

权 利 要 求 书

1、基于大数据的人力资源文本信息管理方法，包括：

获取从多渠道采集的待分析对象的文本数据；

提取所述文本数据的代表性特征，所述代表性特征基于所述文本数据中的每个特征的基本信息得到，所述基本信息包括每个特征的编码、频率和长度；

对所述代表性特征进行处理以将所述代表性特征转化为另一维度的特征；

对处理后的代表性特征进行特征分类和特征分组，所述特征分类用于评价所述代表性特征的位阶，所述特征分组用于聚类属于相近特征的代表性特征；

根据所述代表性特征的特征分组和特征分类的结果，计算待分析对象的文本数据的特征得分。

2、根据权利要求 1 所述的基于大数据的人力资源文本信息管理方法，其中，提取所述文本数据的代表性特征，所述代表性特征基于所述文本数据中的每个特征的基本信息得到，所述基本信息包括每个特征的编码、频率和长度，包括：

采用公式（1）提取所述文本数据的代表性特征：

$$D = \sum \frac{1}{N} [f_1(w_k, T) * \log(\delta \| w_k \|^2 + \gamma f_2(w_k)) + \frac{f_1(w_k, T) * \sqrt{\gamma f_2(w_k) + 5 \| w_k \|^2}}{L(w_k, T) + K(w_k, T)}]$$

公式（1）

其中， D 表示代表性特征， T 表示文本数据， w_k 表示文本数据 T 中的第 k 个分词， f_1 表示分词编码函数， f_2 表示分词概率密度函数， $L(w_k, T)$ 表示文本数据 T 中的第 k 个分词 w_k 的长度， $K(w_k, T)$ 表示第 k 个分词 w_k 在文本数据

T 中出现的次数, γ 表示分词概率影响因子, δ 表示第 k 个分词 w_k 的缩放因子, N 表示文本数据 T 中的分词的总数, ||| 表示范数符号。

3、根据权利要求 2 所述的基于大数据的人力资源文本信息管理方法, 其中, 对所述代表性特征进行处理以将所述代表性特征转化为另一维度的特征, 包括:

对所述代表性特征 D , 采用公式 (2) 进行处理:

$$D_1 = \sum D * \frac{\log \frac{D + \sqrt{D^2 - 4\mu c}}{2\sigma} - a_1 \log \sqrt{D^2 + 5Db^2}}{e^{\sqrt{D^2 - 4\mu c}} + \alpha e^{-\beta D^2}} \quad \text{公式 (2)}$$

其中, D 表示代表性特征, D_1 表示预处理后的代表性特征, 即, 另一维度的特征, μ 表示代表性特征 D 的均值, σ 表示代表性特征 D 的方差, a_1 表示代表性特征 D 的平滑因子, α 表示代表性特征 D 的正则化因子, β 表示代表性特征 D 的惩罚因子, b 表示代表性特征 D 的缩放因子, c 表示代表性特征 D 的均值平滑因子。

4、根据权利要求 1 所述的基于大数据的人力资源文本信息管理方法, 其中, 对处理后的代表性特征进行特征分类和特征分组中的特征分组, 包括:

$$R_1 = \sum \| D_1 - V_j \|^2 * \log \frac{D + \sqrt{D^2 - 4\mu c}}{2\sigma} \quad \text{公式 (3)}$$

其中, R_1 表示特征分组的结果, D_1 表示预处理后的代表性特征, 即, 另一维度的特征, μ 表示代表性特征 D 的均值, V_j 表示预处理后的代表性特征 D_1 中的第 j 个中心特征, σ 表示代表性特征 D 的方差, D 表示代表性特征, ||| 表示范数符号, c 表示代表性特征 D 的均值平滑因子。

5、根据权利要求 1 所述的基于大数据的人力资源文本信息管理方法, 其中, 对处理后的代表性特征进行特征分类和特征分组中的特征分类, 包括:

$$R_2 = \sum \frac{D_1 w_k}{\sqrt{\|w_k\|^2 + \varepsilon}} + p_i \frac{D_1 \theta_i}{\|D_1\| * \|\theta_i\|} \log \sqrt{D_1^2 - D b^2}$$

公式 (4)

其中, R_2 表示特征分类的结果, D_1 表示预处理后的代表性特征, 即, 另一维度的特征, b 表示代表性特征 D 的缩放因子, $\|\cdot\|$ 表示范数符号, w_k 表示文本数据 T 中的第 k 个分词, p_i 表示预处理后的代表性特征 D_1 中的第 i 个特征的权重因子, ε 表示预处理后的代表性特征 D_1 的惩罚因子, θ_i 表示预处理后的代表性特征 D_1 中的第 i 个特征的缩放因子。

6、根据权利要求 1 所述的基于大数据的人力资源文本信息管理方法, 其中, 根据所述代表性特征的特征分组和特征分类的结果, 计算待分析对象的文本数据的特征得分, 包括:

采用公式 (5) 计算所述待分析对象的特征得分:

$$R = \frac{\tau_1 e^{X_1 R_1} + \tau_2 e^{X_2 R_2}}{\tau_3 + \tau_4 (\|R_1\|^2 + \|R_2\|^2)}$$

公式 (5)

其中, R 表示特征得分, R_1 表示特征分组的结果, R_2 表示特征分类的结果, X_1 表示特征得分的结果 R_1 的权重影响因子, X_2 表示特征分组的结果 R_2 的权重影响因子, τ_1 表示特征得分的结果 R_1 的特征缩放系数, τ_2 表示特征分组的结果 R_2 的特征缩放系数, τ_3, τ_4 表示特征调节因子。

7、根据权利要求 1-6 中任一项所述的基于大数据的人力资源文本信息管理方法, 其中, 在根据所述代表性特征的特征分组和特征分类的结果, 计算待分析对象的特征得分之后, 所述基于大数据的人力资源文本信息管理方法还包括:

将所述待分析对象的特征得分推送至分析用户。

8、根据权利要求 1-6 中任一项所述的基于大数据的人力资源文本信息管理方法, 其中, 在提取所述文本信息的代表性特征的步骤之前, 所述基于

大数据的人力资源文本信息管理方法还包括：

对所述文本数据进行数据清洗。

9、基于大数据的人力资源文本信息管理装置，包括：

存储器，所述存储器被配置成存储指令；以及

处理器，所述处理器被配置成从所述存储器调用所述指令，以执行根据权利要求 1-8 中任一项所述的基于大数据的人力资源文本信息管理方法。

10、介质，所述介质上存储有指令，所述指令用于使得机器执行根据权利要求 1-8 中任一项所述的基于大数据的人力资源文本信息管理方法。