2014 International Joint Conference on Neural Networks (IJCNN)
July 6-11, 2014, Beijing, China

Differentially Private Feature Selection

Jun Yang
College of Computer Science
Nanjing University of Posts and Telecommunications
Nanjing, China

Yun Li
College of Computer Science
Nanjing University of Posts and Telecommunications
Nanjing, China
Email: liyun@njupt.edu.cn

***Abstract*—The privacy-preserving data analysis has been gained significant interest across several research communities. The current researches mainly focus on privacy-preserving classification and regression. However, feature selection is also an essential component for data analysis, which can be used to reduce the data dimensionality and can be utilized to discover knowledge, such as inherent variables in data. In this paper, in order to efficiently mine sensitive data, a privacy preserving feature selection algorithm is proposed and analyzed in theory based on local learning and differential privacy. We also conduct some experiments on benchmark data sets. The Experimental results show that our algorithm can preserve the data privacy to some extent.**

I. Introduction

Privacy is the sensitive information that data owner reluctant to disclose, which has been a growing concern in medical records, financial records, web search histories and social network data. Thus, an emerging challenge for machine learning and data mining is how to learn from these data sets without privacy leak. The current researches mainly focus on privacy-preserving classification and regression [1]. However, feature selection is also one of the key problems in machine learning and data mining [2], [3]. Feature selection brings the immediate effects of speeding up a machine learning or data mining algorithm, improving learning accuracy, and enhancing model comprehensibility. Various studies show that features can be removed without performance deterioration [4]. Moreover, feature selection also leads to better data visualization, reduction of measurement and storage requirements. So the feature selection with privacy preserving is a very important issue and need to be deeply addressed. Before introducing the concrete privacy preserving feature selection algorithm, we will simply discuss the common properties of feature selection and basic knowledge for privacy preserving.

Roughly speaking, a feature selection algorithm is usually associated with two important aspects: search strategy and evaluation criterion. According to the criterion, algorithms can be categorized into filter model, wrapper model and embedded model [2], [3]. In the wrapper model, the selection method tries to directly optimize the performance of a specific predictor (algorithm). The main drawback of this method is its computational deficiency. The embedded model incorporates feature selection as a part of the training process, and features' utility is obtained based on analyzing their utility for optimizing the objective function of the learning model. Compared to wrapper and embedded models, feature selection algorithms with filter model are independent to any learning model, therefore do not have bias on specific learner models, which is believed to be one advantage of the filter model. Another advantage of the filter model is that it has very simple structure, which usually contains a straightforward search strategy, such as backward elimination or forward selection, and a feature evaluation criterion designed according to certain criteria. The benefits of the simple structure are two folds. First, it is easy to design, and after being implemented, it is also easy to understand for other researchers. This explains that why most feature selection algorithms are of filter model. Second, since its structure is simple, it is usually very fast [4]. On the other hand, if the categorization is based on output characteristics, feature selection algorithms can be divided into either feature weighting/ranking algorithms or subset selection algorithms [4]. A comprehensive survey of existing feature selection techniques and a general framework for their unification can be found in [2], [3], [4]. In this paper, we focus on feature weighting algorithm, and it belongs to filter model.

For the privacy preserving, according to the processing stages of information flowing, the methods of privacy preserving are divided into four categories, namely, input perturbation, transformed data release, query auditing and query answer perturbation, access control [5]. In our case, the proposed privacy preserving feature selection method belongs to query answer perturbation. For privacy measure, we adopt ε-differential privacy model [6], which is a measure of quantifying the privacy-risk associated with computing functions of sensitive data. A statistical procedure satisfies ε-differential privacy if changing a single data point does not shift the output distribution by too much. Therefore, from the output of the algorithm, it is difficult to infer the value of any particular data point [1]. And ε-differential privacy model is robust to known attacks, such as those involving side information [7]. ε-differential privacy model is a strong, cryptographically-motivated definition of privacy that has recently received a significant amount of research attention, such as differentially private empirical risk minimization for classification [1].

In this paper, we apply the sensitivity-based method to design feature selection algorithm while guaranteeing ε-differential privacy. The original feature selection algorithm-FWELL is based on local learning, and logistic loss with L2-regularizer is adopted to design the evaluation criterion of feature selection. In differentially private feature selection method-Private FWELL, noise is added as output perturbation according to the sensitivity analysis of FWELL. We also give proof for its ε-differential privacy.

The paper is organized as follows, the feature weighting algorithm based on local learning FWELL is introduced in section 2. Section 3 presents and analyzes the differentially private feature selection algorithm Private FWELL. The exper-

978-1-4799-1484-5/14/$31.00 ©2014 IEEE

imental results on bench mark data sets are shown in section 4. The paper concludes in section 5.

II. Feature weighting algorithm based on local learning

For feature weighting, we are given a training sample set *D*, which contains *n* samples, $D = \{\mathbf{X}, \mathbf{Y}\} = \{\mathbf{x}_i, y_i\}_{i=1}^n$, where $\mathbf{x}_i$ is the input for the *i*-th training sample, and y_i is the label, and each sample is represented by a *d*-dimensional vector $\mathbf{x}_i = (x_{i1}, x_{i2}, \cdots, x_{id}) \in \mathbb{R}^d$.

Based on local learning, for sample $\mathbf{x}_i$, it should be close to the nearest neighbor sample with the same label to $\mathbf{x}_i$ (i.e., near hit sample $NH(\mathbf{x}_i)$) and away from the nearest neighbor sample with different class label (i.e., near miss sample $NM(\mathbf{x}_i)$) [8]. For the purposes of this paper, we use the Manhattan distance to find the nearest neighbors (i.e., $NH(\mathbf{x}_i)$ and $NM(\mathbf{x}_i)$) and to define their closeness, while other standard distance definitions may also be used. The logistic regression loss is adopted to model the fit of data for its simplicity and effectiveness. In addition, the logistic loss is twice differentiable and strongly convex, which is good for faster optimizations [9]. Then for any sample $\mathbf{x}_i$, the logistic loss function is defined as follows,

$$\mathcal{L}(\mathbf{w}^T\mathbf{z}_i) = \log(1 + \exp(-\mathbf{w}^T\mathbf{z}_i)) \tag{1}$$

In the Eqn. (1), T is the transpose, $\mathbf{w}$ is the feature weight vector, $\mathbf{z}_i = |\mathbf{x}_i - NM(\mathbf{x}_i)| - |\mathbf{x}_i - NH(\mathbf{x}_i)|$ and $|.|$ is an element-wise absolute operator. $\mathbf{z}_i$ can be considered as the mapping point of $\mathbf{x}_i$. $\mathbf{w}^T\mathbf{z}_i$ is the local margin for $\mathbf{x}_i$, which belongs to hypothesis margin [10] and an intuitive interpretation of this margin is a measure of the proportion of the features in $\mathbf{x}_i$ that can be corrupted by noise (or how much $\mathbf{x}_i$ can move in the feature space) before $\mathbf{x}_i$ is being misclassified [8]. In other words, the feature weighting based on local learning is like to scale each feature, and thus obtain a weighted feature space parameterized by a vector $\mathbf{w}$, so that a local margin-based loss function in the induced feature space is minimized. Thus by the large margin theory [11], a classifier trained on weighted feature space that minimizes a margin-based loss function usually generalizes well on unseen test data.

Moreover, in order to prevent from overfitting, the regularization is always used. Thus, the evaluation criterion for feature weighting on the training data set *D* is defined as follows,

$$L(\mathbf{w}, D) = \frac{1}{n}\sum_{i=1}^{n}\mathcal{L}(\mathbf{w}^T\mathbf{z}_i) + \lambda\mathcal{R}(\mathbf{w}), \tag{2}$$

where λ is the cost parameter balancing the importance of the two terms, $\mathcal{R}(\mathbf{w})$ in (2) is a regularizing term. Then feature selection aims to find the target model $\mathbf{w}$, which minimizes the loss function in Eqn.(2). Then we obtain the feature selection algorithm based on local learning shown in Algorithm 1. Note that, as an example, the gradient descent algorithm is used to illustrate the minimization of evaluation function (2). Of course, the optimal feature weights can be found by many other optimization approaches.

In the following analysis and experiments, the L2 regularizer is used as $\mathcal{R}(\mathbf{w})$ in Eqn. (2) for its rotational invariance and strong stability property [12]. Then the concrete evaluation criterion considered in this paper is as follows,

$$L(\mathbf{w}, D) = \frac{1}{n}\sum_{i=1}^{n}\mathcal{L}(\mathbf{w}^T\mathbf{z}_i) + \lambda||\mathbf{w}||^2. \tag{3}$$

And the gradient descent algorithm is used to minimize the evaluation function (3) to obtain the feature weights as described in Algorithm 1.

Algorithm 1 Feature WEighting algorithm based on Local Learning-FWELL

Step 1. Input training data set $D = \{\mathbf{x}_i, y_i\}_{i=1}^n$, $\mathbf{x}_i \in \mathbb{R}^d$ and regularization parameter λ in Eqn. (2).
Step 2. Initialize $\mathbf{w} = (1, 1, \cdots, 1) \in \mathbb{R}^d$.
Step 3. For $i = 1, 2, \cdots, n$
(a) Given $\mathbf{x}_i$, find the $NH(\mathbf{x}_i)$ and $NM(\mathbf{x}_i)$.
(b) Based on Eqn. (1) to obtain $\mathcal{L}(\mathbf{w}^T\mathbf{z}_i)$
(c) $\nabla = \frac{1}{n}\frac{\partial\mathcal{L}(\mathbf{w}^T\mathbf{z}_i)}{\partial\mathbf{w}} + \lambda\frac{\partial\mathcal{R}(\mathbf{w})}{\partial\ \mathbf{w}}$.
(d) $\mathbf{w} = \mathbf{w} - \frac{\nabla}{||\nabla||_2}$.
Step 4. Output the feature weighting vector $\mathbf{w}$.

III. Differentially private feature selection

A. Sensitivity analysis

We like to propose a privacy preserving FWELL in terms of differential privacy definition in [6], then the *sensitivity* of the FWELL should be analyzed. So before introducing the privacy preserving feature selection method, we need a definition-*sensitivity* as follows [13], [14].

Definition 1: For any function A with *n* inputs, we define the *sensitivity* ΔQ as the maximum, over all inputs, of the difference in the value of A when one input of A is changed. That is,

$$\Delta Q = \max_{D,D'} \|A(D) - A(D')\| \tag{4}$$

Data sets D and D' differ by at most one element. Based on the Definition 1, we will give the Corollary 1 for the sensitivity analysis of Algorithm 1 with L2 regularizer.

Corollary 1 *The feature weighting algorithm described in Algorithm 1 with L2 regularizer has the sensitivity* $\frac{2}{\lambda n}$.

Proof. Let $D = \{(\mathbf{x}_1, y_1), \cdots, (\mathbf{x}_n, y_n)\}$ and $D' = \{(\mathbf{x}_1, y_1), \cdots, (\mathbf{x}_{n'}, y_{n'})\}$ be two data sets that differ in the value of the *n*-th individual. Suppose $\mathbf{w}_1$ and $\mathbf{w}_2$ are the solutions respectively to Algorithm 1 when data sets are D and D'.

$$\begin{cases} \mathbf{w}_1 = \arg\min_{\mathbf{w}} L(\mathbf{w}, D), \\ \mathbf{w}_2 = \arg\min_{\mathbf{w}} L(\mathbf{w}, D') \end{cases} \tag{5}$$

Then according to the Definition 1, the sensitivity of Algorithm 1 is the upper bound of $\|\mathbf{w}_1 - \mathbf{w}_2\|$.

We define a function $\ell(\mathbf{w})$ as

$$\ell(\mathbf{w}) = L(\mathbf{w}, D') - L(\mathbf{w}, D). \tag{6}$$

Because logistic loss and L2 regularizer are adopted, functions L and ℓ are continuous and differentiable. Since the $\mathbf{w}_1$ and $\mathbf{w}_2$ are minimizers of $L(\mathbf{w}, D)$ and $L(\mathbf{w}, D')$,

respectively, and they are obtained through gradient descent, then their gradients are equal to zero, i.e., $\nabla L(\mathbf{w}_1, D) = 0$ and $\nabla L(\mathbf{w}_2, D') = 0$. Based on Eqn. (6), we can obtain $\ell(\mathbf{w}_2) = L(\mathbf{w}_2, D') - L(\mathbf{w}_2, D)$, then

$$\nabla L(\mathbf{w}_2, D) + \nabla \ell(\mathbf{w}_2) = 0 \tag{7}$$

Since logistic loss and L2 regularizer are 1-strongly convex, then L satisfies λ-strongly convex [1]. We obtain

$$(\nabla L(\mathbf{w}_1, D) - \nabla L(\mathbf{w}_2, D))^T (\mathbf{w}_1 - \mathbf{w}_2) \geq \lambda \|\mathbf{w}_1 - \mathbf{w}_2\|^2 \tag{8}$$

while based on Eqn. (7) and $\nabla L(\mathbf{w}_1, D) = 0$,

$$\begin{aligned}(\nabla L(\mathbf{w}_1, D) - \nabla L(\mathbf{w}_2, D))^T (\mathbf{w}_1 - \mathbf{w}_2) \\ = (\mathbf{w}_1 - \mathbf{w}_2)(\nabla \ell(\mathbf{w}_2))^T\end{aligned} \tag{9}$$

According to Cauchy-Schwartz inequality, we obtain

$$\|\mathbf{w}_1 - \mathbf{w}_2\| \cdot \|\nabla \ell(\mathbf{w}_2)\| \geq (\mathbf{w}_1 - \mathbf{w}_2)(\nabla \ell(\mathbf{w}_2))^T \tag{10}$$

So, combining Eqns. (8), (9) and (10)

$$\|\mathbf{w}_1 - \mathbf{w}_2\| \cdot \|\nabla \ell(\mathbf{w}_2)\| \geq \lambda \|\mathbf{w}_1 - \mathbf{w}_2\|^2 \tag{11}$$

Namely,

$$\|\mathbf{w}_1 - \mathbf{w}_2\| \leq \frac{1}{\lambda} \|\nabla \ell(\mathbf{w}_2)\| \tag{12}$$

Suppose $\mathbf{z}_n$ and $\mathbf{z}_{n'}$ only depends on $\mathbf{x}_n$ and $\mathbf{x}_{n'}$ respectively, and based on Eqns. (3) and (6), for any $\mathbf{w}$, we can approximately obtain

$$\ell(\mathbf{w}) = \frac{1}{n}(\mathcal{L}(\mathbf{w}^T \mathbf{z}_{n'}) - \mathcal{L}(\mathbf{w}^T \mathbf{z}_n)) \tag{13}$$

According to Eqn. (1), for any point $\mathbf{z}$

$$\mathcal{L}(\mathbf{w}^T \mathbf{z}) = \log(1 + \exp(-\mathbf{w}^T \mathbf{z})) \tag{14}$$

Then for the normalized $\|\mathbf{z}\| \leq 1$

$$\begin{aligned}\left\|\nabla \mathcal{L}(\mathbf{w}^T \mathbf{z})\right\| &= \left\| \frac{-\mathbf{z}}{1 + \exp(\mathbf{w}^T \mathbf{z})} \right\| \\ &\leq \left\| \frac{1}{1 + \exp(\mathbf{w}^T \mathbf{z})} \right\| \|-\mathbf{z}\| \\ &\leq 1\end{aligned} \tag{15}$$

Based on Eqns. (13) and (15), we can achieve

$$\begin{aligned}\|\nabla \ell(\mathbf{w})\| &= \frac{1}{n} \left\|\nabla \mathcal{L}(\mathbf{w}^T \mathbf{z}_{n'}) - \nabla \mathcal{L}(\mathbf{w}^T \mathbf{z}_n)\right\| \\ &\leq \frac{1}{n}(\left\|\nabla \mathcal{L}(\mathbf{w}^T \mathbf{z}_{n'})\right\| + \left\|\nabla \mathcal{L}(\mathbf{w}^T \mathbf{z}_n)\right\|) \\ &\leq \frac{2}{n}\end{aligned} \tag{16}$$

So $\|\nabla \ell(\mathbf{w_2})\|$ is less than $\frac{2}{n}$. Based on Eqn. (12), we can obtain

$$\|\mathbf{w}_1 - \mathbf{w}_2\| \leq \frac{2}{\lambda n} \tag{17}$$

■

B. Differential privacy feature selection

Now we turn to propose the corresponding differentially private feature selection for Algorithm 1 with L2 regularizer. For our privacy measure, we use a differential privacy definition [6] to quantify the privacy-risk associated with feature selection evaluation functions for sensitive data.

Definition 2: A randomized mechanism A provides ε-*differential privacy*, if, for all data sets D and D' which differ by at most one element, and for all output subsets $S \subseteq Range(A)$,

$$Pr[A(D) \in S] \leq \exp(\varepsilon) \times Pr[A(D') \in S] \tag{18}$$

Then the algorithm is called satisfying differential privacy. The probability Pr is taken over the coin tosses of A, and $Range(A)$ denotes the output range of A. The privacy parameter ε measures the disclosure.

Our proposed differential privacy feature selection algorithm-Private FWELL is described in Algorithm 2.

Algorithm 2 Differentially Private Feature Selection-Private FWELL

Step 1. Input data set $D = \{\mathbf{x}_i, y_i\}_{i=1}^n$, regularization parameter λ in Eqn. (2), privacy parameter ε and parameter a.

Step 2. Obtain $\mathbf{w}$ according to Algorithm 1.

Step 3. Draw a random noise vector $\mathbf{b}$ with density $v(\mathbf{b})$ based on the sensitivity analysis in Corollary 1.

$$v(\mathbf{b}) = \frac{1}{a} e^{-\frac{n\varepsilon\lambda}{2}\|\mathbf{b}\|}.$$

Step 4. Compute $\mathbf{w}' = \mathbf{w} + \mathbf{b}$.

Step 5. Output $\mathbf{w}'$.

For Private FWELL, the following Theorem 1 is obtained.

Theorem 1 *Private FWELL is ε-differential privacy.*

Proof: We know that D and D' are any two data sets that differ in one individual. For $\mathbf{w}'$ derived from Algorithm 2 on D and D', we obtain $\mathbf{w}' = \mathbf{w}_1 + \mathbf{b}_1$ and $\mathbf{w}' = \mathbf{w}_2 + \mathbf{b}_2$ where $\mathbf{w}_1$ and $\mathbf{w}_2$ are unique outputs of Algorithm 1 on data sets D and D' respectively, $\mathbf{b}_1$ and $\mathbf{b}_2$ are the corresponding noise vectors in Algorithm 2.

$$\frac{v(\mathbf{w}'|D)}{v(\mathbf{w}'|D')} = \frac{v(\mathbf{b}_1)}{v(\mathbf{b}_2)} = e^{\frac{n\varepsilon\lambda}{2}(\|\mathbf{b}_2\| - \|\mathbf{b}_1\|)} \tag{19}$$

where $v(\mathbf{w}'|D)(v(\mathbf{w}'|D'))$ is the probability density of the output of Algorithm 2 at $\mathbf{w}'$ when the input is data set D (D'). Since $\mathbf{b}_1 - \mathbf{b}_2 = \mathbf{w}_2 - \mathbf{w}_1$, we obtain following equation using a triangle inequality,

$$\|\mathbf{b}_2\| - \|\mathbf{b}_1\| \leq \|\mathbf{b}_2 - \mathbf{b}_1\| = \|\mathbf{w}_1 - \mathbf{w}_2\| \tag{20}$$

Combining Eqns. (17), (19) and (20),

$$\frac{v(\mathbf{w}'|D)}{v(\mathbf{w}'|D')} \leq e^{\varepsilon} \tag{21}$$

■

Therefore, our algorithm-Private FWELL is ε-differential privacy in terms of Definition 2.

IV. EXPERIMENTS

In this section, we will give some experimental results for our proposed differentially private feature selection algorithm. The experiments consists of two parts: validate the effect of privacy parameter ε and the classification performance for different selected feature subsets under a given privacy degree. The classifiers used in the experiment are linear support vector machine (SVM) with C=1 and the 3-nearest neighbor classifier ($3NN$). Classification accuracy was assessed using a 10-fold cross-validation. For each fold, the proposed FWELL and Private FWELL algorithm were applied to the training part of the data to obtain the feature weighting result, and the features are ranked in descending order. Then different numbers of important features are selected with top ranks one by one to create classifiers. For all experiments, the parameter λ in our proposed methods are tuned by cross-validation. Six bench-mark data sets from UCI repository (http://archive.ics.uci.edu/ml/datasets.html) are used in our experiments. The data sets are summarized in following Table.

TABLE I. DESCRIPTION OF EXPERIMENTAL DATA SETS

Data sets	Samples	Features	Classes
$BASEHOCK$	1993	4863	2
$Soybean$	307	34	2
$Sonar$	208	60	2
$Waveform$	3343	20	2
$Lung$	203	12600	2
$Wdbc$	569	30	2

A. Experiments for parameter ε

For the first part of our experiment, we study the tradeoff between the privacy preserving degree and the classification accuracy. The number of selected features for training classifier is 10 percent of original feature dimensions. The privacy degree for Private FWELL is quantified by the value of ε. The increasing of ε implies a higher change in the belief of adversary when one entry in D changes, and thus lower privacy preserving. The experimental results are shown in Fig.1-6 for six data sets using 3NN classifier, and Fig.7-12 for six data sets using SVM. The X-axis is the log value of selected privacy parameter ε and the Y-axis is the classification accuracy using different classifiers.

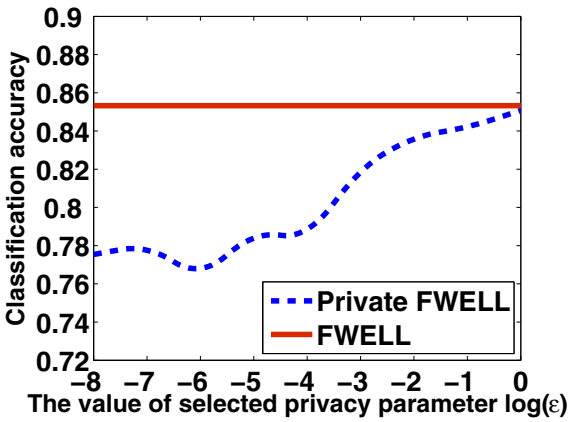

Fig. 1. Experimental results on 3NN classifier for BASEHOCK

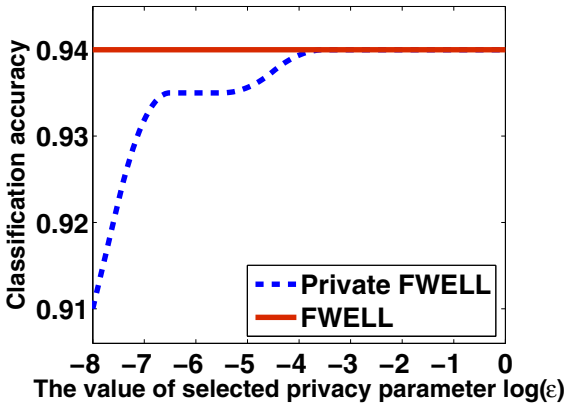

Fig. 2. Experimental results on 3NN classifier for Lung

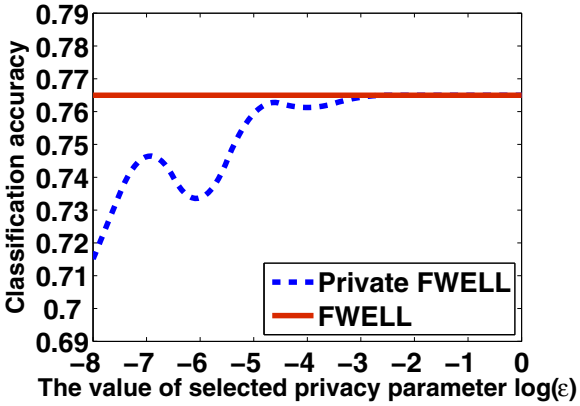

Fig. 3. Experimental results on 3NN classifier for Sonar

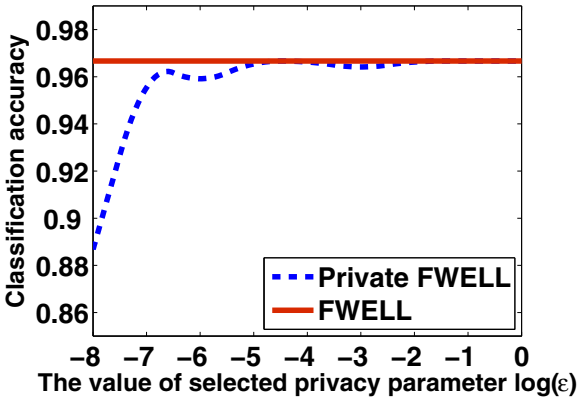

Fig. 4. Experimental results on 3NN classifier for Soybean

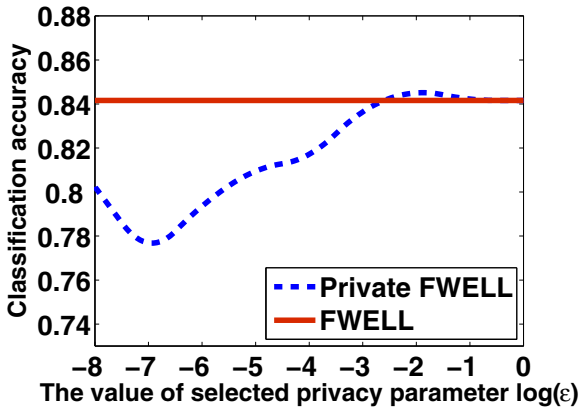

Fig. 5. Experimental results on 3NN classifier for Waveform

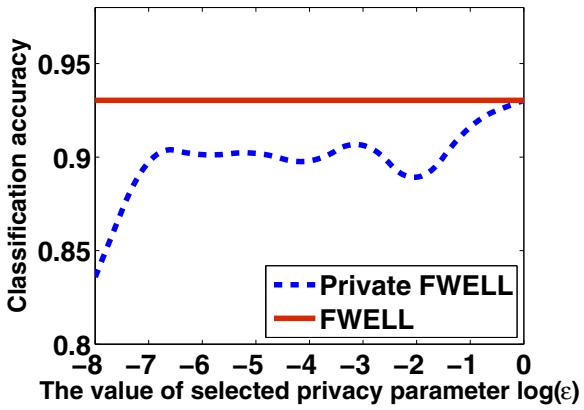

Fig. 6. Experimental results on 3NN classifier for WDBC

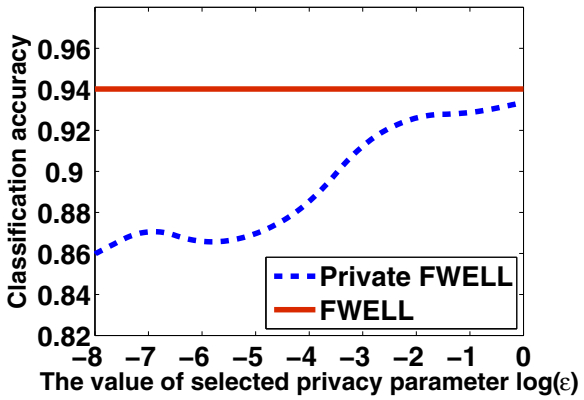

Fig. 7. Experimental results on SVM classifier for BASEHOCK

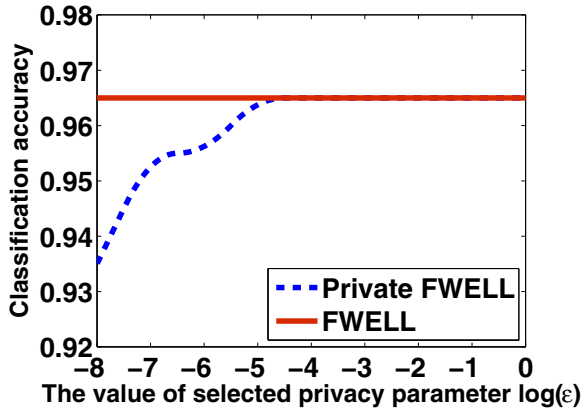

Fig. 8. Experimental results on SVM classifier for Lung

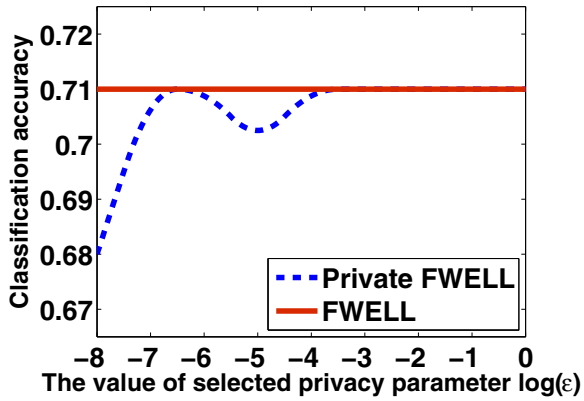

Fig. 9. Experimental results on SVM classifier for Sonar

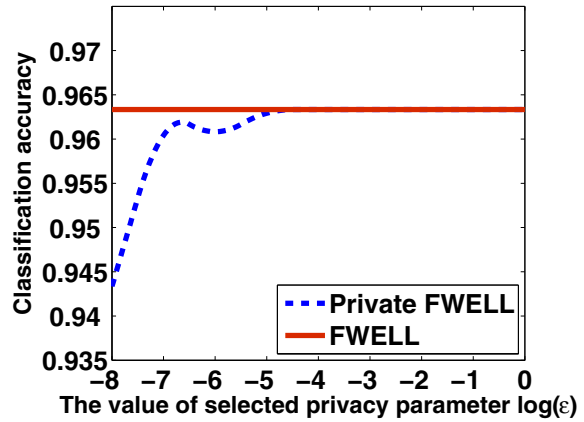

Fig. 10. Experimental results on SVM classifier for Soybean

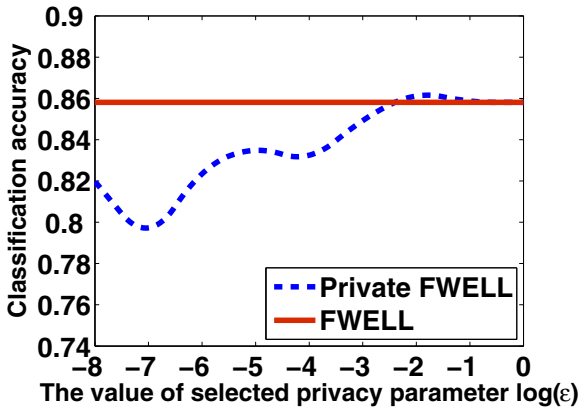

Fig. 11. Experimental results on SVM classifier for Waveform

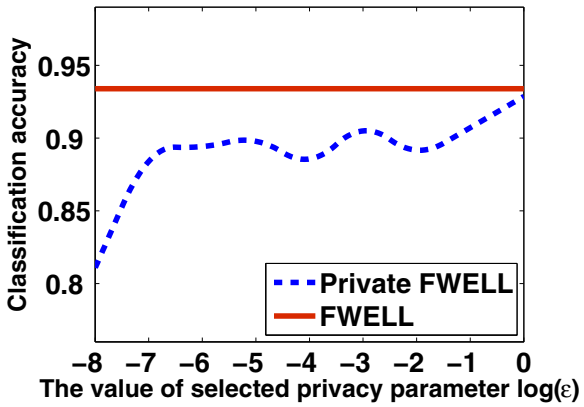

Fig. 12. Experimental results on SVM classifier for WDBC

From the results, we can observe that in the case of no privacy-preserving, i.e., $\varepsilon = 1$, the Private FWELL consistently obtains similar classification accuracy to the FWELL's. The performance of Private FWELL will drop along with the decline of ε, i.e., the data privacy preserving degree is increasing. When $\varepsilon = 0.00000001$, the extent of privacy preserving is pretty high, then the value of classification accuracy for Private FWELL is very low. These results are consistent with our intuition.

B. Experiments for differentially private feature selection

For the second part of our experiment, we study the classification accuracy when the classifiers are trained with different numbers of selected features. And the privacy parameter for Private FWELL is constant, i.e., $\varepsilon = 0.01$. The experimental results for FWELL and Private FWELL on six data sets are shown in Fig.13-24. The X-axis is the number of selected features and the Y-axis is the classification accuracy using different classifiers.

From the results, we can observe that the performance of privacy preserving feature selection-Private FWELL is very close to the non-privacy preserving one-FWELL in most cases. Then our proposed differentially private feature selection algorithm can obtain approximate classification performance to non-privacy preserving feature selection under the constraint

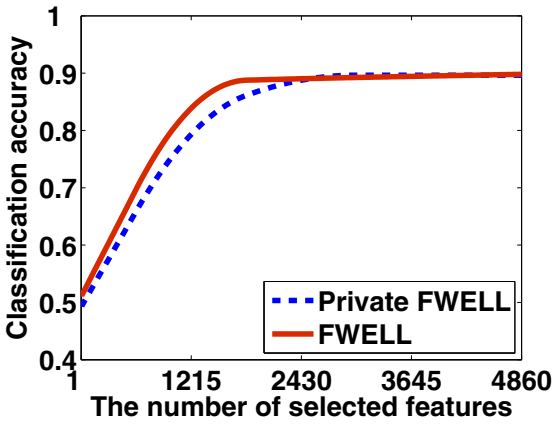

Fig. 13. Experimental results on 3NN classifier for BASEHOCK

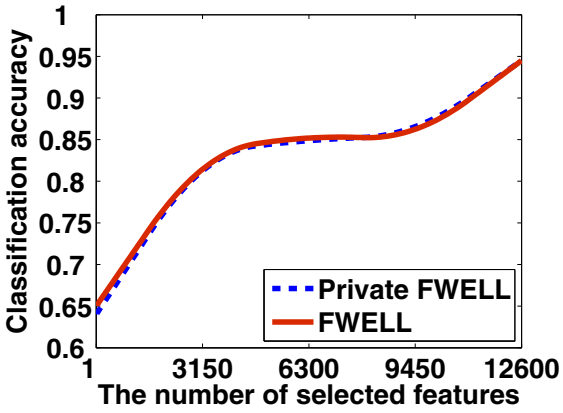

Fig. 14. Experimental results on 3NN classifier for Lung

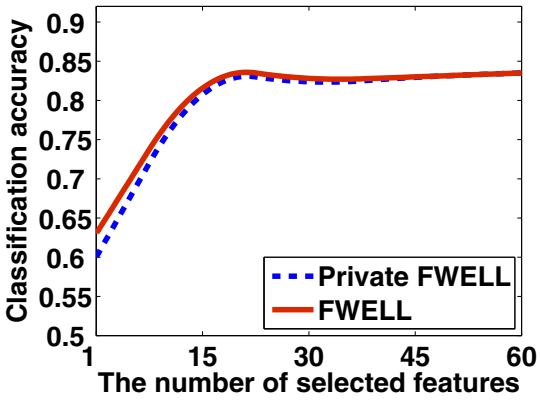

Fig. 15. Experimental results on 3NN classifier for Sonar

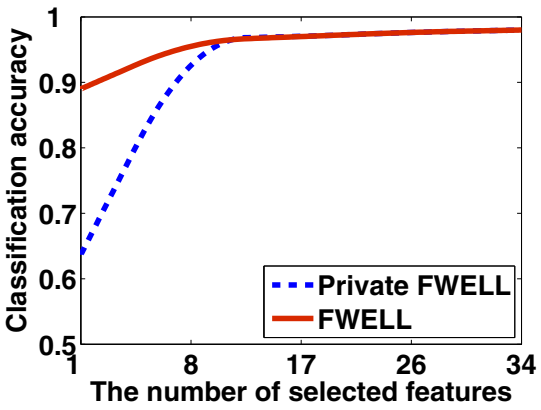

Fig. 16. Experimental results on 3NN classifier for Soybean

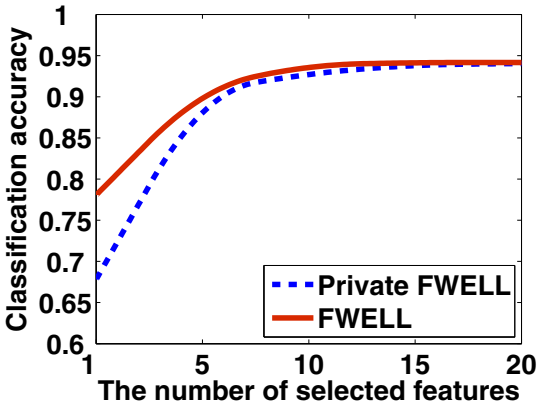

Fig. 17. Experimental results on 3NN classifier for Waveform

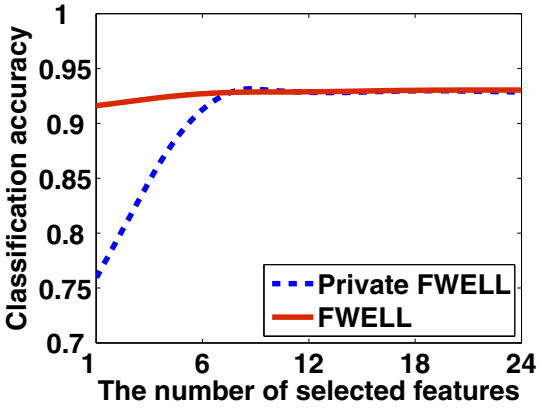

Fig. 18. Experimental results on 3NN classifier for WDBC

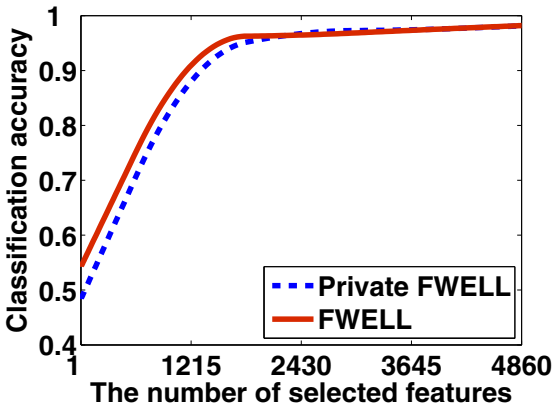

Fig. 19. Experimental results on SVM classifier for BASEHOCK

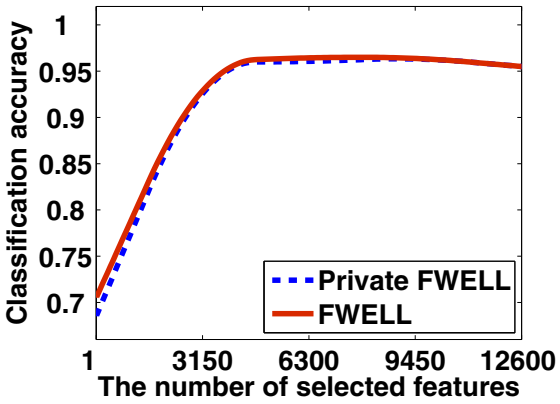

Fig. 20. Experimental results on SVM classifier for Lung

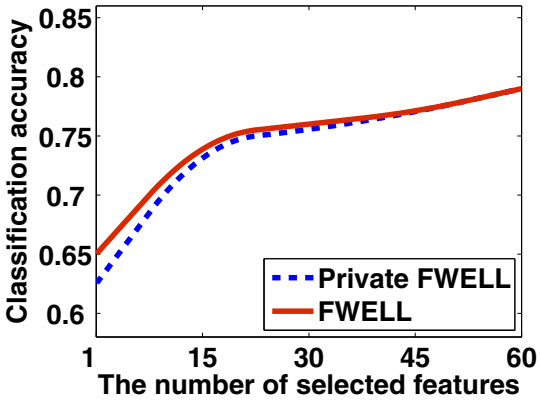

Fig. 21. Experimental results on SVM classifier for Sonar

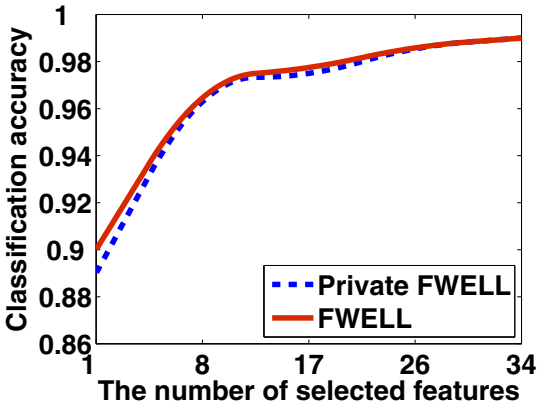

Fig. 22. Experimental results on SVM classifier for Soybean

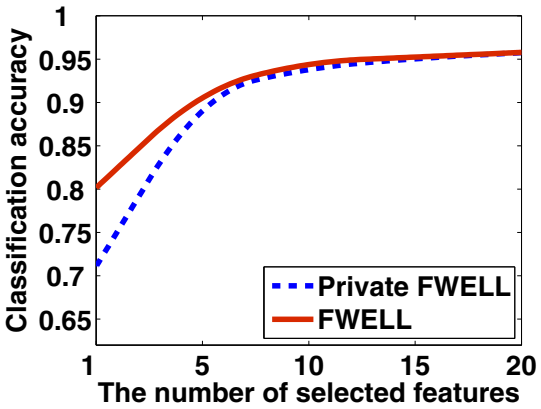

Fig. 23. Experimental results on SVM classifier for Waveform

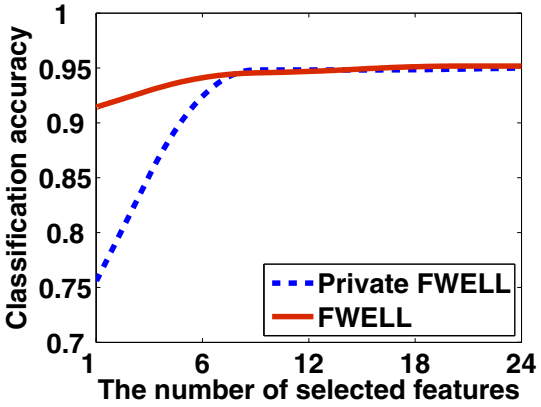

Fig. 24. Experimental results on SVM classifier for WDBC

of privacy, such as with $\varepsilon = 0.01$. And the proposed algorithm-Private FWELL is effective and efficient.

V. Conclusions

In this paper, we study the problem of privacy preserving feature selection, and a corresponding feature selection algorithm based on local learning and differential privacy is proposed and analyzed in theory. We also conduct many experiments to validate its performance on different benchmark data sets and classifiers. In the experiments, the results for different privacy preserving degree ε and different numbers of selected features under privacy constraint are shown. Our experiments as well as theoretical results indicate that in general, the proposed algorithm-Private FWELL can obtain high performance under some privacy constraint, and it can preserve data privacy to some extent.

Acknowledgment

This work was supported by a NSFC 61073114, 61300165, 61300164 and NSF of Jiangsu BK20131378

References

[1] K. Chaudhuri, C. Monteleoni and A. D. Sarwate, "Differentially Private Empirical Risk Minimization". *Journal of Machine Learning Research*, vol. 12, pp. 1069-1109, 2011.

[2] H. Liu and L. Yu, "Toward Integrating Feature Selection Algorithms for Classification and Clustering". *IEEE Transactions on Knowledge and Data Engineering*, vol. 17,no. 4, pp. 491-502, 2005.

[3] M. Dash and H. Liu, *Feature Selection for Classification.* Intelligent Data Analysis, pp. 131-156, The IOS Press, 1997.

[4] Z. Zhao. *Spectral feature selection for mining ultrahigh dimensional data*. PhD Dissertation, Arizona State University, 2010.

[5] M. Hay, "Enabling accurate analysis of private network data". Ph.D. Dissertation, 2010.

[6] C. Dwork, "Differential privacy," *International Colloquium on Automata, Languages and Programming*, pp. 1-12, 2006.

[7] S. R. Ganta, S. P. Kasiviswanathan and A. Smith. "Composition attacks and auxiliary information in data privacy". *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pp. 265-273, 2008.

[8] Y. J. Sun, S. Todorovic and S. Goodison. "Local learning based feature selection for high dimensional data analysis". *IEEE Trans. Pattern Analysis and Machine Intelligence*, pp. 1610-1626, 2010.

[9] M. K. Tan, I. W. Tsang, and L. Wang, "Minimax sparse logistic regression for very high dimensional feature selection". *IEEE Transactions on Neural Networks and Learning Systems*, vol. 24, pp. 1609-1622, 2013.

[10] K. Crammer, R. G. Bachrach, A. Navot and N. Tishby. "Margin analysis of the lvq algorithm". *Advances in Neural Information Processing Systems*, La Jolla CA, 2002.

[11] R. E. Schapire, Y. Freud, P. Bartlett and W. S. Lee. "Boosting the margin: A new explanation for the effectiveness of voting methods". *The Annals of Statistics*, vol.26, pp. 1651-1686, 1998.

[12] A. Y. Ng. "Feature selection, l1 vs. l2 regularization, and rotational invariance". *Proceedings of International Conference on Machine Learning*, Banff, Canada, 2004

[13] C. Dwork, F. McSherry, K. Nissim and A. Smith, "Calibrating noise to sensitivity in private data analysis". *Proceedings of the Third conference on Theory of Cryptography* , pp. 265-284, 2006.

[14] K. Chaudhuri and C. Monteleoni, "Privacy-preserving logistic regression". *The Twenty-Second Annual Conference on Neural Information Processing Systems*, pp. 289-296, 2008.