Supplementary Materials To: "fMBN-E: Efficient Unsupervised Network Structure Ensemble and Selection for Clustering"

Xiao-Lei Zhang

Abstract

This supplementary material is the full version of Section V-H in the main text. It discusses the experimental effect of the candidate meta functions of MBN-E, and candidate ensemble selection methods of MBN-SO on performance.

I. Discussions

A. On candidate meta-clustering functions of MBN-E

MBN-E concatenates the learned representations from the MBN base models as a new meta-representation for clustering, while a conventional clustering ensemble method usually uses a meta-clustering function to fuse the predictions produced from a number of base clusterings. From the perspective of ensemble learning, we may also adopt other candidate meta-clustering functions to fuse the clustering results of the MBN base models. In this section, we study the effect of the meta-clustering approaches on performance.

We adopted 12 meta-clustering functions, which are CSPA [1], HGPA [1], MCLA [1], DREC [2], LinkClueE [3], [11], ARA1 [4], ARA2 [4], Borda [5], Cvote [6], Vote [7], ECPCS_MC [8], and ECPCS_HC [8], respectively. The predictions of data for the meta-clustering functions here is obtained by applying agglomerative hierarchical clustering to the learned representations of the MBN base models.

Table I lists the comparison results of the standard MBN-E and 12 meta-clusterings that use the same MBN base models. From the table, we find that the proposed MBN-E ranks the second place, which is slightly worse than Vote [7]. If we look at the details, we find that MBN-E performs only 0.1% worse than Vote on Dermatology, COIL20, and MNIST, which accounts for the inferiority of MBN-E over Vote. We further observe that MBN-E wins the best performance on three datasets, which has the same highest number of championships as ECPCS_MC [8]. To summarize, considering the "Occam's Razor" as the principle for designing algorithms, the simple MBN-E is recommended as the best choice of fusing multiple MBN base models.

If we further compare the results in Table I with MBN†, we find that none of the 13 comparison methods achieve comparable performance with MBN†—one of the base models that has been applied to all of the comparison methods. This phenomenon suggests that, if we could find MBN† from the candidate base models, then the performance could at least outperform the comparison methods, which motivates the invention of MBN-SO and MBN-SD.

B. On candidate ensemble selection methods of MBN-SO

MBN-SO simply selects the MBN base models with the highest weights. In literature, there are many studies on how to select the base models given the weights, which may lead to higher performance and lower computational power than the proposed method.

This section applies five representative clustering ensemble selection functions to MBN-SO, given the same MBN base models. They can be categorized into two classes. The first class conducts the ensemble selection according to the clustering results of the base models only. It consists of the sum of the normalized mutual information (SNMI) [9], joint criterion (JC) [9], and cluster and select (CAS) [9]. The selection criteria of the methods consider both the accuracy and diversity of the clustering results.

The second class [10] picks the base models according to an optimization-like criterion, which is closely related to the proposed MBN-SO. Here we compare with the following representative ones:

- **Single index selection (SIS) [10]**: Contrary to MBN-SO which uses the predicted label from MBN-E as a reference to evaluate the discriminability of the output representation of each base model, SIS uses the predicted label from each base clustering as a reference to evaluate the discriminability of the original data representation, and uses a

Xiao-Lei Zhang is with the Research & Development Institute of Northwestern Polytechnical University in Shenzhen, Shenzhen, China, and the School of Marine Science and Technology and the Center for Intelligent Acoustics and Immersive Communications, Northwestern Polytechnical University, Xi'an, China. E-mail: xiaolei.zhang@nwpu.edu.cn.

TABLE I
ACC COMPARISON BETWEEN MBN-E AND THE META-CLUSTERING FUNCTIONS THAT USE THE SAME MBN BASE MODELS AS MBN-E. THE ABBREVIATIONS "DERM.", "NT", "YALE B", AND "20-NG" ARE SHORT FOR DERMATOLOGY, NEW-THYROID, EXTENDED-YALE B, AND 20-NEWSGROUPS, RESPECTIVELY. THE TERM "N/A" MEANS THAT A SINGLE RUN CANNOT BE FINISHED IN 24 HOURS.

	Dermatology	New-Thyroid	UMIST	Extended-Yale B	COIL20	COIL100	20-Newsgroups	MNIST	**Rank**
CSPA [1]	0.721	0.491	0.592	0.966	0.816	0.677	0.581	0.106	8.125
HGPA [1]	0.306	0.698	0.083	0.027	0.050	0.010	0.053	0.113	12.000
MCLA [1]	0.791	**0.949**	0.602	0.961	0.830	0.726	0.586	**0.965**	5.125
DREC [2]	0.669	0.777	0.500	0.684	0.619	0.545	0.401	N/A	10.875
LinkClueE [3]	0.891	0.948	0.651	0.917	0.894	0.796	N/A	N/A	5.875
ARA1 [4]	0.866	0.897	0.587	0.921	0.837	0.586	0.578	N/A	7.750
ARA2 [4]	0.848	0.937	0.431	0.834	0.757	0.399	0.494	N/A	9.875
Borda [5]	0.922	0.940	0.539	0.888	0.656	0.536	0.516	**0.965**	7.375
Cvote [6]	0.685	0.683	0.631	0.965	**0.981**	0.831	0.204	**0.965**	5.750
Vote [7]	0.867	0.880	0.649	0.968	0.930	0.825	0.618	**0.965**	**3.250**
ECPCS_MC [8]	**0.935**	0.940	0.598	0.947	0.884	0.784	**0.633**	**0.965**	4.125
ECPCS_HC [8]	0.852	0.943	0.597	0.816	0.857	0.765	0.431	0.694	7.000
MBN-E	0.866	0.860	**0.670**	**0.973**	0.929	**0.832**	0.584	0.964	3.875
MBN†	0.971	0.964	0.770	0.969	0.994	0.901	0.623	0.965	

TABLE II
ACC COMPARISON BETWEEN MBN-SO AND THE CLUSTERING ENSEMBLE SELECTION FUNCTIONS THAT USE THE SAME CANDIDATE MBN BASE MODELS AS MBN-SO.

	Dermatology	New-Thyroid	UMIST	Extended-Yale B	COIL20	COIL100	20-Newsgroups	MNIST	**Rank**
SNMI [9]	0.708	0.485	0.555	0.823	0.726	0.608	0.534	0.106	15.375
JC [9]	0.746	0.537	0.546	0.947	0.873	0.800	0.556	0.106	11.250
CAS [9]	0.734	0.479	0.560	0.940	0.698	0.617	0.462	0.106	14.250
SIS (SWC) [10]	0.686	0.528	0.559	0.929	0.880	0.776	0.544	0.106	13.000
SIS (PB) [10]	0.682	0.494	0.572	0.930	0.898	0.771	0.544	0.106	12.875
SIS (PBM) [10]	0.658	0.486	0.587	0.910	0.892	0.808	0.483	0.106	13.250
SIS (VRC) [10]	0.643	0.522	0.634	0.909	0.963	0.809	0.545	0.106	11.125
SR [10]	0.645	0.509	0.567	0.924	0.889	0.790	0.532	0.106	13.625
MBN-SO (SWC)	0.854	0.859	0.717	**0.968**	0.957	0.857	0.602	0.964	**4.500**
MBN-SO (PB)	0.851	0.880	0.699	0.960	0.956	0.884	0.591	0.964	5.250
MBN-SO (PBM)	0.852	0.630	0.718	0.961	0.990	0.866	0.602	0.962	4.750
MBN-SO (VRC)	0.714	0.771	**0.767**	0.941	**0.995**	**0.908**	**0.623**	0.964	4.750
MBN-SD	0.849	**0.940**	0.519	0.891	0.958	0.760	0.607	0.841	9.750
rSNMI	0.730	0.565	0.552	0.949	0.873	0.796	0.556	0.106	11.500
rMBN-SO (SWC)	0.867	0.885	0.625	0.966	0.920	0.823	0.611	**0.965**	5.125
rMBN-SO (PB)	0.806	0.938	0.656	0.934	0.965	0.852	0.617	**0.965**	4.625
rMBN-SO (PBM)	**0.905**	0.937	0.626	0.954	0.953	0.821	0.605	0.964	5.625
rMBN-SO (VRC)	0.855	0.937	0.654	0.945	0.952	0.830	0.611	0.962	5.875

meta-clustering function to fuse the predicted labels from the top B base clusterings into the final prediction result. Because the original data representation is very noisy, we replaced it with the output representation of MBN-E, which improves SIS to a fair experimental setting with MBN-SO. Here we apply the criteria of SWC, PB, PBW, and VRC to SIS for a point-to-point comparison with MBN-SO. Following [10], we used CSPA as the meta-clustering function of SIS.

- **Sum of ranks (SR) [10]** It runs SIS with different optimization-like criteria, each of which produces a ranking of the base models. Then, it averages the rankings for the final ranking of the base models. At last, it uses a meta-clustering function to fuse the predicted labels from the top B base clusterings into the final prediction result. Following [10], we used CSPA as the meta-clustering function of SR.

The top 2 parts of Table II lists the comparison result between MBN-SO and the referenced methods [9], [10]. From the ranking list of the table, we see that the variants of MBN-SO behave similarly with each other, and outperform the referenced methods apparently. The variants of SIS perform similarly as well, which outperform SNMI and CAS, and

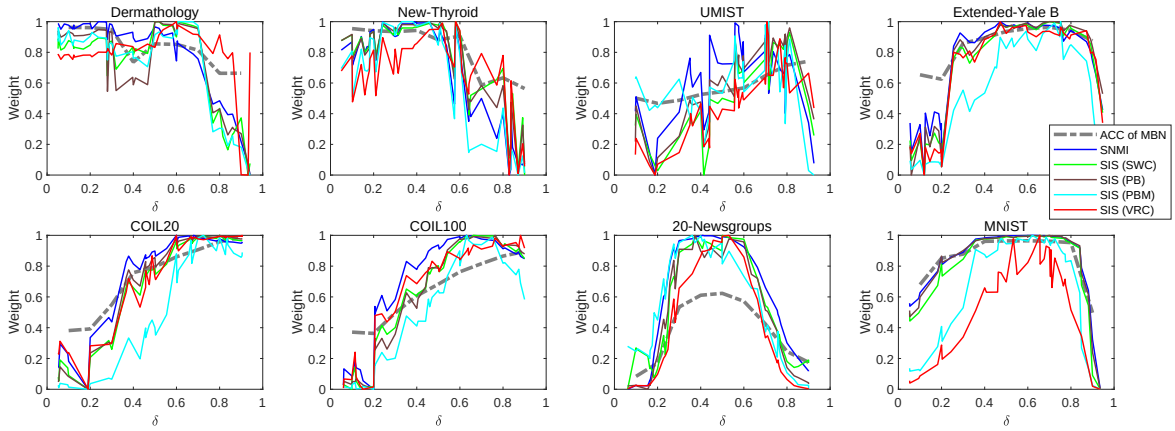

Fig. 1. Weights of the MBN base models of the SNMI and SIS functions.

are inferior to JC. If we look at the details, we find that "MBN-SO (VRC)" achieves the top performance in five out of the eight datasets. As for the referenced methods, most of them do not behave fundamentally different. Particularly, they have failed to achieve reasonable results on MNIST, comparing to random guess.

Fig. 1 shows the weights of the MBN base models of the SNMI and SIS functions in a single run. After comparing the curves of the weights with the clustering accuracy of the MBN base models, we see that although the weights of MBN-SO are more accurate than the weights of the SIS variants, the performance gap between SIS and MBN-SO in Table II seem unnecessarily to be so large.

To investigate why the proposed MBN-SO has such a large advantage over the referenced methods, we first removed the ensemble selection criterion based on diversity in SNMI by simply picking the B base models that have the largest weights. The new method is named *revised SNMI* (rSNMI). From the result in Table II, we see that rSNMI significantly outperforms SNMI and CAS, and performs as good as JC. That is to say, a simple ensemble selection strategy like MBN-SO is enough, while further exploring the diversity between the base models via complicated algorithms is unnecessary.

Then, we replaced the meta-clustering function of SIS by simply concatenating the output representations of the selected base models. Because the only difference between the revised algorithm and MBN-SO is that the revised algorithm uses the data representation produced by MBN-E as a reference to evaluate the clustering quality of each MBN base model, while MBN-SO uses the clustering result of MBN-E as a reference to evaluate the data representation learned by each MBN base model, we name the revised algorithm as *revised MBN-SO* (rMBN-SO). The bottom 2 parts of Table II lists the comparison result between MBN-SO and rMBN-SO. From the apple-to-apple comparison, we see that the ensemble selection strategy of MBN-SO is better than rMBN-SO. By comparing rMBN-SO and SIS, we see that the meta-clustering function is responsible for the large performance gap between MBN-SO and SIS.

REFERENCES

[1] A. Strehl and J. Ghosh, "Cluster ensembles—a knowledge reuse framework for combining multiple partitions," *Journal of Machine Learning Research*, vol. 3, pp. 583–617, 2003.

[2] J. Zhou, H. Zheng, and L. Pan, "Ensemble clustering based on dense representation," *Neurocomputing*, vol. 357, pp. 66–76, 2019.

[3] N. Iam-on, S. Garrett *et al.*, "Linkclue: A matlab package for link-based cluster ensembles," *Journal of Statistical Software*, vol. 36, no. 9, pp. 1–36, 2010.

[4] B. G. Mirkin and A. Shestakov, "Least square consensus clustering: criteria, methods, experiments," in *European Conference on Information Retrieval*. Springer, 2013, pp. 764–767.

[5] X. Sevillano, F. Alías, and J. C. Socoró, "Bordaconsensus: a new consensus function for soft cluster ensembles," in *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, 2007, pp. 743–744.

[6] H. G. Ayad and M. S. Kamel, "On voting-based consensus of cluster ensembles," *Pattern Recognition*, vol. 43, no. 5, pp. 1943–1953, 2010.

[7] E. Dimitriadou, A. Weingessel, and K. Hornik, "A combination scheme for fuzzy clustering," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 16, no. 07, pp. 901–912, 2002.

[8] D. Huang, C.-D. Wang, H. Peng, J. Lai, and C.-K. Kwoh, "Enhanced ensemble clustering via fast propagation of cluster-wise similarities," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2018.

[9] X. Z. Fern and W. Lin, "Cluster ensemble selection," *Statistical Analysis and Data Mining: The ASA Data Science Journal*, vol. 1, no. 3, pp. 128–141, 2008.

[10] M. C. Naldi, A. Carvalho, and R. J. Campello, "Cluster ensemble selection based on relative validity indexes," *Data Mining and Knowledge Discovery*, vol. 27, no. 2, pp. 259–289, 2013.

[11] N. Iam-On, T. Boongoen, S. Garrett, and C. Price, "A link-based approach to the cluster ensemble problem," *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 12, pp. 2396–2409, 2011.