

基于长短期记忆模型的疾病预测方法、装置和计算机设备

本申请要求于2019年6月27日提交中国专利局、申请号为201910570055.9，发明名称为“基于长短期记忆模型的疾病预测方法、装置和计算机设备”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本申请涉及到计算机领域，特别是涉及到一种基于长短期记忆模型的疾病预测方法、装置、计算机设备和存储介质。

背景技术

疾病风险预测是人工智能和医学结合产生的，它的核心是预测未来一段时间内患某种疾病的概率，主要的计算方法包括经典的回归分析，传统的机器学习方法和新兴的深度学习方法。但是真实世界的医学数据质量差，数据维度高，数据不平衡和医学数据不连续的时序性，导致准确地预测疾病的风险难度很大。现有的疾病风险预测系统需要患者的体检数据，但是体检过程耗时费力；同时现有产品只能分析单一疾病的风险，不能考虑疾病之间的关联；此外，现有产品的风险预测只能预测未来一段时间的患病风险，不能考虑输入变量的时间信息；并且，现有的疾病预测只考虑被测对象的医疗数据，预测准确度有待提高；现有的疾病预测产品仅采用单一的预测模型，预测准确度不足。

技术问题

本申请的主要目的为提供一种基于长短期记忆模型的疾病预测方法、装置、计算机设备和存储介质，旨在提高疾病预测的准确度。

技术方案

为了实现上述发明目的，本申请提出一种基于长短期记忆模型的疾病预测方法，包括以下步骤：

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。

本申请提供一种基于长短期记忆模型的疾病预测装置，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。

本申请提供一种计算机设备，包括存储器和处理器，所述存储器存储有计算机程序，所述处理器执行所述计算机程序时实现上述任一项所述方法的步骤。

本申请提供一种计算机可读存储介质，其上存储有计算机程序，所述计算机程序被处理器执行时实现上述任一项所述的方法的步骤。

有益效果

本申请的基于长短期记忆模型的疾病预测方法、装置、计算机设备和存储介质，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；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；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；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；输出所述疾病预测结果与所述关联疾病，从而提高预测的准确性。

附图说明

图1 为本申请一实施例的基于长短期记忆模型的疾病预测方法的流程示意图；

图2 为本申请一实施例的基于长短期记忆模型的疾病预测装置的结构示意框图；

图3 为本申请一实施例的计算机设备的结构示意框图。

本申请目的的实现、功能特点及优点将结合实施例，参照附图做进一步说明。

本发明的最佳实施方式

为了使本发明的目的、技术方案及优点更加清楚明白，以下结合附图及实施例，对本发明进行进一步详细说明。应当理解，此处描述的具体实施例仅仅用以解释本发明，并不用于限定本发明。

参照图 1，本发明实施例提供一种基于长短期记忆模型的疾病预测方法，包括以下步骤：

S1、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；

S2、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；

S3、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；

S4、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；

S5、输出所述疾病预测结果与所述关联疾病。

如上述步骤 S1 所述，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。本发明用于预测目标对象的疾病，而关联对象的第二医疗数据用于辅助预测目标对象的疾病，第二医疗数据包括遗传病治疗史。由于遗传病与血缘相关联，具有血缘关系的对象即使没有表现出显现的遗传病特征，也可能具有隐性的生理特征（隐疾），因此关联对象的遗传病治疗史有助于预测目标对象的疾病。其中第一医疗数据包括用药史、疾病史和手术史，由于用药史、疾病史和手术史会对人体具有影响，可以作为疾病预测的依据。例如，对于曾经采用吡格列酮、卡托普利和尼群地平作为治疗糖尿病、高血压和房颤的历史用药的病人，其在未来有可能患有心肌梗、冠心病、脑卒中等风险。传统技术仅采取一对一分析的策略，即仅考虑目标对象的医疗数据对目标对象将来的疾病进行预测，而本发明还采用关联对象的第二医疗数据作为预测用数据，以增加预测准确度。并且传统技术仅对单一疾病进行建模，因此只能预测单一疾病，而本发明能实现多种疾病在不同时间区段内的预测。并且本发明的医疗数据不仅包括目标对象的第一医疗数据，还考虑关联对象的第二医疗，使预测更精准，且克服了目标对象谎报病史、隐瞒家族病史的缺陷（即即使目标对象身上的遗传病为隐性表示，但在基因层面上也是存在的，在将来可能呈显性表性，并天也可能存在对身体机能、其他疾病产生影响，因此通过关联对象的第二医疗数据，以获知目标对象的遗传病等数据，从而提高疾病预测的准确性）。

如上述步骤 S2 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, 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。长短期记忆模型是使用长短期记忆网络的模型, 其中长短期记忆网络是一种时间递归神经网络, 适合于处理和预测时间序列中间隔和延迟相对较长的重要事件, 相对于普通的循环神经网络, 其加入了一个判断信息有用与否的“处理器”, 只有符合算法认证的信息才会留下, 不符的信息则通过遗忘门被遗忘, 从而解决了长序依赖问题。本申请采用的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, 以实现多种疾病的时序性预测。其中, 本申请中的编码指将输入的信息转换为指定长度的向量序列, 解码指将输入的向量序列转换为预测的向量序列。

如上述步骤 S3 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。其中第二长短期记忆网络可采用任意的方法进行运算, 例如采用公式: $\hat{s}_t = \tanh(W_c[c_i; s_t])$, $c_i = \sum_{j=1}^n \alpha_{ij} h_j$, $\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^n \exp(e_{ik})}$, $e_{ij} = \text{score}(s_i, h_j)$, $p(y_t | y < t, x) = \text{softmax}(W_s \hat{s}_t)$, 其中 c_i 为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, W_c 为权值, p 为输出概率, y_t 为第二长短期记忆网络的对应第 t 个时间段的输出, x 为输入(与第一医疗数据和第二医疗数据直接相关)。

如上述步骤 S4 所述, 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。其中, 所述疾病关联网络可以为任意关联网络, 例如为知识图谱网络, 所述知识图谱网络的构建方式例如为: 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。从而在长短期记忆模型的基础上, 进一步进行预测, 进一步提高预测的准确性。

如上述步骤 S5 所述, 输出所述疾病预测结果与所述关联疾病。其中所述疾病预测结果是所述长短期记忆模型的输出结果, 关联疾病是疾病关联网络的输出结果, 从而通过结合所述长短期记忆模型和疾病关联网络, 进一步提高预测的准确性。

在一个实施方式中, 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 S2, 包括:

S201、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；

S202、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；

S203、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。

如上所述，实现了得到所述第一长短期记忆网络中的隐藏状态向量序列。其中所述指定影响因子在不同时间段中取值相同或者取值动态变化（因为在随着时间的变化，遗传病对其他疾病的影响程度也随之变化）。其中第一医疗数据是一段时间内的医疗数据，将其分为不同时间段内的多段数据序列，并且为了将第二医疗数据容纳入长短期记忆模型中，本申请将第二医疗数据反应为指定影响因子，从而将所述多段数据序列与所述指定影响因子共同作为长短期记忆模型的输入。具体地，单段数据与对应的指定影响因子将生成一个高维向量，因此多段数据序列与分别对应的指定影响因子将生成高维向量序列，从而作为已训练完成的长短期记忆模型中的计算基础。其他疾病指除所述遗传病之外的疾病。所述影响因子指将遗传病对其他疾病的影响数值化得到的数据，用于在长短期记忆网络进行计算，所述影响因子可以任意形式存在，例如以分离的向量形式。

在一个实施方式中，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 S203，包括：

S2031、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；

S2032、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。

如上所述，实现了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。本申请采用公式： $h_t = LSTM_{enc}(x_t, h_{t-1})$ ，获得所述第一长短期记忆网络中的隐藏状态向量 h_t ，并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。据此，第一长短期记忆网络将所述多段数据序列与所述指定影响因子编码为隐藏状态向量序列，作为第二长短期记忆网络的解码基础。

在一个实施方式中,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 S2032, 包括:

S20321、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;

S20322、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 $c_1, c_2, \dots, c_n$ 。

如上所述, 实现了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 $h_1, h_2, \dots, h_n$ 。本申请根据公式: $c_i = \sum_{j=1}^n \alpha_{ij} h_j$, $\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^n \exp(e_{ik})}$, $e_{ij} = \text{score}(s_i, h_j)$, 获取所述第一长短期记忆网络中最终隐藏状态向量 c_i , 也即是引入了注意力机制自动捕捉对结局重要的信息, 从而将最终隐藏状态向量序列作为第二长短期记忆网络的解码基础。由于采用了注意力机制, 使得权重分配更为准确, 有利于改善预测的准确性。

在一个实施方式中, 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 S3, 包括:

S301、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;

S302、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。

如上所述, 实现了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。其中所述第二长短期记忆网络输出的是高维向量序列, 所述高维向量序列代表了不同时间段的预测结果, 其中的高维向量的分向量代表了预测的疾病类型、对应的患病机率。根据预设的分向量与预测结果含义对应关系, 即可获知将来不同时间段内的预测疾病类型与对应的发病机率。

在一个实施方式中, 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 S3 之后, 包括:

S311、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;

S312、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;

S313、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。

如上所述，实现了生成治疗推荐方案。由于本申请的输入为包括用药史和手术史的医疗数据，因此也可以接受假设的用药或者手术，进而本申请的长短期记忆模型可以模拟治疗方案的疗效。因此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。其中预设的选择规则例如为预测的疾病类型最少，或者预测的疾病类型的发病率均小于预设的阈值等。

在一个实施方式中，所述疾病关联网络为知识图谱网络，所述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的步骤 S4 之前，包括：

S321、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；

S322、对所述初始实体进行去重处理，从而获取最终实体；

S323、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。

如上所述，实现了构建所述包括所述指定成员的知识图谱。其中预设的知识图谱构建工具可以为任意工具，任如为现有的 SPSS、UcinetNetDraw、VOSviewer 等等，由于上述工具是现有的知识图谱构建工具，因此不再赘述。其中指定信息记载了疾病信息，据此可以获知疾病之间的关联关系。其中所述实体即是知识图谱中的知识节点，初始实体指未过去重处理的知识节点。识别出初始实体的过程例如为：对指定信息进行分词处理，从而获得由多个词构成的词序列，将所述词序列输入预设的语句结构模型，从而在所述词序列中获取初始实体。再对所述初始实体进行去重处理，从而获取最终实体。去重处理的过程例如为：对所有初始实体进行同义词判断，将属于同一同义词组的初始实体替换为所述同义词组中的一个词汇。再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包括所述指定成员的知识图谱。其中三元组例如指两个实体之间的关系。其中，所述从所述指定信息中提取出最终实体之间的关系的方法例如为：将所述指定信息套入预设的语句结构，从而通过所述语句结构将表述多个实体间的关系的词汇提取出来。据此，将疾病之间的相互关系以知识图谱的网络形式

表示出来,其中疾病类型作为知识图谱中的一个知识节点。进一步地,所述知识图谱的知识节点还可以包括非疾病类型的实体。

本申请的基于长短期记忆模型的疾病预测方法,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;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;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;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;输出所述疾病预测结果与所述关联疾病,从而提高预测的准确性。

参照图2,本申请实施例提供一种基于长短期记忆模型的疾病预测装置,包括:

医疗数据获取单元10,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20,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30,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40,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50,用于输出所述疾病预测结果与所述关联疾病。

其中上述单元分别用于执行的操作与前述实施方式的基于长短期记忆模型的疾病预测方法的步骤一一对应,在此不再赘述。

在一个实施方式中,所述隐藏状态向量序列获取单元20,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。

其中上述子单元分别用于执行的操作与前述实施方式的基于长短期记忆模型的疾病预测方法的步骤一一对应,在此不再赘述。

在一个实施方式中,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。

其中上述模块分别用于执行的操作与前述实施方式的基于长短期记忆模型的疾病预测方法的步骤一一对应,在此不再赘述。

在一个实施方式中,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 $c_1, c_2, \dots, c_n$ 。

其中上述子模块分别用于执行的操作与前述实施方式的基于长短期记忆模型的疾病预测方法的步骤一一对应,在此不再赘述。

在一个实施方式中,所述疾病预测结果获取单元 30,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。

其中上述子单元分别用于执行的操作与前述实施方式的基于长短期记忆模型的疾病预测方法的步骤一一对应,在此不再赘述。

在一个实施方式中,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。

其中上述单元分别用于执行的操作与前述实施方式的基于长短期记忆模型的疾病预测方法的步骤一一对应,在此不再赘述。

在一个实施方式中,所述疾病关联网络为知识图谱网络,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。

其中上述单元分别用于执行的操作与前述实施方式的基于长短期记忆模型的疾病预测方法的步骤一一对应,在此不再赘述。

本申请的基于长短期记忆模型的疾病预测装置,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;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;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;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;输出所述

疾病预测结果与所述关联疾病，从而提高预测的准确性。

参照图3，本发明实施例中还提供一种计算机设备，该计算机设备可以是服务器，其内部结构可以如图所示。该计算机设备包括通过系统总线连接的处理器、存储器、网络接口和数据库。其中，该计算机设计的处理器用于提供计算和控制能力。该计算机设备的存储器包括非易失性存储介质、内存储器。该非易失性存储介质存储有操作系统、计算机程序和数据库。该内存储器为非易失性存储介质中的操作系统和计算机程序的运行提供环境。该计算机设备的数据库用于存储基于长短期记忆模型的疾病预测方法所用数据。该计算机设备的网络接口用于与外部的终端通过网络连接通信。该计算机程序被处理器执行时以实现一种基于长短期记忆模型的疾病预测方法。

上述处理器执行上述基于长短期记忆模型的疾病预测方法，其中所述方法包括的步骤分别与执行前述实施方式的基于长短期记忆模型的疾病预测方法的步骤一一对应，在此不再赘述。

本领域技术人员可以理解，图中示出的结构，仅仅是与本发明方案相关的部分结构的框图，并不构成对本发明方案所应用于其上的计算机设备的限定。

本发明一实施例还提供一种计算机可读存储介质，其上存储有计算机程序，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，其中所述方法包括的步骤分别与执行前述实施方式的基于长短期记忆模型的疾病预测方法的步骤一一对应，在此不再赘述。

本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程，是可以通过计算机程序来指令相关的硬件来完成，所述的计算机程序可存储于一非易失性计算机可读存储介质中，该计算机程序在执行时，可包括如上述各方法的实施例的流程。其中，本发明所提供的和实施例中所使用的对存储器、存储、数据库或其它介质的任何引用，均可包括非易失性和/或易失性存储器。非易失性存储器可以包括只读存储器（ROM）、可编程 ROM（PROM）、电可编程 ROM（EPROM）、电可擦除可编程 ROM（EEPROM）或闪存。易失性存储器可包括随机存取存储器（RAM）或者外部高速缓冲存储器。作为说明而非局限，RAM 以多种形式可得，诸如静态 RAM（SRAM）、动态 RAM（DRAM）、同步 DRAM（SDRAM）、双速据率 SDRAM（SSRSDRAM）、增强型 SDRAM（ESDRAM）、同步链路（Synchlink）DRAM（SLDRAM）、存储器总线（Rambus）直接 RAM（RDRAM）、直接存储器总线动态 RAM（DRDRAM）、以及存储器总线动态 RAM（RDRAM）等。