

权 利 要 求

1、一种基于知识驱动和数据驱动的分群方法，其中，所述方法包括：

根据医学规则数据创建初始知识分群决策树，并将样本数据填充至所述初始知识分群决策树各分支的末端知识子节点，得到目标知识分群决策树，所述医学规则数据用于指示对已归档的医学文献、医学共识、医学指南和专家的临床经验的医学学科知识数据进行规则提取处理所得的第一分群规则，所述第一分群规则用于对临床上的相似特征人群进行分群，所述样本数据包括遵循医学知识规则的临床上的相似特征人群对应的临床结局数据；

获取待分类人群的特征数据，并通过所述目标知识分群决策树对所述特征数据进行分群，得到目标末端知识子节点对应的处理方案，并对所述处理方案的类别数量进行识别，所述目标末端知识子节点为所述特征数据在所述目标知识分群决策树上的节点位置，所述处理方案用于指示对相似特征人群的解决渠道；

若识别到所述处理方案的类别数量为一个，则确定所述待分类人群属于第一特征相似人群，所述第一特征相似人群为所述目标知识分群决策树中目标末端知识子节点对应的处理方案所属的特征相似人群；

若识别到所述处理方案的类别数量为至少两个，则基于第二分群规则生成数据分群决策树，并确定每个处理方案对应所述数据分群决策树的数据子节点所处的目标层数，所述第二分群规则指示根据递归分区算法和处理方案的有效性进行分类；

若所述目标层数为第一层，则将所述待分类人群划分为所述第一特征相似人群；

若所述目标层数不为第一层，则将所述待分类人群划分为第二特征相似人群，所述第二特征相似人群为所述数据分群决策树中目标数据节点对应的处理方案所属的特征相似人群，每个目标数据节点对应不同的特征相似人群。

2、根据权利要求1所述的基于知识驱动和数据驱动的分群方法，其中，所述基于第二分群规则生成数据分群决策树，并确定每个处理方案对应所述数据分群决策树的数据子节点所处的目标层数，包括：

步骤一：将所述目标末端知识子节点作为所述数据分群决策树的数据根节点，计算所述数据根节点对应的样本数据中处理方案的有效性分值；

步骤二：根据每个处理方案的有效性分值对所述数据根节点进行分支，根据递归分区算法将每个处理方案依次划分到每个分支的数据子节点中，并继续对每个数据子节点对应的处理方案进行有效性分析；

步骤三：若当前数据子节点所对应的处理方案的有效性分值大于第一阈值，则对所述当前数据子节点进行分支；

步骤四：重复执行步骤三，直至每个数据子节点所对应的处理方案的有效性分值小于或等于所述第一阈值，得到数据分群决策树，所述数据分群决策树包括多个数据子节点，所述多个数据子节点中每个末端数据叶子节点对应有一个处理方案；

步骤五：确定每个处理方案对应的末端数据叶子节点在所述数据分群决策树中所处的目标层数。

3、根据权利要求2所述的基于知识驱动和数据驱动的分群方法，其中，所述将所述目标末端知识子节点作为所述数据分群决策树的数据根节点，计算所述数据根节点对应的处理方案的有效性分值，包括：

将所述目标末端知识子节点作为所述数据分群决策树的数据根节点；

获取所述数据根节点对应的样本数据中处理方案的目标特征信息，将所述目标特征信息拆分成因素因子；

计算所述因素因子在不同水平上的关联度、对预置影响因素的影响度以及所述因素因子不控制的边缘关联度；

通过预置的可压缩准则和可比较准则分析所述关联度、所述影响度和所述边缘关联度，得到混杂因子；

计算所述混杂因子与所述预置影响因素发生概率的回归系数的指数变换值，并以所述指数

变换值作为判定得分, 得到所述数据根节点对应的处理方案的有效性分值, 所述指数变换值的计算公式如下:

$$p(y=k|x;w)=\frac{\exp(w_k^T x)}{\sum_{l=1}^K \exp(w_l^T x)},$$

所述 $p(y=k|x;w)$ 为所述指数变换值, 所述 w 为所述混杂因子与所述预置影响因素发生的概率, 所述 x 为所述混杂因子, 所述 k 为所述预置影响因素的类别数量, 所述 y 为输出的预测值, 所述 y 是 k 的条件概率, 所述 K 为 k 的最大值, 所述 T 为时间周期。

4、根据权利要求 2 所述的基于知识驱动和数据驱动的分群方法, 其中, 所述若所述目标层数不为第一层, 则将所述待分类人群划分为第二特征相似人群, 包括:

若所述目标层数不为第一层, 判断所述目标层数对应处理方案的有效性分值是否大于第二阈值, 所述第二阈值大于所述第一阈值;

若所述目标层数对应处理方案的有效性分值大于第二阈值, 则将有效性分值大于第二阈值的处理方案所位于的数据叶子节点作为目标数据节点, 并将所述待分类人群划分为所述目标数据节点对应的处理方案所属的特征相似人群, 得到第二特征相似人群;

若所述目标层数对应处理方案的有效性分值小于或者等于第二阈值, 则分析所述目标层数的上一层数对应处理方案的有效性分值;

若所述目标层数的上一层数对应处理方案的有效性分值大于第二阈值, 则将所述目标层数的上一层数对应处理方案所位于的数据叶子节点作为目标数据节点, 并将所述待分类人群划分为所述目标数据节点对应的处理方案所属的特征相似人群, 得到第二特征相似人群。

5、根据权利要求 1 所述的基于知识驱动和数据驱动的分群方法, 其中, 所述根据医学规则数据创建初始知识分群决策树, 并将样本数据填充至所述初始知识分群决策树各分支的末端知识子节点, 得到目标知识分群决策树之前, 所述方法还包括:

从多个系统、多个平台和多个医疗系统获取医学学科知识数据和临床数据, 并对所述医学学科知识数据和所述临床数据进行数据处理, 获得医学规则数据, 所述数据处理包括数据清理处理、数据集成处理、数据规约处理、数据变换处理、特征提取处理、数据脱敏处理和分类处理。

6、根据权利要求 1 所述的基于知识驱动和数据驱动的分群方法, 其中, 所述对所述处理方案的类别数量进行识别, 包括:

计算所述处理方案之间的相似度获得多个相似度值, 并比较多个所述相似度值是否均在预设范围内;

若比较结果为是, 则统计所述处理方案的类别为一个;

若比较结果为否, 则统计所述处理方案的类别为多个。

7、根据权利要求 1-6 中任一项所述的基于知识驱动和数据驱动的分群方法, 其中, 所述若所述目标层数不为第一层, 则将所述待分类人群划分为第二特征相似人群之后, 所述方法还包括:

调用预置的解释模型对所述第一特征相似人群或所述第二特征相似人群进行解释处理和可视化图表统计分析处理, 得到可解释性特征相似人群, 所述解释处理包括有效性排序处理、匹配度计算处理和匹配度排序处理。

8、一种基于知识驱动和数据驱动的分群装置, 其中, 所述装置包括:

第一创建单元, 用于根据医学规则数据创建初始知识分群决策树, 并将样本数据填充至所述初始知识分群决策树各分支的末端知识子节点, 得到目标知识分群决策树, 所述医学规则数据用于指示对已归档的医学文献、医学共识、医学指南和专家的临床经验的医学学科知识数据进行规则提取处理所得的第一分群规则, 所述第一分群规则用于对临床上的相似特征人群进行分群, 所述样本数据包括遵循医学知识规则的临床上的相似特征人群对应的临床结局数据;

识别单元, 用于获取待分类人群的特征数据, 并通过所述目标知识分群决策树对所述特征数据进行分群, 得到目标末端知识子节点对应的处理方案, 并对所述处理方案的类别数量进行识别, 所述目标末端知识子节点为所述特征数据在所述目标知识分群决策树上的节点位置, 所

述处理方案用于指示对相似特征人群的解决渠道；

确定单元，用于若识别到所述处理方案的类别数量为一个，则确定所述待分类人群属于第一特征相似人群，所述第一特征相似人群为所述目标知识分群决策树中目标末端知识子节点对应的处理方案所属的特征相似人群；

- 5 第二创建单元，用于若识别到所述处理方案的类别数量为至少两个，则基于第二分群规则生成数据分群决策树，并确定每个处理方案对应所述数据分群决策树的数据子节点所处的目标层数，所述第二分群规则指示根据递归分区算法和处理方案的有效性进行分类各分支的末端子节点的；

- 10 第一划分单元，用于若所述目标层数为第一层，则将所述待分类人群划分为所述第一特征相似人群；

第二划分单元，用于若所述目标层数不为第一层，则将所述待分类人群划分为第二特征相似人群，所述第二特征相似人群为所述数据分群决策树中目标数据节点对应的处理方案所属的特征相似人群，每个目标数据节点对应不同的特征相似人群。

- 15 9、一种基于知识驱动和数据驱动的分群设备，其中，所述基于知识驱动和数据驱动的分群设备包括：

存储器、处理器及存储在所述存储器上并可在所述处理器上运行的计算机程序，所述处理器执行所述计算机程序时实现如下所述的基于知识驱动和数据驱动的分群方法的步骤：

- 20 根据医学规则数据创建初始知识分群决策树，并将样本数据填充至所述初始知识分群决策树各分支的末端知识子节点，得到目标知识分群决策树，所述医学规则数据用于指示对已归档的医学文献、医学共识、医学指南和专家的临床经验的医学学科知识数据进行规则提取处理所得的第一分群规则，所述第一分群规则用于对临床上的相似特征人群进行分群，所述样本数据包括遵循医学知识规则的临床上的相似特征人群对应的临床结局数据；

- 25 获取待分类人群的特征数据，并通过所述目标知识分群决策树对所述特征数据进行分群，得到目标末端知识子节点对应的处理方案，并对所述处理方案的类别数量进行识别，所述目标末端知识子节点为所述特征数据在所述目标知识分群决策树上的节点位置，所述处理方案用于指示对相似特征人群的解决渠道；

若识别到所述处理方案的类别数量为一个，则确定所述待分类人群属于第一特征相似人群，所述第一特征相似人群为所述目标知识分群决策树中目标末端知识子节点对应的处理方案所属的特征相似人群；

- 30 若识别到所述处理方案的类别数量为至少两个，则基于第二分群规则生成数据分群决策树，并确定每个处理方案对应所述数据分群决策树的数据子节点所处的目标层数，所述第二分群规则指示根据递归分区算法和处理方案的有效性进行分类；

若所述目标层数为第一层，则将所述待分类人群划分为所述第一特征相似人群；

- 35 若所述目标层数不为第一层，则将所述待分类人群划分为第二特征相似人群，所述第二特征相似人群为所述数据分群决策树中目标数据节点对应的处理方案所属的特征相似人群，每个目标数据节点对应不同的特征相似人群。

10、根据权利要求9所述的基于知识驱动和数据驱动的分群设备，其中，所述基于第二分群规则生成数据分群决策树，并确定每个处理方案对应所述数据分群决策树的数据子节点所处的目标层数，包括以下步骤：

- 40 步骤一：将所述目标末端知识子节点作为所述数据分群决策树的数据根节点，计算所述数据根节点对应的样本数据中处理方案的有效性分值；

步骤二：根据每个处理方案的有效性分值对所述数据根节点进行分支，根据递归分区算法将每个处理方案依次划分到每个分支的数据子节点中，并继续对每个数据子节点对应的处理方案进行有效性分析；

- 45 步骤三：若当前数据子节点所对应的处理方案的有效性分值大于第一阈值，则对所述当前数据子节点进行分支；

步骤四：重复执行步骤三，直至每个数据子节点所对应的处理方案的有效性分值小于或等

于所述第一阈值,得到数据分群决策树,所述数据分群决策树包括多个数据子节点,所述多个数据子节点中每个末端数据叶子节点对应有一个处理方案;

步骤五:确定每个处理方案对应的末端数据叶子节点在所述数据分群决策树中所处的目标层数。

- 5 11、根据权利要求 10 所述的基于知识驱动和数据驱动的分群设备,其中,所述将所述目标末端知识子节点作为所述数据分群决策树的数据根节点,计算所述数据根节点对应的处理方案的有效性分值,包括以下步骤:

将所述目标末端知识子节点作为所述数据分群决策树的数据根节点;

- 10 获取所述数据根节点对应的样本数据中处理方案的目标特征信息,将所述目标特征信息拆分成因素因子;

计算所述因素因子在不同水平上的关联度、对预置影响因素的影响度以及所述因素因子不控制的边缘关联度;

通过预置的可压缩准则和可比较准则分析所述关联度、所述影响度和所述边缘关联度,得到混杂因子;

- 15 计算所述混杂因子与所述预置影响因素发生概率的回归系数的指数变换值,并以所述指数变换值作为判定得分,得到所述数据根节点对应的处理方案的有效性分值,所述指数变换值的计算公式如下:

$$p(y=k|x;w)=\frac{\exp(w_k^T x)}{\sum_{i=1}^K \exp(w_i^T x)},$$

- 20 所述 $p(y=k|x;w)$ 为所述指数变换值,所述 w 为所述混杂因子与所述预置影响因素发生的概率,所述 x 为所述混杂因子,所述 k 为所述预置影响因素的类别数量,所述 y 为输出的预测值,所述 y 是 k 的条件概率,所述 K 为 k 的最大值,所述 T 为时间周期。

12、根据权利要求 10 所述的基于知识驱动和数据驱动的分群设备,其中,所述若所述目标层数不为第一层,则将所述待分类人群划分为第二特征相似人群,包括以下步骤:

- 25 若所述目标层数不为第一层,判断所述目标层数对应处理方案的有效性分值是否大于第二阈值,所述第二阈值大于所述第一阈值;

若所述目标层数对应处理方案的有效性分值大于第二阈值,则将有效性分值大于第二阈值的处理方案所位于的数据叶子节点作为目标数据节点,并将所述待分类人群划分为所述目标数据节点对应的处理方案所属的特征相似人群,得到第二特征相似人群;

- 30 若所述目标层数对应处理方案的有效性分值小于或者等于第二阈值,则分析所述目标层数的上一层数对应处理方案的有效性分值;

若所述目标层数的上一层数对应处理方案的有效性分值大于第二阈值,则将所述目标层数的上一层数对应处理方案所位于的数据叶子节点作为目标数据节点,并将所述待分类人群划分为所述目标数据节点对应的处理方案所属的特征相似人群,得到第二特征相似人群。

- 35 13、根据权利要求 9 所述的基于知识驱动和数据驱动的分群设备,其中,所述根据医学规则数据创建初始知识分群决策树,并将样本数据填充至所述初始知识分群决策树各分支的末端知识子节点,得到目标知识分群决策树之前,还包括以下步骤:

从多个系统、多个平台和多个医疗系统获取医学学科知识数据和临床数据,并对所述医学学科知识数据和所述临床数据进行数据处理,获得医学规则数据,所述数据处理包括数据清理处理、数据集成处理、数据规约处理、数据变换处理、特征提取处理、数据脱敏处理和分类处理。

- 40 14、根据权利要求 9 所述的基于知识驱动和数据驱动的分群设备,其中,所述对所述处理方案的类别数量进行识别,包括以下步骤:

计算所述处理方案之间的相似度获得多个相似度值,并比较多个所述相似度值是否均在预设范围内;

- 45 若比较结果为是,则统计所述处理方案的类别为一个;

若比较结果为否,则统计所述处理方案的类别为多个。

15、根据权利要求 9-14 中任一项所述的基于知识驱动和数据驱动的分群设备，其中，所述若所述目标层数不为第一层，则将所述待分类人群划分为第二特征相似人群之后，还包括以下步骤：

5 调用预置的解释模型对所述第一特征相似人群或所述第二特征相似人群进行解释处理和可视化图表统计分析处理，得到可解释性特征相似人群，所述解释处理包括有效性排序处理、匹配度计算处理和匹配度排序处理。

16、一种计算机可读存储介质，其中，包括指令，当所述指令在计算机上运行时，使得计算机执行如下所述的基于知识驱动和数据驱动的分群方法的步骤：

10 根据医学规则数据创建初始知识分群决策树，并将样本数据填充至所述初始知识分群决策树各分支的末端知识子节点，得到目标知识分群决策树，所述医学规则数据用于指示对已归档的医学文献、医学共识、医学指南和专家的临床经验的医学学科知识数据进行规则提取处理所得的第一分群规则，所述第一分群规则用于对临床上的相似特征人群进行分群，所述样本数据包括遵循医学知识规则的临床上的相似特征人群对应的临床结局数据；

15 获取待分类人群的特征数据，并通过所述目标知识分群决策树对所述特征数据进行分群，得到目标末端知识子节点对应的处理方案，并对所述处理方案的类别数量进行识别，所述目标末端知识子节点为所述特征数据在所述目标知识分群决策树上的节点位置，所述处理方案用于指示对相似特征人群的解决渠道；

20 若识别到所述处理方案的类别数量为一个，则确定所述待分类人群属于第一特征相似人群，所述第一特征相似人群为所述目标知识分群决策树中目标末端知识子节点对应的处理方案所属的特征相似人群；

若识别到所述处理方案的类别数量为至少两个，则基于第二分群规则生成数据分群决策树，并确定每个处理方案对应所述数据分群决策树的数据子节点所处的目标层数，所述第二分群规则指示根据递归分区算法和处理方案的有效性进行分类；

25 若所述目标层数为第一层，则将所述待分类人群划分为所述第一特征相似人群；

若所述目标层数不为第一层，则将所述待分类人群划分为第二特征相似人群，所述第二特征相似人群为所述数据分群决策树中目标数据节点对应的处理方案所属的特征相似人群，每个目标数据节点对应不同的特征相似人群。

17、根据权利要求 16 所述的计算机可读存储介质，其中，所述基于知识驱动和数据驱动的分群的指令被计算机执行所述基于第二分群规则生成数据分群决策树，并确定每个处理方案对应所述数据分群决策树的数据子节点所处的目标层数的步骤时，包括以下步骤：

30 步骤一：将所述目标末端知识子节点作为所述数据分群决策树的数据根节点，计算所述数据根节点对应的样本数据中处理方案的有效性分值；

35 步骤二：根据每个处理方案的有效性分值对所述数据根节点进行分支，根据递归分区算法将每个处理方案依次划分到每个分支的数据子节点中，并继续对每个数据子节点对应的处理方案进行有效性分析；

步骤三：若当前数据子节点所对应的处理方案的有效性分值大于第一阈值，则对所述当前数据子节点进行分支；

40 步骤四：重复执行步骤三，直至每个数据子节点所对应的处理方案的有效性分值小于或等于所述第一阈值，得到数据分群决策树，所述数据分群决策树包括多个数据子节点，所述多个数据子节点中每个末端数据叶子节点对应有一个处理方案；

步骤五：确定每个处理方案对应的末端数据叶子节点在所述数据分群决策树中所处的目标层数。

18、根据权利要求 17 所述的计算机可读存储介质，其中，所述基于知识驱动和数据驱动的分群的指令被计算机执行所述将所述目标末端知识子节点作为所述数据分群决策树的数据根节点，计算所述数据根节点对应的处理方案的有效性分值的步骤时，包括以下步骤：

45 将所述目标末端知识子节点作为所述数据分群决策树的数据根节点；

获取所述数据根节点对应的样本数据中处理方案的目标特征信息，将所述目标特征信息拆

分成因素因子；

计算所述因素因子在不同水平上的关联度、对预置影响因素的影响度以及所述因素因子不控制的边缘关联度；

- 5 通过预置的可压缩准则和可比较准则分析所述关联度、所述影响度和所述边缘关联度，得到混杂因子；

计算所述混杂因子与所述预置影响因素发生概率的回归系数的指数变换值，并以所述指数变换值作为判定得分，得到所述数据根节点对应的处理方案的有效性分值，所述指数变换值的计算公式如下：

$$p(y = k|x; w) = \frac{\exp(w_k^T x)}{\sum_{l=1}^K \exp(w_l^T x)},$$

- 10 所述 $p(y = k|x; w)$ 为所述指数变换值，所述 w 为所述混杂因子与所述预置影响因素发生的概率，所述 x 为所述混杂因子，所述 k 为所述预置影响因素的类别数量，所述 y 为输出的预测值，所述 y 是 k 的条件概率，所述 K 为 k 的最大值，所述 T 为时间周期。

- 15 19、根据权利要求 17 所述的计算机可读存储介质，其中，所述基于知识驱动和数据驱动的分群的指令被计算机执行所述若所述目标层数不为第一层，则将所述待分类人群划分为第二特征相似人群的步骤时，包括以下步骤：

若所述目标层数不为第一层，判断所述目标层数对应处理方案的有效性分值是否大于第二阈值，所述第二阈值大于所述第一阈值；

- 20 若所述目标层数对应处理方案的有效性分值大于第二阈值，则将有效性分值大于第二阈值的处理方案所位于的数据叶子节点作为目标数据节点，并将所述待分类人群划分为所述目标数据节点对应的处理方案所属的特征相似人群，得到第二特征相似人群；

若所述目标层数对应处理方案的有效性分值小于或者等于第二阈值，则分析所述目标层数的上一层数对应处理方案的有效性分值；

- 25 若所述目标层数的上一层数对应处理方案的有效性分值大于第二阈值，则将所述目标层数的上一层数对应处理方案所位于的数据叶子节点作为目标数据节点，并将所述待分类人群划分为所述目标数据节点对应的处理方案所属的特征相似人群，得到第二特征相似人群。

20、根据权利要求 16 所述的计算机可读存储介质，其中，所述基于知识驱动和数据驱动的分群的指令被计算机执行所述根据医学规则数据创建初始知识分群决策树，并将样本数据填充至所述初始知识分群决策树各分支的末端知识子节点，得到目标知识分群决策树之前，还包括以下步骤：

- 30 从多个系统、多个平台和多个医疗系统获取医学学科知识数据和临床数据，并对所述医学学科知识数据和所述临床数据进行数据处理，获得医学规则数据，所述数据处理包括数据清理处理、数据集成处理、数据规约处理、数据变换处理、特征提取处理、据脱敏处理和分类处理。