基于区块链的医疗影像数据人工智能检测模型*

陈思源[1,2],谭艾迪[3],魏双剑[3],盖珂珂[2,4]
(1.北京理工大学 计算机学院,北京 100081;2.北京理工大学长三角研究院(嘉兴),浙江 嘉兴 314019;
3.中国船舶工业综合技术经济研究院,北京 100081;4.北京理工大学 网络空间安全学院,北京 100081)

摘 要:基于深度学习的目标检测技术被广泛应用于医疗检测领域,该技术依赖大量医疗影像训练分类模型,从而为医生决策提供有力的辅助医疗手段。因涉及患者隐私并直接关系到医生诊断,所以医疗影像数据的共享必须保护患者隐私并确保数据准确不被篡改,而现有中心化的医疗数据存储方案面临隐私泄露等诸多安全问题。提出了一种基于区块链的医疗影像数据人工智能检测模型。该模型针对目标检测技术辅助医生诊断的问题,采用区块链技术实现去中心化、不可篡改的训练参数聚合,通过加密和签名技术保护数据隐私,利用智能合约评估服务器诊断准确率,有助于解决医疗数据壁垒和医疗隐私泄露问题。

关键词:深度学习;区块链;安全数据共享;人工智能检测
中图分类号:TP311 **文献标识码**:A DOI:10.19358/j.issn.2097-1788.2022.04.003
引用格式:陈思源,谭艾迪,魏双剑,等.基于区块链的医疗影像数据人工智能检测模型[J].网络安全与数据治理,2022,41(4):21-25.

Blockchain-based artificial intelligence detection model for medical data

Chen Siyuan[1,2], Tan Aidi[3], Wei Shuangjian[3], Gai Keke[2,4]
(1.School of Computer Science, Beijing Institute of Technology, Beijing 100081, China;
2.Yangtze Delta Region Academy of Beijing Institute of Technology, Jiaxing 314019, China;
3.China Institute of Marine Technology and Economy, Beijing 100081, China;
4.School of Cyberspace Science and Technology, Beijing Institute of Technology, Beijing 100081, China)

Abstract: Deep learning-based target detection technology is being widely used in the field of medical detections. For training a large number of medical images, we can construct an effective classification model to effectively predict the disease situation of patients and provide a powerful auxiliary medical means of decision-making. In order to improve the prediction accuracy, massive training data are the premise to construct an effective learning model. However, medical data involve patients' privacy and are directly related to diagnoses. Sharing medical data needs to guarantee privacy, accuracy and tamper-proof. Existing centralized medical storage schemes face many security issues, e.g., privacy disclosure. This paper proposes a blockchain-based artificial intelligence detection model for medical data that uses a target detection technology to assist physicians during the diagnosis process. In our model, blockchain technology supports realizing the decentralized and un-tampered aggregation of training parameters. Encryption and signature technology are used to protect privacy and smart Contract is implemented to evaluate the accuracy of server diagnosis. The proposed model will contribute to solving the issues in medical data barriers and privacy disclosure.

Key words: deep learning; blockchain; secure data sharing; artificial intelligence detection

0 引言

医院每天产生和诊断大量的医疗影像,据统计在医疗数据中,影像数据占数据总量的90%以上。随着医疗检测设备的更新换代和不断增加,影像数据以每年超过30%的增长速度急剧增加。与此形成鲜明对比的是,医生数量缓慢增长,这使得影像诊断如阅读分析CT(计算机断层扫描)等工作对医生造成的负担日益加剧,经验缺乏与工作量增大容易

*基金项目:国防基础科研计划(JCKY2020206C058)

造成误诊。随着大数据和人工智能技术的发展,利用计算机辅助诊断,使用基于人工智能的目标检测技术帮助医生做出快速判断,对减轻医生负担、增加诊断准确率、提高就诊效率而言就显得十分必要且具有现实意义。

目标检测技术因其广泛的现实应运用场景备受学术界和工业界关注。随着计算机算力的不断提升,目标检测技术蓬勃发展,衍化出双阶段和单阶段两大类。

Faster R-CNN[1]、R-FCN[2]和 FPN[3]都属于双阶段,该类算法在基于特征提取的基础上,在目标对象上生成候选区域,再对其区域进行分类和回归,用矩形框确定准确位置信息并判定类别。双阶段类算法中,R-CNN[4]是最早的深度学习目标检测算法,显著提升了目标检测平均精准度(mean Average Precison,mAP);He[5]通过引入空间金字塔池化层(Spatial Pyramid Pooling,SPP)提出SPP-NET,检测效率较R-CNN大幅提高;Fast R-CNN[9]改进了R-CNN的池化核,将候选区域映射成统一尺度的特征向量,提高了模型训练效率。一般的,双阶段类算法具有较高的精度。

YOLO[6],SSD[7]和 RetinaNet[8]等属于单阶段目标检测算法。该类算法一次性实现候选区域的生成、分类和回归。YOLO网络结构简单,具有速度快、实时性较好的特点,是单阶段类算法的代表。SSD通过对不同卷积层预设锚盒提取特征,优化了对多尺度目标的检测。YOLOv3延续了之前版本Darknet结构的使用,同时引入残差结构,提高了小尺度目标的检测精度。总的看来单阶段类算法具有较高的效率。

然而,目标检测确保预测精准的必要前提是需要大量准确数据。训练样本过小,会导致学习效果不佳。错误的数据更是可能产生误导性的结果。为了获得大量数据,在医院、研究中心等机构间共享医疗数据是简单高效的解决方案,但是中心化的医疗数据存储不利于数据分享,容易催生倒卖数据的黑色产业,更不利于对患者隐私的保护。区块链是一种分布式的数据账本,可以在零信任基础的用户之间实现安全的数据共享。将区块链用于医疗数据存储和共享,这意味着用户可以得到更多的训练数据,产生更好的训练模型,得到更准确的诊断预测结果,并有助于数据集的丰富和更新。除此之外,区块链有助于帮助AI解释自己,使AI的"黑匣子"存在被人类理解的可能。区块链的数据公布于链上且不可篡改,拥有清晰的数据审计基础,不仅可以提高数据和模型的可信度,还可以提供一条清晰的路径来追溯计算机决策的过程,消除人们对AI诊断的顾虑。

本文提出了一种新的医疗影像数据人工智能检测模型,该模型使用区块链网络,利用智能合约实现医疗影像数据共享,通过密码学技术保证数据安全和隐私保护。

1 基于深度学习的目标检测

基于深度学习的目标检测算法,其性能主要取决于网络结构,常见的网络结构结构有卷积神经网络和残差网络结构等。

1.1 卷积神经网络

卷积神经网络结构一般分为下列层级:

(1)输入层

输入一般为图片,经过输入层的运算处理,可转化为三维像素矩阵,三个维度分别代表图像的长宽和色彩的通道数。黑白图片通道数为1,彩色图片通道数为3。经过输入层后,三维矩阵再通过不同网络结构进行变换,直到最后的全连接层为止。

(2)卷积层

卷积层用于提取特征。输入矩阵通过卷积核得到输出矩阵。卷积核也是一个三维矩阵,它的长和宽是通常为3×3或5×5。因为卷积操作两个矩阵的深度必须是一致的,所以滤波器的深度不可以改变,必须和输入层矩阵的深度一致。卷积核还需要人工设置个数,它的个数决定了输出层矩阵深度的大小。

(3)池化层

池化层用于压缩数据尺寸大小,减少参数个数,在加快网络计算速度的同时又能够防止过拟合问题的出现。

池化层通过使用过滤器在矩阵上进行滑动,根据过滤器计算方式为取最大值还是平均值,分为平均池化和最大池化。平均池化与目标区域所有值都有关,所以对图像背景信息的保留更为完整,而最大值池化并非与所有值有关,因此突出保留边缘纹理信息,这是识别目标的重要特征,所以图像识别大多使用最大值池化法。

(4)全连接层

全连接层的作用类似于分类器,用于实现特征到样本空间的映射,可以通过采用尺寸1×1的卷积核的卷积层替代。

(5)softmax 层

如图 1 所示,softmax 层往往作为深度神经网络的最后一层,作为多分类的输出,输出值为不同分类的概率。最后一次全连接神经元的个数 n 即为需要分类的种数。之后将所有的经过 softmax 运算得到最终的输出,softmax 运算公式为:

$$y_i=\text{softmax}(z_i)=\frac{e^i}{\sum_i e^i} \tag{1}$$

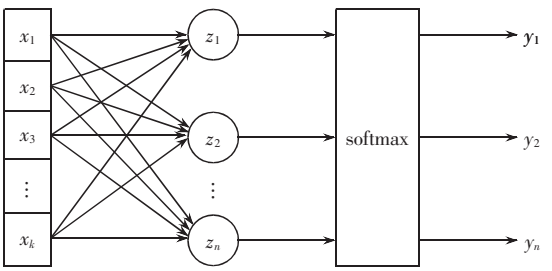

图 1 softmax 层结构

softmax 运算相当于将输入值映射为 0~1 之间的实数,并且所有的输出值和为 1,正好可以把输出值看作为判断为该类别的概率,所以它可以作为分类问题中卷积神经网络的输出。

1.2 残差网络结构

随着深层卷积网络深度不断增加,出现了准确率反而降低的网络的退化现象,其本质就是因为信息丢失而出现了过拟合问题。在网络层数较多的情况,经卷积处理的图片会使得原本差别很大的图像变得相似,那么这种情况下的分类结果过会出现误差,问题的解决办法就是提高泛化性和差异性,由此产生了残差网络,其结构如图 2 所示。残差结构通过断开式设计,将之前层级中的数据直接输入到后面数据层中,设 x 为输入,拟合函数映射设为 $H(x)$,定义残差映射 $F(x)=H(x)-x$,则原始的映射函数 $H(x)$ 可表示为 $F(x)+x$。

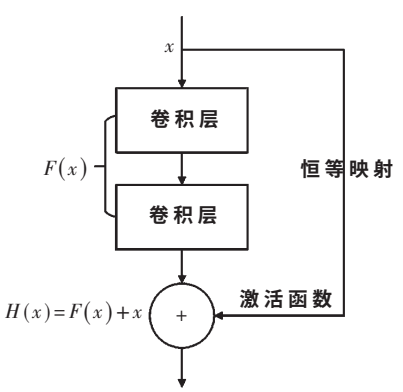

图 2 残差网络结构

2 区块链

区块链是一种变革式的数据存储技术,又称分布式账本,是比特币的底层技术,利用数字签名、散列函数等密码学技术,结合 P2P 网络实现的去信任、无法篡改的分布式数据库[10]。数据存储在称为区块的结构中,这些区块通过散列以链形式相互连接(通常每个块还包括时间戳和通过其散列指向前一个块的链接)。这些块具有报头和内容。由于区块之间按一定规则顺序相连,随着参与者和区块数量的增加,在没有网络共识的情况下,要修改任何信息都是极其困难的。

从区块链技术实现的层面来看,区块链并不是计算机领域某一单一技术的改良或革新,而是多种跨领域成熟技术的完美结合。当这些技术不再以独立的身份孤立存在,而是以新的方式相互融合,创造了一种全新的数据承载模式[11]。

区块链迅速为人们所熟知,并广泛应用于大数据、物联网[12-16]等多个领域,得益于其本身具有的多种出众特性,其中去中心、透明性和可溯源性、不可篡改性是其最为突出的优势:

(1)去中心化

相较于传统的中心化服务器存储,区块链避免了使用单一节点存储、更新数据的弊端,取而代之的是由所有参与方共同维护网络数据。这种方式不仅提高了系统的安全性,同时使网络数据不会因为某一寡头存在而使整个网络被更改和控制。

(2)透明性和可溯源性

区块链面向所有接入节点开放,以比特币为例,所有接入节点均可以查询区块链中的任意交易,每一节点都有权保存全部交易账本,因此是面对交易透明的。于此同时,区块之间以链式结构首尾相连,按上链时间先后顺序记录数据,因此实现了交易可溯源[17]。

(3)不可篡改性

区块链上的数据上链前需运行共识算法,达成共识后方可上链。共识算法种类很多,以比特币为例,其采用工作量证明(POW)共识算法,链上交易需经过 6 个区块后才能被确认为交易成功,这使得算力在 50%以下的攻击者想要篡改链上数据几乎不可能实现[18],由此保证了不可篡改性。

3 模型设计

模型结构如图 3 所示,目的是要充分利用各医

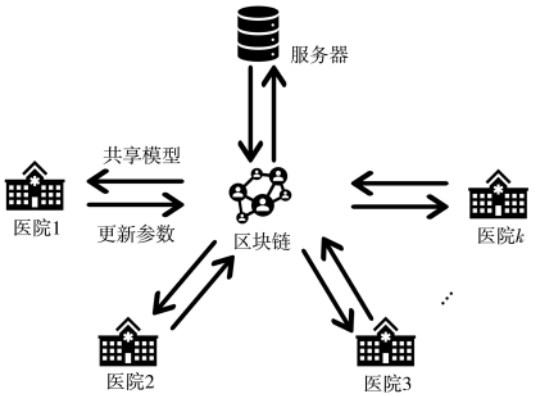

图 3 模型结构

院的医疗影像数据,在众多机构参与的情况下形成海量的数据集,从而训练得到更好的全局模型。对于某一训练任务,服务器对任务进行细分,明确学习的对象、相关设备、应用的类型(如 CT 图)、训练集的格式、训练模型(如卷积神经网络)、训练要求(如学习率)等。为了保证各方原始数据不泄露,对 k 家医院,它们各自拥有 n_k 个数据点,设本地模型权重为 w_t,学习率为 η,医院在本地训练其模型,并将更新参数记录在区块链上,再由服务器聚合得到全局模型。

参数更新公式如下:

$$g_k=\nabla F_k(w_t) \tag{2}$$

$$F_k(w_t)=\frac{1}{n_k}\sum_{i\in P_k} f_i(w_t) \tag{3}$$

本地权重更新:

$$w_{t+1}^{k}=w_t-\eta g_k \tag{4}$$

全局权重更新:

$$w_{t+1}=\sum_{k=1}^{K}\frac{n_k}{n}w_{t+1}^{k} \tag{5}$$

各医院节点充当区块链矿工,以工作量证明的共识机制(PoW)将更新参数记录到区块链中,以此基于贡献的原则激励数据共享。

为了保证数据安全,采用加密数据上链。各医院拥有自身密钥$\{PK_j, SK_j\}\{j=1,2,3,\cdots,K\}$。服务器提出任务 tx_i 时,生成临时密钥串$\{\{PK_1^i, SK_1^i\}, \{PK_2^i, SK_2^i\}\cdots\{PK_K^i, SK_K^i\}\}$,上标代表任务批次,下标代表医院编号,并使用对应医院的密钥加密后广播,因此各医院只能得到自己的临时密钥。各医院使用自身临时密钥对数据加密并做哈希后再将数据上链。在区块链中,每个区块对应某一学习任务 tx_i,对每家医院而言,只能通过解密方式查看自己的数据,其他链上数据呈现加密乱码,由此保护各医院数据的独立安全。

除此之外,考虑两种现实需求:(1)医院想借助服务器诊断 CT 图,除服务器给出的结果外,医院还需判断该结果的准确率;(2) 医院 A 需要医院 B 的链上数据以优化自己的本地模型。

对于(1),过程如图 4 所示,设医院 H 需服务器对 CT 图 x_0 进行诊断,得到病因 y_0。为了不暴露患者隐私,医院 H 不想服务器知道需要诊断的具体为哪张 CT 图,为此引入图 $x_1, x_2, \cdots, x_t$ 作为混淆输入。$x_1, x_2, \cdots, x_t$ 为已被正确诊断的 CT 图,$x_1, x_2, \cdots, x_t$ 对应诊断结果 $y_1, y_2, \cdots, y_t$。$\{x_0, \{x_1, y_1\}, \{x_2, y_2\}, \cdots, \{x_t, y_t\}\}$被 H 提交至智能合约,智能合约再将$\{x_0, x_1, x_2, \cdots, x_t\}$发送服务器。由于没有收到 $y_1, y_2, \cdots, y_t$,服务器无法识别真实需要判断的 CT 图 x_0,无差别地对所有输入进行判断,将所有判断结果$\{y_0, y_1, \cdots, y_t\}$返回智能合约。此时,设$y_i=y_i^{'}\{i=1,\cdots,t\}$的个数为 m,则 $y_i\neq y_i\{i=1,\cdots,t\}$的个数为 $t-m$,可认为预测准确率为 m/t。设智能合约判断阈值为 M,当 $m/t\geq M$ 时,智能合约向医院 H 返回 x_0,否则返回 false。

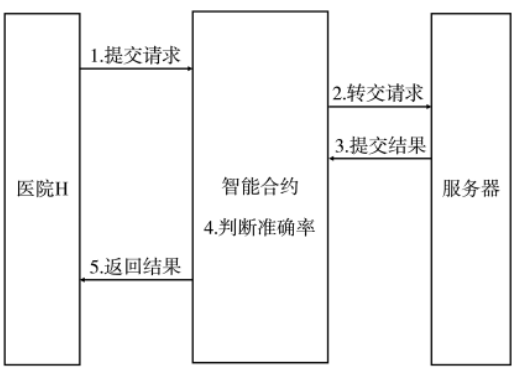

图 4 申请服务器诊断流程

对于(2),过程如图 5 所示,设医院 A 向智能合约发送请求 $SK_A(request)$,希望得到区块 tx_i 上医院 B 的数据 $data_B$,智能合约验证医院 A 的签名后,转发请求至医院 B。医院 B 将链上对应数据解密,使用 PK_A^i 加密 $data_B$,并用 SK_B 签名后发送给智能合约。随后智能合约验证签名并转发医院 A,医院 A 使用自己的临时密钥 SK_A 解密得到 $data_B^{'}$。医院 A 对解密后的数据用医院 B 的临时密钥 PK_B^i 加密并做哈希运算,查看结果 $h(PK_B^i(data_B^{'}))$与链上数据是

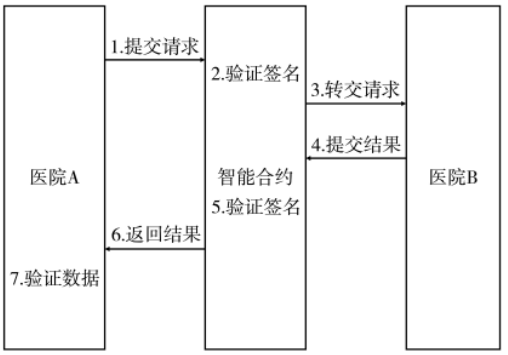

图 5 申请数据流程

否一致，从而可以判断医院 B 是否交付真实数据。

4 结论

本文提出了一种基于区块链的医疗影像数据人工智能检测模型。该模型基于区块链网络结合加密和签名技术，采用各方本地训练模型、在链上聚合更新权重的方式实现数据安全共享并避免了原始数据泄露。本文还细化提出了医院共享数据、判断服务器诊断准确率的方法。

参考文献

[1] REN S Q, He K M, GIRSHICK R B, et al. Faster R-CNN: towards real-time object detection with region proposal networks[C]//IEEE Trans. Pattern Anal. Mach. Intell., 2017(39): 1137-1149.

[2] DAI J F, LI Y, HE K M, et al. R-FCN: object detection via region-based fully convolutional networks[C]// Proceedings of Conference on Advances in Neural Information Processing Systems, Barcelona, 2016: 379-387.

[3] LIN T, DOLLAR P, GIRSHICK R B, et al. Feature pyramid networks for object detection[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, 2017: 936-944.

[4] AGRAWAL P, GIR SHICK R, MALIK J. Analyzing the performance of multilayer neural networks for object recognition[C]//European Conference on Computer Vision. Switzerland: ECCV, 2014: 329-344.

[5] HE K, ZHANG X, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 37(9): 1904-16.

[6] REDMON J, FARHADI A. YOLOv3: an incremental improvement[J]. ArXiv preprint arXiv: 1804.02767, 2018.

[7] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]//Proceedings of European Conference on Computer Vision, Amsterdam, 2016: 21-37.

[8] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[J]. IEEE Trans. Pattern Anal. Mach. Intell., 2020, 42: 318-327.

[9] Niu Zuodong, Qin Tao, Li Handong, et al. Improved algorithm of RetinaFace for natural scene mask wear detection[J]. Computer Engineering and Applications, 2020, 56(12): 1-7.

[10] 金鑫. 基于区块链技术的网络信息安全研究[J]. 信息系统工程, 2018(12): 79-82.

[11] 袁勇, 王飞跃. 区块链技术发展现状与展望[J]. 自动化学报, 2016, 42(4): 481-494.

[12] GAI K, QIU M, ZHAO H. Privacy-preserving data encryption strategy for Big Data in mobile cloud computing[J]. IEEE Transactions on Big Data, 2017: 1-1.

[13] GAI K, WU Y, ZHU L, et al. Blockchain-enabled trustworthy group communications in UAV networks[J]. IEEE Transactions on Intelligent Transportation Systems, 2020(99): 1-13.

[14] GAI K, GUO J, ZJU L, et al. Blockchain meets cloud computing: a survey[J]. IEEE Communications Surveys & Tutorials, 2020(99): 1-1.

[15] GAI K, WU Y, ZHU L, et al. Differential privacy-based blockchain for industrial Internet-of-Things[J]. IEEE Transactions on Industrial Informatics, 2020, 16(6): 4156-4165.

[16] GAI K, WU Y, ZHU L, et al. Permissioned blockchain and edge computing empowered privacy-preserving smart grid networks[J]. IEEE Internet of Things Journal, 2019: 7992-8004.

[17] 邵奇峰, 金澈清, 张召, 等. 区块链技术: 架构及进展[J]. 计算机学报, 2018, 41(5): 969-988.

[18] 汪垚. 区块链技术在互联网安全中的应用探究[J]. 企业科技与发展, 2019(2): 133-134.

(收稿日期: 2021-08-21)

作者简介:

陈思源(1994-), 男, 硕士, 主要研究方向: 区块链隐私保护。

谭艾迪(1991-), 男, 博士, 主要研究方向: 大数据资源建设与管理应用。

盖珂珂(1982-), 通信作者, 男, 博士, 主要研究方向: 网络空间安全、区块链、云计算/边缘计算安全。E-mail: gaikeke@bit.edu.cn。

版权声明

凡《网络安全与数据治理》录用的文章，如作者没有关于汇编权、翻译权、印刷权及电子版的复制权、信息网络传播权与发行权等版权的特殊声明，即视作该文章署名作者同意将该文章的汇编权、翻译权、印刷权及电子版的复制权、信息网络传播权与发行权授予本刊，本刊有权授权本刊合作数据库、合作媒体等合作伙伴使用。同时，本刊支付的稿酬已包含上述使用的费用，特此声明。

《网络安全与数据治理》编辑部