A Note on the effectiveness of the LEAST SQUARES CONSENSUS CLUSTERING

B. Mirkin, A. Shestakov

12 December 2013

Abstract

We develop a consensus clustering framework proposed three decades ago in Russia and experimentally demonstrate that our least squares consensus clustering algorithm consistently outperforms several recent consensus clustering methods.

keywords: consensus clustering, ensemble clustering, least squares

1 Introduction

The problem of finding a partition reconciling a set of pre-specified partitions has been stated, developed and applied by B. Mirkin and L. Cherny in the beginning of 70es in the context of "nominal factor analysis" [1–4]. Yet this work remained largely unknown until M. Meila mentioned the so-called Mirkin's distance, an "iceberg tip" of the work [5].

Perhaps the grand start for consensus clustering approach on the international scene was made by A. Strehl and J. Ghosh [10]. Since then consensus clustering has become popular in bioinformatics, web-document clustering and categorical data analysis. According to [6] consensus clustering algorithms can be organized in three main categories: probabilistic approach [11], [12]; direct approaches [10,13,15,16], and pairwise similarity-based approach [9,14]. The (i,j)-th entry a_{ij} in the consensus matrix $A = (a_{ij})$ shows the number of partitions in which objects y_i and y_j are in the same cluster.

Here we invoke a least-squares consensus clustering approach from the paper [7] predating the above developments, update it with a more recent clustering procedure to obtain an algorithm for concensus clustering and compare the results on synthetic data of Gaussian clusters with those by the more recent methods. It appears our method outperforms those with a good margin.

2 Least squares criterion for consensus clustering

Given a partition of N-element dataset Y on K non-overlapping classes $S = \{S_1, \ldots, S_K\}$, its binary membership $N \times K$ matrix $Z = (z_{ik})$ is defined so that $z_{ik} = 1$ if y_i belongs to S_k and $z_{ik} =$ otherwise. As is known, the orthogonal projection matrix over the linear space spanning the columns of matrix Z is defined as $P_Z = Z(Z^T Z)^{-1} Z^T = (p_{ij})$ where $p_{ij} = \frac{1}{N_k}$, if $\{y_i, y_j\} \in S_k$ and 0 otherwise.

Given a profile of T partitions $R = \{R^1, R^2, \ldots, R^T\}$, its ensemble consensus partition is defined as that with a matrix Z minimizing the sum of squared residuals in equations

$$x_{il}^t = \sum_{k=1}^{K} c_{kl}^t z_{ik} + e_{ik}^t, \tag{1}$$

over the coefficients c_{kl}^t and matrix elements z_{ik} where X^t, $t = 1, \ldots, T$ are binary membership matrices for partitions in the given profile R. The criterion can be equivalently expressed as

$$E^2 = \|X - P_Z X\|^2, \tag{2}$$

where X is concatenation of matrices $X^1, \ldots, X_t$ and $\|\cdot\|^2$ denotes the sum of squares of the matrix elements. This can be further transformed into an equivalent criterion to be maximized:

$$g(S) = \sum_{k=1}^{K} \sum_{i,j \in S_k} \frac{a_{ij}}{N_k}, \tag{3}$$

where $A = (a_{ij})$ is the consensus matrix A from the pairwise similarity-based approach.

To (locally) maximize (3), we use algorithm AddRemAdd(j) from Mirkin in [8] which finds clusters one-by-one. Applied to each object y_j this method outputs a cluster with a high within cluster similarity according to matrix A. AddRemAdd(j) runs in a loop over all $j = 1 \ldots N$ and takes that of the found clusters at which (3) is maximum. When it results in cluster $S(j)$, the algorithm is applied on the remaining dataset $Y' = Y/S(j)$ with a correspondingly reduced matrix A'. It halts when no unclustered entities remain. The least squares ensemble consensus partition consists of the AddRemAdd cluster outputs: $S^* = \bigcup S(j)$. It should be pointed out that the number of clusters is not pre-specified at AddRemAdd.

3 Experimental results

All evaluations are done on synthetic datasets that have been generated using Netlab library [17]. Each of the datasets consists of 1000 twelve-dimensional objects comprising

nine randomly generated spherical Gaussian clusters. The variance of each cluster lies in $0.1 - 0.3$ and its center components are independently generated from the Gaussian distribution $\mathcal{N}(0, 0.7)$.

Let us denote thus generated partition as Λ with $k_\Lambda = 9$ clusters. The profile of partitions $R = \{R^1, R^2, \ldots, R^T\}$ for consensus algorithms is constructed as a result of $T = 50$ runs of k-means clustering algorithm starting from random k centers. We carry out the experiments in four settings: a) $k = 9 = k_\Lambda$, b) $k = 6 < k_\Lambda$, c) $k = 12 > k_\Lambda$, d) k is uniformly random on the interval $(6, 12)$. Each of the settings results in 50 k-means partitions. After applying consensus algorithms, Adjusted Rand Index (ARI) [6] for the consensus partitions S and generated partition Λ is computed as $\varphi^{ARI}(S, \Lambda)$.

3.1 Comparing consensus algorithms

The least squares consensus results have been compared with the results of the following algorithms (see Tables 1-4):

- Voting Scheme (Dimitriadou, Weingessel and Hornik - 2002) [13]
- cVote (Ayad - 2010) [16]
- Fusion Transfer (Guenoche - 2011) [14]
- Borda Consensus (Sevillano, Carrie and Pujol - 2008) [15]
- Meta-CLustering Algorithm (Strehl and Ghosh - 2002) [10]

Table 1: *The average values of $\phi^{ARI}(S, \Lambda)$ and the number of classes if $k_\Lambda = k = 9$ over 10 experiments in each of the settings.*

Algorithm	Average ϕ^{ARI}	Std. ϕ^{ARI}	Avr. # of classes	Std. # of classes
ARA	**0.9578**	0.0246	7.6	0.5164
Vote	0.7671	0.0624	8.9	0.3162
cVote	0.7219	0.0882	8.1	0.7379
Fus	0.7023	0.0892	11.6	1.8379
Borda	0.7938	0.1133	8.5	0.7071
MCLA	0.7180	0.0786	8.6	0.6992

Tables 1-4 consistently show that:

Table 2: *The average values of $\phi^{ARI}(S,\Lambda)$ and the number of classes at $k_\Lambda > k = 6$ over 10 experiments in each of the settings.*

Algorithm	Average ϕ^{ARI}	Std. ϕ^{ARI}	Avr. # of classes	Std.# of classes
ARA	**0.8333**	0.0586	6.2	0.6325
Vote	0.7769	0.0895	5.9	0.3162
cVote	0.7606	0.0774	5.6	0.6992
Fus	**0.8501**	0.1154	7.7	1.3375
Borda	0.7786	0.0916	6	0
MCLA	0.7902	0.0516	6	0

Table 3: *The average values of $\phi^{ARI}(S,\Lambda)$ and the number of classes at $k_\Lambda < k = 12$ over 10 experiments in each of the settings.*

Algorithm	Average ϕ^{ARI}	Std. ϕ^{ARI}	Avr. # of classes	Std.# of classes
ARA	**0.9729**	0.0313	9	0.9428
Vote	0.6958	0.0796	11.4	0.5164
cVote	0.672	0.0887	10.9	0.7379
Fus	0.6339	0.0827	16	4
Borda	0.7132	0.074	11.1	0.7379
MCLA	0.6396	0.0762	11.9	0.3162

Table 4: *The average values of $\phi^{ARI}(S,\Lambda)$ and the number of classes at $k \in (6,12)$ over 10 experiments in each of the settings.*

Algorithm	Average ϕ^{ARI}	Std. ϕ^{ARI}	Avr. # of classes	Std.# of classes
ARA	**0.9648**	0.019	6.8	0.7888
cVote	0.5771	0.1695	10.4	1.2649
Fus	0.62	0.0922	11.6	2.0656
MCLA	0.6567	0.1661	10.6	1.3499

- The least-squares consensus clustering algorithm have outperformed the other consensus clustering algorithms consistently – the average ϕ^{ARI} is higher while it's standard deviation is closer to zero;

- The only exception, at option (c), with $k_\Lambda > k = 6$ the Fusion Transfer algorithm demonstrated a little better result probably because of the transfer procedure (see Table 2) .

- The average number of clusters in the consensus clustering is lower than k in the profile R and k_Λ

Conclusion

This paper revitalizes a 30-years old approach to consensus clustering proposed by Mirkin and Muchnik in Russian. When supplemented with an update algorithmic procedures, the method shows a very good competitiveness over a set of recent cluster consensus techniques. Our further work will include: (a) extension of the experimental series to a wider set of consensus clustering procedures, including those based on probabilistic modeling, (b) attempts at using the approach as a device for choosing "the right number of clusters", (c) exploring various devices, such as random intializations in k-means or bootstrapping of variables, for generation of ensembles of partitions, etc.

This work was supported by the research grant “Methods for the analysis and visualization of texts” No. 13-05-0047 under The National Research University Higher School of Economics Academic Fund Program in 2013.

References

[1] B.G. Mirkin. A new approach to the analysis of sociology data. - In: Y. Voronov (Ed.) Measurement and Modeling in Sociology, Novosibirsk, Nauka Pubishers, 51-61, 1969 (In Russian).

[2] L.B. Cherny. The method of the partition space in the analysis of categorical features. A PhD thesis, Institute of Control Problems, Moscow, 1973. (In Russian)

[3] L. B. Cherny. Relationship between the method of the partition space and other methods of data analysis. - In: B. Mirkin (Ed.) Issues in Analysis of Complex Systems, Novosibirsk, Nauka Pubishers, 84-89, 1974. (In Russian)

[4] B.G. Mirkin. Analysis of Categorical Features. Moscow, Statistika Publishers, 166 p., 1976 (In Russian).

[5] M. Meila. Comparing clusteringsвЂ”an information based distance. - The Journal of Multivariate Analysis, 98(5), 873-881, 2007.

[6] J. Ghosh, A. Acharya. "Cluster ensembles". - Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2011.

[7] B. Mirkin, I. Muchnik. Geometrical interpretation of clustering scoring functions. In: B. Mirkin (Ed.) Methods for the Analysis of Multivariate Data in Economics, Nauka Publisher, Novosibirsk, 1981, 3 – 11 (In Russian).

[8] B. Mirkin. Core concepts in Data Analysis: Summarization, Correlation, Visualization. - Springer, 2011.

[9] B. Mirkin. Clustering: A Data Recovery Approach. - Chapmam and Hall / CRC Press, 2012.

[10] A. Strehl, J. Ghosh. Cluster ensembles - a knowledge reuse framework for combining multiple partitions. - Journal on Machine Learning Research, 2002.

[11] A. Topchy, A. K. Jain, and W. Punch A mixture model for clustering ensembles. In Proceedings SIAM International Conference on Data Mining, 2004.

[12] H. Wang, H. Shan, A. Banerjee Bayesian cluster ensembles. In: Proceedings of the Ninth SIAM International Conference on Data Mining, 211пїS222, 2009.

[13] E. Dimitriadou, A. Weingessel and K. Hornik. A Combination Scheme for Fuzzy Clustering. - Journal of Pattern Recognition and Artificial Intelligence, 2002.

[14] A. Guenoche. Consensus of partitions : a constructive approach.
- Adv. Data Analysis and Classification 5, pp. 215-229, 2011.

[15] X. Sevillano Dominguez, J. C. Socoro Carrie and F. Alias Pujol. Fuzzy clusterers combination by positional voting for robust document clustering. - Procesamiento del lenguaje natural, пїS 43, pp. 245-253.

[16] H. Ayad, M. Kamel On voting-based consensus of cluster ensembles. - Pattern Recognition, pp. 1943-1953, 2010.

[17] Netlab Neural Network software, <http://www.ncrg.aston.ac.uk/netlab/index.php>