



(12)发明专利

(10)授权公告号 CN 106295186 B

(45)授权公告日 2019.03.15

(21)申请号 201610658754.5

(22)申请日 2016.08.11

(65)同一申请的已公布的文献号

申请公布号 CN 106295186 A

(43)申请公布日 2017.01.04

(73)专利权人 中国科学院计算技术研究所

地址 100080 北京市海淀区中关村科学院南路6号

(72)发明人 刘鹏鹤 张书涵 孙晓平 孙毓忠

(74)专利代理机构 北京律诚同业知识产权代理有限公司 11006

代理人 祁建国 梁挥

(51)Int.Cl.

G16H 50/20(2018.01)

(56)对比文件

CN 102110192 A,2011.06.29,

CN 104866979 A,2015.08.26,

CN 105260598 A,2016.01.20,

CN 105808931 A,2016.07.27,

CN 105608091 A,2016.05.25,

侯桂英等.应用智能医疗诊断系统诊断高血压的研究.《哈尔滨医科大学学报》.2003,第37卷(第3期),摘要,正文第1-3节.

李蕾等.基于案例与模糊推理的中医诊断系统研究.《信阳师范学院学报:自然科学版》.2014,第27卷(第4期),第585-588页.

Riccardo Miotto等.Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records.《Scientific Reports》.2016,第1-10页.

Ms.Rupali R.Patil.Heart Disease Prediction System using Naive Bayes and Jelinek-mercer smoothing.《International Journal of Advanced Research in Computer and Communication Engineering》.2014,第3卷(第5期),第6787-6792页.

(续)

审查员 何守兵

权利要求书1页 说明书10页 附图2页

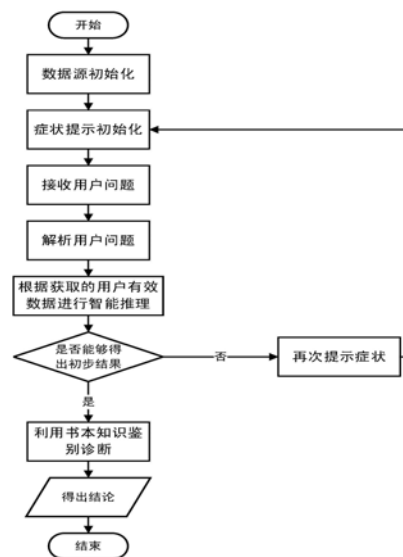
(54)发明名称

一种基于智能推理的辅助疾病诊断的系统

(57)摘要

本发明提出一种基于智能推理的辅助疾病诊断的方法及系统,涉及计算机领域,该方法包括获取用户的输入信息并进行解析,将输入信息中的有效信息进行存储,计算所述有效信息在病例库中的信息匹配度以及在规则库中的规则覆盖率,对所述信息匹配度与所述规则覆盖率进行加权求和,获取疾病库中每种疾病对所述输入信息的解释值p;按照所述解释值p,将与所述解释值p相对应的疾病进行排序,将所述解释值p大于等于阈值E的疾病作为初始推理结论返回给用户,获取用户输入的检查结果信息并结合所述有效信息,将书本知识库中与所述初始推理结论中的疾病相对应的诊断标准显示给所述用户。本发明模拟临床诊断真实场景下问诊的过程和形式,

能够较好的解释确诊结果。



CN 106295186 B

[接上页]

**(56)对比文件**

向艳等.基于模糊推理模型的专家系统的研究与应用.《计算机工程》.2005,第31卷(第10期),第180-181、187页.

陈再旺等.一个医疗辅助诊断专家系统的设计与实现.《计算机系统应用》.2001,(第12期),第50-51、34页.

杨丽等.基于案例推理的中医临床诊疗决策支持系统.《世界科学技术—中医药现代化》.2014,第16卷(第3期),正文第1节.

王靖飞等.基于专家思维的动物疾病诊断专家系统知识组织及推理策略研究.《黑龙江畜牧兽医》.2004,(第10期),第5-7页.

1. 一种基于智能推理的辅助疾病诊断的系统,其特征在于,包括

获取解释值模块,用于获取用户的输入信息并进行解析,将所述输入信息中的有效信息进行存储,计算所述有效信息在病例库中的信息匹配度以及在规则库中的规则覆盖率,对所述信息匹配度与所述规则覆盖率进行加权求和,获取疾病库中每种疾病对所述输入信息的解释值 $p$ ;

显示诊断标准模块,用于按照所述解释值 $p$ ,将与所述解释值 $p$ 相对应的疾病进行排序,将所述解释值 $p$ 大于等于阈值 $E$ 的疾病作为初始推理结论返回给用户,获取用户输入的检查结果信息并结合所述有效信息,将书本知识库中与所述初始推理结论中的疾病相对应的诊断标准显示给所述用户;

其中计算所述信息匹配度的公式为

$$m = \sum_{r=1} \lambda_{X_r} \cdot \left( \sum_{i=1} \sum_{j=1} \omega_{X_{ri}} \cdot g(X_{ri}, Y_j) \right)$$

其中, $m$ 为所述信息匹配度, $X$ 为病例知识图谱, $X_r$ 为处于 $X$ 中级别为 $r$ 的三元组集合, $\lambda_{X_r}$ 为级别为 $r$ 的三元组集合的系数, $X_{ri}$ 为级别为 $r$ 的三元组集合中的第 $i$ 个三元组, $\omega_{X_{ri}}$ 为级别为 $r$ 的三元组集合中的第 $i$ 个三元组的权重, $Y$ 为输入信息的三元组集合, $Y_j$ 为输入信息的三元组集合中的第 $j$ 个三元组。

2. 如权利要求1所述的基于智能推理的辅助疾病诊断的系统,其特征在于,计算所述规则覆盖率的公式为

$$o = \sum_i \sum_j \theta_{R_i} \cdot f(R_i, Y_j)$$

其中, $o$ 为规则覆盖率, $R$ 为疾病的规则集合, $R_i$ 为规则集合中的第 $i$ 条规则, $\theta_{R_i}$ 代表第 $i$ 条规则的权重, $Y$ 为输入信息的三元组集合, $Y_j$ 表示输入信息的三元组集合中的第 $j$ 个三元组, $f$ 为标识函数。

3. 如权利要求1所述的基于智能推理的辅助疾病诊断的系统,其特征在于,在所述获取解释值模块之前还包括,获取用户的病例知识图谱并进行初始化,获取所述病例知识图谱中时间段 $T$ 内的病例,抽取所述病例中最热门的 $h$ 个症状作为用户的初始提示症状。

4. 如权利要求1所述的基于智能推理的辅助疾病诊断的系统,其特征在于,还包括如果未存在疾病的解释值 $p$ 达到阈值 $E$ ,则选择 $k$ 个解释值 $p$ 对应的疾病,并在知识图谱中搜索除输入信息之外的权重大于阈值 $F$ 的三元组作为症状返回给用户,并返回所述获取解释值模块。

## 一种基于智能推理的辅助疾病诊断的系统

### 技术领域

[0001] 本发明涉及计算机领域,特别涉及一种基于智能推理的辅助疾病诊断的方法与系统。

### 背景技术

[0002] 疾病辅助诊断系统作为医生诊断的辅助工具,通过学习、继承和发扬医学专家的宝贵理论知识以及丰富的临床经验知识,借助信息技术和智能计算技术,在临床诊断过程中,根据患者当前的病症信息,依据系统知识库和推理分析计算机制,对病情进行分析提示,对诊断治疗方案决策提供辅助支持信息。智能辅助诊断系统可以帮助年轻无经验的医生在临床诊断决策过程中不遗漏、不错过重要的信息和线索,为疑难杂症寻找更多的解决方案。

[0003] 国内医疗辅助诊断系统的智能推理机制采用以下几种方式:第一种为基于产生式规则的推理机制,如文献(陈再旺和陈景长,“一个医疗辅助诊断专家系统的设计与实现,”计算机系统应用,pp.50-51,2001.)中构建了一个基于产生式规则以及冲突消解的急性腹痛辅助诊断专家系统,该系统采用正向推理以及深度优先结合剪枝的搜索策略,一步步匹配用户的输入直到寻找一个答案,然而基于规则的匹配有很大的缺陷,如果规则量级比较大,则匹配的过程的开销就会很大,规则的每一次维护与修改都需要专业人员的参与,不够灵活。基于案例推理的自动推理机制自提出以来受到了较大的关注,这种推理的主要思路是从案例库中学习出一种指导思想,之后再利用该指导思想去解决遇到的新的问题,如文献(杨丽,周雪忠,毕澜馨,张润顺,王映辉,刘保延,et al.,“基于案例推理的中医临床诊疗决策支持系统,”世界科学技术:中医药现代化,pp.474-480,2014.)基于临床实际的中医病例数据,提出一基于案例推理的中医临床诊疗决策支持系统,该系统从中医临床数据仓库中筛选加工形成中医临床效验案例库,通过计算真实案例与效验案例间的相似度来作智能诊断,然而基于案例推理的诊断方法也有其自身的局限性,例如如何有效表达真实的病例;如何使计算机像人一样去理解病例?还有一种常用的推理机制为使用神经网络来实现辅助诊断系统的智能推理,这得益于神经网络具有的联想记忆与联想映射以及不断学习的能力,如文献(蔡航,“基于神经网络的医疗诊断专家系统,”数理医药学杂志,vol.15,pp.294-295,2002.)中提出了一个基于神经网络的肺癌诊断专家系统,使用病例中的年龄、性别、家族史、既往史、吸烟史、X线检查、CT检查、纤维支气管镜检查等8项作为神经网络的输入,通过病历训练模型来根据特征输入判断是否患肺癌,又如文献(侯桂英,孙佰清,关振中,and冯英俊,“应用智能医疗诊断系统诊断高血压的研究,”哈尔滨医科大学学报,vol.37,pp.223-225,2003.)中,作者利用人工神经网络和专家系统相结合的方法,对100份高血压病例提取特征进行训练构建预测模型,再使用该模型根据高血压病症的特征进行诊断,然而,这两种专家诊断系统有个最明显的缺陷是神经系统的不可解释性,它并不能告诉用户其背后的推理过程,另外,这两个系统特征选取过于简单,在文献(蔡航,“基于神经网络的医疗诊断专家系统,”数理医药学杂志,vol.15,pp.294-295,2002.)中病例中的既往

史、X线检查等都只是二元变量(是/否),这很容易造成由于特征过于简单而出现拟合的情况,因而,基于神经系统的专家系统的功能还很有限,其目前只适用于解决规模较小的问题,且效果与性能明显受训练数据集限制,除了推理机制的不同外,目前国内大多数疾病诊断专家系统交互不够友好,如文献(陈再旺 and 陈景长,“一个医疗辅助诊断专家系统的设计与实现,”《计算机系统应用》,pp. 50-51,2001.;杨丽,周雪忠,毕澜馨,张润顺,王映辉,刘保延,et al.,“基于案例推理的中医临床诊疗决策支持系统,”《世界科学技术:中医药现代化》,pp. 474-480,2014.;侯桂英,孙佰清,关振中,and 冯英俊,“应用智能医疗诊断系统诊断高血压的研究,”《哈尔滨医科大学学报》,vol. 37,pp. 223-225,2003.)等并没有向用户呈现一个良好的交互过程,其通常需要用户有较高的专业知识并且需要用户一次性输入较多信息,而通常情况下用户并不能准确地描述自身的所有有效信息,同时,这些系统在一次性获取用户信息后往往直接将诊断结果返回给用户,缺少了与用户的交互过程,这非常不利于用户理解系统推理的过程以及自身患病的诊断流程。

### 发明内容

[0004] 针对现有技术的不足,本发明提出一种基于智能推理的辅助疾病诊断的方法与系统。

[0005] 本发明提供一种基于智能推理的辅助疾病诊断的方法,包括

[0006] 步骤1,获取用户的输入信息并进行解析,将所述输入信息中的有效信息进行存储,计算所述有效信息在病例库中的信息匹配度以及在规则库中的规则覆盖率,对所述信息匹配度与所述规则覆盖率进行加权求和,获取疾病库中每种疾病对所述输入信息的解释值 $p$ ;

[0007] 步骤2,按照所述解释值 $p$ ,将与所述解释值 $p$ 相对应的疾病进行排序,将所述解释值 $p$ 大于等于阈值 $E$ 的疾病作为初始推理结论返回给用户,获取用户输入的检查结果信息并结合所述有效信息,将书本知识库中与所述初始推理结论中的疾病相对应的诊断标准显示给所述用户。

[0008] 计算所述信息匹配度的公式为

$$[0009] \quad m = \sum_{r=1} \lambda_{X_r} \cdot \left( \sum_{i=1} \sum_{j=1} \omega_{X_{ri}} \cdot g(X_{ri}, Y_j) \right)$$

[0010] 其中, $m$ 为所述信息匹配度, $X$ 为病例知识图谱, $X_r$ 为处于 $X$ 中级别为 $r$ 的三元组集合, $\lambda_{X_r}$ 为级别为 $r$ 的三元组集合的系数, $X_{ri}$ 为级别为 $r$ 的三元组集合中的第 $i$ 个三元组, $\omega_{X_{ri}}$ 为级别为 $r$ 的三元组集合中的第 $i$ 个三元组的权重, $Y$ 为输入信息的三元组集合, $Y_j$ 为输入信息的三元组集合中的第 $j$ 个三元组。

[0011] 计算所述规则覆盖率的公式为

$$[0012] \quad o = \sum_i \sum_j \theta_{R_i} \cdot f(R_i, Y_j)$$

[0013] 其中, $o$ 为规则覆盖率, $R$ 为疾病的规则集合, $R_i$ 为规则集合中的第 $i$ 条规则, $\theta_{R_i}$ 代

表第*i*条规则的权重,*Y*为输入信息的三元组集合,*Y<sub>j</sub>*表示输入信息的三元组集合中的第*j*个三元组,*f*为标识函数。

[0014] 在所述步骤1之前还包括,获取用户的病例知识图谱并进行初始化,获取所述病例知识图谱中时间段*T*内的病例,抽取所述病例中最热门的*h*个症状作为用户的初始提示症状。

[0015] 还包括如果未存在疾病的解释值*p*达到阈值*E*,则选择*k*个解释值*p*对应的疾病,并在知识图谱中搜索除输入信息之外的权重大于阈值*F*的三元组作为症状返回给用户,并返回所述步骤1。

[0016] 本发明还提出一种基于智能推理的辅助疾病诊断的系统,包括

[0017] 获取解释值模块,用于获取用户的输入信息并进行解析,将所述输入信息中的有效信息进行存储,计算所述有效信息在病例库中的信息匹配度以及在规则库中的规则覆盖率,对所述信息匹配度与所述规则覆盖率进行加权求和,获取疾病库中每种疾病对所述输入信息的解释值*p*;

[0018] 显示诊断标准模块,用于按照所述解释值*p*,将与所述解释值*p*相对应的疾病进行排序,将所述解释值*p*大于等于阈值*E*的疾病作为初始推理结论返回给用户,获取用户输入的检查结果信息并结合所述有效信息,将书本知识库中与所述初始推理结论中的疾病相对应的诊断标准显示给所述用户。

[0019] 计算所述信息匹配度的公式为

$$[0020] \quad m = \sum_{r=1} \lambda_{X_r} \cdot \left( \sum_{i=1} \sum_{j=1} \omega_{X_{ri}} \cdot g(X_{ri}, Y_j) \right)$$

[0021] 其中,*m*为所述信息匹配度,*X*为病例知识图谱,*X<sub>r</sub>*为处于*X*中级别为*r*的三元组集合, $\lambda_{X_r}$ 为级别为*r*的三元组集合的系数,*X<sub>ri</sub>*为级别为*r*的三元组集合中的第*i*个三元组, $\omega_{X_{ri}}$ 为级别为*r*的三元组集合中的第*i*个三元组的权重,*Y*为输入信息的三元组集合,*Y<sub>j</sub>*为输入信息的三元组集合中的第*j*个三元组。

[0022] 计算所述规则覆盖率的公式为

$$[0023] \quad o = \sum_i \sum_j \theta_{R_i} \cdot f(R_i, Y_j)$$

[0024] 其中,*o*为规则覆盖率,*R*为疾病的规则集合,*R<sub>i</sub>*为规则集合中的第*i*条规则, $\theta_{R_i}$ 代表第*i*条规则的权重,*Y*为输入信息的三元组集合,*Y<sub>j</sub>*表示输入信息的三元组集合中的第*j*个三元组,*f*为标识函数。

[0025] 在所述获取解释值模块之前还包括,获取用户的病例知识图谱并进行初始化,获取所述病例知识图谱中时间段*T*内的病例,抽取所述病例中最热门的*h*个症状作为用户的初始提示症状。

[0026] 还包括如果未存在疾病的解释值*p*达到阈值*E*,则选择*k*个解释值*p*对应的疾病,并在知识图谱中搜索除输入信息之外的权重大于阈值*F*的三元组作为症状返回给用户,并返回所述获取解释值模块。

[0027] 由以上方案可知,本发明的优点在于:

[0028] 本发明在方法上首先采用基于带权重的疾病知识图谱搜索与规则推理相结合的方式根据用户信息给出用户初始推理结果,然后使用专业书本知识中的鉴别诊断方法为用户作最终的疾病诊断;在系统层面上,本发明的系统具有良好的用户交互界面,用户可以有多种输入方式,针对每种输入方式系统采用不同方法解析出用户输入的有效信息,并根据当前推理情况循序渐进的询问用户的进一步症状等,模拟临床诊断真实场景下问诊的过程和形式,系统最终确诊时也会将本次问诊的推理过程返回给用户,可以较好的解释确诊结果。

### 附图说明

[0029] 图1为本发明一个实施例的基于智能推理的辅助疾病诊断的系统框架图;

[0030] 图2为本发明一个实施例的基于智能推理的辅助疾病诊断的系统流程图。

### 具体实施方式

[0031] 本发明提供一种基于智能推理的辅助疾病诊断的方法与系统,用于根据获取的用户病症信息为其提供准确的疾病诊断分析结果。

[0032] 为实现上述目的,本发明提供了一种基于智能推理的辅助疾病诊断的方法,方法说明如下:

[0033] 步骤S1:初始化提示信息,使用潜在狄利克雷分布(LDA)主题模型,根据时间段T(最近一月或更长)的病例,抽取出当前最热门的h个症状作为用户的初始提示症状;

[0034] 步骤S2:接收用户的输入,允许用户采用不同的格式输入;

[0035] 步骤S3:解析用户的输入,储存用户的有效信息,有效信息包括用户的性别年龄,所患症状与症状的相关描述如症状持续时间等以及用户的检查信息;

[0036] 步骤S4:根据获得的用户有效信息,计算其在病例库中的信息匹配度以及在规则库中的规则覆盖率,信息匹配度的计算公式为:

$$[0037] \quad m = \sum_{r=1} \lambda_{X_r} \cdot \left( \sum_{i=1} \sum_{j=1} \omega_{X_{ri}} \cdot g(X_{ri}, Y_j) \right)$$

[0038] 其中m为信息匹配度,X代表某一疾病的病例知识图谱, $X_r$ 表示处于X中级别为r的三元组集合, $\lambda_{X_r}$ 表示级别为r的三元组集合的系数, $X_{ri}$ 表示级别为r的三元组集合中的第i个三元组, $\omega_{X_{ri}}$ 表示级别为r的三元组集合中的第i个三元组的权重,Y为用户信息的三元组集合, $Y_j$ 表示用户信息三元组集合中的第j个三元组,g为一个标志函数,即当满足一定条件时其函数值为1,否则函数值为0,这里表示当分别来自知识图谱中的某条三元组和来自用户信息集合中的某条三元组匹配时,g函数取值为1,否则取值0,即

$$[0039] \quad g(x, y) = \begin{cases} 1, & \text{if}(x = y) \\ 0, & \text{others} \end{cases}, \text{其中} x, y \text{表示来自不同集合的两个三元组。}$$

[0040] 规则覆盖率的计算公式为:

$$[0041] \quad o = \sum_i \sum_j \theta_{R_i} \cdot f(R_i, Y_j)$$

[0042] 其中 $o$ 为规则覆盖率, $R$ 表示某一疾病规则集合, $R_i$ 为规则集合中的第 $i$ 条规则, $\theta_{R_i}$ 代表第 $i$ 条规则的权重, $Y$ 为用户信息的三元组集合, $Y_j$ 表示用户信息三元组集合中的第 $j$ 个三元组,另外, $f$ 也是一个标志函数,即当满足一定条件时函数值为1,否则函数值为0,这里表示当来自用户信息集合中的某条三元组满足来自规则集合中的某个规则时, $f$ 函数取值为1,否则取值0,即

$$[0043] \quad f(x, y) = \begin{cases} 1, & \text{if}(x \text{ 满足 } y) \\ 0 & \text{others} \end{cases}, \text{其中 } x \text{ 表示来自用户信息集合中的某条三元组, } y$$

表示来自规则集合中的某条规则。

[0044] 步骤S5:对计算出的信息匹配度与规则覆盖率进行加权求和,求出疾病库中每种疾病对用户的输入信息的解释值 $p$ ,

$$[0045] \quad p = \mu m + \nu o$$

[0046]  $\mu, \nu$ 分别代表信息匹配度与规则覆盖率的系数,由实验给出,解释值表示某一疾病对用户信息的解释程度, $p$ 的值越大,表示该疾病越能解释用户的输入信息;

[0047] 步骤S6:按解释值 $p$ 由大到小的顺序对对应的疾病进行排序,如果没有疾病的解释值达到阈值 $E$ ,说明此时用户的有效信息还不足以作出判断,则进入步骤S7向用户推荐其他症状;如果有某种疾病的解释值达到阈值 $E$ ,则将前 $k$ 个疾病作为初始推理结论返回给用户,进入步骤S8;

[0048] 步骤S7:选择前 $k$ 个解释值对应的疾病的知识图谱中搜索除用户已输入症状之外的权重大于一阈值 $F$ 的三元组症状作为症状提示返回给用户,回到步骤 S2;

[0049] 步骤S8:提示用户输入检查结果信息,根据用户检查结果信息并结合用户已输入的有效信息,将书本知识库中与所述初始推理结论中的疾病相对应的诊断标准显示给所述用户;

[0050] 本发明还提出一种基于智能推理的辅助疾病诊断的系统,包括

[0051] 获取解释值模块,用于获取用户的输入信息并进行解析,将所述输入信息中的有效信息进行存储,计算所述有效信息在病例库中的信息匹配度以及在规则库中的规则覆盖率,对所述信息匹配度与所述规则覆盖率进行加权求和,获取疾病库中每种疾病对所述输入信息的解释值 $p$ ;

[0052] 显示诊断标准模块,用于按照所述解释值 $p$ ,将与所述解释值 $p$ 相对应的疾病进行排序,将所述解释值 $p$ 大于等于阈值 $E$ 的疾病作为初始推理结论返回给用户,获取用户输入的检查结果信息并结合所述有效信息,将书本知识库中与所述初始推理结论中的疾病相对应的诊断标准显示给所述用户。

[0053] 计算所述信息匹配度的公式为

$$[0054] \quad m = \sum_{r=1} \lambda_{X_r} \cdot \left( \sum_{i=1} \sum_{j=1} \omega_{X_{ri}} \cdot g(X_{ri}, Y_j) \right)$$

[0055] 其中, $m$ 为所述信息匹配度, $X$ 为病例知识图谱, $X_r$ 为处于 $X$ 中级别为 $r$ 的三元组集

合,  $\lambda_{X_r}$  为级别为  $r$  的三元组集合的系数,  $X_{ri}$  为级别为  $r$  的三元组集合中的第  $i$  个三元组,  $\omega_{X_{ri}}$  为级别为  $r$  的三元组集合中的第  $i$  个三元组的权重,  $Y$  为输入信息的三元组集合,  $Y_j$  为输入信息的三元组集合中的第  $j$  个三元组。

[0056] 计算所述规则覆盖率的公式为

$$[0057] \quad o = \sum_i \sum_j \omega_{X_i} \cdot f(X_i, Y_j)$$

[0058] 其中,  $o$  为规则覆盖率,  $X$  为疾病的规则集合,  $X_i$  为规则集合中的第  $i$  条规则,  $\omega_{X_i}$  代表第  $i$  条规则的权重,  $Y$  为输入信息的三元组集合,  $Y_j$  表示输入信息的三元组集合中的第  $j$  个三元组。

[0059] 在所述获取解释值模块之前还包括, 获取用户的病例知识图谱并进行初始化, 获取所述病例知识图谱中时间段  $T$  内的病例, 抽取所述病例中最热门的  $h$  个症状作为用户的初始提示症状;

[0060] 还包括如果未存在疾病的解释值  $p$  达到阈值  $E$ , 则选择  $k$  个解释值  $p$  对应的疾病, 并在知识图谱中搜索除输入信息之外的权重大于阈值  $F$  的三元组作为症状返回给用户, 并返回所述获取解释值模块。

[0061] 以下为本发明实施例, 如下所示:

[0062] 另外, 为实现上述方法, 本发明还公开了一种基于智能推理的辅助疾病诊断的系统, 包括:

[0063] 病例库: 病例库由采用国际通用的资源描述框架 (RDF) 标准对来自于一线医院的真实住院及门诊病例构成的一个个疾病知识图谱组成, 图谱中的每条边为一条表示 [主语, 谓语, 宾语] 的三元组, 进一步地, 为了后续的推理过程, 这里本发明对疾病图谱中的每一个三元组添加了级别和权重, 级别设定方法为: 如果某条三元组的主语为疾病, 那么本发明规定其为 1 级三元组, 也就是说该三元组中的谓词是直接修饰疾病的; 如果某条三元组的主语为 1 级三元组的宾语, 那么规定其为 2 级三元组; 以此类推, 本发明可以使用这种方法对疾病图谱中的每条三元组定级, 同时, 采用以下公式为疾病图谱中的三元组分配权重:

$$[0064] \quad \omega_{ij} = \frac{n_{ij}}{\sum_k n_{kj}} \times \log \frac{|D|}{1 + |\{d_m : n_i \in d_m \mid m = 1, 2, \dots, |D|\}|}$$

[0065] 其中,  $\omega_{ij}$  即为疾病集合中第  $j$  个疾病  $d_j$  中第  $i$  条三元组的权重,  $n_{ij}$  表示某条三元组宾语在疾病  $d_j$  的所有病例中出现的次数,  $\sum_k n_{kj}$  表示疾病  $d_j$  病例中所有三元组宾语出现的总次数,  $D$  表示疾病病种集合,  $|D|$  表示疾病的总种类数,  $|\{d_m : n_i \in d_m \mid m = 1, 2, \dots, |D|\}|$  表示在所有的疾病种类中包含宾语  $n_i$  的疾病种类数, 其中  $n_i$  表示在疾病  $d_j$  中出现的某个三元组宾语,  $m$  为疾病病种编号,  $d_m$  表示疾病病种集合中第  $m$  个病种。

[0066] 书本知识库: 使用 XML (eXtensible Markup Language) 格式存储的专业医疗书本知识。

[0067] 专家规则库: 由来自一线医生根据诊断经验编写而成;

[0068] 语义知识库: 包括词典文件, 停用词表, 正则表达式规则集合等;

[0069] 人机交互模块: 用来接受用户的输入以及显示返回结果;

[0070] 问句处理模块:用来处理用户的输入,利用自然语言处理方法识别用户输入中的有效信息。

[0071] 答句生成模块:根据用户的输入返回给用户不同的回答。

[0072] 智能推理模块:根据当前获取的用户的的有效信息进行推理,提示用户继续输入或者给出可以做进一步鉴别诊断的结论。进一步地,智能推理模块中包括以下步骤:

[0073] (1) 在病例库中计算用户信息与各个疾病的信息匹配度。定义用户信息与某一疾病病例的信息匹配度为 $m$ ,通过以下公式计算 $m$ 的值:

$$[0074] \quad m = \sum_{r=1} \lambda_{X_r} \cdot \left( \sum_{i=1} \sum_{j=1} \omega_{X_{ri}} \cdot g(X_{ri}, Y_j) \right)$$

[0075] 其中, $X$ 代表某一疾病的病例知识图谱, $X_r$ 表示处于 $X$ 中级别为 $r$ 的三元组集合, $\lambda_{X_r}$ 表示级别为 $r$ 的三元组集合的系数, $X_{ri}$ 表示级别为 $r$ 的三元组集合中的第 $i$ 个三元组, $\omega_{X_{ri}}$ 表示级别为 $r$ 的三元组集合中的第 $i$ 个三元组的权重, $Y$ 为用户信息的三元组集合, $Y_j$ 表示用户信息三元组集合中的第 $j$ 个三元组,

$$[0076] \quad g(x, y) = \begin{cases} 1, & \text{if}(x = y) \\ 0, & \text{others} \end{cases}, \text{其中} x, y \text{表示来自不同集合的两个三元组。}$$

[0077] (2) 在专家规则库中计算用户信息对每个疾病规则覆盖率,使用小写字母 $o$ 表示,计算公式如下:

$$[0078] \quad o = \sum_i \sum_j \omega_{X_i} \cdot f(X_i, Y_j)$$

[0079] 其中, $X$ 表示某一疾病规则集合, $X_i$ 为规则集合中的第 $i$ 条规则, $\omega_{X_i}$ 代表第 $i$ 条规则的权重, $Y$ 为用户信息的三元组集合, $Y_j$ 表示用户信息三元组集合中的第 $j$ 个三元组,另外,

$$[0080] \quad f(x, y) = \begin{cases} 1, & \text{if}(x = y) \\ 0 & \text{others} \end{cases}, \text{其中} x, y \text{表示来自不同集合的两个三元组。}$$

[0081] (3) 对计算出的信息匹配度与规则覆盖率进行加权求和,求出每个疾病对用户信息的解释值 $p$ ,

$$[0082] \quad p = \mu m + \nu o$$

[0083]  $\mu, \nu$ 分别代表信息匹配度与规则覆盖率的系数,由实验给出。解释值表示某一疾病对用户信息的解释程度, $p$ 的值越大,表示该疾病越能解释用户的输入信息。

[0084] 症状提示模块:症状提示模块不仅可以帮助用户便捷输入,也可以有助于准确获取用户的表达。系统在两个地方会触发症状提示,(1)系统在初始化时,会根据潜在狄利克雷分配(LDA)主题模型对病例图谱建模分析,找出热度最高 $h$ 个症状作为初始提示返回给用户;(2)当系统根据当前获取的用户信息不足以为用户作出初步推理结果时,系统进一步地搜索出当前最匹配的前 $k$ 个疾病图谱中的其他的权重较大的三元组症状,并将这些症状作为再一此的症状提示返回给用户。

[0085] 鉴别诊断模块:根据用户检查信息并结合用户已输入的有效信息对初始推理结论

中k个疾病使用书本知识库中相应疾病的诊断标准对用户作出最终鉴别诊断,并将诊断结果与系统推理过程由答句生成模块产生答案返回到人机交互界面展示给用户。

[0086] 以下为本发明的实施例,如下所示:

[0087] 下面将结合附图1,2对本发明中的技术方案进行进一步地详细且清晰的阐述。图1为基于智能推理的疾病辅助诊断系统框架图,包含下列模块:

[0088] 病例库:病例库由采用国际通用的资源描述框架(RDF)标准对来自于一线医院的真实住院及门诊病例构成的一个个疾病知识图谱组成,图谱中的每条边为一条表示[主语,谓语,宾语]的三元组;

[0089] 书本知识库:使用XML(eXtensible Markup Language)格式存储的专业医疗书本知识。

[0090] 专家规则库:由来自一线医生根据诊断经验编写而成。

[0091] 语义知识库:包括词典文件,停用词表,正则表达式规则集合等。

[0092] 人机交互模块:用来接受用户的输入以及显示返回结果。

[0093] 问句处理模块:用来处理用户的输入,利用自然语言处理方法识别用户输入中的有效信息。

[0094] 答句生成模块:根据用户的输入返回给用户不同的回答。

[0095] 智能推理模块:根据当前获取的用户的有效信息进行推理,提示用户继续输入或者给出可以做进一步鉴别诊断的结论。

[0096] 症状提示模块:症状提示模块不仅可以帮助用户便捷输入,也可以有助于准确获取用户的表达。系统在两个地方会触发症状提示,(1)系统在初始化时,会根据潜在狄利克雷分配(LDA)主题模型对病例图谱建模分析,找出热度最高h个症状作为初始提示返回给用户;(2)当系统根据当前获取的用户信息不足以为用户作出初步推理结果时,系统进一步地搜索出当前最匹配的前k个疾病图谱中的其他的权重最大的三元组症状,并将这些症状作为再一次的症状提示返回给用户。

[0097] 鉴别诊断模块:根据用户检查信息并结合用户已输入的有效信息对初始推理结论中k个疾病使用书本知识库中相应疾病的诊断标准对用户作出最终鉴别诊断,并将诊断结果与系统推理过程由答句生成模块产生答案返回到人机交互模块展示给用户。

[0098] 本系统的执行流程图如附图2所示,下面结合详细的步骤阐述系统的诊断流程。

[0099] 步骤S1:数据源初始化,对病例知识图谱中的三元组进行定级和权重分配处理,级别设定方法为:如果某条三元组的主语为疾病,那么本发明规定其为1级三元组,也就是说该三元组中的谓词是直接修饰疾病的,如[“肺炎”,“症状”,“咳嗽”],该条三元组的主语为“肺炎”,那么其被划分为1级三元组;如果某条三元组的主语为1级三元组的宾语,那么规定其为2级三元组,例如三元组[“咳嗽”,“诱因”,“着凉”]就为2级三元组;以此类推,本发明可以使用这种方法对疾病图谱中的每条三元组定级。同时,采用以下公式为疾病图谱中的三元组分配权重:

$$[0100] \quad \omega_{ij} = \frac{n_{ij}}{\sum_k n_{kj}} \times \log \frac{|D|}{1 + |\{m : n_i \in d_m\}|}$$

[0101] 其中, $n_{ij}$ 表示某条三元组宾语在疾病 $d_j$ 的所有病例中出现的次数, $\sum_k n_{kj}$ 表示疾病 $d_j$ 所有病例中三元组宾语出现的总次数, $|D|$ 表示疾病的总种类数, $|\{m : n_i \in d_m\}|$ 表示包含

宾语 $n_i$ 的疾病种类数。

[0102] 步骤S2:系统初始化,调用症状提示模块,根据系统当前执行的位置点,得知此时是在系统初始化时机,使用潜在狄利克雷分配(LDA)主题模型对病例知识图谱建模分析,找出不同主题下(本系统是疾病类型)最热的症状,从这之中选取 $h$ 个症状作为初始症状由人机交互界面返回给用户。

[0103] 步骤S2:用户在系统人机交互界面输入问句,问句的格式由复杂程度分为3种,包括:T1-用户症状关键词组合,示例:“咳嗽,发热”;T2-用户症状的详细描述组合,这里还需说明一下,T2-类型的输入是由系统获取用户的某一症状关键词后提示用户输入的关于该症状的详细描述,示例:“诱因:着凉,开始时间:3天前,最高体温:39度”;T3-用户自由输入,示例:“我家小孩 3岁了,这两天老是咳嗽,这是怎么了?”。

[0104] 步骤S3:获取步骤S1中用户的输入,对其做相应处理,方法如下:

[0105] 步骤S31:根据用户输入判断其所属类型(T1/T2/T3)。

[0106] 步骤S32:对不同类型的输入采用不同的解析方式,但最终都是以三元组形式返回。针对T1类型的输入,S1中的示例返回[[‘unknown’,‘症状’,‘咳嗽’],[‘unknown’,‘症状’,‘发热’]];针对T2类型的输入,假设用户之前获取的症状关键词为“发热”,则S1中的示例返回[[‘发热’,‘诱因’,‘着凉’],[‘发热’,‘开始时间’,‘3天前’],[‘发热’,‘最高体温’,‘39度’]];针对T3类型的输入,S1中的示例返回[[‘unknown’,‘症状’,‘咳嗽’],[‘咳嗽’,‘持续时间’,‘2天’],[‘unknown’,‘年龄’,‘3岁’]]。

[0107] 步骤S33:将步骤S22中对用户输入的处理结果返回给答句生成模块。

[0108] 步骤S4:答句生成模块获取问句处理模块的返回值,针对不同结果作不同的措施。

[0109] 步骤S41:如果在一次交互中,返回值中仅包含用户输入的症状关键词,则将返回值加入全局用户信息字典中并这些症状关键词依次加入一全局队列中。对该全局队列中的所有症状关键词,依次弹出,针对每一个弹出的症状关键词,生成描述该症状关键词的提示信息,作为系统的输出返回到人机交互界面,返回步骤S1;

[0110] 步骤S42:如果返回值中是关于用户某一症状关键词的描述三元组,则将这些症状描述三元组依次加入全局用户信息字典中。如果此时存储症状关键词的全局队列仍不为空,则继续弹出一症状关键词,生成描述该症状关键词的提示信息,作为系统的输出返回到人机交互界面,返回步骤S1;如果此时症状关键词全局队列为空,则跳至步骤S4,进入智能推理模块。

[0111] 步骤S5:根据上述步骤获取的全局用户信息,分别在病例库,规则库中计算其与各个疾病的信息匹配度以及规则覆盖率。

[0112] 步骤S51:在病例库中计算用户信息与各个疾病的信息匹配度。

[0113] 定义用户信息与某一疾病病例的信息匹配度为 $m$ ,通过以下公式计算 $m$ 的值:

$$[0114] \quad m = \sum_{r=1} \lambda_{X_r} \cdot \left( \sum_{i=1} \sum_{j=1} \omega_{X_{ri}} \cdot g(X_{ri}, Y_j) \right)$$

[0115] 其中, $X$ 代表某一疾病的病例知识图谱, $X_r$ 表示处于 $X$ 中位置为 $r$ 的三元组集合, $\lambda_{X_r}$ 表示位置为 $r$ 的三元组集合的系数, $X_{ri}$ 表示位置为 $r$ 的三元组集合中的第 $i$ 个三元组, $\omega_{X_{ri}}$ 表

示级别为 $r$ 的三元组集合中的第 $i$ 个三元组的权重, $Y$ 为用户信息的三元组集合, $Y_j$ 表示用户信息三元组集合中的第 $j$ 个三元组,

$$[0116] \quad g(x, y) = \begin{cases} 1, & \text{if}(x = y) \\ 0, & \text{others} \end{cases}, \text{其中} x, y \text{表示来自不同集合的两个三元组。}$$

[0117] 根据上面的公式可以计算出用户信息相对于疾病库中每一个疾病的匹配度。

[0118] 步骤S52:在规则库中计算用户信息对每个疾病规则的覆盖率,使用小写字母 $o$ 表示,计算公式如下:

$$[0119] \quad o = \sum_i \sum_j \omega_{X_i} \cdot f(X_i, Y_j)$$

[0120] 其中, $X$ 表示某一疾病规则集合, $X_i$ 为规则集合中的第 $i$ 条规则, $\omega_{X_i}$ 代表第 $i$ 条规则的权重, $Y$ 为用户信息的三元组集合, $Y_j$ 表示用户信息三元组集合中的第 $j$ 个三元组,另外,

$$[0121] \quad f(x, y) = \begin{cases} 1, & \text{if}(x = y) \\ 0, & \text{others} \end{cases}, \text{其中} x, y \text{表示来自不同集合的两个三元组。}$$

[0122] 步骤S53:对用户信息针对每个疾病计算出的信息匹配度与规则覆盖率进行加权求和,求出每个疾病对用户信息的解释值 $p$ ,

$$[0123] \quad p = \mu m + \nu o$$

[0124]  $\mu, \nu$ 分别代表信息匹配度与规则覆盖率的系数,由实验给出。解释值表示某一疾病对用户信息的解释程度, $p$ 的值越大,表示该疾病越能解释用户的输入信息。

[0125] 步骤S54:将疾病根据解释值 $p$ 的值由大到小排序,考虑前 $k$ 个疾病,默认 $k=3$ ,给定一阈值 $E$ ,由实验和专家给出,如果第一个疾病的解释值小于 $E$ ,转至步骤S6;如果某一疾病的解释值已经超过设定的阈值 $E$ ,则取前 $k$ 个解释值最大的疾病作为初步推理结果,调用答句生成模块,生成初步推理结果返回给人工交互模块供用户查看。随后,转至步骤S7。

[0126] 步骤S6:如果最大疾病的解释值也没有达到阈值 $E$ ,说明系统当前获取到的用户信息还不足以使系统做出初始推理结论,这时调用症状提示程序,系统进一步地搜索出当前最匹配的前 $k$ 个疾病图谱中的其他权重最大的症状,并将这些症状作为再一次的症状提示由人机交互模块返回给用户。

[0127] 步骤S7:进入鉴别诊断模块,提示用户输入检查结果信息,调用问句处理模块对用户输入的检查信息作解析,然后利用知识库中 $k$ 个疾病的诊断标准对用户作出鉴别诊断,并将诊断结果与系统推理过程由答句生成模块产生答案返回到人机交互模块展示给用户。

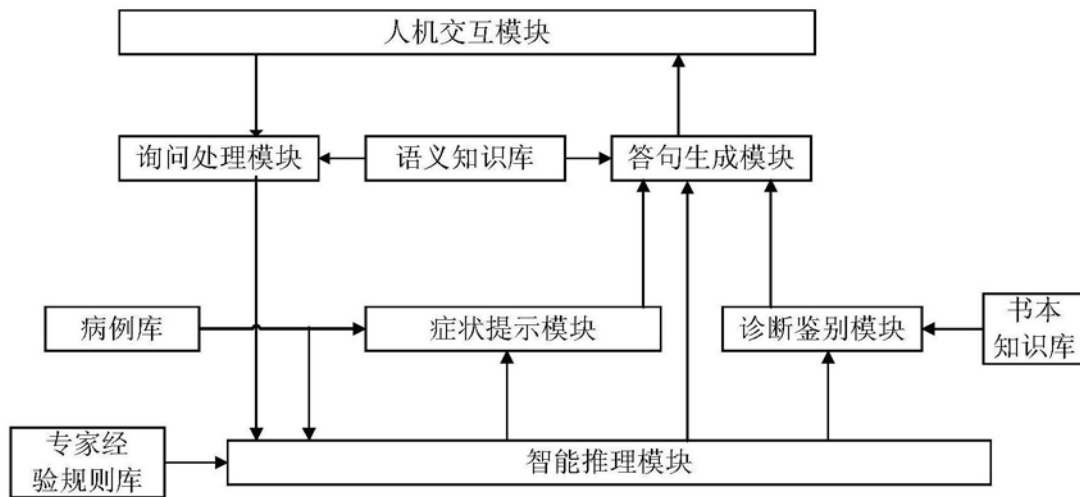


图1

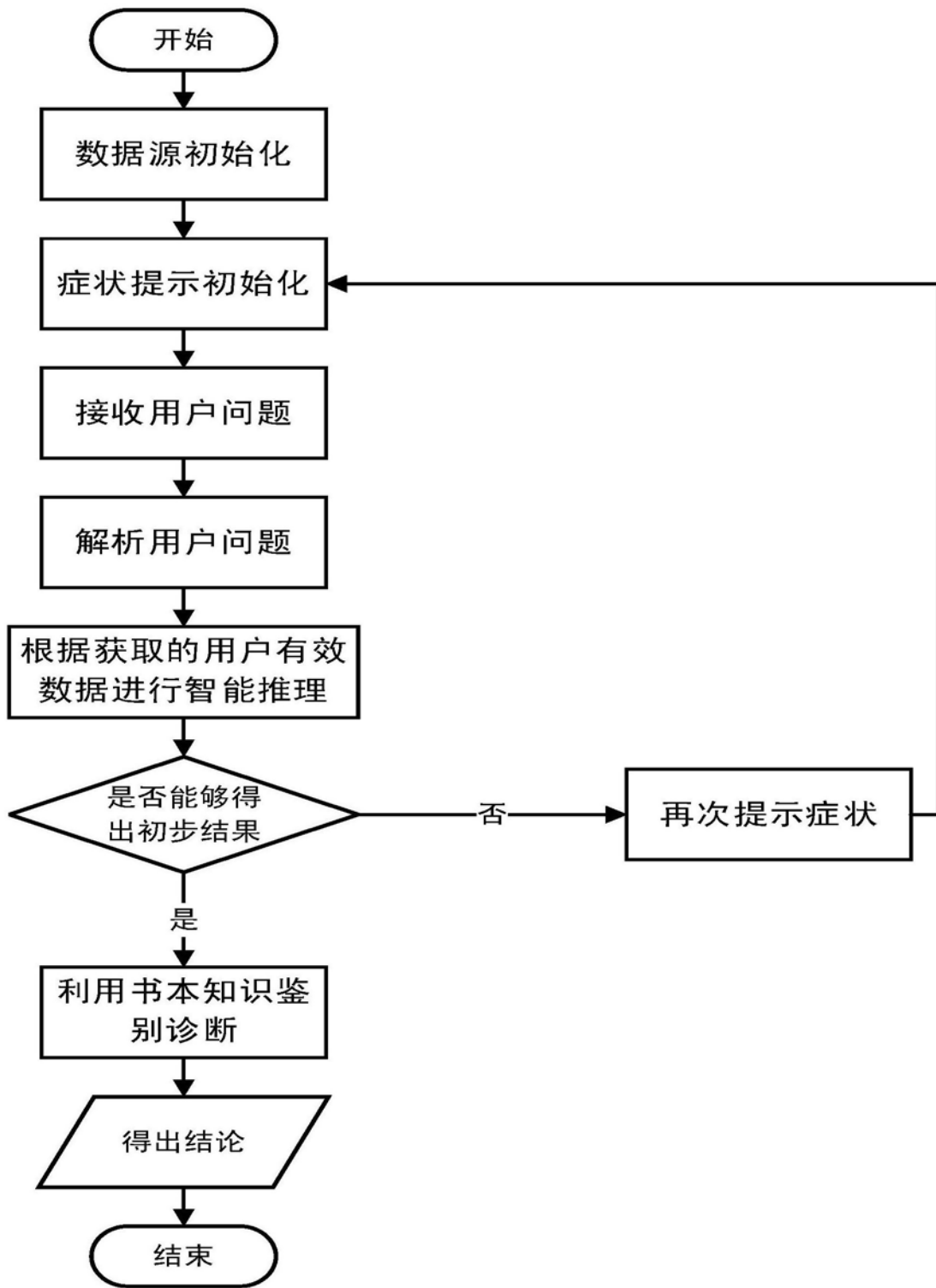


图2