A Supervised Fuzzy Online Sparse Streaming Feature Selection

Anonymous

I. Introduction

This is the supplementary file for the paper entitled "*A Supervised Fuzzy Online Sparse Streaming Feature Selection*". It mainly contains the figures of experimental results.

II. Supplementary Tables

TABLE S(I) Symbol Annotations

Symbol	Explanation
F	The streaming features set, F={F_1, F_2, …, F_T }.
F_t	The t-th column vector from F, t∈{1, 2, …, T}.
$f_{m,t}$	The F_t of m-th row vector, m∈{1, 2, …, M}.
F'_t	The sparse streaming feature F' of t-th column vector.
$f'_{m,t}$	The F'_t of m-th row vector.
Λ	The known values.
L	The buffer size of B.
w	The inertia weight of PSO-based feature evaluation.
α,β	The threshold value of three-way feature analysis.
$x_{m,k}$	The X of m-th row and k-th column vector.
$y_{j,k}$	The Y of j-th row and k-th column vector.
$\hat{B}_t$	The completed sparse streaming features matrix.
$\hat{F}'_t$	The $\hat{B}$ of t-th column vector.

TABLE S(II) The details of selected datasets.

Mark	Dataset	#(Features)	#(Instances)	#(Class)
D1	USPS	242	1500	2
D2	Madelon	501	2600	6
D3	COIL20	1025	1440	20
D4	Marti1	1025	500	2
D5	Colon	2001	62	2
D6	Lung	3313	83	5
D7	lymphoma	4027	96	9
D8	DriveFace	6401	606	3
D9	ALLAML	7130	72	2
D10	Lungcancer	12534	181	2

TABLE S(III) All the algorithm parameters used in the experiments.

Mark	Algorithm	Parameter
D1	SF-OS^{2}FS	*Z test*, *Alpha is* 0.05, L=200, w=1, β=0.5, and α=0.9.
D2	LOSSA	*Z test*, *Alpha is* 0.05. (TSMC, 2022)
D3	Fast-OSFS	*Z test*, *Alpha is* 0.05. (TPAMI, 2013)
D4	SAOLA	*Z test*, *Alpha is* 0.05. (TKDD, 2016)
D5	SFS-FI	*Z test*, *Alpha is* 0.01, γ=0.01. (TNNLS, 2021)
D6	OFS-A3M	(Information Sciences, 2019)
D7	RHOFS	(TPDS, 2023)

TABLE S(IV) All the classifier parameters used in the experiments.

Classifier	Parameter
KNN	*The neighbors are 3.*
Random Forest	*6 decision trees*.
CART	*Default parameters values.*

TABLE S(V) 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(VI) The statistical results of the Wilcoxon signed-ranks test on the accuracies recorded in Tables S(V) (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 p-value∈[0.1, 0.9] at the 0.1 significance level.

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

ψ	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

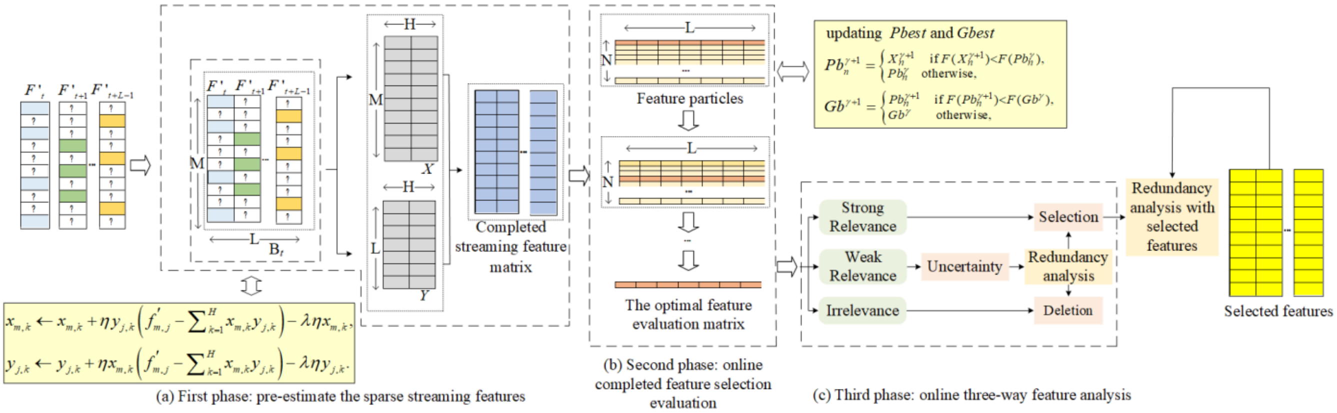


Fig.S1. Flowchart of SF-OS^{2}FS model

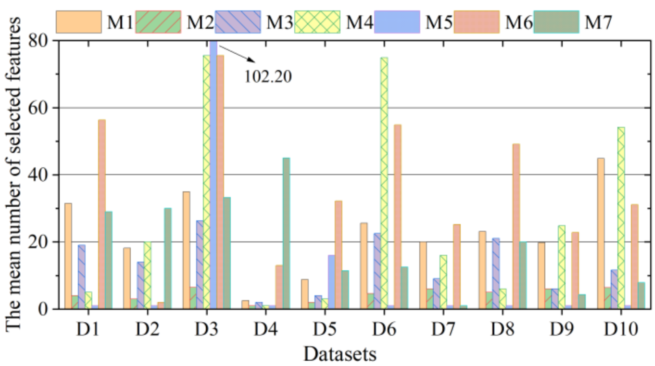
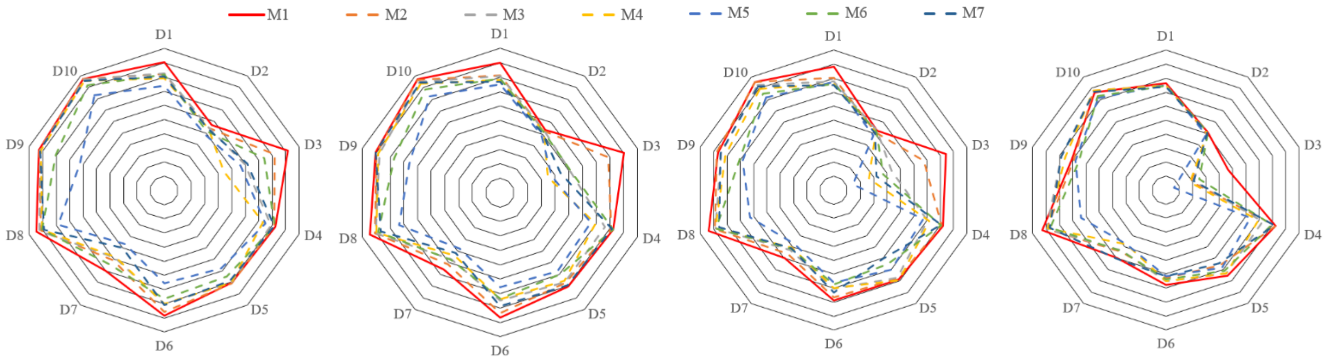


Fig. S2. Mean number of selected features varying with different algorithms.



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

Fig. S3. The average accuracy comparison of OSFS and OS^{2}FS model on each dataset.

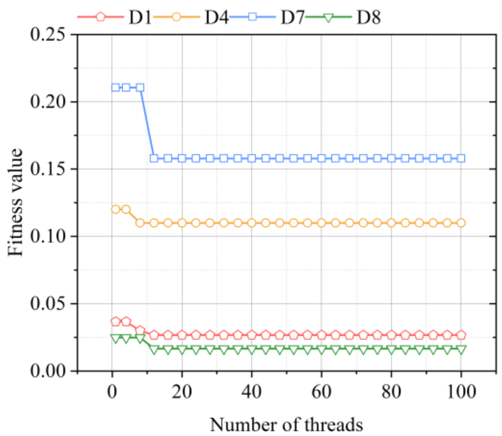


Fig. S4. The convergence curve of PSO-based feature evaluation

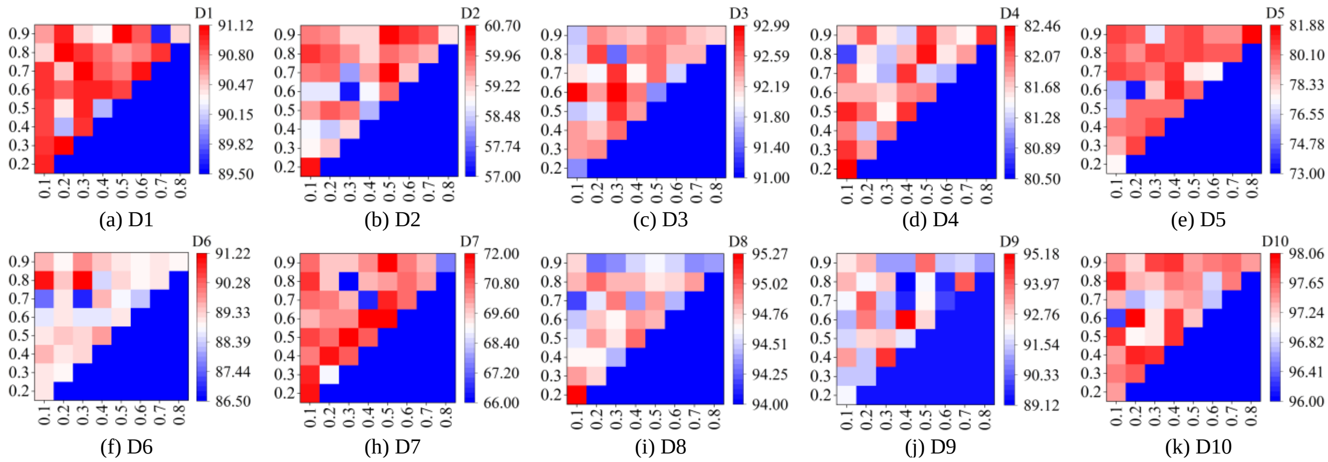


Fig. S5. The average accuracy of different parameter α and β.

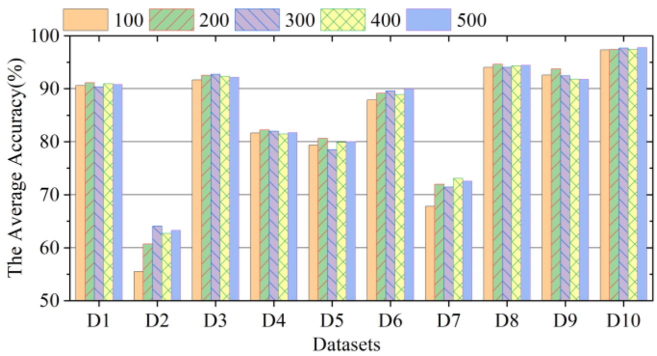


Fig. S6. The average accuracy of different parameter L.

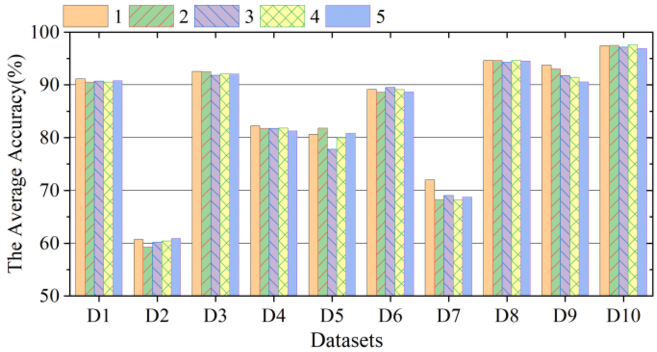


Fig. S7. The average accuracy of different parameter w.

IV. Supplementary Algorithm

Algorithm 1. Pseudocode of finding the optimal feature evaluation matrix

Input $\gamma_{\max}$, c_1, c_2, N, L

Initialize a cluster of particles, X_n (n=1,2, …, N)
Initialize the fitness, Pb_n of particles, and set Gb
for $\gamma=1$ to $\gamma_{\max}$
  for n=1 to N
    for l=1 to L
      update the velocity of particle by (12)
      update the position of particle by (13)
    end for
  end for
  evaluate the fitness of particle using (16)
  update Pb_n and Gb by (14) and (15)
end for

Output The optimal feature evaluation matrix Gb

Algorithm 2. The SF-OS^{2}FS Algorithm

initialize: L, w, α, β
repeat
a new sparse streaming feature F'_t flows in B_t at timestamp t
if $L \neq 0$
  $\mathrm{B}_t = \mathrm{B}_t \cup F'_t$, L=L-1
end if
  repeat
  /* pre-estimate the sparse streaming features */
  while $\Psi_T = maxT$
    for j=t to t+L-1
      for $\forall f'_{m,j} \in \Lambda_t$
        pre-estimate $x_{m,k}$, $y_{j,k}$ following (11)
      end for
    end for $\Psi_T = \Psi_T - 1$
  end while $\hat{\mathrm{B}}_t = XY^{\mathrm{T}}$
  /* online completed feature selection evaluation */
  obtain the Gb according to *Algorithm* 1
  /* online features redundancy analysis */
  for l=1 to L
    fetch gb_l form Gb
    if $gb_l \leq \beta$ discard $\hat{F}'_{t+l\text{-}1}$
    else if $gb_l \geq \alpha$, $S_t = S_t \cup \hat{F}'_{t+l\text{-}1}$
    else accept $\hat{F}'_{t+l\text{-}1} \in S_t$ following (18)
    end if l=l+1
  end for
$S_T = S_T \cup S_t$
for $\forall \hat{F}'_{t+\psi\text{-}1} \in S_T$
  if $\hat{F}'_{t+\psi\text{-}1}$ satisfies (18)
    put $\hat{F}'_{t+\psi,d}$ into S_T
    for each feature $\Omega_F \subseteq S_T - \hat{F}'_{t+\psi\text{-}1}$
      if $\exists \delta \subseteq S_T - \hat{F}'_{t+\psi,d}$ s.t. $Ind\ (C, \Omega_F\,|\delta)$
        $S_T = S_T - \delta$
      end if
    end for
  end if
end for
until no features are available

Output S_T

V. SUPPLEMENTARY APPENDIX

The actions a_P, a_B and a_E categorize the features in set X into the total counts of φ^t_{PP}, φ^t_{BP}, and φ^t_{EP}, respectively. Likewise, these actions also partition the features in set X_C into the total counts of φ^t_{PE}, φ^t_{BE}, and φ^t_{EE}, respectively. Table S(VIII) presents the cardinal number of three-way feature analysis. $\varphi^{t,(2)}_{*}$ ($* \in \{PP, PE, EP, EE\}$) denotes the cardinal number of two-way feature analysis.

Proposition 4. Based on the indicators mentioned earlier, the three-way relevance analysis and the two-way comparison are as follows: *Precision*$^{(2)}$ is less than *Precision*, *Recall*$^{(2)}$ equals *Recall*, *Accuracy*$^{(2)}$ equals *Accuracy*, and $F_1^{(2)}$ is less than F_1.

Proof of Proposition 4. According to Table S(VIII).

a) Given that $\varphi^{t,(2)}_{PP}=\varphi^t_{PP}$ and $\varphi^{t,(2)}_{PE}>\varphi^t_{PE}$, we can conclude:

$$Precision^{(2)} = \frac{\varphi^{t,(2)}_{PP}}{\varphi^{t,(2)}_{PP}+\varphi^{t,(2)}_{PE}} < \frac{\varphi^t_{PP}}{\varphi^t_{PP}+\varphi^t_{PE}} = Precision\ ;$$

b) Considering $\varphi^{t,(2)}_{PP}=\varphi^t_{PP}$, and $\varphi^{t,(2)}_{EP}=\varphi^t_{EP}+\varphi^t_{BP}$, we can derive:

$$Recall^{(2)} = \frac{\varphi^{t,(2)}_{PP}}{\varphi^{t,(2)}_{PP}+\varphi^{t,(2)}_{EP}} = \frac{\varphi^t_{PP}}{\varphi^t_{PP}+\varphi^t_{BP}+\varphi^t_{EP}} = Recall\ ;$$

c) Based on $\varphi^{t,(2)}_{PP}=\varphi^t_{PP}$, and $\varphi^{t,(2)}_{EE}=\varphi^t_{EE}$, with the total number of samples being N, we can state:

$$Accuracy^{(2)} = \frac{\varphi^{t,(2)}_{PP}+\varphi^{t,(2)}_{EE}}{N} = \frac{\varphi^t_{PP}+\varphi^t_{EE}}{N} = Accuracy\ ;$$

d) According to $\varphi^{t,(2)}_{PP}=\varphi^t_{PP}$, $\varphi^{t,(2)}_{EP}=\varphi^t_{EP}+\varphi^t_{BP}$, and $\varphi^{t,(2)}_{PE}>\varphi^t_{PE}$, we can conclude:

$$F_1^{(2)} = \frac{2\varphi^{t,(2)}_{PP}}{2\varphi^{t,(2)}_{PP}+\varphi^{t,(2)}_{EP}+\varphi^{t,(2)}_{PE}} < \frac{2\varphi^t_{PP}}{2\varphi^t_{PP}+\varphi^t_{BP}+\varphi^t_{EP}+\varphi^t_{PE}} = F_1.$$

As a result, the three-way relevance analysis outperforms the two-way analysis in terms of *Precision* and F_1 score, while showing similar performance for *Recall* and *Accuracy*. Thus, *Proposition* 1 holds true. □

TABLE S(VIII) CARDINAL NUMBER.

Action	Cardinal Number	
	X	X_C
a_P	φ^t_{PP}	φ^t_{PE}
a_B	φ^t_{BP}	φ^t_{BE}
a_E	φ^t_{EP}	φ^t_{EE}