Knowledge-Based Systems 000 (2017) 1–11

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Private classification with limited labeled data

Xiaoqian Liu[a], Qianmu Li[a,*], Tao Li[b,c,*]

[a] School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China
[b] Jiangsu BDSIP Key Lab, School of Computer Science, Nanjing University of Posts and Telecommunications, Nanjing 210023, China
[c] School of Computing and Information Sciences, Florida International University, Miami, FL 33199, United States

ARTICLE INFO

Article history:
Received 20 January 2017
Revised 28 June 2017
Accepted 2 July 2017
Available online xxx

Keywords:
Unlabeled data
TSVM
Differential privacy
Exponential mechanism

ABSTRACT

Differentially private Support Vector Machines (SVMs) have been extensively studied in recent years. Most design mechanisms are focused on perturbing the solution to a decent convex optimization problem under the theory of Empirical Risk Minimization (ERM). To preserve the accuracy, a large number of labeled data is needed for training the model. However, in most cases, the labeled data is limited. Constructing private SVMs in such cases often suffers from low accuracy. The situation gets worse if the given privacy budget is small. In this paper, we make use of Transductive Support Vector Machines (TSVMs) to learn from the unlabeled data. Through minimizing the overall loss on both labeled and unlabeled data, we generate a label assignment pool. Each label assignment in the pool is first evaluated as an output candidate, then selected with uncertainty for privacy consideration. The proposed algorithm provides high classification accuracy, when the labeled data is limited and when the privacy budget is small, under differential privacy. Extensive experiments show the effectiveness of the proposed algorithm on both real datasets and synthetic datasets.

1. Introduction

Privacy preservation is crucial. In a society that knowledge discovery techniques are ubiquitously applied, there is a great necessity to take extra care to privacy leakage [1–3]. Privacy preservation is frequently concerned about in traditional data mining tasks, such as classification and recommender systems [4–8].

SVMs aim to find the solution to a regularized convex optimization problem based on ERM. In SVM, support vectors are the informative data points in the training set to support the maximum margin decision boundary. The model weight vector $\mathbf{w}$ is returned as the solution of the optimization problem. As seen from Eq. (1), returning $\mathbf{w}$ implies the blatant release of these informative points, which is privacy leaking. In Eq. (1), I denotes the index set for the specific data points served as support vectors [9,10], and α_i denotes the point-wise dual variable.

$$\mathbf{w} = \sum_{i \in I} \alpha_i y_i \mathbf{x}_i \tag{1}$$

Researchers have made great efforts to prevent privacy leakage. The state-of-the-art privacy model is *differential privacy* [11–14]. It ensures no distinguishable contribution of a specific individual can be observed by the adversary. Differential privacy provides both algorithmic and semantic preservation for privacy.

In the area of private classification through building differentially private SVMs, several pieces of work has been done [15–18] in recent decades. Two popular techniques, *output perturbation* and *objective perturbation*, initiated by Chaudhuri et al., [15,18], are frequently applied. In the above perturbation-based techniques, to restrict the influence of an individual data point, the convexity and smoothness of the objective function has to be guaranteed [18,19].

In output perturbation, uncertainty is introduced into the returned weight vector $\mathbf{w}$. The magnitude of the uncertainty depends on the sensitivity of the weight vector $\mathbf{w}$ [13]. Output perturbation is performed after the optimization, in the way of "*optimize – then – perturb*", as illustrated in Fig. 1. Therefore, it is independent of the optimization process.

In contrast, in objective perturbation, the uncertainty is introduced into the objective function before the optimization process. As a result, the converging process may get affected. Also, the magnitude of the uncertainty is not related to the sensitivity of $\mathbf{w}$. Instead, it is carefully chosen so that the probability ratio for returning indistinguishable objective values is bounded. As illustrated in Fig. 2, objective perturbation is performed before the optimization, in the way of "*perturb – then – optimize*".

Both of the aforementioned techniques are done in the setting of inductive learning. They have the characteristics described below. First, randomness is added feature-wisely to the model vector

[*] Corresponding author.
E-mail addresses: lxqlara@gmail.com (X. Liu), qianmu@njust.edu.cn (Q. Li), taoli@cs.fiu.edu (T. Li).

http://dx.doi.org/10.1016/j.knosys.2017.07.006
0950-7051/© 2017 Elsevier B.V. All rights reserved.

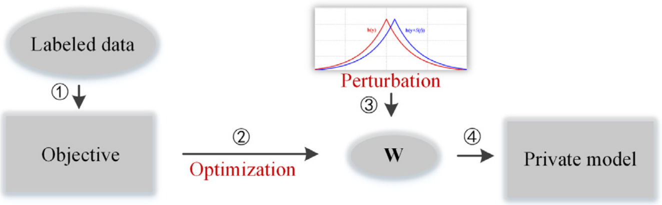

Fig. 1. The process of output perturbation in SVM.

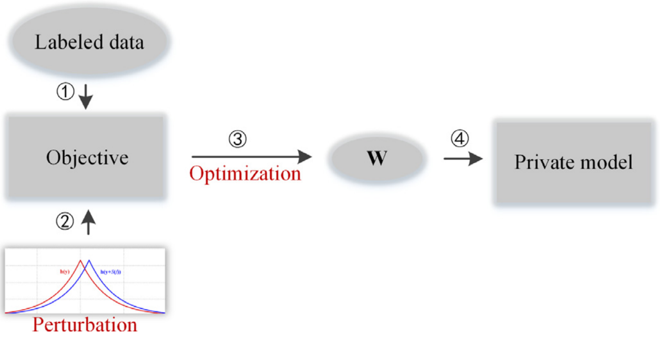

Fig. 2. The process of objective perturbation in SVM.

w. Higher dimension implies larger noise addition. Therefore, the information loss is larger, causing the training data to lose its representative capability. Second, to provide reasonable accuracy, the sample complexity of the training data (labeled) is often large. It is impractical when the labeled data is considered to be expensive. Third, the accuracy is unacceptable when the privacy budget is relatively small. Fourth, the sensitivity, in either output perturbation or objective perturbation, is uneasy to analyze.

Motivation. Trying to output a general, fit-all model requires a large number of labeled data. However, the labeled data is often expensive. In addition, privacy is preserved through noise addition. Large noise (corresponding to small privacy budget) causes massive accuracy degeneration. Another consideration is that the knowledge conveyed in the unlabeled data is often overlooked. Transductive learning [20] makes use of the unlabeled data to improve the performance. It provides efficient learning and requires smaller labeled sample complexity. In addition, it ensures “free” label privacy for the unlabeled data. Note that the unlabeled data is the data to be classified in the data mining process. This motivates us to extend the idea of TSVMs [21,22] into our task of private classification.

Our contributions. The contributions of our work are summarized as follows. We propose a private (binary) classification algorithm with high accuracy when the labeled data is limited and the privacy budget is small. First, we construct a pool of approximately correct label assignments using both the labeled and unlabeled data. In the pool, each label assignment is evaluated with an authority learned with a “safe” TSVM [23]. By safe, we indicate the accuracy of the authority is definitely better than the accuracy provided in the inductive way (trained with the same labeled data). The authority serves as a moderate reference to prioritize available label assignments in the pool. Under the exponential mechanism, a good label assignment (i.e., in large agreement with the authority) is returned with exponentially exaggerated confidence and is guaranteed to satisfy differential privacy. Extensive experiments are executed to show the effectiveness of prioritizing available label assignments with the learned authority.

Our algorithm differs from the aforementioned perturbation-based methods in that we implement random sampling under the exponential mechanism [24], instead of introducing uncertainty into the objective function. In the way of leveraging high-level randomness, the analysis of the sensitivity is greatly simplified. With limited labeled data, we effectively implement private classification. Unlike conventional situations, unlabeled data itself prioritizes all available label assignments.

The rest of the paper is organized as below. The related work is discussed in Section 2. The background and problem formulation are presented in Section 3. We present the detailed design of the algorithm in Section 4. We provide the utility analysis of the proposed algorithm in Section 5. In Section 6, we conduct extensive experiments on real datasets and synthetic datasets to compare the performance of the proposed algorithm with related ones. We conclude the work in Section 7.

2. Related work

2.1. Differentially private ERM

ERM is often talked about for performing classification tasks in the area of machine learning and data mining. Both SVM and TSVM methods highly rely on the theory of ERM. A rich line of work [18,19], [25–27] has been done based on the ERM theory to provide differential privacy. Chaudhuri et al., initiated the work on differentially private regularized ERM [18] and proposed both *output perturbation* and *objective perturbation* techniques. In their work, the loss function and the regularizer need to satisfy certain convexity and differentiability criteria.

Both the two above techniques have been applied to build differentially private SVMs in [18]. Chaudhuri’s work shows that large sample complexity is inevitable in building private learners if certain accuracy level is required. Wang et al., demonstrate the equivalence between private learnability and private Asymptotic Empirical Risk Minimization (AERM) in [27]. Their theory claims that if there exists an algorithm which can minimize the asymptotic empirical risk, the model is privately learnable.

2.1.1. Private SVMs

High-dimensional data is more difficult to handle since noise is often added feature-wisely. As a result, high dimension implies larger noise magnitude. To solve this problem, Rubinstein et al., combine output perturbation with Fourier transformation for feature mappings [17]. To further boost the accuracy, Li et al., propose a hybrid frame work in [16] based on feature mapping in [17]. Their work uses a portion of public data for calculating the Fourier mapping vectors. The goal is to make sure feature mappings do not depend on the private data.

The performance can also be improved through relaxing the privacy requirement. Kifer et al., present a more accurate objective perturbation algorithm to provide Approximate Differential Privacy (ADP) [28] in [25]. ADP is formulated as (ϵ, λ)-differential privacy. Here, λ is a negligible confidence parameter. ADP guarantees that with high confidence $1-\lambda$, the algorithm preserves ϵ-differential privacy. In the implementation of ADP, Gaussian noise is used instead of the Laplacian noise. The slack parameter λ enables less noise, thus higher accuracy.

Other extensions include private kernel methods. Jain and Thakurta propose solutions for non-translation invariant kernels (polynomial kernel) in [26] with differential privacy guarantee. In comparison, the methods in [17,18] handle translation invariant kernels.

Bassily et al., talk about the implementation of exponential mechanism in [19]. In their work, the sampling confidence of a model hypothesis is evaluated with the opposite of the objective value. The hypothesis returning lower objective value is exponentially more likely to be output. However, assumptions of the convexity of the objective function, e.g., the Lipschitz continuity and