



# (12)发明专利申请

(10)申请公布号 CN 107403067 A

(43)申请公布日 2017. 11. 28

(21)申请号 201710642039.7

(22)申请日 2017.07.31

(71)申请人 京东方科技集团股份有限公司

地址 100015 北京市朝阳区酒仙桥路10号

(72)发明人 李慧

(74)专利代理机构 北京金信知识产权代理有限公司

公司 11225

代理人 黄威 夏东栋

(51)Int.Cl.

G06F 19/00(2011.01)

G06N 5/02(2006.01)

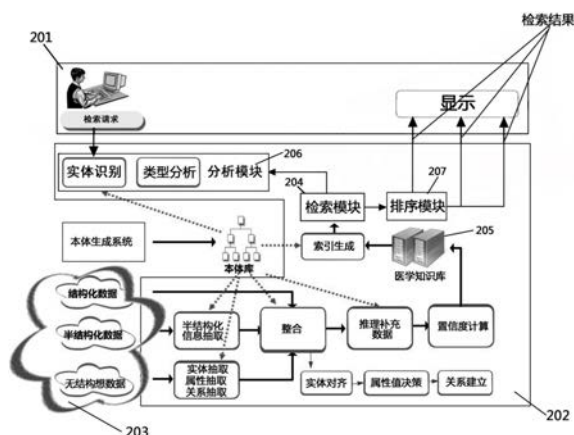
权利要求书1页 说明书8页 附图6页

## (54)发明名称

基于医学知识库的智能分诊服务器、终端及系统

## (57)摘要

本发明提供一种基于医学知识库的智能分诊服务器、终端及系统。所述智能分诊服务器包括：接收模块，配置为接收来自所述智能分诊终端的检索请求；分析模块，配置为对由所述接收模块所接收的检索请求进行实体识别和类型分析；检索模块，配置为根据由所述分析模块所识别的实体和分析出的类型，利用所述医学知识库的索引在所述医学知识库中检索出一条或更多条知识；排序模块，配置为对由所述检索模块所检索出的一条或更多条知识按照置信度高低进行排序，并将排序在前的知识作为分析结果传输给输出模块；所述输出模块，用于将所获得的分析结果传输到所述智能分诊终端。



1. 一种基于医学知识库的智能分诊服务器,其特征在于,所述智能分诊服务器包括:  
接收模块,配置为接收来自智能分诊终端的检索请求;  
分析模块,配置为对由所述接收模块所接收的检索请求进行实体识别和类型分析;  
检索模块,配置为根据由所述分析模块所识别的实体和分析出的类型,利用所述医学知识库的索引在所述医学知识库中检索出一条或更多条知识;  
排序模块,配置为对由所述检索模块所检索出的一条或更多条知识按照置信度高低进行排序,并将排序在前的知识作为检索结果传输给输出模块;  
所述输出模块,用于将所获得的检索结果传输到所述智能分诊终端。
2. 根据权利要求1所述的基于医学知识库的智能分诊服务器,其特征在于,所述分析模块通过隐式马尔可夫模型、最大熵模型、支持向量机、条件随机场中的至少一种方法或其组合来进行实体识别。
3. 根据权利要求1所述的基于医学知识库的智能分诊服务器,其特征在于,所述分析模块通过步步为营方法来进行实体识别。
4. 根据权利要求1-3中任何一项所述的基于医学知识库的智能分诊服务器,其特征在于,所述分析模块利用决策树、随机森林、逻辑回归、梯度提升、支持向量机中的至少一种分类器或其组合来进行类型分析。
5. 根据权利要求1-3中任何一项所述的基于医学知识库的智能分诊服务器,其特征在于,所识别的实体包括症状类实体、时间类实体、疾病类实体、个人信息类实体和医疗地点类实体。
6. 根据权利要求4所述的基于医学知识库的智能分诊服务器,其特征在于,所述类型包括预诊类型、分诊类型和保健类型。
7. 根据权利要求1-3所述的基于医学知识库的智能分诊服务器,其特征在于,所述医学知识库由结构化数据构成,并通过如下步骤来建立和更新所述医学知识库:  
从数据源中的半结构化数据和无结构数据抽取实体、属性以及关系作为知识;  
对抽取的知识进行融合;  
对融合后的知识进行置信度计算,将置信度超过阈值的知识存入所述医学知识库。
8. 根据权利要求7所述的基于医学知识库的智能分诊服务器,其特征在于,所述建立和更新所述医学知识库还包括,在对抽取的知识进行融合后:  
对融合后的知识进行推理以得到隐含的知识;  
对所得到的隐含的知识进行置信度计算,将置信度超过阈值的隐含的知识也存入所述医学知识库。
9. 根据权利要求7所述的基于医学知识库的智能分诊服务器,其特征在于,从半结构化数据和无结构数据抽取实体和属性通过监督算法、半监督算法、无监督算法和远距离监督算法中的任何一种来执行,从半结构化数据和无结构数据抽取关系通过监督算法来实现。
10. 根据权利要求7所述的基于医学知识库的智能分诊服务器,其特征在于,对抽取的知识进行融合包括:  
对从半结构化数据和从无结构数据抽取的知识进行整合,  
所述整合包括实体对齐、属性值的决策和关系建立。

## 基于医学知识库的智能分诊服务器、终端及系统

### 技术领域

[0001] 本发明涉及计算机领域,尤其涉及一种基于医学知识库的智能分诊服务器、终端及包括两者的系统。

### 背景技术

[0002] 中国医疗目前面临很多问题,包括:患者缺乏医疗健康知识;患者对于自身的症状无法正确评估,不知道何时需要去医院,不清楚就诊什么科室,经常发生挂错科室、小病到大医院就医等情况,从而降低医疗效率并浪费了医疗资源;医生工作负荷大,没有足够时间回答患者所有问题,造成医患关系紧张;医疗资源不均衡且缺乏,导致就医难等。

[0003] 整体说来,医学是一门知识驱动类学科,只要从多维度收集足够多、足够可靠的知识,就能在辅助决策层面发挥巨大的价值,把医疗人员从繁杂的重复性劳动中解放出来,进而去做更多有创造性的事。传统的诊断学思维大多使用人工编辑的医学知识库,医学知识库规模较小,在知识和信息更新上比较滞后,且采用决策树的结构,在诊断效果上很快就会达到瓶颈。通常现有的医学知识库由知识条目构成,这些知识条目通常是自由的文本格式,严重影响了检索效率。

### 发明内容

[0004] 发明人发现,需要利用目前人工智能的方式进行从用户终端到医学知识库的智能分诊系统的构建,使得患者可以方便地以自然语言提出检索请求,并准确高效地获得其想要得到的检索结果。这里的智能分诊,主要针对的是诊前咨询,主要解决用户在“身体不适”到“去医院”之间的需求,充当家庭医生,给诊疗做出指导。

[0005] 本发明人提出了本发明的例如如下技术方案来满足以上需求。

[0006] 根据本发明的第一方案,提供一种基于医学知识库的智能分诊服务器,其特征在于,所述智能分诊服务器包括:

[0007] 接收模块,配置为接收来自智能分诊终端的检索请求;

[0008] 分析模块,配置为对由所述接收模块所接收的检索请求进行实体识别和类型分析;

[0009] 检索模块,配置为根据由所述分析模块所识别的实体和分析出的类型,利用所述医学知识库的索引在所述医学知识库中检索出一条或更多条知识;

[0010] 排序模块,配置为对由所述检索模块所检索出的一条或更多条知识按照置信度高低进行排序,并将排序在前的知识作为检索结果传输给输出模块;

[0011] 所述输出模块,用于将所获得的检索结果传输到所述智能分诊终端。

[0012] 该智能分诊服务器能够利用自然语言处理技术对所接收的检索请求进行分析,依据分析结果在医学知识库中高效检索出相关的可靠知识并传输给所述智能分诊终端。

[0013] 优选地,所述分析模块通过隐式马尔可夫模型、最大熵模型、支持向量机、条件随机场中的至少一种方法或其组合来进行实体识别。优选地,所述分析模块通过步步为营

(“bootstrapping”)方法来进行实体识别。优选地,所识别的实体包括症状类实体、时间类实体、疾病类实体、个人信息类实体和医疗地点类实体。

[0014] 优选地,所述分析模块利用决策树、随机森林、逻辑回归、梯度提升、支持向量机中的至少一种分类器或其组合来进行类型分析。优选地,所述类型包括预诊类型、分诊类型和保健类型。

[0015] 优选地,所述医学知识库由结构化数据构成,所述计算机可执行指令被所述处理器执行时用于通过如下步骤来建立和更新所述医学知识库:从数据源中的半结构化数据和无结构数据抽取实体、属性以及关系作为知识,对抽取的知识进行融合,对融合后的知识进行置信度计算,将置信度超过阈值的知识存入所述医学知识库。相较现有技术的人工编辑的无结构数据的医学知识库,这种由结构化数据构成的具备自主学习和扩展能力的医学知识库能够从广泛的数据源和数据抽取知识,优选地能够对知识进行融合形成新的知识,还能够选取比较可信的新知识来合理地扩展医学知识库。

[0016] 优选地,所述建立和更新所述医学知识库还包括:在对抽取的知识进行融合后,对融合后的知识进行推理,得到隐含的知识,用于进行置信度计算,将置信度超过阈值的隐含的知识也存入所述医学知识库。

[0017] 优选地,从半结构化数据和无结构数据抽取实体和属性通过监督算法、半监督算法、无监督算法和远距离监督算法中的任何一种来执行,从半结构化数据和无结构数据抽取关系通过监督算法来实现。

[0018] 优选地,对抽取的知识进行融合包括:对从半结构化数据和从无结构数据抽取的知识进行整合,所述整合能够包括实体对齐、属性值的决策和关系建立。

[0019] 根据本发明的第二方案,提供了可与智能分诊服务器配合使用的用户友好且携带方便的智能分诊终端,使得用户能够利用自然语言的形式输入检索请求,且能够将从医学知识库检索的检索结果以自然语言的形式显示给用户,从而给患者提供个性化的咨询服务。

[0020] 根据本发明的第三方案,提供了一种智能分诊系统,包括上述的智能分诊终端和与之通信连接的上述的基于医学知识库的智能分诊服务器。

## 附图说明

[0021] 为使本领域技术人员更好地理解本发明,下面参照附图对本发明的实施例进行详细说明,但不作为对本发明的限定。以下将参照附图来描述本发明的示例性实施例的特征、优点,以及技术和适用性,在附图中,相同的附图标记指代相同的元件,并且在附图中:

[0022] 图1示出根据本发明第一实施例的基于医学知识库的智能分诊系统的主要原理的示意图;

[0023] 图2示出根据本发明第二实施例的基于医学知识库的智能分诊系统200的框图,其中框201表示智能分诊系统中的智能分诊终端,框202表示智能分诊系统中的智能分诊服务器,框203表示数据来源;

[0024] 图3示出根据本发明第三实施例的基于医学知识库的智能分诊系统200的分层架构图;

[0025] 图4示出根据本发明第四实施例的融合后的医学知识的示意图;

[0026] 图5(a)-图5(e)示出根据本发明第五实施例的基于医学知识库的智能分诊终端201的界面的示意图。

### 具体实施方式

[0027] 以下结合附图对本发明进行详细描述。图1示出根据本发明第一实施例的基于医学知识库的智能分诊系统的主要原理的示意图,如图1所示,所述智能分诊系统通信链接到数据源,所述数据源包括诸如医学文献的无结构的文本数据和诸如电子病历的半结构化数据等。所述智能分诊系统接收用户的输入,所述输入包括自然语言形式的用户的个人信息、检索请求等,并输出检索结果,检索结果可以包含用户可能患有的疾病(针对用户的预诊类型的检索请求)、对应的科室(针对用户的分诊类型的检索请求)以及健康教育信息(针对用户的保健类型的检索请求)中的至少一种。例如,所述智能分诊系统可以包括:处理器;和存储器,所述存储器存储有计算机可执行指令,所述计算机可执行指令被所述处理器执行时,实现基于自然语言处理和机器学习技术的分诊算法,所述分诊算法具体说来包含:对用户输入的检索请求进行实体识别和类型分析;根据所识别的实体和分析出的类型,利用所述医学知识库的索引在所述医学知识库中检索出一条或更多条知识;选择那些置信度较高的知识作为检索结果,并提供给用户。优选地,所述分诊算法还可以包括:对来自数据源的数据进行挖掘、分析和抽取,可选地可对抽取的知识进行融合,从而建立和/或更新所述医学知识库。注意,虽然图1中将医学知识库包含在智能分诊系统中作为示例,但医学知识库未必需要存储在存储器中,只要智能分诊系统可以调用之即可;此外,系统不仅可以使其自身建立和/或更新的医学知识库,也可以与其他系统共享医学知识库。由此,所述分诊算法整合了自然语言处理、数据挖掘、数据分析和机器学习(例如深度学习)等功能。优选地,所述医学知识库可以包括健康知识库、健康信息库、慢病管理知识库中的至少一种。

[0028] 图2示出根据本发明第二实施例的基于医学知识库的智能分诊系统200的框图,其中框201表示智能分诊系统200中的智能分诊终端,框202表示智能分诊系统200中的智能分诊服务器,框203表示数据来源。请注意,图2中,方形框表示物理意义上的构件(包含软件和/或硬件),例如各种程序模块、智能分诊终端、本体库等;圆角框表示这些构件所执行的步骤,经由存储在智能分诊服务器202的存储器(图中未示出)上的计算机可执行指令的执行来实现;气泡框表示各种类型的数据。

[0029] 如图2的框202所示,所述智能分诊服务器202连接到智能分诊终端201,接收来自其的用户的个人信息(可选,简化版也可以不含此信息)和检索请求,所述检索请求是由用户向所述智能分诊终端201输入的语句,输入方式例如可以经由选择框勾选、手写输入等等或其组合。例如,检索请求可以由用户输入的个人信息“2岁女”、经由选择框勾选的症状信息“高烧39度持续2小时不退”、手写输入的问题“要不要去医院”构成。当然,通常检索请求可以更简单,例如仅仅是用户输入的症状信息。

[0030] 所述智能分诊服务器202据此从医学知识库205获取针对检索请求的检索结果,并传输给所述智能分诊终端201供其显示给用户。

[0031] 具体说来,如图2所示,所述智能分诊服务器202包括:接收模块(图中未示出),配置为接收来自智能分诊终端201的检索请求;分析模块206,配置为对由所述接收模块所接收的检索请求进行实体识别和类型分析;检索模块204,配置为根据由所述分析模块206所

识别的实体和分析出的类型,利用所述医学知识库205的索引在所述医学知识库中检索出一条或更多条知识;排序模块207,配置为对由所述检索模块204所检索出的一条或更多条知识按照置信度高低进行排序,并将排序在前的知识作为检索结果传输给输出模块;所述输出模块(图中未示出),用于将所获得的检索结果传输到所述智能分诊终端201,以供显示。

[0032] 目前分析模块206可以采用有监督的方法和半监督的方法来进行实体识别。例如,有监督的方法包括隐式马尔可夫模型、最大熵模型、支持向量机、条件随机场中的至少一种方法或其组合。例如,半监督的方法包括步步为营 (bootstrapping) 方法。通常,有监督算法在进行实体识别之前,需要先利用标注好的语料进行训练。

[0033] 例如,对“1岁女”,“突发高烧39度2天持续不退”,“什么病?要不要去医院?”的检索请求进行实体识别,识别出的实体包括:症状类实体“突发”、“高烧39度”;时间类实体“2天”;疾病类实体“病”;个人信息类实体“1岁”和“女”;还有医疗地点类实体“医院”。检索模块204可根据所识别的这些实体,利用医学知识库205的索引在医学知识库205中检索出知识,利用索引可迅速定位包含这些实体的知识条目,从而大大提高检索效率,各条知识是由作为节点的实体、实体的属性和实体之间的关系构成的。

[0034] 为了从检索出的包含所识别的实体的知识中选择更符合用户意图的知识,可以还对检索请求进行类型分析,并从按照实体检索出的知识中进一步选择出属于分析所得类型的那些知识,从而更精准地满足用户的需求。具体说来,所述分析模块206可以利用决策树、随机森林、逻辑回归、梯度提升、支持向量机中的至少一种分类器或其组合来进行类型分析,这些分类器在应用之前需要利用标注语料进行训练。通常说来,所述类型包括:预诊类型,例如,用户想要预测可能的疾病、疾病概率等;分诊类型,例如,用户想知道该去什么医院什么科室挂号;和保健类型,例如,用户想要得到一些疾病的科普知识、还有如何保健和如何护理的知识。还以“1岁女”,“突发高烧39度2天持续不退”,“什么病?要不要去医院?”的检索请求为例,利用分类器将其分类为预诊类型、分诊类型和保健类型。又例如,当检索请求中不包含任何诸如“什么病?要不要去医院?”的语句时,默认用户需要得到所有三种类型的检索结果,以免遗漏用户的需求。

[0035] 以“1岁女”,“突发高烧39度2天持续不退”,“什么病?要不要去医院?”的检索请求为例,按照识别的实体和分析的类型,最后得出的检索结果是:

[0036] 可能疾病:婴儿玫瑰疹

[0037] 概率:70%

[0038] 推荐科室:儿科门诊

[0039] .....

[0040] 护理建议:可留家查看,配合物理降温,体温大于38.5度时服用退烧药,一般3-5天疹出退烧,皮疹为粉红色或鲜红色斑点;如果逾期不退烧出疹,可去儿科门诊挂号就诊。

[0041] 例如,当检索请求不包括“要不要去医院”时,判定其类型为预诊类型,最后得出的检索结果中仅仅包括上文中的可能疾病和概率。

[0042] 排序模块207可配置为对由所述检索模块204所检索出的一条或更多条知识计算置信度,并按照置信度的高低进行排序。可以利用多种方法来计算置信度,例如可以利用知识中的实体和关系在医学知识库205中出现的频率、各条知识的子节点个数等或其组合来

计算置信度。

[0043] 所述数据源203包括结构化数据、半结构化数据和无结构数据,结构化数据例如当前的医学知识库205,半结构化数据例如电子病历,无结构化数据例如来自医学网页、医学文献和医学指南等的文本数据。所述医学知识库205由结构化数据构成,可通过如下步骤来建立和更新所述医学知识库205:从数据源203中的半结构化数据和无结构数据抽取实体、属性以及关系作为知识,对抽取的知识进行融合,对融合后的知识进行置信度计算,将置信度超过阈值的知识存入所述医学知识库。

[0044] 下面结合图2对抽取和融合的几个步骤进行具体说明。对半结构化数据和无结构化数据分别进行实体、属性和关系的抽取,以得到结构化数据(作为知识),所得的结构化数据可以进行整合,这个对异构数据进行抽取和整合的过程实现了异构数据整合,整合后的数据也称为融合后的知识。抽取的具体示例实现在下文中会结合图3进行详细描述,在此不赘述。对知识的融合,即需要对实体、属性和关系的冲突、重复冗余、偏差进行数据的清理工作,包括实体对齐、属性值决策、关系建立等。

[0045] 例如,以实体“长时间断断续续肚子疼”和“腹痛1周”为例,判定实体“肚子疼”和“腹痛”是同一个实体,则可以将两者对齐为“腹痛”,“长时间断断续续肚子疼”可以确定为其属性是“慢性的”,对于“腹痛1周”就需要进行决策,从而确定“腹痛”是“慢性的”,例如,已知属性“慢性”的实体“腹痛1周”、实体“无偏头痛家族史”、实体“无癫痫家族史”、属性“绞痛”的实体“腰部疼痛”与实体“肾结石”之间的关系为:“肾结石”的症状表现通常为“慢性腹痛”、“无偏头痛家族史”、“无癫痫家族史”、“腰部绞痛”,实体“肾结石”与实体“泌尿外科”之间的关系为:“肾结石”需要去“泌尿外科”就诊。由此,使得两个实体“长时间断断续续肚子疼”和“腹痛1周”对齐为属性为“慢性”的“腹痛”,建立其通常为“肾结石”的表现症状之一且需要去“泌尿外科”就诊的关系,融合后的知识如图4所示。

[0046] 融合后的知识数量可观,如果将其全部存入医学知识库205会导致医学知识库205的急剧和不当扩充、存储空间被价值不高的知识占据等问题。针对这些问题,可以对融合后的知识进行上文中记载的方法进行置信度计算,通过设置置信度阈值,将置信度超过阈值的知识存入所述医学知识库205,从而优先将存储空间留给那些可靠的知识,确保所述智能分诊服务器202的存储效率以及医学知识库205的合理扩展及其中所含的知识品质。优选地,所述建立和更新所述医学知识库205还包括:在对抽取的知识进行融合后,对融合后的知识进行推理,得到隐含的知识(推理的具体示例在下文中结合图3进行详细描述),用于进行置信度计算,将置信度超过阈值的隐含的知识也存入所述医学知识库205;如此,加强了医学知识库205的自主学习能力,挖掘出更多有价值的知识。

[0047] 在所建立或更新的医学知识库205的基础上,生成索引以供处理器204使用。通常说来,在本体库的引导下,实现对来自用户的输入的实体识别、索引生成、对半结构化数据和无结构数据的知识抽取、异构数据整合、数据的推理补充中的至少一个。本体库可以利用各种本体生成系统来生成,请注意,如图2所示,本体生成系统和本体库无需驻留在所述智能分诊服务器202上,可以与之通信连接供其应用即可。医学知识库205并不一定在智能分诊服务器202本地,只要检索模块204可以调用其即可,也就是说,医学知识库205可以存储在与智能分诊服务器202通信连接的其他服务器等。

[0048] 优选地,所述智能分诊服务器202可以实现为各种具有处理器和足够运算能力的

装置,例如集中式的服务器、分布式(云端)的服务器等。虽然在这里对包括所述的智能分诊终端201和与之通信连接的所述的基于医学知识库的智能分诊服务器202的智能分诊系统200进行了说明,要注意,智能分诊终端201和智能分诊服务器202可以分别实现为彼此独立和远程的装置,可以分别使用、制造和出售,不一定非要整合到一起。

[0049] 图3示出根据本发明第三实施例的基于医学知识库的智能分诊系统200的分层架构图。如图3所示,智能分诊系统200包括数据层、模型层和应用层。下面分别对各层执行的步骤进行具体说明。

[0050] 1、数据层:从医学指南、医学文献、电子病历、医学网页等数据源中进行知识抽取和知识融合。让计算机阅读理解海量医疗文献,抽取针对某种知识总结出上万条书写规律。如针对“疾病-症状”的关系,文献中描述方式为:“(X疾病)的症状有(Y症状)”,“(X疾病)临床表现为(Y症状)”,“(X疾病)容易引起(Y症状)等不适症状”等。针对两类数据进行分析和抽取:半结构化数据(如电子病历)和无结构的文本数据(如医学文献)。具体来说,利用自然语言处理技术从半结构或者无结构数据中抽取实体、属性以及关系;然后对抽取的知识进行融合(主要包括实体对齐与链接、属性值的决策和确定以及关系建立);基于融合后的知识进行推理,寻找隐含的知识,例如,已有两条知识“西兰花中含有钙元素”和“钙能预防骨质疏松”,则可以得到隐含的知识“西兰花能预防骨质疏松”。基于上述步骤建立一个医学知识库。

[0051] 知识抽取通常包括实体抽取、关系抽取和属性抽取,优选地还包括质量评估,例如对所抽取得到的知识进行置信度评估,只将那些比较可信的知识存入医学知识库205以对其进行扩展。优选地,实体/属性抽取采取自动抽取方式,主要有监督算法(如svm、maxent、CRF)、无监督算法、半监督算法(如步步为营(bootstrapping))、远距离监督(如Distant Supervision)等;关系抽取主要采用监督方法,即利用标注数据训练模型,进行分类。

[0052] 下面以用于实体/属性抽取的步步为营(bootstrapping)算法的具体应用进行说明:

[0053] 输入:给定某一类别的种子实例

[0054] 输出:给定类别的其他实例

[0055] 步骤:

[0056] 1.将依据输入的种子实例扩展进种子集;

[0057] 2.依据种子集中的种子实例,从文献中抽取模版;

[0058] 3.对抽取的模版进行打分并排序,将前n个模版扩展进模版集,n为自然数;

[0059] 4.依据模版集中的模版,从文献中抽取实例;

[0060] 5.对抽取的实例进行打分并排序,将前m个实例扩展进种子集,m为自然数;

[0061] 6.重复步骤2-5直到达到指定停止条件或者实例集不再发生变化。

[0062] 例如给定“便血”对应的疾病名称“憩室炎”,种子集 $S = \{\}$ ,模版集 $P = \{\}$ ;

[0063] 步骤1:令 $S = \{\text{憩室炎}\}$ , $P = \{\}$ ;

[0064] 步骤2:从文献中的句子“憩室炎的病人常有便秘的情形”得到模版“\*的病人常有便秘的情形”,其中\*是通配符。从句子“腹痛伴有便血者应该考虑消化性溃疡或憩室炎”得到模版“腹痛伴有便血者应该考虑消化性溃疡或\*”。这样从文献大数据中可以得到很多模版;



[0065] 步骤3:对上面的模版进行打分,得分的方法有提取特征采用监督方法建立回归模型进行得分预测、或者采用二部图使用图传播方法进行打分等等;

[0066] 步骤4:假设步骤3将模版“\*的病人常有便秘的情形”等扩展进模版集 $P = \{“*的病人常有便秘的情形”, \dots\}$ ,基于该模版从文献中的句子“痔疮的病人常有便秘的情形”得到实例“痔疮”。通过这样的方法可以得到很多实例;

[0067] 步骤5:对步骤4得到的实例打分。得分方法和上面对模版的打分方法类似。将得分高的前 $m$ 个实例扩展进种子集;

[0068] 步骤6:重复上述步骤。输出最后得到的种子集。

[0069] 知识融合包含两个方面,数据层面和语义层面。在结合图2和图4进行的描述中已经对如何进行知识融合进行了举例说明,在此不赘述。

[0070] 2、模型层:该层包含上文所述的智能分诊服务器202,以及实现智能分诊服务器202对医学知识库205的调用,在结合图2进行的描述中已经举例说明,在此不赘述。

[0071] 3、应用层:应用层包含智能分诊终端201,该智能分诊终端201的界面提供交互页面,所述智能分诊终端201连接到基于医学知识库205的智能分诊服务器202,接收用户的检索请求并传输给所述智能分诊服务器202,然后接收来自所述智能分诊服务器202基于所述检索请求从所述医学知识库205获得的检索结果并提供给所述用户,所述用户的检索请求以及所述检索结果均采用自然语言的形式,这样用户可以采用方便地使用自然语言来查询,也容易地以自然语言解读所显示的分析结果。用户可以采用各种方法输入检索请求,包括键盘输入、手写输入、经由选择框勾选或其组合等。优选地,所述检索请求包括用户向所述智能分诊终端201经由选择框勾选或手写输入的问题。如此用户通过智能分诊终端201提出问题,进行查询,也可以通过对智能分诊终端201所显示的一系列症状的选择,从所述智能分诊服务器202获得最终的检索结果,包括疾病建议、概率、建议科室等,并在界面上以自然语言显示给用户。

[0072] 图5示出根据本发明第五实施例的基于医学知识库的智能分诊终端201的界面的示意图。举例说来,所述智能分诊终端201可以实现为各种装置,例如各种便携式的智能装置,包括PDA、手机、平板PC、笔记本电脑等,也可以实现为非便携式的智能装置,包括台式PC等。图5以实现为手机的智能分诊终端201为例,对其交互界面进行说明。

[0073] 该智能分诊终端201连接到基于医学知识库205的智能分诊服务器202,接收用户输入的检索请求并传输给所述智能分诊服务器202,所述检索请求可以包括个人信息(参见图5(a))、症状信息(参见图5(b))和提问(参见图5(c)),然后接收来自所述智能分诊服务器202基于所述检索请求从所述医学知识库205检索的结果并显示给所述用户(参见图5(e)),可以看到所述用户的检索请求以及所述检索结果均采用自然语言的形式。所述检索请求包括用户向所述智能分诊终端经由选择框勾选或打字输入的提问,例如,图5(c)示出了用户输入的问题“我最可能得的病?该去哪里挂号?”。所述智能分诊终端201上提供所述症状信息的选择框,当所述症状信息的选择框被用户选中时,所述智能分诊终端将相应的症状信息传输给所述智能分诊服务器202,例如,如图5(b)所示,当症状信息“心悸”的选择框被用户选中时,就将“心悸”的症状信息传输给智能分诊服务器202。优选地,如图5(b)所示,所述选择框包括生理部位的选择框及对应的症状信息选择框,当一生理部位的选择框“胸部”被用户选中时,显示相应的症状信息选择框,包括“呼吸困难、心悸、咳嗽、胸痛、胸闷”等。再优

选地,当症状信息选择框例如“心悸”被用户选中时,所述智能分诊终端201在界面上提供与所选择的症状信息关联的症状信息的选择框,比如“发烧、高烧”、“头晕头昏”、“眼前发黑”、“畏寒寒战”等,以供用户选择,从而更全面地反映用户的所有症状信息,以便让智能分诊服务器202据此提供更准确的检索结果。优选让用户通过选择框来选中症状信息,因为普通用户对于症状的表述差异较大且有时不太规范,增加了智能分诊服务器202对症状信息进行实体识别和类型分析的难度,通过选择框来选中症状信息,尤其是包括生理部位的选择框及对应的症状信息选择框的复合选择框,能够提示用户更准确、规范和全面地输入其症状信息,减少了智能分诊服务器202对症状信息进行实体识别和类型分析的工作量。用户的提问“我最可能得的病?该去哪里挂号?”传输给智能分诊服务器202,使得后者利用自然语言处理技术对其进行分类,例如预诊类型和分诊类型,来获知用户的真实意图,来从医学知识库205中得到精准的检索结果并提供给所述智能分诊终端201,如图5(e)所示。

[0074] 优选地,所述智能分诊终端201可以依序将所述检索结果显示给用户,显示的次序能够由用户指定。在检索请求至少属于预诊类型的情况下,更优选地,可以将检索结果按照可能疾病的概率从高到低排序,以便在终端有限的显示区域上用户能够直观地获知那些最可能患的疾病的条目,在各个疾病的条目中,可以按照用户的意图依序显示可能疾病、概率和推荐科室等,如图5(e)所示。

[0075] 以上实施例仅为本发明的示例性实施例,不用于限制本发明,本发明的保护范围由权利要求书限定。本领域技术人员可以在本发明的实质和保护范围内,对本发明做出各种修改或等同替换,这种修改或等同替换也应视为落在本发明的保护范围内。

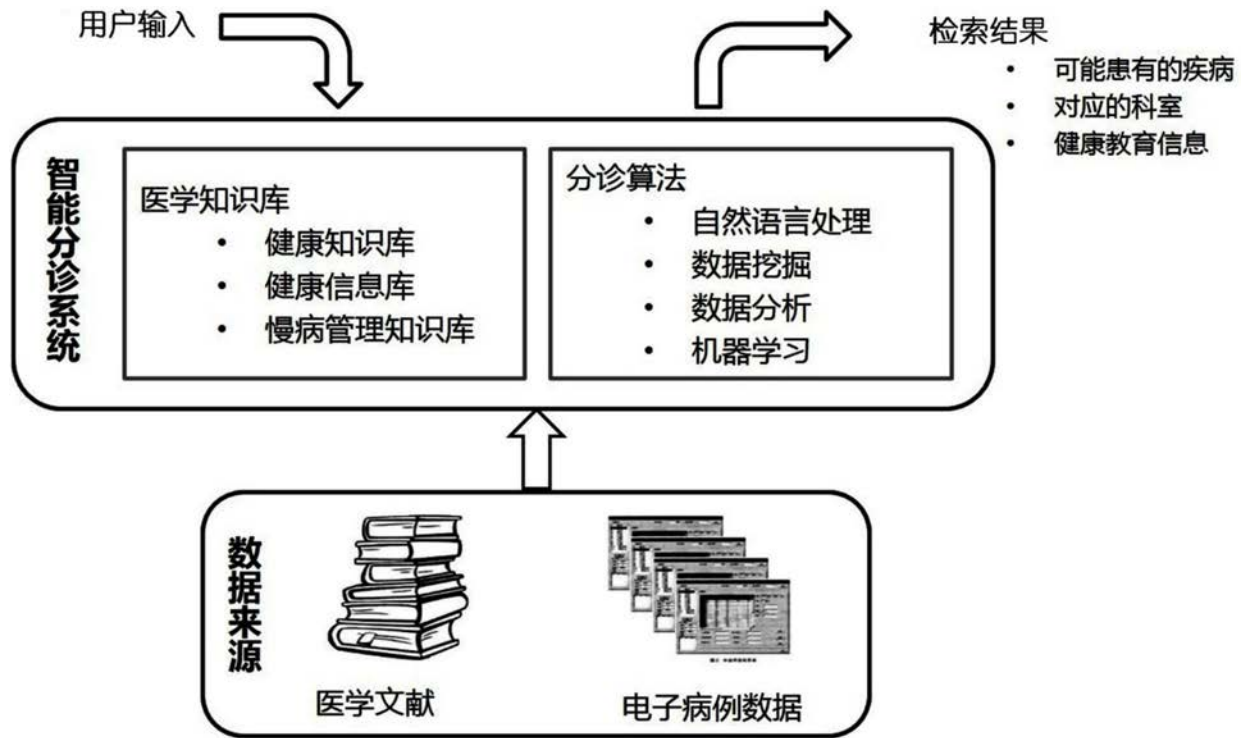


图1

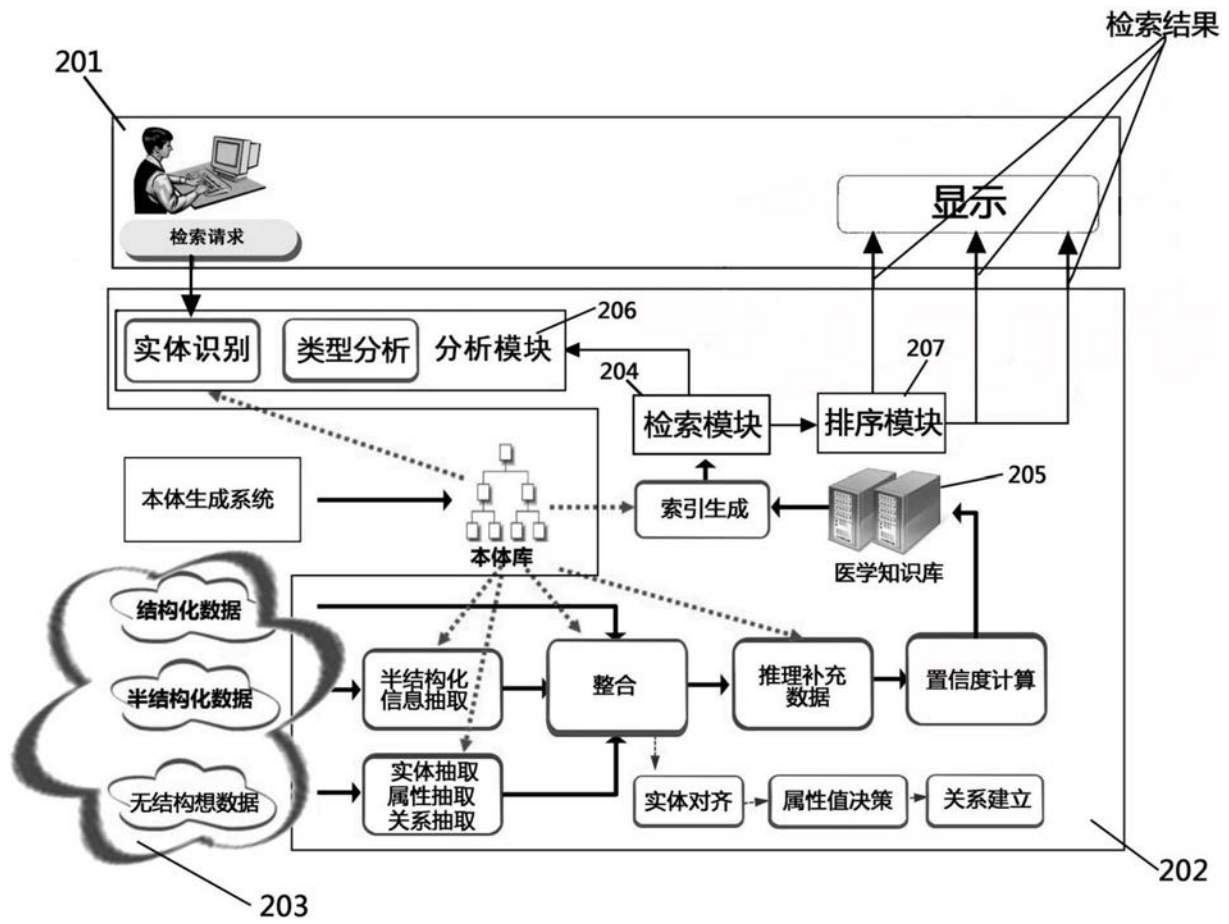


图2

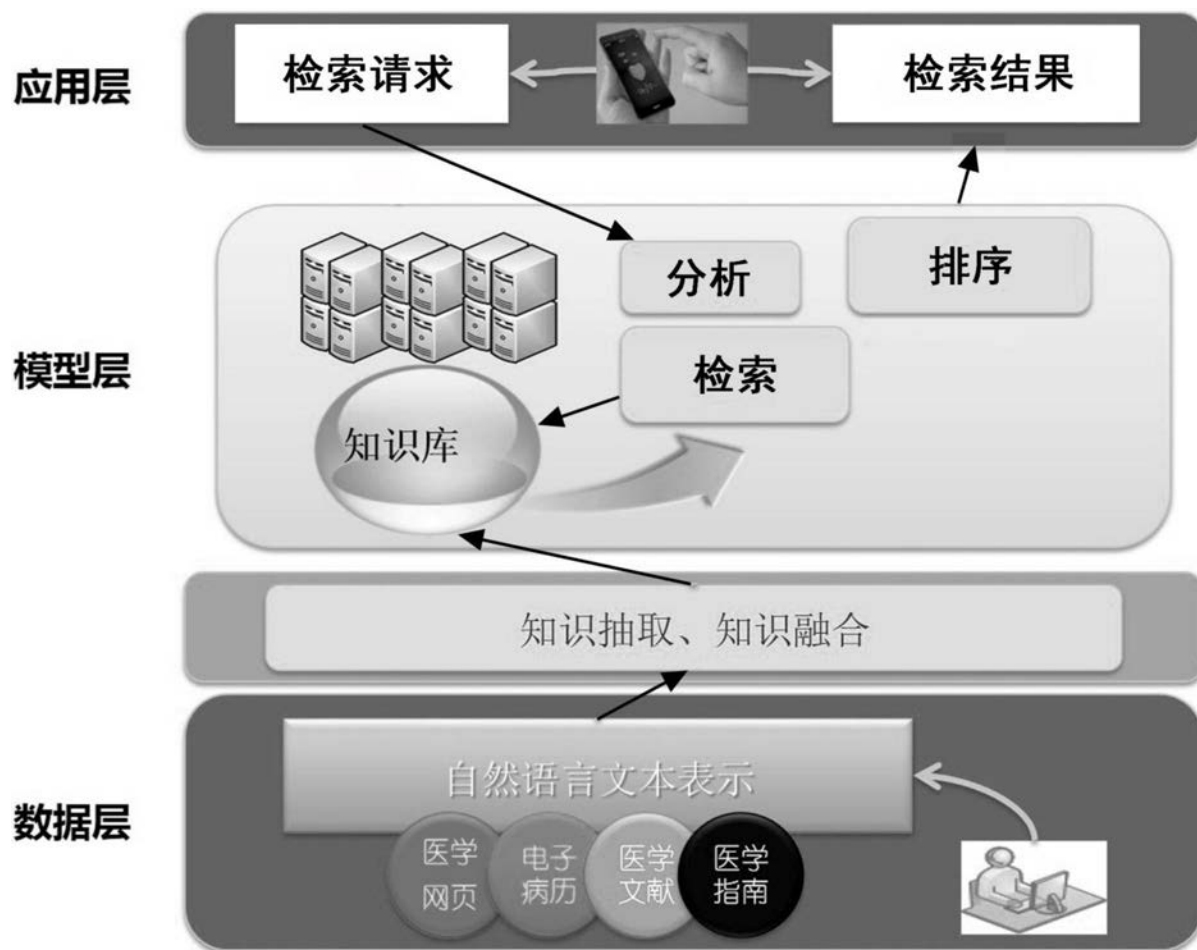


图3

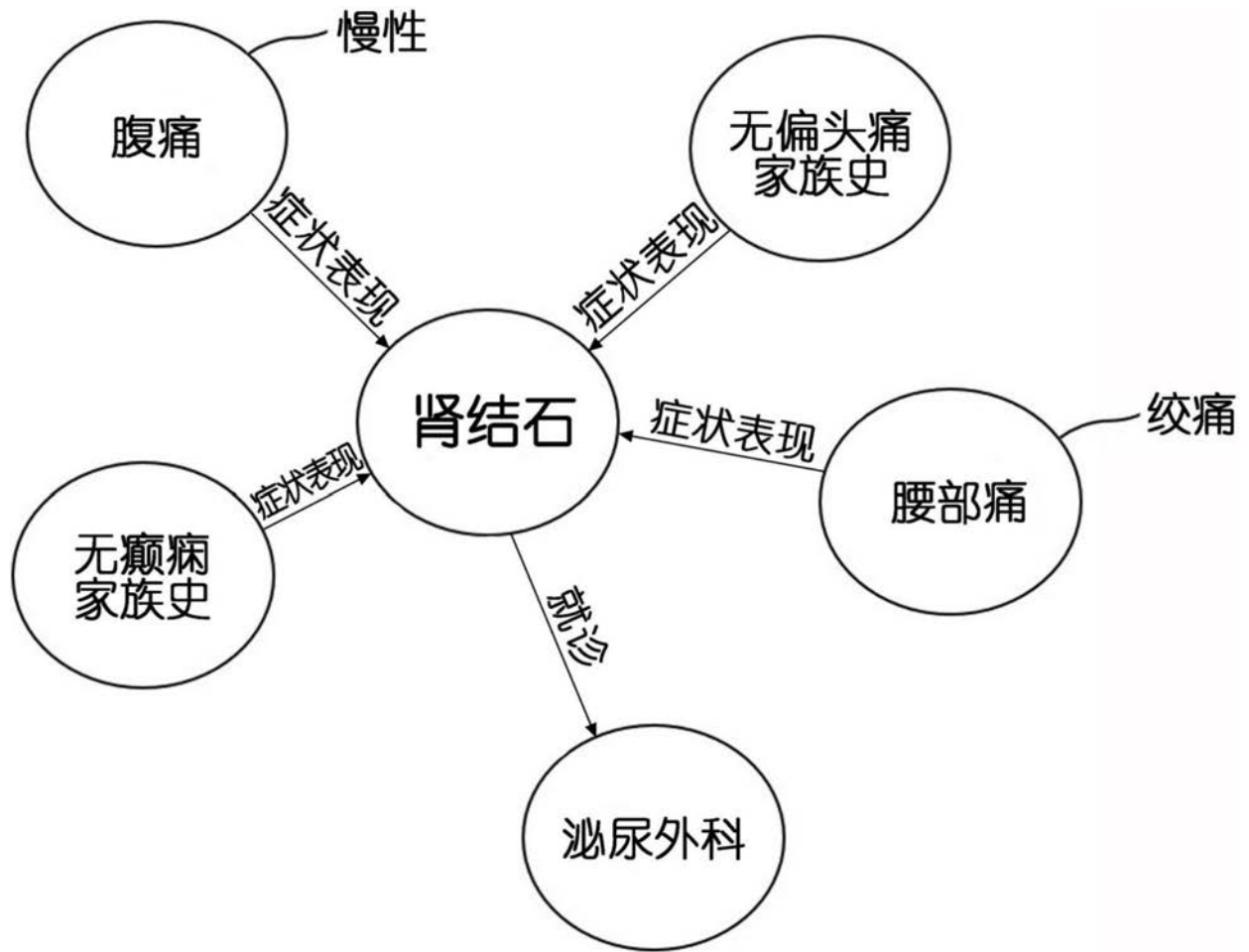


图4



图5 (a)



图5 (b)



图5 (c)



图5 (d)



图5 (e)