Clustering Ensemble Selection
with Analytic Hierarchy Process

Wei Liu[1], Xiaodong Yue[1,2(✉)], Caiming Zhong[3], and Jie Zhou[4]

[1] School of Computer Engineering and Science, Shanghai University,
Shanghai 200444, China
`ldachuan@outlook.com, yswantfly@shu.edu.cn`
[2] Shanghai Institute for Advanced Communication and Data Science,
Shanghai University, Shanghai, China
[3] College of Science and Technology, Ningbo University, Ningbo 315211, China
`zhongcaiming@nbu.edu.cn`
[4] College of Computer Science and Software Engineering, Shenzhen University,
Guangdong 518060, China
`jie_jpu@163.com`

Abstract. Existing clustering ensemble selection methods adopt internal and external evaluation indexes to measure the quality and diversity of base clusterings. The significance of base clustering is quantified by the average or weighted average of multiple evaluation indexes. However, there exist two limitations in these methods. First, the evaluation of base clusterings in the form of linear combination of multiple indexes lacks the structural analysis and relative comparison between clusterings and measures. Second, the consistency between the final evaluation and the multiple evaluations from different measures cannot be guaranteed. To tackle these problems, we propose a clustering ensemble selection method with Analytic Hierarchy Process (AHPCES). Experimental results validate the effectiveness of the proposed method.

Keywords: Clustering ensemble selection · Analytic hierarchy process · Multiple clustering evaluation indexes

1 Introduction

Clustering ensemble selection creates the final clustering by evaluating and selecting a subset of base clusterings, and it performs as good as even better than full clustering ensemble [6,13,17,18]. Quality and diversity of base clustering are two critical issues of clustering ensemble selection. A subset of base clusterings of both high quality and diversity will produce a precise ensemble result.

Most existing clustering ensemble selection methods adopt internal and external evaluation indexes to measure the quality and diversity of base clusterings. The significance of base clustering is directly quantified by the average

(or weighted average) of multiple evaluation indexes [18]. However, there exist the following problems in these methods. First, the final evaluation of base clusterings is obtained by the linear combination of multiple indexes, which is a single level evaluation assemble and lacks the structural analysis and relative comparison between clusterings and measures. Second, the consistency between the final evaluation of base clusterings and the multiple evaluations from different measures cannot be guaranteed.

To tackle the problems above, we propose a clustering ensemble selection method based on Analytic Hierarchy Process (AHPCES) [15] in this paper. AHP is a structured technique for organizing and analyzing complex group decisions. As to the characteristics of AHP for group decision making, we apply it to formulate the clustering ensemble selection process. Specifically, we design an analytic hierarchy process to evaluate base clusterings, which achieves the hierarchical ensemble of the weights of evaluations from base clustering level to criteria level. Considering the significances of evaluation measures and integrating the weights of base clusterings under multiple measures in hierarchy, AHP-based method can generate more reasonable and consistent evaluations of base clustering results which facilitates the selective clustering ensemble.

The rest of this paper is organized as follows. Section 2 introduces the proposed method in detail, which includes comparison matrix construction, weight computation and algorithm implementation. In Sect. 3, experiment results validate the effectiveness of the AHP-based selection for clustering ensemble.

2 Clustering Ensemble Selection with AHP

The framework of the proposed AHPCES consists of three levels. On the bottom level of base clusterings, all the base clusterings are measured with multiple evaluation indexes under the criteria of quality and diversity; On the second level of evaluation index, multiple evaluation indexes are grouped and composed to form the overall evaluation of base clusterings in the aspects of quality and diversity respectively; On the top level of criteria, the final significance of base clusterings is obtained by integrating the significances of the criteria of quality and diversity. The detailed computation of the base clustering significance is introduced below.

On the top level, the significance of a clustering result C_i is obtained by

$$sig(C_i) = \lambda_q \cdot q_{C_i} + \lambda_d \cdot d_{C_i},$$
$$\lambda_q, \lambda_d \in [0,1] \text{ and } \lambda_q + \lambda_d = 1 \tag{1}$$

in which the weights λ_q and λ_d denote the significances of the criteria of clustering quality and diversity and we set $\lambda_q = 0.7, \lambda_d = 0.3$ as default. q_{C_i} and d_{C_i} denote the evaluations of the quality and diversity of the clustering C_i.

On the second level of evaluation index, given L quality measures and T diversity measures, q_{C_i} and d_{C_i} are obtained by

$$q_{C_i} = \sum_{l=1}^{L} wq_l \cdot q_{C_i}^l, \ d_{C_i} = \sum_{t=1}^{T} wd_t \cdot d_{C_i}^t \tag{2}$$

where $0 \leq wq_l, wd_t \leq 1$ and $\sum_{l=1}^{L} wq_l = \sum_{t=1}^{T} wd_t = 1$. $q_{C_i}^l$ and $d_{C_i}^t$ denote the evaluations of the clustering C_i with the lth quality measure and the tth diversity measure respectively, wq_l and wd_t are the normalized significance weights of the measures.

On the bottom level of base clustering results, adopting L quality and T diversity measures to evaluate M base clusterings, we suppose $\{q_{C_i}^1, ..., q_{C_i}^l, ..., q_{C_i}^L\}$ are L quality evaluations and $\{d_{C_i}^1, ..., d_{C_i}^t, ..., d_{C_i}^T\}$ are T diversity evaluations of a clustering C_i, $1 \leq i \leq M$ and normalize all the evaluations to $[0, 1]$. The pairwise comparison of clusterings C_i and C_j in quality can be formulated by

$$L(i,j)_q^l = q_{C_i}^l - q_{C_j}^l, \ \ q_{C_i}^l \geq q_{C_j}^l \text{ and } 1 \leq l \leq L \tag{3}$$

and the pairwise comparison of clusterings in diversity can be formulated by

$$L(i,j)_d^t = d_{C_i}^t - d_{C_j}^t, \ \ d_{C_i}^t \geq d_{C_j}^t \text{ and } 1 \leq t \leq T \tag{4}$$

Based on the pairwise comparison, given an evaluation measure, we can construct a comparison matrix of all the clusterings.

Definition 1. *Pairwise-Comparison Matrix of Clustering Results.* *Given* M *clusterings and an evaluation measure, the pairwise comparison of clusterings* C_i *and* C_j *is* $L(i,j)_{q|d}^{l|t}$ *denoted by* $L(i,j)$. *Suppose a mapping* $f : L(i,j) \rightarrow [1,9]$ *ranks continuous pairwise comparison into 9 levels through sorting and partitioning all the values, the comparison matrix of clusterings is constructed as follows.*

$$\begin{aligned} \mathbf{P} &= [\mathbf{P}(i,j) \,|\, 1 \leq i,j \leq M] \\ &= \begin{pmatrix} 1 & f(L(1,2)) & \cdots & f(L(1,M)) \\ \vdots & \vdots & \ddots & \vdots \\ 1/f(L(1,M)) & 1/f(L(2,M)) & \cdots & 1 \end{pmatrix} \end{aligned} \tag{5}$$

For convenience, we use $\mathbf{P}_q^l$ to denote the comparison matrix of the lth quality measure and $\mathbf{P}_d^t$ to denote the matrix of the tth diversity measure.

According to the AHP theory, the elements of the comparison matrix indicate the levels of relative importance between the different clustering results. '1 | 9' means 'equal | extreme importance'. Moreover, because $\mathbf{P}(i,j) \cdot \mathbf{P}(j,i) = 1$, the comparison matrix $\mathbf{P}$ is a positive reciprocal matrix. Referring to [15], we can further define the consistent comparison matrix of clusterings.

Definition 2. *Consistent Comparison Matrix.* *If the pairwise-comparison matrix* $\mathbf{P} = [\mathbf{P}(i,j) \,|\, 1 \leq i,j \leq M]$ *is a positive reciprocal matrix, it is consistent if and only if the following relationship holds.*

$$\mathbf{P}(i,k) \cdot \mathbf{P}(k,j) = \mathbf{P}(i,j), \ 1 \leq i,j \text{ and } k \leq M. \tag{6}$$

A comparison matrix $\mathbf{P}$ is 'perfectly consistent', if $\mathbf{P}$ satisfies the $C.R. = 0$. A comparison matrix $\mathbf{P}$ is 'near consistent', if the consistent ratio $0 \leq C.R. \leq 0.1$. The ratio $C.R. = \frac{C.I.}{R.I.}$, $R.I.$ is a random index whose value depends on M and $C.I. = \frac{\lambda_{\max} - M}{M - 1}$, in which $\lambda_{\max}$ is the maximum eigenvalue of the matrix. If a comparison matrix is consistent (or near), the principle eigenvector of the matrix is the consistent low-rank representation of all the pairwise comparisons [15]. The experiments will validate the consistency of the constructed comparison matrix.

Due to the principal eigenvector of a consistent comparison matrix $\mathbf{P}$ is a necessary representation of the priorities derived from the pairwise comparisons, we can calculate the evaluations of clusterings through the eigendecomposition of the comparison matrix. The elements of the obtained principal eigenvector represent the significance of the corresponding clustering results, which are consistent to the pairwise comparisons. For the $M \times M$ pairwise comparison matrix $\mathbf{P}_q^l$ of the lth quality measure, we normalize the element sum of the principle eigenvector $\mathbf{q}^l$ to one and obtain the quality weights of M clustering results.

$$\mathbf{P}_q^l \cdot \mathbf{q}^l = \lambda_{\max} \cdot \mathbf{q}^l, \ \mathbf{q}^l = \{q_{C_1}^l, \cdots, q_{C_M}^l\} \ \text{and} \ \sum_{i=1}^{M} q_{C_i}^l = 1 \qquad (7)$$

For multiple L quality measures, the principle eigenvectors of L comparison matrices lead to L groups of quality weights $\{\mathbf{q}^1, ..., \mathbf{q}^L\}$ as the overall quality evaluation of the clustering results. In a similar way, we can calculate the diversity weights of clustering results $\mathbf{d}^t = \{d_{C_1}^1, \cdots, d_{C_M}^T\}$ through the eigendecomposition of comparison matrices of diversity measures $\mathbf{P}_d^t, 1 \leq t \leq T$.

Besides constructing the comparison matrices to compute the weights of clustering results $q_{C_i}^l$ and $d_{C_i}^t$, we should also construct comparison matrices to calculate the significance weights of evaluation measures, i.e. wq_l, wd_t. In this paper, we adopt four internal indexes as the quality measures ($L = 4$) and three external indexes as the diversity measures ($T = 3$). The internal indexes include Calinski Harabasz index (CHI) [3], Davies-Bouldin index (DBI) [5], Compactness (CP) and Separation (SP). The external indexes include Cluster Accuracy (CA), Adjusted Rand index (ARI) and Normalized Mutual Information (NMI). Referring to [14], we can construct a 4×4 comparison matrix $\mathbf{P}_{eq}$ for quality measures and a 3×3 comparison matrix $\mathbf{P}_{ed}$ for diversity measures.

$$\mathbf{P}_{eq} = \begin{pmatrix} & \text{DB} & \text{CH} & \text{CP} & \text{SP} \\ \text{DB} & 1 & 3 & 8 & 9 \\ \text{CH} & 1/3 & 1 & 7 & 8 \\ \text{CP} & 1/8 & 1/7 & 1 & 3 \\ \text{SP} & 1/9 & 1/8 & 1/3 & 1 \end{pmatrix}, \ \mathbf{P}_{ed} = \begin{pmatrix} & \text{CA} & \text{NMI} & \text{ARI} \\ \text{CA} & 1 & 2 & 3 \\ \text{NMI} & 1/2 & 1 & 2 \\ \text{ARI} & 1/3 & 1/2 & 1 \end{pmatrix} \qquad (8)$$

Similar to the calculation of weights of clusterings, the significance weights of evaluation measures are computed through eigendecomposition of $\mathbf{P}_{eq}$ and $\mathbf{P}_{ed}$.

$$\mathbf{P}_{eq} \cdot \mathbf{eq} = \lambda_{\max} \cdot \mathbf{eq}, \ \mathbf{P}_{ed} \cdot \mathbf{ed} = \lambda_{\max} \cdot \mathbf{ed} \qquad (9)$$

Algorithm 1. Clustering Ensemble Selection with AHP (AHPCES)

Input: M base clustering results $\mathbf{C} = \{C_1, C_2, \cdots, C_M\}$;
 1: Construct $\mathbf{P}_{eq}$, $\mathbf{P}_{ed}$ and compute $\mathbf{eq} = \{wq_1, \cdots, wq_L\}$, $\mathbf{ed} = \{wd_1, \cdots, wd_T\}$;
 2: **for** $l = 1$ to L **do**
 3:     Construct $\mathbf{P}_q^l$ and compute $\mathbf{q}^l = \{q_{C_1}^l, \cdots, q_{C_M}^l\}$;
 4: **end for**
 5: **for** $t = 1$ to T **do**
 6:     Construct $\mathbf{P}_d^t$ and compute $\mathbf{d}^t = \{d_{C_1}^t, \cdots, d_{C_M}^t\}$;
 7: **end for**
 8: Construct $\mathbf{Q} = \{\mathbf{q}^1, ..., \mathbf{q}^L\}$ and $\mathbf{D} = \{\mathbf{d}^1, ..., \mathbf{d}^T\}$;
 9: Compute $\mathbf{Q} \cdot \mathbf{eq} = \{q_{C_1}, q_{C_2}, ..., q_{C_M}\}$ and $\mathbf{D} \cdot \mathbf{ed} = \{d_{C_1}, d_{C_2}, ..., d_{C_M}\}$;
10: Compute the significance of clustering results $sig(C_i) = \lambda_q \cdot q_{C_i} + \lambda_d \cdot d_{C_i}, 1 \leq i \leq M$;

11: Rank the clustering results according to their significances.

Table 1. Description of Data Sets

Data sets	Instance	Class	Sources	Data sets	Instance	Class	Sources
DS1(S1)	5000	15	[7]	DS7(D31)	3100	31	[1]
DS2(Jain)	373	2	[11]	DS8(Heart)	270	2	[1]
DS3(Flame)	240	2	[8]	DS9(Wine)	178	3	[1]
DS4(Spiral)	312	3	[4]	DS10(Ecoli)	336	8	[1]
DS5(Pathbased)	300	3	[4]	DS11(Australian Credit Approval)	690	2	[1]
DS6(Aggregation)	788	7	[9]				

where $\mathbf{eq} = \{wq_1, \cdots, wq_L\}, \sum_{l=1}^{L} wq_l = 1$ and $\mathbf{ed} = \{wd_1, \cdots, wd_T\}, \sum_{t=1}^{T} wd_t = 1$. The normalized principal eigenvectors $\mathbf{eq}$ and $\mathbf{ed}$ provide the comparative significance weights of the evaluation measures in quality and diversity respectively.

   As introduced above, synthesizing the weights of clustering results and evaluation measures as shown in Formulas (1) and (2), we can quantify the overall evaluation of clustering results under multiple measures and thereby select the significant clusterings for ensemble. The process of clustering ensemble selection based on AHP is presented in Algorithm 2.

3   Experimental Results

We implement three experiments to validate AHPCES. In these experiments, we adopt kmeans to generate 50 base clusterings, set the selection ratio ranges from 20% to 60%, use LinkCluE [10] as the consensus function and adopt three external criteria (CA, ARI, NMI) and three internal criteria (SC, DI, CHI) to evaluate the clusterings, high evaluation values indicate the high performances of the clusterings. All the experiments are performed on 11 data sets. The detailed descriptions of the data sets are shown in Table 1.
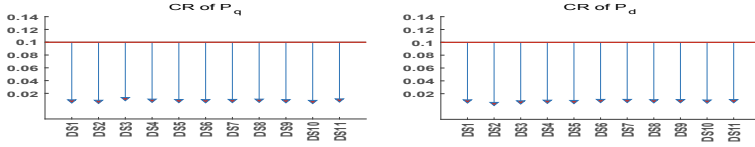
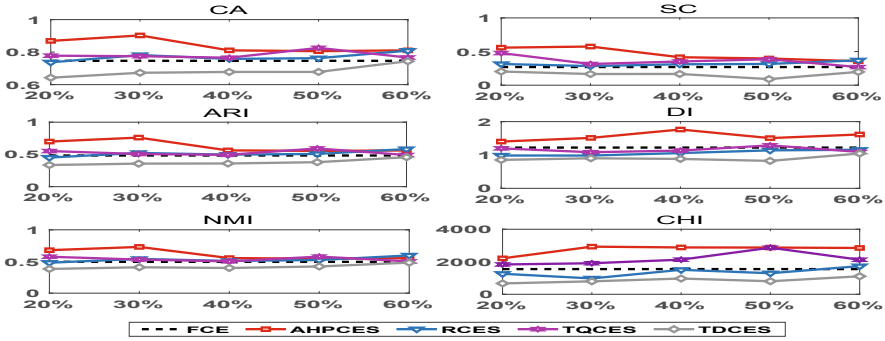**Fig. 1.** Consistency ratios of comparison matrices.

Fig. 2. Comparison of different clustering selection strategies.

3.1 Test of Comparison Matrix Consistency

In the first experiment, in order to demonstrate the consistency of the constructed pairwise comparison matrices of clustering quality and diversity (see Sect. 2), we compute the consistency ratio $C.R.$ of all the pairwise comparison matrices for the base clusterings on each data set.

Figure 1 shows the average consistency ratio of the quality and diversity comparison matrices $C.R. \leq 0.1$ (namely near consistent), which indicates that the principal eigenvectors of comparison matrices can be used to quantify the significance of base clusterings in quality and diversity.

3.2 Test of AHP-based Clustering Selection

In the second experiment, we implement two tests to verify the AHPCES. In the first test, we compare AHPCES with four clustering selection strategies, which include Full Clustering Ensemble (FCE), Random Clustering Ensemble Selection (RCES), Top-k Quality Clustering Ensemble Selection (TQCES) and Top-k Diversity Clustering Ensemble Selection (TDCES).

Figure 2 shows the average evaluations on all the data sets against the selection ratios ranging from 20% to 60%. We can find that, under the same selection proportions, AHPCES achieves the most precise ensemble results than the other selection strategies, which means the AHPCES is more effective to select the significant base clusterings for ensemble.
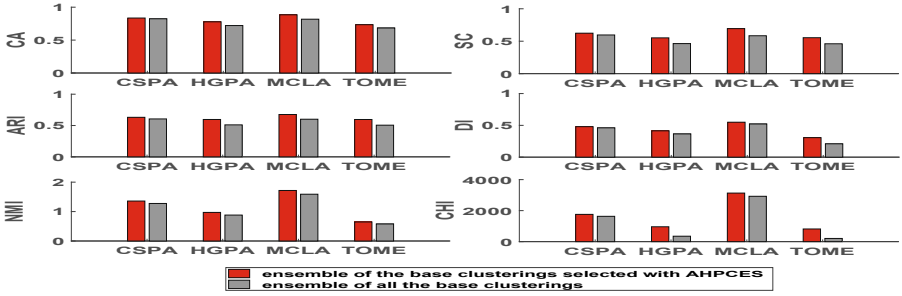
Fig. 3. Comparison of ensemble with/without AHPCES on four consensus functions.

Table 2. Comparison of elegant clustering ensemble selection methods

Criterion	Method			
	AHPCES	CAS	ACES	SELSCE
CA	0.8904(30%)	0.6607(50%)	0.7871(30%)	0.7628(50%)
ARI	0.7701(30%)	0.3588(60%)	0.5845(30%)	0.5421(50%)
NMI	0.7519(30%)	0.4238(60%)	0.5920(30%)	0.5647(50%)
SC	0.5337(30%)	0.3382(60%)	0.4049(30%)	0.3615(50%)
DI	1.6723(50%)	0.9082(20%)	1.2681(20%)	1.2538(30%)
CHI	2933(40%)	1890(40%)	1935(20%)	1341(50%)

In the second test, we adopt four consensus functions to ensemble the clusterings selected by AHPCES and the full clusterings without selection. The consensus functions include CSPA, HGPA, MCLA [16] and TOME [19]. With different consensus functions, Fig. 3 shows the comparison between the ensemble results with AHPCES and without selection. We can find that, for all the consensus functions, the AHPCES can improve the full ensemble results, which validates the robustness of the AHP strategy for base clustering selection.

3.3   Overall Evaluation

In the final experiment, we expect to overall evaluate the proposed AHPCES method through comparing with three elegant clustering ensemble selection methods. The methods for comparison include Cluster and Selection algorithm (CAS) [6], Adaptive Cluster Ensemble Selection algorithm (ACES) [2] and Selective Spectral Clustering Ensemble algorithm (SELSCE) [12]. Because of the limitation of paper length, we just present the average evaluations generated by different selection strategies on all the data sets.

Table 2 presents the comparison between AHPCES and other clustering ensemble selection methods. We show the max average value evaluated by each criterion and attach the corresponding selected proportion. It is obvious that

AHPCES achieves the best performances. Abundant experiments indicate that our proposed method is effective to select the qualified and diverse base clusterings, and thereby produce precise clustering ensemble results.

4   Conclusions

we propose a clustering ensemble selection method with analytic hierarchy process (AHPCES) to address the limitations of the most existing clustering ensemble selection methods. Experimental results validate AHPCES can generate reasonable and consistent evaluations of base clusterings for selective ensemble. Our future work will focus on the theoretical analysis of the consistency of pairwise comparison matrices and try to improve the construction of the comparison matrices of evaluations indexes.

References

1. Asuncion, A., Newman, D.: UCI machine learning repository (2007)
2. Azimi, J., Fern, X.: Adaptive cluster ensemble selection. In: Twenty-First International Joint Conference on Artificial Intelligence, **9**, pp. 992–997 (2009)
3. Caliński, T., Harabasz, J.: A dendrite method for cluster analysis. Commun. Stat. Theory Methods **3**(1), 1–27 (1974)
4. Chang, H., Yeung, D.Y.: Robust path-based spectral clustering. Pattern Recogn. **41**(1), 191–203 (2008)
5. Davies, D.L., Bouldin, D.W.: A cluster separation measure. IEEE Trans. Pattern Anal. Mach. Intell. **2**, 224–227 (1979)
6. Fern, X.Z., Lin, W.: Cluster ensemble selection. Stat. Anal. Data Min. ASA Data Sci. J. **1**(3), 128–141 (2008)
7. Fränti, P., Virmajoki, O.: Iterative shrinking method for clustering problems. Pattern Recogn. **39**(5), 761–775 (2006)
8. Fu, L., Medico, E.: Flame, a novel fuzzy clustering method for the analysis of dna microarray data. BMC Bioinform. **8**(1), 3 (2007)
9. Gionis, A., Mannila, H., Tsaparas, P.: Clustering aggregation. ACM Trans. Knowl. Discov. from Data (TKDD) **1**(1), 4 (2007)
10. Iam-on, N., Garrett, S., et al.: Linkclue: a matlab package for link-based cluster ensembles. J. Stat. Softw. **36**(9), 1–36 (2010)
11. Jain, A.K., Law, M.H.C.: Data clustering: a user's dilemma. In: Pal, S.K., Bandyopadhyay, S., Biswas, S. (eds.) PReMI 2005. LNCS, vol. 3776, pp. 1–10. Springer, Heidelberg (2005). https://doi.org/10.1007/11590316_1
12. Jia, J., Xiao, X., Liu, B., Jiao, L.: Bagging-based spectral clustering ensemble selection. Pattern Recogn. Lett. **32**(10), 1456–1467 (2011)
13. Jie, Z., Zhihui, L., Duoqian, M., Can, G., Xiaodong, Y.: Multigranulation rough-fuzzy clustering based on shadowed sets. Inf. Sci. **507**, 553–573 (2018)

14. Liu, Y., Li, Z., Xiong, H., Gao, X., Wu, J., Wu, S.: Understanding and enhancement of internal clustering validation measures. IEEE Trans. Cybern. **43**(3), 982–994 (2013)
15. Saaty, T.L.: Decision-making with the ahp: why is the principal eigenvector necessary. Eur. J. Oper. Res. **145**(1), 85–91 (2003)
16. Strehl, A., Ghosh, J.: Cluster ensembles-a knowledge reuse framework for combining multiple partitions. J. Mach. Learn. Res. **3**(Dec), 583–617 (2002)
17. Yue, X., Miao, D., Cao, L., Wu, Q., Chen, Y.: An efficient color quantization based on generic roughness measure. Pattern Recogn. **47**(4), 1777–1789 (2014)
18. Zhao, X., Liang, J., Dang, C.: Clustering ensemble selection for categorical data based on internal validity indices. Pattern Recogn. **69**, 150–168 (2017)
19. Zhong, C., Yue, X., Zhang, Z., Lei, J.: A clustering ensemble: two-level-refined co-association matrix with path-based transformation. Pattern Recogn. **48**(8), 2699–2709 (2015)