Neurocomputing 174 (2016) 1107–1115

Contents lists available at ScienceDirect

Neurocomputing

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

# Local learning-based feature weighting with privacy preservation

Yun Li [a,*], Jun Yang [a], Wei Ji [b,c]

[a] College of Compter Science and Jiangsu High Technology Research Key Laboratory for Wireless Sensor Networks, Nanjing University of Posts and Telecommunications, Nanjing, China
[b] College of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing, China
[c] Key Laboratory of Cloud Computing & Complex System, Guilin University of Electronic Technology, Guilin, China

ARTICLE INFO

*Article history:*
Received 4 March 2015
Received in revised form
14 August 2015
Accepted 12 October 2015
Communicated by Feiping Nie
Available online 20 October 2015

*Keywords:*
Local learning
Feature weighting
Privacy preservation

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. On the other hand, feature selection is also one of the key problems in data mining and machine learning. However, for privacy-preserving feature selection, the relevant papers are few. In this paper, a local learning-based feature weighting framework is introduced. Moreover, in order to preserve the data privacy during local learning-based feature selection, the objective perturbation and output perturbation strategies are used to produce local learning-based feature selection algorithms with privacy preservation. Meanwhile, we give deep analysis about their privacy preserving property based on the differential privacy model. Some experiments are conducted on benchmark data sets. The experimental results show that our algorithms can preserve the data privacy to some extent and the objective perturbation always obtains higher classification performance than output perturbation when the privacy preserving degree is constant.

© 2015 Elsevier B.V. All rights reserved.

## 1. Introduction

Feature selection is one of the key problems in machine learning and data mining [1,2], which 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 [3]. 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 [1,2]. On the other hand, if the categorization is based on output style, feature selection algorithms can be divided into either feature weighting/ranking algorithms or subset selection algorithms [3]. The output of feature selection algorithms discussed in this paper is feature weighting. A comprehensive survey of existing feature selection techniques and a general framework for their unification can be found in [1–3].

Current feature selection research focuses on the classification accuracy and stability of selected features, however, the privacy preservation property is also very important for feature selection. The privacy preservation means the selected features cannot leak the privacy information of data, and the privacy information is the sensitive one that data owner reluctant to disclose. The privacy information has been a growing concern in medical records, financial records, web search histories, social network data, etc. The privacy-preserving classification and regression [4–7] have been deeply analyzed. However, the privacy preserving feature selection algorithms are very few. In this paper, we will present some works on the privacy preserving feature selection. Concretely, two strategies, i.e., output perturbation and objective perturbation, are adopted to add privacy preserving property for local learning-based feature selection algorithm, and the $\varepsilon$-differential privacy [8] is chosen as privacy model. For the local learning-based feature selection, the logistic loss with L2-regularizer is utilized to design the evaluation criterion of feature selection.

This paper is the expand of our previous work [9] and it is organized as follows, the feature weighting algorithm based on local learning FWELL is introduced in Section 2. Section 3 presents privacy model. Section 4 describes the differentially private feature selection algorithm based on output perturbation Output-FWELL. Section 5 introduces the differentially private feature selection algorithm based on objective perturbation Objective-FWELL. The experimental results on bench mark data sets are shown in Section 6. The paper concludes in Section 7.

## 2. 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 $\mathbf{x}_i=(x_{i1},x_{i2},\ldots,x_{id})\in\mathbb{R}^d$, and $y_i$ is the corresponding label.

* Corresponding author. Tel.: +86 25 85866421; fax: +86 25 85866151.
*E-mail address:* liyun@njupt.edu.cn (Y. Li).

http://dx.doi.org/10.1016/j.neucom.2015.10.038
0925-2312/© 2015 Elsevier B.V. All rights reserved.

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 labels (i.e., near miss sample $NM(\mathbf{x}_i)$) [10]. 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 [11]. 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 Eq. (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 $|\cdot|$ 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 [12] 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 [10]. In other words, the feature weighting based on local learning is like to scale each feature, and thus obtains 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 [13], 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 $\lambda$ 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 Eq. (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 Eq. (2). Of course, the optimal feature weights can be found by many other optimization approaches.

**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 $\lambda$ in Eq. (2).

*Step* 2. Initialize $\mathbf{w} = (1, 1, \ldots, 1) \in \mathbb{R}^d$.

*Step* 3. For $i = 1, 2, \ldots, n$

(a) Given $\mathbf{x}_i$, find the $NH(\mathbf{x}_i)$ and $NM(\mathbf{x}_i)$.

(b) Based on Eq. (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 $\mathbf{w}$.

In the following analysis and experiments, the L2 regularizer is used as $\mathcal{R}(\mathbf{w})$ in Eq. (2) for its rotational invariance and strong stability property [14]. 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 with name FWELL.

## 3. Privacy model

For privacy measure, we adopt $\varepsilon$-differential privacy model [8], which is a measure of quantifying the privacy-risk associated with computing functions of sensitive data. A statistical procedure satisfies $\varepsilon$-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 [5]. And $\varepsilon$-differential privacy model is robust to known attacks, such as those involving side information [15]. $\varepsilon$-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 and regression [4–7].

**Definition 1.** A randomized mechanism $A$ provides *$\varepsilon$-differential privacy*, if, for all data sets $D$ and $D'$ which differ by at most one element, and for all output subsets $S \subseteq Range(A)$:

$$Pr[A(D) \in S] \le \exp(\varepsilon) \times Pr[A(D') \in S] \tag{4}$$

The probability $Pr$ is taken over the coin tosses of $A$, and $Range(A)$ denotes the output range of $A$. The privacy parameter $\varepsilon$ measures the disclosure. When data sets which are identical except for a single entry are input to the algorithm $A$, the two distributions on the algorithm's output are close. That is, any single entry of the data set does not affect the output. This means that an adversary, who knows all but one entry of the data set, cannot gain much additional information about this entry by observing the output of the algorithm. So the privacy of this entry is preserved. In other words, suppose $D = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)\}$ and $D' = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_{n'}, y_{n'})\}$ be two data sets that differ in the value of the $n$th individual. The two distributions on the differentially private algorithm $A$'s output are close. Then an adversary, who knows all but the $n$th entry of the data set, cannot gain much additional information about this entry by observing the output of the algorithm.

## 4. Differentially private feature selection based on output perturbation

### 4.1. Sensitivity analysis

In order to propose privacy preserving FWELLs in terms of the differential privacy in Eq. (4), we like to adopt the output perturbation and objective perturbation strategy. In this section, we will present the differentially private FWELL with output perturbation. This algorithm depends on the FWELL's sensitivity. In general, the sensitivity is always defined as follows [5,16].

**Definition 2.** For any function $A$ with $n$ inputs, we define the *sensitivity* $\Delta Q$ as the maximum, over all inputs, of the difference in the value of $A$ when one input of $A$ is changed. That is,

$$\Delta Q = \max_{D,D'} \|A(D) - A(D')\| \tag{5}$$

Data sets $D$ and $D'$ differ by at most one element.

According to Definition 2, we can analyze the sensitivity of FWELL with L2 regularizer and obtain Corollary 1.

**Corollary 1.** *The feature weighting algorithm described in Algorithm 1 (FWELL) with L2 regularizer has the sensitivity $2/\lambda n$.*

**Proof.** Let $D = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)\}$ and $D' = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_{n'}, y_{n'})\}$ 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 FWELL when

Download English Version:

https://daneshyari.com/en/article/411644

Download Persian Version:

https://daneshyari.com/article/411644

Daneshyari.com