Supplementary File of "Clustering Ensemble Based on Fuzzy Matrix Self-Enhancement"

Xia Ji, Jiawei Sun, Jianhua Peng, Yue Pang, Peng zhou*

I. ADDITIONAL EXPERIMENTS

The following content serves as supplementary material to the main manuscript, aimed at providing a more comprehensive understanding of our research. This section firstly compares the clustering performance of different clustering ensemble algorithms across all benchmark datasets based on the ARI metric and provides the results of two statistical tests for the comparison experiment, supplementing the content of Section IV-B in the main manuscript. Additionally, we also provide the ablation study results of the FMSE-v algorithm on the ACC, NMI, and ARI metrics to further support the discussion in Section IV-G of the main manuscript. Through these additional experimental results, we aim to provide readers with a deeper understanding of our research contributions and findings.

A. Comparison Experiment with Other Clustering Ensemble Methods

In this subsection, we provide the clustering performance of different methods with respect to ARI, along with their corresponding ranking.

TABLE I
THE ARI RESULTED FROM DIFFERENT ALGORITHMS, PRESENTED AS "MEAN ± STANDARD DEVIATION"

Method	Seeds	Heart	Ecoli	Climate	Diabetes	WDBC
HGPA [*MLR*, 2002]	0.7061 ± 0.0188	0.0421 ± 0.0157	0.3258 ± 0.0310	0.0019 ± 0.0023	0.0003 ± 0.0044	0.1576 ± 0.1125
PTA-CL [*TKDE*, 2016]	0.3313 ± 0.1453	0.0283 ± 0.0275	0.3341 ± 0.0303	0.0105 ± 0.0603	0.0212 ± 0.0317	0.2338 ± 0.2109
CESC [*CC*, 2021]	0.6332 ± 0.1406	0.0304 ± 0.0280	0.5176 ± 0.0996	−0.0043 ± 0.0230	0.0238 ± 0.0224	0.3948 ± 0.1622
CMSE [*TNNLS*, 2023]	0.5665 ± 0.1220	0.0379 ± 0.0227	**0.7134 ± 0.1106**	0.0098 ± 0.0433	0.0005 ± 0.0036	0.3039 ± 0.2373
HC-a [*FUZZ-IEEE*, 2014]	0.2552 ± 0.0224	0.0083 ± 0.0109	0.0793 ± 0.0672	0.0127 ± 0.0408	0.0222 ± 0.0206	0.1726 ± 0.1206
HC-c [*FUZZ-IEEE*, 2014]	0.2743 ± 0.0373	0.0124 ± 0.0106	0.1194 ± 0.0701	0.0090 ± 0.0180	0.0198 ± 0.0142	0.1580 ± 0.1034
RPFCM [*FUZZ-IEEE*, 2015]	0.6555 ± 0.0537	0.0223 ± 0.0122	0.4486 ± 0.0385	0.0036 ± 0.0069	0.0002 ± 0.0023	**0.5959 ± 0.0138**
FCSPA [*AAI*, 2008]	0.6687 ± 0.0239	0.0446 ± 0.0032	0.3115 ± 0.0052	0.0003 ± 0.0000	−0.0005 ± 0.0002	0.4291 ± 0.0470
RBEACE_AL [*FSS*, 2021]	0.5455 ± 0.0770	0.0348 ± 0.0246	0.5091 ± 0.0717	0.0050 ± 0.0401	0.0177 ± 0.0208	0.2797 ± 0.2178
RBEACE_CL [*FSS*, 2021]	0.5058 ± 0.0895	0.0361 ± 0.0245	0.4205 ± 0.0655	0.0079 ± 0.0398	0.0046 ± 0.0080	0.2740 ± 0.2077
SCPP-a [*TFS*, 2023]	0.5543 ± 0.1518	0.0310 ± 0.0315	0.5528 ± 0.0787	0.0221 ± 0.0263	0.0170 ± 0.0176	0.3006 ± 0.1770
SCPP-v [*TFS*, 2023]	0.5335 ± 0.1624	0.0225 ± 0.0244	0.5742 ± 0.0792	0.0239 ± 0.0384	0.0171 ± 0.0182	0.3039 ± 0.1847
FMSE-a	**0.7202 ± 0.0208**	**0.0457 ± 0.0169**	**0.6138 ± 0.0224**	**0.0303 ± 0.0365**	**0.0536 ± 0.0353**	**0.4938 ± 0.0593**
FMSE-v	**0.7132 ± 0.0594**	**0.0514 ± 0.0164**	0.6073 ± 0.0283	**0.0224 ± 0.0405**	**0.0509 ± 0.0353**	0.4701 ± 0.0712

Method	CMC	BCW	Waveform	Spambase	SMK	LR
HGPA [*MLR*, 2002]	0.0228 ± 0.0054	0.1530 ± 0.0970	0.1705 ± 0.0601	0.0014 ± 0.0024	0.0499 ± 0.0186	0.1100 ± 0.0069
PTA-CL [*TKDE*, 2016]	0.0178 ± 0.0049	0.2173 ± 0.1764	0.3072 ± 0.0318	0.0979 ± 0.0587	0.0098 ± 0.0268	**0.1853 ± 0.0000**
CESC [*CC*, 2021]	0.0197 ± 0.0101	0.3656 ± 0.1535	0.3138 ± 0.1157	0.0096 ± 0.0165	0.0006 ± 0.0056	0.1126 ± 0.0000
CMSE [*TNNLS*, 2023]	0.0062 ± 0.0142	0.1096 ± 0.1519	0.3357 ± 0.1391	0.0031 ± 0.0072	0.0164 ± 0.0312	N/A
HC-a [*FUZZ-IEEE*, 2014]	0.0115 ± 0.0112	0.1707 ± 0.1443	0.1767 ± 0.0634	0.0405 ± 0.0415	0.0153 ± 0.0149	0.0047 ± 0.0004
HC-c [*FUZZ-IEEE*, 2014]	0.0141 ± 0.0115	0.2053 ± 0.1400	0.1060 ± 0.0553	0.0216 ± 0.0319	0.0182 ± 0.0081	0.0193 ± 0.0077
RPFCM [*FUZZ-IEEE*, 2015]	0.0200 ± 0.0052	**0.5953 ± 0.0129**	0.3364 ± 0.0127	0.1084 ± 0.0037	0.0409 ± 0.0189	0.0072 ± 0.0055
FCSPA [*AAI*, 2008]	0.0204 ± 0.0009	0.4491 ± 0.0190	0.0001 ± 0.0002	**0.1427 ± 0.0011**	0.0400 ± 0.0078	0.0389 ± 0.0235
RBEACE_AL [*FSS*, 2021]	0.0150 ± 0.0059	0.3205 ± 0.1817	0.2344 ± 0.0251	0.0548 ± 0.0506	−0.0014 ± 0.0008	0.0154 ± 0.0022
RBEACE_CL [*FSS*, 2021]	0.0119 ± 0.0092	0.2723 ± 0.1371	0.2158 ± 0.0318	0.0906 ± 0.0556	0.0016 ± 0.0115	0.0168 ± 0.0033
SCPP-a [*TFS*, 2023]	0.0143 ± 0.0106	0.3722 ± 0.1627	0.2213 ± 0.0314	0.0220 ± 0.0066	0.0376 ± 0.0376	0.1074 ± 0.0085
SCPP-v [*TFS*, 2023]	0.0203 ± 0.0084	0.3759 ± 0.1162	0.2407 ± 0.0265	0.0092 ± 0.0033	0.0093 ± 0.0233	0.0981 ± 0.0166
FMSE-a	**0.0263 ± 0.0107**	**0.5103 ± 0.0430**	**0.3653 ± 0.0018**	**0.1361 ± 0.0497**	**0.0532 ± 0.0193**	0.1128 ± 0.0168
FMSE-v	**0.0272 ± 0.0061**	0.4987 ± 0.0513	**0.3642 ± 0.0033**	0.1196 ± 0.0559	**0.0524 ± 0.0157**	**0.1134 ± 0.0260**

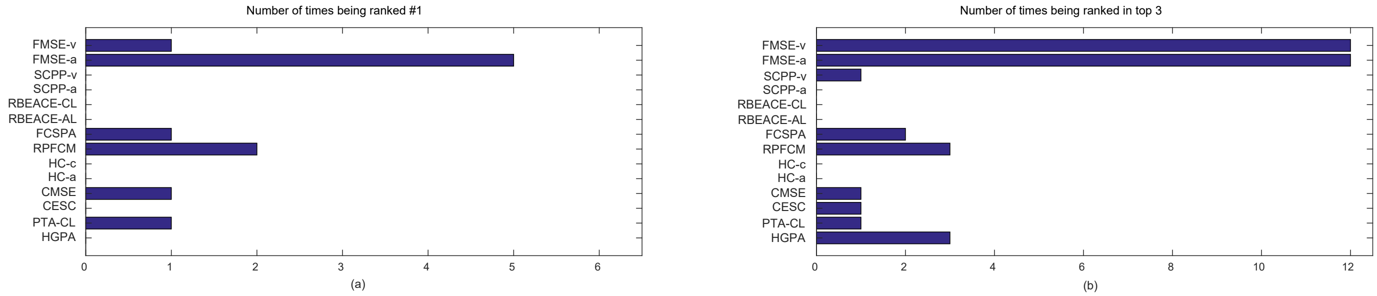

Fig. 1. Number of times that each method is ranked in the (a) first position and (b) top three with respect to Table I.

B. Statistical Test

In this subsection, we conducted statistical tests to better analyze the experimental results in Tables III and IV, including Significance comparison [1] and the Nemenyi test [2]. Significance comparison assesses the superiority of algorithms by comparing the difference in the number of "significantly better" and "significantly worse" methods between different methods. For methods a and b with t experiments on dataset k, method a is significantly better than method b if the following inequality holds:

$$\mu_{(a,k)} - 1.96\frac{\delta_{(a,k)}}{\sqrt{t}} > \mu_{(b,k)} - 1.96\frac{\delta_{(b,k)}}{\sqrt{t}} \quad (1)$$

where μ represents the mean value and δ represents the deviation value. The results of the significance comparison are shown in Fig.2. In Fig.2, each bar corresponds to a method, with higher bars indicating that the corresponding method obtains "significantly better" more times and "significantly worse" fewer times. From Fig.2, the proposed two FMSE methods have the highest two bars for both ACC and NMI.

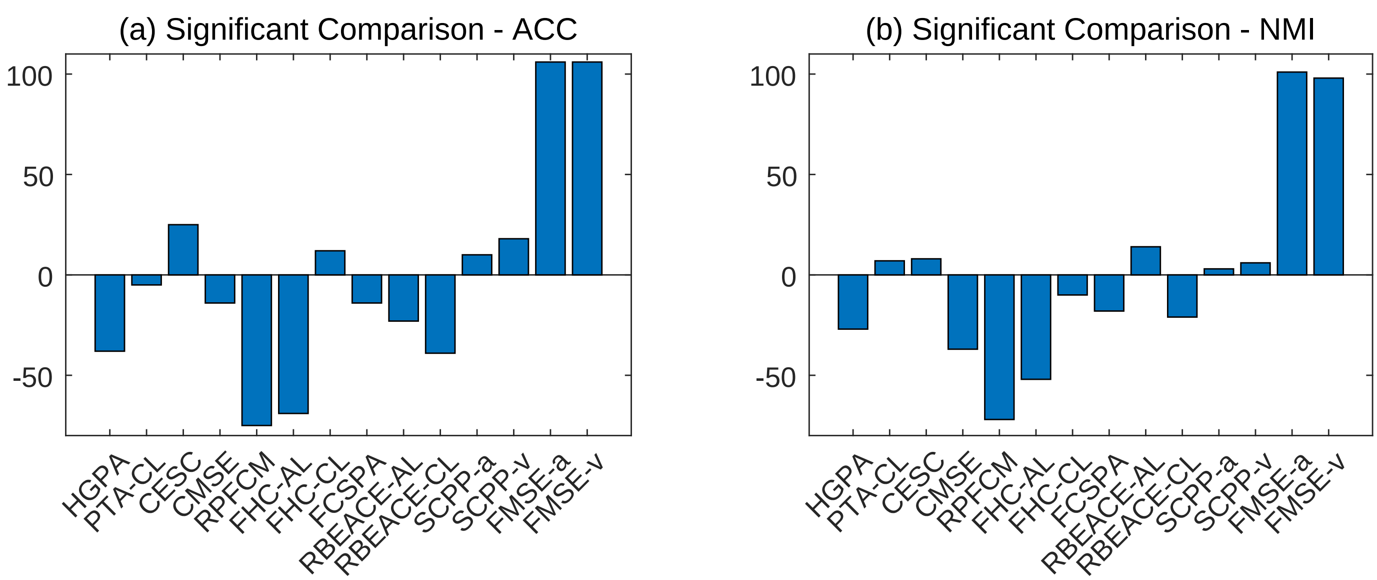

Fig. 2. The Significant comparison on Table III and Table IV.

The Nemenyi test is a non-parametric test that indicates the difference between the mean rankings, and the critical difference (CD) of the Nemenyi test is calculated as follows:

$$\text{CD} = q_a\sqrt{\frac{k'(k'+1)}{6N}} \quad (2)$$

where the confidence level $\alpha = 0.10$, $q_a = 2.920$, N is the number of datasets, and $k^{'}$ is the number of comparison methods. If the difference in average ranks between two methods exceeds the CD, the method with the smaller rank is significantly better than the other method. The results of the Nemenyi test are shown in Fig.3. In Fig.3, each red dot represents the average rank of the corresponding method, the length of the blue line is the CD, and the vertical coordinate of the black dashed line is: $\max\{\text{AVE-r(FMSE-a)}, \text{AVE-r(FMSE-v)}\} + \frac{\text{CD}}{2}$. From Fig.3, the average rank of the FMSE method is significantly better than the comparison methods.

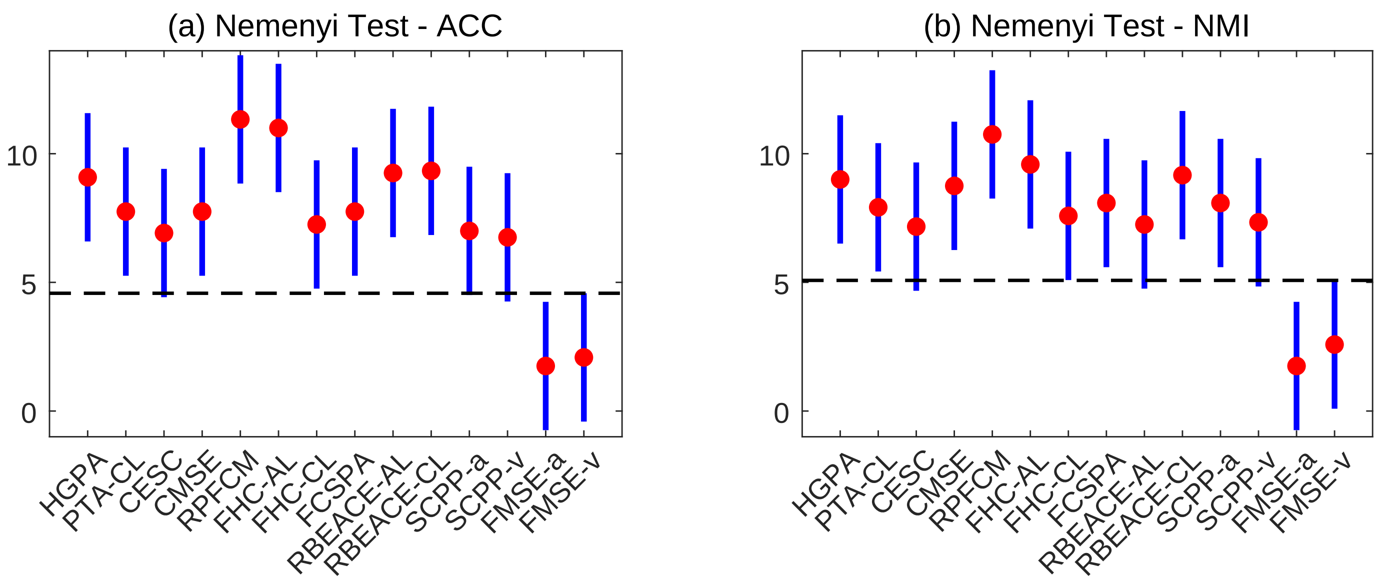

Fig. 3. The Nemenyi test on Table III and Table IV.

C. Ablation Study

In this subsection, we provide the ablation experimental results of the proposed FMSE-v algorithm on the metrics ACC, NMI and ARI.

TABLE II
ABLATION STUDY IN ACC, NMI AND ARI FOR THE PROPOSED FMSE-V

Metric	ACC			NMI			ARI		
Algorithm	w/o SVD	w/o FRI	FMSE-v	w/o SVD	w/o FRI	FMSE-v	w/o SVD	w/o FRI	FMSE-v
Seeds	0.8378	0.7779	**0.8862**	0.6684	0.6227	**0.6945**	0.6692	0.5766	**0.7132**
Heart	0.5706	0.6080	**0.6220**	0.0289	0.0301	**0.0359**	0.0163	0.0424	**0.0514**
Ecoli	0.5802	0.7077	**0.7134**	0.5047	**0.5790**	0.5583	0.3824	0.5754	**0.6073**
Climate	0.5670	0.8827	**0.8837**	**0.0132**	0.0130	0.0130	0.0046	0.0041	**0.0224**
Diabetes	**0.6207**	0.5763	0.6135	0.0187	0.0171	**0.0202**	0.0415	0.0325	**0.0509**
WDBC	0.7970	0.7559	**0.8508**	0.3936	0.2771	**0.4669**	0.3945	0.2836	**0.4701**
CMC	0.4193	0.4188	**0.4295**	0.0255	**0.0310**	0.0305	0.0201	0.0236	**0.0272**
BCW	0.8270	0.7480	**0.8562**	0.3922	0.2064	**0.4145**	0.4530	0.2040	**0.4987**
Waveform	0.5047	0.6353	**0.6363**	**0.3782**	0.3036	0.3090	0.2903	0.3631	**0.3642**
Spambase	0.6416	0.6756	**0.6791**	0.0188	0.0730	**0.0805**	0.0223	0.1049	**0.1196**
SMK	0.6180	0.6100	**0.6225**	0.0406	0.0359	**0.0464**	0.0403	0.0423	**0.0524**
LR	0.2495	**0.2525**	0.2480	0.3342	0.3359	**0.3467**	**0.1258**	0.1164	0.1134

REFERENCES

[1] N. Iam-On and T. Boongoen, “Comparative study of matrix refinement approaches for ensemble clustering,” *Machine Learning*, vol. 98, pp. 269–300, 2015.
[2] J. Demšar, “Statistical comparisons of classifiers over multiple data sets,” *The Journal of Machine learning research*, vol. 7, pp. 1–30, 2006.