



## (12)发明专利申请

(10)申请公布号 CN 111180061 A

(43)申请公布日 2020.05.19

(21)申请号 201911251667.8

(22)申请日 2019.12.09

(71)申请人 广东工业大学

地址 510062 广东省广州市大学城外环西路100号

(72)发明人 冷杰武 张虎 阮国磊 叶仕德  
景建波 李龙章

(74)专利代理机构 广东广信君达律师事务所  
44329

代理人 杨晓松

(51)Int.Cl.

G16H 50/20(2018.01)

G16H 10/60(2018.01)

G06N 3/08(2006.01)

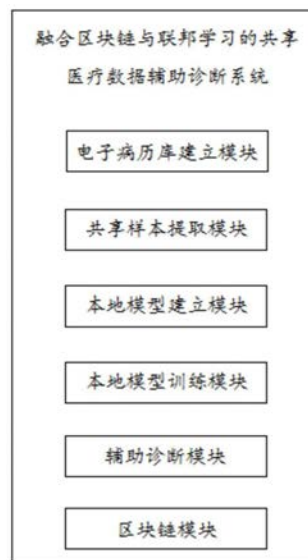
权利要求书1页 说明书5页 附图4页

(54)发明名称

融合区块链与联邦学习的共享医疗数据智能辅助诊断系统

(57)摘要

本申请公开了一种融合区块链与联邦学习的共享医疗数据智能辅助诊断系统,包括:电子病历库建立模块,用于获取医院的医疗数据,并建立电子病例库;共享样本提取模块,用于从参与医疗数据共享的医院的电子病例库中,筛选出病理特征相同样本作为共享样本;区块链模块,用于将本地模型训练模块发送给中央服务器的信息存储在区块链中,并利用区块链的共识机制,对贡献量大的医院进行补偿;本地模型建立模块,用于建立本地模型;本地模型训练模块,用于进行本地模型的加密训练;辅助诊断模块,用于保存训练好的本地模型,以进行疾病的辅助诊断。该系统可以在保证隐私、安全的前提下解决医疗数据的孤岛问题,从而为疾病的诊断提供良好的辅助。



1. 一种融合区块链与联邦学习的共享医疗数据智能辅助诊断系统,其特征在于,包括:

电子病历库建立模块,用于获取医院的医疗数据,并建立电子病例库;所述电子病例库来自于医院的医疗数据,电子病例库包含多个样本,每个样本包括一个患者的病理特征和诊疗结果,其中诊疗结果包括诊断结论和治療措施;本地电子病例库建立好之后,对其中的样本进行特征提取和向量化处理,得到病理特征向量和诊疗结果向量;

共享样本提取模块,用于从参与医疗数据共享的医院的电子病例库中,筛选出病理特征相同样本作为共享样本;

本地模型建立模块,用于建立本地模型,其中,参与医疗数据共享的医院均设置有本地模型建立模块;

本地模型训练模块,用于根据所述共享样本进行本地模型的加密训练;加密训练过程为:将参与医疗数据共享的各个医院和中央服务器组成网络,中央服务器将公钥分发给各个医院的本地训练模块;各个医院的本地模型训练模块根据共享样本训练本地模型,计算训练梯度,然后通过公钥加密的方式共享梯度的更新,并将加密的结果发送到中央服务器;中央服务器根据汇总结果计算总梯度并将其解密,将解密后的总梯度分别传送给每个医院的本地模型训练模块,本地模型训练模块根据总梯度更新本地模型的参数;

辅助诊断模块,辅助诊断模块用于保存训练好的本地模型,训练好的本地模型用于进行疾病的辅助诊断。

2. 根据权利要求1所述的融合区块链与联邦学习的共享医疗数据智能辅助诊断系统,其特征在于,所述的系统还包括:

区块链模块,用于将本地模型训练模块发送给中央服务器的信息存储在区块链中,并利用区块链的共识机制,对贡献量大的医院进行补偿。

3. 根据权利要求1所述的融合区块链与联邦学习的共享医疗数据智能辅助诊断系统,其特征在于,所述病理特征包括**症状名称、症状持续时间、症状出现条件和症状严重程度**;诊断结论是医生针对于患者病情作出的诊断结果,包括疾病名称、并发症;治療措施是医生针对病情做出的各种治療方案,包括药物名称、用药方式、剂量、频率、时间。

4. 根据权利要求1所述的融合区块链与联邦学习的共享医疗数据智能辅助诊断系统,其特征在于,所述共享样本提取模块,采用样本对齐技术,使得不同医院在不公开各自数据的前提下确认双方的共享样本,且不暴露不相互重叠的病历,以便于利用共享样本进行本地模型训练。

5. 根据权利要求1所述的融合区块链与联邦学习的共享医疗数据智能辅助诊断系统,其特征在于,所述本地模型为神经网络模型或深度学习模型,包括BP神经网络模型或VGG深度学习模型。

6. 根据权利要求1所述的融合区块链与联邦学习的共享医疗数据智能辅助诊断系统,其特征在于,在本地模型训练模块根据共享样本训练本地模型时,对于共享样本,将共享样本中的病理特征向量作为本地模型的输入,将诊疗结果向量作为本地模型的输出,对本地模型进行训练。

## 融合区块链与联邦学习的共享医疗数据智能辅助诊断系统

### 技术领域

[0001] 本申请涉及大数据、医疗技术领域，具体涉及一种融合区块链与联邦学习的共享医疗数据智能辅助诊断系统。

### 背景技术

[0002] 海量的数据是人工智能发展的必备要素，而数据的获取与使用需要有两个条件。一是拥有大量的数据来源，二是如何实现高效，安全可信，合理合法的共享。

[0003] 我国虽然在医院病例数方面有很大优势，但由于医疗数据没有共享，存在“数据孤岛”现象，使得这些数据不能合理的利用，从而阻碍了医疗人工智能的发展。但通常又不能简单的将各方的数据进行组合，由于医疗数据的隐私性，相关法律法规的保护性，以及各机构之间的竞争性等等原因，实现医疗数据的简单组合是不现实的。如何在满足数据隐私、安全和监管要求的前提下，设计一个机器学习框架，让人工智能系统能够更加高效、准确的共同使用各自的数据，是当前人工智能发展的一个重要问题。

### 发明内容

[0004] 本申请的目的是提供一种融合区块链与联邦学习的共享医疗数据智能辅助诊断系统，该系统可以在保证隐私、安全的前提下解决医疗数据的孤岛问题，从而为疾病的诊断提供良好的辅助。

[0005] 为了实现上述任务，本申请采用以下技术方案：

[0006] 一种融合区块链与联邦学习的共享医疗数据智能辅助诊断系统，包括：

[0007] 电子病历库建立模块，用于获取医院的医疗数据，并建立电子病例库；所述电子病例库来自于医院的医疗数据，电子病例库包含多个样本，每个样本包括一个患者的病理特征和诊疗结果，其中诊疗结果包括诊断结论和治療措施；本地电子病例库建立好之后，对其中的样本进行特征提取和向量化处理，得到病理特征向量和诊疗结果向量；

[0008] 共享样本提取模块，用于从参与医疗数据共享的医院的电子病例库中，筛选出病理特征相同样本作为共享样本；

[0009] 本地模型建立模块，用于建立本地模型，其中，参与医疗数据共享的医院均设置有本地模型建立模块；

[0010] 本地模型训练模块，用于根据所述共享样本进行本地模型的加密训练；加密训练过程为：将参与医疗数据共享的各个医院和中央服务器组成网络，中央服务器将公钥分发给各个医院的本地训练模块；各个医院的本地模型训练模块根据共享样本训练本地模型，计算训练梯度，然后通过公钥加密的方式共享梯度的更新，并将加密的结果发送到中央服务器；中央服务器根据汇总结果计算总梯度并将其解密，将解密后的总梯度分别传送给每个医院的本地模型训练模块，本地模型训练模块根据总梯度更新本地模型的参数；

[0011] 辅助诊断模块，辅助诊断模块用于保存训练好的本地模型，训练好的本地模型用于进行疾病的辅助诊断。

[0012] 进一步地,所述的系统还包括:

[0013] 区块链模块,用于将本地模型训练模块发送给中央服务器的信息存储在区块链中,并利用区块链的共识机制,对贡献量大的医院进行补偿。

[0014] 进一步地,所述病理特征包括症状名称、症状持续时间、症状出现条件和症状严重程度;诊断结论是医生针对于患者病情作出的诊断结果,包括疾病名称、并发症;治疗措施是医生针对病情做出的各种治疗方案,包括药物名称、用药方式、剂量、频率、时间。

[0015] 进一步地,所述共享样本提取模块,采用样本对齐技术,使得不同医院在不公开各自数据的前提下确认双方的共享样本,且不暴露不相互重叠的病历,以便于利用共享样本进行本地模型训练。

[0016] 进一步地,所述本地模型为神经网络模型或深度学习模型,包括BP神经网络模型或VGG深度学习模型。

[0017] 进一步地,在本地模型训练模块根据共享样本训练本地模型时,对于共享样本,将共享样本中的病理特征向量作为本地模型的输入,将诊疗结果向量作为本地模型的输出,对本地模型进行训练。

[0018] 本申请具有以下技术特点:

[0019] 1.为了解决各医院间医疗数据的孤岛问题,本申请利用联邦学习,通过加密机制下的参数交换方式,使各医院的医疗数据都保留在本地,在不泄露隐私也不违反法规的前提下,联合各方的医疗数据建立一个共同的神经网络模型,该神经网络模型具有帮助医生进行辅助诊断的功能,即通过输入病人的病理特征,能够通过算法在病例库中匹配出最为相似的病理,并按照相关病例的记录,给出推荐的治疗方案。

[0020] 2.本申请通过使用区块链来代替该中央服务器,将各医院本地训练模型的更新永远记录在区块链上,减少因中央服务器的故障而对整个联盟系统的损坏,并利用区块链的共识机制对数据贡献量多的医院进行奖励,以激发各医院的积极性。

## 附图说明

[0021] 图1为本申请系统的结构示意图;

[0022] 图2为本地模型训练过程中的交互示意图;

[0023] 图3为根据共享样本训练本地模型的示意图;

[0024] 图4为本地模型进行辅助诊断时的流程示意图;

[0025] 图5为融合区块链的本地模型交互示意图。

## 具体实施方式

[0026] 本申请中为了各方的医疗数据能够有效的联合,各医院方需要建立一个相似的病例库数据以供本地模型训练学习。本地模型采用相识度匹配的算法,即通过比较目标病例与数据库中的相识程度来判断两病例的匹配程度。当各方的本地模型训练完毕后,通过加密机制下的参数交换方式,使各方的数据都保留在本地,在不泄露隐私也不违反法规的前提下,建立一个全局联合模型,以解决数据孤岛问题。同时进一步的,由于联邦学习模型依赖于一个单一的中央服务器,若该服务器发生故障,则会对所有本地的训练模型造成影响,因此本申请通过使用区块链来代替该中央服务器。基于此,本申请的融合区块链与联邦学

习的共享医疗数据智能辅助诊断系统,如图1所示,包括:

[0027] 1.电子病历库建立模块

[0028] 电子病历库建立模块用于获取医院的医疗数据,并建立电子病例库。

[0029] 所述电子病例库来自于医院的医疗数据,电子病例库包含多个样本,每个样本包括一个患者的病理特征和诊疗结果,其中诊疗结果包括诊断结论和治療措施。所述病理特征包括症状名称、症状持续时间、症状出现条件和症状严重程度等;诊断结论是医生针对于患者病情作出的诊断结果,包括疾病名称、并发症等;治療措施是医生针对病情做出的各种治療方案,包括药物名称、用药方式、剂量、频率、时间等。

[0030] 本地电子病例库建立好之后,对其中的样本进行特征提取和向量化处理,得到病理特征向量V1和诊疗结果向量V2。

[0031] 2.共享样本提取模块

[0032] 共享样本提取模块用于从参与医疗数据共享的医院的电子病例库中,筛选出病理特征相同样本作为共享样本。

[0033] 由于病例的特殊性,每家医院的病例库中,拥有相同病理特征的数量较多,但病人的重合度却较小,故采用横向联邦学习的结构,将医疗数据按照病人维度进行切分,取出不同医院中病理特征相同而病人不完全相同的那部分数据进行训练。

[0034] 采用样本对齐技术,使得不同医院在不公开各自数据的前提下确认双方的共享样本,且不暴露不相互重叠的病历,以便于利用共享样本进行本地模型训练。

[0035] 3.本地模型建立模块

[0036] 本地模型建立模块用于建立本地模型,其中,参与医疗数据共享的医院均设置有本地模型建立模块。所述本地模型为神经网络模型或深度学习模型,例如BP神经网络模型或VGG深度学习模型。

[0037] 4.本地模型训练模块

[0038] 如图2所示,本地模型训练模块用于根据所述共享样本进行本地模型的加密训练;同样地,在参与医疗数据共享的医院均设置有本地模型训练模块;具体的加密训练过程为:将参与医疗数据共享的各个医院和中央服务器组成网络,中央服务器将公钥分发给各个医院的本地训练模块;各个医院的本地模型训练模块根据共享样本训练本地模型,计算训练梯度,然后通过公钥加密的方式共享梯度的更新,并将加密的结果发送到中央服务器;中央服务器根据汇总结果计算总梯度并将其解密,将解密后的总梯度分别传送给每个医院的本地模型训练模块,本地模型训练模块根据总梯度更新本地模型的参数。

[0039] 保存训练好的本地模型用于疾病的辅助诊断。

[0040] 在本地模型训练模块根据共享样本训练本地模型时,对于共享样本,将共享样本中的病理特征向量V1作为本地模型的输入,将诊疗结果向量V2作为本地模型的输出,对本地模型进行训练,如图3所示。

[0041] 5.辅助诊断模块

[0042] 辅助诊断模块用于保存训练好的本地模型,其中训练好的本地模型用于进行疾病的辅助诊断。

[0043] 在各个医院的本地模型均训练完成后,各个医院可以利用本地模型进行辅助诊断,如图4所示,步骤如下:

[0044] 步骤1,病人在诊断完成后,根据检测数据,提取相应的病理特征,包括症状名称、症状持续时间、症状出现条件和症状严重程度。其中症状严重程度中可以是相关的检查数据,例如血常规,体温、咳嗽等对疾病诊断起主要作用的症状数据。将病理特征进行向量化处理后,作为输入向量,输入到已经训练好的本地模型中。

[0045] 步骤2,在已训练好的本地模型中,根据算法匹配可能的相似病历,并输出对应的诊疗结果。

[0046] 算法的大致步骤如下。将病例分为病例库中的原病例*i*和目标病例*j*,其中原病例*i*共有*n*个病症。 $\text{Sim}(i, j)$ 表示病例库中病例*i*和新病例*j*间的综合相似度。 $w_p$ 表示第*p*个病症在整个病例中*i*的权重, $\text{Sim}(i_p, j_p)$ 表示两个病例*i*、*j*在属性*p*上的相似度,因此分母就表示目标病例中所有属性对应的权重和,分子表示的是目标病例中所有属性与新病例属性相似度与对应权重积的和,公式如下:

$$[0047] \quad \text{Sim}(i, j) = \frac{\sum_{p=1}^n [w_p * \text{Sim}(i_p, j_p)]}{\sum_{p=1}^n w_p}$$

[0048] 当两病例的相识度 $\text{Sim}(i, j)$ 高于某一阈值时,则说明该病人所患病症很有可能与病例库中匹配的病症相似,并且可以根据匹配的病症给病人提供推荐诊断方案,以帮助医生进行最终诊断。

[0049] 步骤3,医生利用专业知识和经验,集合病人的实际情况,判断诊疗结果是否可行;若方案不可行,则对诊疗方案进行适当调整。若调整后的结果应用临床后结果合理,则将此病历作为新的样本存储到本地电子病例库中,以便于后续的本地模型训练。

[0050] 在上述技术方案的基础上,本系统还可以包括:

#### [0051] 6. 区块链模块

[0052] 区块链模块用于将本地模型训练模块发送给中央服务器的信息存储在区块链中,并利用区块链的共识机制,对贡献量大的医院进行补偿。

[0053] 由于上述建立的联邦学习模型依赖于一个单一的中央服务器,该服务器聚合所有本地模型的更新,并进行一系列的计算,从而产生一个全局模型的更新,并将更新传递给本地模型,再进行本地模型更新。若该服务器发生故障,则会对所有本地的训练模型造成影响。同时,由于提供数据量多的医院对全局训练模型的贡献更大,若不对这些机构进行有效的补偿,则拥有数据量大的医院就不太愿意与那些拥有数据量小的医院进行联合,最终造成整个联盟系统的崩塌。因此,本方案中引入了区块链,将本地训练模型的更新永远记录在区块链上,减少因中央服务器的故障而对整个联盟系统的损坏。

[0054] 其逻辑结构由医院以及维护区块链的矿工组成,如图5所示,医院与上述联邦学习模型一样,利用本地数据进行训练并提供本地模型的更新;矿工与某一医院相对应,将各医院本地模型的更新进行打包并写入区块链进行保存,并与医院共同维护区块链的一致性。在一段时间后,由于随着各医院病例的增多,其本地训练的模型也会发生改变,进而会影响整个联邦学习的全局模型。区块链在这样的迭代过程中不断储存各医院的更新记录,实现了去中心化,解决了传统联邦学习依赖单一服务器的缺点。同时,在每次更新中,对贡献量大的机构进行补偿,对维护和激励整个联邦学习模型有着重要作用。在某一次更新中,其流程如下所示:

[0055] 6.1本地模型的更新与上传:每个医院计算并上传本地模型的更新到它在区块链

中相应的矿工上。

[0056] 6.2交叉验证,各个矿工交换并验证各医院上传的模型更新,然后将通过验证的更新写入候选区块上中。

[0057] 6.3生成区块,各矿工通过一定的共识机制获得打包区块权,获得打包权后将候选块写入区块链。

[0058] 6.4广播区块,将生成的区块广播给其他的矿工与节点。

[0059] 6.5全局模型的下载与更新,每个医院通过区块链下载相应的更新,并在本地进行全局模型的更新。

[0060] 通过引入区块链,可以避免由于中央服务器的故障而使整个系统的瘫痪。同时,由于全局模型的更新是在各个医院的本地进行的,单个设备或矿工的问题,不会影响其他节点的更新。在交叉验证阶段,根据各医院提供的数据量多少以及矿工的验证打包过程提供奖励来提高机构各成员的积极性。

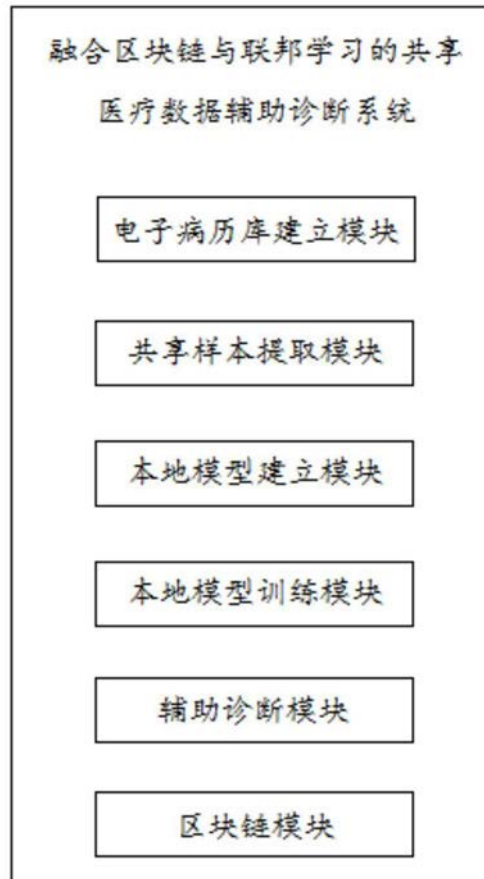


图1



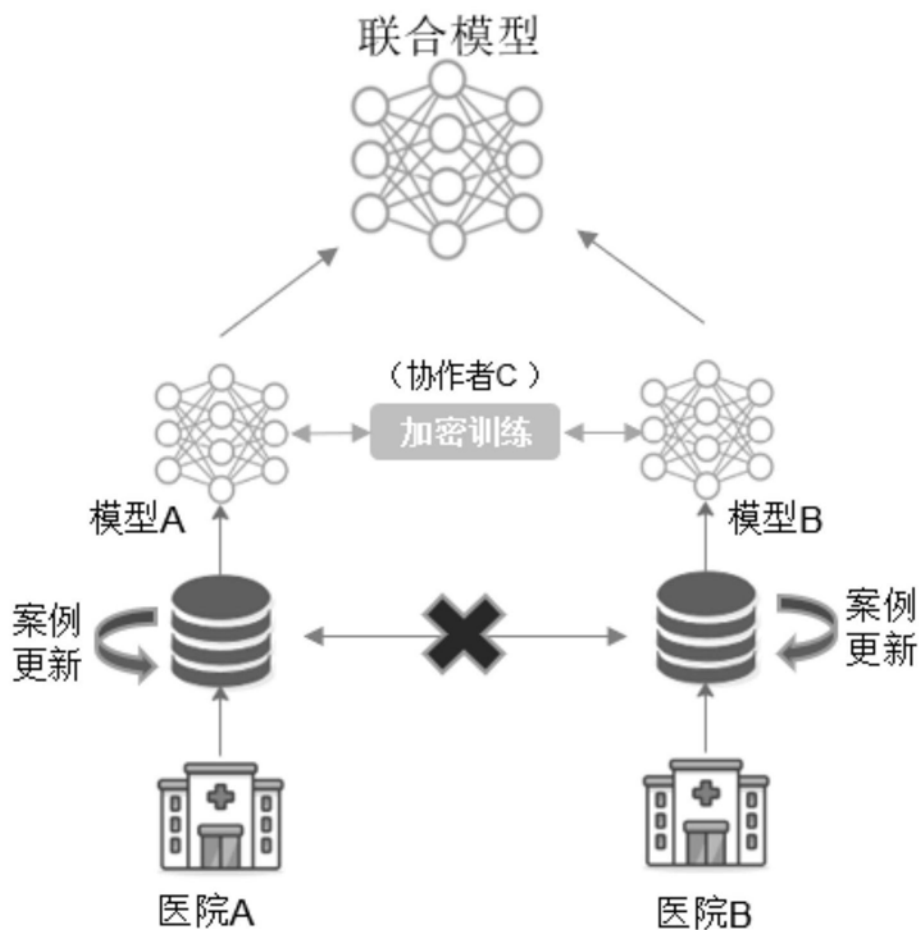


图2

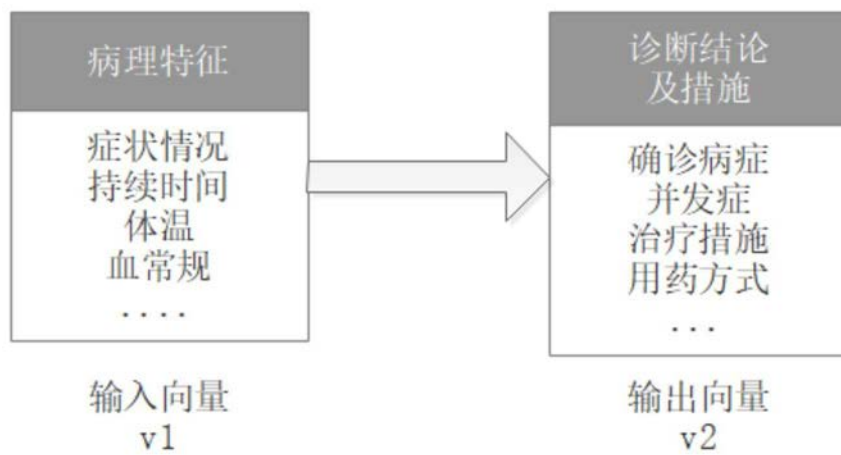


图3

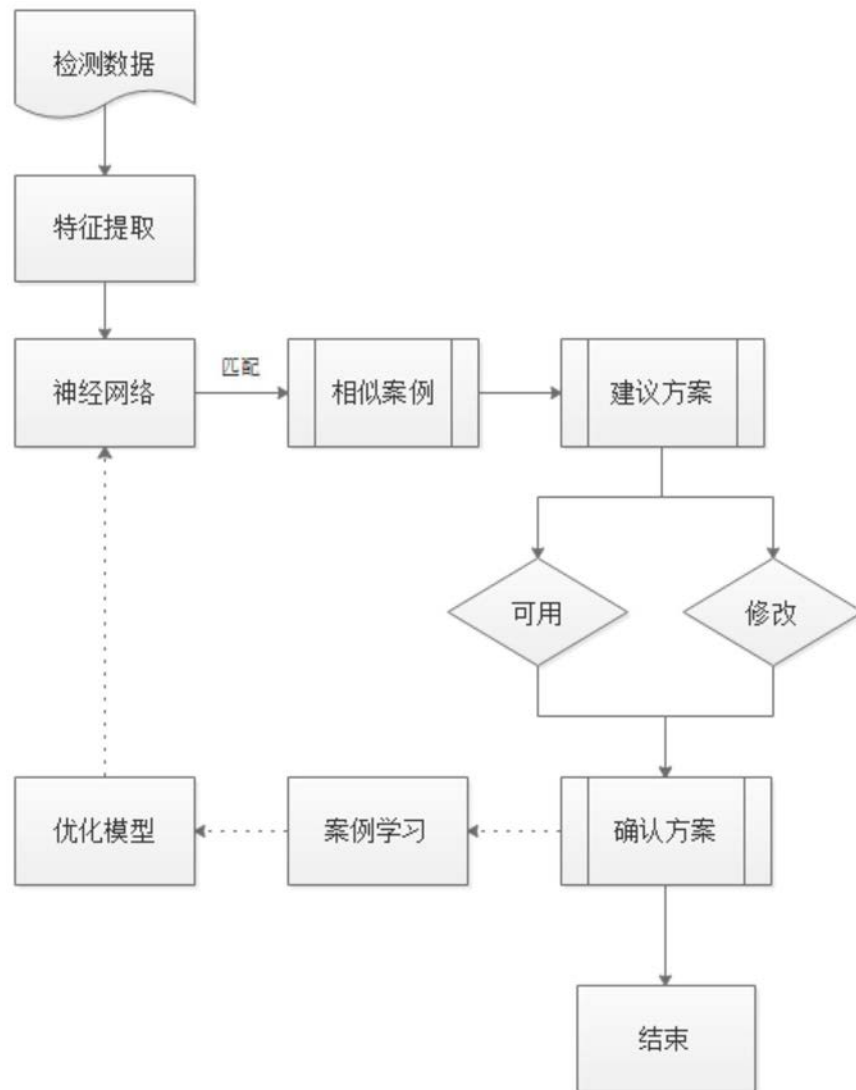


图4

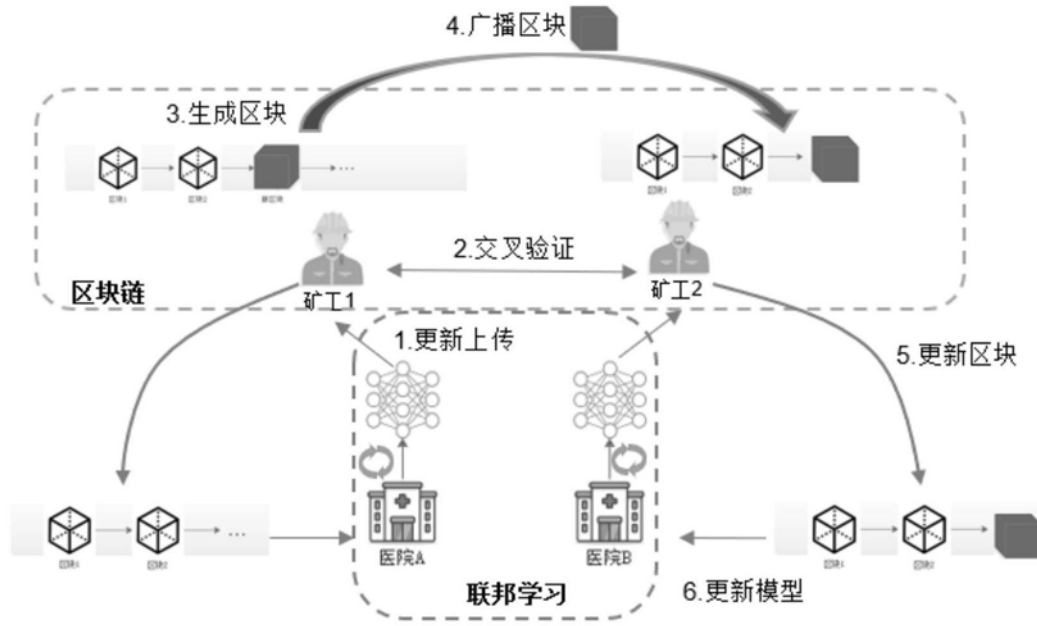


图5