Yin *et al. Hum. Cent. Comput. Inf. Sci.* (2019) 9:33
https://doi.org/10.1186/s13673-019-0195-4

Human-centric Computing and Information Sciences

RESEARCH Open Access

Local privacy protection classification based on human-centric computing

Chunyong Yin[1], Biao Zhou[1], Zhichao Yin[2] and Jin Wang[3*]

*Correspondence:
jinwang@csust.edu.cn
[3] School of Computer & Communication Engineering, Changsha University of Science and Technology, Changsha 410004, China
Full list of author information is available at the end of the article

Abstract

Human-centric computing is becoming an important part of data-driven artificial intelligence (AI) and the importance of data mining under Human-centric computing is getting more and more attention. The rapid development of machine learning has gradually increased its ability to mine data. In this paper, privacy protection is combined with machine learning, in which a logistic regression is adopted for local differential privacy protection to achieves classification task utilizing noise addition and feature selection. The design idea is mainly divided into three parts: noise addition, feature selection and logistic regression. In the part of noise addition, the way of adding noise using Laplace mechanism to original data achieves the purpose of disturbing data. The part of feature selection is to highlight the impact of noised data on the classifier. The part of logistic regression is to use logistic regression to implement classification task. The experimental results show that an accuracy of 85.7% can be achieved for the privacy data by choosing appropriate regularization coefficients.

Keywords: Data-driven, Human-centric computing, Privacy protection, Machine learning, Classification

Introduction

Human-centric computing is the wave of the future, which not only helps bring us convenience but also presents new challenges of processing and analyzing massive data. With the rapid development of the Internet and the arrival of the era of big data, our privacy data is also unconsciously leaked [1–5]. From the existing research situation, the ways of privacy protection mainly include data distortion, encryption and block chain [6]. The classical k-anonymity method was proposed by Sweeney and Samarati [7]. Since this method does not include randomized processing, attackers can still infer privacy information related to individuals from data sets satisfying the anonymity nature of this method [8]. At the same time, k-anonymity method can not deal with consistency attacks and background knowledge attacks. In view of this, Machanavajjhala et al. proposed an improved algorithm [9]. Based on such privacy protection schemes, no reasonable assumptions about attack models are made. In 2006, Dwork et al. proposed a differential privacy model to solve this problem [10]. Dwork analyzed the sensitivity calculation method of each query function in the differential privacy k-means algorithm in detail, proposed different allocation methods of privacy budget in two cases, and gave the total sensitivity of the whole query sequence. In [10], differential privacy, a method

of an attacker cannot distinguish the encrypted results under the state of the inscription was proposed. Because the model does not need to rely on the attacker's background knowledge, and provides a higher level of semantic security for the privacy information, this model is widely used at present.

Around the people-centered data generation sources are also more and more widely, the current society is in the forefront of the development of artificial intelligence. It can be predicted that future computing will be human-centered computing. Data collected from different electronic devices can be processed and analyzed to better design information systems, improve transportation systems, promote social and economic development, and also provide better personalized services [11, 12]. What needs to be emphasized is that the tremendous progress of data analysis and data mining technology in recent years has enabled attackers to mine a lot of privacy-related information from massive data, so the protection of privacy data is an important and arduous task [13]. With the emergence of big data and the gradual maturity of deep learning technologies [14–19], AI is gradually becoming a popular solution for data analysis and prediction [20–23], cutting-edge technologies and Human-centric computing [24]. In [24], Rabaey thinks that the sense of miniaturization, the computational and driving devices, and the appearance of interfaces that fit the human body provide the basis for the symbiosis between biological functions and electronic devices. Hence, the data driven artificial intelligence is required to perform data processing and analyzing in human-centric computing [25]. The security development of AI is the security and development of data in the final analysis. Therefore, in the process of data processing and analysis in AI, the security and privacy issues involved in computation cannot be ignored. Faced with the huge amount of data, especially in the area of health care, more personal privacy data are involved, such as medical history [26, 27], cases, education, family income and so on [28]. Especially for the privacy data, it is more important. Ideally, it not only protects the privacy data, but also realizes a certain degree of data mining without exposing the privacy data [29]. At present, how to find a balance between protecting privacy data and making protected data available is a hot research direction. Therefore, privacy protection is a cross-cutting technology that combines multiple disciplines [30]. It is a natural combination to mine private data through machine learning technology.

In this scheme, the relative balance between data protection and data mining is achieved. In short, the privacy data generated by adding noise is still usable and can be classified by machine learning. The basic design idea is as follows. Firstly, the original data is preprocessed to eliminate useless information, then the local differential privacy is realized by Laplace mechanism, and then the normalization operation is carried out. Next, in order to highlight the impact of noise data and improve the classification effect of the classifier, the necessary feature selection is carried out. Finally, the classical logistic regression in machine learning is used to realize the classification task. There are three key points. The first is the mechanism of adding noise, the second is the method of feature selection, and the third is the realization of logical regression. Noise is added to protect data. The main purpose of feature selection is to highlight the impact of noise data on the subsequent classifier. Logic regression is the way to achieve classification.

The rest of the paper is organized as follows: Relevant work is presented in "Relevant work and theory". After that, In "Proposed method" the proposed method will be

introduced in detail. The experiment and result analysis will be given in "Experiments". Finally in "Conclusion", we summarize the research content and look forward to the future research direction.

The main contributions are as follows:

1. Compared with the research status of privacy protection, we do not use the mainstream clustering methods, but used classification methods. The experimental results show that choosing a reasonable model can not only protect data, but also achieve a certain degree of data mining effect.
2. As far as the degree of privacy protection is concerned, considering the actual situation, some data need not be protected. At the same time, in order not to destroy the distribution of real data sets and ensure the authenticity of classification results, we have not added noise to all data. In reality, the data needed to be protected should often be specified by the user. So we only add noise to local sensitive data to achieve the purpose of protection.
3. In order to highlight the impact of local privacy data on classification results, we have made the necessary feature selection, but also to ensure that the noise feature is not eliminated. The experimental results show that although this method can reduce the accuracy of classification, the data is still available in general.

Relevant work and theory

This section mainly introduces the related work and the related concepts of privacy protection. In the related work part, the application of machine learning method in privacy protection is emphasized. In the related theory part, the key concepts of privacy protection are taken as the focus of introduction.

Relevant work

In the business and medical fields, there are often a large number of private data, and data mining and machine learning will have to deal with such large-scale data. How to mine and analyze useful information is a research topic of privacy protection while protecting privacy information. So the application of data mining and machine learning technology under differential privacy is an important research.

Vadidya et al. proposed a naive Bayesian model based on differential privacy [31]. The model calculated the sensitivity of discrete classification data and continuous classification attributes. This method realized differential privacy protection by adding noise to the parameters of naive Bayesian classifier. Another algorithm for differential privacy protection is a non-interactive anonymous data mining algorithm based on generalization technology proposed by Mohammed et al. [32]. Its main idea is to add exponential noise to achieve differential privacy protection. To increase the similarity in the same class and the discrimination between different classes [33], uses original training samples and their corresponding random-filtering virtual samples by adding random noise to construct a new training set, then exploits the new training set to perform collaborative representation classification.

In the field of privacy protection, machine learning mainly focuses on the classification of supervised learning and clustering of unsupervised learning. In the field of privacy protection, machine learning mainly focuses on the classification of supervised learning and clustering of unsupervised learning. Supervise learning methods in machine learning, such as support vector machines and logistic regression and other methods, to achieve the task of classifying private data [34, 35]. In [36], Jia proposes an approach to preserve the model privacy of the data classification and similarity evaluation for distributed systems. The clustering method of unsupervised learning is more widely used in privacy protection. The problem mainly focuses on the research of wireless sensor networks [37–40], wireless multi-hop networks [41, 42] and smart grid [43] where the AI technologies are widely used to solve the route, service prediction and service selection problem [44–47]. In [48], Gao thinks traditional clustering approaches are directly performed on private data and fail to cope with malicious attacks in massive data mining tasks against attackers' arbitrary background knowledge. To address these issues, the authors propose an efficient privacy-preserving hybrid k-means under Spark. In [49], Kai et al. propose a mutual privacy preservation K-means clustering scheme. It neither discloses personal privacy information nor discloses community characteristic data (clusters).

Relevant theory

This part mainly introduces ε-differential privacy [50], Laplace mechanism and Laplace sensitivity [51, 52].

ε-differential privacy: For a random algorithm M, P_m is a set of all the values that algorithm M can output. If for any pair of adjacent data sets D and $D^{'}$, and any subset P_m of S_m, algorithm M satisfies in formula (1):

$$\Pr\left[M(D) \in S_m\right] \leq e^{\varepsilon} \Pr\left[M\left(D^{'}\right) \in S_m\right] \tag{1}$$

The algorithm M satisfies ε-differential privacy, where ε is the privacy protection budget. The smaller the parameter ε is, the closer the probability distribution is and the higher the degree of protection is. On the contrary, the larger the parameter ε is, the lower the degree of protection.

Laplace mechanism: By adding Laplace noise to the data set to change the real value, the differential privacy is satisfied before and after adding noise, thus protecting the privacy data. It is the most widely used noise mechanism to disturb data and achieve the purpose of privacy protection.

For query function f, the Laplace mechanism satisfies differential privacy by adding noise satisfying $Lap\left(0, \frac{\nabla f}{\varepsilon}\right)$ distribution, in short, $F(D) = f(D) + noise$. Noise is to satisfy Laplace distribution. It should be noted that in order to ensure unbiased estimation, the mean μ of noise added to satisfy Laplace distribution should be 0. That is, $nosie \sim Lap(b)$. The probability density of noise is defined as in formula (2):

$$f(x|b) = \frac{1}{2b} \mathrm{e}^{-\frac{|x|}{b}} \tag{2}$$

Given ∇f, the smaller ε is, the larger b is. Conversely, the larger ε is, the smaller b is.

Laplace sensitivity $D(F)$: If the existence function $F(C) \in R$, the sensitivity of F is defined as in formula (3):

$$D(F) = \max_{C_1, C_2} \|F(C_1) - F(C_2)\|_1 \tag{3}$$

According to the noise level and probability density, the smaller the privacy budget ε is, the higher the degree of protection is.

Proposed method

The design idea is mainly divided into three parts: noise addition, feature selection and logistic regression. The part of noise addition is to realize the disturbance of the original data and protect the privacy data. The part of feature selection is to highlight the impact of noised data on the classifier, so that it can form a stronger contrast with the original data without noise. The part of logistic regression is to use supervised logistic regression to implement classification task. In this section, we need to note the effect of the regularization coefficient on the classification model. In order to highlight the classification results with or without noise under the same classification model, it is necessary to pay attention to the relationship between the noise intensity and the regularization coefficient, taking into account the features selected by the feature selection process.

Main design ideas

The method presented in this paper can be simplified in Fig. 1.

As shown in the Fig. 1, the preprocessing section deals mainly with useless information in raw data. Firstly, the noise addition part mainly adds the noise satisfying Laplace mechanism to the data, which aims to protect the local privacy data. Then the feature selection part is to highlight the performance of privacy data in classifiers on the one hand, on the other hand, to remove redundant features, so that the classifier has a better generalization effect. Then normalization is used to improve the convergence rate of the classifier. Finally, the logical regression part is used as the final classifier to classify the data, and the classification indicators are obtained to evaluate whether the method achieves the desired classification effect and whether it can protect the privacy data while making these data available.

Noise addition part

This section focuses on adding noise to the previously mentioned data set. Firstly, the probability density function of Laplace function is given, and then the influence of parameters ε and b on the intensity of noise addition and the degree of privacy protection is discussed.

The distribution function of noise under different ε is shown in Fig. 2.

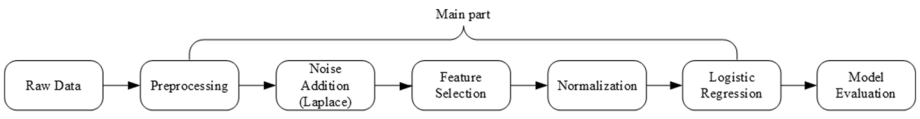

Fig. 1 The simplified structure

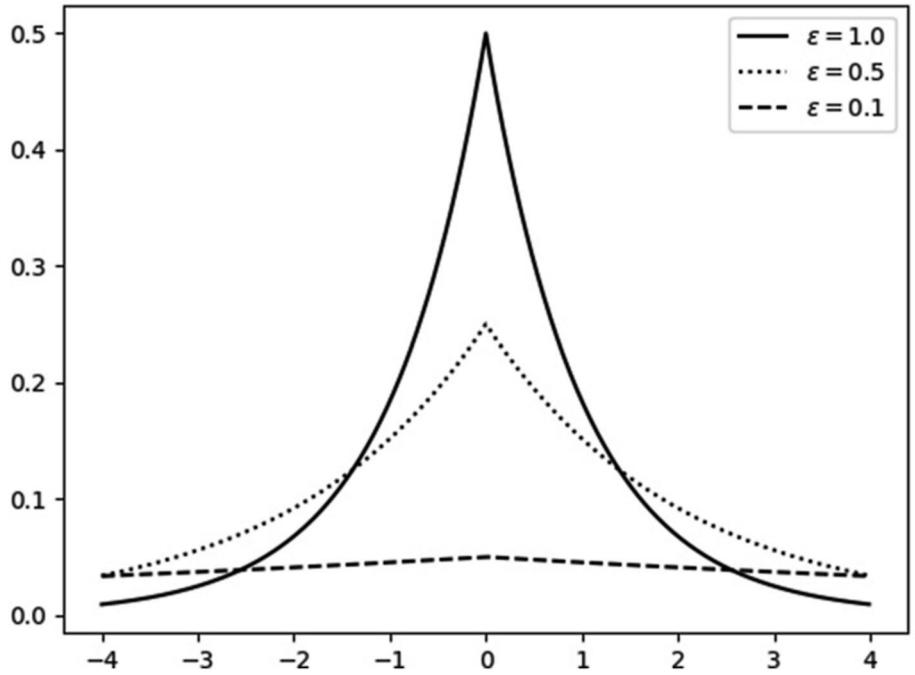

Fig. 2 The distribution function

Assuming $b = \frac{\nabla f}{\varepsilon}$, as shown in the Fig. 2, the smaller ε is, the bigger b is the more uniform the added noise, the smaller the probability of adding noise to 0, the greater the degree of confusion and the higher the degree of protection. Then the privacy budget ε can be modified through b, and then control the intensity of privacy protection.

Feature selection part

In this part, we mainly deal with the methods for feature selection. Here we use Random Forest (RF) in ensemble learning and Classified and Regression Decision Tree (CART) for each decision tree [53]. Firstly, CART is introduced briefly, and then the flow chart of feature selection is given.

CART forms a tree structure, and the splitting of nodes is based on the minimum *Gini* index. The methods of constructing CART are as follows:

Step 1: If the condition of stopping classification is satisfied, the splitting is stopped. The stopping condition is that the *Gini* coefficient is less than the threshold or the number of samples is less than the threshold.

Step 2: Otherwise, the minimum *Gini* coefficient is selected for segmentation. The features for splitting are preserved and the feature set with selection is added sequentially.

Step 3: Steps 1–2 is performed recursively until the splitting is completed.

Step 4: The output is a set of features arranged from strong to weak according to the importance of features.

The *Gini* index is defined as follows in formula (4):

$$Gini(t) = 1 - \sum_{k} \left[p(c_k|t)^2 \right] \tag{4}$$

A represents the classification feature and divides data set D into D_1 and D_2, then *Gini* index of set D is defined as in formula (5):

$$Gini(D, A) = \frac{|D_1|}{|D|} Gini(D_1) + \frac{|D_2|}{|D|} Gini(D_2) \tag{5}$$

The minimum *Gini* index reduces the uncertainty of samples to the greatest extent, so it is the basis for classification. In the process of gradual splitting, features for splitting are preserved at one time until all splitting is completed, and feature sets for classification are output, then the importance of features is consistent with the order of storage.

Random Forest is an integrated algorithm, which belongs to bagging type. It combines several weak classifiers, and the final result is obtained by voting or taking the mean. The algorithm of feature selection using the RF is as follows:

Algorithm: Feature Selection Based on RF for Local Privacy Data

Input:

All feature sets S, The Number of Features in the Selected Feature Set K

S comes from the feature set constructed by the i-th CART. $S = \bigcup_{1}^{n} S_i$

Output: $S_{min(k)}$

$S_{min(k)}$ represents a set of features with K features with minimum out-of-pocket errors.

begin:

Set $S_{min(k)}$ to empty and $Error_{min}$ to 0

for S_i do

step1: Calculate the importance of features and arrange them in descending order.

step2: Determine the removal ratio R and Remove low importance features.

step3: Forming a new feature set S_i'.

step4: If $|S_i'| = K$, $S_i' \rightarrow S'_{min(k)}$, perform step 5. Otherwise, perform steps 1-3.

step5: If $i = 1$, $Error_1 \rightarrow Error_{min}$, $S^1_{min(k)} \rightarrow S_{min(k)}$.

Otherwise, $S^i_{min(k)} \rightarrow S_{min(k)}$, if $i > 1$ and $Error_i < Error_{min}$.

end for

output $S_{min(k)}$

end

Logistic regression design part

The essence of logistic regression is to determine the cost function J for a classification problem, then choose the optimal iteration method to solve the parameters of the model, and finally verify the model. Logical regression is very suitable for binary classification

Table 1 Part of data set

ID	Attribute	Data type
1	Age	Int
2	BMI	Float
3	Glucose	Int
4	HOMA	Float
5	Resistin	Float
6	MCP.1	Float

problems, and the training of the model is faster. In this part, firstly the linear boundary conditions are given, secondly the prediction function is selected and the cost function is constructed, and finally the parameter training method of the model is given.

The boundary conditions can be expressed as in formula (6):

$$z = \sum_{i=1}^{n} \theta_i x_i \tag{6}$$

θ_i is the parameter to be trained in the model, x_i is the attribute value of the sample and z is the linear output.

The predictive function can be expressed as in formula (7):

$$h_\theta = \frac{1}{1+e^{-z}} = \frac{1}{1+e^{-\sum \theta_i x_i}} \tag{7}$$

h_θ is the probability value of z prediction.

The cost function with regularization can be expressed as in formula (8):

$$J(\theta) = -\left[\frac{1}{m}\sum_{i}^{m}\left(y^{(i)}\log\left(h_\theta\left(x^{(i)}\right)\right) + \left(1-y^{j}\right)\log\left(1-h_\theta\left(x^{(i)}\right)\right)\right)\right] + C\sum_{j=1}^{n}\theta_j^2 \tag{8}$$

$J(\theta)$ is the cost function, θ is the model parameter, and C is the regularization coefficient. The stochastic gradient descent algorithm is used to minimize the cost function, and then the model parameters are obtained.

Experiments

In this section, the experimental environment, the source and processing of data sets and three different groups of experiments are described.

Experimental environment

The experimental environment includes Inter (R) Core (TM) i5-4210 M CPU @2.60 Ghz, RAM 4.00G, operating system win10 and programming language python. The experimental data set is from UCI repository. In the data set, there are 116 samples, each of which s has 10 attributes and the classification label of which is 1 (healthy controls) or 2 (patients).

Table 2 Confusion matrix

	Predicted value		Total
	1	0	
True value			
1	True positive (TP)	False negative (FN)	True positive (TP + FN)
0	False positive (FP)	True negative (TN)	True negative (FP + TN)
Total	Predicted Positive (TP + FP)	Predicted Negative (FN + TN)	TP + FN + FP + TN

The data set URL is http://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Coimbra. Part of the data set information is shown in Table 1.

In experiments, BMI index is selected as the privacy data, which is processed by adding noise to form privacy protection data.

Evaluations index of classification result

In this part, the accuracy (ACC), the true positive rate (TPR) and the false positive rate (FPR) are calculated by using the confusion matrix, and then the meaning of the receiver operating characteristic curve (ROC) drawing and the area under the curve (AUC) is explained. These indicators will serve as the basis for evaluating classification models. A detailed description of the confusion matrix is shown as Table 2.

The accuracy calculation is shown in formula (9).

$$ACC = \frac{TP+TN}{TP + FN + FP + TN} \tag{9}$$

Because the accuracy of a classification model is relatively one-sided, we need to introduce the values of FPR and TPR, and form a set of FPR and TPR values according to different thresholds, then draw ROC curve, and then calculate the area under the curve, that is, the value of AUC. The closer the corresponding AUC value is to 1, indicating that the better the effect of the classifier.

The definitions of FPR is shown in formula (10) and TPR is shown in formula (11).

$$FPR = \frac{FP}{FP + TN} \tag{10}$$

$$TPR = \frac{TP}{TP + FN} \tag{11}$$

As the threshold decreases, more and more instances are classified into positive classes, but these positive classes are also doped with real negative instances, that is, TPR and FPR will increase simultaneously. As the threshold increases gradually, fewer instances will be predicted as positive classes, and the values of FPR and TPR will decrease and tend to 0 at the same time. When the threshold decreases gradually, more instances will be predicted as positive classes, then FPR and TPR will gradually increase, and tend to 1 at the same time. The threshold here refers to the threshold that an instance is predicted to be positive, an instance larger than the threshold is predicted

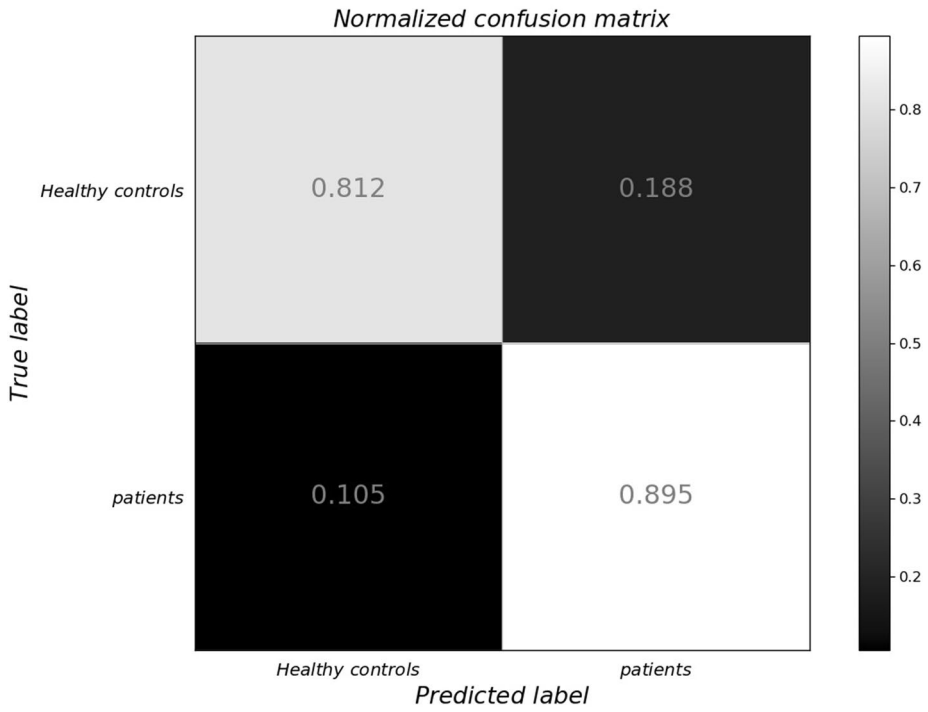

Fig. 3 Confusion matrix for a given noise intensity of 1.0

to be positive, and a case smaller than the threshold is predicted to be negative. Ideally, *FPR* is 0 and *TPR* is 1.

Experimental results and analysis

In this experiment part, we mainly design the influence of data on classification effect under the condition of adding noise or not, the classification effect of the same regularization coefficient C under the condition of adding different intensity of noise, and the classification situation of different regularization coefficient under the same intensity of noise.

The confusion matrix for a given noise intensity of 1.0 is given in Fig. 3.

As shown in Fig. 3, the probability of correct prediction is relatively high. Especially, the accuracy of predicting real patients to be sick is 89.5%, which is much higher than 81.2% for health control. However, it was noted that the probability of predicting real patients to be healthy and controllable was 18.8%, which indicated that there were a certain number of false positive cases (*FP*), which would have a negative impact in the actual environment. After analysis, it is known that the data after noise processing does have a better accuracy, but there may also be a risk of high false positive rate.

The effect of data on Classification under noisy and non-noisy conditions, and gives it in the form of *ROC* is shown in Fig. 4.

Firstly, in Fig. 4, it is pointed out that the stronger the noise is, the higher the interference intensity is. That is to say, the smaller the noise intensity is, the closer the new data is to the original data. In the absence of noise, the classification accuracy is 88.6%, while

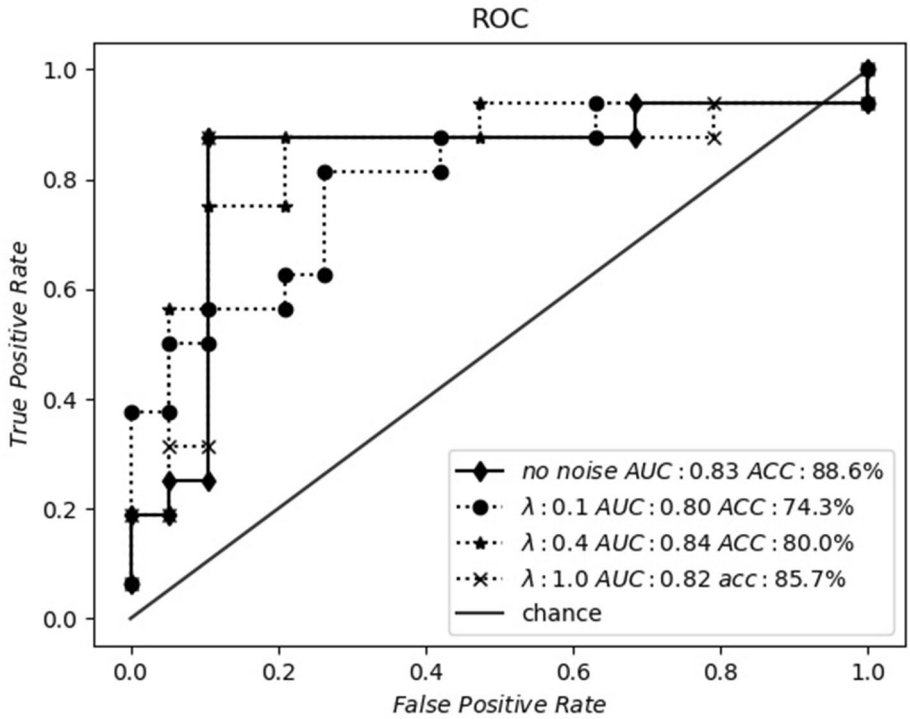

Fig. 4 Different ROC and corresponding accuracy

in the presence of noise intensity of 1.0, the classification accuracy is only reduced by about 3%. With the noise intensity increasing, the accuracy is also gradually reduced. When the noise intensity is 0.1, the noise interference degree is the largest, resulting in the classification accuracy reduced by about 14.3%. This shows that the noise interference ability does affect the classification accuracy of the classifier. If the noise interference ability is stronger, the classification accuracy is lower. On the contrary, the lower the noise interference ability, the higher the classification accuracy will be guaranteed. In addition, it should be noted that the change trend of *AUC* value is basically consistent with that of accuracy. The experimental results show that the value of *AUC* with or without noise can be above 0.8.

Figure 5 shows the comparison of the accuracy of different regularization coefficients *C* without adding noise and the average accuracy after adding different noise.

Firstly, as a whole, with the regularization coefficient *C* decreasing gradually, the accuracy will decrease whether noise is added or not. The reason is that the regularization coefficient decreases due to the decision of the classifier itself, which leads to the overfitting of the classification model and reduces the classification effect. Locally, given the positive deterioration coefficient, that is, given a certain classifier, the classification result without noise is always better than the average case with added noise. From the experimental results, the classification accuracy without noise will be improved by 8.6% to 9.7%. This shows that the addition of noise does affect the classification effect of the classifier. However, it is noteworthy that even with the addition of noise, the mining of privacy data can still be achieved after choosing the appropriate positive deterioration

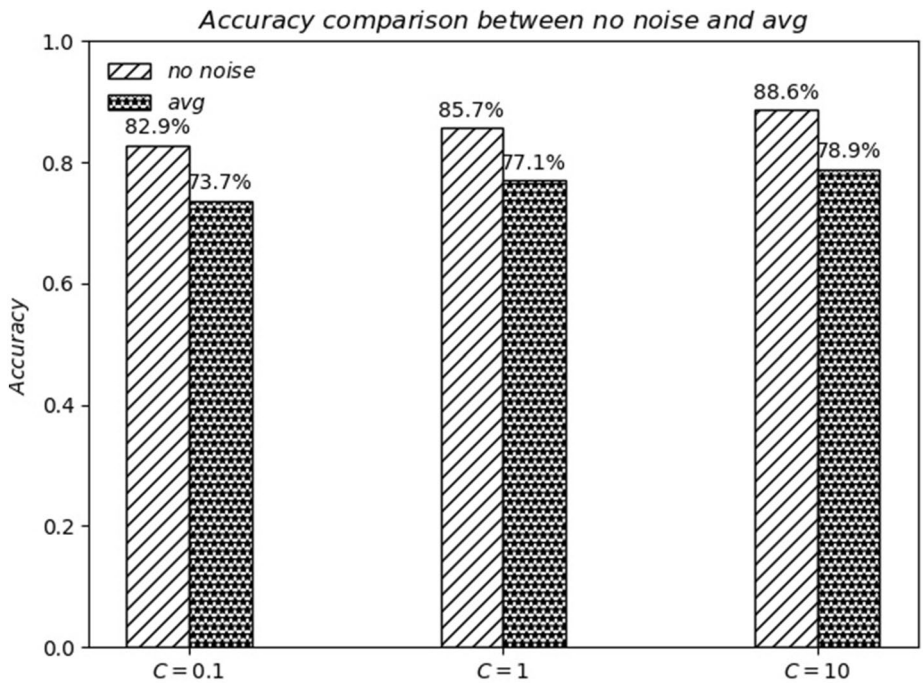

Fig. 5 Accuracy trends under different regularization coefficients and with or without added noise

coefficient. For example, in the case of $C = 10$ and adding noise, the average classification accuracy of privacy data can still reach about 80%.

Conclusion

In this paper, we combine privacy protection with data mining to implement a local privacy protection classification based on Human-centric computing. First, perform the necessary pre-processing on the original data. Then, in order to achieve disturbance of the source data, the protected data is formed by adding noise conforming to the Laplace mechanism to the privacy data. Secondly, in order to highlight the influence of the noisy data and also to better classify the generalization ability, the random forest strategy is adopted for feature selection. Then use the classic logistic regression model in machine learning to implement the classification task. Finally, the experiment shows that after selecting the appropriate regularization coefficient, the accuracy of the noise-added data can reach 85.7%.

Considering the actual situation, the deficiency of this paper is that the number of patients predicted as health controllable is relatively large and the overall prediction accuracy is not high. The next step is to improve the degree of mining privacy data without affecting the machine learning model as much as possible.

Abbreviations

AI: artificial intelligence; RF: random forest; CART: classified and regression decision tree; ACC: accuracy; TPR: true positive rate; FPR: false positive rate; ROC: receiver operating characteristic curve; AUC: area under the curve; TP: true positive; FN: false negative; TN: true negative; FP: false positive.

Acknowledgements
It was supported by the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD), Postgraduate Research & Practice Innovation Program of Jiangsu Province (KYCX18_1032).

Authors' contributions
CY conceptualized the study. BZ performed all experiments and wrote the manuscript. ZY analyzed all the data. JW advised on the manuscript preparation and technical knowledge. All authors read and approved the final manuscript.

Funding
This work was supported by the National Natural Science Foundation of China (61772282, 61772454, 61811530332).

Availability of data and materials
We declared that materials described in the manuscript will be freely available to any scientist wishing to use them for non-commercial purposes.

Competing interests
The authors declare that they have no competing interests.

Author details
[1] School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing 210044, China. [2] College of Information Science and Technology, Nanjing Forestry University, Nanjing 210037, China. [3] School of Computer & Communication Engineering, Changsha University of Science and Technology, Changsha 410004, China.

Received: 9 June 2019 Accepted: 26 August 2019
Published online: 04 September 2019

References

1. Zhaojun L, Gang Q, Zhenglin L (2019) A survey on recent advances in vehicular network security, trust, and privacy. IEEE Trans Intell Transp Syst 20(2):760–776
2. Xiong L, Jian N, Bhuiyan MZA (2018) A robust ECC-based provable secure authentication protocol with privacy preserving for industrial internet of things. IEEE Trans Ind Inf 14(8):3599–3609
3. Sangaiah AK, Medhane DV, Han T, Hossain MS, Muhammad G (2019) Enforcing position-based confidentiality with machine learning paradigm through mobile edge computing in real-time industrial informatics. IEEE Trans Ind Inf 15(7):4189–4196
4. Perez AJ, Zeadally S, Jabeur N (2018) Security and privacy in ubiquitous sensor networks. J Inf Process Syst 14(2):286–308
5. Kang WM, Moon SY, Park JH (2017) An enhanced security framework for home appliances in smart home. Human Centric Comput Inf Sci 7(1):6
6. Kim HW, Jeong YS (2018) Secure authentication-management human-centric scheme for trusting personal resource information on mobile cloud computing with blockchain. Human Centric Comput Inf Sci 8(1):11
7. Sweeney L (2002) K-anonymity: a model forprotecting privacy. Int J Uncertainty Fuzziness Knowl Based Syst 10(5):557–570
8. Sweeney L (2002) Achieving k-anonymity privacy protection using generalization, and suppression. Int J Uncertainty Fuzziness Knowl Based Syst 10(5):571–588
9. Machanavajjhala A., Gehrke J., Kifer D., and Venkitasubramaniam M. (2006) L-diversity: privacy beyond k-anonymity. In: 22nd international conference on data engineering. IEEE Computer Society
10. Dwork C (2006) Differential privacy. International colloquium on automata, languages, and programming. Springer, Berlin, pp 1–12
11. Jian S, Chen W, Tong L (2018) Secure data uploading scheme for a smart home system. Inf Sci 453:186–197
12. Zhou SW, He Y, Xiang SZ, Li KQ, Liu YH (2019) Region-based compressive networked storage with lazy encoding. IEEE Trans Parallel Distrib Syst 30(6):1390–1402
13. Hu P, Dhelim S, Ning H (2017) Survey on fog computing: architecture, key technologies, applications and open issues. J Netw Comput Appl 98:27–42
14. He S, Li Z, Tang Y, Liao Z, Wang J, Kim HJ (2019) Parameters compressing in deep learning. Comput Mater Con (accepted)
15. Zhang J, Jin X, Sun J, Wang J, Sangaiah AK (2018) Spatial and semantic convolutional features for robust visual object tracking. Multimedia Tools Appl. https://doi.org/10.1007/s11042-018-6562-8
16. Zhang J, Jin X, Sun J, Wang J, Li K (2019) Dual model learning combined with multiple feature selection for accurate visual tracking. IEEE Access 7:43956–43969
17. Zhang J, Wu Y, Feng W, Wang J (2019) Spatially attentive visual tracking using multi-model adaptive response fusion. IEEE Access 7:83873–83887
18. Zhang J, Wang W, Lu C, Wang J, Sangaiah AK (2019) Lightweight deep network for traffic sign classification. Ann Telecommun. https://doi.org/10.1007/s12243-019-00731-9
19. Zhang J, Lu C, Li X, Kim H-J, Wang J (2019) A full convolutional network based on DenseNet for remote sensing scene classification. Math Biosci Eng 16(5):3345–3367
20. Yin Y, Chen L, Xu Y, Wan J, Zhang H, Mai Z (2019) QoS prediction for service recommendation with deep feature learning in edge computing environment. Mobile Netw Appl. https://doi.org/10.1007/s11036-019-01241-7
21. Yin Y, Chen L, Xu Y, Wan J (2018) Location-aware service recommendation with enhanced probabilistic matrix factorization. IEEE Access 6:62815–62825

22. Yin Y, Xu Y, Xu W, Gao M, Yu L, Pei Y (2017) Collaborative service selection via ensemble learning in mixed mobile network environments. Entropy 19(7):358
23. Yin Y, Xu W, Xu Y, Li H, Yu L (2017) Collaborative QoS prediction for mobile service with data filtering and SlopeOne model. Mobile Inf Syst 2017:1–14
24. Rabaey JM (2018) Towards true human-centric computation. Commun Comput 131:73–76
25. Heng Z, Wenchao M, Jun Q (2019) Distributed load sharing under false data injection attack in an inverter-based microgrid. IEEE Trans Ind Electron 66(2):1543–1551
26. Guo Kehua, He Yan, Kui Xiaoyan, Sehdev Paramjit, Chi Tao, Zhang Ruifang, Li Jialun (2018) LLTO: towards efficient lesion localization based on template occlusion strategy in intelligent diagnosis. Pattern Recogn Lett 116:225–232
27. Guo Kehua, Li Ting, Huang Runhe, Kang Jian, Chi Tao (2018) DDA: a deep neural network-based cognitive system for IoT-aided dermatosis discrimination. Ad Hoc Netw 80:95–103
28. Ozatay M, Verma N (2019) Exploiting emerging sensing technologies toward structure in data for enhancing perception in Human-Centric applications. IEEE Internet Things J 6(2):3411–3422
29. Gergely A, Luca M, Castelluccia C (2019) Differentially private mixture of generative neural networks. IEEE Trans Knowl Data Eng 31(6):1109–1121
30. Yin C, Xi J, Sun R, Wang J (2018) Location privacy protection based on differential privacy strategy for big data in industrial internet-of-things. IEEE Trans Ind Inf 14(8):3628–3636
31. Vaidya J, Shafiq B, Basu A, Hong Y (2013) Differentially private Naive Bayes classification. In: Proceedings of the 2013 IEEE/WIC/ACM international joint conferences on web intelligence (WI) and intelligent agent technologies (IAT)
32. Mohammed N (2011) Differentially private data release for data mining. In: ACM SIGKDD international conference on knowledge discovery and data mining ACM; 2011. pp 493–501
33. Tang DY, Zhou SW, Yang WJ (2019) Random-filtering based sparse representation parallel face recognition. Multimedia Tools Appl 78(2):1419–1439
34. Yin C, Ding S, Wang J (2019) Mobile marketing recommendation method based on user location feedback. Human Centric Comput Inf Sci 9(1):14–31
35. Yin C, Shi L, Sun R, Wang J (2019) Improved collaborative filtering recommendation algorithm based on differential privacy protection. J Supercomput 7:1–14
36. Jia Q, Guo L, Jin Z (2018) Preserving model privacy for machine learning in distributed systems. IEEE Trans Parallel Distrib Syst 29(8):1808–1822
37. Wang J, Gao Y, Liu W, Wu W, Lim S (2019) An asynchronous clustering and mobile data gathering schema based on timer mechanism in wireless sensor networks. Comput Mater Continua 58(3):711–725
38. Wang J, Gao Y, Liu W, Sangaiah AK, Kim H-J (2019) Energy efficient routing algorithm with mobile sink support for wireless sensor networks. Sensors 19(7):1494
39. Pan J-S, Kong L, Sung T-W, Tsai P-W, Snasel V (2018) Alpha-fraction first strategy for hierarchical wireless sensor networks. J Internet Technol 19(6):1717–1726
40. Pan J-S, Lee C-Y, Sghaier A, Zeghid M, Xie J (2019) Novel systolization of subquadratic space complexity multipliers based on Toeplitz matrix-vector product approach. IEEE Trans Very Large Scale Integr Syst 27(7):1614–1622
41. He S, Xie K, Xie K, Xu C, Wang J (2019) Interference-aware multi-source transmission in multi-radio and multi-channel wireless network. IEEE Syst J. https://doi.org/10.1109/JSYST.2019.2910409
42. He S, Xie K, Chen W, Zhang D, Wen J (2018) Energy-aware routing for SWIPT in multi-hop energy-constrained wireless network. IEEE Access 6:17996–18008
43. He S, Zeng W, Xie K, Yang H, Lai M, Su X (2017) PPNC: privacy preserving scheme for random linear network coding in smart grid. KSII Trans Internet Inf Syst 11(3):1510–1533
44. Wang J, Gao Y, Yin X, Li F, Kim HJ (2018) An enhanced PEGASIS algorithm with mobile sink support for wireless sensor networks. Wire Commun Mobile Comput. https://doi.org/10.1155/2018/9472075
45. Nguyen T-T, Pan J-S, Dao T-K (2019) An improved flower pollination algorithm for optimizing layouts of nodes in wireless sensor network. IEEE Access. https://doi.org/10.1109/ACCESS.2019.2921721
46. Meng Z, Pan J-S, Tseng K-K (2019) PaDE: an enhanced differential Evolution algorithm with novel control parameter adaptstion schemes for numerical optimization. Knowl Based Syst 168:80–99
47. Pan J-S, Kong L, Sung T-W, Tsai P-W, Snáel V (2018) A clustering scheme for wireless sensor networks based on genetic algorithm and dominating set. J Internet Technol 19(4):1111–1118
48. Gao Z, Sun Y, Cui X (2018) Privacy-preserving hybrid K-means. Int J Data Warehouse Min 14(2):1–17
49. Xing K, Hu C, Yu J (2017) Mutual privacy preserving k-means clustering in social participat-ory sensing. IEEE Trans Ind Inf 13(4):2066–2076
50. Gong M, Pan K, Xie Y (2019) Differential privacy preservation in regression analysis based on relevance. Knowl Based Syst 173:140–149
51. Wang J, Gao Y, Liu W, Sangaiah AK, Kim HJ (2019) An improved routing schema with special clustering using PSO algorithm for heterogeneous wireless sensor Network. Sensors 19(3):671–688
52. Wang J, Ju C, Gao Y, Sangaiah AK, Kim G-J (2018) A PSO based energy efficient coverage control algorithm for wireless sensor networks. Comput Mater Continua 56:433–446
53. Chen W, Xie X, Wang J (2017) A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. CATENA 151:147–160

Publisher's Note