Proceedings of 2024 IEEE 12th International Conference on Intelligent Systems (IS)

A Constrained Cluster Ensemble Using Hierarchical Clustering Methods

Francis J. Williams, Samuel L. Hennessey
and Ludmila I. Kuncheva
School of Computer Science and Engineering
Bangor University
Bangor, UK
Email: {f.j.williams, sml18vly, l.i.kuncheva}@bangor.ac.uk

José F. Diez-Pastor
and Juan J. Rodríguez
Departamento de Ingeniería Informática
Universidad de Burgos
Burgos, Spain
Email: {jdfpastor, jjrodriguez}@ubu.es

***Abstract*—Unsupervised classification of data is an ongoing challenge in many areas. With evolving stream data, hierarchical clustering methods have proved effective, especially with non-spherical clusters. Additionally, incorporating pairwise constraints has been shown to further improve clustering accuracy.**

We propose a cluster ensemble using constrained hierarchical methods. The experiment was performed on a collection of 52 Synthetic and 96 Real datasets.

Our analysis shows that our constrained cluster ensemble method results in a high accuracy across various proportions of constraints without sacrificing speed.

***Keywords*—constrained clustering; hierarchical clustering; pairwise constraints; semi-supervised learning**

I. INTRODUCTION

Object classification within images and video is an ongoing challenge in many fields [1], [2], [3]. Annotating images and labelling the data within is a tedious task which takes countless man-hours if performed manually. In recent years, with the development of smart technologies, the search for automated object classification methods has been a continuous pursuit. Object clustering is a form of unsupervised learning, in which previously unseen data are grouped based on similar attributes.

Real-life datasets often include stream data, such as videos. As the data evolves in adjacent frames, a minor change is expected to an object's appearance and, thus, to its feature representation. This leads to the assumption that the data will produce string-shaped clusters. On the other hand, an object that appears several times in the video data may give rise to multiple string-shaped clusters in possibly distinct parts of the feature space. For example, in recognising faces in video, there may be a cluster of points coming from a frontal-view shot, and another cluster where the face is in profile [4]. As the two views will have very different feature representations, the two clusters for the same identity will be in different parts of the feature space.

Unlike standard clustering tasks, object clustering from video allows for using additional information in the form of constraints. Typically, these are pairwise constraints: must-link (ML) constraints, where two objects have to belong to the same cluster, and cannot-link (CL) constraints, where two objects must not be in the same cluster. Constrained clustering (CC) has been developed to incorporate additional information about the data into standard clustering algorithms [5], [6], [7].

Current comparisons of CC methods do not specifically address the type of the data being clustered. Most comparisons rely on benchmark Real data sourced from well-established repositories. It is well-known that such data feature largely hyper-spherical clusters with good correspondence to class labels. Centroid-based clustering methods would be a good fit for such data.

For the type of data we are considering here, hierarchical clustering is more likely to produce good results. Successful hierarchical methods for CC have been summarised by Gonzalez-Almagro et al [5]. However, most of these require human intervention or a form of active learning. We are interested in an automatic hierarchical method for constrained clustering with a view to extend it to online clustering in the future.

Years of research on clustering have led to the observation that, generally, cluster ensembles are more successful than single clusterers [8], [9], [10]. We propose a clustering ensemble comprised of elementary semi-supervised agglomerative hierarchical methods. To demonstrate the advantages of our method, we ran a large experimental evaluation on our bespoke collection of 52 Synthetic datasets sourced from the literature on clustering, and 96 Real datasets from the KEEL-dataset repository [11].

The rest of the paper is organised as follows. Section II discusses related constrained clustering and ensemble solutions. Section III details our proposed method. Section IV describes our experiment and Section V shows the results. Finally, we offer our conclusions in Section VI.

II. RELATED WORK

CC methods have emerged as powerful tools in machine learning and data analysis, offering the ability to incorporate domain knowledge in the form of constraints to guide the clustering process [12], [13], [14].

The integration of constraints in hierarchical clustering methods (HCC) has been the focus of various studies, accommodating a range of applications [12], [15]. One advantage of

979-8-3503-5098-2/24/$31.00 ©2024 IEEE

HCC is its ability to cluster non-spherical data more accurately where centroid-based methods such as k-means could not produce an adequate partition [16]. Many applications of HCC are active methods, where an oracle is consulted to determine which constraint applies to a pair of instances [14], [17], [18], [19]. Such methods are not suited to our needs, as we are interested in fully autonomous methods. Other methods take different approaches to incorporating constraints, such as through cost functions [13], distance metrics [20], [21], and restricting constraints to certain parts of the hierarchical structure [7].

Meanwhile, cluster ensembles offer several advantages over single clustering methods due to their ability to harness the collective intelligence of multiple clustering algorithms. By aggregating diverse clustering solutions, ensembles can mitigate the limitations of individual algorithms, such as sensitivity to initialisation, parameter settings, and data characteristics. Moreover, ensembles inherently capture a broader range of data perspectives, leading to enhanced robustness and stability in the clustering results [8], [9], [10]. The diversity among ensemble members allows for a more comprehensive exploration of the data space, thereby increasing the likelihood of identifying complex cluster structures and outliers. Overall, the ensemble approach results in improved clustering quality, greater reliability, and enhanced performance across a wide range of datasets and applications [22], [23], [24], [25], [26].

Combining constraint-based clustering with ensemble methods has been shown to yield more reliable, accurate, and meaningful clustering results across a wide range of applications [27], [28], [29], [30].

Over time, with a growing emphasis on clustering accuracy, contemporary constrained clustering methods have evolved to become exceedingly intricate, often at the expense of speed and transparency. This evolution has rendered them less suitable for online learning in real-time applications. In response, we have developed a method that combines the advantages of constraint-based clustering with those of ensemble methods, while preserving the simplicity of basic hierarchical agglomerative clustering and ensuring swift execution.

III. Proposed Method

A. Klein, Kamvar and Manning's Algorithm

To understand the proposed ensemble method, we first detail Klein et al's algorithm from [31].

Let $X = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ be a dataset of N objects each described by a feature vector. Two sets of constraints are defined over X: ML and CL. A pair of objects, i and j, belongs to ML,$(i, j) \in ML$, if they must share the same cluster label in the final partition. Conversely, $(i, j) \in CL$, if they must be assigned to different clusters. Not every constrained clustering algorithm can guarantee that all the constraints are satisfied.

Klein et al. proposed that integrating constraints into hierarchical clustering methods did not necessitate modifications to the unsupervised algorithm itself. Due to the inherent characteristics of hierarchical methods, which enable the creation of a clustering partition from a distance matrix, they could adjust the distance matrix to accommodate the constraints instead.

First, ML not always represents a full set of constraints. There may be pairs of objects linked through transitive closure, which are not present in the original ML. Therefore, we expand ML into MLa to include all possible pairs that must be connected through must-link. To this end, we build a graph with N nodes and place an edge between all pairs of nodes corresponding to ML. From this graph, we derive the connected components. For example, if pairs (a, b) and (b, c) are in ML, the respective connected component of the graph will contain all three a, b and c. Thus, ML is augmented with all pairs in each connected component.

Consider a distance matrix $M_{N\times N} = \{a_{ij}\}$, where a_{ij} indicates the distance between points i and j. Constraints can be incorporated into M by modifying entry a_{ij} based on the constraint between points i and j. If $(i, j) \in MLa$, then $a_{ij} = 0$; conversely, if $(i, j) \in CL$, then $a_{ij} = \infty$.

After adapting the distance matrix to include MLa and CL, unsupervised hierarchical clustering methods like average linkage, single linkage, and complete linkage can be employed on a dataset to produce a partition. The respective variants will be called in the rest of this paper `CAL` (Constrained Average Linkage), `CCL` (Constrained Complete Linkage), and `CSL` (Constrained Single Linkage).

B. Constrained Cluster Ensemble

Our proposed method is an ensemble of our implementation of the algorithms in Section III-A. We refer to it as `CCEN` in the remainder of this paper. The input to the algorithm consists of: the dataset X, the constraint sets ML and CL, the chosen constrained clustering base method $CCBM$, the number of clusters k, and the number of ensemble members L.

The `CCEN` algorithm proceeds by initialising an empty cumulative adjacency matrix $CuAd$ of size $N \times N$. Each ensemble member will contribute an adjacency matrix that will be added to $CuAd$. Thus, the pairs of points which are labelled together by all ensemble members will have an entry of L in $CuAd$. Points which were never linked by any ensemble member will score a zero in $CuAd$.

We decided to "recycle" the simple constrained hierarchical methods from Section III-A by using them as potential base clusterers. We opted for CAL as the $CCBM$ in our experiment. Furthermore, to save time, we can run the method only once, and use the resultant dendrogram to create the ensemble members. To do so, we cut the dendrogram to arrive at $k, k+1, \dots, k+L$ clusters, which represent the output of the ensemble members.

Once $CuAd$ is complete, we convert it back to cluster labels. First, we chose to cut the matrix at $\frac{L}{2}$. All elements of $CuAd$ greater than the threshold turn to value 1, and those less than or equal to the threshold turn to value 0. The new ensemble adjacency matrix $EnsAd$ defines a graph, whose connected components are returned by our `CCEN` algorithm as the ensemble cluster labels cl. The steps of this algorithm are outlined in Algorithm 1.

Algorithm 1 Constrained Cluster Ensemble (CCEN)

Require: $X, ML, CL, CCBM, k, L$
Ensure: ecl
$CuAd = [\,]$
for i in $1, \ldots, L$ **do**
  $cl \leftarrow CCBM(X, ML, CL, k + i - 1)$
  $ad \leftarrow$ adjacency matrix from cl
  $CuAd \leftarrow CuAd + ad$
end for
if $CuAd(i,j) > \frac{L}{2}$ **then** $EnsAd(i,j) = 1$
else$EnsAd(i,j) = 0$
end if
$EnsAd = CuAd > \frac{L}{2}$
ecl = CONVERT($EnsAd$)

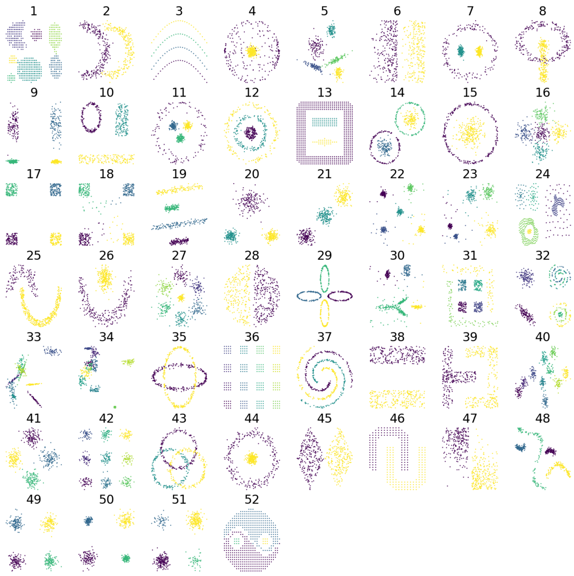


Fig. 1: All Synthetic datasets used in the experimental study visualised in 2-Dimensional space.

IV. Experimental Setup

A. Data

One of our contributions in this study is a collection of Synthetic datasets featuring a variety of cluster shapes and varying levels of clustering difficulty. All datasets have been used before in publications on clustering. A MATLAB library for generating the Synthetic datasets is provided on GitHub[1]. The Synthetic dataset collection contains 47 2D datasets and 5 3D datasets. A 2D view of all 52 datasets is shows in Figure 1.

The Real dataset collection consists of 96 datasets with a minimum of 2 classes and a maximum of 26 classes with a dimensionality from 3 to 262 [11]. For the purposes of our experiment we capped the number of objects of the datasets at 1000.

Tables of data names and attributes are available at https://github.com/frankmnb/Ensemble-hierarchical-method-for-constrained-clustering.

B. Metrics

Normalised Mutual Information (NMI) and Adjusted Rand Index (ARI) were used here to compare the obtained cluster labels with the true labels of the data. $NMI(X,Y) = 0$ signifies no relationship between variables X and Y, and $NMI(X,Y) = 1$ indicates perfect relationship between the two variables. ARI yields a score between -1 and 1, where a score close to 1 indicates a high degree of agreement between the clusterings, while a score near 0 suggests random clustering, and negative scores imply disagreement. ARI is suitable when the number of clusters varies or when the sizes of clusters differ.

[1]Synthetic dataset code available at https://github.com/LucyKuncheva/Synthetic-datasets-for-classification-and-clustering

C. Experimental Protocol

We experimented with different proportions of constraints relative to the number of objects N. The proportions we used were $[0, 1, 2, 3, 4, 5, 10, 15, 20]\%$. For the proportion PC the number of constraints was calculated as: $NC = z(z-1)/2$, where $z = \text{round}(N \times PC/100)$.

After determining the number of constraints, we generate pairs of points $(P_i, P_j)_1, \ldots, (P_i, P_j)_{NC}$ where $i \neq j$ and proceed to identify the nature of the constraints by comparing the true labels of P_i and P_j. If the true labels for both points match, they are designated as ML constraints. Conversely, if they do not belong to the same cluster, they are allocated to the CL constraint set. We generated 5 different sets of NC constraints for each value of PC. This was done to account for the randomness of the constraints. All results are an average across the 5 constraint sets.

D. Methods

The methods we chose for the experiments are:

- COP-kmeans (`COP`) [32] – as a baseline
- COP-kmeans - improved (`COPI`) – this method differs to the original by comparing the new and old labels between iterations, rather than comparing the new and old means
- Constrained Spectral Clustering (`CSP`) – it is not a centroid-based method, which makes it a candidate for the task considered in our study
- Complete Average Linkage (`CAL`)
- Complete Complete Linkage (`CCL`)
- Complete Single Linkage (`CSL`)
- Constrained Clustering Ensemble (`CCEN`)

We deliberately did not include in the comparisons some recent, successful constrained clustering methods. While be-

TABLE I: Execution time (in milli-seconds) averaged across repetitions and datasets.

(a) Synthetic Data

PC	COP	COPI	CSP	CCL	CSL	CAL	CCEN
0	32	23	1164	8	6	5	12
1	33	29	1051	6	6	5	10
2	29	52	891	6	6	5	10
3	17	88	803	5	6	5	10
4	15	137	834	5	6	5	10
5	13	182	946	5	6	5	10
10	11	474	1254	6	6	6	14
15	11	832	1256	6	6	6	17
20	12	1317	1264	6	6	6	18

(b) Real Data

PC	COP	COPI	CSP	CCL	CSL	CAL	CCEN
0	40	39	6624	17	17	16	33
1	56	64	6285	17	18	17	28
2	28	166	6165	17	18	16	28
3	17	342	6403	17	18	16	28
4	15	578	6982	17	18	16	30
5	15	803	7300	17	18	17	32
10	12	2566	7471	18	17	18	39
15	13	5988	7473	18	17	18	41
20	15	10729	7499	18	18	18	41

longing to the hierarchical group, 3SHACC [33] proved to be too intricate and time-consuming to run in our experimental setup. Our main focus here is on identifying a fast and uncomplicated potential candidate for future online constrained clustering. We also dismissed centroid-based methods such as PCCC [34]. We acknowledge that they may perform well on spherical data such as our Real data collection.

The code used in the experiment is available on GitHub[2].

V. Results

Figure 2 show the ARI and NMI scores for each method across various proportions of constraints, averaged across the Synthetic and Real datasets. From these plots, it is evident that the simpler methods, viz. CAL, CCL and CSL, are more accurate than COP, COPI, CSP. Moreover, they exhibit progressive improvement as the number of constraints increases. Our proposed method, CCEN, enhances the accuracy compared to the simpler methods across most proportions of constraints, particularly when applied to the Real datasets.

Table I presents the execution times for the Synthetic and real datasets, respectively. Simpler clustering methods can produce a partition much quicker compared to their competitors. Moreover, their speed remains stable regardless of the number of pairwise constraints given to the algorithm. Notice that CCEN does not take much more time than its base method ($CCBM$) - CAL. This makes CCEN the preferred choice for real-time applications.

Additionally, we carried out an experimental study to find out whether increasing the number of clusterers in the ensemble has a significant effect. We compared the three hierarchical variants as base clusterers ($CCBM$) for the ensemble. The same experimental set-up was used as the one detailed above, with both the Synthetic and Real datasets. The results are shown in Figure 3. Each plot shows a cluster quality metric as a function of percentage of constraints, for ensembles of four different sizes: 1, 2, 4 and 6 clusterers. Smaller ensembles are shown with smaller markers and in a lighter colour. The largest ensemble is shown with largest markers. The three methods are displayed on the same plot to facilitate a visual comparison between the three candidates for their respective base clustering methods.

The figures show that larger ensembles are typically better than the single clusterer (ensemble of size 1), that is CCEN is better than any of the individual constrained clustering methods. The only exception is for the small proportion of constraints for the Synthetic data (Figures 3a and 3c), where smaller ensembles fared better in the experiments. We also observe that the difference between ensembles of sizes 4 and 6 is not substantial, which suggests that small ensembles can be both accurate and efficient for the purposes of future real-time clustering. The second observation is that CEAL is better than the others. This is expected, given that CAL was superior to CCL and CSL, as seen in Figure 2.

VI. Conclusion

We introduce a constrained cluster ensemble (CCEN) using basic hierarchical methods. Our extensive experimental analysis demonstrates accuracy gain, particularly as the number of constraints increases. Notably, it maintains efficiency suitable for online and real-time applications.

Our future project includes extending CCEN to real-time clustering for animal reidentification in videos.

Acknowledgements

This work is supported by the UKRI Centre for Doctoral Training in Artificial Intelligence, Machine Learning and Advanced Computing (AIMLAC), funded by grant EP/S023992/1.

J.F. Díez-Pastor was supported by a mobility grant PRX22/00634 from the Spanish Ministry of Universities.

References

[1] F. Liu, D. Chen, F. Wang, Z. Li, and F. Xu, "Deep learning based single sample face recognition: a survey," *Artificial Intelligence Review*, vol. 56, no. 3, pp. 2723–2748, 2023.

[2] S. Sivaraman and M. M. Trivedi, "Looking at vehicles on the road: A survey of vision-based vehicle detection, tracking, and behavior analysis," *IEEE transactions on intelligent transportation systems*, vol. 14, no. 4, pp. 1773–1795, 2013.

[3] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.

[2] Code available at https://github.com/frankmnb/Ensemble-hierarchical-method-for-constrained-clustering

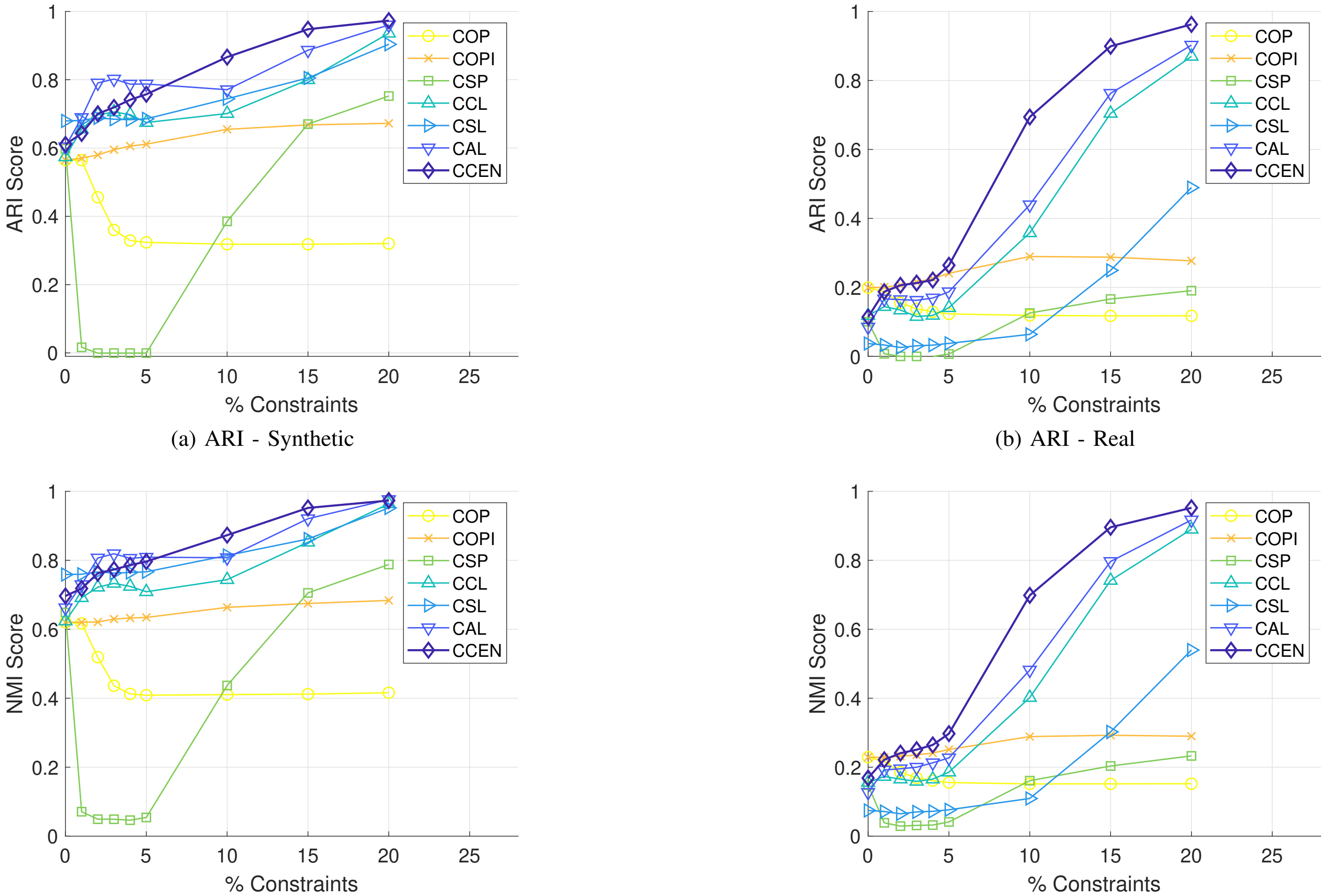


Fig. 2: Metric scores for the experimental methods for various values of PC, averaged across the datasets.

[4] P. Kulshreshtha and T. Guha, "An online algorithm for constrained face clustering in videos," in *2018 25th IEEE International Conference on Image Processing (ICIP)*. IEEE, 2018, pp. 2670–2674.

[5] G. González-Almagro, D. Peralta, E. De Poorter, J.-R. Cano, and S. García, "Semi-supervised constrained clustering: An in-depth overview, ranked taxonomy and future research directions," *arXiv*, 2023.

[6] K. Wagstaff and C. Cardie, "Clustering with instance-level constraints," *Proceedings of the Seventeenth International Conference on Machine Learning*, pp. 1103–1110, 2000.

[7] H. A. Kestler, J. M. Kraus, G. Palm, and F. Schwenker, "On the effects of constraints in semi-supervised hierarchical clustering," in *Artificial Neural Networks in Pattern Recognition*. Springer, 2006, pp. 57–66.

[8] T. Boongoen and N. Iam-On, "Cluster ensembles: A survey of approaches with recent extensions and applications," *Computer Science Review*, vol. 28, pp. 1–25, 2018.

[9] K. Golalipour, E. Akbari, S. S. Hamidi, M. Lee, and R. Enayatifar, "From clustering to clustering ensemble selection: A review," *Engineering Applications of Artificial Intelligence*, vol. 104, p. 104388, 2021.

[10] L. I. Kuncheva and D. P. Vetrov, "Evaluation of stability of k-means cluster ensembles with respect to random initialization," *IEEE transactions on pattern analysis and machine intelligence*, vol. 28, no. 11, pp. 1798–1808, 2006.

[11] J. Derrac, S. Garcia, L. Sanchez, and F. Herrera, "Keel data-mining software tool: Data set repository, integration of algorithms and experimental analysis framework," *J. Mult. Valued Logic Soft Comput*, vol. 17, pp. 255–287, 2015.

[12] I. Davidson and S. S. Ravi, "Using instance-level constraints in agglomerative hierarchical clustering: theoretical and empirical results," *Data Mining and Knowledge Discovery*, vol. 18, pp. 257–282, 2009.

[13] N. Grira, M. Crucianu, and N. Boujemaa, "Fuzzy clustering with pairwise constraints for knowledge-driven image categorisation," *IEE Proceedings-Vision, Image and Signal Processing*, vol. 153, no. 3, pp. 299–304, 2006.

[14] W. Zhao, Q. He, H. Ma, and Z. Shi, "Effective semi-supervised document clustering via active learning with instance-level constraints," *Knowledge and Information Systems*, vol. 30, pp. 569–587, 2012.

[15] I. Davidson and S. Ravi, "Agglomerative hierarchical clustering with constraints: Theoretical and empirical results," in *Knowledge Discovery in Databases: PKDD 2005*, 2005, pp. 59–70.

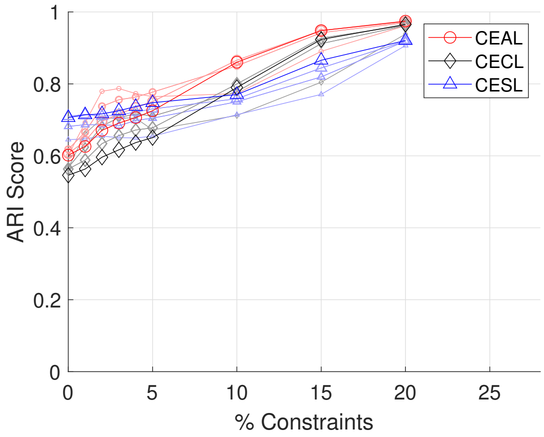
[16] S. Miyamoto and A. Terami, "Constrained agglomerative hierarchical clustering algorithms with penalties," in *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2011)*. IEEE, 2011, pp. 422–427.

[17] S. Gilpin and I. Davidson, "Incorporating sat solvers into hierarchical clustering algorithms: an efficient and flexible approach," in *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2011, pp. 1136–1144.

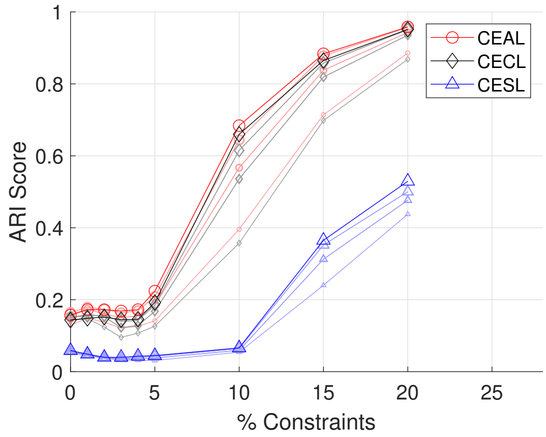
[18] A. Biswas and D. Jacobs, "Active image clustering with pairwise constraints from humans," *International Journal of Computer Vision*, vol. 108, pp. 133–147, 2014.

[19] T. Van Craenendonck, S. Dumančić, E. Van Wolputte, and H. Blockeel, "Cobras: Interactive clustering with pairwise queries," in *Advances in Intelligent Data Analysis XVII*. Springer, 2018, pp. 353–366.

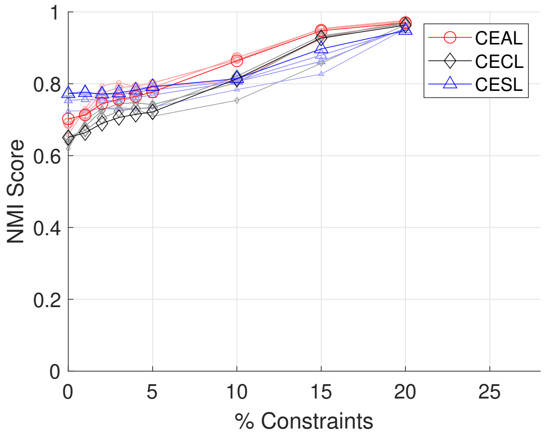
[20] G. González-Almagro, J. L. Suarez, J. Luengo, J.-R. Cano, and S. García, "Agglomerative constrained clustering through similarity



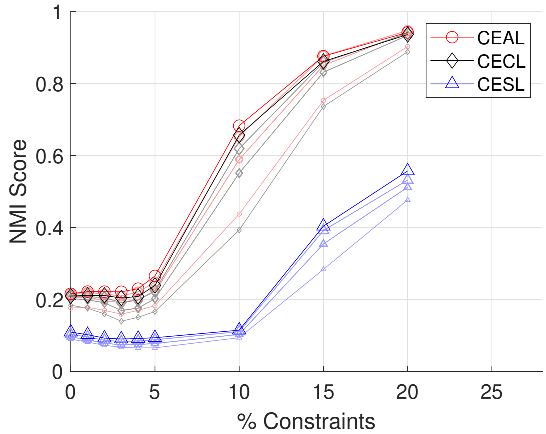
(a) ARI - Synthetic



(b) ARI - Real



(c) NMI - Synthetic



(d) NMI - Real

Fig. 3: Metric scores for the ensemble methods with $L = [1, 2, 4, 6]$ clusterers.

and distance recalculation," in *Hybrid Artificial Intelligent Systems*. Springer, 2020, pp. 424–436.

[21] E. Xing, M. Jordan, S. J. Russell, and A. Ng, "Distance metric learning with application to clustering with side-information," *Advances in Neural Information Processing Systems*, vol. 15, 2002.

[22] C. Zhong, L. Hu, X. Yue, T. Luo, Q. Fu, and H. Xu, "Ensemble clustering based on evidence extracted from the co-association matrix," *Pattern Recognition*, vol. 92, pp. 93–106, 2019.

[23] L. I. Kuncheva, J. J. Rodríguez, C. O. Plumpton, D. E. Linden, and S. J. Johnston, "Random subspace ensembles for fmri classification," *IEEE transactions on medical imaging*, vol. 29, no. 2, pp. 531–542, 2010.

[24] X. Z. Fern and C. E. Brodley, "Random projection for high dimensional data clustering: A cluster ensemble approach," in *Proceedings of the 20th international conference on machine learning (ICML-03)*, 2003, pp. 186–193.

[25] W. Zhuang, Y. Ye, Y. Chen, and T. Li, "Ensemble clustering for internet security applications," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 6, pp. 1784–1796, 2012.

[26] X. Hu and I. Yoo, "Cluster ensemble and its applications in gene expression analysis," in *Proceedings of the second conference on Asia-Pacific bioinformatics-Volume 29*, 2004, pp. 297–302.

[27] Y. Wang and Y. Pan, "Semi-supervised consensus clustering for gene expression data analysis," *BioData mining*, vol. 7, pp. 1–13, 2014.

[28] Y. Yang, F. Teng, T. Li, H. Wang, H. Wang, and Q. Zhang, "Parallel semi-supervised multi-ant colonies clustering ensemble based on mapreduce methodology," *IEEE Transactions on Cloud Computing*, vol. 6, no. 3, pp. 857–867, 2015.

[29] Z. Yu, H.-S. Wongb, J. You, Q. Yang, and H. Liao, "Knowledge based cluster ensemble for cancer discovery from biomolecular data," *IEEE transactions on nanobioscience*, vol. 10, no. 2, pp. 76–85, 2011.

[30] Y. Lai, S. He, Z. Lin, F. Yang, Q. Zhou, and X. Zhou, "An adaptive robust semi-supervised clustering framework using weighted consensus of random k k-means ensemble," *IEEE Transactions on Knowledge and data engineering*, vol. 33, no. 5, pp. 1877–1890, 2019.

[31] D. Klein, S. D. Kamvar, and C. D. Manning, "From instance-level constraints to space-level constraints: Making the most of prior knowledge in data clustering," in *Proceedings of the Nineteenth International Conference on Machine Learning*, 2002, pp. 307–314.

[32] K. Wagstaff, C. Cardie, S. Rogers, and S. Schrödl, "Constrained k-means clustering with background knowledge," in *Proceedings of the Eighteenth International Conference on Machine Learning*, 2001, pp. 577–584.

[33] G. Gonzalez-Almagro, J. L. Suárez, J. Luengo, J.-R. Cano, and S. García, "3SHACC: Three stages hybrid agglomerative constrained clustering," *Neurocomputing*, vol. 490, pp. 441–461, 2022.

[34] P. Baumann and D. S. Hochbaum, "PCCC: The pairwise-confidence-constraints-clustering algorithm," *arXiv preprint arXiv:2212.14437*, 2022.