(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国 际 局

(43) 国际公布日
2021 年 7 月 15 日 (15.07.2021)

WIPO | PCT

(10) 国际公布号
**WO 2021/139232 A1**

(51) 国际专利分类号:
*G16H 50/20* (2018.01) *G06F 16/36* (2019.01)

(21) 国际申请号: PCT/CN2020/118139

(22) 国际申请日: 2020 年 9 月 27 日 (27.09.2020)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202010621760.X 2020年6月30日 (30.06.2020) CN

(71) 申请人: **平安科技（深圳）有限公司 (PING AN TECHNOLOGY (SHENZHEN)CO., LTD.)** [CN/CN]; 中国广东省深圳市福田区福田街道福安社区益田路5033号平安金融中心23楼, Guangdong 518000 (CN)。

(72) 发明人: **林桂 (LIN, Gui)**; 中国广东省深圳市福田区福田街道福安社区益田路5033号平安金融中心23楼, Guangdong 518000 (CN)。 **黎旭东 (LI, Xudong)**; 中国广东省深圳市福田区福田街道福安社区益田路5033号平安金融中心23楼, Guangdong 518000 (CN)。

(74) 代理人: **深圳市明日今典知识产权代理事务所（普通合伙） (SHENZHEN MINGRIJINDIAN INTELLECTUAL PROPERTY AGENCY FIRM (GENERAL))**; 中国广东省深圳市南山区南头街道安乐社区关口二路15号智恒产业园30栋4层406室, Guangdong 518000 (CN)。

(81) 指定国(除另有指明，要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU,

(54) **Title:** MEDICAL KNOWLEDGE GRAPH-BASED TRIAGE METHOD AND APPARATUS, DEVICE, AND STORAGE MEDIUM

(54) **发明名称**: 基于医疗知识图谱的分诊方法、装置、设备及存储介质

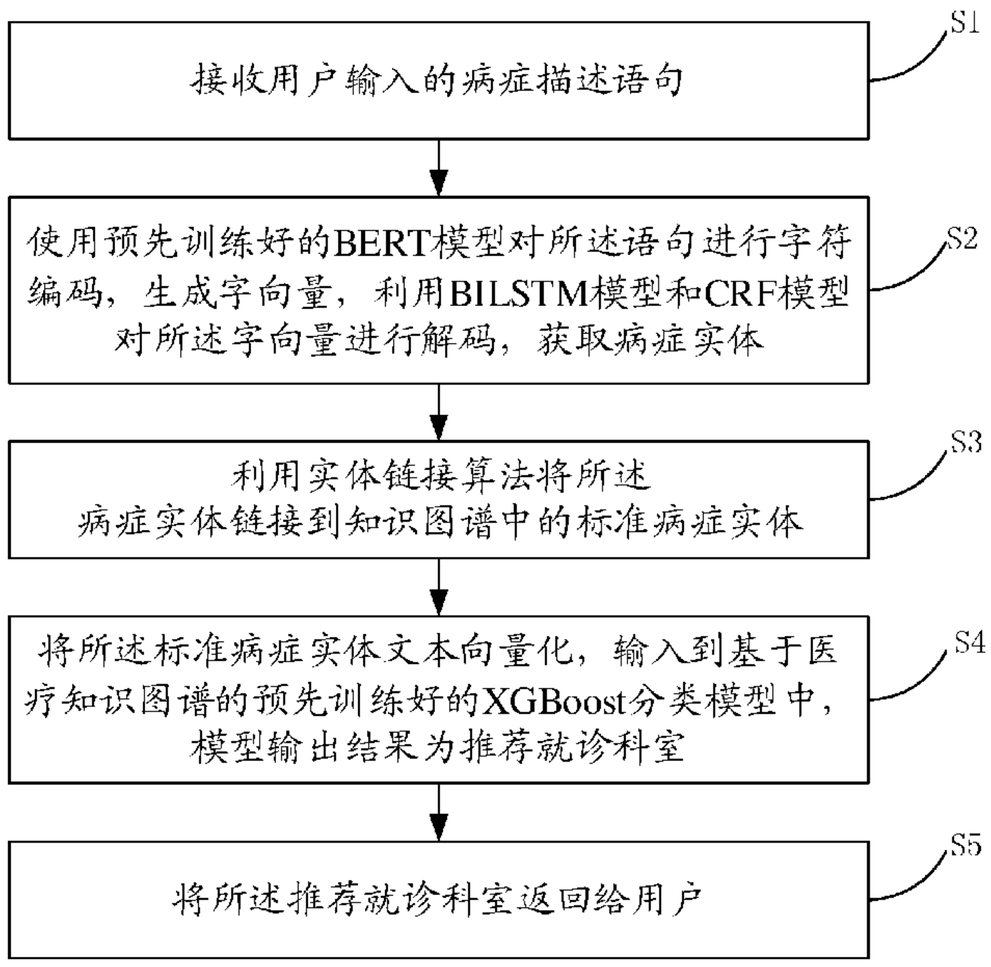



图 1

S1 Receive disease description sentences inputted by a user

S2 Perform character encoding on the sentences by using a pre-trained BERT model so as to generate word vectors, and decode the word vectors by using a BILSTM model and a CRF model so as to obtain a disease entity

S3 Link the disease entity to a standard disease entity in a knowledge graph by using an entity linking algorithm

S4 Perform text vectorization of the standard disease entity, and input same into a medical knowledge graph-based pre-trained XGBoost classification model, the model then outputting a result that is a recommended treatment department

S5 Feed back to the user the recommended treatment department

(57) **Abstract:** A medical knowledge graph-based triage method and apparatus, a device, and a storage medium, relating to the field of digital medicine. The method comprises: receiving disease description sentences inputted by a user (S1); performing character encoding on the sentences by using a pre-trained BERT model so as to generate word vectors, and decoding the word vectors by using a BILSTM model and a CRF model so as to obtain a disease entity (S2); linking the disease entity to a standard disease entity in a knowledge graph by using an entity linking algorithm (S3); performing text vectorization of the standard disease entity, and inputting same into a medical knowledge graph-based pre-trained XGBoost classification model, the model then outputting a result that is a recommended treatment

[见续页]

WO 2021/139232 A1

CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW。

**(84) 指定国**(除另有指明，要求每一种可提供的地区保护)：ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

**本国际公布**:

— 包括国际检索报告(条约第21条(3))。

— 在修改权利要求的期限届满之前进行，在收到该修改后将重新公布(细则48.2(h))。

— 根据申请人的请求，在条约第21条(2)(a)所规定的期限届满之前进行。

department (S4); and feeding back to the user the recommended treatment department (S5). In the method, the chief complaint of the user is directly understood as serialized text, an XGBoost classification model is used for classification, and the classification result is a recommended treatment department. The invention is user friendly, has the advantage of outstanding speech recognition capabilities, and produces accurate triage results.

**(57) 摘要**：一种基于医疗知识图谱的分诊方法、装置、设备及存储介质，涉及数字医疗领域，方法包括：接收用户输入的病症描述语句（S1）；使用预先训练好的BERT模型对所述语句进行字符编码，生成字向量，利用BILSTM模型和CRF模型对所述字向量进行解码，获取病症实体（S2）；利用实体链接算法将所述病症实体链接到知识图谱中的标准病症实体（S3）；将所述标准病症实体文本向量化，输入到基于医疗知识图谱的预先训练好的XGBoost分类模型中，模型输出结果为推荐就诊科室（S4）；将所述推荐就诊科室返回给用户（S5）。该方法直接将用户的主诉视为序列化文本，采用XGBoost分类模型进行分类，分类结果即为推荐就诊科室，使用体验好，并在口语化识别方面优势明显，分诊结果准确。

# 基于医疗知识图谱的分诊方法、装置、设备及存储介质

[0001] 本申请要求于2020年06月30日提交中国专利局、申请号为202010621760.X，发明名称为“基于医疗知识图谱的分诊方法、装置、设备及存储介质”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

[0002] 本申请涉及到数字医疗中的智慧医疗领域，特别是涉及到一种基于医疗知识图谱的分诊方法、装置、设备及存储介质。

## 背景技术

[0003] 医院传统的分诊方式为设置分诊台，对医院的科室分类不了解、不清楚自己病症的患者可以去分诊台向工作人员咨询。但是，患者在向工作人员咨询时，一方面患者通常只能口语化表达自己的病情而无法使用医学术语，另一方面很多分诊台工作人员并非医师专业出身，所以，分诊台工作人员给出的科室推荐的准确性难以得到保证，且此种方式效率较低，人力成本较大。

[0004] 发明人发现，目前市场上已经存在了一些智能分诊产品，如导诊机器人和分诊应用程序等，但在实际使用体验过程中，主要存在以下不足：现有产品在功能上更偏向于用户症状问询自诊，即通过与用户进行多轮交互，常多于三轮，用户通过多次点选应用推荐的相关症状，利用预先设计好的分类算法，推荐出最贴近用户描述的就诊科室，即推荐结果往往来自于较为冗长的用户与应用交互过程，当用户实际症状较简单或者明显时，仍需进行多轮交互；症状识别的智能化程度不高，分诊产品定位于普通用户，输入以口语化症状为主，难以直接映射到标准症状，效果较差，灵活性欠缺；同时，因为产品主要依赖于规则推荐科室，如将“发烧”的病人推荐到呼吸内科等硬性规则，当用户输入复杂的症状组合时，较难推荐正确答案。

## 发明概述

## 技术问题

[0005] 本申请的主要目的为提供一种基于医疗知识图谱的分诊方法、装置、设备及存

储介质，旨在解决目前的分诊产品与用户交互过程冗长、口语化症状识别效果差、导诊效果差的技术问题。

## 问题的解决方案

### 技术解决方案

[0006] 为了实现上述发明目的，特提出了以下技术方案：

[0007] 第一方面，本申请提出一种基于医疗知识图谱的分诊方法，包括：

[0008] 接收用户输入的病症描述语句；

[0009] 使用预先训练好的BERT模型对所述语句进行字符编码，生成字向量，利用BILSTM模型和CRF模型对所述字向量进行解码，获取病症实体；

[0010] 利用实体链接算法将所述病症实体链接到知识图谱中的标准病症实体；

[0011] 将所述标准病症实体文本向量化，输入到基于医疗知识图谱的预先训练好的XGBoost分类模型中，模型输出结果为推荐就诊科室；

[0012] 将所述推荐就诊科室返回给用户。

[0013] 第二方面，本申请实施例还提供一种基于医疗知识图谱的分诊装置，包括：

[0014] 接收单元，用于接收用户输入的症状描述语句；

[0015] 实体获取单元，用于使用预先训练好的BERT模型对所述语句进行字符编码，生成字向量，利用BILSTM模型和CRF模型对所述字向量进行解码，获取病症实体；

[0016] 实体链接单元，利用实体链接算法将所述病症实体链接到知识图谱中的标准病症实体；

[0017] 分类单元，用于将所述标准病症实体文本向量化，输入到基于医疗知识图谱预先训练好的XGBoost分类模型中，模型输出结果为推荐就诊科室；

[0018] 返回单元，用于将所述推荐就诊科室返回给用户。

[0019] 第三方面，本申请还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一种基于医疗知识图谱的分诊方法，所述基于医疗知识图谱的分诊方法的步骤包括：

[0020] 接收用户输入的病症描述语句；

[0021] 使用预先训练好的BERT模型对所述语句进行字符编码，生成字向量，利用BILSTM模型和CRF模型对所述字向量进行解码，获取病症实体；

[0022] 利用实体链接算法将所述病症实体链接到知识图谱中的标准病症实体；

[0023] 将所述标准病症实体文本向量化，输入到基于医疗知识图谱的预先训练好的XGBoost分类模型中，模型输出结果为推荐就诊科室；

[0024] 将所述推荐就诊科室返回给所述用户。

[0025] 第四方面，本申请还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一种基于医疗知识图谱的分诊方法，所述基于医疗知识图谱的分诊方法的步骤包括：

[0026] 接收用户输入的病症描述语句；

[0027] 使用预先训练好的BERT模型对所述语句进行字符编码，生成字向量，利用BILSTM模型和CRF模型对所述字向量进行解码，获取病症实体；

[0028] 利用实体链接算法将所述病症实体链接到知识图谱中的标准病症实体；

[0029] 将所述标准病症实体文本向量化，输入到基于医疗知识图谱的预先训练好的XGBoost分类模型中，模型输出结果为推荐就诊科室；

[0030] 将所述推荐就诊科室返回给所述用户。

## 发明的有益效果

## 有益效果

[0031] 本申请的基于医疗知识图谱的分诊方法、装置、计算机设备及可读存储介质，直接将用户的主诉视为序列化文本，采用XGBoost分类模型进行分类，分类结果即为推荐就诊科室，减少用户多轮输入带来的麻烦，使用体验好，并在口语化识别方面优势明显，分诊结果准确。

## 对附图的简要说明

## 附图说明

[0032] 图1为本申请一实施例的基于医疗知识图谱的分诊方法的流程示意图；

[0033] 图2 为本申请一实施例的基于医疗知识图谱的分诊装置的结构示意框图；

[0034] 图3 为本申请一实施例的计算机设备的结构示意框图。

## 实施该发明的最佳实施例

## 本发明的最佳实施方式

[0035] 为了使本申请的目的、技术方案及优点更加清楚明白，以下结合附图及实施例，对本申请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申请，并不用于限定本申请。

[0036] 本申请涉及到数字医疗中的智慧医疗领域，参照图1，本申请实施例中提供一种基于医疗知识图谱的分诊方法，包括步骤：

[0037] S1、接收用户输入的症状描述语句；

[0038] S2、使用预先训练好的BERT模型对所述语句进行字符编码，生成字向量，利用BILSTM模型和CRF模型对所述字向量进行解码，获取病症实体；

[0039] S3、利用实体链接算法将所述病症实体链接到知识图谱中的标准病症实体；

[0040] S4、将所述标准病症实体文本向量化，输入到基于医疗知识图谱预先训练好的XGBoost分类模型中，模型输出结果为推荐就诊科室；

[0041] S5、将所述推荐就诊科室返回给用户。

[0042] 本方案提出了一种基于医疗知识图谱的分诊方法，可以应用于医院的智能导诊机器人、自助挂号机，也可以应用于用户的智能移动终端设备如手机、平板电脑等，下文统称为用户终端，这些设备可以通过互联网与服务器连接，在服务器端完成数据的存储与分类计算，提高运行与相应速度。

[0043] 如上述步骤S1所述，智能分诊系统是根据用户的症状描述自动推荐就诊科室的系统，首先需要获取用户的症状。目前用户常用的输入方式有文字输入和语音输入，为方便分类模型对用户输入的症状描述语句进行分析处理，需要将用户输入的语音转化为对应的文字。

[0044] 如上述步骤S2所述，用户输入的病症描述语句中通常会含有病症实体，其中病症实体可以是用户自行诊断的疾病实体或症状实体。例如，用户发烧流涕，稍有常识的病人都会首先怀疑自己是感冒。本方案中，利用BERT（Bidirectional Encoder Representations from Transformers）模型对病症实体进行实体识别可以识别到口语化的病症实体。以“头很痛怎么办”为例，使用预先训练好的BERT模型对该语句进行字符级别编码，分别形成字向量作为输入。核心模型使用经典的BILSTM+CRF架构，其中BILSTM是指双向LSTM（Long Short Term Memory），GRF是Conditional Random Field的缩写。其中双向LSTM主要对语句

进行encode操作，值得一提的是，双向LSTM效果比单向LSTM或GRU（Gate Recurrent Unit）更好，由于对语句进行前后遍历，更能捕获语义特征，起到特征提取的作用。然后将其输入CRF层进行解码操作，计算序列中每个字的标签。经过上述操作，模型输出“头很痛”三个字的标签分别为B，I，I，其中B为Begin，表示名词短语的开始；I为Intermediate，表示名词短语的中间；该句的其他字标签均为0，即Other，表示非名词短语。因此“头很痛”为名词短语，在此为症状实体。

[0045] 如上述步骤S3所述，在步骤S2中，利用BERT模型可以识别出一些口语化的症状，识别更加准确，应用范围更广。在此步骤中，根据识别出的口语化症状利用实体链接算法将口语化的症状链接到知识图谱中的标准症状。如“头很痛”这个症状实体就会被链接到知识图谱中的“头痛”这个症状实体。

[0046] 如上述步骤S4所述，利用基于知识图谱预先训练好的XGBoost分类模型对症状序列化文本进行分类，模型输出结果为推荐就诊科室。所述分类模型是通过使用机器学习算法对知识图谱中的训练数据进行训练得到的，所述训练数据包括症状数据和推荐就诊科室数据，将症状数据作为输入，将推荐就诊科室作为输出对分类模型进行训练。XGBoost模型可以很好地完成分类和回归任务，XGBoost最大的特点在于它能够自动利用 CPU 的多线程进行并行，同时在算法上加以改进提高了精度。本方案采用XGBoost分类模型对用户病状序列文本直接进行分类，区别于传统的症状抽取到疾病诊断到科室的流程，分类结果即为推荐就诊科室，算法整体准确率高达94.8%，实际落地准确率高达85%，可以避免直接使用症状实体识别、疾病推理、科室推荐的常规分诊算法流程，减少了用户多轮输入带来的麻烦，使用体验良好。

[0047] 如上述步骤S5所述，在模型给出推荐的就诊科室之后，将所述推荐就诊科室返回给用户。

[0048] 在一个具体的实施例中，所述接收用户输入的症状描述语句的步骤，包括：

[0049] 判断用户输入的语句是否为语音；

[0050] 若是，则将语音转换成文字，并将所述文字作为所述病症描述语句。如上所述，本方案的分诊方法可以应用于应用程序，在医院的设备或用户的移动终端设

备上实现。目前，语音输入是一种常用的语言输入方式，可以给用户带来便利。如果用户输入症状描述语句是语音输入，为了将其转化为计算机或服务器可以处理的语言，需要对语音进行语音识别，转化为对应的文字格式的病状描述语句。

[0051] 在一个具体的实施例中，所述XGBoost分类模型的训练方法，包括：

[0052] 在医疗知识图谱中进行数据抽取，获取训练数据；

[0053] 将训练数据按训练集：开发集：测试集为8:1:1的比例拆分，对XGBoost分类模型进行训练，其中，模型主要参数最大树深max_depth=6，学习率eta=0.5，分类类别num_class=32。

[0054] 如上所述，XGBoost分类模型的训练数据来自于完善的医疗知识图谱，训练过程中，将训练数据按随机抽样的形式将训练集、开发集、测试集切分为8:1:1，其中，训练集数据用于训练模型参数，开发集数据用于调整参数，测试集数据用于对模型算法进行评估。输入到模型中，模型进行特征选择，节点分裂，最终经过softmax层输出。训练完成的模型主要参数：“max_depth”:6,

“eta”:0.5,

“num_class”:32，即最大树深为6，学习率为0.5，输出类别为32，即32个国标一级科室，其中max_depth是指树的最大深度，max_depth越大，模型学习的更加具体，eta表示学习率，也称为迭代率，此数值影响模型的准确度和运行速度，num_class是指输出的类别数量。

[0055] 在一个具体的实施例中，所述在医疗知识图谱中进行数据抽取，获取训练数据的步骤包括：

[0056] 获取知识图谱中的科室实体及各科室实体的分诊属性，所述分诊属性是指该科室实体可以分诊治疗的疾病实体或/和症状属性；

[0057] 将所述科室实体向量化并储存为标签数据集，将所述分诊属性向量化并储存为特征值数据集；

[0058] 根据所述科室实体和所述分诊属性的对应关系建立所述标签数据集和所述特征值数据集的映射关系，得到所述训练数据。

[0059] 如上步骤所述，上述疾病实体和症状属性均可以作为知识图谱中区分该疾病实

体与其他疾病实体的分诊属性，从而根据该分诊属性去进行分诊，以确定分诊之后所对应的科室实体；比如科室实体为呼吸内科，该科室实体的分诊属性包括肺炎这种疾病实体，同时还可以包括肺炎的症状属性“呼吸困难”、“食欲减退”等。若自然问询语句中包含“肺炎”、“呼吸困难”这两个病症关键词，此时可以根据上述病症关键词(该病症关键词中的肺炎为分诊属性中的疾病实体，呼吸困难为分诊属性中的症状属性)确定其疾病实体为“肺炎”，进而根据该疾病实体(也即分诊属性)确定其分诊科室(科室实体)为呼吸内科。故在对分类模型进行训练时，训练数据为科室实体和分诊属性的映射数据集，其中以科室实体作为标签数据，以分诊属性作为特征值数据。

[0060] 在一个具体的实施例中，所述将推荐就诊科室返回给用户的步骤包括：

[0061] 将所述科室推荐结果发送至用户终端；

[0062] 在用户终端的界面上显示或通过语音播报给用户。

[0063] 如上所述，将生成的科室推荐结果发送至用户终端，具体可以应用、短信、微信通知或邮件等方式进行发送，然后在用户终端的界面上显示或通过语音播报给用户。

[0064] 在一个具体的实施例中，所述将所述推荐就诊科室返回给用户的步骤之后包括：

[0065] 接收所述用户终端发送的就诊反馈结果，从所述就诊反馈结果中提取实际就诊科室；

[0066] 当所述推荐就诊科室与所述实际就诊科室不一致时，根据所述标准病症实体与所述实际就诊科室生成修正样本数据；

[0067] 根据所述修正样本数据对知识图谱进行更新。

[0068] 如上所述，用户通过终端接收到科室推荐结果后进行挂号就诊，并根据实际的就诊情况即就诊科室是否正确、是否需要进行转诊等进行就诊反馈。用户向终端输入就诊反馈结果，当就诊科室与实际诊疗科室不一致时，用户可以将实际诊疗科室的科室信息也输入终端中，用户可以通过文字、语音等方式进行输入实际就诊科室信息，当用户终端获取到用户输入的实际就诊科室信息后，根据实际就诊科室信息生成就诊反馈结果并发送给服务器。服务器接收终端发送的

就诊反馈结果后，从就诊反馈结果中提取用户的实际就诊科室。服务器将提取到的实际就诊信息与对应的科室推荐结果中的推荐科室进行比较，判断推荐科室与实际就诊科室是否一致，当推荐科室与实际就诊科室不一致时，服务器根据所述标准病症实体与所述实际就诊科室生成修正样本数据，将修正样本数据更新至知识图谱中，保证知识图谱的数据准确性。

[0069] 在一个具体的实施例中，所述将推荐就诊科室返回给用户的步骤之后还包括：

[0070] 获取用户预约时间和用户位置信息；

[0071] 查找与所述科室推荐结果、所述用户预约时间对应的在诊医生；

[0072] 获取所述在诊医生对应的医院评分、医院位置信息和医生评分；

[0073] 根据所述医院位置信息和所述用户位置信息计算医院距离得分；

[0074] 根据所述医院评分、所述医生评分和所述医院距离得分计算医生推荐得分；

[0075] 根据所述医生推荐得分对所述在诊医生进行排序，根据排序后的所述在诊医生生成医生预约推荐结果。

[0076] 如上所述，用户预约时间为用户的预约就诊时间，用户位置信息为用户的常用居住地址，用户可以向用户终端输入预约就诊时间和常用居住地地址，用户终端也可以自动定位用户的位置信息，用户终端将获取的居住地地址转换为经纬度位置信息并生成用户位置信息，用户终端将获取的用户预约时间和用户位置信息发送给服务器。服务器接收用户终端发送的用户预约时间和用户位置信息，查找与科室推荐结果中的推荐科室对应的在诊医生，获取在诊医生对应的值班时间，筛选出值班时间与用户预约时间相匹配的在诊医生。服务器获取筛选出的在诊医生对应的医院评分、医院位置信息和医生评分，医院位置信息为医院所在地经纬度数据，服务器根据用户位置信息与医院位置信息计算出医院与用户的位置距离，并查找与位置距离对应的医院距离得分，医院距离得分与位置距离的对应关系事先设定并存储，如位置距离5km以内对应的得分为10分，位置距离5km-10km对应的距离得分为8分等。服务器分别获取预先设定的医院评分、医生评分和医院距离得分对应的评分权重，对各项评分和对应的评分权重进行加权求和后得到筛选出的各在诊医生的医生推荐得分。并根据医生推荐得分由高到低的顺序对筛选出的在诊医生进行排序，将排在预设排序位置之前的在

诊医生提取出来，获取提取出的在诊医生的医生信息，医生信息可以包括医生姓名、出诊时间、所属医院、医院地址、医生职称、执医年限、擅长领域等信息，服务器根据获取到的医生信息生成医生预约推荐结果，并将医生预约推荐结果返回给用户终端。服务器通过获取用户的预约时间和位置信息，可以自动为用户匹配综合评分较高的在诊医生，便于用户进行医院和医生的选择。

[0077] 本申请实施例的基于医疗知识图谱的分诊方法，直接将用户的主诉视为序列化文本，采用XGBoost分类模型进行分类，分类结果即为推荐就诊科室，减少用户多轮输入带来的麻烦，使用体验好，并在口语化识别方面优势明显，分诊结果准确。

[0078] 参照图2，本申请实施例中还提供一种基于医疗知识图谱的分诊装置，包括：

[0079] 接收单元1，用于接收用户输入的症状描述语句；

[0080] 实体获取单元2，用于使用预先训练好的BERT模型对所述语句进行字符编码，生成字向量，利用BILSTM模型和CRF模型对所述字向量进行解码，获取病症实体；

[0081] 实体链接单元3，利用实体链接算法将所述病症实体链接到知识图谱中的标准病症实体；

[0082] 分类单元4，用于将所述标准病症实体文本向量化，输入到基于医疗知识图谱预先训练好的XGBoost分类模型中，模型输出结果为推荐就诊科室；

[0083] 返回单元5，用于将所述推荐就诊科室返回给用户。

[0084] 如上所述，可以理解地，本申请中提出的所述基于医疗知识图谱的分诊装置的各组成部分可以实现如上所述基于医疗知识图谱的分诊方法任一项的功能，具体结构不再赘述。

[0085] 参照图3，本申请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3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

于医疗知识图谱等数据。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医疗知识图谱的分诊方法。

[0086] 上述处理器执行上述的基于医疗知识图谱的分诊方法，包括：接收用户输入的病症描述语句；使用预先训练好的BERT模型对所述语句进行字符编码，生成字向量，利用BILSTM模型和CRF模型对所述字向量进行解码，获取病症实体；利用实体链接算法将所述病症实体链接到知识图谱中的标准病症实体；将所述标准病症实体文本向量化，输入到基于医疗知识图谱的预先训练好的XGBoost分类模型中，模型输出结果为推荐就诊科室；将所述推荐就诊科室返回给用户。

[0087] 在一个实施例中，所述接收用户输入的症状描述语句的步骤，包括：判断用户输入的语句是否为语音；若是，则将语音转换成文字，并将所述文字作为所述病症描述语句。在一个实施例中，所述XGBoost分类模型的训练方法，包括：在医疗知识图谱中进行数据抽取，获取训练数据；将训练数据按训练集：开发集：测试集为8:1:1的比例拆分，对XGBoost分类模型进行训练，其中，模型主要参数max_depth=6，eta=0.5，num_class=32。

[0088] 在一个具体的实施例中，所述在医疗知识图谱中进行数据抽取，获取训练数据的步骤包括：获取知识图谱中的科室实体及各科室实体的分诊属性，所述分诊属性是指该科室实体可以分诊治疗的疾病实体或/和症状属性；将所述科室实体向量化并储存为标签数据集，将所述分诊属性向量化并储存为特征值数据集；根据所述科室实体和所述分诊属性的对应关系建立所述标签数据集和所述特征值数据集的映射关系，得到所述训练数据。

[0089] 在一个具体的实施例中，所述将推荐就诊科室返回给用户的步骤包括：将所述科室推荐结果发送至用户终端；在用户终端的界面上显示或通过语音播报给用户。

[0090] 在一个具体的实施例中，所述将所述推荐就诊科室返回给用户的步骤之后包括：接收所述用户终端发送的就诊反馈结果，从所述就诊反馈结果中提取实际就诊科室；当所述推荐就诊科室与所述实际就诊科室不一致时，根据所述标准病症实体与所述实际就诊科室生成修正样本数据；根据所述修正样本数据对知识

图谱进行更新。

[0091] 在一个具体的实施例中，所述将推荐就诊科室返回给用户的步骤之后还包括：获取用户预约时间和用户位置信息；查找与所述科室推荐结果、所述用户预约时间对应的在诊医生；获取所述在诊医生对应的医院评分、医院位置信息和医生评分；根据所述医院位置信息和所述用户位置信息计算医院距离得分；根据所述医院评分、所述医生评分和所述医院距离得分计算医生推荐得分；根据所述医生推荐得分对所述在诊医生进行排序，根据排序后的所述在诊医生生成医生预约推荐结果。

[0092] 本申请一实施例还提供一种计算机可读存储介质，所述计算机可读存储介质可以是非易失性，也可以是易失性的，其上存储有计算机程序，计算机程序被处理器执行时实现一种基于医疗知识图谱的分诊方法，包括步骤：接收用户输入的病症描述语句；使用预先训练好的BERT模型对所述语句进行字符编码，生成字向量，利用BILSTM模型和CRF模型对所述字向量进行解码，获取病症实体；利用实体链接算法将所述病症实体链接到知识图谱中的标准病症实体；将所述标准病症实体文本向量化，输入到基于医疗知识图谱的预先训练好的XGBoost分类模型中，模型输出结果为推荐就诊科室；将所述推荐就诊科室返回给用户。

[0093] 上述执行的基于医疗知识图谱的分诊方法，直接将用户的主诉视为序列化文本，采用XGBoost分类模型进行分类，分类结果即为推荐就诊科室，减少用户多轮输入带来的麻烦，使用体验好，并在口语化识别方面优势明显，分诊结果准确。

[0094] 在一个实施例中，所述接收用户输入的症状描述语句的步骤，包括：判断用户输入的语句是否为语音；若是，则将语音转换成文字，并将所述文字作为所述病症描述语句。

[0095] 在一个实施例中，所述XGBoost分类模型的训练方法，包括：在医疗知识图谱中进行数据抽取，获取训练数据；将训练数据按训练集：开发集：测试集为8:1:1的比例拆分，对XGBoost分类模型进行训练，其中，模型主要参数max_depth=6，eta=0.5，num_class=32。

[0096] 在一个具体的实施例中，所述在医疗知识图谱中进行数据抽取，获取训练数据

的步骤包括：获取知识图谱中的科室实体及各科室实体的分诊属性，所述分诊属性是指该科室实体可以分诊治疗的疾病实体或/和症状属性；将所述科室实体向量化并储存为标签数据集，将所述分诊属性向量化并储存为特征值数据集；根据所述科室实体和所述分诊属性的对应关系建立所述标签数据集和所述特征值数据集的映射关系，得到所述训练数据。

[0097] 在一个具体的实施例中，所述将推荐就诊科室返回给用户的步骤包括：将所述科室推荐结果发送至用户终端；在用户终端的界面上显示或通过语音播报给用户。

[0098] 在一个具体的实施例中，所述将所述推荐就诊科室返回给用户的步骤之后包括：接收所述用户终端发送的就诊反馈结果，从所述就诊反馈结果中提取实际就诊科室；当所述推荐就诊科室与所述实际就诊科室不一致时，根据所述标准病症实体与所述实际就诊科室生成修正样本数据；根据所述修正样本数据对知识图谱进行更新。

[0099] 在一个具体的实施例中，所述将推荐就诊科室返回给用户的步骤之后还包括：获取用户预约时间和用户位置信息；查找与所述科室推荐结果、所述用户预约时间对应的在诊医生；获取所述在诊医生对应的医院评分、医院位置信息和医生评分；根据所述医院位置信息和所述用户位置信息计算医院距离得分；根据所述医院评分、所述医生评分和所述医院距离得分计算医生推荐得分；根据所述医生推荐得分对所述在诊医生进行排序，根据排序后的所述在诊医生生成医生预约推荐结果。

[0100] 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取存储介质中，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，本申请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。

[0101] 需要说明的是，在本文中，术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含，从而使得包括一系列要素的过程、装置、物品或者方法不仅包括那些要素，而且还包括没有明确列出的其他要素，或者是还包括为这种过程、装置、物品或者方法所固有的要素。在没有更多限制的情况下，由语句“包括一个……”限定的要素，并不排除在包括该要素的过程、装置、物品或者方法中还存在另外的相同要素。

[0102] 以上所述仅为本申请的优选实施例，并非因此限制本申请的专利范围，凡是利用本申请说明书及附图内容所作的等效结构或等效流程变换，或直接或间接运用在其他相关的技术领域，均同理包括在本申请的专利保护范围内。

# 权利要求书

[权利要求 1] 一种基于医疗知识图谱的分诊方法，其中，包括：

接收用户输入的病症描述语句；

使用预先训练好的BERT模型对所述语句进行字符编码，生成字向量，

利用BILSTM模型和CRF模型对所述字向量进行解码，获取病症实体；

利用实体链接算法将所述病症实体链接到知识图谱中的标准病症实体；

将所述标准病症实体文本向量化，输入到基于医疗知识图谱的预先训练好的XGBoost分类模型中，模型输出结果为推荐就诊科室；

将所述推荐就诊科室返回给所述用户。

[权利要求 2] 根据权利要求1所述的基于医疗知识图谱的分诊方法，其中，所述接收用户输入的症状描述语句的步骤，包括：

判断用户输入的语句是否为语音；

若是，则将语音转换成文字，并将所述文字作为所述病症描述语句。

[权利要求 3] 根据权利要求1所述的基于医疗知识图谱的分诊方法，其中，所述XGBoost分类模型的训练方法，包括：

在医疗知识图谱中进行数据抽取，获取训练数据；

将训练数据按训练集：开发集：测试集为8:1:1的比例拆分，对XGBoost分类模型进行训练，其中，模型主要参数最大树深max_depth=6，学习率eta=0.5，分类类别num_class=32。

[权利要求 4] 根据权利要求3所述的基于医疗知识图谱的分诊方法，其中，所述在医疗知识图谱中进行数据抽取，获取训练数据的步骤包括：

获取知识图谱中的科室实体及各科室实体的分诊属性，所述分诊属性是指所述科室实体可以分诊治疗的疾病实体或/和症状属性；

将所述科室实体向量化并储存为标签数据集，以及将所述分诊属性向量化并储存为特征值数据集；

根据所述科室实体和所述分诊属性的对应关系建立所述标签数据集和所述特征值数据集的映射关系，得到所述训练数据。

[权利要求 5] 根据权利要求1所述的基于医疗知识图谱的分诊方法，其中，所述将推荐就诊科室返回给用户的步骤包括：

将所述科室推荐结果发送至用户终端；

在用户终端的界面上显示或通过语音播报给用户。

[权利要求 6] 根据权利要求1所述的基于医疗知识图谱的分诊方法，其中，所述将所述推荐就诊科室返回给用户的步骤之后包括：

接收所述用户终端发送的就诊反馈结果，从所述就诊反馈结果中提取实际就诊科室；

当所述推荐就诊科室与所述实际就诊科室不一致时，根据所述标准病症实体与所述实际就诊科室生成修正样本数据；

根据所述修正样本数据对知识图谱进行更新。

[权利要求 7] 根据权利要求1所述的基于医疗知识图谱的分诊方法，其中，

所述将推荐就诊科室返回给用户的步骤之后还包括：

获取用户预约时间和用户位置信息；

查找与所述科室推荐结果、所述用户预约时间对应的在诊医生；

获取所述在诊医生对应的医院评分、医院位置信息和医生评分；

根据所述医院位置信息和所述用户位置信息计算医院距离得分；

根据所述医院评分、所述医生评分和所述医院距离得分计算医生推荐得分；

根据所述医生推荐得分对所述在诊医生进行排序，根据排序后的所述在诊医生生成医生预约推荐结果。

[权利要求 8] 一种基于医疗知识图谱的分诊装置，其中，包括：

接收单元，用于接收用户输入的症状描述语句；

实体获取单元，用于使用预先训练好的BERT 模型对所述语句进行字符编码，生成字向量，利用BILSTM模型和CRF模型对所述字向量进行解码，获取病症实体；

实体链接单元，利用实体链接算法将所述病症实体链接到知识图谱中的标准病症实体；

分类单元，用于将所述标准病症实体文本向量化，输入到基于医疗知识图谱预先训练好的XGBoost分类模型中，模型输出结果为推荐就诊科室；

返回单元，用于将所述推荐就诊科室返回给用户。

[权利要求 9] 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，其中，所述处理器执行所述计算机程序时实现一种基于医疗知识图谱的分诊方法，所述基于医疗知识图谱的分诊方法的步骤包括：

接收用户输入的病症描述语句；

使用预先训练好的BERT模型对所述语句进行字符编码，生成字向量，利用BILSTM模型和CRF模型对所述字向量进行解码，获取病症实体；

利用实体链接算法将所述病症实体链接到知识图谱中的标准病症实体；

将所述标准病症实体文本向量化，输入到基于医疗知识图谱的预先训练好的XGBoost分类模型中，模型输出结果为推荐就诊科室；

将所述推荐就诊科室返回给所述用户。

[权利要求 10] 根据权利要求9所述的计算机设备，其中，所述接收用户输入的症状描述语句的步骤，包括：

判断用户输入的语句是否为语音；

若是，则将语音转换成文字，并将所述文字作为所述病症描述语句。

[权利要求 11] 根据权利要求9所述的计算机设备，其中，所述XGBoost分类模型的训练方法，包括：

在医疗知识图谱中进行数据抽取，获取训练数据；

将训练数据按训练集：开发集：测试集为8:1:1的比例拆分，对XGBoost分类模型进行训练，其中，模型主要参数最大树深max_depth=6，学习率eta=0.5，分类类别num_class=32。

[权利要求 12] 根据权利要求11所述的计算机设备，其中，所述在医疗知识图谱中进行数据抽取，获取训练数据的步骤包括：

获取知识图谱中的科室实体及各科室实体的分诊属性，所述分诊属性

是指所述科室实体可以分诊治疗的疾病实体或/和症状属性；

将所述科室实体向量化并储存为标签数据集，以及将所述分诊属性向量化并储存为特征值数据集；

根据所述科室实体和所述分诊属性的对应关系建立所述标签数据集和所述特征值数据集的映射关系，得到所述训练数据。

[权利要求 13] 根据权利要求9所述的计算机设备，其中，所述将推荐就诊科室返回给用户的步骤包括：

将所述科室推荐结果发送至用户终端；

在用户终端的界面上显示或通过语音播报给用户。

[权利要求 14] 根据权利要求9所述的计算机设备，其中，所述将所述推荐就诊科室返回给用户的步骤之后包括：

接收所述用户终端发送的就诊反馈结果，从所述就诊反馈结果中提取实际就诊科室；

当所述推荐就诊科室与所述实际就诊科室不一致时，根据所述标准病症实体与所述实际就诊科室生成修正样本数据；

根据所述修正样本数据对知识图谱进行更新。

[权利要求 15] 一种计算机可读存储介质，其上存储有计算机程序，其中，所述计算机程序被处理器执行时实现一种基于医疗知识图谱的分诊方法，所述基于医疗知识图谱的分诊方法的步骤包括：

接收用户输入的病症描述语句；

使用预先训练好的BERT模型对所述语句进行字符编码，生成字向量，

利用BILSTM模型和CRF模型对所述字向量进行解码，获取病症实体；

利用实体链接算法将所述病症实体链接到知识图谱中的标准病症实体；

将所述标准病症实体文本向量化，输入到基于医疗知识图谱的预先训练好的XGBoost分类模型中，模型输出结果为推荐就诊科室；

将所述推荐就诊科室返回给所述用户。

[权利要求 16] 根据权利要求15所述的计算机可读存储介质，其中，所述接收用户输

入的症状描述语句的步骤，包括：

判断用户输入的语句是否为语音；

若是，则将语音转换成文字，并将所述文字作为所述病症描述语句。

[权利要求 17] 根据权利要求15所述的计算机可读存储介质，其中，所述XGBoost分类模型的训练方法，包括：

在医疗知识图谱中进行数据抽取，获取训练数据；

将训练数据按训练集：开发集：测试集为8:1:1的比例拆分，对XGBoost分类模型进行训练，其中，模型主要参数最大树深max_depth=6，学习率eta=0.5，分类类别num_class=32。

[权利要求 18] 根据权利要求17所述的计算机可读存储介质，其中，所述在医疗知识图谱中进行数据抽取，获取训练数据的步骤包括：

获取知识图谱中的科室实体及各科室实体的分诊属性，所述分诊属性是指所述科室实体可以分诊治疗的疾病实体或/和症状属性；

将所述科室实体向量化并储存为标签数据集，以及将所述分诊属性向量化并储存为特征值数据集；

根据所述科室实体和所述分诊属性的对应关系建立所述标签数据集和所述特征值数据集的映射关系，得到所述训练数据。

[权利要求 19] 根据权利要求15所述的计算机可读存储介质，其中，所述将推荐就诊科室返回给用户的步骤包括：

将所述科室推荐结果发送至用户终端；

在用户终端的界面上显示或通过语音播报给用户。

[权利要求 20] 根据权利要求15所述的计算机可读存储介质，其中，所述将所述推荐就诊科室返回给用户的步骤之后包括：

接收所述用户终端发送的就诊反馈结果，从所述就诊反馈结果中提取实际就诊科室；

当所述推荐就诊科室与所述实际就诊科室不一致时，根据所述标准病症实体与所述实际就诊科室生成修正样本数据；

根据所述修正样本数据对知识图谱进行更新。

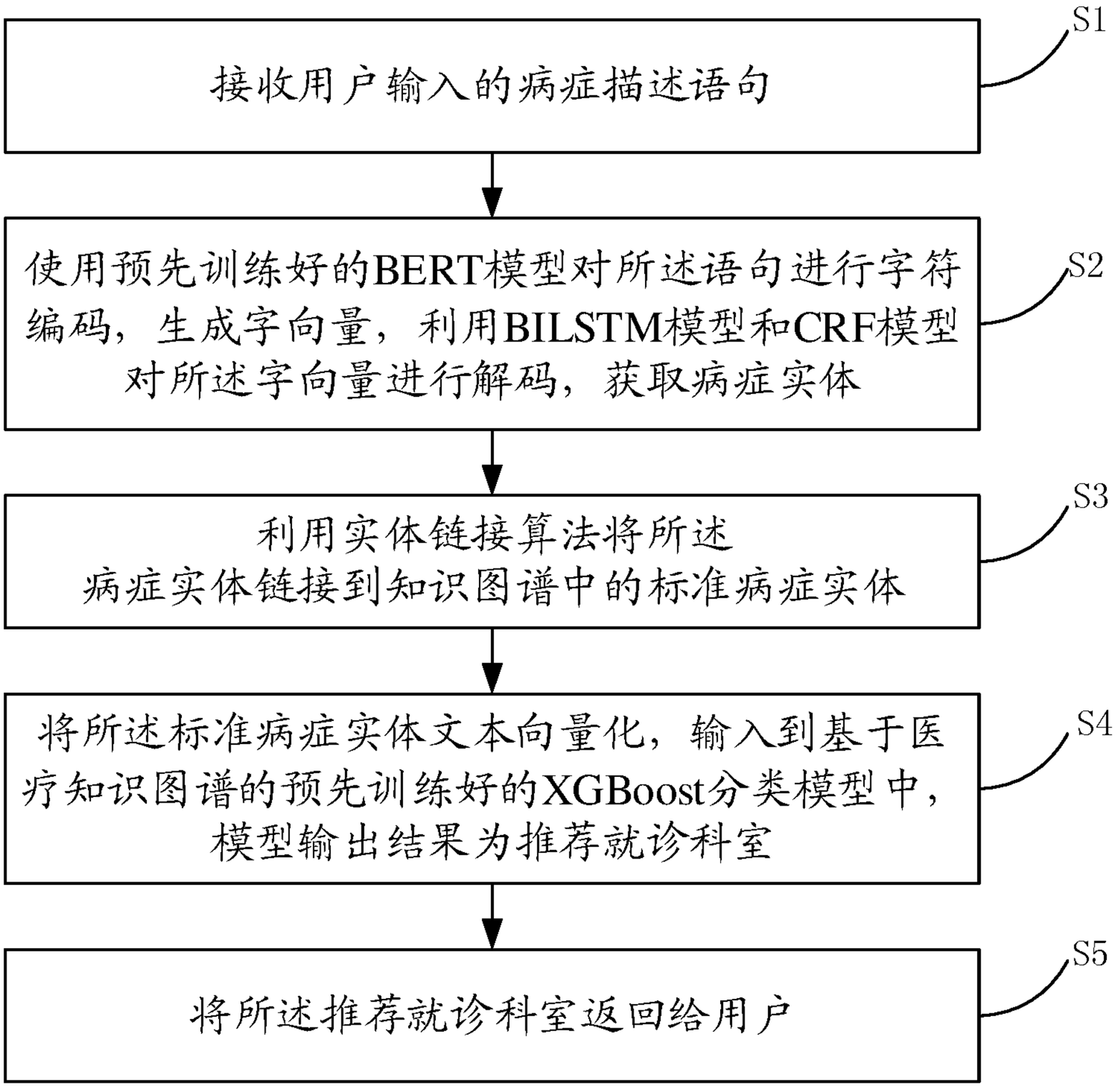

接收用户输入的病症描述语句
S1
使用预先训练好的BERT模型对所述语句进行字符编码，生成字向量，利用BILSTM模型和CRF模型对所述字向量进行解码，获取病症实体
S2
利用实体链接算法将所述病症实体链接到知识图谱中的标准病症实体
S3
将所述标准病症实体文本向量化，输入到基于医疗知识图谱的预先训练好的XGBoost分类模型中，模型输出结果为推荐就诊科室
S4
将所述推荐就诊科室返回给用户
S5

图 1

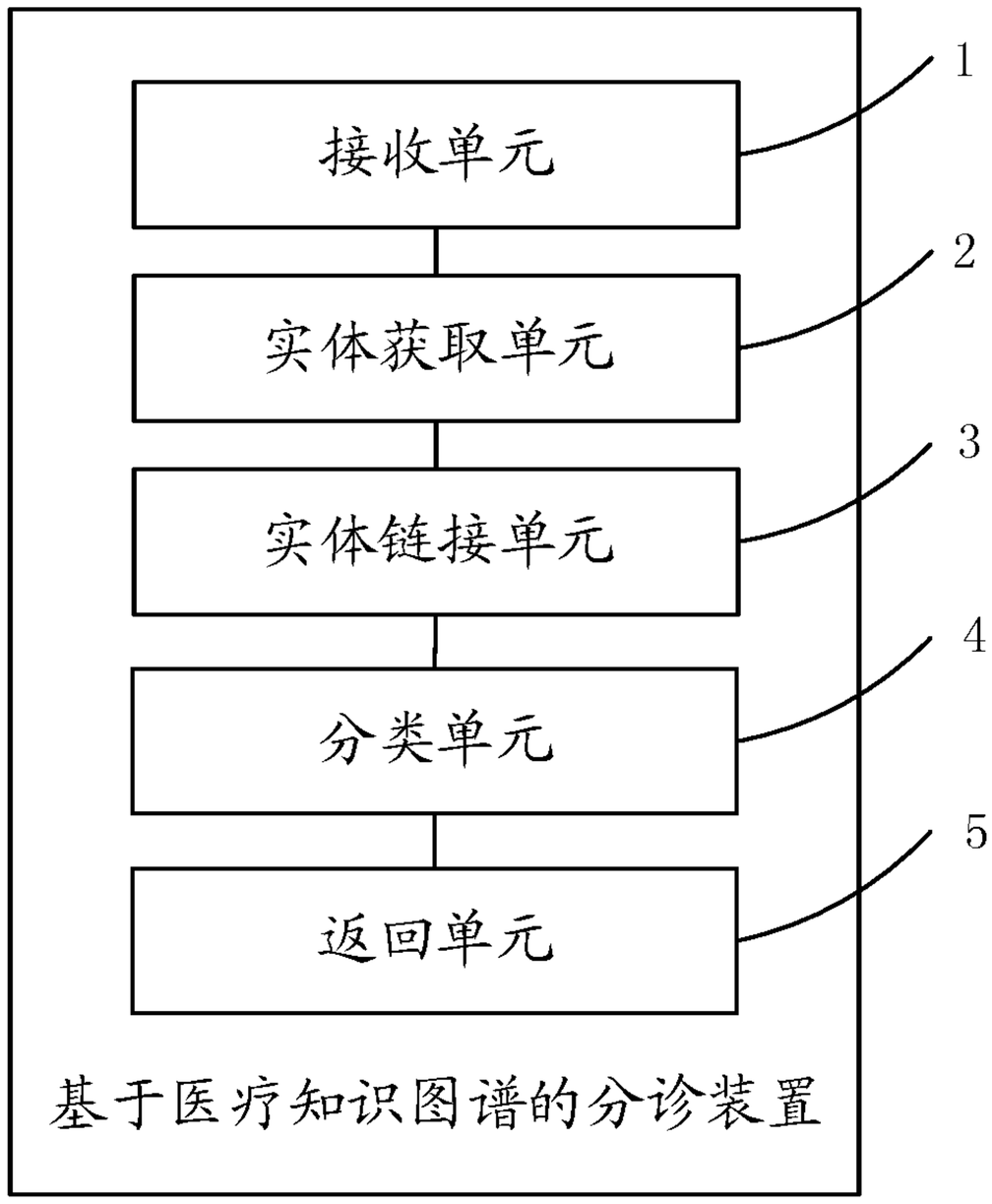

接收单元
1
实体获取单元
2
实体链接单元
3
分类单元
4
返回单元
5
基于医疗知识图谱的分诊装置

图 2

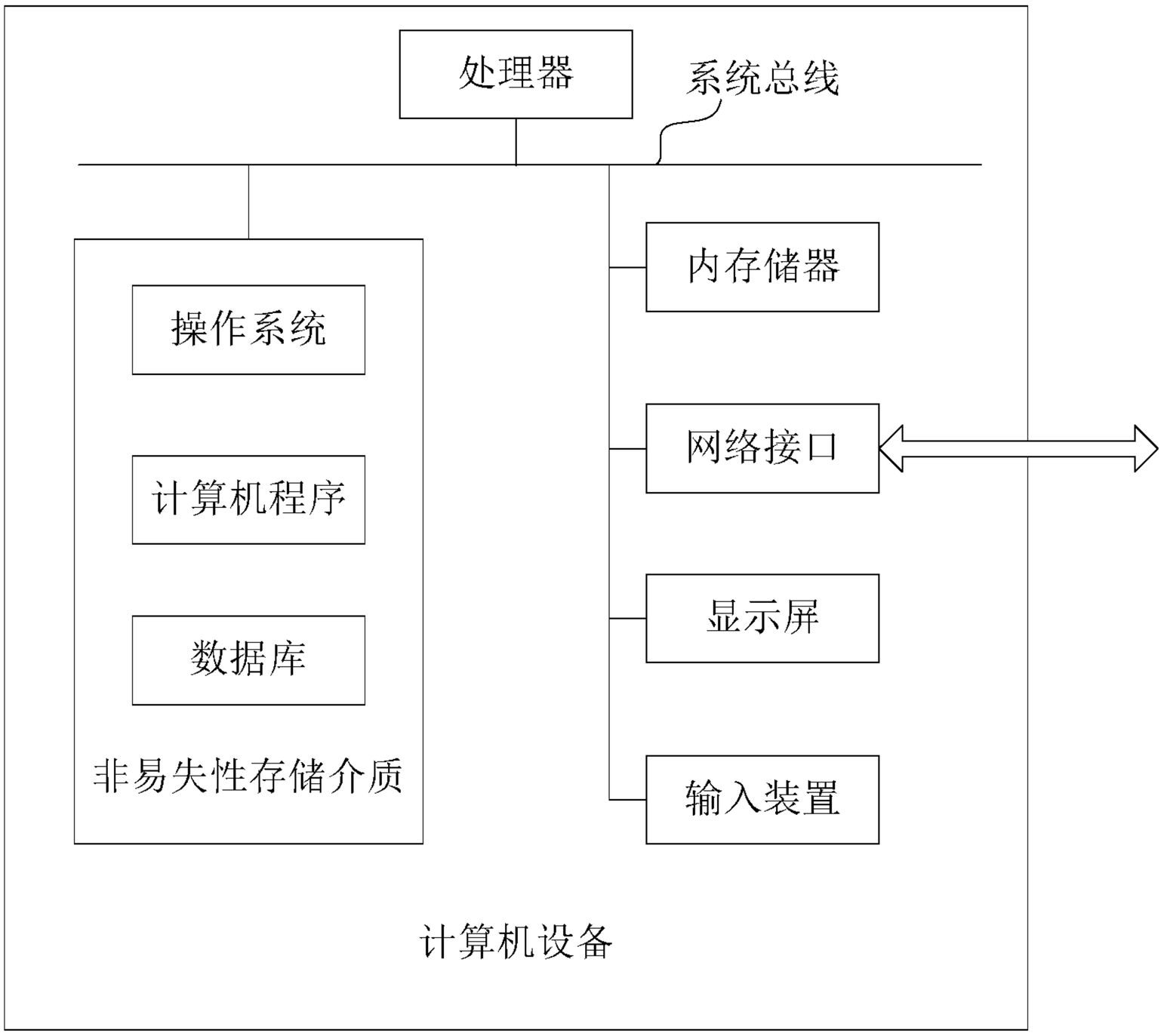

处理器
系统总线
操作系统
计算机程序
数据库
非易失性存储介质
内存储器
网络接口
显示屏
输入装置
计算机设备

图 3

# INTERNATIONAL SEARCH REPORT

International application No.
**PCT/CN2020/118139**

## A. CLASSIFICATION OF SUBJECT MATTER

G16H 50/20(2018.01)i; G06F 16/36(2019.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G16H, G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNPAT, WPI, EPODOC, CNKI, IEEE: 知识图谱, 导诊, 分诊, 推荐, 医生, 科室, 问答, 条件随机场, 双向长短时记忆, 解码, 编码, 实体链接算法, 实体, 分类, knowledge, graph, guidance, triage, recommend, doctor, department, question, answer, CRF, XGBoost, BiLSTM, BERT, entity, link, algorithm, classify

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| Y | CN 110069631 A (TENCENT TECHNOLOGY SHENZHEN CO., LTD.) 30 July 2019 (2019-07-30)<br>description, paragraphs [0142]-[0204], claims 1-15 | 1-20 |
| Y | CN 111310471 A (SHAANXI NORMAL UNIVERSITY) 19 June 2020 (2020-06-19)<br>description, paragraph [0039] | 1-20 |
| Y | CN 110085307 A (EAST CHINA UNIVERSITY OF SCIENCE AND TECHNOLOGY et al.) 02 August 2019 (2019-08-02)<br>claim 9 | 1-20 |
| Y | CN 111339277 A (INDUSTRIAL AND COMMERCIAL BANK OF CHINA LIMITED) 26 June 2020 (2020-06-26)<br>description, paragraphs [0051]-[0069] | 1-20 |
| A | CN 110516260 A (TENCENT TECHNOLOGY SHENZHEN CO., LTD.) 29 November 2019 (2019-11-29)<br>entire document | 1-20 |

☑ Further documents are listed in the continuation of Box C. ☑ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

| Date of the actual completion of the international search | Date of mailing of the international search report |
|---|---|
| **18 March 2021** | **30 March 2021** |
| Name and mailing address of the ISA/CN<br>**China National Intellectual Property Administration (ISA/CN)**<br>**No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing 100088**<br>**China**<br>Facsimile No. **(86-10)62019451** | Authorized officer<br><br>Telephone No. |

# INTERNATIONAL SEARCH REPORT

International application No.
**PCT/CN2020/118139**

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| A | CN 110993078 A (UNION HOSPITAL OF TONGJI MEDICAL COLLEGE, HUAZHONG UNIVERSITY OF SCIENCE AND TECHNOLOGY) 10 April 2020 (2020-04-10)<br>entire document | 1-20 |
| A | CN 109659013 A (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) 19 April 2019 (2019-04-19)<br>entire document | 1-20 |
| A | US 2019317994 A1 (TATA CONSULTANCY SERVICES LIMITED) 17 October 2019 (2019-10-17)<br>entire document | 1-20 |

**INTERNATIONAL SEARCH REPORT**
**Information on patent family members**

International application No.
**PCT/CN2020/118139**

| Patent document cited in search report | | | Publication date (day/month/year) | Patent family member(s) | | | Publication date (day/month/year) |
|---|---|---|---|---|---|---|---|
| CN | 110069631 | A | 30 July 2019 | None | | | |
| CN | 111310471 | A | 19 June 2020 | None | | | |
| CN | 110085307 | A | 02 August 2019 | None | | | |
| CN | 111339277 | A | 26 June 2020 | None | | | |
| CN | 110516260 | A | 29 November 2019 | None | | | |
| CN | 110993078 | A | 10 April 2020 | None | | | |
| CN | 109659013 | A | 19 April 2019 | None | | | |
| US | 2019317994 | A1 | 17 October 2019 | EP | 3557439 | A1 | 23 October 2019 |
| | | | | BR | 102019007765 | A2 | 05 November 2019 |
| | | | | AU | 2019202632 | A1 | 31 October 2019 |
| | | | | CA | 3040373 | A1 | 16 October 2019 |
| | | | | MX | 2019004407 | A | 25 November 2019 |
| | | | | JP | 2019215857 | A | 19 December 2019 |
| | | | | IN | 201821014473 | A | 18 October 2019 |

# 国际检索报告

国际申请号
PCT/CN2020/118139

A. 主题的分类

G16H 50/20(2018.01)i；G06F 16/36(2019.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

G16H，G06F

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称，和使用的检索词（如使用））

CNPAT，WPI，EPODOC，CNKI，IEEE：知识图谱，导诊，分诊，推荐，医生，科室，问答，条件随机场，双向长短时记忆，解码，编码，实体链接算法，实体，分类，knowledge，graph，guidance，triage，recommend，doctor，department，question，answer，CRF， XGBoost，BiLSTM，BERT，entity，link，algorithm，classify

C. 相关文件

| 类 型* | 引用文件，必要时，指明相关段落 | 相关的权利要求 |
|---|---|---|
| Y | CN 110069631 A (腾讯科技深圳有限公司) 2019年 7月 30日 (2019-07-30)<br>说明书第[0142]-[0204]段，权利要求1-15 | 1-20 |
| Y | CN 111310471 A (陕西师范大学) 2020年 6月 19日 (2020-06-19)<br>说明书第[0039]段 | 1-20 |
| Y | CN 110085307 A (华东理工大学 等) 2019年 8月 2日 (2019-08-02)<br>权利要求9 | 1-20 |
| Y | CN 111339277 A (中国工商银行股份有限公司) 2020年 6月 26日 (2020-06-26)<br>说明书第[0051]-[0069]段 | 1-20 |
| A | CN 110516260 A (腾讯科技深圳有限公司) 2019年 11月 29日 (2019-11-29)<br>全文 | 1-20 |
| A | CN 110993078 A (华中科技大学同济医学院附属协和医院) 2020年 4月 10日 (2020-04-10)<br>全文 | 1-20 |
| A | CN 109659013 A (平安科技深圳有限公司) 2019年 4月 19日 (2019-04-19)<br>全文 | 1-20 |

☑ 其余文件在C栏的续页中列出。

☑ 见同族专利附件。

* 引用文件的具体类型：

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件，或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件（如具体说明的）

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布，与申请不相抵触，但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件，单独考虑该文件，认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件，当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时，要求保护的发明不具有创造性

“&” 同族专利的文件

| 国际检索实际完成的日期 | 国际检索报告邮寄日期 |
|---|---|
| 2021年 3月 18日 | 2021年 3月 30日 |
| ISA/CN的名称和邮寄地址<br>中国国家知识产权局(ISA/CN)<br>中国北京市海淀区蓟门桥西土城路6号 100088<br>传真号 (86-10)62019451 | 受权官员<br>郑宁<br>电话号码 86-(10)-53961313 |

# 国际检索报告

国际申请号
PCT/CN2020/118139

C. 相关文件

| 类 型* | 引用文件，必要时，指明相关段落 | 相关的权利要求 |
|---|---|---|
| A | US 2019317994 A1 (TATA CONSULTANCY SERVICES LIMITED) 2019年 10月 17日 (2019 - 10 - 17)<br>全文 | 1-20 |

# 国际检索报告
## 关于同族专利的信息

国际申请号
PCT/CN2020/118139

| 检索报告引用的专利文件 | | | 公布日<br>(年/月/日) | 同族专利 | | | 公布日<br>(年/月/日) |
|---|---|---|---|---|---|---|---|
| CN | 110069631 | A | 2019年 7月 30日 | 无 | | | |
| CN | 111310471 | A | 2020年 6月 19日 | 无 | | | |
| CN | 110085307 | A | 2019年 8月 2日 | 无 | | | |
| CN | 111339277 | A | 2020年 6月 26日 | 无 | | | |
| CN | 110516260 | A | 2019年 11月 29日 | 无 | | | |
| CN | 110993078 | A | 2020年 4月 10日 | 无 | | | |
| CN | 109659013 | A | 2019年 4月 19日 | 无 | | | |
| US | 2019317994 | A1 | 2019年 10月 17日 | EP | 3557439 | A1 | 2019年 10月 23日 |
| | | | | BR | 102019007765 | A2 | 2019年 11月 5日 |
| | | | | AU | 2019202632 | A1 | 2019年 10月 31日 |
| | | | | CA | 3040373 | A1 | 2019年 10月 16日 |
| | | | | MX | 2019004407 | A | 2019年 11月 25日 |
| | | | | JP | 2019215857 | A | 2019年 12月 19日 |
| | | | | IN | 201821014473 | A | 2019年 10月 18日 |