***附件：***

**创意说明书**

|  |  |  |  |
| --- | --- | --- | --- |
| 创意名称 | 医疗场景下基于区块链的信息共享和多方计算 | | |
| 创意发明人 |  | 学校 |  |
| 联系方式 |  | | |
| 创意亮点 | 1. **一句话解释你的创意：**   医疗场景下基于区块链的信息共享和多方计算   1. **一幅图解释你的创意：**   图1. 算法框架 | | |
| 与本创意最相近似的现有技术方案 | 图2.一种基于区块链的安全多方计算模型框架  互联⽹上每天都在产⽣⼤量的数据，数据⼤多被⼀些公司拥有，数据逐渐成为公司的重要资产，⽽数据的共享和流通能产⽣更⼤价值，但是出于对⽤户隐私的保护，各个公司之间不能直接进⾏数据交互使⽤。所以数据的隐私保护和共享使⽤之间的⽭盾越加明显。随着这⼏年⼈们对隐私的关注度越来越⾼，隐私保护技术蓬勃发展传统的隐私保护技术有差分隐私、匿名算法、同态加密等，现在的隐私保护主要技术是多⽅安全计算（MPC）和基于硬件的可信执⾏环境（TEE）。区块链的去中心化的性质与MPC契合，并且还具有公开不可篡改的功能，所以将区块链和MPC技术结合，链上做计算存证，链下做多方安全计算的模型就产生了。图2就是一种基于区块链的多方安全计算方案。该方案是基于联盟链的环境下，MPC是采用的同态加密，用户将数据加密后将密文上传到区块链，计算节点在密文上直接进行计算，计算后的结果也被存储在区块链上，只有拥有私钥的用户能解密。最早在2014年，文献[2]中，实现了将比特币系统和MPC结合；文献[3]，一个基于区块链的可验证公平的多方安全计算。文献[4]，在参与方都是foresighted的情况下，引入了荣誉系统(reputation system)作为激励，这项工作重点在保证参与方的公平性。文献[5]提出一种通用的基于区块链的多方安全计算，链下采用的是基于Gennaro方案的不公平通用MPC协议，在秘密重建阶段和智能合约进行多轮交互，实现公平性。  [1] Yuhan Yang, Lijun Wei, Jing Wu, and Chengnian Long. 2020. Block-SMPC: A Blockchain-based Secure Multi-party Computation for Privacy-Protected Data Sharing. In Proceedings of the 2020 The 2nd International Conference on Blockchain Technology (ICBCT'20). Association for Computing Machinery, New York, NY, USA, 46–51.  [2] M. Andrychowicz, S. Dziembowski, D. Malinowski and L. Mazurek, "Secure Multiparty Computations on Bitcoin," 2014 IEEE Symposium on Security and Privacy, San Jose, CA, 2014, pp. 443-458.  [3] Ranjit Kumaresan and Iddo Bentov. 2014. How to Use Bitcoin to Incentivize Correct Computations. In Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security (CCS '14). Association for Computing Machinery, New York, NY, USA, 30–41.  [4] H. Gao, Z. Ma, S. Luo and Z. Wang, "BFR-MPC: A Blockchain-Based Fair and Robust Multi-Party Computation Scheme," in IEEE Access, vol. 7, pp. 110439-110450, 2019, doi: 10.1109/ACCESS.2019.2934147.  [5] 黄建华, 江亚慧， 李忠诚 《利用区块链构建公平的安全多方计算》计算机应用研究 2020年1月 37卷 第一期 | | |
| 现有技术的缺点，本创意解决的问题及有益效果 | 1. **现有技术的缺点**   在文献[1]中，该模型是在联盟链的环境下，链下的多方安全计算使用的是基于ElGamal的全同态加密方案，一方面是全同态加密算法能够对密文直接进行操作，模型简单，但是全同态的复杂度太高，使用效率太低，目前还不能商业化的应用，另一方面，由于做同态计算的密文要求是同一个公钥加密的，而各个用户使用的密钥并不相同，所以该模型中需要有可信的第三方。文献[4]假设参与方都是理性的，通过引入reputation system作为激励，来保证协议的公平性，但是没有考虑效率。文献[5]的链下执行的是一个通用的公平的MPC协议，不公平指的是，在秘密重建的过程中，有的恶意节点获取了其他节点的信息分片后自己可以恢复出秘密，且不公开自己的分片导致诚实节点恢复不出秘密。文献[5]的秘密重建部分在链上执行，需要与智能合约进行多轮交互，不公开分片的节点将被判断为恶意节点进行惩罚，从而保证公平性。但是由于该MPC协议是基于Gennaro方案，一方面是效率低，特别是当参与方增加时，复杂度很高，另一方面是，该协议中的恶意方不能超过n/2。本创意将着重解决如何提高链下的MPC协议的效率，降低在参与方增多时的复杂度，提高对恶意节点的抵抗。   1. **本创意需要解决的技术问题**   要解决基于区块链的多方协作的效率问题，需要考虑到具体的应用场景，采用适合的安全多方计算协议，考虑区块链的  要解决视频文本描述生成，需要结合计算机视觉、机器学习、自然语言处理等多个学科的方法。如何挖掘和表示视觉知识，是本创意要解决的基本问题。这里的知识主要包括属性知识和关系知识。其中属性知识表示描述物体和行为等视觉对象的语义属性，关系知识表示视觉对象之间的相互关系。如何利用物体、行为等视觉对象的属性知识和关系知识来指导特定视频的文本描述生成是本创意要解决的核心问题。 | | |
| 本创意技术方案的详细阐述（创意核心内容，请尽可能详实描述） | 目前已有的最有效的视频文本描述方法都是受益于机器翻译技术的进展。通过最大化条件概率p(T|S)，一个用源语言描述的句子S可以被翻译成为用目标语言描述的句子T。在早期研究中，机器翻译一般是通过多个子任务（对单词进行翻译，单词校准，重排序等）来完成。近年来，随着深度学习技术的兴起，机器翻译领域的研究人员普遍采用基于循环神经网络的编码-解码模型。一个用于编码的LSTM把输入句子转换成为一个特征向量，然后另外一个解码LSTM基于这个特征向量生成目标语言的句子。由于这种方法简单有效，在本创意中，我们也基于这种编码和解码的算法框架。传统的视频文本描述生成方法需要基于大量的有标记训练样本，为了减小这种依赖性，本创意采用视觉知识记忆模型来辅助LSTM生成文本描述。这样可以使得生成的句子包含视频中的所有物体、行为以及它们相互之间的联系。  本创意采用的视频文本描述生成算法的框架如图1所示。在生成每个单词之前，首先采用视觉注意力模型计算视频在时间和空间上不同局部区域的注意力系数，根据注意力系数对局部区域的特征计算加权和得到当前时刻的视觉特征表示。然后采用视觉知识记忆模型得到与视频内容相关的知识表示向量。最后视频的视觉特征和知识表示向量被同时输入到循环神经网络（LSTM）来生成单词。整个算法可以采用目前较为流行的Pytorch，Caffe和Tensorflow等深度学习框架来实现。算法框架中最重要的视觉注意力模型，视觉语义表示和视觉知识记忆模型将在下面进行详细介绍。   1. **视觉注意力模型**   视觉注意力模型可以在神经网络训练过程中自适应地从整个视频中选择最重要的视觉特征表示。考虑到在图像分类、物体检测等视觉任务中，卷积网络的高层特征可以表达和物体相关的高层语义信息，我们采用预训练的残差网络的最后一个卷积层来提取视频帧的视觉特征。如果用 表示所有视频帧的不同局部位置的视觉特征向量，根据视觉注意力计算的视觉特征可以被定义为，这里表示t时刻第i个局部特征的注意力系数。这些系数可以通过一个线性变换函数和Softmax函数计算得到),其中是t-1时刻的LSTM网络的隐含层特征向量。   1. **视觉语义表示**   相比传统的数据驱动的方法，知识驱动的视频内容理解需要学习与视频中的物体、行为等视觉对象相关的大量知识来指导视频内容的理解。本创意主要研究视觉对象的属性知识，以及视觉对象之间的关系知识。具体来说，我们采用斯坦福大学建立的视觉知识图谱数据集Visual Genome来学习属性和关系知识的视觉语义表示。Visual Genome数据集是目前最大的人工标记的视觉知识图谱数据集，其中包含10万张图片，500万个图片区域描述，100万个视觉问答，300万个物体，200万个属性和200万个关系标注。在本创意中，我们主要利用属性标注和关系标注的图片。    图3. 视觉语义表示学习算法框架  本创意采用图3中所示的对比性约束神经网络算法来学习视觉语义表示。对比性约束神经网络由两支分别用于图片特征表示和文本特征表示的神经网络构成。这里文本输入可以是图片的属性标注或者关系标注。输入图片经过多层卷积网络以及全连接网络最终输出视觉特征表示。对于输入的文本，首先根据预训练的word2vector模型得到每个单词（word）的语义向量，这些语义向量经过池化（pooling）操作得到文本的语义特征表示。在训练的时候，我们基于文本和图片的语义特征建立排序损失函数。我们所采用的对比性约束神经网络的主要目标是学习把图片和文本映射到共享语义特征空间的变换。  假设有N个图片-属性和图片-关系的多模态样本对，我们采用的对比性约束神经网络算法可以被公式化地描述为：  （1）  这里表示图片和属性（关系）文本之间的距离。表示计算图片的视觉语义表示的神经网络，表示计算文本的视觉语义表示的神经网络。表示第i个图片-文本对的配对标记。这里表示文本和图片是相关的，而表示它们是随机生成的不相关的样本对。表示处理图片的卷积网络和处理文本的卷积网络中的参数，表示参数的L2正则，是用于控制两项比例的权重。  在图3所示的对比性约束神经网络训练完成以后，我们就可以用计算图片的视觉语义表示，用文本的视觉语义表示。   1. **视觉知识记忆模型**   基于步骤(1)中得到的视频的视觉特征表示，我们在步骤(2)中介绍的Visual Genome数据集中找到视觉特征距离最近的K个图片。然后利用步骤(2)中的神经网络和计算K个图片的视觉语义表示和对应的属性（关系）文本的视觉语义表示。K个图片的视觉语义表示作为视觉知识记忆模型中的K个Key向量，K个文本的视觉语义表示作为K个Value向量。在预测第t个单词的时候，知识表示向量可以用下式来计算:  , （2）  其中表示第i个Value向量的权重，按照如下方式计算得到：  （3） | | |
| 检索结果及对比文件 | **（1）检索相关文件**  [7] Arnav Kumar Jain, Abhinav Agarwalla, Kumar Krishna Agrawal, Pabitra Mitra: Recurrent Memory Addressing for Describing Videos. CVPR Workshops 2017: 2200-2207  [8] Junbo Wang, Wei Wang, Yan Huang, Liang Wang, Tieniu Tan: Multimodal Memory Modelling for Video Captioning. CoRR abs/1611.05592 (2016)  [9] Cesc Chunseong Park, Byeongchang Kim, Gunhee Kim: Attend to You: Personalized Image Captioning with Context Sequence Memory Networks. CVPR 2017  **（2）本创意的创新技术点**  目前已有少量基于记忆网络的视频文本描述方法[7,8]. 但这两个工作所采用的记忆网络模型并没有引入外部知识，写入到内存的只是训练集上的视频和文本的向量表示。这种记忆网络模型可以帮助循环神经网络生成更长的句子。但无法表达训练视频中出现次数较少的视觉对象以及他们之间的关系。和这两种方法相比，我们为视频文本描述生成网络设计了视觉知识记忆模型来辅助提取视频中视觉对象的属性信息和视觉对象之间的关系信息。 我们在记忆网络中保存了Visual Genome数据集中的图片和属性（关系）文本的视觉语义表示向量。在单词生成过程中，记忆网络可以自适应的选取与视频内容有关的视觉属性和关系信息来帮助预测下一个单词。在与视频文本描述生成紧密相关的图片文本描述领域，文献[9]提出了基于记忆网络的个性化图片描述方法。这种方法采用记忆网络的主要目的是在生成图片描述的同时，考虑图片的用户信息。因此保存在记忆网络内存的是图片对应用户产生的历史语句。通过用户的历史语句，图片描述生成可以学习到用户的语言习惯。和这种方法相比，我们采用记忆网络有不同的目的。我们采用视觉知识记忆模型的目的是提取视频内容相关的属性和关系信息，并对视频文本描述的生成进行指导。 | | |