Semi-Supervised Classification via Both Label and Side Information

Rui Zhang, Feiping Nie, and Xuelong Li, Fellow, IEEE

ruizhang8633@gmail.com, feipingnie@gmail.com, xuelong_li@ieee.org

CONTRIBUTION: UTILIZING BOTH LABEL AND SIDE INFORMATION

As for the semi-supervised learning, both label and side information serve as pretty significant indicators for the classification. However, majority of the associated works only focus on one side of the road. To tackle this issue, SC method is proposed with taking both of them into consideration simultaneously.

PARAMETER-FREE SIMILARITY

To achieve sparse parameter-free similarity, we introduce the following optimization w.r.t. $a_i$ as

$$
\min_{a_i \in [-1,1], \sum a_i = 1} \text{Tr}(X^T (D-A) X) + \sum_{i,j=1}^n \frac{e}{2} a_i^2 \| y_i - y_j \|^2 \tag{1}
$$

Accordingly, the Lagrangian function could be illustrated as

$$
\frac{1}{2} \eta a_i + \frac{e}{2} \| y_i - y_j \|^2 - \eta (a_i^T 1 - 1) - \beta_i a_i \tag{2}
$$

We could achieve a sparse parameter-free similarity $a_i = (-\frac{1}{\beta_i} + \eta) i$.

SEMISUPERVISED LEARNING

Generally speaking, the classification problem is to minimize the intrinsic graph problem $G$ with maximizing the penalty graph problem $G'$ simultaneously. Therefore, the classification problem can be further represented as

$$
\min_{Y} \sum_{i,j} a_{ij}^2 \| y_i - y_j \|^2 = \min_{Y} 2 \text{Tr}(Y^T L Y) \tag{3}
$$

In particular, the graph-based semisupervised learning (GSL) problem can be represented as

$$
\min_{F_a} \frac{\text{Tr}(Q^T A Q)}{\text{Tr}(Q^T B Q)} \tag{4}
$$

We could utilize the label information in $F_a$ and side information in $L^u$ and $L^s$ simultaneously.

CHARACTERISTIC FUNCTION

Apparently, the GSL problem (4) is equivalent to the following quadratic trace ratio (QTR) problem

$$
\min_{Q \in \mathbb{R}^{n \times n}} \text{Tr}(Q^T A_b Q) + 2 \text{Tr}(Q^T C) + e - \lambda \text{Tr}(Q^T B Q) + 2 \text{Tr}(Q^T D) + f \tag{5}
$$

where $A = L^u_{aa}$, $B = L^a_{uu}$, $C = L^u_{uu}$, $D = L^u_{uu} Y 1$, $e = \text{Tr}(Y^T L^s Y 1)$ and $f = \text{Tr}(Y^T L^s Y 1)$ with $\text{Tr}(Y^T Q^T L^s Y Q) > 0$.

To solve the QTR problem (5), we introduce the characteristic function $p(\lambda)$ as

$$
p(\lambda) = \min_{Q} \left( \text{Tr}(Q^T A_b Q) + 2 \text{Tr}(Q^T C) + e - \lambda \text{Tr}(Q^T B Q) + 2 \text{Tr}(Q^T D) + f \right) \tag{6}
$$

where $\lambda \leftarrow \frac{\text{Tr}(Q^T A_b Q) + 2 \text{Tr}(Q^T C) + e}{\text{Tr}(Q^T B Q) + 2 \text{Tr}(Q^T D) + f}$ is to be iteratively updated.

CORE ALGORITHM

while $p > 0$ do

Update $\lambda \leftarrow \frac{1}{\lambda + \lambda}$

Update $Q \leftarrow (A - \lambda B)^{-1} (A D - C)$

Update $p \leftarrow \text{Tr}(Q^T (A - \lambda B) Q) + 2 \text{Tr}(Q^T (C - \lambda D)) + (e - \lambda f)$

if $p > 0$ then

[Replace $\lambda \leftarrow \lambda$]

while not converge do

Update $Q \leftarrow (A - \lambda B)^{-1} (A D - C)$

Update $\lambda \leftarrow \frac{\text{Tr}(Q^T A_b Q) + 2 \text{Tr}(Q^T C) + f}{\text{Tr}(Q^T B Q) + 2 \text{Tr}(Q^T D) + f}$

return $Q$.

Our algorithm performs much better via utilizing both label and side information. We choose 6 datasets as AR, AT&T, COIL20, FEI, FLOWER17 and IMM for the classification comparison with equal labeled data shared by each class. We could observe that the proposed SC method performs much better than other approaches on the classification accuracy with minor exceptions.

COMPARATIVE RESULTS ON TOY DATASETS.

Perfect classification results on toy datasets by virtue of both label and side information. We utilize two-spirals and three-rings synthetic databases to compare the classification results. We could observe that the proposed SC method could achieve the optimal classification results based on utilizing both label and side information. Besides, we notice that the SC method performs better than the LP method and the LLGC method.

COMPARATIVE RESULTS ON BENCHMARK DATASETS.