An Incremental Broad Learning Approach for Semi-Supervised Classification

Conference Paper - August 2019
DOI: 10.1109/DASC/PiCom/CBDCom/CyberSciTech.2019.00053

5 authors, including:

Tie Qiu
Tianjin University
207 PUBLICATIONS 2,865 CITATIONS
SEE PROFILE

Huansheng Ning
University of Science and Technology Beijing
200 PUBLICATIONS 3,929 CITATIONS
SEE PROFILE

Chen Chen
Institute of Electrical and Electronics Engineers
98 PUBLICATIONS 643 CITATIONS
SEE PROFILE

Ning Chen
Tianjin University
9 PUBLICATIONS 347 CITATIONS
SEE PROFILE

Some of the authors of this publication are also working on these related projects:

- conference papers
- Cyberspace

All content following this page was uploaded by Chen Chen on 10 July 2020.

The user has requested enhancement of the downloaded file.
An Incremental Broad Learning Approach for Semi-Supervised Classification

Xize Liu†, Tie Qiu†, Chen Chen‡, Huansheng Ning§, and Ning Chen∗
†School of Software, Dalian University of Technology, Dalian 116620, China
‡School of Computer Science and Technology, Tianjin University, Tianjin 300350, China
§State Key Laboratory of Integrated Service Networks, Xidian University, Xi’an 710071, China
∗Corresponding Email: qiutie@ieee.org

Abstract—Broad Learning System (BLS) is a fast and accurate supervised learning method without deep structure. However, the classifier trained by BLS cannot achieve expected accuracy if the labeled data are insufficient. In this paper, we develop an Incremental Semi-supervised Broad Learning method (ISBL) based on BLS to classify a partially labeled dataset, which applies manifold regularization to explore the underlying data distribution and improve accuracy. ISBL applies to scenarios where data is generated over time. While new patterns are added to the learning algorithm, the proposed method updates the classification model sustainability without retraining. By comparing with original BLS and other semi-supervised classification techniques on various datasets with different dimensions, we verified that ISBL outperforms these methods on accuracy. Experimental results demonstrate that ISBL utilizes unlabeled data effectively and achieves high accuracy. Meanwhile, the incremental learning method reduces the learning time and storage of historical data.

Index Terms—Semi-supervised learning, Broad Learning System (BLS), incremental learning, manifold regularization.

I. Introduction

Supervised learning methods based on the artificial neural network have been widely applied in diverse areas, including Image Recognition [1], Intelligent Vehicular Networks [2], [3], Natural Language Processing [4] and so on. With the rapid development of computer intelligence and data acquisition technology, how to make better use of unlabeled data to improve the performance of machine learning becomes extremely important.

Recently, an effective Broad Learning System (BLS) was proposed by Chen et al. [5]. BLS only calculate the weight parameter that connected to the output layer, which significantly shortens the network training time. At the same time, the accuracy of the training model is improved by the increment of the enhancement nodes. Although BLS is powerful in solving the problem of large volumes and high dimensions of data, it is mainly applied in supervised learning tasks. In some cases like network intruders detection [6], email filtering [7] and social spammer detection [8], most of the labels need to be manually labeled, while a multitude of unlabeled data is easy and cheap to collect. Therefore, the semi-supervised scheme has emerged as a popular strategy due to the expensiveness of getting sufficient labeled data [9], [10].

In many application scenarios, such as computer vision [11] robotics [12], Emergency Internet of Things [13], data is generated over time. While all data arrive at the same time, the classical batch machine learning methods cannot process the data in a short time and result in more and more unprocessed data, which makes the learning method limited by time and memory [14], [15]. Besides, they do not constantly integrate new information into trained models, but regularly rebuild new models. Incremental machine learning methods can update the training model according to the distribution changes of the newly added data so that the classifier model maintains high prediction accuracy for the data of current period [16], [17].

To overcome the disadvantage of BLS cannot make use of unlabeled data, we design an Incremental Semi-Supervised Broad Learning (ISBL) algorithm to classify the partially labeled datasets. Besides, it is able to dynamically update the classification model according to newly added data without retraining.

II. Incremental Semi-supervised Broad Learning Algorithm

ISBL improves the classification accuracy compared with BLS by utilizing manifold regularization [18] with both labeled and unlabeled data. The functions that learned on the dataset is constrained by the high-dimensional structure of the dataset.

The structure of the proposed ISBL is shown in Fig. 1. In the first stage, the mapped features are generated from the training patterns according to the auto-encoder model. Meanwhile, the Laplacian matrix is obtained from labeled and unlabeled data. Then enhancement nodes are generated from feature nodes to expand mapped features. Finally, the weight W connected to the output layer is calculated by manifold regularization.

ISBL is able to learn the high-dimensional structure of the data from the data itself, without the use of predetermined classification. The patterns are represented by vertices of the graph and the similarity of different patterns is expressed as the weighted edges. If the similarity of patterns is high, its most likely that they have the same label. Minimization of the following cost function based on the graph ensures that points close to each other on the manifold are mapped close.
to each other in the low dimensional space, preserving local distances\cite{19, 20}.

\[ L_{ae} = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(y_i - y_j)^2 = Tr(Y^TLY) \] 

(1)

\( y_i, y_j \) are the prediction results with respect to pattern \( x_i \) and \( x_j \), \( n \) is the number of patterns. \( Y = [y_1, y_2, ..., y_n]^T \). \( Tr() \) denotes the trace of a matrix. \( L \) is a Laplacian matrix, which is calculated by \( D - W \). \( D \) is a diagonal matrix with its diagonal elements \( D_{ii} = \sum_{j=1}^{n} w_{ij}, W_{ij} \) represents the similarity of patterns \( x_i \) and \( x_j \), which can be calculated by K-nearest neighbor method (KNN). The similarity \( w_{ij} \) between \( x_i \) and its neighbors can be calculated by using a Radial Basis Function, which is shown as follows:

\[ w_{ij} = \begin{cases} \exp(-\frac{\|x_i - x_j\|^2}{2\sigma^2}), & \text{if } x_i \text{ and } x_j \text{ are neighbors;} \\ 0, & \text{otherwise.} \end{cases} \] 

(2)

A. Semi-supervised Broad Learning Model

ISBL utilizes both labeled data and unlabeled data to train the classification model. We assume that labeled data set is \( \{X_l \in \mathbb{R}^{N \times l}, Y_l \in \mathbb{R}^{N \times L}\} = \{x_{i1}, y_{i1}\}_{i=1}^l \), and the unlabeled data set is \( \{X_u \in \mathbb{R}^{N \times l}, Z_u \}_{i=1}^u \). \( N \) and \( u \) indicate the number of labeled and unlabeled data respectively. \( N \) represents the number of patterns in the training set, \( M \) is the dimension of the input data, and \( D \) is the dimension of the output.

First, the training data set \( X \) is mapped to \( n \) groups of mapped features. The number of feature nodes in group \( i \) is \( k_i \). The \( i \)th mapped features can be denoted as the following equation.

\[ Z_i = \phi_i(XW_{ei} + \theta_{ei}), i = 1, 2, ..., n \] 

(3)

where the weight \( W_{ei} \in \mathbb{R}^{N \times k_i} \) and bias \( \theta_{ei} \in \mathbb{R}^{k_i} \) are randomly generated. \( Z^e \) represents \( n \) groups of mapped features, which is used to generate enhancement nodes.

\[ Z^e = [Z_1, Z_2, ..., Z_n] \] 

(4)

The second step is to generate the enhancement nodes. Enhancement nodes are used to expand mapped features in BLS. The purpose of introducing enhancement nodes is to add non-linear factors to the neural network. We assumed that the model has \( m \) groups enhancement nodes, each group contains \( p_j \) enhancement nodes. The \( j \)th group of enhancement nodes is denoted as Eq. (5).

\[ H_j = \xi_j(Z^eW_{hj} + \theta_{hj}), j = 1, 2, ..., m \] 

(5)

where \( \xi_j \) represents the activation function (e.g., logsig function) of the \( j \)th group of enhancement nodes. The weight \( W_{hj} \in \mathbb{R}^{N \times k_i \times p_j} \) is a normalized random matrix, so that the feature nodes can be mapped to a high-dimensional features by non-linear mapping. The deviation \( \theta_{hj} \in \mathbb{R}^{k_i \times p_j} \) is randomly generated. Donate all the enhancement nodes as \( H^m \).

\[ H^m = [H_1, H_2, ..., H_m] \] 

(6)

Finally, all the feature nodes and enhancement nodes are connected, denoted as \( A^m_n \).

\[ A^m_n = [Z_1, Z_2, ..., Z_n, H_1, H_2, ..., H_m] = [Z^e, H^m] \] 

(7)

ISBL calculates the weight connected to the output layer by utilizing the manifold regularization framework. The similarity of data is represented by Laplacian matrix \( L \) according to Eq. (2). The following cost function obtains \( w \).

\[ \arg \min_{w} L_{ISBL} : C \frac{1}{2} \|AW - Y\|^2 + \frac{1}{2} \|W\|^2 + \frac{1}{2} Tr(W^TLY) \] 

(8)

where the first term is the error vector between the predicted and real results, the second term is the regularization against overfitting. \( L \in \mathbb{R}^{(k \times m) \times (k \times m)} \) is the Laplacian matrix calculated from the labeled and unlabeled data. \( C \) and \( \lambda \) are two different regularization parameters. \( Y \) is the prediction result. Due to \( Y = AW \), we get Eq. (9).

\[ \arg \min_{w} L_{ISBL} : C \frac{1}{2} \|AW - Y\|^2 + \frac{1}{2} \|W\|^2 + \frac{1}{2} Tr(W^TAW) \] 

(9)
In order to solve the minimum value of Eq. (9), we set the gradient of $L_{ISBL}$ to zero,

$$\nabla L_{ISBL} = A^T C (AW - Y) + W + \lambda W^T A^T L A W = 0$$  \hspace{1cm} (10)$$

Our proposed learning model is designed for semi-supervised learning. Therefore, it is common that the number of labeled patterns is fewer than the number of enhancement nodes and feature nodes. If Eq. (10) is an underdetermined equation, it has either no solution or infinitely many solutions. To solve the problem, we restrict $W$ to be a linear combination of the rows of $A$. Then we can derive the following equation:

$$W = A^T (I + CA^T + \lambda L A^T)^{-1} CY$$  \hspace{1cm} (11)$$

where $Y = [Y_{obs}\ 0_{obs}]^T$, $n$ is the dimensions of output results.

If the number of labeled patterns is larger than that of unlabeled pattern, Eq. (10) is an overdetermined equation, we can get the following approximate solution,

$$W = (I + A^T CA + \lambda A^T L A)^{-1} A^T CY$$  \hspace{1cm} (12)$$

After calculating the weight connected to the output layer, the label of test dataset can be predicted according to $AW$.

B. Increment Learning Model

Data is continuously generated over time in many scenarios of the real world. Most machine learning algorithms are limited by available resources in terms of time and memory. To enable ISBL to update the training model without retraining while data distribution changes over time, we designed an incremental semi-supervised broad learning algorithm.

To update the model through incremental learning method, Eq. (12) is divided into two parts.

$$K = I + A^T (C + \lambda A)$$  \hspace{1cm} (13)$$

$$W = K^{-1} A^T CY$$  \hspace{1cm} (14)$$

Suppose that the initial training data set is $X_0 = \{(x_0, y_0), \tau_0\}$. $\tau_0$ is the unlabeled patterns. The intermediate variable $K_0$ and weight $W_0$ is shown as follows.

$$K_0 = I + A_0^T (C_0 + \lambda A_0)$$  \hspace{1cm} (15)$$

$$W_0 = K_0^{-1} A_0^T C_0 Y_0$$  \hspace{1cm} (16)$$

The new added patterns are represented as $X_1 = \{(x_1, y_1), \tau_1\}$. In order to explore the relationship of $W_0$ and updated weight $W_1$, we calculate $W_1$ by utilizing partitioned matrix. Thus we get the following equations:

$$W_1 = K_1^{-1} A_1^T \begin{bmatrix} A_0 \ C_0 \ 0 \ A_1 \ 0 \ C_1 \end{bmatrix} \begin{bmatrix} Y_0 \ Y_1 \ \end{bmatrix}$$  \hspace{1cm} (17)$$

where,

$$K_1 = I + A_1^T \begin{bmatrix} A_0 \ 0 \ A_1 \ \end{bmatrix} \begin{bmatrix} C_0 \ 0 \ C_1 \end{bmatrix} + \lambda L_{(X_0, X_1)}$$  \hspace{1cm} (18)$$

$$L_{(X_0, X_1)} = \begin{bmatrix} L_{X_0} + D_{(X_0, X_1)} & -Q_{(X_0, X_1)} \\ -Q_{(X_0, X_1)}^T & L_{X_1} + D_{(X_0, X_1)} \end{bmatrix}$$  \hspace{1cm} (19)$$

$L_{(X_0, X_1)}$ represents the Laplacian matrix calculated from the historical patterns and the newly added patterns. $C_1$ is the regularization parameter of the newly added data. $Q_{(X_0, X_1)}$ and $D_{(X_0, X_1)}$ are Laplacian operators between data set $X_0$ and $X_1$. Due to Laplacian matrix aims to improve the accuracy of the model, the impact on updating weight $W_1$ for $Q_{(X_0, X_1)}$ and $D_{(X_0, X_1)}$ can be ignored. We simplify Eq. (19) to the following equation.

$$L_{(X_0, X_1)} = \begin{bmatrix} L_{X_0} & 0 \\ 0 & L_{X_1} \end{bmatrix}$$  \hspace{1cm} (20)$$

Then $K_1$ can be calculated by Eq. (18) and Eq. (20)

$$K_1 = I + A_1^T (C_0 + \lambda L_{X_0}) A_0 + A_1^T (C_1 + \lambda L_{X_1}) A_1 = K_0 + A_1^T (C_1 + \lambda L_{X_1}) A_1$$  \hspace{1cm} (21)$$

To explore the relationship of $W_0$ and $W_1$, we transform the last three matrices in Eq. (17) to the following equation.

$$\begin{bmatrix} A_0 \ A_1 \ \end{bmatrix} \begin{bmatrix} C_0 \ 0 \ C_1 \ Y_0 \ Y_1 \ \end{bmatrix} = A_1^T C_0 Y_0 + A_1^T C_1 Y_1$$  \hspace{1cm} (22)$$

$$= K_0^{-1} A_1^T C_0 Y_0 + A_1^T C_1 Y_1$$

$$= A_1^T C_0 Y_0 + A_1^T C_1 Y_1$$

$$= K_1 W_0 - A_1^T (C_1 + \lambda L_{X_1}) A_1 W_0 + A_1^T C_1 Y_1$$

At last, we left multiply Eq. (22) by $K_1^{-1}$ and get the following equation:

$$W_1 = W_0 - K_1^{-1} A_1^T [C_1 Y_1 - C_1 + \lambda L_{X_1}] A_1 W_0$$  \hspace{1cm} (23)$$

When adding new patterns to ISBL, the classification model can continuously updated by calculating parameters $W_1$ and $K_1$.

III. EXPERIMENTAL RESULTS

We compared the performance of ISBL with related algorithms, including Broad Learning System (BLS) [5], Laplacian Support Vector Machine (LapSVM), Laplacian Ridge Regression (LapRR) [21]. All algorithms were implemented on MATLAB software plat on a laptop that equips with Intel i7 2.4 GHz CPU, 64 GB memory.

A. Data Sets

The experiments are applied on NSL-KDD [22] and MNIST datasets. NSL-KDD dataset is designed to build a predictive model capable of distinguishing between intrusions and good normal connections. After pre-processing, each pattern consists of 121 attributes. The data labels range from 0 to 4, which represent five different connections. MNIST is one of the most popular deep learning data sets. It is a handwritten digit recognition dataset with 784 attributes each pattern.
B. Experimental Results

We compare the test accuracy between ISBL and BLS by training classifiers with the different number of labeled patterns. Fig. 2 and Fig. 3 show the test accuracy on NSL-KDD and MNIST data sets respectively. The total number of labeled and unlabeled data trained by ISBL is 7000. The number of mapped features and enhancement nodes in ISBL are set to 100 and 2000 respectively. In both figures, we can see that the test accuracy of ISBL is higher than BLS in the different number of labeled patterns. The test accuracy increases with the number of labeled data increasing. While the number of labeled data is small, ISBL is much more accurate than the original BLS. Therefore ISBL is an efficient semi-supervised learning method that is suitable for scenarios in the absence of labeled pattern.

Next, we compare the classification ability of ISBL with existing semi-supervised learning methods LapRR and LapSVM. Fig. 4 and Fig. 5 show the test accuracy on the NSL-KDD and MNIST data sets respectively. We can see that the performance of ISBL is better than the other two algorithms, especially on MNIST data set. Therefore ISBL can maintain high accuracy on high-dimensional data sets.

To show the effectiveness of the proposed incremental algorithms, we test the accuracy and training time of incremental ISBL. The number of mapped feature and enhancement nodes are set to 100 and 5000 respectively. The results of the experiments are shown in TABLE I. Five thousand labeled patterns and 5000 unlabeled patterns are added to update the

<table>
<thead>
<tr>
<th>N_labeled</th>
<th>N_unlabeled</th>
<th>Test Accuracy</th>
<th>Training Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>5000</td>
<td>5000</td>
<td>96.87%</td>
<td>270s</td>
</tr>
<tr>
<td>10000</td>
<td>10000</td>
<td>97.36%</td>
<td>100s</td>
</tr>
<tr>
<td>15000</td>
<td>15000</td>
<td>97.65%</td>
<td>101s</td>
</tr>
<tr>
<td>20000</td>
<td>20000</td>
<td>97.77%</td>
<td>98s</td>
</tr>
<tr>
<td>25000</td>
<td>25000</td>
<td>97.90%</td>
<td>101s</td>
</tr>
<tr>
<td>30000</td>
<td>30000</td>
<td>97.92%</td>
<td>101s</td>
</tr>
</tbody>
</table>
learning model each time. \( N_{\text{label}} \) represents the total number of labeled patterns for training, which includes the amount of historical and newly added data. \( N_{\text{unlabel}} \) denotes the total number of unlabeled patterns for training. The test accuracy increases with times of incremental learning. The accuracy of prediction is updated without retraining, which saves storage and training time. Therefore ISBL is suitable for scenarios where data is generated continuously over time.

IV. Conclusion

In this paper, an incremental semi-supervised broad learning algorithm is proposed for the classification of partially labeled datasets. ISBL explores the distribution of training data according to the Laplacian matrix. Manifold regularization is applied to guaranteed the accuracy of the classification model. Then we design an incremental method for scenarios that data are generated over time. When new patterns come to the learning algorithm, ISBL can update the model without retraining. It doesn’t save the patterns that already trained, which reduces the storage space of data. At the same time, it only calculates the Laplacian matrix of newly added patterns, which saves training time. Compared with BLS and other classical semi-supervised algorithms LapRR and LapSVM, ISBL achieves high performance on datasets with various dimensions. In the future, we'll focus on optimizing the computational complexity of the algorithm so that the training time is shortened.

Acknowledgment

This work is supported by National Natural Science Foundation of China (No. 61672131).

References


