Supplementary Material: Similarity Fusion via Exploiting High Order Proximity

Wentao Rong, Hong Peng, Guihua Tao, Wanlin Weng, Bin Zhang, and Hongmin Cai (✉)

Department of Computer Science and Engineering, South China University of Technology, Guangzhou, China
hmcai@scut.edu.cn

1 Optimization

The optimization problem is nonconvex, but the objective function for each variable conditional on the other two variables being fixed is convex. So we apply alternating direction method of multipliers (ADMM)[1] to solve this problem efficiently. We optimize over $\boldsymbol{S}$ and $\boldsymbol{w}$. Auxiliary variables $\boldsymbol{S_1}$, $\boldsymbol{S_2}$ are introduced to replace $\boldsymbol{S}$ and following equivalent optimization problem is considered:

$$
\begin{aligned}
\min_{\boldsymbol{S},\boldsymbol{w}} \sum_{v=1}^{m} & \Big(- w_v tr(\boldsymbol{S}\hat{\boldsymbol{P}}^v) + \beta w_v log w_v \Big) \\
& s.t.\ diag(\boldsymbol{S_2}) = \boldsymbol{0}, \boldsymbol{S_1}^{\mathsf{T}}\boldsymbol{1} = \boldsymbol{1}, \boldsymbol{S_2} = \boldsymbol{S_2}^{\mathsf{T}}, \boldsymbol{S_2} \geq \boldsymbol{0}, \boldsymbol{S} = \boldsymbol{S_1}, \boldsymbol{S} = \boldsymbol{S_2} \\
& \quad \boldsymbol{w}^{\mathsf{T}}\boldsymbol{1} = 1, 0 \leq w_v \leq 1.
\end{aligned}
$$

The augmented Lagrangian is :

$$
\begin{aligned}
\mathcal{L}(\boldsymbol{S}, \boldsymbol{S_1}, \boldsymbol{S_2}, \boldsymbol{w}) = & \sum_{v=1}^{m} \Big(- w_v tr(\boldsymbol{S}\hat{\boldsymbol{P}}^v) + \beta w_v log w_v \Big) \\
& + \frac{\mu_1}{2} \|\boldsymbol{S} - \boldsymbol{S_1} + \frac{\boldsymbol{\Lambda_1}}{\mu_1}\| + \frac{\mu_2}{2} \|\boldsymbol{S} - \boldsymbol{S_2} + \frac{\boldsymbol{\Lambda_2}}{\mu_2}\| \\
& s.t.\ diag(\boldsymbol{S_2}) = \boldsymbol{0}, \boldsymbol{S_1}^{\mathsf{T}}\boldsymbol{1} = \boldsymbol{1}, \boldsymbol{S_2} = \boldsymbol{S_2}^{\mathsf{T}}, \boldsymbol{S_2} \geq \boldsymbol{0}, \\
& \quad \boldsymbol{w}^{\mathsf{T}}\boldsymbol{1} = 1, 0 \leq w_v \leq 1.
\end{aligned}
$$

Here, $\|\cdot\|$ is the Frobenius-norm; μ_1 and μ_2 are the penalty scalars and $\boldsymbol{\Lambda_1}$ and $\boldsymbol{\Lambda_2}$ are the corresponding Lagrangian multipliers.

Initialization of $\boldsymbol{S}$, $\boldsymbol{S_1}$, $\boldsymbol{S_2}$ and $\boldsymbol{w}$. The weight of multi-omics data , $\boldsymbol{w}$, is initialized as an uniform distribution vector; i.e.,

$$
\boldsymbol{w} = (\frac{1}{m}, \frac{1}{m}, ..., \frac{1}{m}),
$$

where m is the number of omic. The similarity matrices $\boldsymbol{S}$, $\boldsymbol{S_1}$ and $\boldsymbol{S_2}$ are initialized as zero matrix.

Step 1: *fixing* $\boldsymbol{S_1}, \boldsymbol{S_2}$ *and* $\boldsymbol{w}$ *to update* $\boldsymbol{S}$. The following objective should be optimized:

$$\min_{\boldsymbol{S}} \sum_{v=1}^{m} (-w_v tr(\boldsymbol{S}\hat{\boldsymbol{P}}^v)) + \frac{\mu_1}{2} \|\boldsymbol{S} - \boldsymbol{S_1} + \frac{\boldsymbol{\Lambda_1}}{\mu_1}\| + \frac{\mu_2}{2} \|\boldsymbol{S} - \boldsymbol{S_2} + \frac{\boldsymbol{\Lambda_2}}{\mu_2}\|.$$

Differentiating the objective function with respect to $\boldsymbol{S}$ and then setting the derivative to zero:

$$\boldsymbol{S} = \frac{\mu_1 \boldsymbol{S_1} - \boldsymbol{\Lambda_1} + \mu_2 \boldsymbol{S_2} - \boldsymbol{\Lambda_2} + \sum_{v=1}^{m} (w_v \hat{\boldsymbol{P}}^v)}{\mu_1 + \mu_2},$$

Step 2: *fixing* $\boldsymbol{S}, \boldsymbol{S_2}$ *and* $\boldsymbol{w}$ *to update* $\boldsymbol{S_1}$. Since each row of $\boldsymbol{S_1}$ is independent in this optimization framework, we can parallel the solution by solving each row of $\boldsymbol{S_1}$:

$$\frac{\mu_1}{2} \|\boldsymbol{s}_i - (\boldsymbol{s_1})_i + \frac{\boldsymbol{\Lambda_1}}{\mu_1}\|. \quad s.t. \ (\boldsymbol{s_1})_i^{\top} \mathbf{1} = 1$$

According to KKT condition, the optimal solution can be attained [2].

Step 3: *fixing* $\boldsymbol{S_1}, \boldsymbol{S},$ *and* $\boldsymbol{w}$ *to update* $\boldsymbol{S_2}$. The following objective should be optimized:

$$\begin{aligned} \boldsymbol{S_2} = & arg \min_{\boldsymbol{S_2}} \frac{\mu_2}{2} \|\boldsymbol{S} - \boldsymbol{S_2} + \frac{\boldsymbol{\Lambda_2}}{\mu_2}\|. \\ & s.t. \quad diag(\boldsymbol{S_2}) = \mathbf{0}, \boldsymbol{S_2} = \boldsymbol{S_2}^{\top}, \boldsymbol{S_2} \geq \mathbf{0} \end{aligned}$$

This problem has a closed form solution given as follows: $\boldsymbol{S_2} = [(\widetilde{\boldsymbol{S}} + \widetilde{\boldsymbol{S}}^{\top})/2]_+$, where $\widetilde{\boldsymbol{S}} = (\boldsymbol{S} - \frac{\boldsymbol{\Lambda_2}}{\boldsymbol{\mu_2}}) - Diag(diag(\boldsymbol{S} - \frac{\boldsymbol{\Lambda_2}}{\boldsymbol{\mu_2}}))$. We define $[\widetilde{\boldsymbol{S}}]_+ = max(0, \widetilde{\boldsymbol{S}})$ which gives the nonnegative part of the matrix.

Step 4: *fixing* $\boldsymbol{S}, \boldsymbol{S_1}$ *and* $\boldsymbol{S_2}$ *to update* $\boldsymbol{w}$. We need to solve the following problem:

$$\begin{aligned} & \min_{\boldsymbol{w}} \sum_{v=1}^{m} \big(- w_v tr(\boldsymbol{S}\hat{\boldsymbol{P}}^v) + \beta w_v log w_v \big). \\ & s.t. \ \boldsymbol{w}^{\top} \mathbf{1} = 1, 0 \leq w_v \leq 1 \end{aligned}$$

The Lagrangcian function is

$$\mathcal{L}(\boldsymbol{w}) = \sum_{v=1}^{m} \big(- w_v tr(\boldsymbol{S}\hat{\boldsymbol{P}}^v) + \beta(w_v log w_v) \big) - \delta(\boldsymbol{w}^{\top}\mathbf{1} - 1) - \boldsymbol{\sigma}^{\top}\boldsymbol{w}.$$

The optimal solution holds when

$$\frac{\partial \mathcal{L}(\boldsymbol{w})}{\partial \boldsymbol{w}_v} = tr(\boldsymbol{S}\hat{\boldsymbol{P}}^v) - \beta(1 + log w_v) + \delta + \sigma_v = 0.$$

Along with the constraint $\boldsymbol{w}^\top \mathbf{1} = 1$, we can obtain a closed-form solution for the weights given by

$$\boldsymbol{w}_v = \frac{exp(\frac{tr(\boldsymbol{S}\hat{\boldsymbol{P}}^v)}{\beta})}{\sum_v exp(\frac{tr(\boldsymbol{S}\hat{\boldsymbol{P}}^v)}{\beta})}.$$

Step 5: *update multipliers and penalty scalars*. The multipliers and penalty scalars can be updated with the following rule:

$$\begin{cases} \boldsymbol{\Lambda_1} = \boldsymbol{\Lambda_1} + \mu_1(\boldsymbol{S} - \boldsymbol{S_1}) \\ \boldsymbol{\Lambda_2} = \boldsymbol{\Lambda_2} + \mu_2(\boldsymbol{S} - \boldsymbol{S_2}) \\ \mu_1 = \min(1.2 * \mu_1, 10^6) \\ \mu_2 = \min(1.2 * \mu_2, 10^6). \end{cases}$$

SFHOP iterates steps above until reaching the convergence criterions: $\|\boldsymbol{S} - \boldsymbol{S_1}\|_\infty < 10^{-4}$ and $\|\boldsymbol{S} - \boldsymbol{S_2}\|_\infty < 10^{-4}$.

References

1. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J., et al.: Distributed optimization and statistical learning via the alternating direction method of multipliers. Foundations and Trends® in Machine learning **3**(1), 1–122 (2011)
2. Nie, F., Cai, G., Li, X.: Multi-view clustering and semi-supervised classification with adaptive neighbours. In: Thirty-First AAAI Conference on Artificial Intelligence (2017)