

基于改进的 GBDT 算法与模拟退火策略的母婴健康优化

摘要

本文首先通过基于 Apriori 的多因素关联性模型研究母亲身心健康对婴儿成长的影响，然后建立基于改进的 GBDT 关系模型研究母亲心理指标对婴儿行为特征和睡眠质量的影响，最后基于模拟退火的治疗策略分级规划模型求解最优心理治疗方案。

针对问题一，本文建立基于 Apriori 的多因素关联性模型研究母亲身体和心理指标对婴儿行为特征和睡眠质量的影响。首先，通过 IQR 法、KNNmputer 等方法进行异常值筛选与处理。然后，利用 Apriori 算法求解母婴指标间的关联度。最终求得母亲身体、心理指标与婴儿的行为特征和睡眠质量有影响，且教育程度、心理指标的影响较大。

针对问题二，本文建立基于改进的 GBDT 关系模型研究婴儿行为特征与母亲身体、心理指标的关系。本文通过引入 SMOTE 过采样和 Bayes 优化对 GBDT 算法进行改进，将附件中的前 390 项母婴指标数据作为训练集进行训练，利用五折交叉验证求得模型准确率为 **73.73%**。最后，利用训练好的模型对最后 20 组婴儿的行为特征进行预测，求得行为特征为中等型、安静型、矛盾型婴儿的组数分别为 **10、9、1**。

针对问题三，本文建立基于模拟退火的治疗策略分级规划模型求取最优治疗方案。首先，计算患病程度与治疗费用间的指数关系。然后，以治疗费用最小为目标函数，以行为特征、心理指标为约束条件建立多元非线性规划模型，利用模拟退火算法求解。改善至中间型的治疗方案为 EPDS 得分由 **22** 降低至 **21** 分，最少治疗费用为 **1472.15** 元。一级策略无法得出结果时，采用二级策略，以前 390 组中的安静型样本作为参考方案。改善至安静型的最优方案为 CBTS 得分由 **15** 降低至 **13** 分，最少治疗费用为 **1965.05** 元。

针对问题四，本文建立基于 FCM 的综合评判模型对婴儿睡眠质量进行预测。首先利用 FCM 算法将婴儿的睡眠质量分为优、良、中、差四类。然后，基于问题二对最后 20 组婴儿睡眠质量进行预测，求得睡眠质量为优良中差的组数分别为 **6、12、1、1**。最后对模型进行检验，求得模型的轮廓系数为 **0.6618**，CH 指数为 **1294.9**，准确率为 **86.59%**。

针对问题五，283 号婴儿的睡眠质量为中，与问题三相似，利用一级策略求得 283 号婴儿的睡眠质量改善至良的治疗方案为 HADS 得分由 **18** 降低至 **17** 分，最少治疗费用为 **1332.50** 元，但无法通过一级策略将睡眠质量变为优，利用二级策略求得改善至优的治疗方案为 CBTS 得分由 **15** 降低至 **13** 分，最少治疗费用为 **1965.05** 元。

本文的亮点有：1、本文在 GBDT 算法中引入 SMOTE 过拟合和 Bayes 优化进行改进，模型准确率更高；2、本文在求取最优治疗方案时采用分级策略，模型适用性更强。

关键词：Apriori 改进 GBDT 多元非线性规划 模拟退火 FCM

目录

1	问题重述	1
1.1	问题背景	1
1.2	问题提出	1
2	模型假设	2
3	符号说明	2
4	问题一模型的建立及求解	3
4.1	问题一分析	3
4.2	数据预处理	3
4.3	初步分析单因素间相关性	4
4.4	基于 Apriori 的多因素关联性模型	4
4.5	结果分析	6
5	问题二模型的建立与求解	8
5.1	问题二分析	8
5.2	基于改进的 GBDT 的关系模型	8
5.2.1	GBDT 算法求解	8
5.2.2	改进之处	9
5.3	结果分析	10
5.4	模型检验	11
6	问题三模型的建立与求解	12
6.1	问题三分析	12
6.2	患病程度与治疗费用间的关系	12
6.3	一级策略——基于模拟退火算法的治疗策略规划模型	13
6.3.1	多元非线性规划模型	13
6.3.2	模拟退火算法求解	14
6.3.3	求解结果及分析	15
6.4	二级策略——基于已知方案寻优的治疗策略规划模型	15
6.4.1	多元非线性规划模型	16

6.4.2 求解结果及分析	16
7 问题四模型的建立与求解	17
7.1 问题四分析	17
7.2 基于 FCM 的综合评判模型	17
7.3 评判结果	18
7.4 GBDT 综合睡眠质量预测	19
7.5 结果分析	20
7.6 模型检验	20
8 问题五模型的建立与求解	21
8.1 问题五分析	21
8.2 一级策略——基于模拟退火的多元非线性规划模型	21
8.2.1 求解及结果分析	22
8.3 二级策略	22
8.3.1 多元非线性规划模型	22
8.3.2 求解结果及分析	22
8.4 结果总结	23
9 模型总结与评价	23
9.1 模型总结	23
9.1.1 模型优点	23
9.1.2 模型缺点	23
9.1.3 模型改进	23
9.2 模型推广	24
参考文献	25
附录 A 事项关联度统计表（行为特征）	26
附录 B 事项关联度统计表（睡眠时间）	31
附录 C 事项关联度统计表（睡醒次数）	33
附录 D 事项关联度统计表（入睡方式）	36
附录 E 1-390 号婴儿睡眠质量信息表	41
附录 F 问题一 Python 源代码	45

附录 G	问题二 Python 代码	48
附录 H	问题三 Python 源代码	50
附录 I	问题四 Python 源代码	52
附录 J	问题五 Python 源代码	54

1 问题重述

1.1 问题背景

婴儿期是儿童睡眠习惯培养和睡眠模式发展的关键时期，婴儿睡眠对影响中枢神经系统发育和成熟有关键作用，甚至会对儿童发育产生长远的影响。而母亲作为婴儿的主要看护者，心理特征和身体状况会对婴儿具有最直接的影响，母亲身心健康情况的变化很可能会牵动者婴儿睡眠质量的改变，从而影响婴儿睡眠行为的发展。由于婴儿睡眠方式具有极高可塑性，清楚了解母亲身心健康对婴儿成长的影响是进行早期综合护理干预的关键，从而可以有效避免婴儿睡眠问题和儿童期忽视问题的发生。

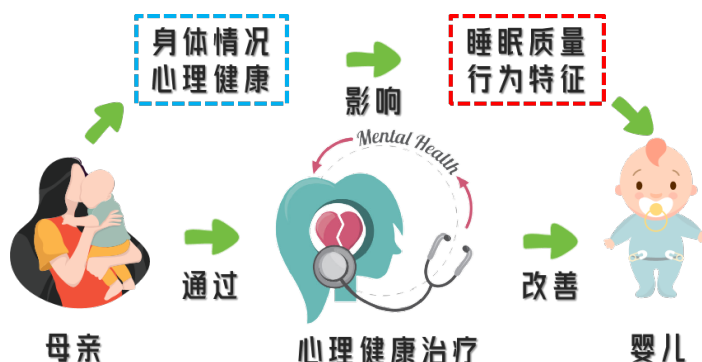


图 1 问题背景图

1.2 问题提出

附件为 390 名 1 至 3 个月的婴儿睡眠质量指标数据及母亲身体指标和心理指标数据。根据附件和相关文献解决如下四个问题。

问题一：根据附件数据，研究母亲的体身体指标和心理指标对婴儿行为特征和睡眠质量的影响规律。

问题二：已知婴儿行为特征分为安静型、中等型、矛盾型，要求建立婴儿的行为特征与母亲的体身体指标与心理指标的关系模型，并预测附件中最后 20 组的婴儿行为特征类型。

问题三：已知 CBTS、HADS、EPDS 的治疗费用相对于患病程度的变化率都与治疗费用呈正比和两个分数对应的治疗费用。要求建立模型，分析能够使编号为 238 的婴儿的行为特征从矛盾型变为中等型需要花费的最少治疗费用，并分析如果要使其行为特征变为安静型，治疗方案又需要如何调整。

问题四：根据婴儿的睡眠质量指标进行综合评判，将其分为优、良、中、差四分类，并建立婴儿综合睡眠质量与母亲的体心理指标、体身体指标的关联模型，预测最后 20 组的婴儿的综合睡眠质量。

问题五：基于问题三，分析如果需要对 238 号婴儿的睡眠质量评级为优，问题三求得的治疗策略是否需要调整，若需要调整，要求分析如何调整。

2 模型假设

- (1) 假设表 1 患病得分与治疗费用中的治疗费用是指 CBTS、EPDS、HADS 三项心理指标从患病得分降到 0 分的治疗费用；
- (2) 假设附件中的数据除去个别异常之外其他均真实可靠；
- (3) 假设仅通过改变母亲的心理健康水平便可改善婴儿的行为特征和睡眠质量；
- (4) 假设附件中已有的母亲心理健康信息可作为心理健康改善的现有参考方案。

3 符号说明

符号	说明
N	附件中存在完整数据的样本数量
$f_i(x)$	第 i 项心理指标治疗费用患病程度相关函数
m_i	治疗前 238 号婴儿的母亲三项心理指标初始得分
p_i	238 号婴儿的母亲身体指标
Z	238 号婴儿行为特征由矛盾型变为中等型所需治疗费用
R	238 号婴儿行为特征由矛盾型变为安静型所需治疗费用
U	FCM 中隶属度的模糊矩阵
c	睡眠质量级别数

注：表中未列出及重复的符号均以首次出现处为准。

4 问题一模型的建立及求解

4.1 问题一分析

问题一要求我们根据附件数据研究母亲的身体指标和心理指标与婴儿的行为特征和睡眠质量的关系。

首先，我们通过**人为判断**和**IQR 法**对附件数据进行异常值筛选，用 **KNNmputer 法**、**众数替换**、**平均值替换法**处理异常数据。然后，通过求取母亲的身体指标与心理指标和婴儿的行为特征与睡眠质量的 spearman 相关系数矩阵，对它们之间的关系进行了初步的分析，发现**单个因素间的相关性关系较差**，因此猜测可能存**多因素间的交互效应**，即母亲的多个身心健康指标形成**混杂因素**来影响婴儿的行为特征与睡眠质量。本文决定采用 **Apriori(关联规则挖掘) 算法**^[1]来挖掘存在的规律。最后，根据求得的**关联规则**判断母婴指标之间的关联性。

问题一的思路如下面图2所示。

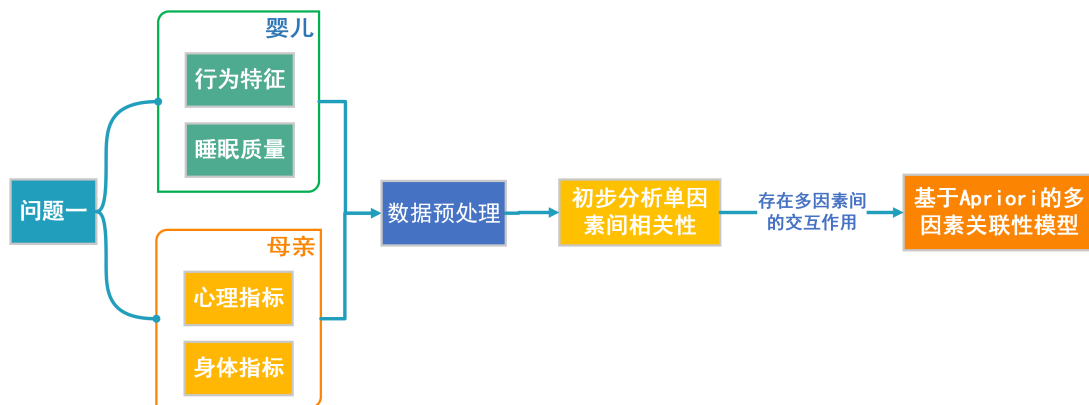


图 2 问题一思路图

4.2 数据预处理

在研究母婴指标间的关联性前，首先进行数据预处理工作。通过观察附件数据，发现无缺失值存在，但可能存在异常值，我们通过人为判断与 IQR 法两种方法对异常数据进行筛选，具体处理情况如下。

人为判断

通过观察附件数据，发现婚姻状况存在填写 3 和 6 的情况，不符合题目要求，本文用该列数据的众数替换这部分数据。整晚睡眠时间存在填写 99 : 99 的情况，明显不符合实际，本文用该列数据的平均值替换此异常数据。

IQR 法判断

通过 IQR 法筛选出每列指标的异常数据，并将其置为空，然后采用 KNNmputer 法^[7]进行填补。通过欧几里德距离矩阵寻找关于最近邻样本，使用最近邻样本的对应位

置的数值的非空均值填补空缺数据。

4.3 初步分析单因素间相关性

本文首先通过计算母亲的身体健康指标与心理指标和婴儿的行为特征与睡眠质量的相关性，计算结果的热力图如下。

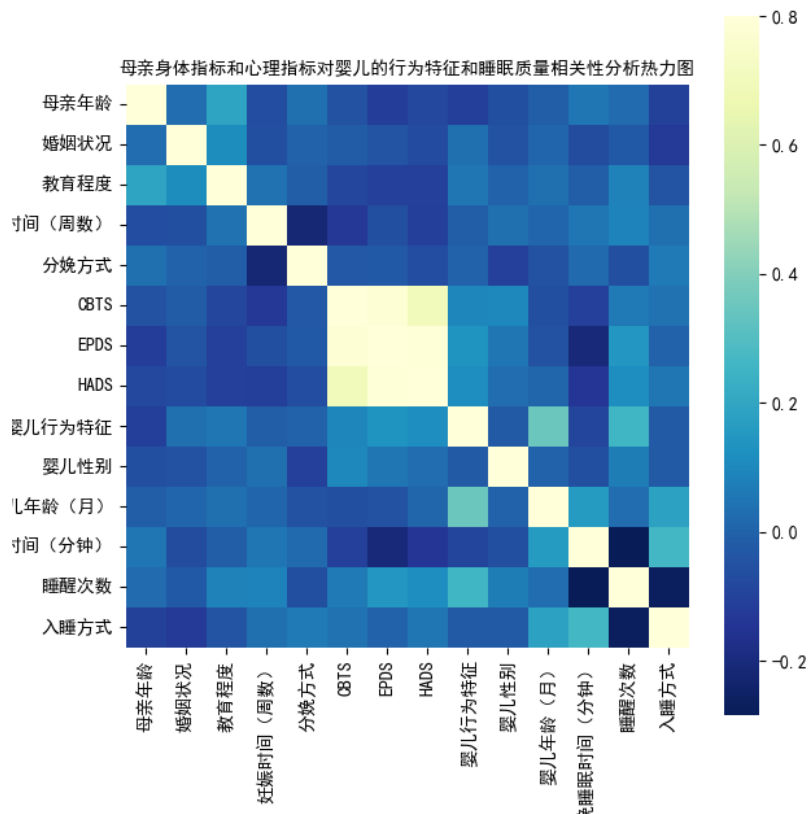


图3 母亲的身体健康指标、心理指标与婴儿的行为特征、睡眠质量的相关性热力图

从图中可以看出母亲的身体健康指标和心理指标与婴儿的行为特征和睡眠质量的相关性系数几乎都小于 0.2，可见单因素之间的相关性较弱。由此，我们猜测母亲身体心理单项指标的改变对婴儿的行为特征和睡眠质量影响较弱，但可能多个指标同时变化会对婴儿的行为特征和睡眠质量产生较大影响。因此本文决定改进研究方向，寻找多因素之间的关联性。

4.4 基于 Apriori 的多因素关联性模型

本文采用 Apriori(关联规则挖掘) 算法研究母亲的身体健康指标和心理指标与婴儿的行为特征和睡眠质量之间的多因素间交互效应。Apriori 算法是一种逐层搜索的迭代方法，

通过在迭代过程中不断链接和剪枝，最终找出数据集中的相继的关系以形成规则。从而，我们可以根据最终形成的规则来判断数据集中各项之间隐藏的属性之间的关系。

该算法中主要概念的含义如下：

(1) 支持度：指数据集中包含某项集 A, B 的记录所占的比例，计算公式为

$$support(A \rightarrow B) = P(A, B). \quad (1)$$

(2) 置信度：指先决条件 A 发生的前提下，另一事务 B 发生的概率，计算公式为

$$Confidence(A \rightarrow B) = P(A|B) = \frac{P(A, B)}{P(A)} \quad (2)$$

(3) 提升度：关联规则的前向事务置信度与后向事务支持度的比值，即事务 B 发生的条件下，同时含有 A 的概率与 A 总体发生的概率的比值，计算公式为

$$Lift(A \rightarrow B) = P(A|B)/P(B) = \frac{Confidence(A \rightarrow B)}{P(B)} \quad (3)$$

如果某个项集大于最小支持度，则称为频繁项集，是指经常出现在一块即关联度较大的项的集合。该算法即是利用频繁项集的先验性质来不断压缩搜索空间，至找到关联规则。算法的主要步骤如下图所示：

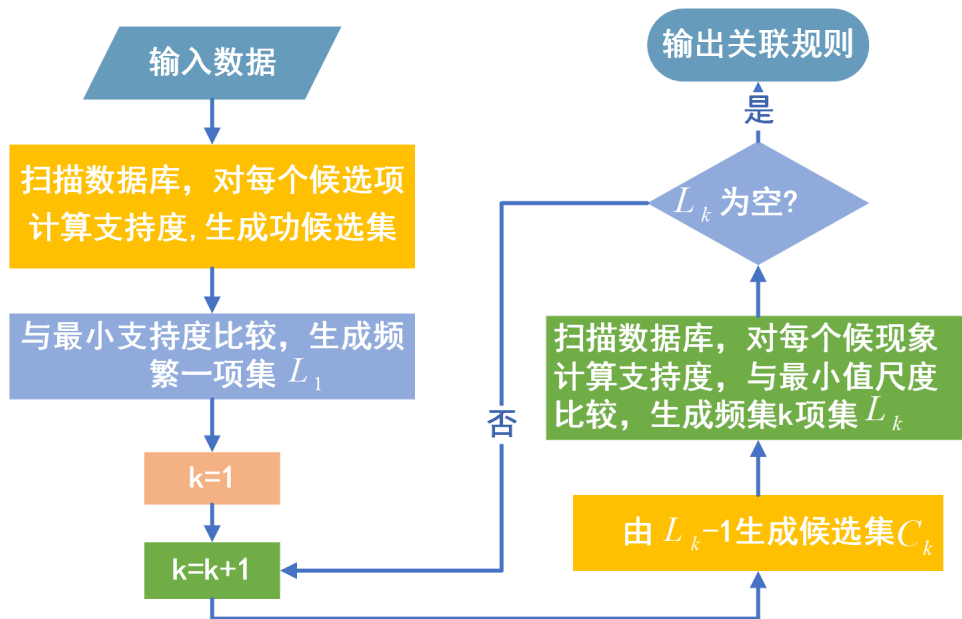


图 4 Apriori 算法的主要步骤流程图

在上述步骤中，对给定频繁集进行关联规则挖掘的具体实现用伪代码表示如下。

Algorithm 1: Mining Regulation

Input: D :dataSets; I :frequentSet; β :Confidence threshold;

Output: R :association rules;

```
1 begin
2   compute the non empty and proper subset of  $I$ : subSets;
3   for  $S$  in subSets do
4     if  $\text{Confidence}(S \rightarrow I - S) \geq \beta$  then
5        $R_+ = (S \rightarrow I - S)$ 
6     end
7   end
8 end
```

在上述伪代码中, β 表示最小置信度。由于附件二数据集中行为特征为矛盾型的婴儿占比最小为 0.12, 为保证频繁项集中涵盖所有行为特征的婴儿, 本文将算法中的最小置信度设置为 0.2.

4.5 结果分析

本文将附件二中的母亲的心理健康指标与婴儿的行为特征睡眠质量数据代入到 Apriori 算法中, 大于最小置信度 0.2 的共计 463 条频繁项集, 本文主要用置信度和关联度量关联规则。根据关联规则可将事务间的关联度表示为前项 \rightarrow 后项, 前项为先决条件, 后项为关联的结果。

整体上, 我们统计婴儿的行为特征与睡眠质量的平均置信度和平均提升度, 结果如下。

表 1 Apriori 算法结果指标

算法结果指标	行为特征	入睡方式	睡眠时间	睡醒次数
平均置信度	0.5124	0.4279	0.2979	0.3409
平均提升度	1.1284	1.0906	1.1469	1.1574

从表中可以看出, 平均置信度几乎都大于 0.3, 说明母亲的身体心理指标的变化对婴儿行为特征、入睡方式、睡眠时间、睡醒次数均有关联, 并且尤其对行为特征、入睡方式的影响较大。平均提升度均大于 1, 说明母亲的身心健康情况与婴儿的行为特征的睡眠质量成正相关关系。局部上, 统计入睡方式、行为特征、睡眠时间、睡醒次数的置

信度和提升程度前三名的频繁项集结果如下，详细请见附录。

表 2 关联度较高的事项统计结果

前项	后项	置信度	提升度
中_EPDS, 已婚, 研究生	婴儿行为特征为中等型	0.656716	1.138308
自然分娩, 中_EPDS, 研究生, 已婚	婴儿行为特征为中等型	0.656716	1.138308
低_CBTS, 研究生, 妊娠时间正常	婴儿行为特征为中等型	0.650794	1.128042
大学, 已婚, 中年	入睡方式为环境营造法	0.590909	1.379967
大学, 中年	入睡方式为环境营造法	0.588235	1.373723
自然分娩, 大学, 中年	入睡方式为环境营造法	0.58209	1.359371
自然分娩, 妊娠时间正常, 研究生, 已婚	睡眠时间为 600 分钟	0.358108	1.305254
已婚, 妊娠时间正常, 研究生, 中年	睡眠时间为 600 分钟	0.357724	1.303852
妊娠时间正常, 研究生, 自然分娩, 中年, 已婚	睡眠时间为 600 分钟	0.357724	1.303852
低_CBTS, 已婚, 低_HADS	睡醒次数为 0	0.452632	1.389971
自然分娩, 低_CBTS, 已婚, 低_HADS	睡醒次数为 0	0.446809	1.372089
低_CBTS, 已婚	睡醒次数为 0	0.426471	1.309634

从表中可以看出，无论在睡眠质量上，还是行为特征上，多个频繁项集中都包括教育程度、心理指标，由此可见两者对婴儿睡眠质量和行为特征影响较大。并且比较来看，教育程度越高，心理指标越好，婴儿的睡眠质量越好、行为情绪越稳定。

5 问题二模型的建立与求解

5.1 问题二分析

问题二要求建立婴儿的行为特征与母亲的身体指标与心理指标的关系模型，并对最后 20 组婴儿的行为特征的分类进行判断。

首先，我们建立了**基于改进的 GBDT 的关系模型**，将前 390 项的母亲的身体健康和心理健康指标与对应的婴儿的行为特征作为训练集，带入到 GBDT 算法中，然后，用训练好的决策树模型，根据最后 20 项母亲的身体健康和心理健康指标数据，对婴儿的行为特征进行预测。最后用**五折交叉验证**检验模型的准确性。

问题二的思路图如下。

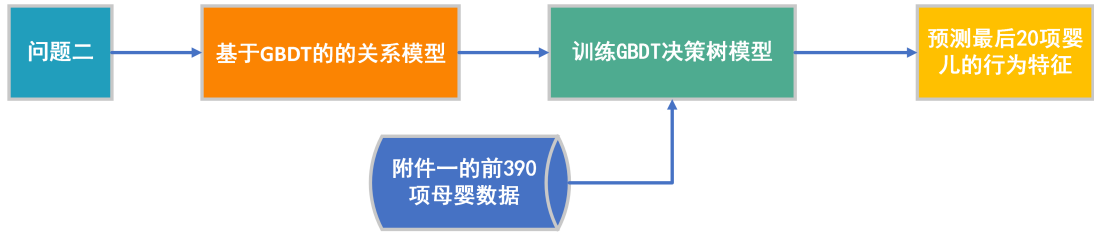


图 5 问题二思路图

5.2 基于改进的 GBDT 的关系模型

本文采用改进的 GBDT (Gradient Boosting Decision Tree) 算法来建立婴儿的行为特征与母亲的身体指标与心理指标的关系模型，并利用训练好的树模型对最后 20 组婴儿的行为特征的分类进行预测。

5.2.1 GBDT 算法求解

梯度增强决策树 GBDT^{[3][9]} 是一种迭代的决策树算法，有多棵决策树组成，其在训练过程中确定一系列合适的决策树作为弱学习器，并将弱学习器线性组合成为强学习器后对测试样本进行判断。GBDT 通过不断的迭代，使得最后得到的输出结果不断靠近训练值并得到相对最优解。

GBDT 可以看作是多棵树组成的加法模型，给定训练集 $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$ ，其中 x 表示母亲的身体指标与心理指标， y 表示对应的婴儿的行为特征， N 为 390，模型公式可表示为

$$F_M(x) = \sum_{m=1}^M T(x, \Phi_m) \quad (4)$$

式中 $F_M(x)$ 表示分类预测函数， x 表示要输入的样本的数据， M 表示树的个数， $T(x, \Phi_m)$ 表示第 m 棵决策树。

利用 GBDT 算法的实现过程主要步骤如下：

(1) 确定初始的提升树为

$$F_0(x) = 0. \quad (5)$$

(2) 结合前项分布算法，则第 s 步的算法模型为

$$F_s(x) = F_{s-1}(x) + T(x, \Phi_s). \quad (6)$$

式中 $F_{s-1}(x)$ 表示第 $s-1$ 步的回归分类函数模型。

(3) 通过得到的预测的分类类别 $F(x)$ 使得损失函数 $L(y, F(x))$ 最小，从而来确定第 s 个决策树的参数 Φ_s ，即

$$\Phi_s = \min \sum_{i=1}^M L[y_i, F_{s-1}(x_i) + T(x_i; \Phi_s)]. \quad (7)$$

通过迭代过程，输出结果不断靠近训练值，最终得到训练良好的决策树模型，即婴儿的行为特征与母亲的身体指标与心理指标的关系模型。根据最后 20 项的母亲的身体指标和心理指标数据预测婴儿的行为特征，并利用 softmax 函数将模型的原始预测值转换为概率，计算公式如下

$$P(F_M(x) = k) = \frac{e^{s_k}}{\sum_{i=1}^3 e^{s_k}} \quad (8)$$

其中 $F_M(x)$ 表示 GBDT 的预测结果分值， $P(F_M(x) = k)$ 表示预测得分为 k 的概率。

5.2.2 改进之处

本文主要从两个方面对 GBDT 进行改进，分别是 SMOTE 过采样与 Bayes 优化。

一方面，由于一些行为特征如矛盾型占比过少，可能会造成数据类别不平衡问题，不平衡样本会导致最终训练模型侧重样本数目较多的类别，而“轻视”样本数目较少类别，这样模型在测试数据上的泛化能力就会受到影响。因此，为解决不平衡样本问题，本文采用 SMOTE 合成少数类过采样技术^[10]对数据进行过采样，这是在随机采样的基础上改进的一种过采样算法，其实现过程为：

(1) 选择少数类样本：从少数类样本中随机选择一个样本 x_i ；

(2) 找到其 k 个最近邻：使用 K 近邻算法找到该样本的 k 个最近邻居，记为 $N_k(x)$ ；

(3) 从 k 个最近邻中随机选择一个邻居：从 $N_k(x)$ 中随机选择一个邻居 x_{ni} ；

(4) 在所选样本和所选邻居之间生成新样本：选择两个样本之间的差异，并乘以一个在 0 到 1 之间的随机数 β ，然后将其加到所选样本上即可得到平衡数据集。合成的计算公式如下

$$x_n = x_i + \beta \times (x_{ni} - x_i) \quad (9)$$

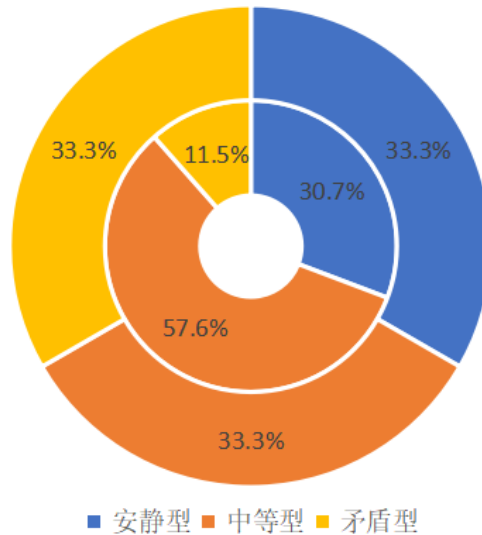


图 6 SMOTE 过采样前后对比图

另一方面，我们采用贝叶斯优化算法^[2]对 GBDT 的超参数进行优化，希望得到最优超参数组合，以获得更好的模型性能。它基于贝叶斯定理，结合代理模型和目标函数的信息，动态地选择下一次探索的点，以在有限的迭代次数内找到可能的最优解。其实现过程为：

- (1) 定义超参数空间：确定要优化的超参数，本文为学习率、树的数量、树的深度、子采样比例；
- (2) 选择目标函数：确定评估模型性能的指标，本文选择的目标函数为交叉验证的准确率；
- (3) 迭代优化：贝叶斯优化会在每一轮迭代中，根据代理模型和目标函数评估，选择下一个要尝试的超参数组合。通常，它会在未知区域（不确定性高）进行探索，或者在代理模型预测较好的地方进行利用。每次迭代后，会根据实际目标函数的值，更新代理模型；
- (4) 终止条件：设置终止条件，例如迭代次数达到一定限制或者目标函数的改进不再显著；
- (5) 应用最优参数：在贝叶斯优化结束后，得到的最优超参数组合可以用于训练 GBDT 模型。

5.3 结果分析

通过 GBDT 算法求解得到 390-410 号婴儿的行为特征预测结果如下，统计预测的结果中各类行为特征的频数如下。

表 3 婴儿行为特征预测结果表

编号	391	392	393	394	395	396	397	398	399	400
行为特征	中等型	中等型	中等型	中等型	中等型	中等型	安静型	中等型	中等型	中等型
编号	391	392	393	394	395	396	397	398	399	400
行为特征	中等型	中等型	中等型	安静型	中等型	安静型	中等型	安静型	安静型	中等型

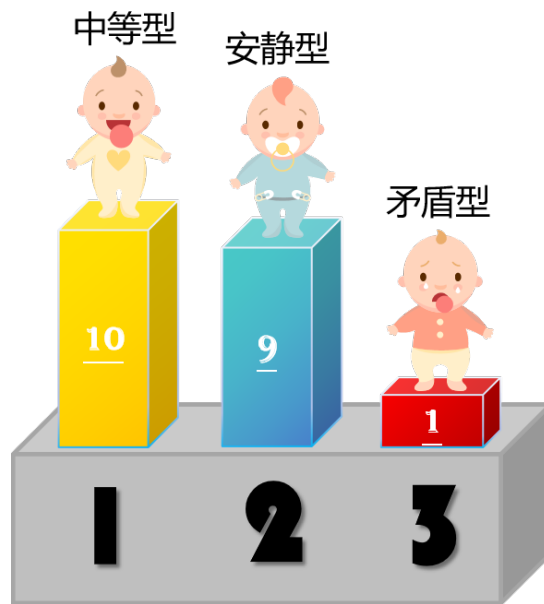


图 7 婴儿行为特征的预测结果直方图

5.4 模型检验

本文采用交叉验证对上述 GBDT 模型性能进行检验。交叉验证是一种统计学上将样本数据切割为较小子集的实用方法，可以重复运用随机生产的子样本的数据进行多次训练验证，本文采用五折交叉验证的方法划分训练集，将原始数据集分成五个大小相等的子集，其中四个子集用作训练数据，剩下的一个子集用作验证数据，以往重复进行五次，每次使用不同的子集作为验证数据，最终得到五个模型结果。本文对得到的五个模型结果求取平均值。本文最终计算得到模型在测试集的平均准确率达到 **73.73%**，效果较好，但还能进一步提升。

6 问题三模型的建立与求解

6.1 问题三分析

问题三是一个规划问题，要求求取 238 号婴儿行为特征由矛盾型变为中等型和安静型分别需要的最小治疗费用和对应的治疗方案。经过分析题目，我们需要通过**干预母亲的心理健康水平来改变婴儿的行为特征**，即通过治疗改善母亲的 CBTS、EPDS、HADS 的三项心理指标得分来促成婴儿的行为特征的改变。

首先，根据题目计算患病程度与治疗费用间的函数关系。然后，我们建立**一级策略规划模型**，以最小治疗费用为目标函数，引入婴儿治疗前后行为特征变化约束和心理指标得分两类约束条件建立**多元非线性规划模型**^[5]，并利用**模拟退火算法**求解。但通过观察 238 号婴儿行为特征的预测概率随单项心理指标的变化情况，发现通过一级策略无法实现从矛盾型转变为安静型。接着，我们建立**二级策略规划模型**，在前 390 组数据婴儿行为特征为安静型样本中，寻找能使 238 号婴儿为达到各样本中母亲心理指标得分需要花费治疗费用最小的样本，将其作为从矛盾型转变为安静型最优治疗方案的参考。

问题三的思路图如下。

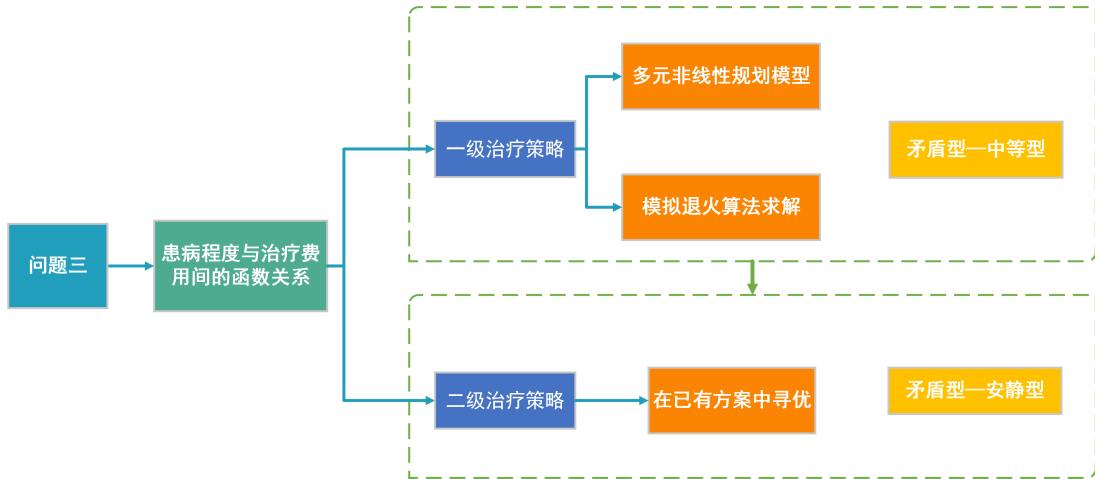


图 8 问题三思路图

6.2 患病程度与治疗费用间的关系

根据题目 CBTS、EPDS、HADS 的治疗费用相对于患病程度的变化率均与治疗费用呈正比，则 CBTS、EPDS、HADS 三项心理指标与患病程度之间的函数关系可以表示为：

$$\begin{cases} \frac{\partial f_i(x_1, x_2)}{\partial x_1} \propto f_i(x_1, x_2) \\ \frac{\partial f_i(x_1, x_2)}{\partial x_2} \propto f_i(x_1, x_2) \end{cases} \quad (10)$$

即可求得

$$f_i(x_1, x_2) = C_i e^{k_i(x_1 - x_2)} \quad (i = 1, 2, 3) \quad (11)$$

式中 x_1 、 x_2 表示产妇治疗前后的患病程度； $f_i(x)$ 代表产妇第 i 项心理指标的治疗费用关于患病程度的函数，其中 $i = 1, 2, 3$ 分别表示 CBTS、EPDS、HAD； C 与 k 表示常数。

从而，通过题目给出的每项产妇心理指标的两个分数对应的治疗费用，我们可以分别求得 CBTS、EPDS、HAD 的治疗费用与患病程度的函数关系为

$$f_1(x_1, x_2) = 200e^{0.8811(x_1 - x_2)} \quad (12)$$

$$f_2(x_1, x_2) = 500e^{0.6649(x_1 - x_2)} \quad (13)$$

$$f_3(x_1, x_2) = 300e^{0.7459(x_1 - x_2)} \quad (14)$$

6.3 一级策略——基于模拟退火算法的治疗策略规划模型

6.3.1 多元非线性规划模型

目标函数

问题要求最少需要花费最小治疗费用改善母亲的心理健康水平使婴儿的行为特征能够从矛盾型变为中等型，即

$$\min Z = \sum_{i=1}^t f_i(m_i, m'_i) \quad (15)$$

其中 Z 表示治疗费用， t 表示心理指标类别数即 3， m_i 表示治疗前 238 号婴儿的母亲的 CBTS、EPDS、HAD 三项心理指标的初始得分； m'_i 表示治疗后母亲的 CBTS、EPDS、HAD 三项心理指标的得分。

约束条件

(1) 婴儿行为特征约束：238 号婴儿的母亲的身体指标已知，并假设治疗过程中仅心理指标发生变化，且婴儿的行为特征为矛盾型，治疗后为中等型，由此可写出以下约束：

$$P(y = b_0) = F(p_1, p_2, \dots, p_n, m_1, m_2, m_3) = \max \{P(y = k) | k = 1, 2, 3\} \quad (16)$$

$$P'(y = b_1) = F(p_1, p_2, \dots, p_n, m'_1, m'_2, m'_3) = \max \{P'(y = k) | k = 1, 2, 3\} \quad (17)$$

其中 b_0 表示治疗前 238 号婴儿的行为特征为矛盾型即取值为 3； b_1 表示治疗前婴儿的行为特征为中等型即取值为 2； p_1, p_2, \dots, p_n 表示 238 号母亲的身体指标数据， n 表示身体指标类别即 5； F 表示 GBDT 的分类预测函数。

(2) 心理指标分数约束：CBTS、EPDS、HAD 三项心理指标的得分最大为 30，最小为 0 分。并且治疗后的心理健康理应得到了优化，CBTS、EPDS、HAD 三项心理指标的得分降低，由此可写出以下约束：

$$m'_i \leq m_i \quad (m'_i \in [S_{\min}, S_{\max}]) \quad (18)$$

其中 S_{\max} 表示各项心理指标的最大得分即 30 分， S_{\min} 表示各项心理指标的最小得分即 0 分。

模型总述:

$$\min Z = \sum_{i=1}^t f_i(m_i, m'_i) \quad (19)$$

$$\begin{cases} P(y = b_0) = F(p_1, p_2, \dots, p_n, m_1, m_2, m_3) = \max\{P(y = k) | k = 1, 2, 3\}; \\ P'(y = b_1) = F(p_1, p_2, \dots, p_n, m'_1, m'_2, m'_3) = \max\{P'(y = k) | k = 1, 2, 3\}; \\ m'_i \leq m_i; \\ m'_i \in [S_{\min}, S_{\max}]. \end{cases} \quad (20)$$

6.3.2 模拟退火算法求解

模拟退火算法^[4]是一种广泛应用的智能优化算法，其原理灵感来源于固体材料的退火过程，其中材料在高温下被加热，然后逐渐冷却，从而达到更低能量状态。类比到优化问题中，模拟退火算法把目标函数作为能量函数，在解空间中搜索全局最优解，并且可通过一系列的随机搜索和接受劣质解的策略，从而避免陷入局部最优解。

模拟退火的主要步骤如:

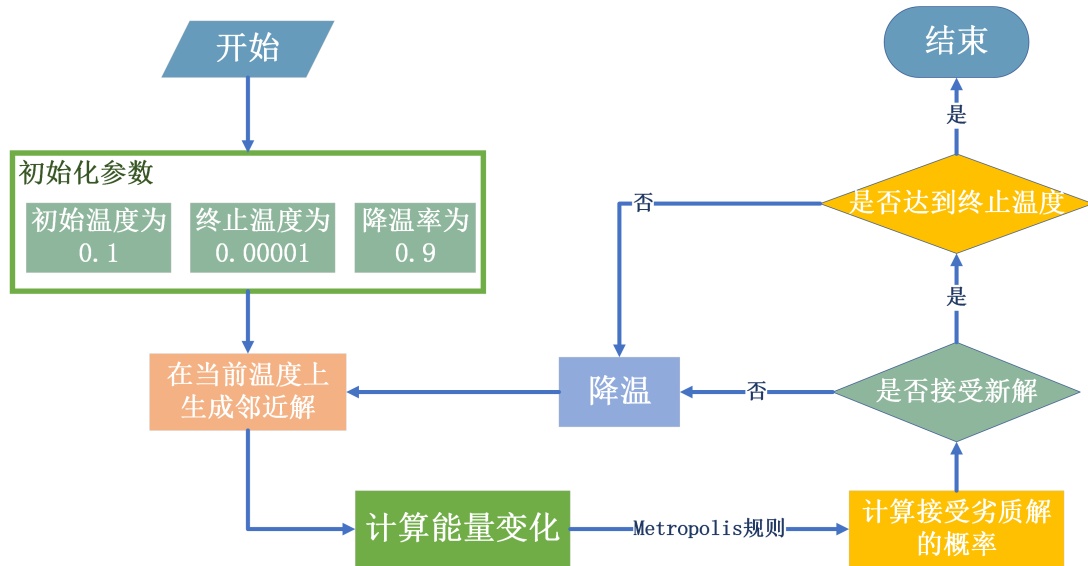


图9 模拟退火算法主要步骤

6.3.3 求解结果及分析

第一小问：矛盾型 → 中等型

基于上述规划模型，通过模拟退火算法，我们可以求解得到为将第 238 号婴儿的行为特征从矛盾型变为中等型，其母亲 EPDS 项心理指标得分需要从 22 变为 21.47 分，其他两项心理指标 CBTS、HADS 不需要发生改变，并求得最少治疗费用为 **1212.55** 元。优化得分出现小数，这是由于为了反映随着治疗心理健康情况的变化是连续的，本文在上述模型中并未引入心理指标得分必须为整数的约束。经考虑实际治疗结果监测情况，本文最终将治疗方案制定为将母亲 EPDS 得分由 22 降低至 21 分，并求得治疗费用为 **1472.15** 元。

第二小问：矛盾型 → 安静型

若使 238 号婴儿的行为特征变为安静型，可以将线性规划模型中的 b_1 改变取值为 1，接着利用上述模型算法求解。但是，本文通过统计 238 号婴儿的不同行为特征的预测概率分别随 CBTS、HADS、EPDS 单个指标的变化情况，发现无法将婴儿的行为特征从矛盾型变为安静型。以 238 号婴儿的不同行为特征的预测概率随 EPDS 指标的变化曲线为例，从下图可以看出无论如何改变 EPDS 指标，安静型概率无法大于矛盾型概率使婴儿的行为特征表现为安静型。因此，本文决定采用二级策略来实现 238 号婴儿行为特征的改变。

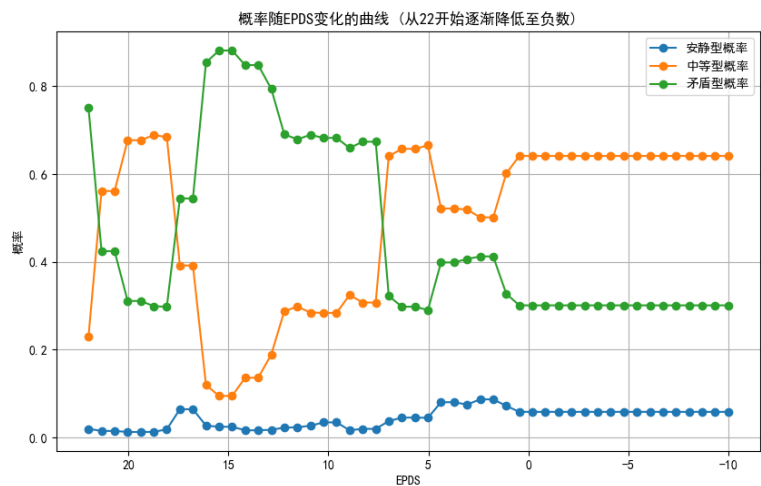


图 10 238 号婴儿的不同行为特征的预测概率随 EPDS 指标的变化曲线

6.4 二级策略——基于已知方案寻优的治疗策略规划模型

当由一级策略无法达到目标时，采取二级策略寻优。将已知的前 390 组中每个安静型的数据样本作为 238 号婴儿治疗后母亲心理指标可能的结果，通过计算达到每一个样

本的三个心理指标得分所需的治疗费用，选出最小的治疗费用的方案，基于此本文建立多元非线性规划模型。

6.4.1 多元非线性规划模型

目标函数

设 W 表示为 238 号婴儿行为特征由矛盾型变为安静型时，母亲改善心理健康指标需要的治疗费用。要求达到已知安静型心理指标的方案需要花费的治疗费用最小。

$$\min R = \sum_{i=1}^t f_i(m_i, M'_{ij}) \quad (i = 1, 2, \dots, w) \quad (21)$$

其中 w 表示前 390 组中安静型的数据样本数量，即 120； m_i 表示治疗前 238 号婴儿的母亲的 CBTS、EPDS、HAD 三项心理指标的初始得分； M'_{ij} 表示治疗后 238 号婴儿的母亲的 CBTS、EPDS、HAD 三项心理指标的得分。

约束条件

(1) 婴儿行为特征约束：婴儿治疗前的行为特征为矛盾型，在已知数据集中选取的方案婴儿的行为特征为安静型。

$$P(y = b_0) = F(p_{1j}, p_{2j}, \dots, p_{nj}, m_{1j}, m_{2j}, m_{3j}) = \max \{P(y = k) | k = 1, 2, 3\} \quad (22)$$

$$P'(y = b_2) = F(p_{1j}, p_{2j}, \dots, p_{nj}, m'_{1j}, m'_{2j}, m'_{3j}) = \max \{P'(y = k) | k = 1, 2, 3\} \quad (23)$$

其中， b_2 表示选取的方案中的婴儿行为特征都为安静型即取值为 1， m'_{ij} 表示第 j 个安静型数据样本的 CBTS、EPDS、HADS 三项心理指标得分。

(2) 心理指标得分约束：治疗后需要达到的心理指标得分取选取方案和初始得分中的较小者。

$$M_{ij} = \min \{m'_{ij}, m_{ij}\} \quad (24)$$

模型总述

$$\min R = \sum_{i=1}^t f_i(m_i, M'_{ij}) \quad (i = 1, 2, \dots, w) \quad (25)$$

$$\begin{cases} P(y = b_0) = F(p_{1j}, p_{2j}, \dots, p_{nj}, m_{1j}, m_{2j}, m_{3j}) = \max \{P(y = k) | k = 1, 2, 3\}; \\ P'(y = b_2) = F(p_{1j}, p_{2j}, \dots, p_{nj}, m'_{1j}, m'_{2j}, m'_{3j}) = \max \{P'(y = k) | k = 1, 2, 3\}; \\ M_{ij} = \min \{m'_{ij}, m_{ij}\} \end{cases} \quad (26)$$

6.4.2 求解结果及分析

通过上述多元非线性规划模型，本文求解得到 260 号为最优方案，由此得出，为将 238 号婴儿的行为特征从矛盾型变为安静型的治疗方案为将 CBTS 心理指标得分由 15

改善为 13，其他两项心理指标 EPDS、HADS 不需要发生改变，并求解得到的最小治疗费用为 **1965.05** 元。

7 问题四模型的建立与求解

7.1 问题四分析

问题四是一个聚类问题，要求我们将附件中的婴儿的睡眠质量进行优、良、中、差四分类综合评判，并根据评判结果建立**婴儿综合睡眠质量与母亲的身体指标、心理指标的关联模型**，最后预测最后 20 组婴儿的综合睡眠质量。

首先，我们利用 **FCM(模糊 C 均值聚类)** 依据整晚睡眠时间、睡醒次数、入睡方式数据对婴儿进行聚类为优、良、中、差四个等级，并利用**轮廓系数**和 **CH 指数**对模型进行评价。然后，以前 390 组母亲的心理指标和身体指标与婴儿的睡眠质量评判结果作为训练集，利用问题二的 GBDT，预测最后 20 组婴儿的**睡眠质量评级**。

问题四的思路图如下。

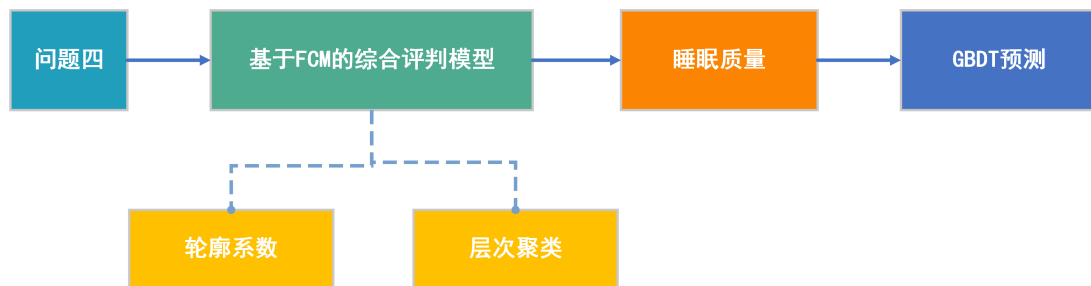


图 11 问题四思路图

7.2 基于 FCM 的综合评判模型

本文需要根据反映婴儿的睡眠质量的各项指标数据，对睡眠质量进行优、良、中、差四分类综合评判。由于附件所有样本的类别标签均未给出，这属于无监督学习范畴，本文决定采用模糊聚类中的模糊 C 均值聚类对应二的睡眠质量类别进行评判。

模糊聚类^{[6][8]}是多元统计分析的一种，也是无监督模式识别的一个重要分支，可将没有类别标记的样本集按照某种准则划分为若干类别。与传统的硬聚类方法（如 K 均值）不同，它允许一个数据样本属于不同的类别的程度，而不是仅属于一个确定的类别。最终可根据每个数据点与每个聚类中心之间的隶属度值来确定该数据点属于该聚类的程度。本文采用的模糊 C 均值聚类法 (FCM) 一种常用的模糊聚类方法，其核心是通过优化隶属度和聚类中心来达到最小化目标函数。

给定的 390 组婴儿的各项睡眠指标数据集 $X = \{x_1, x_2, \dots, x_N\}$ ，假设 $U = \{u_{ij}\}_{c \times N}$ 为隶属度的模糊矩阵，其中 c 表示聚类数，由于题目中已将睡眠质量分为差、中、优、

良四类，因此 c 取 4. 假设类别中心矩阵为 $V = \{v_j\}_{c \times 1}$ 则隶属度的计算公式为

$$\mu_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{d_{ij}}{d_{ik}} \right)^{2/(\tau-1)}} \quad (27)$$

其中 $d(x_i, v_k)$ 表示数据点 x_i 到聚类中心 v_k 的距离即 $d_{ik} = \|x_i - v_k\|$; v_k 表示上次迭代更新所得到的第 k 类数据的类别中心; τ 表示模糊因子设置为 2。

聚类中心的计算公式为

$$\nu_j = \frac{\sum_{i=1}^N \mu_{ij}^\tau x_i}{\sum_{i=1}^N \mu_{ij}^\tau} \quad (28)$$

其中 μ_{ij} 表示数据点 x_i 属于聚类中心 v_j 的隶属度。

基于上述两个公式，目标函数可表示为

$$J_m = \sum_{i=1}^N \sum_{j=1}^c (\mu_{ij})^\tau d^2(x_i, v_j) \quad (29)$$

通过不断对隶属度矩阵和聚类中心不断迭代，直至达到如下要求为止，此时能够获初聚类的结果。

$$\left| \mu_{ij}^{(l+1)} - \mu_{ij}^{(l)} \right| \leq \varepsilon \quad (30)$$

其中 l 表示迭代次数， ε 表示允许的误差值。综上，FCM 的主要步骤如下：

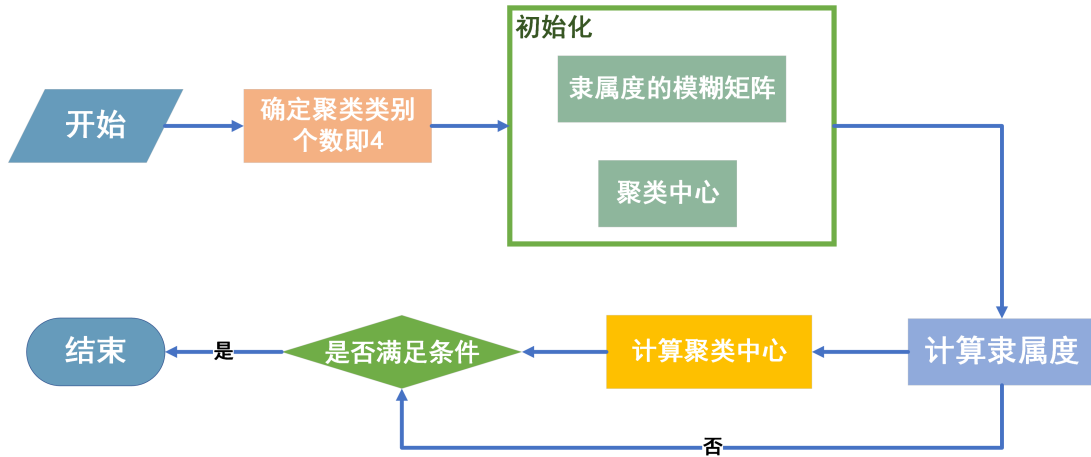


图 12 FCM 算法主要步骤

7.3 评判结果

可以根据聚类的结果绘制聚类散点图，如下所示。

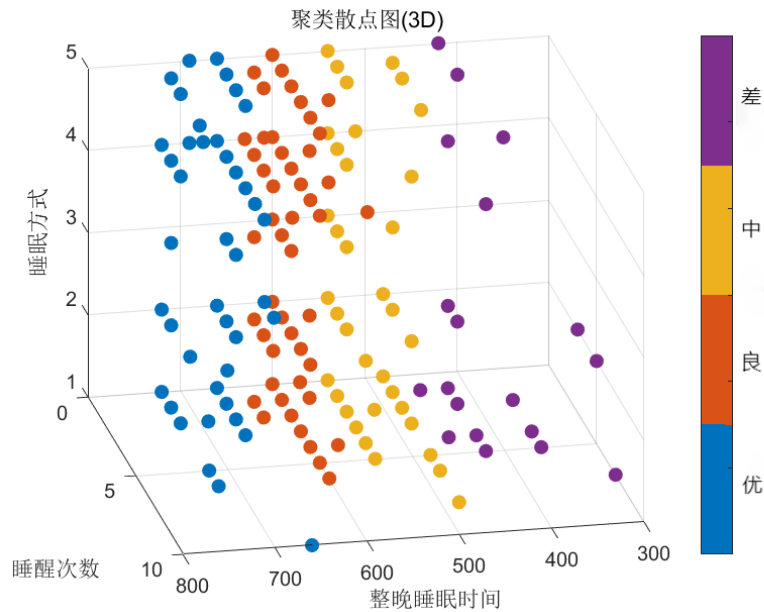


图 13 聚类散点图

我们根据聚类情况对 1-390 号婴儿进行划分优、良、中、差等级，频次统计结果如下图所示，详细结果见附件。

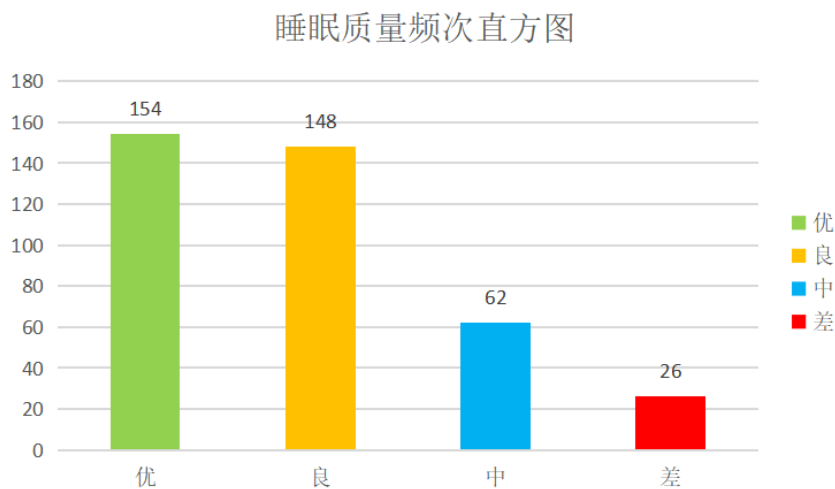


图 14 睡眠质量频次直方图

7.4 GBDT 综合睡眠质量预测

本文采用 GBDT 对最后 20 组婴儿的睡眠质量进行预测。GBDT 的具体原理在问题二中已详述，此处不再赘述。

将前 390 组母亲的心理指标和身体指标与 FCM 求解得出婴儿的睡眠质量评判结果作为训练集代入到 GBDT 中，得到训练好的决策树模型，输入最后 20 组母亲的心理指

标和身体指标，得到的输出即为每组婴儿的睡眠质量评级，并得到 GBDT 关于睡眠质量的分类预测函数 $G(x)$ 。

7.5 结果分析

通过 GBDT 算法求解得到 391-410 号婴儿的睡眠质量如下所示，共计 6 个优、12 个良、1 个中、1 个差。

表 4 婴儿睡眠质量预测结果表

编号	391	392	393	394	395	396	397	398	399	400
睡眠质量	良	良	良	优	良	良	良	良	良	优

编号	401	402	403	404	405	406	407	408	409	410
睡眠质量	优	良	良	优	良	差	优	中	优	良

7.6 模型检验

对于聚类模型，本文采用轮廓系数和 CH 指数对上述基于 FCM 的综合评判模型进行检验。轮廓系数（Silhouette Coefficient）、CH 指数（Calinski-Harabasz 指数）是评估聚类效果的常用指标。对于每一个数据点，轮廓系数是基于它与所属的那一类的其他数据点的相似度以及它与最近的另一类数据点的相似度来计算的。较大的 CH 指数通常意味着聚类效果较好，因为它意味着聚类之间的离散度较大，而聚类内部的离散度较小。最终计算得到的轮廓系数为 **0.6618**，CH 指数为 **1294.9**，聚类效果良好。其中 k 是聚类的数量, N 是数据点的数量, $B(k)$ 是各聚类中心到全局中心的总平方和, $W(k)$ 是每个数据点到其聚类中心的平方和之和。

对于预测模型，本文同样采用交叉验证对模型性能进行检验。本文对五个模型求解结果取平均值。最终计算得到模型在测试集的平均准确率达到 **86.59%**，结果具有可信性。

8 问题五模型的建立与求解

8.1 问题五分析

问题五是一个规划问题，建立以 238 号婴儿的睡眠质量提升至优为目标的治疗策略规划模型。

与问题三思路大体相似，首先利用一级策略即基于模拟退火的多元非线性规划进行求解，发现无法利用一级策略进行求解，我们该用二级策略进行求解。在前 390 组数据婴儿睡眠质量为优的样本中，寻找能使 238 号婴儿为达到各样本中母亲心理指标得分需要花费治疗费用最小的样本，将其作为从中转变为优最优治疗方案的参考。

问题五的思路图如下。

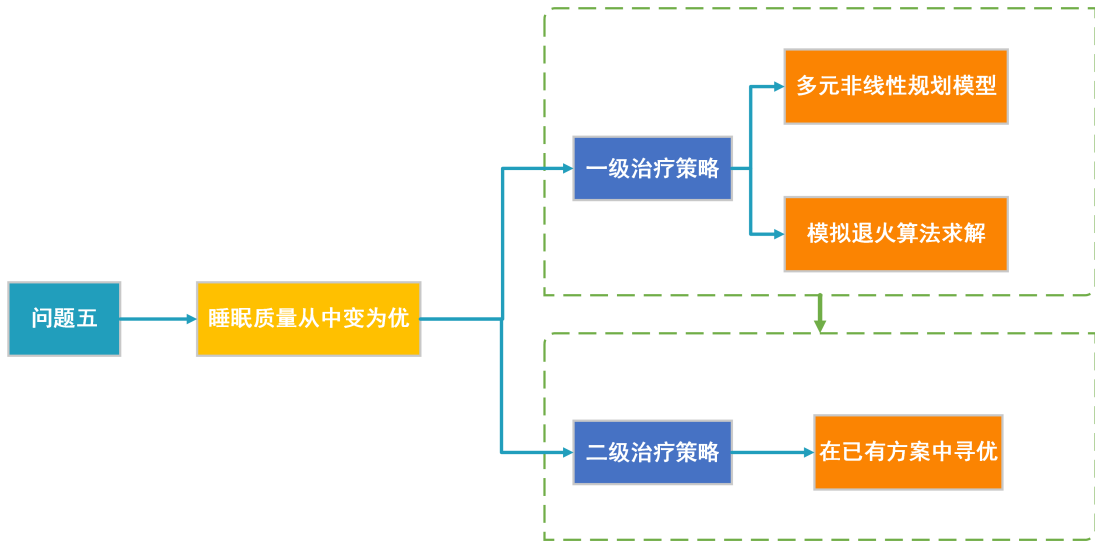


图 15 问题五思路图

8.2 一级策略——基于模拟退火的多元非线性规划模型

根据问题三结果可知 238 号婴儿的睡眠质量评级为中，现需要治疗后评级为优。与问题二类似本文首先采用一级策略，通过改善母亲的身体指标、心理指标来改变婴儿睡眠质量评级，依据此建立多元非线性规划模型，然后用模拟退火算法求解。

$$\min \Gamma = \sum_{i=1}^t f_i(m_i, l_i) \quad (31)$$

$$\begin{cases} P(y = e_0) = G(p_1, p_2, \dots, p_n, m_1, m_2, m_3) = \max \{P(y = k) | k = 1, 2, 3\}; \\ P'(y = e_1) = G(p_1, p_2, \dots, p_n, l_1, l_2, l_3) = \max \{P'(y = k) | k = 1, 2, 3\}; \\ l_i \leq m_i; \\ l_i \in [S_{\min}, S_{\max}]. \end{cases} \quad (32)$$

其中 Γ 表示睡眠质量改变，一级策略所需要花费的最小费用， e_0 表示治疗前 238 号婴儿的睡眠质量等级， e_1 表示治疗号 238 号婴儿的睡眠质量等级， l_i 表示 238 号母亲的诊疗后的心理指标得分， G 表示 GBDT 关于睡眠质量的预测函数。

8.2.1 求解及结果分析

利用一级策略进行求解，可以得到 238 号婴儿的睡眠质量评级从中提升到良的治疗策略，即将 CBTS 调整到 14.90，HADS 调整 17.16，所得到的最低治疗费用为 1289.09 元，考虑到治疗得分为整数，因此最终的治疗策略为 HADS 调整到 17，得到的最低费用为 **1332.50** 元，但是我们无法计算出从中提升到优的方案。

8.3 二级策略

当由一级策略无法达到目标时，采取二级策略寻优。将已知的前 410 组中每个睡眠质量为优的数据样本作为 238 号婴儿治疗后母亲心理指标可能的结果，通过计算达到每一个样本的三个心理指标得分所需的治疗费用，选出最小的治疗费用的方案，基于此本文建立多元非线性规划模型。

多元非线性模型综述如下

8.3.1 多元非线性规划模型

$$\min \Theta = \sum_{i=1}^t f_i(m_i, L_{ij}) \quad (i = 1, 2, \dots, w) \quad (33)$$

$$\begin{cases} P(y = b_0) = G(p_{1j}, p_{2j}, \dots, p_{nj}, m_{1j}, m_{2j}, m_{3j}) = \max \{P(y = k) | k = 1, 2, 3\}; \\ P'(y = b_2) = G(p_{1j}, p_{2j}, \dots, p_{nj}, l_{1j}, l_{2j}, l_{3j}) = \max \{P'(y = k) | k = 1, 2, 3\}; \\ L_{ij} = \min \{l_{ij}, m_{ij}\} \end{cases} \quad (34)$$

其中 Θ 表示睡眠质量改变，二级策略所需要花费的最小费用， L_{ij} 表示 238 号母亲的诊疗后的心理指标得分，选取方案和初始得分中的较小者。

8.3.2 求解结果及分析

通过上述多元非线性规划模型，本文求解得到 260 号为最佳方案，由此得出，为将 238 号婴儿的睡眠质量从中提升到优，制定的治疗策略为将 CBTS 心理指标得分由 15 改善到 13，其他两项心理指标 EPDS、HADS 不需要改变，并求解得到的最小治疗费用为 **1965.05** 元。

8.4 结果总结

根据结果我们也可以得出，在对母亲心理健康进行治疗时，婴儿的行为特征和睡眠质量都有可能得到一定的改善，在对 238 号母亲进行心理健康治疗时，采用将 CBTS 心理指标得分由 15 改善到 13、不改变其他两项心理指标的治疗策略，婴儿的行为特征从矛盾型改善为安静型，睡眠质量从中改善到优，这样的策略不仅能减少花费的费用，还能起到一举多得的功效。

9 模型总结与评价

9.1 模型总结

9.1.1 模型优点

1) 问题一采用基于 Apriori 算法的关联规则挖掘模型，可以有效挖掘稀疏或者小规模数据集因素间的潜在关系。

2) 问题二建立的基于贝叶斯优化算法改进的 GBDT 算法，能处理非线性数据，也可以灵活处理各种类型的数据，在小数据集情况下效果普遍优于其他模型。

3) 问题四采用的模糊均值聚类算法，处理非凸、噪声干扰和模糊数据，适用于复杂的数据分析任务，且可以对样本进行软分类，有利于更准确地对样本进行分类。

4) 问题三和问题五建立的模拟退火算法的治疗策略，算法耗时小，考虑较为全面，能为患者提供最优解，且优化效果良好。

9.1.2 模型缺点

1) 本文基于模拟退火算法的治疗策略存在只能制定患病程度逐级降低的治疗方案的缺陷，无法实现跨级治疗方案的制定。

2) 本文改进的 GBDT 算法在进行分类时候的准确率还有待提升。

9.1.3 模型改进

1) 针对本文基于模拟退火算法的治疗策略，可以考虑对算法进行参数调节，确保算法找到全局最优解。

2) 优于本数据集数据量较少，导致模型训练过程中容易产生过拟合的问题，后续可以通过扩充数据集对模型进行模型剪枝，防止过拟合。

9.2 模型推广

考虑到治疗的因素会同时影响婴儿行为特征和婴儿睡眠质量，本文基于问题三和问题五制定的治疗优化策略具有普适性，可以考虑结合起来进行推广，建立综合治疗策略，同时对婴儿这两方面进行治疗，帮助患者得到更优化的治疗策略，降低治疗费用。

参考文献

- [1] 杜军, 郭慧敏, 曾昭宇, 等. 基于 R 的 Apriori 算法在超长住院患者信息挖掘中的应用 [J]. 中国病案, 2022, 23(6):4.
- [2] 崔佳旭, 杨博. 贝叶斯优化方法和应用综述 [J]. 软件学报, 2018, 29(10):23.
- [3] 刘金元, 丁勇, 李涛. 基于梯度提升决策树的航班延误分类预测 [J]. 数学的实践与认识, 2018, 48(4):7.
- [4] 赖元文张杰. 基于模拟退火-自适应布谷鸟算法的城市公交调度优化研究 [J]. 交通运输系统工程与信息, 2021, 21(1):183-189.
- [5] 陈伟锋, 邵之江, CHEN Weifeng, 等. 基于收敛深度控制的多元混合非线性规划算法 [J]. 化工学报, 2014, 65(6):7.
- [6] 魏亚清, 夏少芳, 胡秀源. 海量船舶故障数据挖掘中的模糊运算聚类算法分析 [J]. 舰船科学技术, 2022, 44(13):4.
- [7] Shen H B , Yang J , Chou K C .Fuzzy KNN for predicting membrane protein types from pseudo-amino acid composition[J].Journal of Theoretical Biology, 2006, 240(1):9-13.
- [8] Ban X , Gao X Z , Huang X ,et al.Stability analysis of the simplest Takagi-Sugeno fuzzy control system using circle criterion[J].Information Sciences, 2007, 177(2):4387-4409.
- [9] Zhang Y, Beudaert X, Argandoña J, et al. A CPPS based on GBDT for predicting failure events in milling[J]. The International Journal of Advanced Manufacturing Technology, 2020, 111: 341-357.
- [10] El Moutaouakil K, Roudani M, El Ouissari A. Optimal Entropy Genetic Fuzzy-C-Means SMOTE (OEGFCM-SMOTE)[J]. Knowledge-Based Systems, 2023, 262: 110235.

附录 A 事项关联度统计表 (行为特征)

前项	后项	置信度	提升度
中年	安静型	0.337792642	1.097826087
低_CBTS	安静型	0.380952381	1.238095238
低_EPDS	安静型	0.380952381	1.238095238
低_HADS	安静型	0.405797101	1.31884058
自然分娩	安静型	0.309090909	1.004545455
低_CBTS, 中年	安静型	0.4	1.3
低_EPDS, 中年	安静型	0.396694215	1.289256198
低_HADS, 中年	安静型	0.447368421	1.453947368
已婚, 中年	安静型	0.328621908	1.068021201
妊娠时间正常, 中年	安静型	0.334645669	1.087598425
自然分娩, 中年	安静型	0.338983051	1.101694915
低_EPDS, 低_CBTS	安静型	0.409090909	1.329545455
低_CBTS, 低_HADS	安静型	0.435643564	1.415841584
低_CBTS, 已婚	安静型	0.375	1.21875
低_CBTS, 妊娠时间正常	安静型	0.353846154	1.15
自然分娩, 低_CBTS	安静型	0.376712329	1.224315068
低_EPDS, 低_HADS	安静型	0.439252336	1.427570093
低_EPDS, 已婚	安静型	0.37037037	1.203703704
低_EPDS, 妊娠时间正常	安静型	0.344	1.118
自然分娩, 低_EPDS	安静型	0.379310345	1.232758621
已婚, 低_HADS	安静型	0.404580153	1.314885496
妊娠时间正常, 低_HADS	安静型	0.394736842	1.282894737
自然分娩, 低_HADS	安静型	0.407407407	1.324074074
低_HADS, 低_CBTS, 中年	安静型	0.476190476	1.547619048
低_CBTS, 已婚, 中年	安静型	0.390909091	1.270454545
低_CBTS, 妊娠时间正常, 中年	安静型	0.367924528	1.195754717
自然分娩, 低_CBTS, 中年	安静型	0.394957983	1.283613445
低_EPDS, 中年, 低_HADS	安静型	0.483146067	1.570224719
低_EPDS, 已婚, 中年	安静型	0.381818182	1.240909091
自然分娩, 低_EPDS, 中年	安静型	0.391666667	1.272916667
低_HADS, 已婚, 中年	安静型	0.444444444	1.444444444

低_HADS, 妊娠时间正常, 中年	安静型	0.434782609	1.413043478
自然分娩, 低_HADS, 中年	安静型	0.446428571	1.450892857
妊娠时间正常, 已婚, 中年	安静型	0.32780083	1.065352697
自然分娩, 已婚, 中年	安静型	0.329749104	1.071684588
自然分娩, 妊娠时间正常, 中年	安静型	0.335968379	1.091897233
低_EPDS, 低_CBTS, 低_HADS	安静型	0.448275862	1.456896552
低_EPDS, 低_CBTS, 已婚	安静型	0.405940594	1.319306931
自然分娩, 低_EPDS, 低_CBTS	安静型	0.403669725	1.311926606
低_CBTS, 已婚, 低_HADS	安静型	0.431578947	1.402631579
自然分娩, 低_CBTS, 低_HADS	安静型	0.43	1.3975
低_CBTS, 妊娠时间正常, 已婚	安静型	0.344537815	1.119747899
自然分娩, 低_CBTS, 已婚	安静型	0.37037037	1.203703704
自然分娩, 低_CBTS, 妊娠时间正常	安静型	0.353846154	1.15
低_EPDS, 已婚, 低_HADS	安静型	0.44	1.43
自然分娩, 低_EPDS, 低_HADS	安静型	0.438095238	1.423809524
低_EPDS, 妊娠时间正常, 已婚	安静型	0.339130435	1.102173913
自然分娩, 低_EPDS, 已婚	安静型	0.368421053	1.197368421
自然分娩, 低_EPDS, 妊娠时间正常	安静型	0.346774194	1.127016129
妊娠时间正常, 已婚, 低_HADS	安静型	0.392523364	1.275700935
自然分娩, 已婚, 低_HADS	安静型	0.40625	1.3203125
自然分娩, 妊娠时间正常, 低_HADS	安静型	0.398230088	1.294247788
自然分娩, 中年, 低_CBTS, 低_HADS	安静型	0.469879518	1.527108434
自然分娩, 低_CBTS, 已婚, 中年	安静型	0.385321101	1.252293578
自然分娩, 低_CBTS, 妊娠时间正常, 中年	安静型	0.367924528	1.195754717
低_EPDS, 中年, 已婚, 低_HADS	安静型	0.481927711	1.56626506
自然分娩, 低_EPDS, 中年, 低_HADS	安静型	0.477272727	1.551136364
自然分娩, 低_EPDS, 已婚, 中年	安静型	0.376146789	1.222477064
自然分娩, 中年, 已婚, 低_HADS	安静型	0.443396226	1.441037736
自然分娩, 中年, 妊娠时间正常, 低_HADS	安静型	0.434782609	1.413043478
自然分娩, 妊娠时间正常, 已婚, 中年	安静型	0.329166667	1.069791667
自然分娩, 低_EPDS, 低_CBTS, 已婚	安静型	0.4	1.3
自然分娩, 低_CBTS, 已婚, 低_HADS	安静型	0.425531915	1.382978723
自然分娩, 低_CBTS, 妊娠时间正常, 已婚	安静型	0.344537815	1.119747899
自然分娩, 低_EPDS, 已婚, 低_HADS	安静型	0.43877551	1.426020408

自然分娩, 低 _EPDS, 妊娠时间正常, 已婚	安静型	0.342105263	1.111842105
自然分娩, 妊娠时间正常, 已婚, 低 _HADS	安静型	0.396226415	1.287735849
低 _HADS, 自然分娩, 低 _EPDS, 中年, 已婚	安静型	0.475609756	1.545731707
中 _CBTS	中等型	0.595588235	1.032352941
中 _EPDS	中等型	0.613138686	1.062773723
中 _HADS	中等型	0.613333333	1.063111111
已婚	中等型	0.577656676	1.001271571
研究生	中等型	0.597826087	1.036231884
高 _CBTS	中等型	0.607476636	1.052959502
高 _HADS	中等型	0.607843137	1.053594771
中 _EPDS, 中 _CBTS	中等型	0.6	1.04
中 _HADS, 中 _CBTS	中等型	0.615384615	1.066666667
中 _CBTS, 已婚	中等型	0.603053435	1.045292621
妊娠时间正常, 中 _CBTS	中等型	0.581196581	1.007407407
中 _CBTS, 研究生	中等型	0.588235294	1.019607843
自然分娩, 中 _CBTS	中等型	0.586466165	1.016541353
中 _EPDS, 中 _HADS	中等型	0.607142857	1.052380952
中 _EPDS, 中年	中等型	0.603773585	1.046540881
中 _EPDS, 已婚	中等型	0.619402985	1.073631841
中 _EPDS, 妊娠时间正常	中等型	0.593220339	1.028248588
中 _EPDS, 研究生	中等型	0.647058824	1.121568627
自然分娩, 中 _EPDS	中等型	0.607407407	1.052839506
中 _HADS, 中年	中等型	0.591304348	1.024927536
中 _HADS, 已婚	中等型	0.626760563	1.086384977
中 _HADS, 妊娠时间正常	中等型	0.614814815	1.065679012
中 _HADS, 研究生	中等型	0.639344262	1.108196721
自然分娩, 中 _HADS	中等型	0.608108108	1.054054054
大学, 中年	中等型	0.588235294	1.019607843
研究生, 中年	中等型	0.588235294	1.019607843
中年, 高 _CBTS	中等型	0.579710145	1.004830918
中年, 高 _HADS	中等型	0.614285714	1.064761905
低 _CBTS, 研究生	中等型	0.605633803	1.049765258
低 _EPDS, 研究生	中等型	0.579710145	1.004830918
已婚, 研究生	中等型	0.594285714	1.030095238

已婚, 高_CBTS	中等型	0.6	1.04
已婚, 高_HADS	中等型	0.595744681	1.032624113
妊娠时间正常, 研究生	中等型	0.592356688	1.026751592
妊娠时间正常, 高_CBTS	中等型	0.583333333	1.011111111
妊娠时间正常, 高_HADS	中等型	0.585365854	1.014634146
自然分娩, 研究生	中等型	0.593406593	1.028571429
自然分娩, 高_CBTS	中等型	0.603773585	1.046540881
自然分娩, 高_HADS	中等型	0.607843137	1.053594771
高_CBTS, 高_HADS	中等型	0.602739726	1.044748858
高_EPDS, 高_HADS	中等型	0.592105263	1.026315789
中_EPDS, 中_CBTS, 已婚	中等型	0.607594937	1.053164557
中_EPDS, 妊娠时间正常, 中_CBTS	中等型	0.577464789	1.000938967
自然分娩, 中_EPDS, 中_CBTS	中等型	0.594936709	1.031223629
中_HADS, 中_CBTS, 已婚	中等型	0.64	1.109333333
中_HADS, 妊娠时间正常, 中_CBTS	中等型	0.628571429	1.08952381
自然分娩, 中_HADS, 中_CBTS	中等型	0.61038961	1.058008658
妊娠时间正常, 中_CBTS, 已婚	中等型	0.578947368	1.003508772
中_CBTS, 已婚, 研究生	中等型	0.590909091	1.024242424
自然分娩, 中_CBTS, 已婚	中等型	0.59375	1.029166667
中_EPDS, 中年, 中_HADS	中等型	0.590909091	1.024242424
中_EPDS, 已婚, 中_HADS	中等型	0.617283951	1.069958848
中_EPDS, 妊娠时间正常, 中_HADS	中等型	0.597402597	1.035497835
自然分娩, 中_EPDS, 中_HADS	中等型	0.597560976	1.035772358
中_EPDS, 已婚, 中年	中等型	0.611650485	1.060194175
自然分娩, 中_EPDS, 中年	中等型	0.596153846	1.033333333
中_EPDS, 妊娠时间正常, 已婚	中等型	0.6	1.04
中_EPDS, 已婚, 研究生	中等型	0.656716418	1.138308458
自然分娩, 中_EPDS, 已婚	中等型	0.613636364	1.063636364
自然分娩, 中_EPDS, 妊娠时间正常	中等型	0.58974359	1.022222222
自然分娩, 中_EPDS, 研究生	中等型	0.647058824	1.121568627
中_HADS, 已婚, 中年	中等型	0.607476636	1.052959502
中_HADS, 妊娠时间正常, 中年	中等型	0.6	1.04
自然分娩, 中_HADS, 中年	中等型	0.584070796	1.012389381
中_HADS, 妊娠时间正常, 已婚	中等型	0.620155039	1.074935401

自然分娩, 中_HADS, 已婚	中等型	0.621428571	1.077142857
自然分娩, 中_HADS, 妊娠时间正常	中等型	0.611940299	1.060696517
自然分娩, 中_HADS, 研究生	中等型	0.639344262	1.108196721
自然分娩, 大学, 中年	中等型	0.582089552	1.008955224
中年, 已婚, 研究生	中等型	0.589041096	1.021004566
已婚, 中年, 高_CBTS	中等型	0.582089552	1.008955224
已婚, 中年, 高_HADS	中等型	0.617647059	1.070588235
自然分娩, 研究生, 中年	中等型	0.585526316	1.014912281
自然分娩, 中年, 高_HADS	中等型	0.614285714	1.064761905
低_CBTS, 研究生, 已婚	中等型	0.606060606	1.050505051
低_CBTS, 研究生, 妊娠时间正常	中等型	0.650793651	1.128042328
自然分娩, 低_CBTS, 研究生	中等型	0.605633803	1.049765258
妊娠时间正常, 已婚, 研究生	中等型	0.590604027	1.023713647
妊娠时间正常, 已婚, 高_CBTS	中等型	0.580246914	1.005761317
妊娠时间正常, 已婚, 高_HADS	中等型	0.576923077	1
妊娠时间正常, 已婚, 高中	中等型	0.582278481	1.0092827
自然分娩, 研究生, 已婚	中等型	0.589595376	1.021965318
自然分娩, 已婚, 高_CBTS	中等型	0.595959596	1.032996633
自然分娩, 已婚, 高_HADS	中等型	0.595744681	1.032624113
高_CBTS, 已婚, 高_HADS	中等型	0.590909091	1.024242424
自然分娩, 妊娠时间正常, 研究生	中等型	0.58974359	1.022222222
自然分娩, 妊娠时间正常, 高_CBTS	中等型	0.583333333	1.011111111
自然分娩, 妊娠时间正常, 高_HADS	中等型	0.585365854	1.014634146
自然分娩, 高_CBTS, 高_HADS	中等型	0.602739726	1.044748858
高_EPDS, 自然分娩, 高_HADS	中等型	0.592105263	1.026315789
中_EPDS, 妊娠时间正常, 中_CBTS, 已婚	中等型	0.585714286	1.015238095
自然分娩, 中_EPDS, 中_CBTS, 已婚	中等型	0.602564103	1.044444444
中_HADS, 妊娠时间正常, 中_CBTS, 已婚	中等型	0.637681159	1.10531401
自然分娩, 中_HADS, 中_CBTS, 已婚	中等型	0.635135135	1.100900901
自然分娩, 中_HADS, 妊娠时间正常, 中_CBTS	中等型	0.623188406	1.080193237
中_EPDS, 妊娠时间正常, 已婚, 中_HADS	中等型	0.608108108	1.054054054
自然分娩, 中_EPDS, 已婚, 中_HADS	中等型	0.607594937	1.053164557
自然分娩, 中_EPDS, 妊娠时间正常, 中_HADS	中等型	0.592105263	1.026315789
中_EPDS, 妊娠时间正常, 已婚, 中年	中等型	0.581395349	1.007751938

自然分娩, 中 _EPDS, 已婚, 中年	中等型	0.603960396	1.046864686
自然分娩, 中 _EPDS, 妊娠时间正常, 已婚	中等型	0.596491228	1.033918129
自然分娩, 中 _EPDS, 研究生, 已婚	中等型	0.656716418	1.138308458
中 _HADS, 妊娠时间正常, 已婚, 中年	中等型	0.606060606	1.050505051
自然分娩, 中 _HADS, 已婚, 中年	中等型	0.6	1.04
自然分娩, 中 _HADS, 妊娠时间正常, 中年	中等型	0.596153846	1.033333333
自然分娩, 中 _HADS, 妊娠时间正常, 已婚	中等型	0.6171875	1.069791667
自然分娩, 研究生, 中年, 已婚	中等型	0.586206897	1.016091954
自然分娩, 已婚, 中年, 高 _HADS	中等型	0.617647059	1.070588235
自然分娩, 低 _CBTS, 研究生, 已婚	中等型	0.606060606	1.050505051
自然分娩, 低 _CBTS, 妊娠时间正常, 研究生	中等型	0.650793651	1.128042328
自然分娩, 妊娠时间正常, 研究生, 已婚	中等型	0.587837838	1.018918919
自然分娩, 妊娠时间正常, 已婚, 高 _CBTS	中等型	0.580246914	1.005761317
自然分娩, 妊娠时间正常, 已婚, 高 _HADS	中等型	0.576923077	1
自然分娩, 妊娠时间正常, 已婚, 高中	中等型	0.576923077	1
自然分娩, 高 _CBTS, 已婚, 高 _HADS	中等型	0.590909091	1.024242424
中 _EPDS, 妊娠时间正常, 中 _CBTS, 自然分娩, 已婚	中等型	0.579710145	1.004830918
妊娠时间正常, 中 _CBTS, 自然分娩, 中 _HADS, 已婚	中等型	0.632352941	1.096078431
中 _EPDS, 妊娠时间正常, 自然分娩, 中 _HADS, 已婚	中等型	0.602739726	1.044748858
妊娠时间正常, 自然分娩, 中 _HADS, 中年, 已婚	中等型	0.602040816	1.043537415
已婚	矛盾型	0.117166213	1.015440509
妊娠时间正常	矛盾型	0.120845921	1.047331319
自然分娩	矛盾型	0.116883117	1.012987013
妊娠时间正常, 已婚	矛盾型	0.124203822	1.076433121
自然分娩, 已婚	矛盾型	0.11878453	1.02946593
自然分娩, 妊娠时间正常	矛盾型	0.121580547	1.053698075
自然分娩, 妊娠时间正常, 已婚	矛盾型	0.125	1.083333333

附录 B 事项关联度统计表 (睡眠时间)

前项	后项	置信度	提升度
中 _EPDS	600	0.328467153	1.197216727
中 _HADS	600	0.306666667	1.117757009

中年	600	0.280936455	1.023973994
已婚	600	0.280653951	1.022944307
妊娠时间正常	600	0.290030211	1.057119462
研究生	600	0.331521739	1.208350264
中_EPDS, 已婚	600	0.335820896	1.224020086
中_EPDS, 妊娠时间正常	600	0.347457627	1.266434342
自然分娩, 中_EPDS	600	0.325925926	1.187954309
中_HADS, 已婚	600	0.316901408	1.155061208
中_HADS, 妊娠时间正常	600	0.333333333	1.214953271
自然分娩, 中_HADS	600	0.304054054	1.108234403
已婚, 中年	600	0.286219081	1.043228427
妊娠时间正常, 中年	600	0.295275591	1.076238134
研究生, 中年	600	0.339869281	1.238775884
自然分娩, 中年	600	0.277966102	1.013147473
妊娠时间正常, 已婚	600	0.296178344	1.079528543
研究生, 已婚	600	0.337142857	1.228838451
自然分娩, 已婚	600	0.279005525	1.016936025
妊娠时间正常, 研究生	600	0.350318471	1.276861718
自然分娩, 妊娠时间正常	600	0.288753799	1.052467119
自然分娩, 研究生	600	0.32967033	1.201602136
中_EPDS, 妊娠时间正常, 已婚	600	0.356521739	1.299471759
自然分娩, 中_EPDS, 已婚	600	0.333333333	1.214953271
自然分娩, 中_EPDS, 妊娠时间正常	600	0.341880342	1.246105919
中_HADS, 妊娠时间正常, 已婚	600	0.341085271	1.243207998
自然分娩, 中_HADS, 已婚	600	0.314285714	1.14552737
自然分娩, 中_HADS, 妊娠时间正常	600	0.328358209	1.19681964
妊娠时间正常, 已婚, 中年	600	0.298755187	1.088920774
已婚, 研究生, 中年	600	0.342465753	1.248239662
自然分娩, 已婚, 中年	600	0.283154122	1.03205708
妊娠时间正常, 研究生, 中年	600	0.356589147	1.299717453
自然分娩, 妊娠时间正常, 中年	600	0.292490119	1.066085479
自然分娩, 研究生, 中年	600	0.335526316	1.222946385
妊娠时间正常, 研究生, 已婚	600	0.355704698	1.296493759
自然分娩, 妊娠时间正常, 已婚	600	0.294871795	1.074766355

自然分娩, 研究生, 已婚	600	0.335260116	1.221976122
自然分娩, 妊娠时间正常, 研究生	600	0.352564103	1.285046729
自然分娩, 中 _EPDS, 妊娠时间正常, 已婚	600	0.350877193	1.27889818
自然分娩, 中 _HADS, 妊娠时间正常, 已婚	600	0.3359375	1.224445093
已婚, 妊娠时间正常, 研究生, 中年	600	0.357723577	1.303852291
自然分娩, 妊娠时间正常, 已婚, 中年	600	0.295833333	1.078271028
自然分娩, 研究生, 中年, 已婚	600	0.337931034	1.231711247
自然分娩, 妊娠时间正常, 研究生, 中年	600	0.356589147	1.299717453
自然分娩, 妊娠时间正常, 研究生, 已婚	600	0.358108108	1.305253852
妊娠时间正常, 研究生, 自然分娩, 中年, 已婚	600	0.357723577	1.303852291
中年	660	0.20735786	1.036789298
自然分娩	660	0.202597403	1.012987013
已婚, 中年	660	0.201413428	1.007067138
妊娠时间正常, 中年	660	0.208661417	1.043307087
自然分娩, 中年	660	0.210169492	1.050847458
自然分娩, 妊娠时间正常	660	0.200607903	1.003039514
妊娠时间正常, 已婚, 中年	660	0.203319502	1.01659751
自然分娩, 已婚, 中年	660	0.204301075	1.021505376
自然分娩, 妊娠时间正常, 中年	660	0.209486166	1.04743083
自然分娩, 妊娠时间正常, 已婚, 中年	660	0.204166667	1.020833333
已婚	720	0.177111717	1.030948798
自然分娩, 已婚	720	0.174033149	1.013028779

附录 C 事项关联度统计表 (睡醒次数)

前项	后项	置信度	提升度
低 _CBTS	0	0.414965986	1.274304998
低 _EPDS	0	0.387755102	1.190744014
低 _HADS	0	0.376811594	1.157137966
已婚	0	0.329700272	1.012465404
低 _CBTS, 中年	0	0.4	1.228346457
低 _EPDS, 中年	0	0.380165289	1.167436715
低 _HADS, 中年	0	0.377192982	1.158309159

低_EPDS, 低_CBTS	0	0.409090909	1.256263422
低_CBTS, 低_HADS	0	0.425742574	1.307398456
低_CBTS, 已婚	0	0.426470588	1.30963409
低_CBTS, 妊娠时间正常	0	0.4	1.228346457
自然分娩, 低_CBTS	0	0.410958904	1.261999784
低_EPDS, 低_HADS	0	0.392523364	1.20538671
低_EPDS, 已婚	0	0.392592593	1.2055993
低_EPDS, 妊娠时间正常	0	0.368	1.13007874
自然分娩, 低_EPDS	0	0.386206897	1.185989682
已婚, 低_HADS	0	0.396946565	1.218969766
妊娠时间正常, 低_HADS	0	0.368421053	1.131371736
自然分娩, 低_HADS	0	0.377777778	1.160104987
自然分娩, 已婚	0	0.328729282	1.009483621
低_CBTS, 已婚, 中年	0	0.409090909	1.256263422
低_CBTS, 中年, 妊娠时间正常	0	0.396226415	1.216758283
自然分娩, 低_CBTS, 中年	0	0.394957983	1.212863098
低_EPDS, 已婚, 中年	0	0.381818182	1.172512527
自然分娩, 低_EPDS, 中年	0	0.375	1.151574803
低_HADS, 已婚, 中年	0	0.398148148	1.222659668
自然分娩, 低_HADS, 中年	0	0.375	1.151574803
低_EPDS, 低_CBTS, 已婚	0	0.425742574	1.307398456
自然分娩, 低_EPDS, 低_CBTS	0	0.403669725	1.23961569
低_CBTS, 已婚, 低_HADS	0	0.452631579	1.38997099
自然分娩, 低_CBTS, 低_HADS	0	0.42	1.28976378
低_CBTS, 已婚, 妊娠时间正常	0	0.411764706	1.264474294
自然分娩, 低_CBTS, 已婚	0	0.422222222	1.296587927
自然分娩, 低_CBTS, 妊娠时间正常	0	0.4	1.228346457
低_EPDS, 已婚, 低_HADS	0	0.42	1.28976378
自然分娩, 低_EPDS, 低_HADS	0	0.39047619	1.199100112
低_EPDS, 妊娠时间正常, 已婚	0	0.382608696	1.174940089
自然分娩, 低_EPDS, 已婚	0	0.390977444	1.200639394
自然分娩, 低_EPDS, 妊娠时间正常	0	0.370967742	1.139192278
妊娠时间正常, 已婚, 低_HADS	0	0.392523364	1.20538671
自然分娩, 已婚, 低_HADS	0	0.3984375	1.223548228

自然分娩, 妊娠时间正常, 低_HADS	0	0.371681416	1.141383876
低_CBTS, 妊娠时间正常, 已婚, 中年	0	0.40625	1.24753937
自然分娩, 低_CBTS, 已婚, 中年	0	0.403669725	1.23961569
自然分娩, 低_CBTS, 妊娠时间正常, 中年	0	0.396226415	1.216758283
自然分娩, 低_EPDS, 已婚, 中年	0	0.376146789	1.155096439
自然分娩, 中年, 已婚, 低_HADS	0	0.396226415	1.216758283
自然分娩, 低_EPDS, 低_CBTS, 已婚	0	0.42	1.28976378
自然分娩, 低_CBTS, 已婚, 低_HADS	0	0.446808511	1.372089127
自然分娩, 低_CBTS, 妊娠时间正常, 已婚	0	0.411764706	1.264474294
自然分娩, 低_EPDS, 已婚, 低_HADS	0	0.418367347	1.284750121
自然分娩, 低_EPDS, 妊娠时间正常, 已婚	0	0.385964912	1.185246581
自然分娩, 妊娠时间正常, 已婚, 低_HADS	0	0.396226415	1.216758283
妊娠时间正常, 自然分娩, 低_CBTS, 中年, 已婚	0	0.40625	1.24753937
中_CBTS	1	0.330882353	1.11244929
中_EPDS	1	0.313868613	1.055247923
中_HADS	1	0.313333333	1.053448276
妊娠时间正常	1	0.311178248	1.046202729
中_CBTS, 已婚	1	0.328244275	1.103579889
自然分娩, 中_CBTS	1	0.315789474	1.061705989
中_EPDS, 已婚	1	0.305970149	1.028692743
自然分娩, 中_EPDS	1	0.311111111	1.045977011
中_HADS, 已婚	1	0.309859155	1.041767848
中_HADS, 妊娠时间正常	1	0.318518519	1.070881226
自然分娩, 中_HADS	1	0.310810811	1.044967381
妊娠时间正常, 中年	1	0.299212598	1.005973391
妊娠时间正常, 已婚	1	0.299363057	1.006479244
自然分娩, 妊娠时间正常	1	0.306990881	1.032124515
自然分娩, 中_CBTS, 已婚	1	0.3125	1.050646552
自然分娩, 中_EPDS, 已婚	1	0.303030303	1.018808777
中_HADS, 妊娠时间正常, 已婚	1	0.310077519	1.042502005
自然分娩, 中_HADS, 已婚	1	0.307142857	1.032635468
自然分娩, 中_HADS, 妊娠时间正常	1	0.313432836	1.05378281
自然分娩, 中_HADS, 妊娠时间正常, 已婚	1	0.3046875	1.024380388
中年	2	0.193979933	1.065523576

已婚	2	0.188010899	1.032735925
研究生	2	0.217391304	1.194121249
自然分娩	2	0.184415584	1.012987013
已婚, 中年	2	0.201413428	1.106355447
妊娠时间正常, 中年	2	0.192913386	1.059665077
自然分娩, 中年	2	0.196610169	1.079971354
妊娠时间正常, 已婚	2	0.184713376	1.014622768
研究生, 已婚	2	0.222857143	1.224144869
自然分娩, 已婚	2	0.190607735	1.047000233
自然分娩, 研究生	2	0.21978022	1.207243461
妊娠时间正常, 已婚, 中年	2	0.199170124	1.094033078
自然分娩, 已婚, 中年	2	0.204301075	1.122217174
自然分娩, 妊娠时间正常, 中年	2	0.193675889	1.063853477
自然分娩, 妊娠时间正常, 已婚	2	0.185897436	1.021126761
自然分娩, 研究生, 已婚	2	0.225433526	1.238296833
自然分娩, 妊娠时间正常, 已婚, 中年	2	0.2	1.098591549

附录 D 事项关联度统计表 (入睡方式)

前项	后项	置信度	提升度
已婚	定时法	0.117166213	1.015440509
自然分娩, 已婚	定时法	0.116022099	1.005524862
已婚	抚触法	0.177111717	1.001066224
研究生	抚触法	0.233695652	1.320888469
研究生, 已婚	抚触法	0.234285714	1.324223602
自然分娩, 研究生	抚触法	0.230769231	1.304347826
自然分娩, 研究生, 已婚	抚触法	0.231213873	1.30686102
中年	哄睡法	0.230769231	1.011235955
妊娠时间正常	哄睡法	0.238670695	1.045860348
研究生	哄睡法	0.233695652	1.024059599
自然分娩	哄睡法	0.231168831	1.012987013
妊娠时间正常, 中年	哄睡法	0.24015748	1.052375476
自然分娩, 中年	哄睡法	0.233898305	1.024947629

妊娠时间正常, 已婚	哄睡法	0.238853503	1.046661418
自然分娩, 已婚	哄睡法	0.229281768	1.00471786
妊娠时间正常, 研究生	哄睡法	0.248407643	1.088527875
自然分娩, 妊娠时间正常	哄睡法	0.240121581	1.052218162
自然分娩, 研究生	哄睡法	0.236263736	1.035313002
妊娠时间正常, 已婚, 中年	哄睡法	0.236514523	1.036411954
自然分娩, 妊娠时间正常, 中年	哄睡法	0.241106719	1.056535062
自然分娩, 妊娠时间正常, 已婚	哄睡法	0.240384615	1.053370787
自然分娩, 妊娠时间正常, 研究生	哄睡法	0.25	1.095505618
自然分娩, 妊娠时间正常, 已婚, 中年	哄睡法	0.2375	1.040730337
中_EPDS	环境营造法	0.445255474	1.039818174
中年	环境营造法	0.451505017	1.054412913
低_CBTS	环境营造法	0.489795918	1.143834779
低_EPDS	环境营造法	0.462585034	1.080288403
低_HADS	环境营造法	0.442028986	1.03228326
大学	环境营造法	0.552941176	1.291299753
高中	环境营造法	0.444444444	1.037924152
中_EPDS, 中年	环境营造法	0.462264151	1.079539035
中_EPDS, 已婚	环境营造法	0.447761194	1.045669854
中_EPDS, 妊娠时间正常	环境营造法	0.43220339	1.009337258
自然分娩, 中_EPDS	环境营造法	0.437037037	1.020625416
中_HADS, 中年	环境营造法	0.469565217	1.09658943
低_CBTS, 中年	环境营造法	0.483333333	1.128742515
低_EPDS, 中年	环境营造法	0.47107438	1.100113822
中年, 低_HADS	环境营造法	0.447368421	1.0447526
大学, 中年	环境营造法	0.588235294	1.373723142
已婚, 中年	环境营造法	0.45229682	1.056262034
妊娠时间正常, 中年	环境营造法	0.44488189	1.038945731
自然分娩, 中年	环境营造法	0.447457627	1.044960926
低_EPDS, 低_CBTS	环境营造法	0.481818182	1.125204137
低_CBTS, 低_HADS	环境营造法	0.465346535	1.086737416
低_CBTS, 已婚	环境营造法	0.485294118	1.133321592
低_CBTS, 妊娠时间正常	环境营造法	0.484615385	1.131736527
自然分娩, 低_CBTS	环境营造法	0.48630137	1.135673858

低_EPDS, 低_HADS	环境营造法	0.457943925	1.069449885
低_EPDS, 已婚	环境营造法	0.459259259	1.072521623
低_EPDS, 妊娠时间正常	环境营造法	0.448	1.046227545
自然分娩, 低_EPDS	环境营造法	0.462068966	1.079083213
已婚, 低_HADS	环境营造法	0.442748092	1.033962609
妊娠时间正常, 低_HADS	环境营造法	0.429824561	1.00378191
自然分娩, 低_HADS	环境营造法	0.444444444	1.037924152
大学, 已婚	环境营造法	0.548780488	1.281583175
大学, 妊娠时间正常	环境营造法	0.541666667	1.26497006
自然分娩, 大学	环境营造法	0.547619048	1.27887083
已婚, 高中	环境营造法	0.428571429	1.000855432
自然分娩, 高中	环境营造法	0.432989691	1.011173529
中_EPDS, 已婚, 中年	环境营造法	0.466019417	1.088308819
中_EPDS, 妊娠时间正常, 中年	环境营造法	0.449438202	1.049586221
自然分娩, 中_EPDS, 中年	环境营造法	0.451923077	1.055389222
中_EPDS, 妊娠时间正常, 已婚	环境营造法	0.434782609	1.015360583
自然分娩, 中_EPDS, 已婚	环境营造法	0.439393939	1.026129559
中_HADS, 已婚, 中年	环境营造法	0.46728972	1.091275393
中_HADS, 妊娠时间正常, 中年	环境营造法	0.457142857	1.067579127
自然分娩, 中_HADS, 中年	环境营造法	0.460176991	1.07466483
低_EPDS, 低_CBTS, 中年	环境营造法	0.494382022	1.154544843
中年, 低_CBTS, 低_HADS	环境营造法	0.464285714	1.084260051
低_CBTS, 已婚, 中年	环境营造法	0.481818182	1.125204137
低_CBTS, 中年, 妊娠时间正常	环境营造法	0.481132075	1.123601853
自然分娩, 低_CBTS, 中年	环境营造法	0.478991597	1.11860313
低_EPDS, 中年, 低_HADS	环境营造法	0.471910112	1.102065532
低_EPDS, 已婚, 中年	环境营造法	0.472727273	1.10397387
低_EPDS, 妊娠时间正常, 中年	环境营造法	0.460784314	1.076083128
自然分娩, 低_EPDS, 中年	环境营造法	0.466666667	1.089820359
中年, 已婚, 低_HADS	环境营造法	0.453703704	1.059547572
中年, 妊娠时间正常, 低_HADS	环境营造法	0.434782609	1.015360583
自然分娩, 中年, 低_HADS	环境营造法	0.446428571	1.042557742
大学, 已婚, 中年	环境营造法	0.590909091	1.379967338
自然分娩, 大学, 中年	环境营造法	0.582089552	1.359370811

妊娠时间正常, 已婚, 中年	环境营造法	0.443983402	1.036847467
自然分娩, 已婚, 中年	环境营造法	0.448028674	1.046294508
自然分娩, 妊娠时间正常, 中年	环境营造法	0.442687747	1.033821685
低_EPDS, 低_CBTS, 低_HADS	环境营造法	0.471264368	1.100557506
低_EPDS, 低_CBTS, 已婚	环境营造法	0.475247525	1.109859489
低_EPDS, 低_CBTS, 妊娠时间正常	环境营造法	0.473684211	1.106208635
自然分娩, 低_EPDS, 低_CBTS	环境营造法	0.47706422	1.114102071
低_CBTS, 已婚, 低_HADS	环境营造法	0.463157895	1.081626221
低_CBTS, 低_HADS, 妊娠时间正常	环境营造法	0.465116279	1.086199694
自然分娩, 低_CBTS, 低_HADS	环境营造法	0.46	1.074251497
低_CBTS, 已婚, 妊娠时间正常	环境营造法	0.478991597	1.11860313
自然分娩, 低_CBTS, 已婚	环境营造法	0.481481481	1.124417831
自然分娩, 低_CBTS, 妊娠时间正常	环境营造法	0.484615385	1.131736527
低_EPDS, 已婚, 低_HADS	环境营造法	0.46	1.074251497
低_EPDS, 妊娠时间正常, 低_HADS	环境营造法	0.449438202	1.049586221
自然分娩, 低_EPDS, 低_HADS	环境营造法	0.457142857	1.067579127
低_EPDS, 妊娠时间正常, 已婚	环境营造法	0.443478261	1.035667795
自然分娩, 低_EPDS, 已婚	环境营造法	0.458646617	1.071090901
自然分娩, 低_EPDS, 妊娠时间正常	环境营造法	0.451612903	1.054664864
妊娠时间正常, 已婚, 低_HADS	环境营造法	0.429906542	1.003973362
自然分娩, 已婚, 低_HADS	环境营造法	0.4453125	1.039951347
自然分娩, 妊娠时间正常, 低_HADS	环境营造法	0.433628319	1.012664936
自然分娩, 大学, 已婚	环境营造法	0.543209877	1.268573963
自然分娩, 大学, 妊娠时间正常	环境营造法	0.541666667	1.26497006
中_EPDS, 妊娠时间正常, 已婚, 中年	环境营造法	0.453488372	1.059044701
自然分娩, 中_EPDS, 已婚, 中年	环境营造法	0.455445545	1.063615344
自然分娩, 中_EPDS, 妊娠时间正常, 中年	环境营造法	0.443181818	1.034975504
自然分娩, 中_EPDS, 妊娠时间正常, 已婚	环境营造法	0.429824561	1.00378191
中_HADS, 妊娠时间正常, 已婚, 中年	环境营造法	0.454545455	1.061513337
自然分娩, 中_HADS, 已婚, 中年	环境营造法	0.457142857	1.067579127
自然分娩, 中_HADS, 妊娠时间正常, 中年	环境营造法	0.451923077	1.055389222
低_EPDS, 低_CBTS, 已婚, 中年	环境营造法	0.49382716	1.153249057
自然分娩, 低_EPDS, 低_CBTS, 中年	环境营造法	0.488636364	1.141126837
低_CBTS, 妊娠时间正常, 已婚, 中年	环境营造法	0.479166667	1.119011976

自然分娩, 低_CBTS, 已婚, 中年	环境营造法	0.47706422	1.114102071
自然分娩, 低_CBTS, 妊娠时间正常, 中年	环境营造法	0.481132075	1.123601853
低_EPDS, 中年, 已婚, 低_HADS	环境营造法	0.481927711	1.125459924
自然分娩, 低_EPDS, 中年, 低_HADS	环境营造法	0.465909091	1.08805117
低_EPDS, 妊娠时间正常, 已婚, 中年	环境营造法	0.462365591	1.079775932
自然分娩, 低_EPDS, 已婚, 中年	环境营造法	0.467889908	1.092677031
自然分娩, 低_EPDS, 妊娠时间正常, 中年	环境营造法	0.460784314	1.076083128
自然分娩, 中年, 已婚, 低_HADS	环境营造法	0.452830189	1.057507626
自然分娩, 中年, 妊娠时间正常, 低_HADS	环境营造法	0.434782609	1.015360583
自然分娩, 妊娠时间正常, 已婚, 中年	环境营造法	0.441666667	1.031437126
自然分娩, 低_EPDS, 低_CBTS, 低_HADS	环境营造法	0.465116279	1.086199694
低_EPDS, 妊娠时间正常, 已婚, 低_CBTS	环境营造法	0.465116279	1.086199694
自然分娩, 低_EPDS, 低_CBTS, 已婚	环境营造法	0.47	1.09760479
自然分娩, 低_EPDS, 妊娠时间正常, 低_CBTS	环境营造法	0.473684211	1.106208635
自然分娩, 低_CBTS, 已婚, 低_HADS	环境营造法	0.457446809	1.068288954
自然分娩, 低_CBTS, 妊娠时间正常, 低_HADS	环境营造法	0.465116279	1.086199694
自然分娩, 低_CBTS, 妊娠时间正常, 已婚	环境营造法	0.478991597	1.11860313
自然分娩, 低_EPDS, 已婚, 低_HADS	环境营造法	0.459183673	1.072345106
自然分娩, 低_EPDS, 妊娠时间正常, 低_HADS	环境营造法	0.454545455	1.061513337
自然分娩, 低_EPDS, 妊娠时间正常, 已婚	环境营造法	0.447368421	1.0447526
自然分娩, 妊娠时间正常, 已婚, 低_HADS	环境营造法	0.433962264	1.013444808
妊娠时间正常, 自然分娩, 中_HADS, 中年, 已婚	环境营造法	0.448979592	1.048515214
自然分娩, 低_EPDS, 低_CBTS, 中年, 已婚	环境营造法	0.4875	1.138473054
妊娠时间正常, 自然分娩, 低_CBTS, 中年, 已婚	环境营造法	0.479166667	1.119011976
低_HADS, 自然分娩, 低_EPDS, 中年, 已婚	环境营造法	0.475609756	1.110705418
妊娠时间正常, 自然分娩, 低_EPDS, 中年, 已婚	环境营造法	0.462365591	1.079775932
妊娠时间正常, 自然分娩, 低_EPDS, 低_CBTS, 已婚	环境营造法	0.465116279	1.086199694

附录 E 1-390 号婴儿睡眠质量信息表

编号	睡眠质量	编号	睡眠质量	编号	睡眠质量
1	良	131	差	261	差
2	优	132	良	262	优
3	优	133	差	263	良
4	优	134	优	264	中
5	良	135	优	265	优
6	优	136	中	266	优
7	良	137	良	267	良
8	良	138	良	268	优
9	良	139	优	269	良
10	优	140	中	270	良
11	优	141	良	271	中
12	良	142	良	272	中
13	优	143	优	273	良
14	优	144	良	274	良
15	良	145	良	275	良
16	优	146	良	276	中
17	优	147	优	277	优
18	中	148	优	278	良
19	优	149	良	279	中
20	差	150	优	280	中
21	优	151	良	281	优
22	差	152	优	282	差
23	中	153	差	283	良
24	优	154	优	284	良
25	优	155	良	285	良
26	优	156	优	286	良
27	良	157	良	287	良
28	良	158	优	288	优
29	良	159	良	289	优
30	优	160	良	290	良
31	优	161	良	291	良

32	良	162	良	292	差
33	良	163	优	293	优
34	优	164	良	294	中
35	优	165	中	295	良
36	中	166	良	296	优
37	优	167	良	297	优
38	中	168	良	298	优
39	差	169	良	299	优
40	优	170	良	300	中
41	优	171	良	301	优
42	良	172	良	302	良
43	良	173	良	303	中
44	良	174	优	304	良
45	良	175	优	305	优
46	优	176	优	306	中
47	良	177	差	307	良
48	优	178	优	308	中
49	优	179	优	309	良
50	差	180	良	310	中
51	良	181	优	311	优
52	良	182	良	312	良
53	良	183	中	313	良
54	优	184	良	314	良
55	差	185	优	315	中
56	良	186	差	316	良
57	良	187	良	317	优
58	优	188	优	318	优
59	中	189	良	319	优
60	优	190	优	320	良
61	良	191	中	321	差
62	良	192	良	322	良
63	良	193	优	323	优
64	优	194	良	324	中
65	良	195	良	325	优

66	优	196	优	326	差
67	中	197	差	327	优
68	中	198	优	328	优
69	优	199	良	329	良
70	良	200	中	330	差
71	优	201	良	331	优
72	中	202	良	332	优
73	良	203	优	333	优
74	优	204	优	334	优
75	优	205	良	335	良
76	优	206	中	336	良
77	中	207	良	337	良
78	优	208	优	338	优
79	优	209	优	339	中
80	优	210	优	340	良
81	优	211	优	341	优
82	优	212	优	342	优
83	优	213	中	343	中
84	良	214	优	344	中
85	优	215	良	345	良
86	优	216	优	346	优
87	优	217	良	347	中
88	优	218	中	348	良
89	优	219	优	349	优
90	优	220	优	350	优
91	优	221	中	351	中
92	中	222	良	352	优
93	良	223	中	353	良
94	良	224	良	354	良
95	差	225	中	355	良
96	中	226	优	356	良
97	优	227	优	357	优
98	优	228	差	358	优
99	差	229	良	359	优

100	优	230	中	360	中
101	良	231	中	361	良
102	良	232	良	362	优
103	中	233	优	363	优
104	中	234	良	364	差
105	优	235	差	365	中
106	良	236	良	366	良
107	良	237	良	367	良
108	良	238	中	368	差
109	良	239	优	369	优
110	中	240	优	370	中
111	良	241	中	371	优
112	优	242	优	372	优
113	良	243	良	373	良
114	优	244	良	374	优
115	良	245	良	375	差
116	优	246	优	376	优
117	优	247	良	377	中
118	良	248	中	378	优
119	中	249	良	379	良
120	优	250	优	380	差
121	良	251	优	381	中
122	良	252	优	382	良
123	优	253	良	383	优
124	良	254	良	384	中
125	中	255	良	385	良
126	优	256	良	386	中
127	优	257	良	387	优
128	良	258	中	388	良
129	优	259	中	389	差
130	良	260	优	390	中

附录 F 问题一 Python 源代码

```
from mlxtend.frequent_patterns import apriori, association_rules
from mlxtend.preprocessing import TransactionEncoder

# Load the data
data = pd.read_csv("E:/Desktop/mnt/附件.csv", encoding="gbk")

# Discretize continuous variables
data['母亲年龄段'] = pd.cut(data['母亲年龄'], bins=[0, 25, 35, 100], labels=['年轻', '中年',
    '年纪较大'])
data['妊娠时间段'] = pd.cut(data['妊娠时间 (周数)'], bins=[0, 36, 41, 100], labels=['较短',
    '正常', '较长'])

for column in ['CBTS']:
    mean_val = data[column].mean()
    std_val = data[column].std()
    data[f'{column}_段'] = pd.cut(data[column],
        bins=[-float('inf'), mean_val-0.5*std_val, mean_val+0.5*std_val,
            float('inf')],
        labels=['低_CBTS', '中_CBTS', '高_CBTS'])

for column in ['EPDS']:
    mean_val = data[column].mean()
    std_val = data[column].std()
    data[f'{column}_段'] = pd.cut(data[column],
        bins=[-float('inf'), mean_val-0.5*std_val, mean_val+0.5*std_val,
            float('inf')],
        labels=['低_EPDS', '中_EPDS', '高_EPDS'])

for column in ['HADS']:
    mean_val = data[column].mean()
    std_val = data[column].std()
    data[f'{column}_段'] = pd.cut(data[column],
        bins=[-float('inf'), mean_val-0.5*std_val, mean_val+0.5*std_val,
            float('inf')],
        labels=['低_HADS', '中_HADS', '高_HADS'])

# Prepare data for Apriori
columns_to_use = ['母亲年龄段', '婚姻状况', '教育程度', '妊娠时间段', '分娩方式', 'CBTS_段',
    'EPDS_段', 'HADS_段', '婴儿行为特征']
for col in columns_to_use:
    data[col] = data[col].astype(str)
transactions = data[columns_to_use].values.tolist()

# Convert data to the right format for Apriori
te = TransactionEncoder()
```

```

te_ary = te.fit(transactions).transform(transactions)
df_apriori = pd.DataFrame(te_ary, columns=te.columns_)

# Use Apriori to find frequent itemsets
frequent_itemsets = apriori(df_apriori, min_support=0.1, use_colnames=True)

# Generate association rules from frequent itemsets
rules = association_rules(frequent_itemsets, metric="lift", min_threshold=1)
#显示所有列
pd.set_option('display.max_columns', None)
#显示所有行
pd.set_option('display.max_rows', None)

rules.to_csv('E:/Desktop/mnt/rules_行为特征.csv',encoding="gbk")

import pandas as pd
from mlxtend.frequent_patterns import apriori, association_rules
from mlxtend.preprocessing import TransactionEncoder

# Discretize continuous variables
data['母亲年龄段'] = pd.cut(data['母亲年龄'], bins=[0, 25, 35, 100], labels=['年轻', '中年',
    '年纪较大'])
data['妊娠时间段'] = pd.cut(data['妊娠时间 (周数)'], bins=[0, 36, 41, 100], labels=['较短',
    '正常', '较长'])

for column in ['CBTS']:
    mean_val = data[column].mean()
    std_val = data[column].std()
    data[f'{column}_段'] = pd.cut(data[column],
        bins=[-float('inf'), mean_val-0.5*std_val, mean_val+0.5*std_val,
            float('inf')],
        labels=['低_CBTS', '中_CBTS', '高_CBTS'])

for column in ['EPDS']:
    mean_val = data[column].mean()
    std_val = data[column].std()
    data[f'{column}_段'] = pd.cut(data[column],
        bins=[-float('inf'), mean_val-0.5*std_val, mean_val+0.5*std_val,
            float('inf')],
        labels=['低_EPDS', '中_EPDS', '高_EPDS'])

for column in ['HADS']:
    mean_val = data[column].mean()
    std_val = data[column].std()
    data[f'{column}_段'] = pd.cut(data[column],
        bins=[-float('inf'), mean_val-0.5*std_val, mean_val+0.5*std_val,
            float('inf')],
        labels=['低_HADS', '中_HADS', '高_HADS'])

```

```

# Prepare data for Apriori
columns_to_use = ['母亲年龄段', '婚姻状况', '教育程度', '妊娠时间段', '分娩方式', 'CBTS_段',
                  'EPDS_段', 'HADS_段', '睡醒次数']
for col in columns_to_use:
    data[col] = data[col].astype(str)
transactions = data[columns_to_use].values.tolist()

# Convert data to the right format for Apriori
te = TransactionEncoder()
te_ary = te.fit(transactions).transform(transactions)
df_apriori = pd.DataFrame(te_ary, columns=te.columns_)

# Use Apriori to find frequent itemsets
frequent_itemsets = apriori(df_apriori, min_support=0.1, use_colnames=True)

# Generate association rules from frequent itemsets
rules = association_rules(frequent_itemsets, metric="lift", min_threshold=1)
#显示所有列
pd.set_option('display.max_columns', None)
#显示所有行
pd.set_option('display.max_rows', None)

rules.to_csv('E:/Desktop/mnt/rules_睡醒次数.csv', encoding="gbk")
#画热力图
import numpy as np
import seaborn as sns
# values = np.random.rand(5, 5)
# ax = sns.heatmap(data, cmap="YlGnBu", annot=True, linewidths=.5) # 修改颜色, 添加线宽
## 1) 相关性分析
# Load data

data = pd.read_csv('E:/Desktop/mnt/源附件2.csv', encoding="gbk")
plt.rcParams['font.sans-serif'] = ['SimHei'] # 用来正常显示中文标签
plt.rcParams['axes.unicode_minus'] = False # 用来正常显示负号
behavior_encoding = {"安静型": 1, "中等型": 2, "矛盾型": 3}
data["婴儿行为特征"] = data["婴儿行为特征"].map(behavior_encoding)
numeric_features = ['母亲年龄', '婚姻状况', '教育程度', '妊娠时间（周数）', '分娩方式', 'CBTS',
                    'EPDS',
                    'HADS', '婴儿行为特征', '婴儿性别', '婴儿年龄（月）', '整晚睡眠时间（分钟）', '睡醒次数',
                    '入睡方式']
price_numeric = data[numeric_features]
correlation = price_numeric.corr()

#添加每个热力块的具体数值,保留两位小数
for i in range(price_numeric.shape[0]):
    for j in range(1, price_numeric.shape[1]):

```

```

        text = ax.text(j-1,i,price_numeric.iloc[i,j].round(2),ha="center", va="center",
            color="w",fontsize=12,fontname='Times New Roman')

f , ax = plt.subplots(figsize = (7, 7))

plt.title('母亲身体指标和心理指标对婴儿的行为特征和睡眠质量相关性分析热力图',y=1,size=10)

sns.heatmap(correlation,square = True, vmax=0.8, cmap="YlGnBu_r")
plt.savefig("E:/Desktop/p1_graph_tables/热力图.png")

```

附录 G 问题二 Python 代码

```

import pandas as pd
import optuna
from sklearn.model_selection import StratifiedKFold
from sklearn.metrics import accuracy_score
from sklearn.ensemble import GradientBoostingClassifier
from imblearn.over_sampling import SMOTE

# 读取数据集
data = pd.read_csv("E:/Desktop/try/processed_data.csv", encoding="gbk")

# 提取特征和目标变量
X = data[['母亲年龄', '婚姻状况', '教育程度', '妊娠时间（周数）', '分娩方式', 'CBTS', 'EPDS',
    'HADS']]
y = data['婴儿行为特征']

# 数据均衡化处理
oversampler = SMOTE(random_state=42)
X_balanced, y_balanced = oversampler.fit_resample(X, y)

# 初始化交叉验证
n_splits = 5
skf = StratifiedKFold(n_splits=n_splits, shuffle=True, random_state=42)

# 定义目标函数（准确率）
def objective(trial):
    accuracy_list = []

    for train_index, test_index in skf.split(X_balanced, y_balanced):
        X_train, X_test = X_balanced.iloc[train_index], X_balanced.iloc[test_index]
        y_train, y_test = y_balanced.iloc[train_index], y_balanced.iloc[test_index]

        # 创建并训练GBDT模型
        max_depth = trial.suggest_int("max_depth", 2, 6)

```

```

learning_rate = trial.suggest_loguniform("learning_rate", 0.001, 0.1)
n_estimators = trial.suggest_int("n_estimators", 50, 200)

model_gbdtd = GradientBoostingClassifier(max_depth=max_depth,
    learning_rate=learning_rate, n_estimators=n_estimators)
model_gbdtd.fit(X_train, y_train)

# 预测并计算准确率
y_pred_gbdtd = model_gbdtd.predict(X_test)

accuracy = accuracy_score(y_test, y_pred_gbdtd)
accuracy_list.append(accuracy)

# 返回平均准确率
return sum(accuracy_list)/len(accuracy_list)

# 使用贝叶斯优化调参
study = optuna.create_study(direction='maximize')
study.optimize(objective, n_trials=100)

best_params = study.best_params
print("Best Parameters:", best_params)

# 打印每一折的准确率和预测概率
params = best_params

accuracy_list = []
probability_list = []

for train_index, test_index in skf.split(X_balanced, y_balanced):
    X_train, X_test = X_balanced.iloc[train_index], X_balanced.iloc[test_index]
    y_train, y_test = y_balanced.iloc[train_index], y_balanced.iloc[test_index]

    # 创建并训练GBDT模型
    model_gbdtd = GradientBoostingClassifier(max_depth=params["max_depth"],
        learning_rate=params["learning_rate"], n_estimators=params["n_estimators"])
    model_gbdtd.fit(X_train, y_train)

    # 预测并计算准确率
    y_pred_gbdtd = model_gbdtd.predict(X_test)
    y_pred_proba_gbdtd = model_gbdtd.predict_proba(X_test)

    accuracy = accuracy_score(y_test, y_pred_gbdtd)
    accuracy_list.append(accuracy)
    probability_list.append({"class_0": y_pred_proba_gbdtd[:, 0], "class_1":
        y_pred_proba_gbdtd[:, 1], "class_2": y_pred_proba_gbdtd[:, 2]})

```

```
# 输出每一折的准确率和预测概率
print("Accuracy List:", accuracy_list)
print("Probability List:", probability_list)
```

附录 H 问题三 Python 源代码

```
import numpy as np
import pandas as pd
from sklearn.ensemble import GradientBoostingClassifier

plt.rcParams['font.sans-serif'] = ['SimHei'] # 用来正常显示中文标签
plt.rcParams['axes.unicode_minus'] = False # 用来正常显示负号

# 数据预处理
features = data.drop(columns=['整晚睡眠时间（分钟）', '睡醒次数', '入睡方式', '婴儿行为特征'])
labels = data['婴儿行为特征']

# 数据标准化
scaler = StandardScaler()
scaled_features = scaler.fit_transform(features)
scaled_features

# 构建模型
from sklearn.ensemble import GradientBoostingClassifier
# 训练模型
# model = LogisticRegression(multi_class='multinomial', solver='lbfgs', max_iter=10000)
model = GradientBoostingClassifier(learning_rate=0.1, n_estimators=300, max_depth=5,
    min_samples_leaf =20,
    min_samples_split =90, max_features='sqrt', subsample=0.9, random_state=80)
model.fit(scaled_features, labels)

#得到概率的函数模型
def update_and_get_probabilities_2(baby_id, cbts, epds, hads):
    baby = babies_data[babies_data["编号"] == baby_id]
    baby = baby.drop(columns=['婴儿行为特征', '整晚睡眠时间（分钟）', '睡醒次数', '入睡方式'])
    baby['CBTS'] = cbts
    baby['EPDS'] = epds
    baby['HADS'] = hads
    scaled_baby = scaler.transform(baby)
#    probabilities = model.predict_proba(scaled_baby)
    probabilities = model.predict_proba(scaled_baby)
    return {"安静型概率": probabilities[0][0], "中等型概率": probabilities[0][1], "矛盾型概率":
        probabilities[0][2]}

# 对每个EPDS值计算概率，从22开始逐渐降低，这次我们考虑负数范围
probabilities_quiet_epds = []
```

```

probabilities_moderate_epds = []
probabilities_conflicting_epds = []
epds_range = np.linspace(22, -10, 50)
current_cbts = 15
current_hads = data[data["编号"] == 238]["EPDS"].values[0]

for epds in epds_range:
    probs = update_and_get_probabilities_2(238, current_cbts, epds, current_hads)
    probabilities_quiet_epds.append(probs['安静型概率'])
    probabilities_moderate_epds.append(probs['中等型概率'])
    probabilities_conflicting_epds.append(probs['矛盾型概率'])

# 绘制概率随EPDS变化的曲线

plt.figure(figsize=(10, 6))
plt.plot(epds_range, probabilities_quiet_epds, label='安静型概率', marker='o')
plt.plot(epds_range, probabilities_moderate_epds, label='中等型概率', marker='o')
plt.plot(epds_range, probabilities_conflicting_epds, label='矛盾型概率', marker='o')
plt.xlabel('EPDS')
plt.ylabel('概率')
plt.title('概率随EPDS变化的曲线（从22开始逐渐降低至负数）')
plt.legend()
plt.grid(True)
plt.gca().invert_xaxis() # 从22开始逐渐降低，所以反转x轴
plt.savefig("E:/Desktop/p1_graph_tables/概率随EPDS变化的曲线（从22开始逐渐降低至负数）.png")
plt.show()

import numpy as np
import pandas as pd

# 加载数据
data=pd.read_excel("E:/Desktop/try/processed_data.xlsx")

# 过滤出安静型的婴儿数据
quiet_babies = data[data['婴儿行为特征'] == '安静型']

# 定义治疗费用计算函数
def cost_function(delta_CBTS, delta_EPDS, delta_HADS):
    return 200 * np.exp(delta_CBTS * 0.8811) + 500 * np.exp(delta_EPDS * 0.6649) + 300 *
        np.exp(delta_HADS * 0.7459)

# 计算从矛盾型转为安静型的最佳治疗费用
def calculate_final_correct_optimal_cost(target):
    x1 = min(target['CBTS'], 15)
    x2 = min(target['EPDS'], 22)
    x3 = min(target['HADS'], 18)

```

```

    return 200 * np.exp((15 - x1) * 0.8811) + 500 * np.exp((22 - x2) * 0.6649) + 300 *
           np.exp((18 - x3) * 0.7459)

# 计算每个安静型婴儿的治疗费用
optimal_costs_last_correction = []
for index, row in quiet_babies.iterrows():
    optimal_costs_last_correction.append((index, calculate_final_correct_optimal_cost(row)))

# 找出治疗费用最低的方案
min_optimal_cost_baby_last_correction = min(optimal_costs_last_correction, key=lambda x: x[1])

min_optimal_cost_baby_last_correction, data.loc[min_optimal_cost_baby_last_correction[0],
        ['CBTS', 'EPDS', 'HADS']]

```

附录 I 问题四 Python 源代码

```

import pandas as pd
from sklearn.cluster import KMeans
from sklearn.model_selection import train_test_split
from sklearn.ensemble import GradientBoostingClassifier

# 加载数据
data = pd.read_excel("E:/Desktop/try/processed_data.xlsx")
last_20_data = pd.read_excel("E:/Desktop/try/问题二结论.xlsx")

# 对睡眠特性执行聚类
sleep_features = data[['整晚睡眠时间（分钟）', '睡醒次数', '入睡方式']]
kmeans = KMeans(n_clusters=4, random_state=0).fit(sleep_features)
data['Sleep Quality Cluster'] = kmeans.labels_

# 定义用于训练的特性和目标变量
X = data[['母亲年龄', '婚姻状况', '教育程度', '妊娠时间（周数）', '分娩方式', 'CBTS', 'EPDS',
        'HADS']]
y = data['Sleep Quality Cluster']

# 将数据分为训练集和测试集
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# 训练梯度增强分类器
gbc = GradientBoostingClassifier(n_estimators=100, random_state=42)
gbc.fit(X_train, y_train)

# 从提供的数据中提取必要的特性进行预测
last_20_features = last_20_data[['母亲年龄', '婚姻状况', '教育程度', '妊娠时间（周数）',

```



```

    '分娩方式', 'CBTS', 'EPDS', 'HADS']]

# 预测最后20个婴儿的睡眠质量类别
predicted_clusters = gbc.predict(last_20_features)
print(predicted_clusters)

import numpy as np
import pandas as pd
from sklearn.ensemble import GradientBoostingClassifier
from sklearn.metrics import calinski_harabasz_score, silhouette_score
import skfuzzy as fuzz

# 加载数据
data = pd.read_excel("E:/Desktop/try/processed_data.xlsx")
last_20_data = pd.read_excel("E:/Desktop/try/问题二结论.xlsx")

# 模糊C均值聚类
sleep_features = data[['整晚睡眠时间（分钟）', '睡醒次数', '入睡方式']].values.T
cntr, u, _, _, _, _ = fuzz.cluster.cmeans(sleep_features, 4, 2, error=0.005, maxiter=1000)
data['Sleep Quality Cluster'] = np.argmax(u, axis=0)

# 计算CH值和轮廓系数
ch_score = calinski_harabasz_score(sleep_features.T, np.argmax(u, axis=0))
silhouette_coeff = silhouette_score(sleep_features.T, np.argmax(u, axis=0))

# 根据聚类中心确定每个聚类的评级
sorted_cluster_centers = sorted(list(range(4)), key=lambda x: (-cntr[x, 0], cntr[x, 1]))
rating_dict = {sorted_cluster_centers[0]: "优",
               sorted_cluster_centers[1]: "良",
               sorted_cluster_centers[2]: "中",
               sorted_cluster_centers[3]: "差"}

# 定义用于训练的特性和目标变量
X = data[['母亲年龄', '婚姻状况', '教育程度', '妊娠时间（周数）', '分娩方式', 'CBTS', 'EPDS',
          'HADS']]
y = data['Sleep Quality Cluster']

# 将数据分为训练集和测试集
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# 训练梯度增强分类器
gbc = GradientBoostingClassifier(n_estimators=100, random_state=42)
gbc.fit(X_train, y_train)

# 从提供的数据中提取必要的特性进行预测
last_20_features = last_20_data[['母亲年龄', '婚姻状况', '教育程度', '妊娠时间（周数）',

```

```

    '分娩方式', 'CBTS', 'EPDS', 'HADS']]

# 预测最后20个婴儿的睡眠质量类别，并转化为评级
predicted_clusters = gbc.predict(last_20_features)
predicted_ratings = [rating_dict[cluster] for cluster in predicted_clusters]

print("CH值:", ch_score)
print("轮廓系数:", silhouette_coeff)
print("预测的睡眠质量评级:", predicted_ratings)

```

附录 J 问题五 Python 源代码

```

#####睡眠质量优化模型#####

import pandas as pd
import numpy as np
from sklearn.ensemble import GradientBoostingClassifier

# 加载数据
# known_babies_data = pd.read_excel("/mnt/dataown_babies_ratings.xlsx")
known_babies_data = pd.read_excel("E:/Desktop/try/known_babies_ratings.xlsx")

# 选择用于训练的特征
features = known_babies_data.columns.drop(["编号", "Sleep Quality Rating"])

# 训练GBDT模型
sleep_model = GradientBoostingClassifier(random_state=42)
X = known_babies_data[features]
y = known_babies_data["Sleep Quality Rating"]
sleep_model.fit(X, y)

# 定义治疗费用函数
def treatment_cost(CBTS, EPDS, HADS):
    return 200 * np.exp((15 - CBTS) * 0.8811) + 500 * np.exp((22 - EPDS) * 0.6649) + 300 *
        np.exp((18 - HADS) * 0.7459)

# 定义模拟退火函数
def simulated_annealing_for_good_sleep(epochs, initial_temperature, cooling_rate, data, model):
    current_solution = data[features].iloc[0].copy()
    current_cost = treatment_cost(current_solution["CBTS"], current_solution["EPDS"],
        current_solution["HADS"])
    print(current_cost)

    best_solution = current_solution.copy()
    best_cost = current_cost

```

```

temperature = initial_temperature
# 生成邻近解
neighbor_solution = current_solution.copy()
neighbor_solution["CBTS"] = max(0, current_solution["CBTS"] - np.random.uniform(0, 1))
neighbor_solution["EPDS"] = max(0, current_solution["EPDS"] - np.random.uniform(0, 1))
neighbor_solution["HADS"] = max(0, current_solution["HADS"] - np.random.uniform(0, 1))
neighbor_cost = treatment_cost(neighbor_solution["CBTS"], neighbor_solution["EPDS"],
                                neighbor_solution["HADS"])
best_cost = neighbor_cost

for epoch in range(epochs):
    # 生成邻近解
    neighbor_solution = neighbor_solution.copy()
    neighbor_solution["CBTS"] = max(0, current_solution["CBTS"] - np.random.uniform(0, 1))
    neighbor_solution["EPDS"] = max(0, current_solution["EPDS"] - np.random.uniform(0, 1))
    neighbor_solution["HADS"] = max(0, current_solution["HADS"] - np.random.uniform(0, 1))

    # 计算新的治疗费用
    neighbor_cost = treatment_cost(neighbor_solution["CBTS"], neighbor_solution["EPDS"],
                                    neighbor_solution["HADS"])

    # 判断新解是否为更好的解
    if model.predict(pd.DataFrame([neighbor_solution]))[0] == "良" and neighbor_cost <
        best_cost :
    # if model.predict(pd.DataFrame([neighbor_solution]))[0] == "良" :
        best_solution = neighbor_solution.copy()
        best_cost = neighbor_cost

    # 否则，根据Metropolis准则判断是否接受新解
    elif np.random.rand() < np.exp(-(neighbor_cost - current_cost) / temperature):
        current_solution = neighbor_solution.copy()
        current_cost = neighbor_cost

    # 降温
    temperature *= cooling_rate

return best_solution, best_cost

# 获取238号婴儿的数据
baby_238_data = known_babies_data[known_babies_data["编号"] == 238]

# 运行模拟退火算法
best_solution, best_cost = simulated_annealing_for_good_sleep(10, 1000, 0.995, baby_238_data,
                                                             sleep_model)
predicted_rating = sleep_model.predict(pd.DataFrame([best_solution]))[0]

```