Supplementary Materials For Reviewer acan

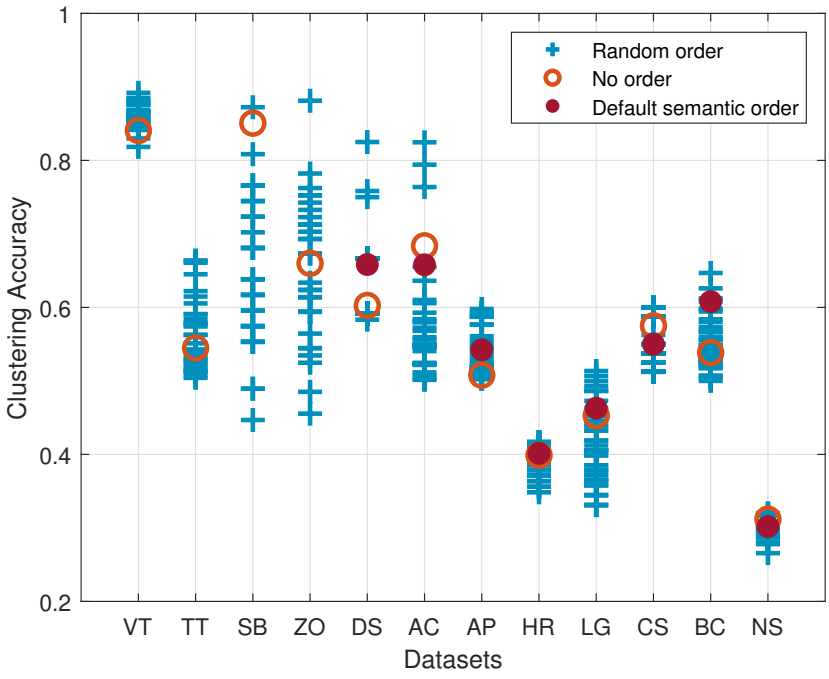

Figure 1: Clustering accuracy of k-modes [1] on four datasets with no semantic order (i.e., VT, TT, SB, ZO), two datasets with obvious semantic order (DS and AC), six mixture datasets (i.e., AP, HR, LG, CS, BC, NS). For an attribute with order, e.g., {accept, neutral, reject}, the original Hamming distance of k-modes is replaced by normalized order distance, e.g., $dist$(accept, neutral) = 0.5, $dist$(neutral, reject) = 0.5, and $dist$(accept, reject) = 1. We also randomly generate attribute order 30 times for all the datasets and show the accuracy under the normalized order distance.

Table 1: Detailed information about 12 datasets. *Num.attribute*, *Num.possible_value*,*Num.mean*, *Num.max*, and *Num.min* are the numbers of attributes, possible values, mean possible values, max possible value, and min possible value, respectively.

Data	*Num.attribute*	*Num.possible_value*	*Num.mean*	*Num.max*	*Num.min*
SB	35	7 2 3 3 2 4 4 2 2 3 2 2 4 4 2 2 2 2 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1	2.09	7	1
NS	8	3 5 4 4 3 2 3 3	3.38	5	2
AP	12	8 5 8 7 8 3 5 6 6 6 3 2	5.58	8	2
DS	5	2 2 2 2 2	2.00	2	2
CS	4	4 3 3 2	3.00	4	2
HR	4	3 4 4 4	3.75	4	3
ZO	15	2 2 2 2 2 2 2 2 2 2 2 2 2 6 2 2 2	2.25	6	2
BC	9	3 3 2 6 2 6 11 7 3	4.78	11	2
LG	18	3 4 8 4 2 2 2 2 2 2 2 3 4 4 8 3 2 2	3.28	8	2
TT	9	3 3 3 3 3 3 3 3 3	3.00	3	3
AC	8	2 2 2 2 3 14 8 3	4.50	14	2
VT	16	3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3	3.00	3	3

REFERENCES

[1] Zhexue Huang. 1998. Extensions to the k-means algorithm for clustering large data sets with categorical values. *Data Mining and Knowledge Discovery* 2, 3 (1998), 283–304.