

Supplementary material: Class-Geometry Aware Correlated Joint Subspace Analysis for Multi-View Data

Sankar Mondal and Pradipta Maji

Remark 1. If $k = 1$, then MGSI reduces to GSI.

Proof. When $k = 1$, the only possible value for ℓ is 1. From the definition of MGSI,

$$\begin{aligned} g(x_i, x_i^1, x_i^2, \dots, x_i^k) &= g(x_i, x_i^1) = \begin{cases} 1 & \text{if } \hat{y}_i = \hat{y}_i^1 \\ 0 & \text{otherwise} \end{cases} \\ &= f(x_i, x_i^1). \end{aligned}$$

Therefore,

$$T' = \frac{1}{n} \sum_{i=1}^n g(x_i, x_i^1, x_i^2, \dots, x_i^k) = \frac{1}{n} \sum_{i=1}^n f(x_i, x_i^1) = T.$$

This concludes the proof. □

Remark 2. $T' \in [0, 1]$.

Proof. From the definition of g ,

$$0 \leq g(x_i, x_i^1, x_i^2, \dots, x_i^k) \leq 1, \quad i = 1, 2, \dots, n$$

$$\begin{aligned} \Rightarrow 0 &\leq \sum_{i=1}^n g(x_i, x_i^1, x_i^2, \dots, x_i^k) \leq n \\ \Rightarrow 0 &\leq \frac{1}{n} \sum_{i=1}^n g(x_i, x_i^1, x_i^2, \dots, x_i^k) \leq 1 \\ \Rightarrow 0 &\leq T' \leq 1 \end{aligned}$$

Therefore, $T' \in [0, 1]$. □

Table 1: Descriptions of Data Sets Used in the Study

Data Sets	Sample	Feature Type(Feature Size)	Class
BRCA	134	ω_1 : RNA Sequence (12268) ω_2 : Protein expression (216) ω_3 : miRNA Sequence (178) ω_4 : Copy number segmentation (2258)	4
CRC	261	ω_1 : RNA Sequence (13465) ω_2 : Protein expression(222) ω_3 : miRNA Sequence(236) ω_4 : DNA methylation (293526)	2
LGG	374	ω_1 : RNA Sequence (11973) ω_2 : Protein expression (181) ω_3 : miRNA Sequence (139) ω_4 : DNA methylation (293965) ω_5 : Copy number segmentation (6261)	3
GBM	213	ω_1 : Gene expression (12042) ω_2 : miRNA expression (534) ω_3 : DNA methylation (21422) ω_4 : Copy number segmentation (4070)	5
Cora	2708	ω_1 : Content (1433) ω_2 : Inbound (2708) ω_3 : Outbound (2708) ω_4 : Citation (2708)	7
Reuters	18758	ω_1 : English (21531) ω_2 : French (24892) ω_3 : German (34251) ω_4 : Italian (15506) ω_5 : Spanish (11547)	6
Caltech-20	2386	ω_1 : Gabor (48) ω_2 : Wavelet moments (40) ω_3 : Cenhist (254) ω_4 : Hog (1984) ω_5 : Gist (512) ω_6 : Local binary patterns (LBP) (928)	20
100Leaves	1600	ω_1 : Shape descriptor (64) ω_2 : Fine scale margin (64) ω_3 : Texture histogram (64)	100

Table 2: View Weights Obtained form SGR-wMCCDA

Data sets	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6
BRCA	0.2754	0.2026	0.2688	0.2532	-	-
CRC	0.2510	0.2487	0.2480	0.2523	-	-
LGG	0.2083	0.1989	0.2058	0.1789	0.2081	-
GBM	0.2508	0.2491	0.2424	0.2577	-	-
Cora	0.2498	0.2544	0.2508	0.2449	-	-
Reuters	0.1977	0.1968	0.2098	0.1952	0.2005	-
Caltech-20	0.1879	0.1863	0.1827	0.1268	0.1632	0.1530
100Leaves	0.2795	0.3604	0.3601	-	-	-

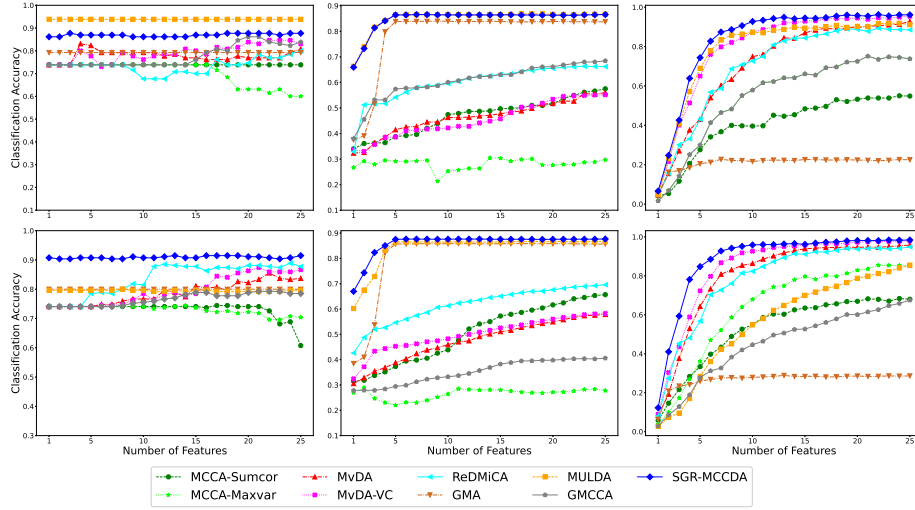


Figure 1: Variation of classification accuracy with respect to the number of extracted features for the proposed (SGR-MCCDA) and existing methods on omics and benchmark data sets (top-row: training-testing, bottom-row: 10-fold CV; left-to-right: CRC, Reuters and 100Leaves).

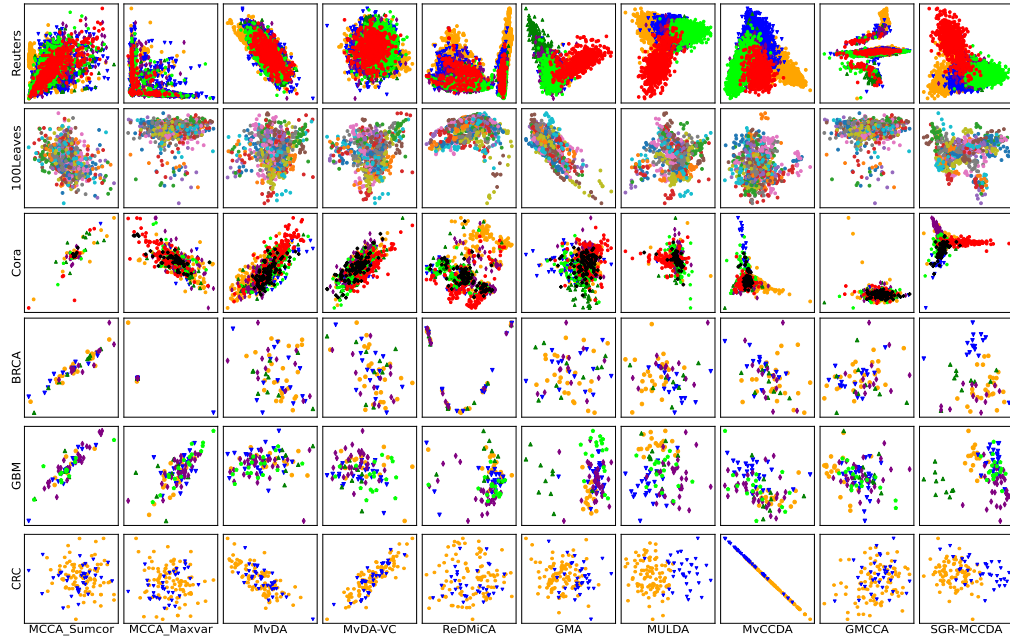


Figure 2: Scatter plot of the proposed method (SGR-MCCDA) and different existing methods, based on the first two extracted features.