Learning Distance Structure for Ordinal Data Clustering

– Experimental Results and Space Complexity Analysis –

Paper ID: 2101

1. Experimental Results with Significance Test: The symbol • indicates that the proposed DSLC performs significantly better than the second best result according to Wilcoxon signed-rank test with 95% confidence interval. All comparative results are averaged by 10 runs of the experiments. The best and second best results are indicated by boldface and underline, respectively. It can be observed that, DSLC significantly outperforms the counterparts on all the data sets in terms of ARI and NMI.

Table 1: ARI performance of the Type-1 approaches and DSLC.

Data	KMDH	KMDC	KMDJ	KMDE	WKMDH	WKMDC	WKMDE	CWO	DSLC
IQ	-0.01±0.02	-0.01±0.01	0.007±0.01	0.044±0.04	-0.01±0.01	-0.02±0.00	0.072±0.09	-0.02±0.02	**0.171±0.12**•
PE	0.105±0.05	0.135±0.09	0.069±0.03	0.222±0.09	0.091±0.08	0.131±0.08	0.166±0.13	0.132±0.10	**0.285±0.03**•
AE	0.140±0.07	0.147±0.08	0.115±0.02	0.270±0.06	0.144±0.07	0.139±0.07	0.279±0.07	0.118±0.05	**0.311±0.10**•
PT	0.073±0.01	0.068±0.01	0.074±0.01	0.078±0.01	0.055±0.02	0.063±0.02	0.045±0.02	0.084±0.00	**0.086±0.00**•
BC	0.015±0.04	0.103±0.07	0.095±0.07	0.042±0.06	0.043±0.06	0.103±0.07	0.057±0.07	0.127±0.07	**0.149±0.07**•
CE	-0.01±0.01	-	0.037±0.03	0.031±0.03	0.010±0.01	-	0.029±0.02	0.013±0.00	**0.072±0.06**•
NS	0.054±0.02	-	0.074±0.03	0.075±0.03	0.084±0.11	-	0.082±0.08	0.003±0.00	**0.150±0.10**•
LE	0.039±0.02	0.034±0.02	0.040±0.02	0.069±0.02	0.038±0.02	0.031±0.01	0.072±0.03	0.050±0.03	**0.081±0.01**•

Table 2: NMI performance of the Type-1 approaches and DSLC.

Data	KMDH	KMDC	KMDJ	KMDE	WKMDH	WKMDC	WKMDE	CWO	DSLC
IQ	0.012±0.01	0.004±0.00	0.014±0.01	0.046±0.04	0.018±0.02	0.003±0.00	0.069±0.07	0.023±0.01	**0.115±0.08**•
PE	0.138±0.07	0.166±0.08	0.095±0.04	0.263±0.09	0.133±0.10	0.176±0.09	0.211±0.12	0.194±0.12	**0.344±0.03**•
AE	0.173±0.08	0.176±0.08	0.124±0.02	0.304±0.05	0.175±0.09	0.171±0.09	0.310±0.06	0.161±0.07	**0.371±0.07**•
PT	0.193±0.02	0.178±0.03	0.184±0.03	0.192±0.02	0.147±0.04	0.159±0.03	0.124±0.05	0.211±0.01	**0.216±0.01**•
BC	0.014±0.01	0.051±0.03	0.051±0.03	0.032±0.03	0.025±0.03	0.050±0.03	0.037±0.03	0.061±0.03	**0.080±0.02**•
CE	0.043±0.02	-	0.075±0.04	0.078±0.04	0.023±0.02	-	0.067±0.04	0.051±0.01	**0.121±0.07**•
NS	0.057±0.02	-	0.077±0.03	0.080±0.03	0.115±0.16	-	0.106±0.09	0.006±0.00	**0.196±0.12**•
LE	0.065±0.02	0.064±0.02	0.075±0.02	0.096±0.02	0.066±0.03	0.058±0.02	0.099±0.03	0.073±0.04	**0.137±0.02**•

Table 3: ARI performance of the Type-2 approaches and DSLC.

Data	KMS	WKMS	NWO	DSLC
IQ	0.090±0.10	0.102±0.11	0.102±0.11	**0.171±0.12**•
PE	0.248±0.05	0.225±0.06	0.248±0.05	**0.285±0.03**•
AE	0.229±0.05	0.234±0.07	0.251±0.07	**0.311±0.10**•
PT	0.084±0.01	0.046±0.02	0.084±0.01	**0.086±0.00**•
BC	0.118±0.04	0.133±0.01	0.136±0.01	**0.149±0.07**•
CE	0.030±0.02	0.024±0.02	0.019±0.02	**0.072±0.06**•
NS	0.110±0.08	0.111±0.11	0.127±0.15	**0.150±0.10**•
LE	0.077±0.02	0.074±0.02	0.067±0.02	**0.081±0.01**•

Table 4: NMI performance of the Type-2 approaches and DSLC.

Data	KMS	WKMS	NWO	DSLC
IQ	0.077±0.08	0.081±0.08	0.081±0.08	**0.115±0.08**•
PE	0.334±0.04	0.310±0.06	0.336±0.06	**0.344±0.03**•
AE	0.309±0.03	0.316±0.05	0.329±0.05	**0.371±0.07**•
PT	0.212±0.01	0.155±0.04	0.209±0.02	**0.216±0.01**•
BC	0.069±0.01	0.072±0.01	0.069±0.01	**0.080±0.02**•
CE	0.085±0.03	0.083±0.04	0.099±0.06	**0.121±0.07**•
NS	0.133±0.09	0.138±0.13	0.159±0.16	**0.196±0.12**•
LE	0.132±0.02	0.127±0.03	0.119±0.03	**0.137±0.02**•

2. Space Complexity Analysis of DSLC: an $n \times m$ matrix $\boldsymbol{X}_{ord}$, an $n \times k$ matrix $\boldsymbol{Q}$, a $k \times m \times V$ matrix $\boldsymbol{U}$, a $1 \times m$ vector $\boldsymbol{W}$, a $V \times m$ matrix $\boldsymbol{L}$, and m matrices $\boldsymbol{M}_1, \boldsymbol{M}_2, ...\boldsymbol{M}_m$, each of which with size $V \times V$, should be maintained during DSLC clustering. Since the adopted $V = \max\{v_1, v_2, ..., v_m\}$ is a small constant in real data sets, overall space complexity of DSLC is $O(nm + nk + km)$.