Transactions on Computer Science and Technology
Volume 9, 2022, PP. 8-13

Research on Privacy Protection of Intelligent Applications

Shengdi Zhao, Yanxin Yao †

School of Information and Communication Engineering, Beijing Information Science & Technology University, Beijing, China

†Email: 1570573856@qq.com

Abstract

There are many privacy protection methods in the field of artificial intelligence. Firstly, this paper summarizes the related secure multi-party privacy computing methods, image retrieval privacy protection methods, and machine learning privacy protection methods. At present, edge computing provides many benefits for various intelligent applications, but at the same time, when end-to-edge distributed computing is carried out during the unloading process of edge computing, privacy disclosure will occur. In this paper, a distributed layout privacy protection strategy is proposed to ensure the two-way tasks of face attribute feature extraction and privacy feature hiding. The main purpose is to avoid the remote transmission of privacy information characters while transmitting the main tasks, and to eliminate the hidden processing on mobile devices as much as possible, so as to improve the effectiveness of privacy protection. From the final experimental results, it can be concluded that the network framework algorithm can effectively achieve the effect of privacy blanking.

Keywords: Edge Computing; Privacy Concealment; Gradient Flipping Layer; Local Fine-tuning

智能应用的隐私保护研究

赵生弟，姚彦鑫

北京信息科技大学，北京 100101

摘 要：关于人工智能领域的隐私保护方法有很多。本文首先对相关的安全多方隐私计算方法、图像检索隐私保护方法、机器学习的隐私保护方法进行了综述。目前边缘计算为多方面智能应用程序提供了很多好处，但与此同时在边缘计算卸载过程中进行端-边分布式计算的时候，会产生隐私泄露问题。本文提出保证人脸属性特征提取、隐私特征消隐的双向任务的分布式布局隐私保护策略，主要目的就是在传输主要任务的同时，避免隐私信息人物向远程传输而将其尽量在移动设备上进行消隐处理，以提高隐私保护的有效性。从最终实验结果可以得出，该网络框架算法可以有效达到隐私消隐的效果。

关键词：边缘计算；隐私消隐；梯度翻转层；局部微调

1 引言

边缘计算[1]（Mobile Edge Computing, MEC）是将用户智能应用的计算任务卸载到临近的边缘服务器中，可以降低本地资源的开销，缓解云计算通信所带来的通信延迟与通信拥塞问题，提高任务计算效率，满足用户应用服务体验质量需求。边缘计算得到了越来越多的重视[2][3]。但由于边缘智能需要卸载的任务到边缘服务器或者其他服务节点上分散计算，而这些服务节点或服务器的所有者可能以各种形态存在，从而引发了更严峻的隐私保护的问题——我们称之为边缘计算的新型隐私保护问题。例如，服务器的中间处理结果可能包括隐私数据，攻击者可能通过攻击卸载到边缘服务器的任务数据而侵犯隐私[4][5]。如果隐私数据

被进一步传输至远程服务器，传输过程中隐私问题也可能泄露。现有边缘计算所提供的隐私属性保护并不完善，不能保证上传到远程服务器上数据的安全性不被侵犯。

到目前为止,很多隐私保护的方法被提出。然而现有的隐私方法都不是针对于边缘计算场景提出的，更多都是考虑一次性的变换、加扰、加密[6][7][8]，然后在云端服务器进行解变换、解扰、解密等，这会引入大量计算量。所以就现有的边缘计算来说，所提供的隐私属性保护并不完善，并不能保证上传到边缘上的数据安全性不被侵犯。目前很多关于人脸的边缘智能应用，为了进行各种人脸属性特征提取，由于身份特征和这些属性存在很多关联性，在进行某种属性的识别的同时，可能存在把身份属性的特征也同时提取出来的隐患。

为了保护任务隐私不被边缘服务器所窃取，在不增加额外计算量的情况下，将任务分散在各个节点上，对保护用户隐私具有很大的挑战性。为此，本文设计了一个适用于边缘智能的人脸图像隐私保护方法——适用于图像边缘智能的隐私特征消隐方法。该方法，一方面通过采用梯度翻转层 GRL[9][10][11]，将身份敏感信息在近用户端设备端进行消隐，使信息传到远程服务器端时，已经再没有敏感信息，而同时采用对抗调整的方式[12][13]，保留执行主要任务的信息。实验结果表明，对于不同的人脸属性，若想在识别一个属性的同时，而识别不出来另一个属性。

2 相关方法

2.1 安全多方隐私计算方法

多方安全计算在无第三方的条件下，采用协议标准的解决办法，让多个参与方共同合作，协同完成计算，计算所需的数据在整个计算过程中始终保存在各参与方的本地数据库，这就保证了输入数据的隐私性，各参与方协同计算，任务完成后返回各自的计算结果，保证了计算的正确性[14][15]。

安全多方计算也常被用来保护生物特征识别系统中敏感数据的隐私。通过安全多方计算的人脸识别隐私保护方案，既能进行人脸识别，又能保证人脸数据的隐私性，此过程通常需要同时利用多个云服务器来实现[16]。Cai 等人[17]提出了一种联合外包云环境下的安全人脸识别方案，可以针对半可信模型有效地保护用户敏感数据的隐私。特征脸算法作为人脸识别的方案并且以隐私保护的方式运行在两个半可信且非共谋的云服务器上，Paillier 密码系统用于安全模型的设计从而保证该方案的识别结果与标准特征脸算法匹配结果完全相同[18]。Ma 等人提出了一个安全人脸认证方案[19]，使用卷积神经网络提取人脸的特征向量，采用两个非共谋的服务器，一个服务器用于存储用户的加密人脸特征，另一个用于完成人脸验证。所有数据都是以密文状态传输，除人脸认证服务器外的其他设备都无法解密数据。Haghighat 等人[20]利用安全电路和同态加密的安全多方计算技术实现加密域下人脸识别，该方案采用 K-D 树结构，不仅有利于实现密文下人脸识别，还能极大地提高系统的识别效率。

2.2 图像检索隐私保护

随着现代科学技术的发展以及数字成像设备的普及，上传到网络上的图像，往往含有用户的隐私信息。采用图像检索方法，包括文本检索和内容检索，可能使上传到服务器图像的隐私数据在未加密的状态下，被窃取和泄露[21][22]。利用内容检索可以窃取图像的视觉特征（颜色、纹理、形状），利用文本检索会根据语义属性标注从海量图片选取图片，造成隐私泄露。

对检索图像创建加密索引是一种典型的解决方案，可以兼顾加密技术带来的计算复杂性过高和密文存储空间过大的问题[23][24]。比如 Ferreira 等人[25]针对带隐私保护的基于内容的图像检索 (PCBIR) 提出一种新颖的加密方案，对一张图像的颜色和纹理信息分开处理，用随机性加密方案加密纹理信息从而保护图像内容，用确定性加密方案处理颜色信息以支持密文下检索。Xia 等人[26]提出了一种安全的 K 最近邻 (KNN) 算法对检索图片创建加密索引，使得服务器能够在不增加通信负担的情况下有效地排列出搜索结果。局部敏

感哈希 (LSH) 技术用来使相似图片聚集，从而提高检索效率。Xu 等人[27]首次提出了一种基于正交分解的 PCBIR 方案，图片通过正交分解被分解为两部分，加密和特征提取分别执行，与其他方案不同的是该方案对加密算法没有特别的要求。

2.3 机器学习的主要隐私保护技术

随着互联网的普及和信息采集技术的提高，大量的敏感数据被收集。机器学习方法可精准推断个人信息或预测个人行为，若此类数据如果被不恰当地使用,将造成非常严重的后果[28]。

Sweeney 提出的匿名方法是一个从数据层面上进行隐私保护的技术匿名化技术，通过将机密数据的关键部分模糊化，来保护了数据隐私，但数据访问者有很多方法比如经过大数据技术，利用其它特征的信息来反推出某人在数据表中对应的那一条数据，从而导致数据准确度受到严重影响，可以分析出隐私信息[29]。机器学习算法通过加密技术在密文的数据上进行操作, 具有隐私保护程度好，数据准确的优点，可以达到原始数据不可获取的目的，保护了数据隐私，但是其加密技术计算延时长，能量消耗高[30][31]，加密技术算法需进一步提升。此外，通用机器学习策略 PATE 满足差分隐私标准，在数据和标签上同时加扰，然后利用训练数据和集成的算法训练外部能识别加扰数据的模型；该算法需要对加扰数据进行集成训练，过程十分复杂[32][33]。基于深度学习的图像隐私感知方法[34],通过注意力机制等准确地区分隐私图像并定位图像中的隐私区域，对其进行选择性保护。但在隐私信息和主要任务信息重叠的应用中不适用。

以上这些隐私方法都不是针对于边缘计算场景提出的，更多是考虑一次性的变换、加扰、加密 [6][7][8]，然后在远程服务器进行解变换、解扰、解密等，会引入大量计算量。

3 方法

将以人脸属性识别为例，将基于人脸的性别识别作为主要任务，人脸的身份信息作为隐私信息来进行研究，以探讨隐私保护策略。因此提出保证人脸属性特征提取、隐私特征消隐的双向任务的深度网络构建方法。设计一个好的特征提取器，将在原始数据域上处于交织（重叠）状态的人物的性别信息和人物身份敏感信息分离开来，兼顾敏感信息两者的处理，提取的中间层特征需要在保护人物身份敏感信息的同时，保留有关主要任务性别的必要信息。

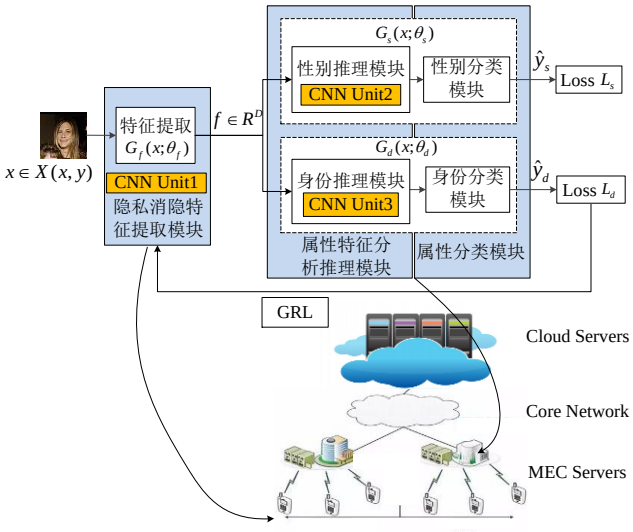

图 1 人脸智能局部微调隐私保护模型

根据隐私消隐目的，先划分网络，可以将整体划分为 3 个模块：具有身份特征消隐和性别特征提取功能的隐私消隐特征提取模块、属性特征分析推理模块和属性分类模块，不同模块在不同的实体上运行，如图 1 所示：隐私消隐特征提取模块放在距离用户比较近的用户终端和距离用户比较近的边缘服务器上处理。以

保证输出的特征在传到距离比较远的边缘服务器的时候已经不存在隐私特征[21][22]。属性特征分析推理模块和属性分类放在远于用户的设备端，以保证在远端服务器的时候不存在隐私特征，在可以识别性别的同时，识别不出身份。

本文目标是期望隐私消隐特征提取模块（主要任务特征提取模块）的输出，经过属性特征分析推理模块后，可以分别达到性别识别和身份隐私消隐的目的。此训练过程需要对属性特征分析推理模块和隐私消隐特征提取网络模块进行对抗微调的目的。此训练过程需要对属性特征分析推理模块和隐私消隐特征提取网络模块进行对抗微调。

最终测试结果：对于性别，进行性别识别的验证选取 16 张图片，测试性别的真实值与预测值对比正确率是 100%。最后看一下身份识别的验证，如混淆矩阵图 2 所示。

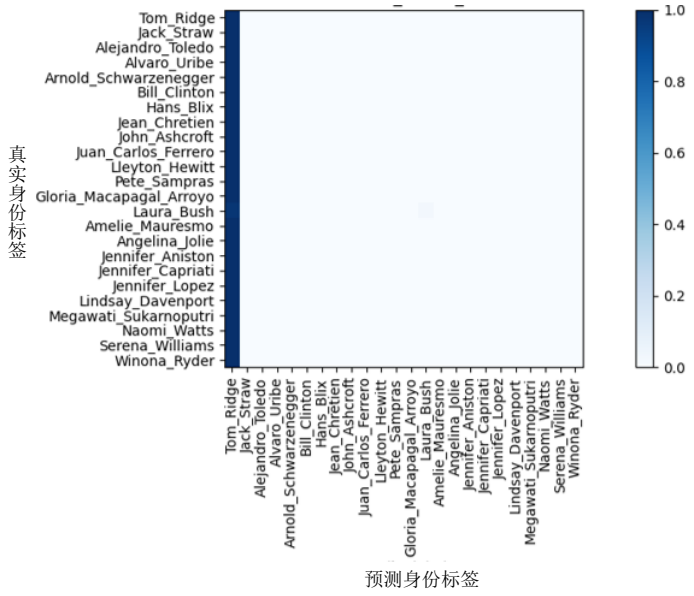

图 2 最终的身份识别混淆矩阵图

可以看出，针对于所有的身份，混淆矩阵图颜色分布并不是斜散分布。测试精确度是 3.19%，表明身份隐私保护取得不错的成效。

4 结论

基于之前的深度学习方法在隐私信息和主要任务信息重叠的应用中不适用。提出了一种基于人脸边缘智能应用隐私保护研究方法。该方法通过网络共享，通过梯度翻转层来进行隐私消隐，使得尽可能在消除隐私特征的同时，保留主要特征。在 LFW 数据集上进行了实验，通过数据集的特征和实验结果对本章提出的模型进行了分析。

基于实验分析，本文研究的便于分布式布局的隐私保护策略——隐私特征消隐方法，为避免隐私信息向远程传输而将其尽量在距离用户比较近的用户终端和边缘服务器上处理进行消隐处理，对提高隐私保护具有很高的有效性。

REFERENCES

[1] H Li, G Shou, Y Hu, et al. Mobile edge computing: Progress and challenges[C]. 2016 4th IEEE International Conference on Mobile Cloud Computing, Services, and Engineering (MobileCloud), 2016: 83-84.

[2] Ma L, Q Pei, H Xiao, et al. Edge computing enhanced privacy preserving for location-based services[C]. IEEE INFOCOM 2019-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), 2019: 1-6.

[3] X Xu, C He, Z Xu, et al. Joint optimization of offloading utility and privacy for edge computing enabled IoT[J]. IEEE Internet of Things Journal, 2019, 7(4): 2622-2629.

[4] M Sharif, S Bhagavatula, L Bauer, et al. Accessorize to a crime: Real and stealthy attacks on state-of-the-art face recognition[C]. Proceedings of the 2016 Acm Sigsac Conference on Computer and Communications Security, 2016: 1528-1540.

[5] D J Robertson, A M Burton. Unfamiliar face recognition: Security, surveillance and smartphones[J]. The Journal of the Homeland Defense and Security Information Analysis Center, 2016: 14-21.

[6] S C Sen-ching, M U Rafique, W Tan. Privacy-preserving distributed deep learning with privacy transformations[C]. 2018 IEEE International Workshop on Information Forensics and Security (WIFS), 2018: 1-7.

[7] Z Chen, L Li, H Peng, et al. An evaluation method of image scrambling degree based on pixel distribution[C]. 2018 International Symposium on Security and Privacy in Social Networks and Big Data (SocialSec), 2018: 206-211.

[8] W Zhang, H Wang, D Hou, et al. Reversible data hiding in encrypted images by reversible image transformation[J]. IEEE Transactions on Multimedia, 2016, 18(8): 1469-1479.

[9] Y Ganin, V Lempitsky. Unsupervised domain adaptation by backpropagation[C]. International Conference on Machine Learning, PMLR, 2015: 1180-1189.

[10] L Cui, Z Chen, A Wang, et al. Development of a robust cooperative adaptive cruise control with dynamic topology[J]. IEEE Transactions on Intelligent Transportation Systems, 2021: 1-12.

[11] Z Chen, B B Park. Cooperative adaptive cruise control with unconnected vehicle in the loop[J]. IEEE Transactions on Intelligent Transportation Systems, 2020: 1-11.

[12] K Karthik, S Kashyap. Transparent hashing in the encrypted domain for privacy preserving image retrieval[J]. Signal, Image and Video Processing, 2013, 7(4): 647-664.

[13] W Lu, A L Varna, M Wu. Confidentiality-preserving image search: A comparative study between homomorphic encryption and distance-preserving randomization[J]. IEEE Access, 2014, 2: 125-141.

[14] H Akbari-Nodehi, M A Maddah-Ali. Secure coded multi-party computation for massive matrix operations[J]. IEEE Transactions on Information Theory, 2021, 67(4): 2379-2398.

[15] P Ah-Fat, M Huth. Optimal accuracy-privacy trade-off for secure computations[J]. IEEE Transactions on Information Theory, 2018, 65(5): 3165-3182.

[16] 杨乾. 一种新的人脸识别隐私保护方案[D]. 华中师范大学, 2020.

[17] Y Cai, C Tang. Securely outsourced face recognition under federated cloud environment[C]. 2016 15th International Symposium on Parallel and Distributed Computing (ISPDC), 2016: 269-276.

[18] S Pearson, A Charlesworth. Accountability as a way forward for privacy protection in the cloud[C]. IEEE International Conference on Cloud Computing, Springer, Berlin, Heidelberg, 2009: 131-144.

[19] Y Ma, L Wu, X Gu, et al. A secure face-verification scheme based on homomorphic encryption and deep neural networks[J]. IEEE Access, 2017, 5: 16532-16538.

[20] M Haghighat, S Zonouz, M Abdel-Mottaleb. CloudID: Trustworthy cloud-based and cross-enterprise biometric identification[J]. Expert Systems with Applications, 2015, 42(21): 7905-7916.

[21] 彭远帆. 隐私保护的图像检索关键技术研究[D]. 北京工业大学, 2015.

[22] H Akbari-Nodehi, M A Maddah-Ali. Secure coded multi-party computation for massive matrix operations[J]. IEEE Transactions on Information Theory, 2021, 67(4): 2379-2398.

[23] Z Huang, M Zhang, Y Zhang. Toward efficient encrypted image retrieval in cloud environment[J]. IEEE Access, 2019, 7: 174541-174550.

[24] Z Xia, L Jiang, D Liu, et al. BOEW: A content-based image retrieval scheme using bag-of-encrypted-words in cloud computing[J]. IEEE Transactions on Services Computing, 2022, 15(1): 202-214.

[25] B Ferreira, J Rodrigues, J Leitão, et al. Towards an image encryption scheme with content-based image retrieval properties[M]. Data Privacy Management, Autonomous Spontaneous Security, and Security Assurance, Springer, Cham, 2014: 311-318.

[26] Z Xia, N N Xiong, A V Vasilakos, et al. EPCBIR: An efficient and privacy-preserving content-based image retrieval scheme in cloud computing[J]. Information Sciences, 2017, 387: 195-204.

[27] Y Xu, J Gong, L Xiong, et al. A privacy-preserving content-based image retrieval method in cloud environment[J]. Journal of Visual Communication and Image Representation, 2017, 43: 164-172.

[28] 唐鹏, 黄征, 邱卫东. 深度学习中的隐私保护技术综述[J]. 信息安全与通信保密, 2019, (6): 55-62.

[29] S Abd Razak, N H M Nazari, A Al-Dhaqm. Data anonymization using pseudonym system to preserve data privacy[J]. IEEE Access, 2020, 8: 43256-43264.

[30] Y Aono, T Hayashi, L Wang, et al. Privacy-preserving deep learning via additively homomorphic encryption[J]. IEEE Transactions on Information Forensics and Security, 2017, 13(5): 1333-1345.

[31] G S Uehara, A Spanias, W Clark. Quantum information processing algorithms with emphasis on machine learning[C]. 2021 12th International Conference on Information, Intelligence, Systems & Applications (IISA), 2021: 1-11.

[32] 李晓东, 韩青, 金鑫. 一种基于深度学习和同态加密的安全高效的人脸识别方法[P]: CN, CN201810973325.6, 2019-1-4.

[33] N Papernot, S Song, I Mironov, et al. Scalable private learning with pate[J]. 2018: arXiv: 1802.08908.

[34] H Wang, Y Zhang, L You, et al. Image privacy perception method based on deep learning[P]: US, US16099836. 2021-7-22.

【作者简介】

[1]赵生弟（1995-），女，汉族，2015 年于河南理工大学获学士学位，现为北京信息科技大学硕士研究生，主要研究方向为智能信号处理。
Email:1570573856@qq.com

[2]姚彦鑫（1982 -），女，汉族，2009 年于北京航空航天大学获博士学位，现为北京信息科技大学教授，主要研究方向为无线通信与节能通信网络、压缩感知与智能信号处理。
Email: yanxin_buaa@126.com。