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

$$\min_{a_i^T\mathbf{1}=1, 0\le a_i\le 1} Tr(X(D-A)X^T) + \sum_{i,j=1}^{n}(\frac{\gamma_i}{2}a_{ij}^2) \quad (1)$$

Accordingly, the Lagrangian function could be illustrated as

$$\frac{1}{2}\|a_i + \frac{e_i}{2\gamma_i}\|_2^2 - \eta(a_i^T\mathbf{1}-1) - \beta_i^T a_i \quad (2)$$

We could achieve a sparse parameter-free similarity $a_{ij} = (-\frac{e_{ij}}{2\gamma_i}+\eta)_+$**.**

Semi-supervised learning

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

$$\begin{cases} \min_Y \sum_{i,j} a_{ij}^w\|y_i - y_j\|_2^2 = \min_Y 2Tr(Y^T L^w Y) \\ \max_Y \sum_{i,j} a_{ij}^b\|y_i - y_j\|_2^2 = \max_Y 2Tr(Y^T L^b Y) \end{cases} \quad (3)$$

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

$$\min_{F_u} \frac{Tr([Y_l;F_u]^T L^w [Y_l;F_u])}{Tr([Y_l;F_u]^T L^b [Y_l;F_u])} = \min_{F_u} \frac{Tr(\begin{bmatrix} Y_l \\ F_u \end{bmatrix}^T \begin{bmatrix} L_{ll}^w & L_{lu}^w \\ L_{ul}^w & L_{uu}^w \end{bmatrix} \begin{bmatrix} Y_l \\ F_u \end{bmatrix})}{Tr(\begin{bmatrix} Y_l \\ F_u \end{bmatrix}^T \begin{bmatrix} L_{ll}^b & L_{lu}^b \\ L_{ul}^b & L_{uu}^b \end{bmatrix} \begin{bmatrix} Y_l \\ F_u \end{bmatrix})} \quad (4)$$

We could utilize the label information in F_u and side information in L^w and L^b simultaneously.

Characteristic function

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

$$\min_{Q\in\mathbb{R}^{n_u\times c}} \frac{Tr(Q^TAQ)+2Tr(Q^TC)+e}{Tr(Q^TBQ)+2Tr(Q^TD)+f} \quad (5)$$

where $A = L_{uu}^w$, $B = L_{uu}^b$, $C = L_{ul}^w Y_l$, $D = L_{ul}^b Y_l$, $e = Tr(Y_l^T L_{ll}^w Y_l)$ and $f = Tr(Y_l^T L_{ll}^b Y_l)$ with $Tr([Y_l;Q]^T L^b [Y_l;Q]) > 0$.
To solve the QTR problem (5), we introduce the **characteristic function** $p(\lambda)$ as

$$p(\lambda) = \min_Q (Tr(Q^TAQ) + 2Tr(Q^TC) + e) - \lambda(Tr(Q^TBQ) + 2Tr(Q^TD) + f) \quad (6)$$

where $\lambda \leftarrow \frac{Tr(Q^TAQ)+2Tr(Q^TC)+e}{Tr(Q^TBQ)+2Tr(Q^TD)+f}$ is to be iteratively updated.

Core algorithm

while $p > 0$ **do**
  Update $\lambda \leftarrow \frac{\lambda_1+\lambda_2}{2}$
  Update $Q \leftarrow (A-\lambda B)^{-1}(\lambda D - C)$
  Update $p \leftarrow Tr(Q^T(A-\lambda B)Q) + 2Tr(Q^T(C-\lambda D)) + (e - \lambda f)$
  if $p > 0$ **then**
    Replace $\lambda_1 \leftarrow \lambda$

while *not converge* **do**
  Update $Q \leftarrow (A-\lambda B)^{-1}(\lambda D - C)$
  Update $\lambda \leftarrow \frac{Tr(Q^TAQ)+2Tr(Q^TC)+e}{Tr(Q^TBQ)+2Tr(Q^TD)+f}$

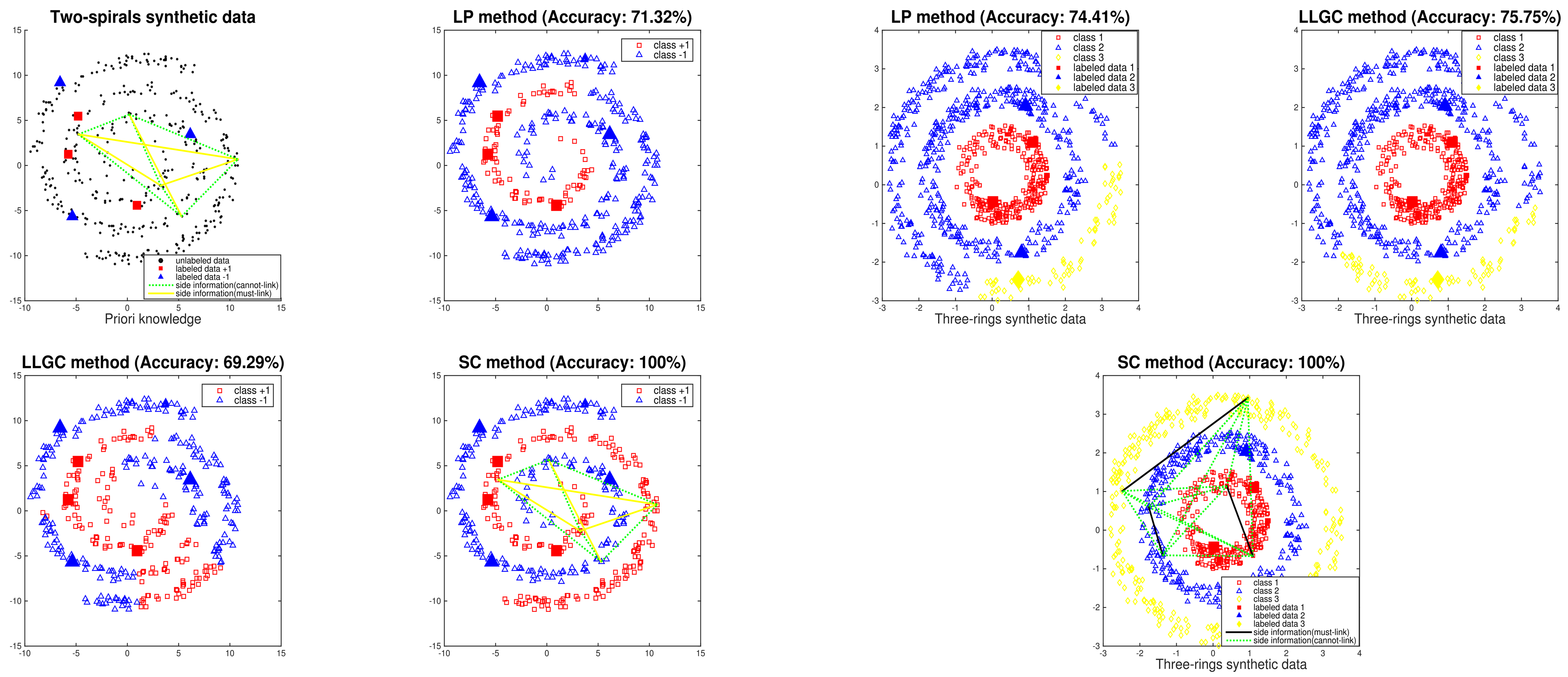
return Q

where Q is the obtained soft label matrix.

The algorithm above can be proposed based on introducing the associated characteristic function $p(\lambda)$. Additionally, theoretical analysis shows that the proposed algorithm monotonically converges to the global optimal solution of GSL problem (4).

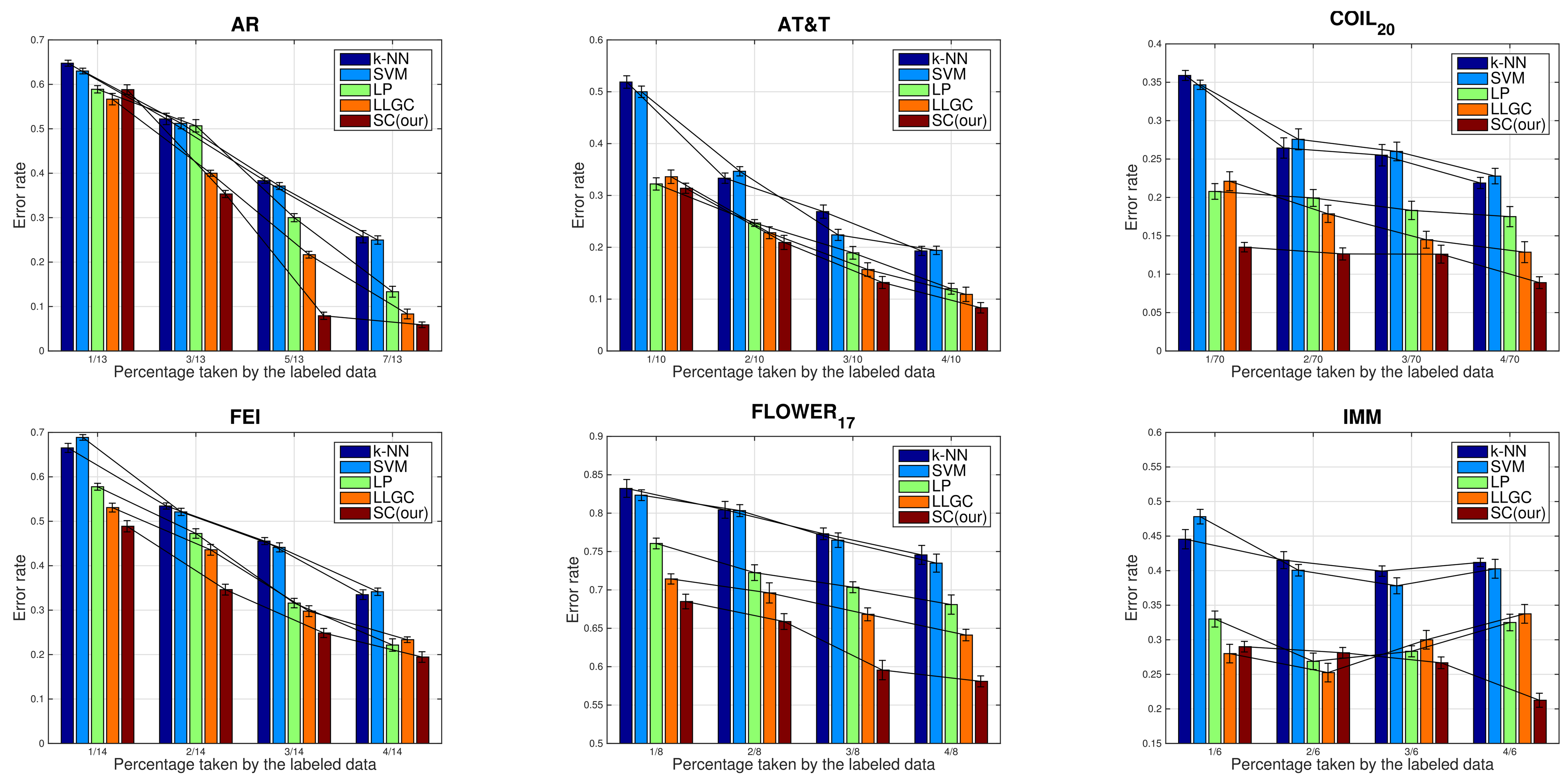
Comparative results on toy datasets.

Two-spirals toy data | Three-rings toy data



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



Our algorithm performs much better via utilizing both label and side information. We choose 6 datasets as AR, AT&T, $COIL_{20}$, FEI, $FLOWER_{17}$ 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.