Supplentary Material

Paper ID: 2689

Proof of Proposition 1

Proof. Note that

$$\max_{\boldsymbol{O},W} \ell(\boldsymbol{O}, W) = \max_{\mathcal{C}_{\boldsymbol{O}}} \max_{W} \ell(\mathcal{C}_{\boldsymbol{O}}, W).$$

Given $\mathcal{C}_{\boldsymbol{O}}$, the optimization problem over W can be solved by optimizing $W_1^1, \dots, W_N^1, W_1^2, \dots, W_N^2$ separately. Due to the constraints w.r.t. W_i^v, the solution $W_i^{v\star}$ is the indicator vector of $\mathcal{N}_{\mathcal{C}_{\boldsymbol{O}}}^t(\mathbf{x}_i^v)$, i.e., $W_{ij}^v = 0$ for $j \notin \mathcal{N}_{\mathcal{C}_{\boldsymbol{O}}}^t(\mathbf{x}_i^v)$ and $W_{ij}^v = 1$ for $j \in \mathcal{N}_{\mathcal{C}_{\boldsymbol{O}}}^t(\mathbf{x}_i^v)$. Thus, we can get:

$$\begin{aligned}
&\max_{\mathcal{C}_{\boldsymbol{O}}} \max_{W} \ell(\mathcal{C}_{\boldsymbol{O}}, W) \\
&= \max_{\mathcal{C}_{\boldsymbol{O}}} \sum_{i \in \mathcal{C}_{\boldsymbol{O}}} \Big(\sum_{j \in \mathcal{N}_{\mathcal{C}_{\boldsymbol{O}}}^t(\mathbf{x}_i^1)} K_{ij}^2 + \sum_{j \in \mathcal{N}_{\mathcal{C}_{\boldsymbol{O}}}^t(\mathbf{x}_i^2)} K_{ij}^1 \Big) \\
&= \max_{\mathcal{C}_{\boldsymbol{O}}} \sum_{i \in \mathcal{C}_{\boldsymbol{O}}} u_{\mathcal{C}_{\boldsymbol{O}}}^t(X_i).
\end{aligned}$$

Thus, $\mathcal{C}_{\boldsymbol{O}^\star}$ is the optimal solution to Eq. 1. □

Proof of Proposition 2

Proof. Since $\boldsymbol{v}$ is the first eigenvector of D^W, $\sqrt{N_0}\boldsymbol{v}$ is one of the optimal solution to Eq. 9. Replacing $\boldsymbol{O}$ with $\sqrt{N_0}|\boldsymbol{v}|$ in the objective of Eq. 9, we can get:

$$N_0 |\boldsymbol{v}^\top| D^W |\boldsymbol{v}| = N_0 |\boldsymbol{v}^\top D^W \boldsymbol{v}| \geq N_0 \boldsymbol{v}^\top D^W \boldsymbol{v}.$$

The first equality comes from the fact that all elements in D^w are nonnegative. Thus, $\sqrt{N_0}|\boldsymbol{v}|$ is one of the optimal solutions to Eq. 9. □

Dataset Description

Tab. 1 gives description of four datasets used in the experiments from UCI Machine Learning Repository. #Class, #Feature and #Instance denote the number of classes, features and instances in each dataset, respectively.

Tab. 2 gives description of the two multi-view benchmark datasets. #Class, #Instance and #View denote the number of classes, instances and views in each dataset and #Feature denotes the number of features in each view.

CASME dataset contains 195 micro-expressions filmed under 60fps. Each micro- expression is in a video clip (frame sequence), where the onset frame (the first frame of the micro-expression), apex frame (the highest intensity frame of the micro-expression) and offset frame (the last frame of the micro-expression) are recorded. Tab. 2 gives the detailed information about the 8 frame sequences chosen from the CASME dataset. Subject represents the id of the participant. Filename represents the name of the video clip. Onset and offset represents the first and the last frame of the micro-expression, respectively. #Frame represents the number of frames in this frame sequence.

Table 1: Description of datasets from UCI Machine Learning Repository.

	Ionosphere	Vowel	Balance	Zoo
#Class	2	11	3	7
#Feature	34	10	4	16
#Instance	351	990	625	101

Table 2: Description of two real world multi-view datasets.

	NewsM	NewsNG
#Class	2	2
#Feature	2000, 2000, 2000	2000, 2000, 2000
#Instance	1200	500
#View	3	3

Table 3: Description of frame sequences from CASME dataset.

	CAS-1	CAS-2	CAS-3	CAS-4	CAS-5	CAS-6	CAS-7	CAS-8
Subject	1	1	2	3	4	5	7	7
Filename	EP12_3	EP16_3	EP07_1	EP07_1	EP11	EP03_1	EP08_2	EP16_1
Onset	25	53	61	83	102	75	39	83
Offset	58	61	81	96	112	100	59	110
#Frame	127	162	144	178	149	154	143	192