CONSENSUS CLUSTERING ON DATA FRAGMENTS

S. Sukhanov, V. Gupta, C. Debes

AGT International
Darmstadt, Germany
{ssukhanov, vgupta, cdebes}@agtinternational.com

A.M. Zoubir

Signal Processing Group
Technische Universität Darmstadt, Germany
zoubir@spg.tu-darmstadt.de

ABSTRACT

Consensus clustering, also known as clustering ensembles is a technique that combines multiple clustering solutions to obtain stable, accurate and novel results. Over the last years several consensus clustering approaches were proposed addressing practical clustering problems with different degrees of success. In this paper, we consider data fragments as elements of a cluster ensemble framework. We propose a new dissimilarity measure on data fragments and build a consensus function that allows handling large scale clustering problems while not compromising on accuracy. We evaluate our proposed consensus function on a number of datasets showing its high performance with respect to other existing consensus functions.

Index Terms— consensus clustering, clustering ensembles, dissimilarity measure

1. INTRODUCTION

Data clustering is a very challenging unsupervised learning problem since it often requires exploratory analysis hindering the discovery of a proper single solution. Consensus clustering has emerged to be a powerful tool to solve practical data clustering problems [1]. Motivated by the success of supervised ensemble learning techniques [2] consensus clustering combines multiple clustering solutions to obtain a single final one. As a result, consensus clustering usually provides a more accurate and stable output [3]. Moreover, consensus clustering also allows obtaining novel solutions that are not achievable by any single clustering method as shown empirically [1, 4] and theoretically [5] in the past.

Given a set of objects, consensus clustering methods consist of two main steps: 1) Generation, in which a set of diverse clusterings is produced and 2) Consensus, where generated clusterings (or some of them) are combined. This combination is usually done without any access to original data on which these clusterings were generated. Finding a proper consensus function that accurately combines given partitions is usually considered to be the biggest challenge in clustering ensembles research [6, 7] and is also the focus of this paper.

In the state of the art there are two main groups of consensus function approaches: median partition and object co-occurrence. Approaches from the first group are searching for a consensus function that is a solution of an optimization problem. Here the objective is to maximize the sum of similarities between the median clustering and all clusterings in the given ensemble. The choice of similarity measure between partitions is the key challenge that is not yet fully addressed [8]. For this reason, there is no answer on which approach should be used to accurately solve practical large-scale consensus clustering problems using median partition-based methods. The second set of methods is based on object co-occurrence which operates on pairs of objects and analyzes whether two objects belong to the same cluster in every partition or not [9]. Despite its solid operational performance the main drawback of co-occurrence based methods is their high computational and memory complexity (usually not less than $\mathcal{O}(N^2)$) and the fact that co-association matrices are not expressive enough to accurately perform aggregation on them (especially when there are not too many ensemble members or they are of limited diversity) [8].

Recently, due to large data volumes and a demand for higher scalability, the data fragment (DF) concept was introduced in several works on consensus clustering. Employing DF allows pruning the search space for median partition-based ensemble aggregation methods [10] as well as to decrease both memory and time complexity for co-occurrence based approaches. In [11] a data fragment-based consensus method called CA-Tree is introduced where both the dendrogram (a tree diagram representing clustering results) and co-association matrix (an indicator matrix reflecting if a pair of data points are coclustered) are used to obtain a consensus solution. The main drawback of the method is high sensitivity to a partition that is taken as the first layer of the dendrogram rendering results unstable. In [12] three methods adopted from data objects to DFs are presented: a bottom-up agglomerative algorithm F-agglomerative, a top-down approach F-Furtherst and a median partition based local-search heuristic F-LocalSearch. These three methods are all adoptions of correspondent object-based approaches and inherit drawbacks of the original methods: they treat all partitions equally even if some of them are nonsense. In addition, they employ the

Hamming distance [13] as a distance measure which results in non-expressive and quantized representations of distances often leading to a non-optimal consensus solution.

In this paper, we propose a consensus clustering framework that addresses the drawbacks of the Hamming distance in co-occurance based method and reduces computational and memory complexity. We employ a DF concept to assure scalability of the consensus clustering function while proposing an expressive distance measure on DFs that leads to a significant improvement in the final solution compared to approaches based on Hamming distance. We further build a consensus function around this measure based on a hierarchical clustering approach.

The paper is structured as follows. In Section 2, the notation and DF concept are introduced. Section 3 presents the proposed dissimilarity measure and the consensus function that is built around it. Further, we evaluate the consensus function in Section 4 both on synthetic and real-world datasets and discuss the results. Section 5 concludes our paper.

2. DATA FRAGMENTS

Let $X = \{x_1, x_2, \ldots, x_n\}$ be a set of n objects, where x_i is a vector in a d-dimensional feature space $\mathcal{R}^d$. $\mathbb{P} = \{P_1, P_2, \ldots, P_H\}$ is a set of H partitions (or clusterings) of X, where each $P_h = \{C_1^h, C_2^h, \ldots, C_{K_h}^h\}$ is a single partition of X with K_h clusters each of them being a disjoint nonempty subset of X with the union of C_k^h, $k = 1, \ldots, K_h$ being X. For each x_i we define a H-dimensional label vector y_i:

$$y_i = [P_1(x_i), P_2(x_i), \ldots, P_H(x_i)] \tag{1}$$

where $P_h(x_i)$ is the cluster label of x_i in partition P_h. Since every clustering P_h assigns symbolic labels to objects x_i the vectors y_i are categorical data vectors. In addition, all vectors y_i are forming a set $Y = \{y_1, y_2, \ldots, y_n\}$. Having introduced Y, a DF can be defined as follows.

Definition. DF $F_l, l = 1, \ldots, L$ is a subset of Y in which all vectors are equal to each other, i.e. $y_i = y_j \; \forall y_i, y_j \in F_l$.

One DF can be considered as a stable group of objects within the ensemble of clusterings where all included points are co-clustered. We will refer to each fragment F_l by its member $f_l \in F_l$. The amount of DFs in an ensemble is usually much less than the number of original data points which allows handling large datasets [10, 12]. Moreover, with the help of DFs it is possible to naturally extend object co-occurrence- and median partition-based frameworks [12] and effectively use them as elements of a consensus function for ensemble aggregation [14]. The fact that every DF represents a set of stable clustered objects a proper (dis)similarity measure between DFs should be established in order to perform a reasonable consensus among partitions. In addition, since f_l is a categorical data vector, the (dis)similarity has to be defined over this type of data which in general is a challenging task [15]. The cardinality of each set F_l corresponds to the amount of data points that are co-clustered by all ensemble members. From the DF definition it is also clear that $\sum_{l=1}^{L} |F_l| = N$. Intuitively, DFs that have large $|F_l|$ are likely to form stable clusters or substantial parts of them. DFs with small $|F_l|$ correspond to objects on which the consensus is weak (outliers and noisy samples). At the same time, it is important to account for the frequency of each label within each clustering since the distribution of them is in general different, can be imbalanced and depends on the ensemble generation scheme and underlying data structure. In the next section, we propose a dissimilarity measure between DFs that addresses this peculiarity.

3. DISSIMILARITY OF DATA FRAGMENTS

In order to be able to operate on DFs and establish a dissimilarity measure between them to find consensus partition we first define an error function on categorical vectors y_i which we want to minimize. For any partition P with K clusters and a dissimilarity measure between two categorical vectors d_c an error function

$$E(P) = \sum_{k=1}^{K} \frac{1}{|C_k| \cdot (|C_k| - 1)} \sum_{\substack{y_i, y_j \in C_k, \\ y_i \neq y_j}} d_c(y_i, y_j) \tag{2}$$

measures the aggregated average dissimilarity between points of every cluster and closely related to optimization criteria of k-means and agglomerative clustering algorithms [16]. The choice of an appropriate dissimilarity measure d_c is critical since it can seriously affect the behavior of the error function. The optimal partition P^* can be then defined as follows:

$$P^* = \underset{P \in \mathcal{P}_x}{\operatorname{argmin}} \, E(P) \tag{3}$$

where $\mathcal{P}_x$ is a search space with all possible clusterings of Y.

In previous works on consensus clustering in which the notion of distance between two categorical vectors y_i and y_j was defined [17] mainly the Hamming distance was considered. In fact, the Hamming distance (sometimes also called as overlap measure when introduced as a similarity measure [15]) provides an easy and understandable way to compare two categorical vectors. However, it suffers from a substantial drawback since it assigns equal significance to dissimilarities for all vectors attributes. For many problems (including consensus clustering) the assumption of equal significance of attributes errors is not valid as the partitions in ensemble could be very diverse (every partition has its own number of clusters K_h and may be generated using different distance metrics and clustering methods). Generally, measuring dissimilarities between categorical vectors is not a straightforward task since the content of dissimilarity is highly application specific and categories are often ambiguous or even arbitrary. As an alternative to the Hamming

distance there are several data-driven dissimilarity functions that take into account the frequency distribution of values of every attribute. In [15] the authors systematically studied 14 measures for categorical values concluding that the choice of them strongly depends on the assumptions that are imposed on the data. In the sequel, we define a dissimilarity measure on DFs introduced in Section 2 and use DFs further in Equation 2 instead of data point labels.

A general distance measure between two categorical vectors f_i and f_j representing DF can be defined as:

$$d(f_i, f_j) = \sum_{h=1}^{H} w_h \cdot d_h(f_i^h, f_j^h) \tag{4}$$

where w_h is the weight for every hth attribute and $d_h(f_i^h, f_j^h)$ defines the dissimilarity between values of this attribute. Since f_i and f_j are vectors representing their respective sets we propose accounting for unequal distributions of attribute values of every attribute. For every DF F_i and partition h we define a significance value S_i^h according to:

$$S_i^h = \frac{|F_i|}{|C_{f_i^h}| \cdot K_h} \tag{5}$$

where $|\cdot|$ is the cardinality of a set and $C_{f_i^h}$ is the cluster with label f_i^h in partition h. A significance value S_i^h shows the relative amount of co-clustered data points in the DF i of the partition h with respect to the number of objects with the same label that are clustered differently assuming that every cluster has equal importance. The dissimilarity d_h between an attribute of two DFs is then defined as:

$$d_h(f_i^h, f_j^h) = \begin{cases} 0, & f_i^h = f_j^h \\ 1 - (\frac{2}{K_h} - S_i^h - S_j^h), & \text{otherwise} \end{cases} \tag{6}$$

which compares the significance of two DFs with doubled significance of a case when a DF occupies whole cluster. In general, the more data points a DF shares with a cluster, the higher the certainty that this DF is a substantial subset of a cluster. As a result, it ends up in a higher distance with other DFs. In addition the proposed dissimilarity considers equal importance of every cluster within a partition independently of the amount of objects assigned to it effectively allowing comparing DFs with different amount of objects. We note that $d_h \in [0, 1]$ and is symmetric (i.e. $d_h(f_i^h, f_j^h) = d_h(f_j^h, f_i^h)$). In Equation 4 the weights w_h are assigned to each attribute h to signify its relative importance. Since the partitions in the ensemble can have different quality level we employ w_h as the degree of agreement of a particular partition P_h with all the partitions in the ensemble $\mathbb{P}$. For that we define a distance measure between two partitions of the given ensemble using their respective connectivity matrices [18], however, formulated on DFs instead of object labels:

$$d(P_1, P_2) = \sum_{i,j}^{L} |M_{ij}(P_1) - M_{ij}(P_2)| \times |F_i| \times |F_j| \tag{7}$$

where $M_{ij}(P_i)$ is the connectivity matrix on DFs that is defined as:

$$M_{ij}(P_h) = \begin{cases} 1, \exists C_k^h \in P_h \mid f_i^h \in C_k^h \text{ and } f_j^h \in C_k^h \\ 0, \quad \text{otherwise} \end{cases} \tag{8}$$

Finally, using Equation 7 for every attribute $h' \in (1, \ldots, H)$ we define its weight $w_{h'}$ as follows:

$$w_{h'} = \frac{\sum_{h=1}^{H} w_h - w'_{h'}}{\sum_{h=1}^{H} w_h} \tag{9}$$

where $w'_{h'}$ is calculated as:

$$w'_{h'} = \sum_{h=1}^{H} d(P_{h'}, P_h) \tag{10}$$

and $\sum_{h=1}^{H} w_h = 1$ holds.

As a result, we obtained a dissimilarity measure over DFs (Equation 4) that provides expressive distance between categorical vectors. The proposed dissimilarity measure is symmetric and non-negative and can be used to construct co-association matrices commonly used in clustering ensembles and create a (dis)similarity matrix that summarizes richer information than the original one.

To establish a consensus function using dissimilarity measure proposed and solve Problem 3 on DFs we employ agglomerative clustering with between-group average linkage [16] on the dissimilarity matrix obtained by applying the proposed dissimilarity measure to all DFs within the ensemble. We call this consensus function DF-based Expressive Consensus (DFEC). The proposed dissimilarity measure can be used with various clustering methods (e.g. EM-based clustering algorithms [8]), however, due to space limitations, in this paper we consider only hierarchical clustering as a consensus function.

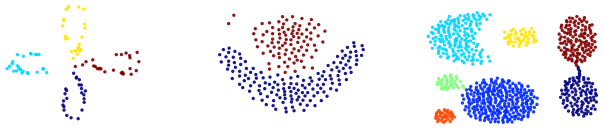
4. EXPERIMENTAL RESULTS

To study the effectiveness of the consensus function based on our proposed dissimilarity measure and compare its performance with the state-of-the-art consensus clustering methods we conduct two sets of experiments using both synthetic and real-world datasets. All datasets are provided with the ground truth (class labels). For both experiments in order to generate diverse input partitions we use multiple clustering algorithms [16] also varying their parameters: k-means (with random initialization), hierarchical clustering (with random linkage and number of neighbors), affinity propagation (with random damping factor and iterations), BIRCH (with random threshold), DBSCAN (with random eps factor) and mean shift (with random bandwidth). For k-means and hierarchical clustering the number of clusters provided was chosen randomly on the uniform interval [2, true cluster count +2]. As a result, every ensemble consists of ten partitions which are assured to be distinct and different from the ground truth. Such diversity in input partitions does not allow particular clustering results to dominate and thus helps to evaluate the stability of a consensus clustering approach and see whether it is capable of

providing a novel solution with respect to the input partition. We chose the cutting threshold for DFEC as well as resulting number of clusters for other consensus clustering methods according to the true number of clusters.

4.1. Synthetic datasets

In this experiment we compare the performance of the proposed DFEC with other DF-based consensus clustering methods: CA-Tree, F-Agglomerative, F-Furthest and F-LocalSearch using four synthetic datasets: petals [19], aggregation [17], flame [20] and dim32 [21]. In Table 1 we report the Adjusted Rand Index (ARI) that is widely used for clustering evaluation and related to the accuracy measure while operating on pairs of elements and adjusted for chance [22]. We also provide graphical results for our proposed method DFEC in Figure 1 to demonstrate its ability to establish consensus (note that a figure for dataset dim32 is not provided since the dimension of the original data is 32).



(a) Petals (b) Flame (c) Aggregation

Fig. 1: Consensus solution using proposed method DFEC

4.2. Real datasets

In our second sets of experiments, we a use number of real-world datasets taken from the UCI repository [23] that are widely used in consensus clustering research, namely breast cancer, thyroid, wine, wdbc and seeds. In addition to the methods evaluated in the synthetic experiment, we evaluate CSPA, HGPA, MCLA [1], CTS, SRS, ASRS [14], HBGF [24] and knowledge based (KB) [25] methods. To evaluate the quality of the final consensus solution besides ARI we employ two other commonly used external validity indexes: Impurity Index (IMP) which reflects the amount of differently labeled points in clusters [17] and Average Entropy (AE) which is defined similarly to the entropy used in traditional decision tree building [26]. Note that for the quality solutions, ARI should be large while IMP and AE should be as low as possible. We summarize the results of this experiment in Table 2.

Table 1: Adjusted Rand Index on synthetic datasets

	Petals	**Flame**	**Aggregation**	**Dim32**
CA-Tree	0.689	0.319	0.649	0.603
F-Agglomerative	0.468	0.078	0.318	0.399
F-Furthest	0.429	0.458	0.615	0.149
F-Local Search	0.091	0.029	0.040	0.000
Proposed approach	**0.973**	**0.876**	**0.876**	**0.925**

4.3. Discussion

Analyzing the evaluation results (Table 1 and 2) it can be seen that in most cases proposed DFEC outperforms other methods in terms of ARI. For some datasets IMP and AE are not the lowest for DFEC (however comparable with the winning ones). The main reason for such behavior is the fact that these measures are biased by different aspects of clustering: IMP considers majority-class points in each cluster, AE focuses on distribution of all labels in each cluster while ARI is related to classical accuracy measure. The opposite effect can be observed for F-Furthest, however, the reason for low ARI for this method is the strong initialization dependence. The superior performance of the proposed DFEC is achieved thanks to the introduced distance measure that takes into account the quality of input partitions as well as significance value for each attribute of DF (Equation 5). The results on synthetic datasets also demonstrate that novel solutions can be found by DFEC. Finally, Table 2 also confirms that the DF concept allows to significantly decrease the amount of points on which aggregation is performed (L is much lower than N) allowing for larger datasets.

5. CONCLUSIONS

In this paper, we addressed the consensus clustering problem by proposing a dissimilarity measure on data fragments. We built a scalable consensus function that utilizes this distance measure and performed evaluation on both synthetic and real-world datasets achieving novel and accurate results that in general outperform the ones of other state-of-the-art methods. Since the proposed approach does not impose any assumptions on the original data distributions it could be applied to clustering problems from different domains. A statement that is also supported by the fact that the real-world datasets evaluated in this paper are coming from various application domains.

Table 2: Evaluation results using real-world data sets (N - original number of data points, L - number of DFs)

	breast cancer ($N = 698, L = 81$)			**thyroid** ($N = 214, L = 16$)			**wine** ($N = 177, L = 21$)			**wdbc** ($N = 568, L = 131$)			**seeds** ($N = 210, L = 22$)		
	ARI	IMP	AE	ARI	IMP	AE	ARI	IMP	AE	ARI	IMP	AE	ARI	IMP	AE
CSPA	0.017	0.967	0.083	0.155	0.453	0.902	0.231	0.588	0.647	0.232	0.599	0.414	0.457	0.426	0.504
HGPA	0.017	0.960	0.094	0.097	0.439	0.885	0.304	0.492	0.676	-0.001	0.746	0.693	0.262	0.536	0.711
MCLA	0.000	0.000	0.000	0.546	0.164	0.522	0.226	0.542	0.560	0.224	0.537	0.295	0.496	0.281	0.388
CA-Tree	0.477	0.388	0.206	0.281	**0.112**	0.325	**0.389**	0.282	0.619	0.490	0.146	0.260	0.545	0.150	0.321
CTS	0.088	0.868	0.064	0.442	0.206	0.523	0.285	0.463	0.557	0.512	0.264	0.294	0.633	0.220	0.357
SRS	0.096	0.853	0.105	0.442	0.206	0.523	0.302	0.475	0.668	0.438	0.405	0.515	0.623	0.239	0.398
ASRS	0.127	0.776	0.056	0.578	0.140	0.394	0.310	0.463	0.534	0.519	0.202	0.142	0.611	0.172	0.289
HBGF	0.042	0.907	0.075	0.386	0.243	0.669	0.261	0.554	0.577	0.338	0.484	0.295	0.520	0.347	0.421
KB	-0.006	0.293	0.051	0.517	0.196	0.498	0.128	0.305	0.590	0.109	0.194	0.457	0.191	0.307	0.499
F-Agglomerative	0.191	0.637	0.087	0.273	0.458	0.606	0.169	0.684	0.312	0.264	0.539	0.150	0.316	0.567	0.261
F-Furthest	0.729	0.165	0.175	0.256	0.336	0.616	0.294	0.418	0.375	0.322	0.431	0.392	0.485	0.277	0.354
F-Local Search	-0.002	**0.013**	**0.016**	0.196	0.117	**0.279**	-0.008	0.311	**0.256**	0.241	0.141	**0.112**	0.107	**0.124**	**0.207**
DFEC	**0.845**	0.040	0.166	**0.601**	0.131	0.371	0.367	**0.291**	0.412	**0.594**	**0.113**	0.254	**0.665**	**0.124**	0.346

6. REFERENCES

[1] A. Strehl and J. Ghosh, "Cluster ensembles — a knowledge reuse framework for combining multiple partitions," *J. Mach. Learn. Res.*, vol. 3, pp. 583–617, 2003.

[2] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996.

[3] A. Goder and V. Filkov, "Consensus clustering algorithms: Comparison and refinement," in *Proceedings of the Meeting on Algorithm Engineering & Expermiments*. Society for Industrial and Applied Mathematics, 2008, pp. 109–117.

[4] I. T. Christou, "Coordination of cluster ensembles via exact methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 2, pp. 279–293, 2011.

[5] A. P. Topchy, M. H. C. Law, A. K. Jain, and A. L. Fred, "Analysis of consensus partition in cluster ensemble," in *Proceedings of the Fourth IEEE International Conference on Data Mining*, 2004, ICDM '04, pp. 225–232.

[6] A. Topchy, A. K. Jain, and W. Punch, "Clustering ensembles: models of consensus and weak partitions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 12, pp. 1866–1881, 2005.

[7] 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.

[8] S. V. Pons and J. R. Shulcloper, "A survey of clustering ensemble algorithms," *International Journal of Pattern Recognition and Artifical Intelligence*, vol. 25, no. 3, pp. 337–372, 2011.

[9] A. L. N. Fred and A. K. Jain, "Combining multiple clusterings using evidence accumulation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 6, pp. 835–850, 2005.

[10] S. Vega-Pons and P. Avesani, "On pruning the search space for clustering ensemble problems," *Neurocomputing*, vol. 150, Part B, pp. 481 – 489, 2015.

[11] T. Wang, "Ca-tree: A hierarchical structure for efficient and scalable coassociation-based cluster ensembles," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 41, no. 3, pp. 686–698, 2011.

[12] O. Wu, W. Hu, S. J. Maybank, M. Zhu, and B. Li, "Efficient clustering aggregation based on data fragments.," *IEEE Trans. Systems, Man, and Cybernetics, Part B*, vol. 42, no. 3, pp. 913–926, 2012.

[13] R. W. Hamming, "Error-detecting and error-correcting codes," in *Bell System Technical Journal*, 1950, vol. 29(2), pp. 147–160.

[14] I. Natthakan and S. Garrett, "Linkclue: A matlab package for link-based cluster ensembles," *Journal of Statistical Software*, vol. 36, no. 1, pp. 1–36, 2010.

[15] S. Boriah, V. Chandola, and V. Kumar, "Similarity measures for categorical data: A comparative evaluation," in *Proceedings of the SIAM International Conference on Data Mining, SDM, Atlanta, Georgia, USA*, 2008, pp. 243–254.

[16] C. Shah and A. Jivani, "Comparison of data mining clustering algorithms," in *2013 Nirma University International Conference on Engineering (NUiCONE)*, 2013, pp. 1–4.

[17] A. Gionis, H. Mannila, and P. Tsaparas, "Clustering aggregation," *ACM Trans. Knowl. Discov. Data*, vol. 1, no. 1, 2007.

[18] T. Li, C. Ding, and M. I. Jordan, "Solving consensus and semi-supervised clustering problems using nonnegative matrix factorization," in *Seventh IEEE International Conference on Data Mining (ICDM 2007)*, 2007, pp. 577–582.

[19] L. Kuncheva, "Artificial data sets [online]," http://pages.bangor.ac.uk/m̃as00a/activities/artificial_data.htm, 2016.

[20] L. Fu and E. Medico, "Flame, a novel fuzzy clustering method for the analysis of dna microarray data," *BMC Bioinformatics*, vol. 8, no. 1, pp. 1–15, 2007.

[21] P. Franti, O. Virmajoki, and V. Hautamaki, "Fast agglomerative clustering using a k-nearest neighbor graph," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, vol. 28, pp. 1875–1881.

[22] L. Hubert and P. Arabie, "Comparing partitions," *Journal of Classification*, vol. 2, no. 1, pp. 193–218, 1985.

[23] M. Lichman, "UCI machine learning repository," 2013.

[24] X. Z. Fern and C. E. Brodley, "Solving cluster ensemble problems by bipartite graph partitioning," in *Proceedings of the International Conference on Machine Learning*, 2004.

[25] Z. Zhi-Hua and T. Wei, "Clusterer ensemble," *Knowl.-Based Syst.*, vol. 19, no. 1, pp. 77–83, 2006.

[26] M. Zhang and Z. Zhou, "Multi-instance clustering with applications to multi-instance prediction," *Applied Intelligence*, vol. 31, no. 1, pp. 47–68, 2009.