

权 利 要 求 书

1、一种基于长短期记忆模型的疾病预测方法，其特征在于，包括：

获取目标对象的第一医疗数据和关联对象的第二医疗数据，其中所述目标对象与关联对象存在血缘关系，第一医疗数据包括用药史、疾病史和手术史，第二医疗数据包括遗传病治疗史；

将第一医疗数据和第二医疗数据输入已训练完成的长短期记忆模型中的第一长短期记忆网络中运算，得到所述第一长短期记忆网络中的隐藏状态向量序列，所述长短期记忆模型包括用于编码的第一长短期记忆网络和用于解码的第二长短期记忆网络；

将所述隐藏状态向量序列输入所述第二长短期记忆网络中运算，得到疾病预测结果，其中所述疾病预测结果包括预测疾病类型与对应的发病机率；

从所述疾病预测结果中筛选出发病机率高于预设阈值的预测疾病，记为指定疾病，并根据预设的疾病关联网络获取与所述指定疾病直接连接的关联疾病，其中关联网络的网络节点为不同的疾病类型；

输出所述疾病预测结果与所述关联疾病。

2、根据权利要求1所述的基于长短期记忆模型的疾病预测方法，其特征在于，所述将第一医疗数据和第二医疗数据输入已训练完成的长短期记忆模型中的第一长短期记忆网络中运算，得到所述第一长短期记忆网络中的隐藏状态向量序列的步骤，包括：

将所述第一医疗数据根据预设的时间段分为多段数据序列；

根据预设的遗传病与其他疾病的影响因子对应关系，获取所述第二医疗数据中的遗传病对其他疾病的指定影响因子；

将所述多段数据序列与所述指定影响因子输入已训练完成的长短期记忆模型中的第一长短期记忆网络中运算，得到所述第一长短期记忆网络中的隐藏状态向量序列。

3、根据权利要求2所述的基于长短期记忆模型的疾病预测方法，其特征在于，所述将所述多段数据序列与所述指定影响因子输入已训练完成的长短期记忆模型中的第一长短期记忆网络中运算，得到所述第一长短期记忆网络中的隐藏状态向量序列的步骤，包括：

根据公式： $h_t = LSTM_{enc}(x_t, h_{t-1})$ ，获得所述第一长短期记忆网络中的隐藏状态向量 h_t ，其中 t 为第 t 个时间段， h_t 为对应于第 t 个时间段的隐藏状态向量， h_{t-1} 为对应于第 $t-1$ 个时间段的隐藏状态向量， X_t 为第 t 个时间段的输入数据， $LSTM_{enc}$ 指利用第一长短期记忆网络进行编码运算，其中 X_t 包括第 t 个时间段的第一医疗数据和第 t 个时间段的指定影响因子；

将多个预设的时间段对应的隐藏状态向量构成隐藏状态向量序列 $h_1, h_2, \dots, h_n$, 其中共有 n 个时间段。

4、根据权利要求 3 所述的基于长短期记忆模型的疾病预测方法, 其特征在于, 所述将多个预设的时间段对应的隐藏状态向量构成隐藏状态向量序列 $h_1, h_2, \dots, h_n$, 其中共有 n 个时间段的步骤, 包括:

根据公式: $c_i = \sum_{j=1}^n \alpha_{ij} h_j$, $\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^n \exp(e_{ik})}$, $e_{ij} = \text{score}(s_i, h_j)$, 获取所述第一长短期记忆网络中最终隐藏状态向量 c_i , α_{ij} 为权重参数, 其中共有 n 个时间段, s_i 为所述第二长短期记忆网络中的第 i 个隐藏状态向量, $\text{score}(s_i, h_j)$ 指采用预设的 score 函数根据 s_i 和 h_j 计算出的分数;

将多个预设的时间段对应的最终隐藏状态向量构成隐藏状态向量序列 $c_1, c_2, \dots, c_n$ 。

5、根据权利要求 1 所述的基于长短期记忆模型的疾病预测方法, 其特征在于, 所述将所述隐藏状态向量序列输入所述第二长短期记忆网络中运算, 得到疾病预测结果, 其中所述疾病预测结果包括预测疾病类型与对应的发病机率的步骤, 包括:

将所述隐藏状态向量序列输入所述第二长短期记忆网络中运算, 从而得到所述第二长短期记忆网络输出的高维向量序列;

根据预设的分向量与预测结果含义对应关系, 解读所述高维向量序列, 从而得到在将来不同时间段内的疾病预测结果, 其中所述疾病预测结果包括预测疾病类型与对应的发病机率。

6、根据权利要求 1 所述的基于长短期记忆模型的疾病预测方法, 其特征在于, 所述将所述隐藏状态向量序列输入所述第二长短期记忆网络中运算, 得到疾病预测结果, 其中所述疾病预测结果包括预测疾病类型与对应的发病机率的步骤之后, 包括:

接收输入的多个改善因素组, 并将所述改善因素组与第一医疗数据和第二医疗数据输入已训练完成的长短期记忆模型中进行计算, 其中所述改善因素组包括在指定时间点进行用药或者手术;

获取所述长短期记忆模型输出的与所述多个改善因素组分别对应的多组改善疾病预测结果, 其中所述改善疾病预测结果包括预测疾病类型与对应的发病机率;

根据预设的选择规则, 从所述多组改善疾病预测结果中筛选出最终改善疾病预测结果, 并生成治疗推荐方案, 其中所述治疗推荐方案附有所述最终改善疾病预测结果对应的改善因素组。

7、根据权利要求 1 所述的基于长短期记忆模型的疾病预测方法, 其特征在于, 所述疾病关联网为知识图谱网络, 所述根据预设的疾病关联网获取与所述指定疾病直接连接的关联疾病, 其中关联网

络的网络节点为不同的疾病类型的步骤之前,包括:

采用预设的知识图谱构建工具从预先收集的指定信息中识别出初始实体,其中所述指定信息至少记载了所述指定疾病,所述初始实体至少包括所述指定疾病;

对所述初始实体进行去重处理,从而获取最终实体;

从所述指定信息中提取出最终实体之间的关系,从而形成三元组,并依据所述三元组生成所述知识图谱网络。

8、一种基于长短期记忆模型的疾病预测装置,其特征在于,包括:

医疗数据获取单元,用于获取目标对象的第一医疗数据和关联对象的第二医疗数据,其中所述目标对象与关联对象存在血缘关系,第一医疗数据包括用药史、疾病史和手术史,第二医疗数据包括遗传病治疗史;

隐藏状态向量序列获取单元,用于将第一医疗数据和第二医疗数据输入已训练完成的长短期记忆模型中的第一长短期记忆网络中运算,得到所述第一长短期记忆网络中的隐藏状态向量序列,所述长短期记忆模型包括用于编码的第一长短期记忆网络和用于解码的第二长短期记忆网络;

疾病预测结果获取单元,用于将所述隐藏状态向量序列输入所述第二长短期记忆网络中运算,得到疾病预测结果,其中所述疾病预测结果包括预测疾病类型与对应的发病机率;

关联疾病获取单元,用于从所述疾病预测结果中筛选出发病机率高于预设阈值的预测疾病,记为指定疾病,并根据预设的疾病关联网络获取与所述指定疾病直接连接的关联疾病,其中关联网络的网络节点为不同的疾病类型;

输出单元,用于输出所述疾病预测结果与所述关联疾病。

9、根据权利要求8所述的基于长短期记忆模型的疾病预测装置,其特征在于,所述隐藏状态向量序列获取单元,包括:

多段数据序列获取子单元,用于将所述第一医疗数据根据预设的时间段分为多段数据序列;

指定影响因子获取子单元,用于根据预设的遗传病与其他疾病的影响因子对应关系,获取所述第二医疗数据中的遗传病对其他疾病的指定影响因子;

隐藏状态向量序列获取子单元,用于将所述多段数据序列与所述指定影响因子输入已训练完成的长短期记忆模型中的第一长短期记忆网络中运算,得到所述第一长短期记忆网络中的隐藏状态向量序列。

10、根据权利要求9所述的基于长短期记忆模型的疾病预测装置,其特征在于,所述隐藏状态向量

序列获取子单元,包括:

隐藏状态向量计算模块,用于根据公式: $h_t = LSTM_{enc}(x_t, h_{t-1})$, 获得所述第一长短期记忆网络中的隐藏状态向量 h_t , 其中 t 为第 t 个时间段, h_t 为对应于第 t 个时间段的隐藏状态向量, h_{t-1} 为对应于第 $t-1$ 个时间段的隐藏状态向量, X_t 为第 t 个时间段的输入数据, $LSTM_{enc}$ 指利用第一长短期记忆网络进行编码运算, 其中 X_t 包括第 t 个时间段的第一医疗数据和第 t 个时间段的指定影响因子;

隐藏状态向量序列获取模块,用于将多个预设的时间段对应的隐藏状态向量构成隐藏状态向量序列 $h_1, h_2, \dots, h_n$, 其中共有 n 个时间段。

11、根据权利要求 10 所述的基于长短期记忆模型的疾病预测装置,其特征在于,所述隐藏状态向量序列获取模块,包括:

最终隐藏状态向量获取子模块,用于根据公式: $c_i = \sum_{j=1}^n \alpha_{ij} h_j$, $\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^n \exp(e_{ik})}$, $e_{ij} = score(s_i, h_j)$, 获取所述第一长短期记忆网络中最终隐藏状态向量 c_i , α_{ij} 为权重参数, 其中共有 n 个时间段, s_i 为所述第二长短期记忆网络中的第 i 个隐藏状态向量, $score(s_i, h_j)$ 指采用预设的 $score$ 函数根据 s_i 和 h_j 计算出的分数;

隐藏状态向量序列获取子模块,用于将多个预设的时间段对应的最终隐藏状态向量构成隐藏状态向量序列 $c_1, c_2, \dots, c_n$ 。

12、根据权利要求 8 所述的基于长短期记忆模型的疾病预测装置,其特征在于,所述疾病预测结果获取单元,包括:

高维向量序列获取子单元,用于将所述隐藏状态向量序列输入所述第二长短期记忆网络中运算,从而得到所述第二长短期记忆网络输出的高维向量序列;

疾病预测结果获取子单元,用于根据预设的分向量与预测结果含义对应关系,解读所述高维向量序列,从而得到在将来不同时间段内的疾病预测结果,其中所述疾病预测结果包括预测疾病类型与对应的发病机率。

13、根据权利要求 8 所述的基于长短期记忆模型的疾病预测装置,其特征在于,所述装置,包括:

改善因素组接收单元,用于接收输入的多个改善因素组,并将所述改善因素组与第一医疗数据和第二医疗数据输入已训练完成的长短期记忆模型中进行计算,其中所述改善因素组包括在指定时间点进行用药或者手术;

改善疾病预测结果获取单元,用于获取所述长短期记忆模型输出的与所述多个改善因素组分别对应

的多组改善疾病预测结果,其中所述改善疾病预测结果包括预测疾病类型与对应的发病机率;

生成治疗推荐方案单元,用于根据预设的选择规则,从所述多组改善疾病预测结果中筛选出最终改善疾病预测结果,并生成治疗推荐方案,其中所述治疗推荐方案附有所述最终改善疾病预测结果对应的改善因素组。

14、根据权利要求 8 所述的基于长短期记忆模型的疾病预测装置,其特征在于,所述疾病关联网为知识图谱网络,所述装置,包括:

初始实体识别单元,用于采用预设的知识图谱构建工具从预先收集的指定信息中识别出初始实体,其中所述指定信息至少记载了所述指定疾病,所述初始实体至少包括所述指定疾病;

最终实体获取单元,用于对所述初始实体进行去重处理,从而获取最终实体;

知识图谱网络生成单元,用于从所述指定信息中提取出最终实体之间的关系,从而形成三元组,并依据所述三元组生成所述知识图谱网络。

15、一种计算机设备,包括存储器和处理器,所述存储器存储有计算机程序,其特征在于,所述处理器执行所述计算机程序时实现基于长短期记忆模型的疾病预测方法,所述基于长短期记忆模型的疾病预测方法,包括:

获取目标对象的第一医疗数据和关联对象的第二医疗数据,其中所述目标对象与关联对象存在血缘关系,第一医疗数据包括用药史、疾病史和手术史,第二医疗数据包括遗传病治疗史;

将第一医疗数据和第二医疗数据输入已训练完成的长短期记忆模型中的第一长短期记忆网络中运算,得到所述第一长短期记忆网络中的隐藏状态向量序列,所述长短期记忆模型包括用于编码的第一长短期记忆网络和用于解码的第二长短期记忆网络;

将所述隐藏状态向量序列输入所述第二长短期记忆网络中运算,得到疾病预测结果,其中所述疾病预测结果包括预测疾病类型与对应的发病机率;

从所述疾病预测结果中筛选出发病机率高于预设阈值的预测疾病,记为指定疾病,并根据预设的疾病关联网获取与所述指定疾病直接连接的关联疾病,其中关联网的网络节点为不同的疾病类型;

输出所述疾病预测结果与所述关联疾病。

16、根据权利要求 15 所述的计算机设备,其特征在于,所述将第一医疗数据和第二医疗数据输入已训练完成的长短期记忆模型中的第一长短期记忆网络中运算,得到所述第一长短期记忆网络中的隐藏状态向量序列的步骤,包括:

将所述第一医疗数据根据预设的时间段分为多段数据序列；

根据预设的遗传病与其他疾病的影响因子对应关系，获取所述第二医疗数据中的遗传病对其他疾病的指定影响因子；

将所述多段数据序列与所述指定影响因子输入已训练完成的长短期记忆模型中的第一长短期记忆网络中运算，得到所述第一长短期记忆网络中的隐藏状态向量序列。

17、根据权利要求16所述的计算机设备，其特征在于，所述将所述多段数据序列与所述指定影响因子输入已训练完成的长短期记忆模型中的第一长短期记忆网络中运算，得到所述第一长短期记忆网络中的隐藏状态向量序列的步骤，包括：

根据公式： $h_t = LSTM_{enc}(x_t, h_{t-1})$ ，获得所述第一长短期记忆网络中的隐藏状态向量 h_t ，其中 t 为第 t 个时间段， h_t 为对应于第 t 个时间段的隐藏状态向量， h_{t-1} 为对应于第 $t-1$ 个时间段的隐藏状态向量， X_t 为第 t 个时间段的输入数据， $LSTM_{enc}$ 指利用第一长短期记忆网络进行编码运算，其中 X_t 包括第 t 个时间段的第一医疗数据和第 t 个时间段的指定影响因子；

将多个预设的时间段对应的隐藏状态向量构成隐藏状态向量序列 $h_1, h_2, \dots, h_n$ ，其中共有 n 个时间段。

18、一种计算机可读存储介质，其上存储有计算机程序，其特征在于，所述计算机程序被处理器执行时实现基于长短期记忆模型的疾病预测方法，所述基于长短期记忆模型的疾病预测方法，包括：

获取目标对象的第一医疗数据和关联对象的第二医疗数据，其中所述目标对象与关联对象存在血缘关系，第一医疗数据包括用药史、疾病史和手术史，第二医疗数据包括遗传病治疗史；

将第一医疗数据和第二医疗数据输入已训练完成的长短期记忆模型中的第一长短期记忆网络中运算，得到所述第一长短期记忆网络中的隐藏状态向量序列，所述长短期记忆模型包括用于编码的第一长短期记忆网络和用于解码的第二长短期记忆网络；

将所述隐藏状态向量序列输入所述第二长短期记忆网络中运算，得到疾病预测结果，其中所述疾病预测结果包括预测疾病类型与对应的发病机率；

从所述疾病预测结果中筛选出发病机率高于预设阈值的预测疾病，记为指定疾病，并根据预设的疾病关联网络获取与所述指定疾病直接连接的关联疾病，其中关联网络的网络节点为不同的疾病类型；

输出所述疾病预测结果与所述关联疾病。

19、根据权利要求18所述的计算机可读存储介质，其特征在于，所述将第一医疗数据和第二医疗

数据输入已训练完成的长短期记忆模型中的第一长短期记忆网络中运算,得到所述第一长短期记忆网络中的隐藏状态向量序列的步骤,包括:

将所述第一医疗数据根据预设的时间段分为多段数据序列;

根据预设的遗传病与其他疾病的影响因子对应关系,获取所述第二医疗数据中的遗传病对其他疾病的指定影响因子;

将所述多段数据序列与所述指定影响因子输入已训练完成的长短期记忆模型中的第一长短期记忆网络中运算,得到所述第一长短期记忆网络中的隐藏状态向量序列。

20、根据权利要求 19 所述的计算机可读存储介质,其特征在于,所述将所述多段数据序列与所述指定影响因子输入已训练完成的长短期记忆模型中的第一长短期记忆网络中运算,得到所述第一长短期记忆网络中的隐藏状态向量序列的步骤,包括:

根据公式: $h_t = LSTM_{enc}(x_t, h_{t-1})$, 获得所述第一长短期记忆网络中的隐藏状态向量 h_t , 其中 t 为第 t 个时间段, h_t 为对应于第 t 个时间段的隐藏状态向量, h_{t-1} 为对应于第 $t-1$ 个时间段的隐藏状态向量, X_t 为第 t 个时间段的输入数据, $LSTM_{enc}$ 指利用第一长短期记忆网络进行编码运算, 其中 X_t 包括第 t 个时间段的第一医疗数据和第 t 个时间段的指定影响因子;

将多个预设的时间段对应的隐藏状态向量构成隐藏状态向量序列 $h_1, h_2, \dots, h_n$, 其中共有 n 个时间段。