应用[J]. 电脑编程技巧与维护,2022,No.443(05):24-26.
4. 廖列法,杨姝,王刊良. 信息系统问卷调查质量评估研究[J]. 科学学与科学技术管理,2008,No.322(07):63-66+74.
5. 胡海锋. 基于大数据分析的温室管控系统优化研究[J]. 农机化研究,2023,45(07):214-218.
6. 相广俐,李林. 基于主成分分析的LSTM神经网络聚乙烯价格预测研究[J]. 计算机时代,2023,No.369(03):67-70+75.
7. 余丽华. 大数据时代人工智能在计算机网络技术中的应用[J]. 中国高新科技,2023,No.133(01):43-45.

[8] 于非,何玉林,贺颖. 人工智能与数字经济专题序言——人工智能赋能数字经济，共创未来无限可能[J]. 深圳大学学报(理工版),2023,40(03):253-257.

# 附录

主成分分析代码：

%% 数据导入处理

clc

clear

A = xlsread("2.xlsx");

%% 数据标准化处理

a = size(A,1);

b = size(A,2);

for i = 1:b

SA(:,i) = (A(:,i) - mean(A(:,i)))/std(A(:,i));

end

%% 计算相关系数矩阵的特征值和特征向量

CM = corrcoef(SA); %计算相关系数矩阵

[V,D] = eig(CM); %计算特征值和特征向量

for j = 1:b

DS(j,1)=D(b+1-j,b+1-j); %对特征值按降序排列

end

for i = 1:b

DS(i,2) = DS(i,1)/sum(DS(:,1)); %贡献率

DS(i,3) = sum(DS(1:i,1))/sum(DS(:,1)); %累计贡献率

end

%% 选择主成分及对应的特征向量

T = 0.9; %主成分保留率

for K = 1:b

if DS(K,3) >= T

Com\_num = K;

break

end

end

%% 提取主成分对应的特征向量

for j = 1:Com\_num

PV(:,j)=V(:,b+1-j);

end

%% 计算个评价对象的主成分的分

new\_score = SA\*PV;

for i = 1:a

total\_score(i,1)= sum(new\_score(i,:));

total\_score(i,2)= i;

end

result\_report = [new\_score,total\_score]; %将各主成分的分与总分放在同一个举证中

result\_report = sortrows(result\_report,-4); %将总分降序排列

%% 输出模型及结果报告

disp('特征值及其贡献率、累计贡献率:')

DS

disp('信息保留率T对应的主成分与特征向量:')

Com\_num

PV

disp('主成分的分及排序)

result\_report

投影寻踪代码：

clear;clc

tic

data=xlsread("数值化处理.xlsx");

data=data(:,[5:30]);

[n,m]=size(data);

% 数据预处理

normal\_data=zeros(n,m);% 归一化矩阵初始化

for j=1:m

for i=1:n

normal\_data(i,j)=(data(i,j)-0.02\*min(data(:,j)))/(0.98\*max(data(:,j))-0.02\*min(data(:,j)));

end

end

% 模拟退火优化投影寻踪

for i=1:15

sensitivity=[];% 灵敏度分析矩阵

temper=10;% 初始温度

iter=10;% 迭代次数

temp=rand(1,m);% 生成一个线性投影初始解

templinear=temp;% 存储最优线性投影

Qa=Target(normal\_data,temp);% 生成初始目标函数值

sensitivity =[Qa];

while temper>1% 退火

for j=1:iter

temp1=rand(1,m);

Qa1=Target(normal\_data,temp1);

if Qa1>Qa

Qa=Qa1;

templinear=temp1;

else

if exp((Qa1-Qa)/temper)>rand()

Qa=Qa1;

templinear=temp1;

sensitivity = [sensitivity, Qa];

end

end

sensitivity = [sensitivity, Qa];

end

temper=temper\*0.999

TempLinear(i,:)=templinear;% 存储各样本的线性投影值

end

Y(i)=Qa;

i

end

plot(sensitivity)% 绘制灵敏度分析图

title('灵敏度分析');

target=TempLinear(find(max(Y)==Y),:)% 先找出最大的评价值，然后在从投影向量矩阵TempLinear中寻找该向量

score=sum(normal\_data.\*target,2);

toc

目标函数代码：

function y=Target(x,a)

%计算目标函数值，见模型步骤3

[m,n]=size(x);

for i=1:m

s1=0;

for j=1:n

s1=s1+a(j)\*x(i,j);

end

z(i)=s1;

end

Sz=std(z);%方差

R=0.1\*Sz;

s3=0;

for i=1:m

for j=1:m

r=abs(z(i)-z(j));

t=R-r;

if t>=0

u=1;

else

u=0;

end

s3=s3+t\*u;

end

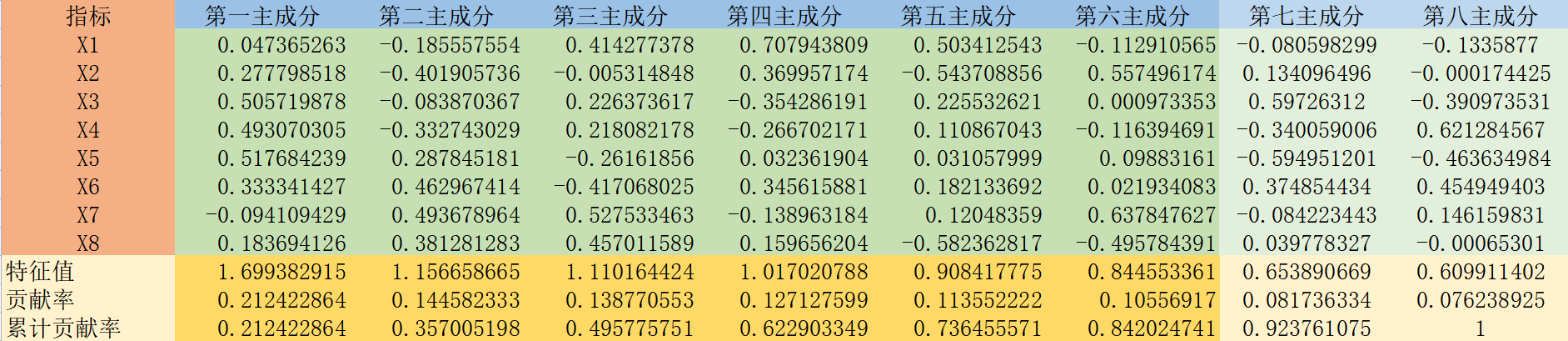
end

Dz=s3;

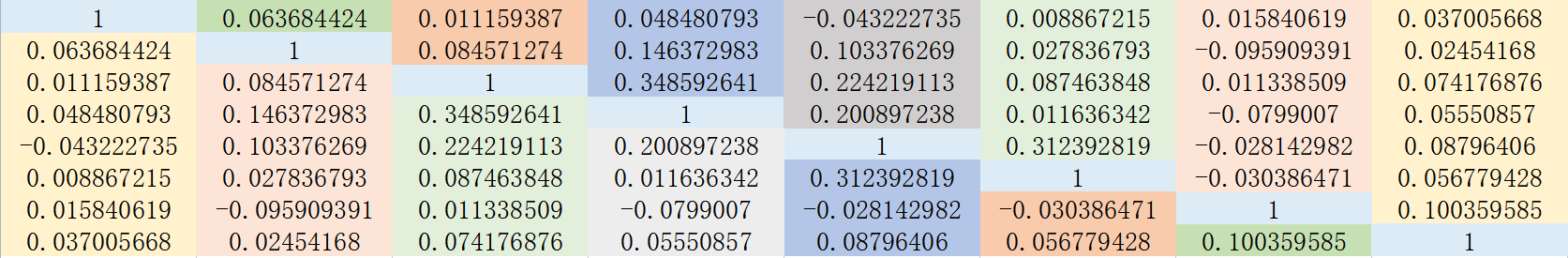
y=Sz\*Dz;

end

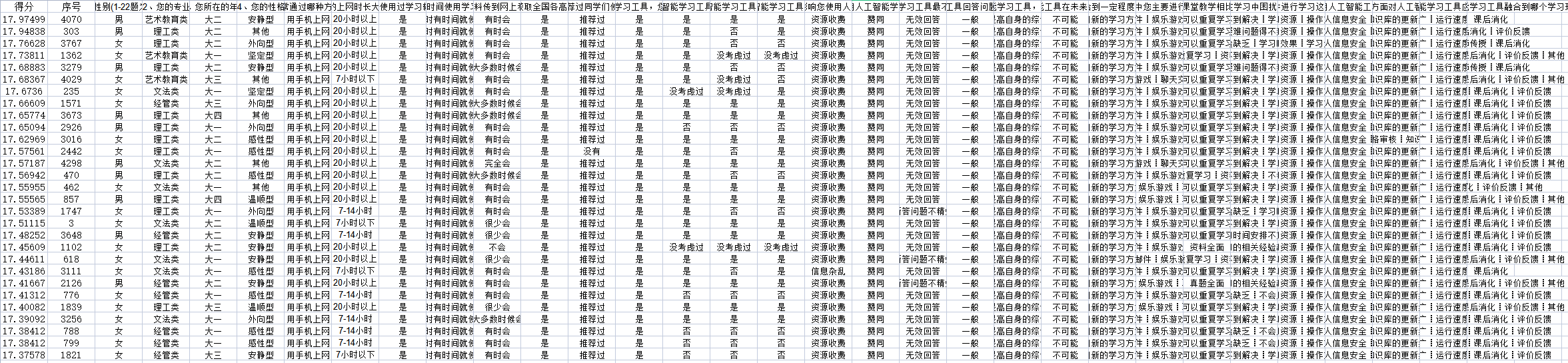
主成分分析结果图：



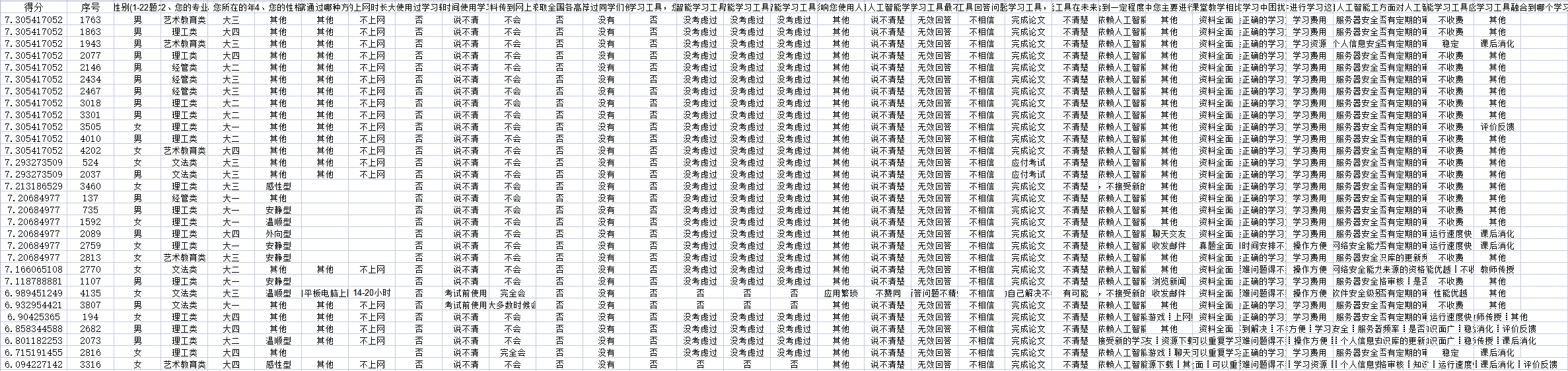
相关系数矩阵图：



得分排名前29名学生答题情况：



得分排名后30名学生答题情况：



遗传算法灵敏度分析：

