A Highly-Accurate Approach to Online Sparse Streaming Feature Selection via Supervised Fuzzy Decision

Anonymous

I. Introduction

This is the supplementary file for the paper entitled "*A Highly-Accurate Approach to Online Sparse Streaming Feature Selection via Supervised Fuzzy Decision*". It mainly contains the figures of experimental results.

II. Supplementary Tables

TABLE S(I) Mean number of selected features varying with different algorithms.

Models/Datasets	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
M1	31.50	18.20	35.00	2.50	8.80	25.60	20.00	23.10	19.80	44.90
M2	4.00	3.00	6.50	1.00	2.00	4.60	6.00	5.00	6.00	6.40
M3	19.00	14.00	26.30	2.00	4.00	22.50	9.00	21.00	6.00	11.60
M4	5.00	20.00	75.50	1.00	3.00	74.90	16.00	6.00	24.80	54.20
M5	1.00	1.00	102.20	1.00	16.00	1.00	1.00	1.00	1.00	1.00
M6	56.30	2.00	75.50	13.00	32.20	54.90	25.20	49.10	22.80	31.10
M7	29.00	30.00	33.30	45.00	11.40	12.50	1.00	20.00	4.30	7.90

TABLE S(II) Using the selected features to train a classifier first and then testing its accuracy (%) when missing data rate is 0.1.

Models /Datasets	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	Average	^Rank
M1	**91.12**±0.54	**60.70**±3.35	**92.52**±0.72	**82.21**±1.61	**80.64**±4.55	**89.13**±2.23	**71.99**±4.60	**94.61**±0.65	**93.77**±2.32	**97.44**±0.66	85.41±2.12	1.00
M2	84.33±0.64	54.97±0.69	84.87±2.63	81.90±0.97	80.45±2.59	84.79±2.77	67.56±3.28	92.29±0.58	92.87±1.71	97.41±0.70	82.14±1.66	3.20
M3	85.48±0.63	54.83±0.97	71.08±2.22	79.17±0.82	78.88±2.59	84.40±2.43	56.67±2.79	93.14±0.67	92.47±2.23	97.00±0.89	79.31±1.62	4.30
M4	80.18±0.54	53.79±0.80	88.59±0.49	79.13±1.05	77.65±3.19	83.38±2.32	61.05±5.16	93.34±0.71	93.14±2.18	96.10±0.84	80.64±1.73	4.40
M5	72.18±0.54	49.33±0.54	80.56±3.48	79.15±0.67	78.63±2.66	62.49±2.96	45.57±1.06	85.51±1.02	62.45±5.53	83.21±2.00	69.91±2.05	6.30
M6	86.90±0.51	50.09±0.62	88.59±0.49	82.20±1.27	74.07±5.52	85.47±3.11	71.03±5.22	93.18±0.85	83.00±5.81	92.99±2.29	80.75±2.57	3.80
M7	84.13±0.71	52.68±0.84	81.38±3.67	81.93±1.34	79.82±2.68	86.78±3.00	49.16±1.78	93.83±0.46	93.47±2.96	96.13±1.18	79.93±1.86	3.60

^ The Average rank.

TABLE S(III) The statistical results of the Wilcoxon signed-ranks test on the accuracies recorded in tables S(II) (on all datasets).

Our algorithm vs. Other algorithms	*R+	*R-	† p-value
M2	55	0	9.7656e-04
M3	55	0	9.7656e-04
M4	55	0	9.7656e-04
M5	55	0	9.7656e-04
M6	55	0	9.7656e-04
M7	55	0	9.7656e-04

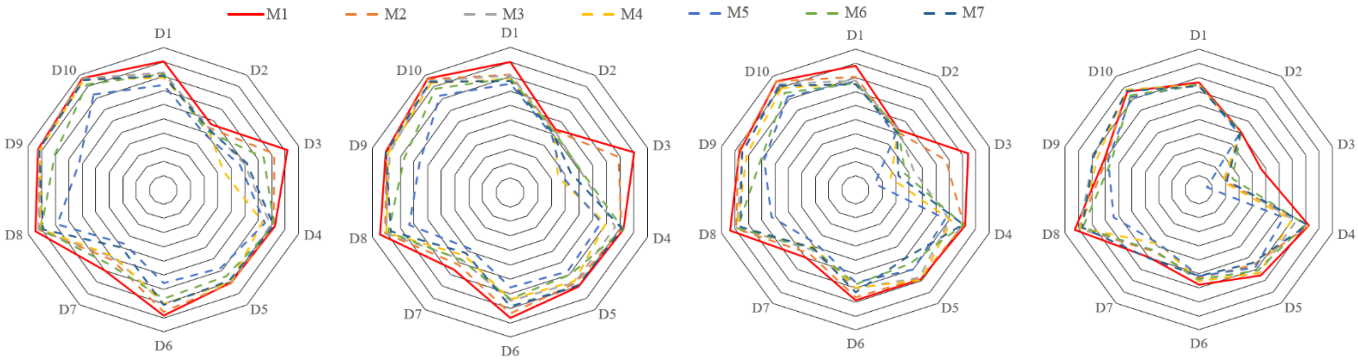
* A larger value denotes a higher accuracy.

† There is no significant difference when ρ-value$\in$[0.1, 0.9] at the 0.1 significance level.

TABLE S(IV) The rank sum of the Wilcoxon signed-ranks on OSFS and OS^{2}FS models.

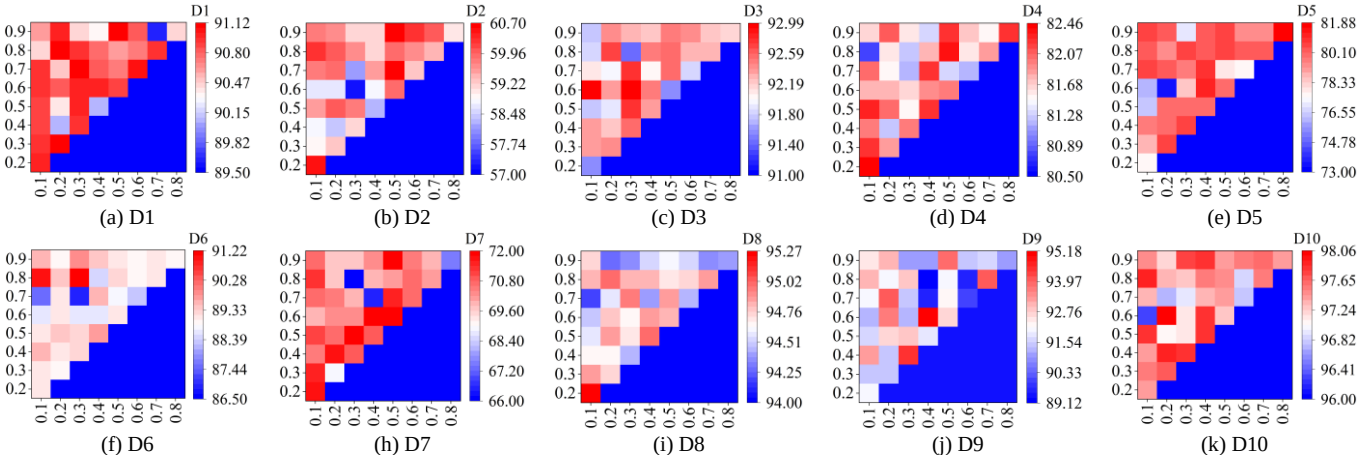
P	M2		M3		M4		M5		M6		M7	
	*R^+	*R^-	*R^+	*R^-	*R^+	*R^-	*R^+	*R^-	*R^+	*R^-	*R^+	*R^-
0.3	53	2	55	0	55	0	55	0	55	0	55	0
0.5	55	0	55	0	55	0	55	0	55	0	55	0
0.7	53	2	55	0	54	1	55	0	55	0	55	0
0.9	52	3	45	10	46	9	55	0	55	0	42	13

III. SUPPLEMENTARY FIGURES



(a) P=0.3 (b) P=0.5 (c) P=0.7 (d) P=0.9

Fig. S1. The average accuracy comparison of OSFS and OS^{2}FS model on each dataset.



(a) D1 (b) D2 (c) D3 (d) D4 (e) D5

(f) D6 (h) D7 (i) D8 (j) D9 (k) D10

Fig. S2. The average accuracy of different parameter α and β.