



(12)发明专利申请

(10)申请公布号 CN 109119160 A

(43)申请公布日 2019.01.01

(21)申请号 201810949348.3

(22)申请日 2018.08.20

(71)申请人 重庆柚瓣家科技有限公司

地址 401121 重庆市两江新区互联网产业
园腾讯众创空间5楼

(72)发明人 薛方正 刘芳利

(74)专利代理机构 重庆强大凯创专利代理事务
所(普通合伙) 50217

代理人 陈家辉

(51)Int.Cl.

G16H 50/20(2018.01)

G16H 40/20(2018.01)

G06N 5/04(2006.01)

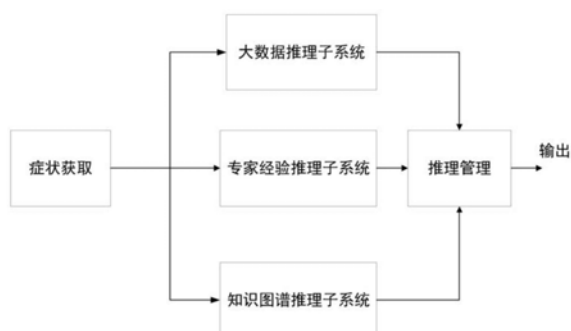
权利要求书2页 说明书17页 附图4页

(54)发明名称

多重推理方式的专家分诊系统及其方法

(57)摘要

本发明公开了一种多重推理方式的专家分诊系统,包括症状获取模块,专家经验推理子系统、大数据推理子系统和推理管理模块;专家经验推理子系统包括,专家经验规则库,模糊化模块,模糊推理模块;大数据推理子系统包括,关键特征值模块,推理模块;推理管理模块,用于从所述专家经验推理子系统和大数据推理子系统所反馈的推理结果中选出对应的发生概率最高的指定数量个疾病作为用于分诊参考的输出。本发明还公开了一种多重推理方式的专家分诊系统的方法,通过各个模式的互补,保证推理结果的输出,提高推理的准确性,从而提高分诊效率。



1. 多重推理方式的专家分诊系统,其特征在于:包括症状获取模块,专家经验推理子系统、大数据推理子系统和推理管理模块;

所述症状获取模块用于获取症状信息,并同时发送至专家经验推理子系统和大数据推理子系统;所述症状信息包括症状名称和症状的严重程度信息;

所述专家经验推理子系统包括,

专家经验规则库,用于存储多条规则,所述规则为某一疾病和该疾病的所有症状及症状的严重程度信息;

模糊化模块,用于对所述症状信息以及所述规则中的严重程度信息通过隶属函数进行模糊化;

模糊推理模块:用于根据所述症状信息及专家经验规则库内的规则,利用模糊推理计算出专家经验规则库内规则所对应的疾病发生的概率,并从中选出概率最高的多个疾病及其概率作为推理结果;

大数据推理子系统包括,

关键特征值模块,用于获取病例文本数据,并从中获取每一病例的关键特征值,进而形成关键特征值库;所述关键特征值库内的数据包括,疾病名及其对应的关键特征值;

推理模块,用于将获取的症状信息以关键特征值的形式表示,并计算症状信息的关键特征值与关键特征值库中各疾病所对应的关键特征值的匹配度,以及每个关键特征值所对应的疾病发生的概率,并从中选出匹配度最高的多个疾病及其概率作为推理结果;

所述推理管理模块,用于从所述专家经验推理子系统和大数据推理子系统所反馈的推理结果中选出对应的发生概率最高的指定数量个疾病作为用于分诊参考的输出。

2. 根据权利要求1所述的多重推理方式的专家分诊系统,其特征在于:还包括知识图谱推理子系统,

所述知识图谱推理子系统包括,存储有医学知识图谱的知识数据库和语义网络查询模块;

所述语义网络查询模块根据其所获取的症状信息生成查询语句,以从所述知识数据库中查询出相应的疾病名称作为推理结果。

3. 根据权利要求2所述的多重推理方式的专家分诊系统,其特征在于:所述推理管理模块,用于从所述专家经验推理子系统和大数据推理子系统所反馈的,对应的发生概率高于一阈值的疾病中选出发生概率最高的指定数量个疾病作为输出;如发生概率高于一阈值的疾病的数量不足,则输出知识图谱推理子系统的推理结果以补足。

4. 根据权利要求3所述的多重推理方式的专家分诊系统,其特征在于:所述推理管理模块,还用于将概率差小于5%的推理结果视为概率相同,并在概率相等的情况下优先选择专家推理系统的推理结果。

5. 根据权利要求1所述的多重推理方式的专家分诊系统,其特征在于:所述推理模块使用等值维度算法计算关键特征值间的匹配度。

6. 多重推理方式的专家分诊系统的方法,其特征在于:症状获取步骤:获取用户的症状信息,所述症状信息包括症状名称和症状的严重程度信息;

同时进行的专家经验推理步骤和大数据推理步骤;

所述专家经验推理步骤包括以下内容,

规则获取步骤:从存储有多条规则的专家经验规则库中获取规则,所述规则为某一疾病和该疾病的所有症状及症状的严重程度信息;

模糊化步骤:对所述症状信息以及所述规则中的严重程度信息通过隶属函数进行模糊化;

模糊推理步骤:根据症状信息以及专家经验规则库内的规则,利用模糊推理计算出专家经验规则库内规则所对应的疾病发生的概率,并从中选出概率最高的多个疾病及其概率作为推理结果;

所述大数据推理步骤包括以下内容,

关键特征值获取步骤:获取病例文本数据,并从中获取每一病例的关键特征值,进而形成关键特征值库,所述关键特征值库内的数据包括,疾病名及其对应的关键特征值;

推理步骤,将获取的症状信息以关键特征值的形式表示,并计算症状信息的关键特征值与关键特征值库中各疾病所对应的关键特征值的匹配度,以及每个关键特征值所对应的疾病发生的概率,并从中选出匹配度最高的多个疾病及其概率作为推理结果;

推理管理步骤,从所述专家经验推理步骤和大数据推理步骤所得到的推理结果中选出对应的发生概率最高的指定数量个疾病用于分诊。

7.根据权利要求6所述的多重推理方式的专家分诊系统的方法其特征在于:还包括与所述专家经验推理步骤和大数据推理同时进行的知识图谱推理步骤,

包括以下内容:

根据所述症状信息生成查询语句,采用该查询语句从知识数据库中查询出相应的疾病名称作为推理结果;所述知识数据库内存储有医学知识图谱。

8.根据权利要求7所述的多重推理方式的专家分诊系统的方法其特征在于:所述推理管理步骤中,还包括从所述专家经验推理步骤和大数据推理步骤所得的,对应的发生概率高于一阈值的推理结果中选出发生概率最高的指定数量个结果,并将这些结果所对应的疾病用作分诊的参考;如发生概率高于一阈值的疾病的数量不足,则输出知识图谱推理子系统的推理结果以补足。

9.根据权利要求8所述的多重推理方式的专家分诊系统的方法其特征在于:所述推理管理步骤中,还包括,将概率差小于5%的推理结果视为概率相同,并在概率相等的情况下优先选择专家推理系统的推理结果。

10.根据权利要求9所述的多重推理方式的专家分诊系统的方法其特征在于:所述推理模块步骤中,使用等值维度计算法计算关键特征值间的匹配度。

多重推理方式的专家分诊系统及其方法

技术领域

[0001] 本发明涉及医疗技术领域,具体涉及一种多重推理方式的专家分诊系统及其方法。

背景技术

[0002] 分诊登记是情况比较严重的节点之一,分诊是指根据病人的主要症状及体征判断病人病情的轻重缓急及其隶属专科,并安排其就诊的过程。但是目前分诊的结果还存在很多不合理的地方。现在大多数医院都只设立就诊指引台,就诊人向护士简要描述症状,由护士的主观经验来初步判断疾病类型;由于工作量大,询问不仔细或医学知识及经验不足,疾病判断准确率低,还是解决不了重复排队、重复诊断的问题。

[0003]

[0004] 现有技术中,已经出现了基于智能设备的分诊系统,通过与患者交互得到症状信息,从而推理出病患可能患有的疾病,从而提示病患前往正确的科室就诊;然而,此类系统多采用单一的推理模式,而每种推理模式都存在着其自己的不足,当病患给出的症状信息不足支撑该推理模式的推理时,推理结果要么是准确率低,要么是无法得到,造成了分诊效率的下降。

发明内容

[0005] 本发明的目的之一是提供一种基于多重推理模式的专家分诊系统,通过各个模式的互补,保证推理结果的输出,提高推理的准确性,从而提高分诊效率。

[0006] 多重推理方式的专家分诊系统,包括症状获取模块,专家经验推理子系统、大数据推理子系统和推理管理模块;

[0007] 所述症状获取模块用于获取症状信息,并同时发送至专家经验推理子系统和大数据推理子系统;所述症状信息包括症状名称和症状的严重程度信息;

[0008] 所述专家经验推理子系统包括,

[0009] 专家经验规则库,用于存储多条规则,所述规则为某一疾病和该疾病的所有症状及症状的严重程度信息;

[0010] 模糊化模块,对所述症状信息以及所述规则中的严重程度信息通过隶属函数进行模糊化;

[0011] 模糊推理模块:用于根据模糊化后的症状信息以及专家经验规则库内的规则,利用模糊推理计算出专家经验规则库内规则所对应的疾病发生的概率,并从中选出概率最高的多个疾病及其概率作为推理结果;

[0012] 大数据推理子系统包括,

[0013] 关键特征值模块,用于获取病例文本数据,并从中获取每一病例的关键特征值,进而形成关键特征值库;所述关键特征值库内的数据包括,疾病名及其对应的关键特征值;

[0014] 推理模块,用于将获取的症状信息以关键特征值的形式表示,并计算症状信息的

关键特征值与关键特征值库中各疾病所对应的关键特征值的匹配度,以及每个关键特征值所对应的疾病发生的概率,并从中选出匹配度最高的多个疾病及其概率作为推理结果;

[0015] 所述推理管理模块,用于从所述专家经验推理子系统和大数据推理子系统所反馈的推理结果中选出对应的发生概率最高的指定数量个疾病作为用于分诊参考的输出。

[0016] 本系统利用模糊化的输入和专家经验规则库可以较准确的推断用户所患疾病,结论足以满足分诊的需求,专家经验规则库的使用避免医学知识及经验不足所造成的疾病判断准确率低;然而专家经验会存在主观性,所以同时引入大数据病例规则库,通过真实的病例数据去弥补专家经验的不足之处,使得对用户所患疾病的推断更为准确;模糊推理为数学方法,属于模糊数学范畴,其原理是在已知规则和规则的结论的基础上,获取的条件从而在已知的结论中匹配得到最接近的结论,同时还能得知条件与规则间的差距,从而估计该结论正确的概率,当今常用的模糊推理方法很多,在此不作赘述,这些方法均可利用计算机等智能设备完成,推理过程不依靠人力,大大提高了分诊的效率,通过两个模式的互补,保证推理结果的输出,提高推理的准确性,从而提高分诊效率。

[0017] 进一步,还包括知识图谱推理子系统包括,存储有医学知识图谱的知识数据库和语义网络查询模块;

[0018] 所述语义网络查询模块根据其所获取的症状信息生成查询语句,以从所述知识数据库中查询出相应的疾病名称作为推理结果。

[0019] 知识图谱在两种推理模式得出的结论的概率较低时进行补充,避免无法在两种推理模式得不到结论时,无法给用户提供建议。

[0020] 进一步,所述推理管理模块,用于从所述专家经验推理子系统和大数据推理子系统所反馈的,对应的发生概率高于一阈值的疾病中选出发生概率最高的指定数量个疾病作为输出;如发生概率高于一阈值的疾病的数量不足,则输出知识图谱推理子系统的推理结果以补足。

[0021] 采用本方案可以避免低概率的结论被发送给用户,同时给出多个结论,降低遗漏疾病的可能性,为用户提供更为多样的建议。

[0022] 进一步,所述推理管理模块,还用于将概率差小于5%的推理结果视为概率相同,并在概率相等的情况下优先选择专家推理系统的推理结果。

[0023] 进一步,所述推理模块使用等值维度计算法计算关键特征值间的匹配度。

[0024] 专家推理系统的优先级被设为最高,同时避免偶然出现的过小的概率差让系统作出错误的判断。

[0025] 本发明的目的之二是提供一种基于多重推理模式的专家分诊系统的方法,通过各个模式的互补,保证推理结果的输出,提高推理的准确性,从而提高分诊效率。

[0026] 一种多重推理方式的专家分诊系统的方法包括以下内容;

[0027] 症状获取步骤:获取用户的症状信息,所述症状信息包括症状名称和症状的严重程度信息;

[0028] 同时进行的专家经验推理步骤和大数据推理步骤;

[0029] 所述专家经验推理步骤包括以下内容,

[0030] 规则获取步骤:从存储有多条规则的专家经验规则库中获取规则,所述规则为某一疾病和该疾病的所有症状及症状的严重程度信息;

[0031] 模糊化步骤:对所述症状信息以及所述规则中的严重程度信息通过隶属函数进行模糊化;

[0032] 模糊推理步骤:根据模糊化后的症状信息以及专家经验规则库内的规则,利用模糊推理计算出专家经验规则库内规则所对应的疾病发生的概率,并从中选出概率最高的多个疾病及其概率作为推理结果;

[0033] 所述大数据推理步骤包括以下内容,

[0034] 关键特征值获取步骤:获取病例文本数据,并从中获取每一病例的关键特征值,进而形成关键特征值库,所述关键特征值库内的数据包括,疾病名及其对应的关键特征值;

[0035] 推理步骤,将获取的症状信息以关键特征值的形式表示,并计算症状信息的关键特征值与关键特征值库中各疾病所对应的关键特征值的匹配度,以及每个关键特征值所对应的疾病发生的概率,并从中选出匹配度最高的多个疾病及其概率作为推理结果;

[0036] 推理管理步骤,从所述专家经验推理步骤和大数据推理步骤所得到的推理结果中选出对应的发生概率最高的指定数量个疾病用于分诊。

[0037] 本方法利用模糊化的输入和专家经验规则库可以较准确的推断用户所患疾病,结论足以满足分诊的需求,专家经验规则库的使用避免医学知识及经验不足所造成的疾病判断准确率低;然而专家经验会存在主观性,所以同时引入大数据病例规则库,通过真实的病例数据去弥补专家经验的不足之处,使得对用户所患疾病的推断更为准确;模糊推理为数学方法,属于模糊数学范畴,其原理是在已知规则和规则的结论的基础上,获取的条件从而在已知的结论中匹配得到最接近的结论,同时还能得知条件与规则间的差距,从而估计该结论正确的概率,当今常用的模糊推理方法很多,在此不作赘述,这些方法均可利用计算机等智能设备完成,推理过程不依靠人力,大大提高了分诊的效率,通过两个模式的互补,保证推理结果的输出,提高推理的准确性,从而提高分诊效率。

[0038] 进一步,还包括与所述专家经验推理步骤和大数据推理同时进行的知识图谱推理步骤,

[0039] 包括以下内容:

[0040] 根据所述症状信息生成查询语句,采用该查询语句从知识数据库中查询出相应的疾病名称作为推理结果;所述知识数据库内存储有医学知识图谱。

[0041] 知识图谱在两种推理模式得出的结论的概率较低时进行补充,避免无法在两种推理模式得不到结论时,无法给用户提供服务。

[0042] 进一步,所述推理管理步骤中,还包括从所述专家经验推理步骤和大数据推理步骤所得到的,对应的发生概率高于一阈值的推理结果中选出发生概率最高的指定数量个结果,并将这些结果所对应的疾病用作分诊的参考;如发生概率高于一阈值的疾病的数量不足,则输出知识图谱推理子系统的推理结果以补足。

[0043] 采用本方法可以避免低概率的结论被发送给用户,同时给出多个结论,降低遗漏疾病的可能性,为用户提供更为多样的建议。

[0044] 进一步,所述推理管理步骤中,还包括,将概率差小于5%的推理结果视为概率相同,并在概率相等的情况下优先选择专家推理系统的推理结果。

[0045] 专家推理系统的优先级被设为最高,同时避免偶然出现的过小的概率差让系统作出错误的判断。

- [0046] 进一步,所述推理模块步骤中,使用等值维度计算法计算关键特征值间的匹配度。
- [0047] 对于文本处理来说,维度算法简单易行,且准确度较高。

附图说明

- [0048] 图1为本发明实施例多重推理方式的专家分诊系统的示意性框图。
- [0049] 图2为本发明实施例中专家经验推理子系统的示意性框图。
- [0050] 图3为本发明实施例中关键特征值模块所用的基于语法分析的文本预处理模型的示意图。
- [0051] 图4为本发明实施例中知识图谱推理子系统的工作流程图。
- [0052] 图5为本发明推理管理模块的工作流程示意图。

具体实施方式

- [0053] 下面通过具体实施方式进一步详细说明:
- [0054] 本实施例中的多重推理方式的专家分诊系统的方法包括以下内容:
- [0055] 症状获取步骤:获取用户的症状信息,所述症状信息包括症状名称和症状的严重程度信息;
- [0056] 同时进行的专家经验推理步骤、大数据推理步骤和知识图谱推理步骤;
- [0057] 所述专家经验推理步骤包括以下内容,
- [0058] 规则获取步骤:从存储有多条规则的专家经验规则库中获取规则,所述规则为某一疾病 和该疾病的所有症状及症状的严重程度信息;
- [0059] 模糊化步骤:对所述症状信息以及所述规则中的严重程度信息通过隶属函数进行模糊化;
- [0060] 模糊推理步骤:根据模糊化后的症状信息以及专家经验规则库内的规则,利用模糊推理 计算出专家经验规则库内规则所对应的疾病发生的概率,并从中选出概率最高的多个疾病及 其概率作为推理结果;
- [0061] 所述大数据推理步骤包括以下内容,
- [0062] 关键特征值获取步骤:获取病例文本数据,并从中获取每一病例的关键特征值,进而形成关键特征值库,所述关键特征值库内的数据包括,疾病名及其对应的关键特征值;
- [0063] 推理步骤,将获取的症状信息以关键特征值的形式表示,并使用等值维度计算法计算症状信息的关键特征值与关键特征值库中各疾病所对应的关键特征值的匹配度,以及每个关键特征值所对应的疾病发生的概率,并从中选出匹配度最高的多个疾病及其概率作为推理结果;
- [0064] 知识图谱推理步骤,
- [0065] 包括以下内容:
- [0066] 根据所述症状信息生成查询语句,采用该查询语句从知识数据库中查询出相应的疾病名称作为推理结果;所述知识数据库内存储有医学知识图谱;
- [0067] 在上述三个推理步骤得出结果后,进入推理管理步骤,包括,从所述专家经验推理步骤 和大数据推理步骤所得的,对应的发生概率高于一阈值的推理结果中选出发生概率最高的指定数量个结果,并将这些结果所对应的疾病用作分诊的参考,其中将概率差小于

5%的推理结果视为概率相同,并在概率相等的情况下优先选择专家推理系统的推理结果;如发生概率高于一阈值的疾病的数量不足,则输出知识图谱推理子系统的推理结果以补足。

[0068] 该方法通过基本如附图1所示多重推理方式的专家分诊系统来实现:该系统包括症状获取模块,专家经验推理子系统、大数据推理子系统、只是图谱推理子系统和推理管理模块,下面逐一说明每一部分。

[0069] 如图2所示,专家经验推理子系统,包括:

[0070] 症状获取模块,用于获取用户的症状信息,所述症状信息包括症状的名称及症状的严重程度信息;

[0071] 专家经验规则库,用于存储多条规则,所述规则为某一疾病和该疾病的所有症状及症状的严重程度;

[0072] 模糊化模块,用于对所述症状信息以及所述规则中的严重程度信息通过隶属函数进行模糊化;

[0073] 模糊推理模块:用于根据模糊化后的症状信息以及专家经验规则库内的规则,利用模糊推理计算出专家经验规则库内规则所对应的疾病发生的概率;

[0074] 本模块中通过计算症状信息与专家经验规则库内的各个规则的贴进度作为各规则所对应的疾病发生的概率;

[0075] 修正模块,用于通过修正函数对所得贴进度进行修正,得出最终修正结果表征这些规则所对应的疾病发生的概率。

[0076] 专家经验规则库

[0077] 本实施例中的专家经验规则库通过数据库体现的知识主要包括疾病、症状及疾病每个症状对应的发生概率,数据库分为症状库、疾病库、疾病隶属度库和规则库。

[0078] 其中,每种疾病和它的所有症状及症状的严重程度就构成一条专家系统的规则,本实施例中的规则库事先由医学专业人士填写制定,也可以从医学书籍中得到,或是从实际病例数据中提取,总的来说,是从既往的经验中提取出来的。以下三个表格分别是专家系统中用到的症状库、疾病库、规则库的示例。

[0079] 表1 症状库

[0080]	编号	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}
	症状	热	咳嗽	咳痰	胸痛	肌肉酸痛	咳嗽血	消瘦	腹痛	腹水	黑便	恶心	呕血
	编号	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}	x_{19}	x_{20}	x_{21}	x_{22}	x_{23}	
	症状	板状腹	血便	黄疸	呕吐	腹胀	头痛	高血压	头晕	贫血	厌食	腹污	

[0081] 表2 疾病库

[0082]

编号	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8
病症	支气管哮喘	急性肺炎	肺炎球菌肺炎	葡萄球菌肺炎	肺炎支原体肺炎	克雷伯杆菌肺炎	肺结核	肺癌
y_9	y_{10}	y_{11}	y_{12}	y_{13}	y_{14}	y_{15}	y_{16}	y_{17}
肠结核	结核性腹膜炎	急性胃炎	慢性胃炎	消化性溃疡并出血	消化性溃疡并穿孔	溃疡性结肠炎	急性胆管炎	急性胰腺炎
y_{18}	y_{19}	y_{20}	y_{21}	y_{22}	y_{23}	y_{24}	y_{25}	y_{26}
肠梗阻	脑出血	脑栓塞	蛛网膜下腔出血	脑肿瘤	硬膜下血肿	脑积水	脑膜炎	脑脓肿

[0083] 表3 规则库(部分)

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}
y_1	7	3	4	7	7	7	7	7	7	7	7	7
y_2	2	3	4	7	7	7	7	7	7	7	7	7
y_3	3	2	5	7	7	7	7	7	7	7	7	7
y_4	4	3	5	6	5	7	7	7	7	7	7	7
y_5	7	3	7	7	4	7	7	7	7	7	7	7
y_6	2	5	6	7	7	7	7	7	7	7	7	7

[0084]

	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}	x_{19}	x_{20}	x_{21}	x_{22}	x_{23}
y_1	7	7	7	7	7	7	7	7	7	7	7
y_2	7	7	7	7	7	7	7	7	7	7	7
y_3	7	7	7	7	7	7	7	7	7	7	7
y_4	7	7	7	7	7	7	7	7	7	7	7
y_5	7	7	7	7	7	7	7	7	7	7	7
y_6	7	7	7	7	7	7	7	7	7	7	7

[0085] 表3中行为症状库中的编号,列为疾病库中的编号;而为了使系统能够更好的进行

疾病 诊断,我们把每个症状遵循疾病医学诊断的评判标准划分为了由低到高的多个等级,每个等 级用一个模糊量词来表示,为了方便,用于表征严重程度的模糊量词用表内的数字记,即“特 别严重”=0,“很严重”=1,“严重”=2,“比较严重”=3,“一般”=4,“有点”=5,“轻微”=6,“无”=7记模糊量词“特别严重”=0,“很严重”=1,“严重”=2,“比较严重”=3,“一 般”=4,“有点”=5,“轻微”=6,“无”=7,症状及其严重程度组成了规则的条件部分,而 疾病则是规则的结论部分。

[0086] 同时,疾病也有严重程度之分,采用同样的等级来区分,而疾病的严重程度信息保存在 如表4所示的疾病隶属度库中。所以,在更为完备结论库和规则库中,疾病及其规则均被进 一步按照严重程度细分,即多个编号对应同一疾病名称,但他们各自在疾病隶属度库中所对 应的严重程度不同,本实施例中,疾病的严重程度和症状的严重程度采用同样的模 糊量词;

[0087] 表4 疾病隶属度库

[0088]

病症	y ₁	y ₂	y ₃	y ₄	y ₅	y ₆	y ₇	y ₈	y ₉	y ₁₀	y ₁₁	y ₁₂	y ₁₃
隶属度	2	3	1	0	3	3	1	2	3	4	2	1	2
y ₁₄	y ₁₅	y ₁₆	y ₁₇	y ₁₈	y ₁₉	y ₂₀	y ₂₁	y ₂₂	y ₂₃	y ₂₄	y ₂₅	y ₂₆	
4	3	0	2	3	4	4	1	1	2	2	3	3	

[0089] 如此,诸如“如果有比较严重的咳嗽和一般程度的咳痰,那么该病人有严重的支气管哮喘病。”这样的经验知识,则可被表示为规则库中的规则1(表3中对应编号y1的第一行),其前提为“有比较严重的咳嗽和一般程度情况下的咳痰”,结论为“有严重的支气管哮喘病”。

[0090] 模糊化

[0091] 模糊化过程主要是把分明值转换成模糊集的隶属函数。由于严重程度是一个模糊的概念,采用的是模糊量词,其模糊度需要用隶属函数来表示;为了使系统能够更好的进行疾病 诊断,也更为贴近实际情况,本实施例中的各个模糊量词的隶属函数采用了如表5所示的正 态分布(0到1之间)。为了便于计算机处理,我们将模糊量词的隶属函数经抽样离散化后 得到表6中的长度为6的隶属度向量(下文中简称隶属度)。

[0092] 表5 各模糊量词所对应的隶属函数

	模糊量词	数值区间	隶属函数
	特别严重	[1.00,1.00]	$A(x) = \begin{cases} 1, & x=1 \\ 0, & x \neq 1 \end{cases}$
	很严重	[0.95,0.99]	$A(x) = 1 - e^{-(0.125/[0.97-x])^{2.5}}, x \in (0,1)$
	严重	[0.80,0.94]	$A(x) = 1 - e^{-(0.125/[0.87-x])^{2.5}}, x \in (0,1)$
[0093]	比较严重	[0.61,0.79]	$A(x) = 1 - e^{-(0.125/[0.70-x])^{2.5}}, x \in (0,1)$
	一般	[0.31,0.60]	$A(x) = 1 - e^{-(0.125/[0.46-x])^{2.5}}, x \in (0,1)$
	有点	[0.11,0.30]	$A(x) = 1 - e^{-(0.125/[0.20-x])^{2.5}}, x \in (0,1)$
	轻微	[0.01,0.10]	$A(x) = 1 - e^{-(0.125/[0.06-x])^{2.5}}, x \in (0,1)$
	无	[0.00,0.00]	$A(x) = \begin{cases} 1, & x=0 \\ 0, & x \neq 0 \end{cases}$

[0094] 表6 各模糊量词值离散化后的隶属度向量

	模糊量词	隶属度
	特别严重	$A(x) = [0,0,0,0,0,1]$
	很严重	$A(x) = [0,0,0,0.07,0.37,1]$
	严重	$A(x) = [0,0,0.04,0.136,0.99,0.98]$
[0095]	比较严重	$A(x) = [0,0.03,0.11,0.83,0.83,0.11]$
	一般	$A(x) = [0,0.15,1,0.53,0.08,0]$
	有点	$A(x) = [0.27,0,0.27,0.05,0.01,0]$
	轻微	$A(x) = [1,0.53,0.08,0,0,0]$
	无	$A(x) = [1,0,0,0,0,0]$

[0096] 模糊推理

[0097] 模糊推理通过各种控制规则的推理,结果被合成在一起,产生一个“模糊推理输出”的集合,本实施例采用贴进度推理算法,具体过程如下:

[0098] 首先,本实施例所使用的隶属度A和隶属度B的贴进度计算公式为:

$$[0099] \quad SM(A, B) = 1 - \left\{ \frac{1}{n} \sum_{i=1}^n [A(u_i) - B(u_i)]^2 \right\}^{1/2},$$

[0100] 其次,模式匹配,也就是将病人出现的症状与规则库中疾病对应的症状进行贴进度计算。

[0101] 病人提供的是一个症状集合X,先用贴进度公式计算病人症状 X_i 与规则库中规则 R_i 的症状 X_i 的贴进度,得到i个症状的贴进度后,将这i个贴进度求和后除以i,得到的就是病人症状集X与规则 R_i 的贴进度,也就是病人患疾病 R_i 的概率。依次计算病人症状集X与规则库中其它规则(R_2, R_3, R_4, \dots)的贴进度,最终得到病人患有规则库中疾病的全部概率。

[0102] 最后,根据事先确定的贴近度的阈值 σ 判断是否激发某一规则,即是否将其纳入“模糊 推理输出”的集合。

[0103] 例如:假设用户出现了比较严重的腹痛,轻微的腹胀,轻微的恶心和有点呕吐的症状。阈值 $\sigma=0.9$,根据表6可知:“无” $= [1, 0, 0, 0, 0, 0]$,“比较严重” $= [0, 0.03, 0.11, 0.83, 0.83, 0.11]$ 。

[0104] 与规则有 R_1 进行比较:

[0105] 设 r 为规则内的症状严重程度隶属度, t 为所得症状集中的症状严重程度隶属度;

[0106] 计算第一个症状:因为症状“热”在规则 R_1 和所得症状中都没有,所以

[0107] $SM_1(r_1, t_1) = 1$

[0108] 计算第二个症状:

[0109]

$$SM_2(r_2, t_2) = 1 - \left\{ \frac{1.00^2 + 0.03^2 + 0.11^2 + 0.83^2 + 0.83^2 + 0.11^2}{6} \right\}^{1/2} = 0.38;$$

[0110] 同理,得出与规则 R_1 所有症状的贴近度的计算。为了方便把记为 S_1 ,为了方便把 $SM_1(r_1, t_1)$ 记为 S_1 , $SM_2(r_2, t_2)$ 记为 S_2 ,以此类推,所得结果如表7所示。

[0111] 表7 输入症状与规则 R_1 的匹配结果

[0112]

S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}	S_{11}	S_{12}
1	0.38	0.38	1	1	1	1	0.38	1	0.78	1	1
S_{13}	S_{14}	S_{15}	S_{16}	S_{17}	S_{18}	S_{19}	S_{20}	S_{21}	S_{22}	S_{23}	
1	1	1	0.68	0.78	1	1	1	1	1	1	

[0113] 将所得的上表结果求和平均计算后得贴进度为0.88,即,与第一条规则匹配后的值为0.88,同理得出与所有规则比较的结果,如表8。

[0114] 表8 输入症状与规则库的匹配结果

[0115]

R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8	R_9	R_{10}	R_{11}	R_{12}	R_{13}
0.88	0.87	0.87	0.84	0.89	0.87	0.85	0.88	0.94	0.94	0.90	0.94	0.89
R_{14}	R_{15}	R_{16}	R_{17}	R_{18}	R_{19}	R_{20}	R_{21}	R_{22}	R_{23}	R_{24}	R_{25}	R_{26}
0.90	0.93	0.91	0.95	0.98	0.88	0.86	0.92	0.90	0.90	0.87	0.88	0.91

[0116] 通过与阈值 $\alpha=0.9$ 比较,当结果大于该阈值时,规则被激发;否则规则不被激发,所得结果如表9。

[0117] 表9 匹配结果与阈值比较结果

[0118]

R_9	R_{10}	R_{11}	R_{12}	R_{14}	R_{15}	R_{16}	R_{17}	R_{18}	R_{21}	R_{22}	R_{23}	R_{26}
0.94	0.94	0.90	0.94	0.90	0.93	0.91	0.95	0.98	0.92	0.90	0.90	0.91

[0119] 最后通过比较得出规则 R_{18} 的值为最大值。

[0120] 修正模块

[0121] 对于基于贴近度计算得到的结果,通常用修正函数加以改进,使系统得出的结果更符合实际情况,本实施例选择衰减型修正函数,遵循Zadeh的模糊数乘法运算。

[0122] 使用所选规则所对应的疾病的严重程度的隶属度,与贴进度进行模糊数乘法运算;例如, R18的严重程度为“比较严重”,而根据表6“比较严重”=[0,0.03,0.11,0.83,0.83,0.11],用该向量与贴进度(0.98)做模糊数乘法运算以得到修正后的贴进度如下:

[0123] $B^* = (0 \wedge 0.98) \vee (0.03 \wedge 0.98) \vee (0.11 \wedge 0.98) \vee (0.83 \wedge 0.98) \vee (0.83 \wedge 0.98) \vee (0.11 \wedge 0.98) = 0.83$

[0124] 所以,患者患有所选规则所对应的疾病的概率为83%。

[0125] 系统及其工作流程

[0126] 本实施例中的系统,连接有输入和输出设备,用户的症状及症状的严重程度作为系统的输入到症状获取模块,例如,用户输入非常头痛,则输入即为症状——头痛和程度——非常。系统根据输入的症状和症状的严重程度进行模糊推理,得出用户可能患有的疾病及患该疾病的概率放入临时数据库,该数据库按疾病概率由大到小排序。当输入的症状数量达到上限(症状数量阈值N,此处暂设为5)时,系统结束推理,根据疾病概率阈值P(即贴进度阈值 σ),输出概率最大且大于P的M个疾病(疾病输出数量M此处暂设为3,例如,经模糊推理后,有4种疾病概率均大于P,则输出概率最大的3种疾病)。

[0127] 其中,整个基于贴进度的推理算法的步骤如下:

[0128] 第一步:初始化

[0129] 1) 确定阈值 σ 的大小。

[0130] 2) 令 $i=1, j=1$ 。其中 $i=1, 2, \dots, n, j=1, 2, \dots, m; n$ 为库内规则的个数, m 为获取的症状的个数。

[0131] 第二步:模式匹配,

[0132] 将病人出现的症状与规则库中疾病对应的症状进行贴进度计算,将用户数据与某条规则进行贴进度计算;

[0133] 第三步:规则的选取;

[0134] 规则选取的过程为:

[0135] 1) 当病人的症状和专家经验规则库中的某一条规则贴进度 $\geq \sigma$ 时,那么这条规则将被激发,此时若 $i=n, j=m$,则转入3),否则转入第二步;若贴进度 $< \sigma$ 时,该规则不被激发,此时若 $i=n, j=m$,则转入3),否则转入第二步。

[0136] 2) 当只有一条规则被激发时,系统会输出该规则的结论部分;若没有一条规则被激发,系统输出“没有符合条件的疾病,建议用户去医院就诊”。

[0137] 3) 当有来自专家经验规则库的多条规则被激发时,系统会自动对满足结果的值进行比较,因为贴进度值越大说明规则越接近,所以选出贴进度值最大的一条或多条规则,并且输出这些规则的结论部分,结论包括了疾病名称、疾病严重程度;若有多条规则的贴进度值同样大时,系统会同时输出这几条规则的结论部分。

[0138] 第四步:计算推理结果

[0139] 用第三步结论中的疾病严重程度的隶属度对所选规则的贴进度进行相应的修正,得出最终修正结果。

[0140] 系统最终输出的结果包括,是否患有疾病、何种疾病、疾病的严重情况以及患病的概率,如“患有疾病肠梗阻比较严重0.83”。

[0141] 大数据推理子系统

[0142] 包括,关键特征值模块,用于获取病例文本数据,并从中获取每一病例的关键特征值,进而形成关键特征值库;所述关键特征值库内的数据包括,疾病名及其对应的关键特征值;

[0143] 推理模块,用于将获取的症状信息以关键特征值的形式表示,并计算症状信息的关键特征值与关键特征值库中各疾病所对应的关键特征值的匹配度,以及关键特征值所对应的疾病发生的概率,并从中选出匹配度最高的多个疾病及其概率作为推理结果.关键特征值模块采用如图3所示的基于语法分析的文本预处理模型,进行病例文本的预处理和特征选取,其中:

[0144] 文本预处理算法步骤包括,

[0145] 删除否定短语,比如:头颅无畸形,淋巴结不肿大,无意义。

[0146] 语法分析,一共分为三类(使用Stanford CoreNLP语法分词工具);

[0147] ①N-A/N-V式,名词+形容词,名词+动词,比如:神情痛苦,面部抽搐;

[0148] ②数量式短语,比如:体温38℃;

[0149] ③其他类型短语;

[0150] 中文分词和噪音消除,根据语法分析的结果分三种情况,

[0151] ①N-A/N-V式,使用键值对分词法.形式如下

[0152] 表10-1 键值对分词的形式

[0153]

键	值1	值2	……	值n
---	----	----	----	----

[0154] 比如:脊柱两侧肌肉紧张有压痛,其分词形式如下表:

[0155] 表10-2 脊柱两侧肌肉紧张有压痛的键值对分词的形式

[0156]

键	值1	值2
脊柱两侧肌肉	紧张	有压痛

[0157] ②数量式短语,从数值判断到键值对。

[0158] 数量短语基本都是测量值,所以依然可以使用键值对来进行表示.同时,根据项目名匹配数据库中标准,将数据替换成偏高,偏低等,比如:

[0159] “WBC (12.1×10^9)”,“HGB (118)”;

[0160] “RBC (5.25×10^{12})”,“PLT (3×10^9)”;

[0161] 其中WBC表示白细胞,HGB表示血红蛋白,RBC表示红细胞,PLT表示血小板,这个数据可以表示如下表:

[0162] 表10-3 数量式短语的键值对表示

[0163]

键	值	键	值
WBC	偏低	RBC	偏高
HGB	正常	PLT	偏高

[0164] ③其他句型,本实施例使用MeCab软件进行中文分词,得到平行单词。

[0165] 特征选择:

[0166] ①对于键值对:直接提取键值加入到关键特征集。

- [0167] ②对于平行单词,使用信息增益算法,选择重要单词加入关键特征集。
- [0168] 其中信息增益算法较为现有,在此不做赘述。
- [0169] 用向量表示文本
- [0170] 首先,将关键特征集里所有内容加上标签;比如:
- [0171] 文本1:斑丘疹,四肢,营养(良好),神智(清醒),体温(正常);
- [0172] 文本2:腹痛,呕吐,营养(欠佳),神智(模糊),体温(偏高);
- [0173] 文本3:斑丘疹,牙龈,营养(良好),体温(偏低),血压(正常);
- [0174] 得到的关键特征集为:“斑丘疹,四肢,营养,神智,体温,腹痛,呕吐,牙龈,血压”,依次标记为W1,W2,⋯,W9,其中的W1,W2,⋯,W9就是标签。
- [0175] 其次,将键值对应的所有属性值从0开始编号:
- [0176] 比如:体温有偏低,正常,偏高,分别表示为1,2,3,数字0表示文本不存在该特征。
- [0177] 最后,对文本进行赋值:
- [0178] 对于前面的示例,赋值如下表所示:
- [0179] 表10-4
- [0180]

	W1	W2	W3	W4	W5	W6	W7	W8	W9
D ₁	1	1	1	1	2	0	0	0	0
D ₂	0	0	2	2	3	1	1	0	0
D ₃	1	0	1	0	1	0	0	1	1

- [0181] 那么文本向量表示为: $D_1 = (1, 1, 1, 1, 2, 0, 0, 0, 0)$, $D_2 = (0, 0, 2, 2, 3, 1, 1, 0, 0)$, $D_3 = (1, 0, 1, 0, 1, 0, 0, 1, 1)$ 。
- [0182] 在操作中,使用每一篇病历的诊断结果即疾病名替换文章名,使用mysql数据库进行疾病名,关键特征集的值,诊断结果的存储,形成特征库。如果多篇病历诊断结果为同一疾病,则进行合并:病历中有相同症状,则症状权重减1,使用负数记录权重(正数用来记录症状情况);有不同症状,则进行症状补充。
- [0183] 于是,将D₁、D₂和D₃合并后文本实际表示为: $D = ((1, -1), (1, -1), (1, -1), (1, -1), (2, -1), 0, 0, 0, 0)$ 。
- [0184] 推理模块
- [0185] 在该模块中,文本向量的等值维度算法如下,

$$S(D_1, D_2) = \sum_{k=1}^n equal(P_{1k}, P_{2k}) \quad ,$$

$$equal(P_{1k}, P_{2k}) = \begin{cases} 1, P_k \text{ 为单值}, P_{1k} = P_{2k} \\ 0, P_k \text{ 为单值}, P_{1k} \neq P_{2k} \\ e, P_k \text{ 为多值}, P_{1k} \text{ 和 } P_{2k} \text{ 中相等值的个数} \end{cases} \quad ,$$

- [0188] 其中,n表示相关特征集纬度,D表示文档,P_k表示第1篇文档中特征k对应的属性值。
- [0189] 比如:
- [0190] $D_1 = (1, 1, 1, 1, 2, 0, 0, 0, 0)$,

[0191] $D_2 = (0, 0, 2, 2, 3, 1, 1, 0, 0)$,

[0192] $D_3 = (1, 0, 1, 0, 1, 0, 0, 1, 1)$,

[0193] 那么 $S(D_1, D_2) = 0, S(D_1, D_3) = 2$ 。即 D_1 和 D_3 相似, 且匹配度为 3。

[0194] 在该模块中, 概率计算的方法如下

[0195] 统计相同病症对应的不同疾病, 比如数据库中发现“腹痛”导致了两个胃病, 两个癌症, 一例胆结石则认为患有腹痛后, 40% 的概率诊断为胃病, 40% 的概率诊断为癌症, 20% 的概率 诊断为胆结石。

[0196] 在本实施例中, 具体做法是:

[0197] 统计特征库中每一列里相同的数字, 根据出现病历库中出现次数即权重值进行概率计算, 如:

[0198] $D_1 = ((1, -1), (1, -1), (1, -1), (1, -1), (2, -1), 0, 0, 0, 0)$,

[0199] $D_2 = ((1, -3), (2, -1), (1, -1), (1, -1), (2, -1), 0, 0, 0, 0)$,

[0200] 则认为症状 1 导致 D_1 的概率为 $1/(1+3) = 25\%$, 导致 D_2 的概率为 $3/(1+3) = 75\%$, 并添加进特征库中, 使用小数保存, 即文本实际表示为:

[0201] $D_1 = ((1, -1, 0.25), (1, -1, 1.00), (1, -1, 0.5), (1, -1, 0.33), (2, -1, 0.1), 0, 0, 0, 0)$

[0202] 本实施例中整个推理的过程如下:

[0203] 1、对用户的症状信息进行文本向量表示,

[0204] 输入症状信息之后, 按照数据预处理方法表示为文本向量, 达到和特征库存储的知识相匹配的目的。

[0205] 2、检索:

[0206] 第一步: 如果查询输入小于 3 个症状, 则向症状获取模块所用病历中最频繁的三个特征词汇, 用于提示, 返回第一步;

[0207] 第二步: 当输入大于 3 个时, 使用等值维度匹配计算出匹配度最高的三篇文档;

[0208] 第三步: 如果用户输入小于 5 个症状, 则返回三篇文档中重复率最高的且用户未包含的病症, 继续提问, 返回第一步; 若输入大于 5 个病症, 则返回三篇文档的诊断信息。

[0209] 在另一些实施例中, 可以利用上述对病历文档的处理, 提取出疾病、症状和症状的严重程度信息, 从而形成形式如同表 3 的规则库, 再利用相同的模糊推理算法得出用户最有可能患有的几种疾病以及每种疾病的患病的概率输出, 此处的推理过程与模糊推理模块中无异, 只是规则库的数据源从专家变成了病例文档。

[0210] 知识图谱推理子系统

[0211] 知识图谱推理包括, 储有医学知识图谱的知识数据库和语义网络查询模块。

[0212] 本实施例中为了综合表达能力和对推理的支持, 选用 OWL DL 子语言 (OWL 子语言的一种, 兼顾表达能力和对推理的支持) 来表示医学知识图谱内的知识。同时选用 Jena 作为平台, 构建医学知识图。Jena 是一个免费开源的支持构建语义网络和数据链接应用的 Java 框架, 由惠普实验室开发, 支持内存和永久存储。

[0213] 本实施例主要针对医学领域的老年病科, 采用“兰州大学临床医学知识服务系统”中老年病科的相关数据作为本体构建根据。

[0214] 知识模式的构建大致分为四部分:

[0215] (1) 类别(CLASS)的定义

[0216] 本实施例定义了最基本的两大类分别是:

[0217] ①“疾病”

[0218] ②“症状”

[0219] (2) 类别层次(上下位关系)的定义

[0220] 本实施例中依附subclass of,在获取到了层次信息后进行添加。

[0221] (3) 对象属性的定义

[0222] 本实施例中对象属性的定义域通常为类别或者实例,值域为某种值,例如string, int 等,以及“疾病”的“别称”、“英文名称”“并发症”属性;而由于症状类实体较为特殊,不存在属性关系。

[0223] (4) 语义关系的定义

[0224] 定义、值域均为类别或者实例。

[0225] 根据设定的两类实体的实际相关性分析,共创建了两种语义关系,分别为:

[0226] ①hazhengzhuang (有症状):疾病--症状,可以用来连接疾病与症状;

[0227] ②hasbingfazheng (并发症):疾病--疾病,可以用来连接疾病与疾病,表示几种疾病在同一患者身上发生,也就是疾病共同出现关系。

[0228] 综合以上对属性提取、语义关系创建,每个类别的最基本关系如下:

[0229] (1) CLASS:疾病

[0230] 属性:别称,类别:string;

[0231] 属性:英文名称,类别:string;

[0232] 属性:haszhengzhuang,类别:症状;

[0233] 属性:hasbingfazheng,类别:疾病;

[0234] (2) CLASS:症状

[0235] 类别:症状。

[0236] 本实施例利用Jena API创建模型,添加和修改属性和属性值、生成本体。

[0237] 本实施例中,知识图谱数据存储需要完成的基本数据存储,具体为主、谓、宾三元组形式的知识,以RDF文本格式存储在硬盘当中,再利用Jena将构建好的知识库固化为TDB的格式进行存储。

[0238] 为了方便管理本体或知识图谱,将RDF固化到一个通用的数据库中是最优做法。基于以上考虑,本实施例选择Jena作为平台,Jena支持RDB、SDB、TDB三种模式来固化RDF。采用TDB(Jena用于RDF存储和查询的模块)来固化RDF。将其他形式的数据转换成RDF(OWL)数据,再进一步存为TDB的形式。

[0239] 知识图谱构建完成后,通过语义网络查询模块接收输入的症状和并发症以及其查询需求。

[0240] 查询处理的过程如下:

[0241] S1、获取病人的症状信息,并生成查询语句,根据查询语句查询出知识图谱DB1中所有与该症状信息匹配的疾病数据Data1进行“ \cap ”运算到疑似疾病集Data;

[0242] S2、将疑似疾病集Data中的疾病数据Data1进行“ \cup ”运算,将运算得到的结果输出显示;

[0243] S21、在S2中,在“ \cup ”运算的结果不为空时,将“ \cup ”运算得到的结果作为查询得到的疾病数据Data1输出,在“ \cup ”运算的结果为空时,输出症状询问信息,在有补充的症状信息时,将补充的症状信息与原有的症状信息作为新的症状信息进行查询,得到疾病数据Data1,若没有补充的症状信息,则输出空的查询结果。

[0244] 具体的,对知识图谱DB1的存储是利用Jena平台先将第三方医学知识存储为OWL文件,然后利用TDB模块将其存储发布到服务器上,上述的TDB模块是Jena平台中用于RDF存储和查询的模块,并且支持所有Jena API。而知识图谱作为语义网络的发展,其存储的格式之一即为RDF,RDF是一个三元组(triple)模型,即每一份知识可以被分解为如下形式:(subject (主),predicate (谓),object (宾))。而虽然OWL比RDF推理能力更强,但本质上还是RDF,完全兼容RDF语法。OWL可理解为一组主谓宾的语句,一条OWL存储一组主谓宾,所以叫三元组,如(感冒,有症状,咳嗽),如果有多个症状就会有多个语句,即(感冒,有症状,咳嗽)、(感冒,有症状,发烧)、(感冒,有症状,食欲不振)等,系统中的所有的数据都用这种形式存在OWL文件中,然后再用Jena平台的TDB模块统一固化,也就完成知识图谱在数据库中的存储;在查询的时候,同样的也是利用TDB模块,可以固定主语,谓语,从而查出比如上面的语句集就可以查出所有症状。

[0245] 表11

[0246]

	疾病 A	疾病 B	疾病 C	疾病 D
症状 a	√	√		√
症状 b		√		
症状 c	√		√	√
症状 d	√		√	

[0247] 如图4和表11所示,患有疾病A的病人在诊断时,医生根据病人提供的信息向交互系统输入病症信息,如病人描述了症状a和症状c,医生此时向系统输入这些症状信息,输入的方式可以有多种,如利用键盘,采用打字输入的方式,也可以使用麦克风,采用语音输入的方式,还可以使用手写板,采用手写输入的方式;

[0248] 然后查询语句生成模块根据输入的症状信息生成查询语句,如系统的查询语句格式为(症状A,症状B,疾病,缺省),输入的症状信息为“症状a,症状c,症状d”,生成的查询语句为(症状a,症状c,症状d,疾病,缺省);查询模块在接收到查询语句后,依次查询各症状所属疾病,其中针对症状a的查询结果,也就是症状a的疑似疾病集为是[疾病A、疾病B、疾病D],症状c的疑似疾病集为[疾病A、疾病C、D],症状d的疑似疾病集为[疾病 A、疾病C],然后运算模块将上述查询得到的三个疑似疾病集进行 \cap 运算,运算结果为[疾病 A],再做 \cup 运算,运算结果还是[疾病A],因此最终的查询结果也就是A,即输出模块输出的结果为(症状a,症状c,疾病,疾病A),从而完成整个系统的查询操作。

[0249] 另外,用户给出的症状信息比较主观,很容易受到情绪的影响,在描述自己的病情时,容易受到悲观情绪影响,进而将症状的严重程度向更严重的方向偏移,如果直接采纳用户的描述进行查询,势必造成查询结果的偏移。因此,这种情况下有必要甄别用户是否

是在受到负面情绪的影响下给出了症状描述,若识别出是确实受到负面情绪的影响,则在进行疾病的推理时,就需要适当的减轻症状。

[0250] 出于这一目的,本实施例中的知识图谱推理子系统还包括有情绪识别模块以及症状修正模块,知识图谱数据库还存储有按照严重程度相互关联的症状信息,程度划分可统一采用如表5中所给出的示例。情绪识别模块对病人的声音、脸部图像信息进行采集识别,在识别出采集到的信息中带有情绪信息后,症状修正模块对症状信息进行修正,将输入的症状信息修正为相关联的严重程度低的症状信息;最后采用修正后的症状信息进行查询。

[0251] 具体的,情绪识别模块通过外设的麦克风或摄像头对用户的声音和表情图像进行采集,当情绪识别模块识别到病人的声音或表情中具有情绪信息时,例如用户此时处于恐慌状态,而如用户此时给出的是“严重”,症状修正模块就会对输入的症状信息进行修正,即将症状“严重”修正为症状“比较严重”,以消除用户受情绪影响给出的症状信息对诊断结果的影响。

[0252] 而在上述过程中,考虑到用户有时可能会因为突发原因产生惊恐情绪,如临时受到了惊吓,而惊吓的影响一直延续,此时不光是需要排除情绪的干扰,还需要分析用户的情绪是受病情的影响还是其他因素的影响,若是其他因素的影响,那么在进行推理的时候,则不需要对症状信息进行修正。

[0253] 为此,本实施例中的知识图谱推理子系统还包括有监控模块和推送模块;

[0254] 监控模块安装在用户终端上,可以为用户的手机,用户采用手机登录本实施例中的知识图谱推理子系统,在手机完成登录后,监控模块就会定期启动外设的麦克风和摄像头对用户的情绪进行监控,若用户使用本实施例的知识图谱推理子系统时,情绪识别模块识别到情绪信息时,如带有恐慌的情绪,同时若监控模块监控到用户长期处于恐慌的状态,此时推送模块还会根据用户联系人中的分组,将这个信息推送给用户的亲朋好友,用户的亲朋好友通过自己用户终端接收到“某某是否长期处于恐慌状态?”的推送信息后,则可以利用自己的用户终端向知识图谱推理子系统反馈,若被推送者这确认用户的确是因为自己的病情而处于恐慌状态的话,则向知识图谱推理子系统反馈表示肯定的信息,若是若被推送者不确认则反馈表示否定的信息;

[0255] 而若知识图谱推理子系统接收到的反馈信息为表示肯定的信息时,知识图谱推理子系统则根据情绪识别模块识别到的情绪信息、监控模块监控到用户长期处于恐慌的状态以及被推送者发来的肯定的反馈信息就可以识别用户的恐慌确实是由自己病情而导致的,此时边控制症状修正模块进行症状的修正,输入的症状信息修正为相关联的严重程度低的症状信息;

[0256] 而若知识图谱推理子系统接收到的反馈信息为表示否定的信息时,根据情绪识别模块识别到的情绪信息、监控模块监控到病人长期处于恐慌的状态以及亲朋好友否定的反馈信息就可以识别用户的恐慌并不是由自己病情所致,则不命令症状修正模块进行修正;

[0257] 若情绪识别模块识别到用户处于恐慌状态时,而监控模块监控到用户之前并没有任何恐慌状态,则表明用户只是当前处于恐慌状态,极可能是在上次监控至当前时间内受到了其他事情突发事件的影响而导致了恐慌情趣,此时症状修正模块同样不进行修正。

[0258] 通过外设麦克风和摄像头采集声音及用户面部图像进行情绪识别的技术对于本

领域的技术人员来说,属于较为现有的技术,在次不做赘述。

[0259] 推理管理模块

[0260] 推理管理模块的工作流程图如图5所示,由于专家经验推理、大数据推理和知识图谱推理是同时进行的,推理管理模块会同时收到3份推理结果,因此需要对这3份结果进行选择。其中专家经验推理系统的准确性是最高的,大数据推理系统次之,所以优先选择专家推理系统结果,然后选择病大数据推理系统结果,最后选择知识图谱结果。

[0261] 由于本系统给出的诊断结果只是对用户的初步诊断,准确度不能完全保证,所以系统每次输出3个诊断结果及其对应的患病概率。如果专家系统和大数据可以给出3个结果,则直接输出概率最大的3个疾病。若专家系统和大数据无法给出结果或结果少于3个,则从知识图谱的结果中进行选择,如果其中某个疾病是专家系统或大数据中所具有的,说明患该病的概率较低,则舍弃该疾病;若某个疾病在另外两个库中都没有,则输出该疾病。

[0262] 如果专家系统和大数据都没有满足各自输出条件的结果,或者两者满足条件的结果很多(此处设置为大于5个)且概率基本一致(此处设置患病概率差值小于5%则认为它们概率基本一致),无法再根据概率选择疾病,这两种情况下都选择输出知识图谱的结果。

[0263] 图5中使用的参数、符号及其对应的含义如下表所示。

[0264] 表12 参数、符号及其含义

[0265]

疾病集A	专家经验推理结果
疾病集B	大数据推理结果
疾病集C	知识图谱结果
疾病集D	系统预输出疾病集
N_0	系统输出疾病数量,
N_1	预输出疾病集数量阈值,
α	患病概率差值阈值,

[0266] 以上所述的仅是本发明的实施例,方案中公知的具体结构及特性等常识在此未作过多描述,所属领域普通技术人员知晓申请日或者优先权日之前发明所属技术领域所有的普通技术知识,能够获知该领域中所有的现有技术,并且具有应用该日期之前常规实验手段的能力,所属领域普通技术人员可以在本申请给出的启示下,结合自身能力完善并实施本方案,一些典型的公知结构或者公知方法不应当成为所属领域普通技术人员实施本申请的障碍。应当指出,对于本领域的技术人员来说,在不脱离本发明结构的前提下,还可以作出若干变形和改进,这些也应该视为本发明的保护范围,这些都不会影响本发明实施的效果和专利的实用性。本申请要求的保护范围应当以其权利要求的内容为准,说明书中的具体实施方式等记载可以用于解释权利要求的内容。

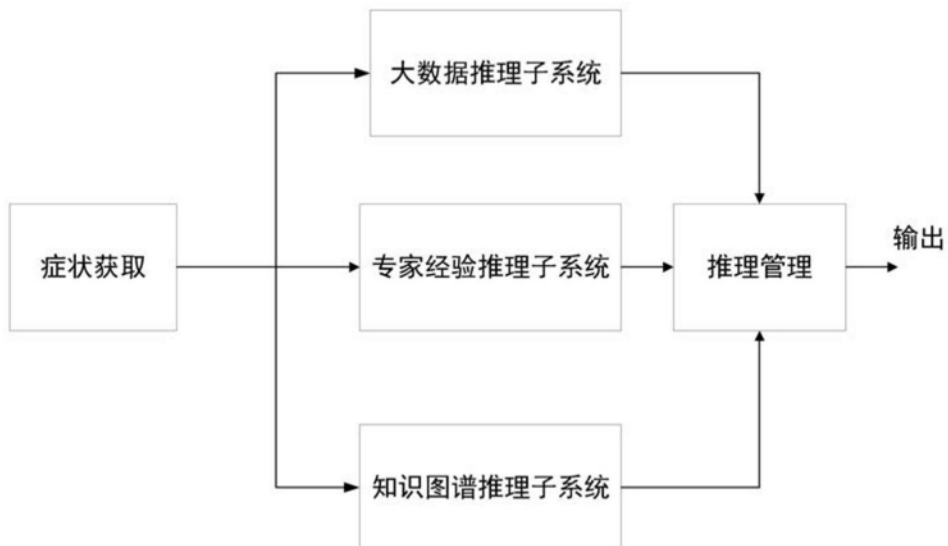


图1

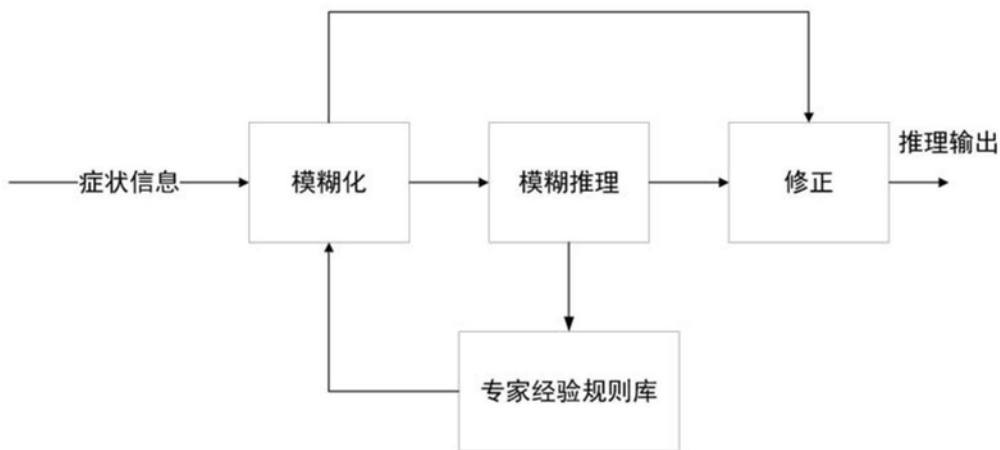


图2

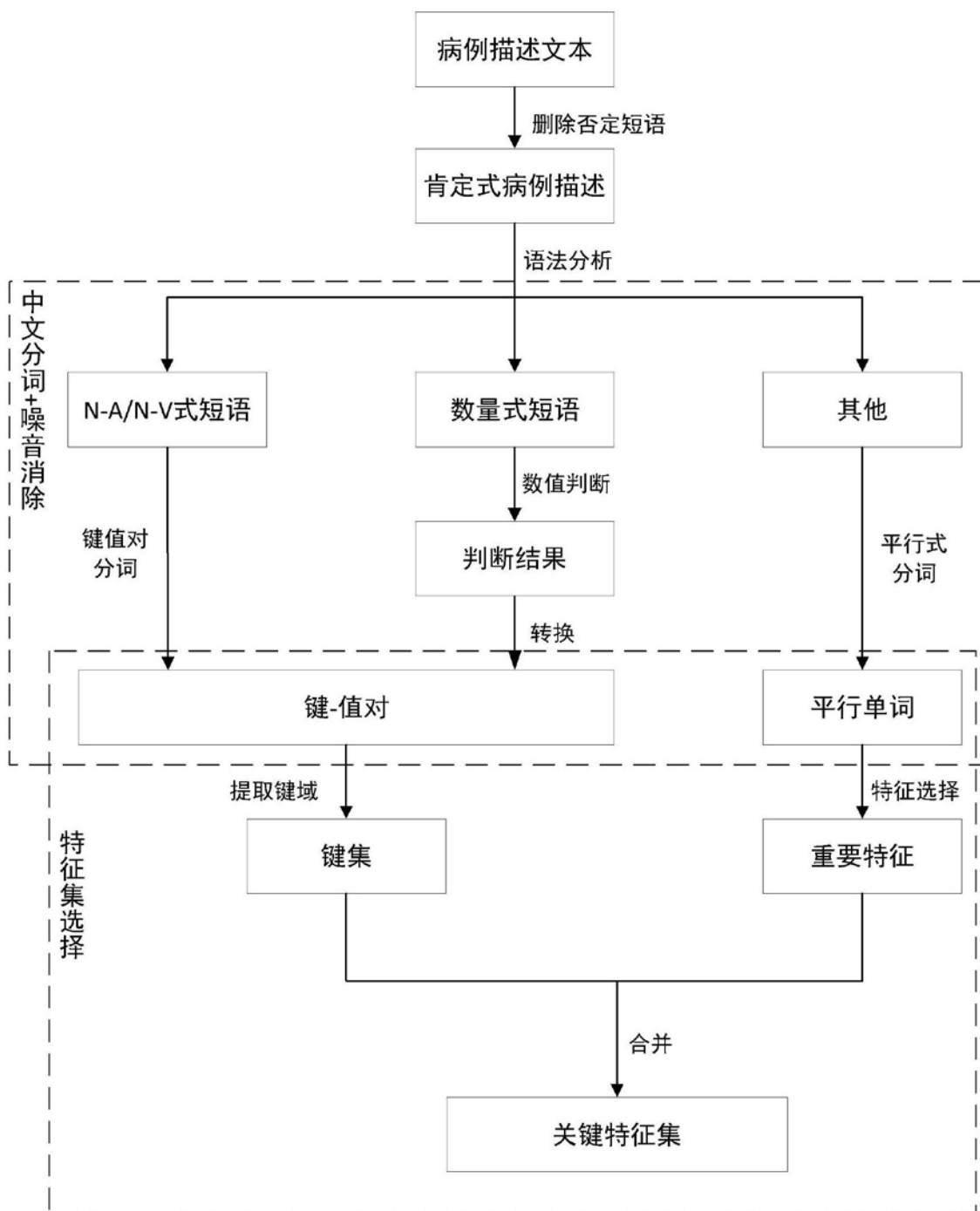


图3

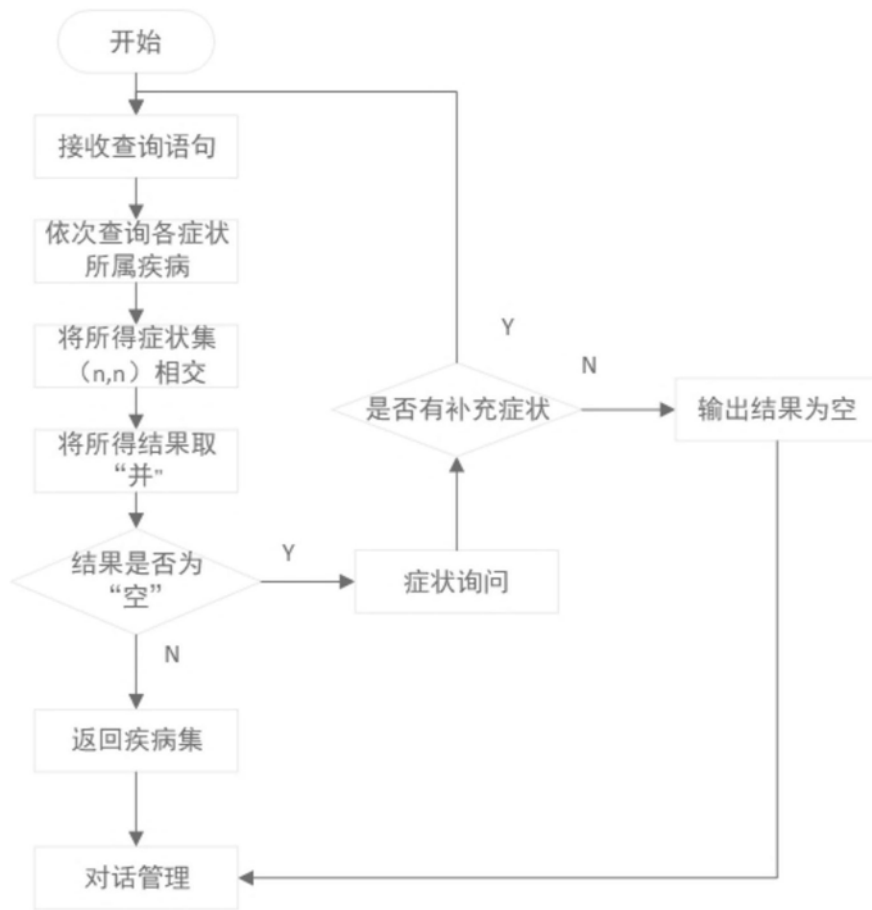


图4

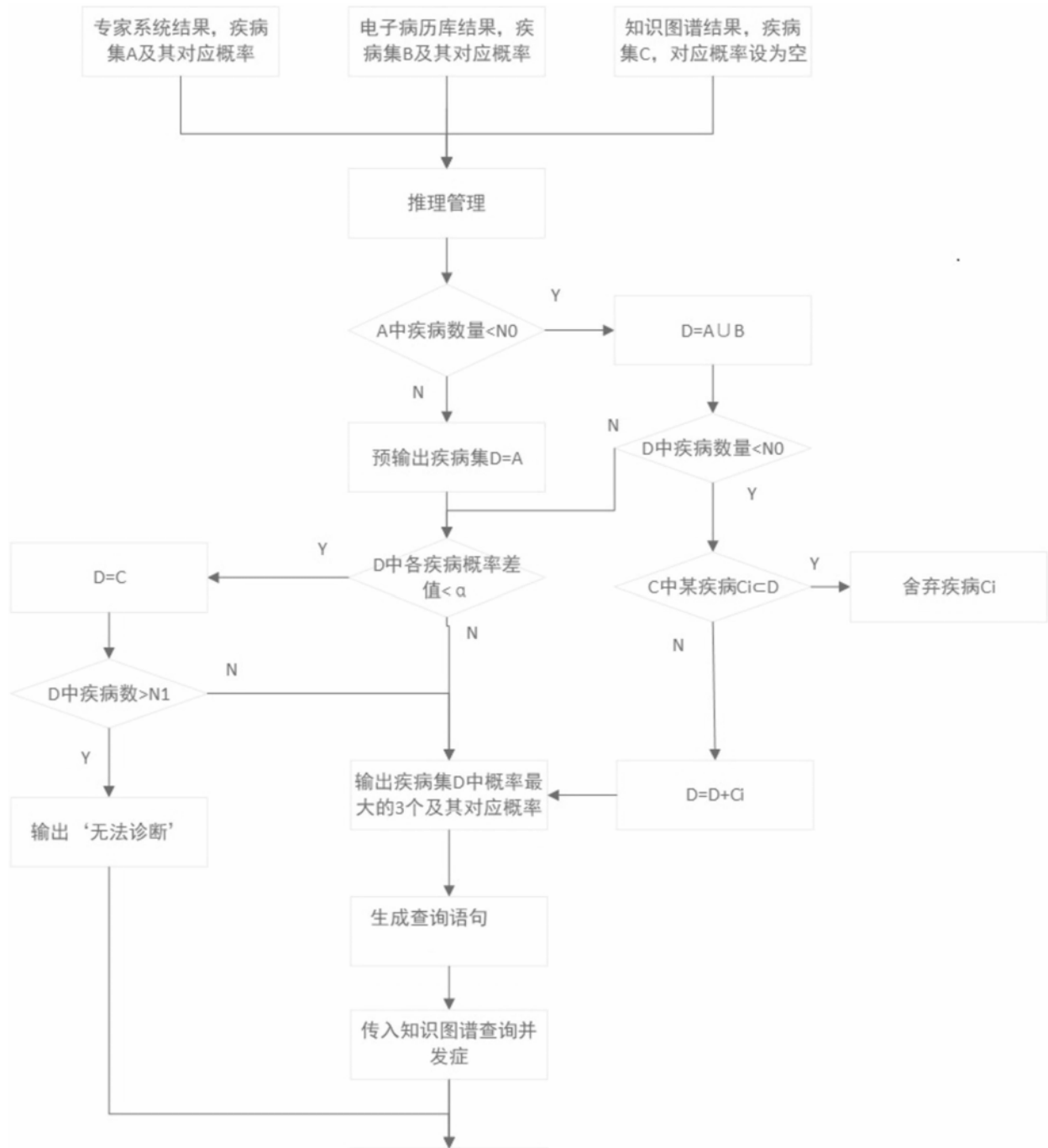


图5