Learning Distance Structure for Ordinal Data Clustering

– Experimental Results and Space Complexity Analysis –

Paper ID: 2101

1. Space Complexity Analysis of DSLC

Suppose $\boldsymbol{X}_{ord}$ is an ordinal data set with n data objects and m attributes. The m attributes have $v_1, v_2, ..., v_m$ categories, respectively. When the number of clusters is set at k, space complexity of the proposed DSLC clustering algorithm is analyzed as follows:

To simplify the space complexity analysis, we adopt $V = \max(v_1, v_2, ..., v_m)$ in the analysis. During DSLC clustering, 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 according to the Algorithm 1 and corresponding discussions in the paper. V is a small constant in most of the real data sets.

■ Therefore, the overall **space complexity of DSLC is $O(nm+nk+km)$, which is not high** compared to the existing clustering algorithms.

2. Comparative Results with Significance Test

The symbol • indicates that the difference between the proposed DSLC and the compared method is statistically significant using Wilcoxon signed-rank test (at 95% confidence interval, which is a commonly used setting). All the compared methods are run 10 times, and all the demonstrated results are the mean of the 10 runs. The best and second best results are indicated by boldface and underline, respectively.

It can be observed from **Table 1** that DSLC outperforms all the **Type-1** counterparts in terms of ARI, and the performance of DSLC is significantly better in 52 comparisons (60 comparisons in total).

■ Therefore, **DSLC significantly outperforms the Type-1 counterparts in most of the comparisons in terms of ARI**, which is **consistent with** the discussion in the paper that "the clustering performance of DSLC is significantly better than the 8 counterparts".

It can be observed from **Table 2** that DSLC outperforms all the **Type-2** counterparts in terms of ARI, and the performance of DSLC is significantly better in 12 comparisons (24 comparisons in total).

■ Therefore, **DSLC significantly outperforms the Type-2 counterparts in a half of the comparisons in terms of ARI**, which is **consistent with** the discussion in the paper that "DSLC still outperforms the Type-2 approaches, but the advantage is not as significant as in the comparison with Type-1 approaches".

Table 1: ARI performance of the Type-1 approaches and DSLC. DSLC achieves 52 "•" in total.

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: ARI performance of the Type-2 approaches and DSLC. DSLC achieves 12 "•" in total.

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**