MULTI-VIEW DOCUMENT CLUSTERING WITH DIFFERENT SIMILARITY MEASUREMENTS VIA ENSEMBLE

P.Divya[1], C.Ranjani[2], M.Kalaiarasu[3], Dr. R.RadhaKrishnan[4]

[1]PG scholar , [2]Assistant Professor,Department of CSE,
Sri Ramakrishna Engineering College, Coimbatore(India)
[3]Associate Professor, Department of IT,Sri Ramakrishna Engineering College, Coimbatore (India)
[4]Principal, VidhyaMandhir Institute of technology, Erode, Tamilnadu (India)

ABSTRACT

Clustering ensembles are a common approach to clustering problem, which combine a collection of clustering into a superior solution. The key issues are how to generate different candidate solutions and how to combine them. Common approach for generating candidate clustering solutions ignores the multiple representations of the data and the standard approach of simply selecting the best solution from candidate clustering solutions ignores the fact that there may be a set of clusters from different candidate clustering solutions which can form a better clustering solution.Multi view clustering can be applied at various stages of the clustering paradigm. This paper proposes a novel multi-view clustering algorithm that combines different ensemble techniques via various similarity metrics have been used to measure the similarity between data objects. Cluster based similarity matrix is used for ensemble clustering by the use of a consensus function. This consensus function is based on the theory of graph partitions. Normalized Discounted Cumulative Gain (NDCG) which is a family of ranking measures widely used in practice. Kullback–Leibler divergence (KLD) is a non-symmetric measure of the difference between two probability distributions P and Q of two distance values between data object.Based on these similarity matrices on the individual datasets and aggregates these to form a combined similarity matrix, which is then used to obtain the final clustering. A proposed ensemble method is compared with existing algorithms on three data sets that have increasing difficulty. The results show that our method significantly outperforms other methods.

Keywords: Affinity matrix, Ensemble clustering, Kullback–Leibler divergence (KLD),Normalized Discounted Cumulative Gain (NDCG), Similarity matrices.

I INTRODUCTION

Clustering is a key issue in intelligence science and is widely used in the field of artificial intelligence. The technique has been studied for several decades in areas of pattern recognition, machine learning, applied statistics, communications and information theory. It is applied to numerous fields of applications including data mining, text mining, bio-informatics, image analysis and segmentation, data compression, and data classification. Clustering is an unsupervised learning technique for organizing similar objects into different groups. Since it is hard to define the similarity especially in high-dimensional data, thousands of clustering algorithms have been proposed in the last 50 years [1]. As no single clustering algorithm is suitable for all types of problems, researchers have been trying different techniquesfor combining different clustering algorithms (clustering ensembles) [2-5].

Multi-view clustering explores and exploits multiple views simultaneously in order to obtain a more accurate and robust partitioning of the data than single view clustering. There exist two methods in multi-view clustering: centralized and distributed [6]. Centralized algorithms simultaneously use all views to cluster the data while distributed algorithms cluster each view independently from others, using a single view algorithm, and then combine the individual clustering to obtain a final partitioning. During the past decade, Bickel and Scheffer [7] developed a two-view EM and a two-view spherical k-means algorithm under the assumption that the views are independent. De Sa [8] proposed a two-view spectral clustering algorithm that creates a bipartite graph and is based on the "minimizing-disagreement" idea. Kumar et al. [9] proposed a co-training approach for multi- view spectral clustering, co-regularized multi-view spectral clustering [10] and kernel-based weighted multi-view clustering [11].

The main goal of clusteringensembles is to solve the problem of producing superior clustering solution from given set of clustering solutions. This problem was previously approached by researchers from different angles and so far the best known approach for clustering ensembles is median partition based approach in which a single candidate clustering solution that has the maximum similarity from all candidate clustering solutions is selected as the final clustering solution. The clustering ensembles methods include two important steps:

1) Generating a set of candidate clustering solutions
2) Combining the set of candidate clustering solutions to generate final clustering solution.

In our evolutionary based clustering approach, step 1 corresponds to an initialization phase in which a set of initial candidate clustering solutions is generated, and step 2 is the evolutionary phase in which the final solution is evolved from the initial candidates. However, by using ensemble clustering can produce a more consistent and more accurate solution.

In this paper, propose a novel multi-view clustering framework based on ensemble clustering. It first generates multiple partitions from each of the single view of a multi-view dataset. Clustering algorithms are applied on the different data matrices to obtain partitions of the data. These partitions are used to generate a set of 5 different similarity matrices such as Cluster Based Similarity Matrix, Affinity Matrix, Pair-wise Dissimilarity Matrix, NDCG and KLD. Ensemble technique to aggregate these and form a new similarity matrix, which is then used for the clustering task

II RELATED WORKS

Clustering has been extensively studied in the literature in many domains, such as in information retrieval to cluster documents, in bio-informatics [12] to cluster genes, in social network analysis [13] to find communities, etc. The basic idea of merging the clustering results from different algorithms evolved as a different field of study for improvement of clustering results. Combining the clustering results of different clustering algorithms is a new clustering framework that is more robust and less susceptible to the adverse effects of each of the single view clustering algorithm.

Most of these clustering frameworks consist of finding similarity/distance matrix on the original dataset and using this to combine data samples into groups or clusters. In recent times a number of authors have proposed different multi-view clustering algorithms [13]. Janssens et al [14] proposed a hybrid clustering method which is based on statistical Meta analysis using Fisher's inverse chi- Square method. In this technique, the distances of data sources are converted into *p*-value by using CDF. These values computed against a randomized data having similar statistical characteristics. The p-values are then converted into a unified p-value using a logarithmic function, which is then used for clustering. Weighted Hybrid Clustering algorithm [13] proposes two steps for combining multiple similarity matrices: Weighted Kernel Fusion clustering and Weighted Ensemble Clustering. In the kernel fusion technique, kernel functions are used to compute the similarity matrices in higher dimensions for each of the data view.

Hong et.al [15] proposes a novel clustering ensembles method, termed as resampling-based selective clustering ensembles method. The proposed selective clustering ensembles method works by evaluating the qualities of all obtained clustering results through resampling technique and selectively choosing part of promising clustering results to build the ensemble committee. Recently, Azimi and Fern [16] proposed an adaptive cluster ensembles method. In contrast, some studies also indicated that medium diversity leads to the best performing ensembles. First generates a diverse set of solutions and combines them into a consensus partition P*. Based on the diversity between the ensemble members and P*, a subset of ensemble members is selected and combined to obtain the final output. Strehl and Ghosh [17] have developed three different consensus functions based on hypergraph for ensemble learning: cluster-based similarity partitioning algorithm, hypergraph-partitioning algorithm, and Meta clustering algorithm. Topchy et al. [3] designed a consensus function based on a finite mixture model. The final partition is found as a solution to a maximum likelihood problem for a given clustering ensembles. Ensemble method, SELective Spectral Clustering Ensemble (SELSCE), is proposed [18].. After the generation of component clustering, the bagging technique, usually applied in supervised learning. Randomly pick part of the available clustering's to get a consensus result and then compute normalized mutual information (NMI) between the consensus result and the component clustering.

III PROPOSED ENSEMBLE TECHNIQUES AND SIMILARITY MEASUREMENTS

We propose an ensemble technique to combine different multi-view clustering algorithms. The proposed work of multi-view clustering consists of both intermediate and late integration.From late integration use the Cluster Based Similarity Matrix and Pairwise Dissimilarity Matrix from the partitions. On the other hand, the Affinity matrix generated form distance matrices is an intermediate integration technique, Normalized discount score cumulative gain value is calculated to determine the similarity value of the data object and the pairwise dissimilarity is also measured using this NDCG. In addition some other additional KLD distance matrix is also used to measure the similarity and dissimilarity value of the data object between two cluster data points' samples in the documents. The proposed ensemble clustering framework with NDCG and KLD performs well than the general multiview clustering methods and each NDCG and KLD with its own benefits as discussed later in this section. The architectural view representation of the proposed ensemble multi-view learning approach for clustering documents is shown in Fig 1

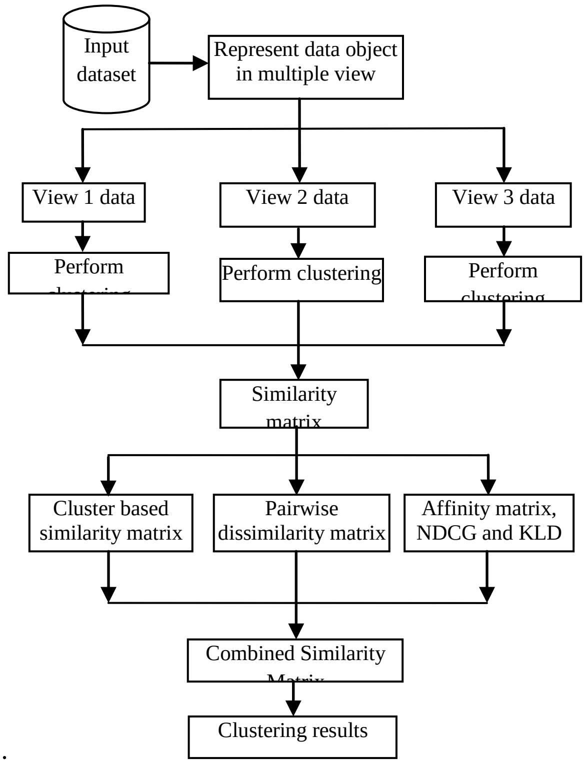

Fig 1: Architectural view representation of the proposed ensemble multi-view learning approach

3.1 Pre-Processing

Standard pre-processing steps typically used in text mining applications which include stop word removal and Stemming. The stop words include a list of commonly used English language words such as *a*, *the*, *was*, etc which do not contribute to the analysis of the categorical information within a document. Similarly, stemming is used to convert different forms of words into a base word using Porter's stemming algorithm. TF/IDF where the IDF factor is computed based on the inverse corpus frequency rather than the document frequency is also applied to pre-process the dataset samples.

Late integration uses the Cluster Based Similarity Matrix (CBSM) and Pairwise Dissimilarity Matrix(PDM) from the partitions. On the other hand, the Affinity matrix generated form distance matrices is an intermediate integration technique. These algorithms have been chosen since they have been shown to perform well and represent diversified approaches, each with its own benefits.

3.2 Cluster based similarity matrix

The Cluster Based Similarity Matrix (CBSM) [19] algorithm uses relationship between objects in the hyper adjacency matrix and calculates a new cluster based similarity matrix. This new matrix, SH, is an $m \times m$ similarity matrix which is obtained by performing a (sparse) matrix multiplication of the hyper adjacency matrix, given by:

$$S_H = \frac{1}{k}(H * H^T) \tag{1}$$

where k is total number of partitions and H^T represents the transpose of the matrix **H**. The values in S_H correspond to the ratio of partitions that have classified objects together. The pair wise dissimilarity matrix [20] tends to differentiate objects that have been partitioned separately by different algorithms. Thus objects that are not partitioned together are pushed further apart in the resulting dissimilarity matrix. The rows (or columns) of the matrix **D** can be thought of as a distribution representing the pair-wise dissimilarity matrix (PDM) of an object x_i with other objects x_j, $j \in 1..m$. Compute the similarity matrix, S_{PDM}, as the similarity between objects in the PDM, D where d_i is a row vector corresponding to x_i :

$$S_{PDM}(i,j) = Cos(\vec{d}_i, \vec{d}_j) = \frac{\vec{d}_i, \vec{d}_j{}'}{\sqrt{(\vec{d}_i, \vec{d}_i{}')(\vec{d}_j, \vec{d}_j{}')}} \tag{2}$$

3.3 Affinity Matrix (AFF)

A data matrix can be represented using a bi-partite graph where one set of vertices represent the objects and the set their features. An edge denotes the strength of relationship between a given object and the corresponding feature. The similarity of any two pair of vertices is then influenced locally by the relationship between edges.The basic equation for the calculation of affinity matrix is given below:

$$S_{AFF}(i,j) = \exp\left(-\frac{d_{ij}^2}{c}\right) \tag{3}$$

Where d_{ij}^2 the distance between pair of is points i and j and c is scaling factor

3.4 Kullback–Leibler Divergence for Similarity and Pairwise Similarity Matrix

Kullback–Leibler divergence (KLD) [22] is a non-symmetric measure of the difference between two probability distributions P and Q of two distance values between data object. Specifically, the Kullback–Leibler divergence of Q from P, denoted DKL (P‖Q), is a measure of the information lost when Q is used to approximate P. The KL divergence measures the expected number of extra bits required to code samples from P when using a code optimized for Q, rather than using the true code optimized for P. Typically P represents the "true" distribution of data, observations, or a precisely calculated theoretical distribution. The measure Q typically represents a theory, model, description, or approximation of P. Given two probability distributions $p(x)$ and $q(x)$, x be the query term belongs to topic p and q over the same alphabet, the KL divergence or relative entropy DKL (P‖Q)is defined as:

$$\text{DKL}\ (\text{P}\|\text{Q}) = E_p + log\frac{P(x)}{q(x)} \tag{4}$$

The KL divergence is often referred to as relative entropy, as it may be regarded as a generalization of the entropy of a distribution, relative to another. If the above mentioned step query terms belongs to topic p is high means then it is considered as the similar object in the cluster .

3.5 Normalized Discounted Cumulative Gain (NDCG) for Similarity and Pairwise Similarity Matrix

Normalized Discounted Cumulative Gain (NDCG) [23] which is a family of ranking measures widely used in practice. Although there are extensive empirical studies of the NDCG family, little is known about its theoretical properties. This result is surprising. On the first sight it seems to mean that the widely used standard NDCG cannot differentiate good and bad ranking systems when the data is of large size. This problem may be serious because huge dataset is common in applications such as clustering. Let f be a ranking function:

$$DCG_D(f, S_n) = \sum_{r=1}^{n} y_{(r)}^{f} D(r) \quad (5)$$

Let $y_1, \ldots, y_n$ $(y_i \in Y)$ be the degree of relevancy associated with $x_1, \ldots, x_n$ cluster data points, denote by $S_n = \{(x_1, y_1), \ldots, (x_n, y_n)\}$the set of data to rank:

$$IDCG_D(S_n) = \max_{f'} \sum_{r=1}^{n} y_{(r)}^{f'} D(r) \quad (6)$$

be the DCG value of the best ranking function on S_n.The NDCG of f on S_n with discount D is defined as:

$$NDCG_D(f, S_n) = \frac{DCG_D(f, S_n)}{DCG_D(S_n)} \quad (7)$$

We proposed a new multi view clustering framework which his based on ensemble clustering. In this framework, each data source is converted into two weighting schemesTF/IDF and TF/ICF. A clustering algorithm is then applied on these matrices to obtain different partitions of the data. Three different matrices are then calculated from these partitions as mentioned above. All the matrices are integrated and form a new similarity matrix, which is then used for the final clustering. Identify and outline 3 fundamental steps in the algorithm which as listed below:

- Feature re-Weighting Scheme
- Extraction of Similarity Matrices
- Combination/Aggregation of Similarity Matrices

3.6 Term Frequency/Inverse Document Frequency (TF/IDF)

The product of term frequency and inverse documents frequency is called Term Frequency/Inverse Documents Frequency (TF/IDF). The mathematical form of TF/IDF, where N is the total number of documents and n_j is the total number of documents containing the word j.

3.6.1.Term Frequency/Inverse Corpus Frequency (TF/ICF)

This weighing scheme [21] is a modified form of TF/IDF where the IDF factor is computed based on the inverse corpus frequency rather than the document frequency. Extraction of similarity matrices extract two different similarity matrices from the partitions and one from datasets because our proposed work based on both late and intermediate integration strategies which are given below:

- Cluster based similarity matrix
- Pair wise dissimilarity matrix
- NDCG
- KLD
- Affinity matrix

These matrices are derived from each of the views using both the term weighting schemes. The first two matrices are derived from partitioning of the data.

3.7 Combination/Aggregation Of Similarity Matrices

These matriceshave so far computed the individual similarity matrices based on different views of the dataset, and then combine these matrices to give a unified similarity matrix that incorporates information from the clustering results of the different views. This new similarity matrix will now be used to get the final consensus clustering of the dataset. The new similarity matrix, S, is calculated by aggregating all the similarity matrices mentioned above and is given by:

$$S = \frac{1}{5}(S_H + S_{PDM} + S_{AFF} + \mathrm{DKL}\left(\mathrm{P}\|\mathrm{Q}\right) + \mathrm{NDCG_D}(\mathrm{f}, \mathrm{S_n})) \quad (8)$$

IV RESULTS AND EVALUATION

In this section we1278nalyse the performance of our proposed algorithm. To this end, perform a series of tests in which cluster datasets represented by multiple views on document as well as movies datasets. Compare our results with those of other state-of the- art multi view clustering algorithms. To do this, we used the well known collection of Movies dataset from the popular movies database website imdb.com as well as the Cornell and Cora datasets from the Universities dataset. A third dataset relates to scholarly articles taken from the site Citeseer.com. We have chosen these datasets as they are easily available and present multiple views of a given data.

4.1 Datasets and Pre-Processing

The movies dataset consists of a collection of movies belonging to 17 categories. There are two views corresponding to this dataset. One is the set of movies represented by the actors which performed in those movies while the other is the same movies described by their keywords on the website. This dataset can be considered as a difficult task since there are 17 categories of unequal sizes. Citeseer dataset corresponds to a collection of 3312 different articles appearing on the website whose views correspond to the document by term and document by citation matrices respectively. The Cornell dataset is a subset of the 4 universities dataset whose views corresponds to the classical document by term matrix and the document by link matrix. This dataset has only 5 different categories. Similarly, the Cora dataset corresponds to an indicator matrix representing the presence or absence of a word within a document while a second view corresponds to the citation information regarding those documents. Summary of the different datasets used in this experimentation along with their statistics is given in Table 1.

Table 1: Details of the datasets

Datasets	Rows/ objects	Features		No of classes
Citeseer	3312	3703	4732	6
Cornell	195	1703	569	5
Cora	2708	1433	5429	6
Movies	617	1398	1878	17

4.2 Performance Evaluation Measures

To evaluate the results of proposed technique and compare it with existing techniques, we utilize an external validation criterion. Broadly have two kinds of validation measures—one that directly compute the measure such as the classification error, accuracy measure, Receiver Operating Characteristics (ROC) curve, etc. and another that indirectly computes a measure of goodness of the solution such as using entropy and mutual information. The ROC and Area under the ROC (AUC) are usually preferred in information retrieval where both precision and recall are important. In our case, where the objects may belong to multiple, hard class, the micro-averaged precision and recall converge to the same value. Chosen accuracy measure as a direct measure to estimate clustering results. Moreover chose another widely used indirect evaluation measure, the Normalized Mutual Information (NMI) as a second measure. The two measures are briefly described below:

4.2.1 Accuracy

It is defined as the percentage of number of correctly classified data to the total data. It is generally used to evaluate classification task using a confusion matrix. The formula for accuracy is defined as:

$$Accuracy = \frac{(a+b)}{(a+b+c+d)} \quad (9)$$

The accuracy measure can be used for any number of classes. The value of accuracy ranges between 0 (all elements are incorrectly classified) and 1 (all elements are correctly classified).

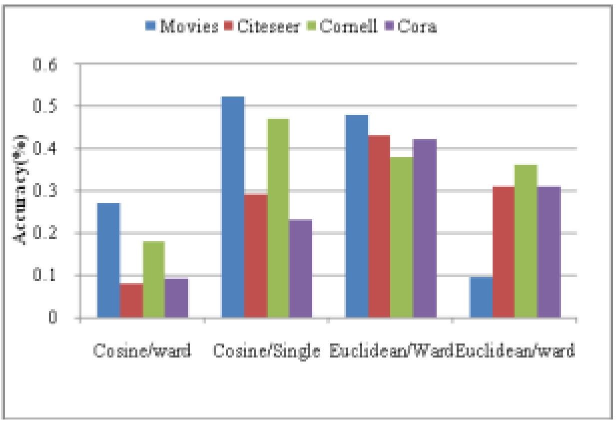

Fig. 2(a) The effect of similarity and linkage using accuracy

Normalized mutual information(NMI)

The NMI is used to measure the quality of a clustering result when the cluster size is small or uneven. This is done by taking the mutual information between two variables X and Y as a ratio of the geometric mean of their entropies. Let X is variable for cluster assignment as predicted by our clustering algorithm and Y is variable for true labels of cluster. Denote $k(X)$ as the number of clusters in X and n_i^X as the number of elements in the i^{th} cluster of X. Similarly, $k(Y)$ denotes the number of clusters in Y and n_i^Y denote the number of elements in the j^{th} cluster of Y. Also, n_{ij} denotes the number of elements classified in the i^{th} cluster by X and the j^{th} cluster by Y. The total number of elements is denoted by n. The formula for NMI is defined as:

$$NMI(X,Y) = \frac{\sum_{i=1}^{k(X)} \sum_{j=1}^{k(Y)} n_{ij} \log_{k(X)k(Y)} \left(\frac{n_{ij}.n}{n_i^X n_i^Y}\right)}{\sqrt{(\sum_{i=1}^{k(X)} n_i^X \log\left(\frac{n_i^X}{n}\right))(\sum_{i=1}^{k(Y)} n_i^Y \log\left(\frac{n_i^Y}{n}\right))}} \quad (10)$$

The value of NMI ranges from 0 to 1 which values closer to 0 representing poor quality clustering and those closer to 1 for high quality clustering.

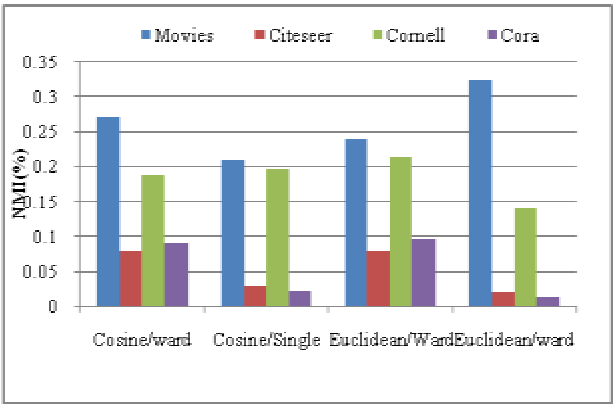

Fig. 2(b) The effect of similarity and linkage using NMI scores

The result of applying the two proximities measures on the various datasets is given in Fig. 2. The reported result is the average from the two views, V_1 and V_2. As can be clearly observed from Fig. 2a, the Cosine similarity gives a significantly better cluster as compared to the Euclidean distance on all the datasets tested. Similarly, using Ward's linkage gives a better average result than Single linkage, even when using Euclidean distance measure.Overall, the Cosine/Ward linkage combination outperforms all other strategies, which is consistent with previous findings in the literature for textual datasets. A similar pattern is observed when considering the NMI score in Fig.2b., thus re-enforcing the previous observations.

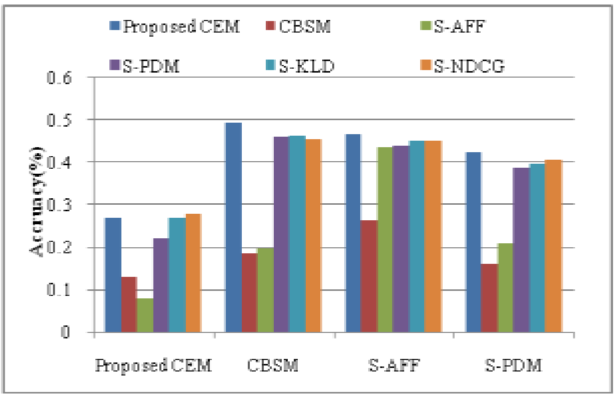

Fig.3 Comparison of proposed algorithm with individual matrices using accuracy

Considered these individual matrices as the resultant matrix and used hierarchical clustering using Ward's linkage to determine the cluster labels. The result is shown in a graphical form (Fig. 3). From these results, it is evident that firstly, the combined result of the matrices is always better than the individual results on all 4 tested datasets. Secondly, among the 5 matrices, the pair-wise similarity matrix (**S_{PDM}**) yields the better results and hence has the greatest contribution towards the combined matrix.

V CONCLUSION AND FUTURE WORK

In this paper, we presented a multi-view ensemble clustering approach for documents. This paper proposes a new ensemble based multi-view clustering algorithm that combines five types of similarity matrices such as the CBSM, the PDM, the AFF, the KLD and NDCG. The major principle of this work is to 1281nalyse the ensemble learning results of the various similarity matrixes are combined and formed as new combined matrix which is combination of all similarity matrix available –knowledge to give a better result. The proposed clustering ensemble framework incorporates both intermediate and late integration strategies for multiple views. Similarly, apply different term weighting methods such as TF/IDF and the TF/ICF to give to improve the clustering accuracy on various –datasets. The experimental evaluation on real-world datasets such Citeseer, Cornell, Cora and Movies for multiple views and demonstrates encouraging results that validate efficiently. Furthermore, the framework is straightforward and combines well-known different similarity matrix, even trivial algorithms to positive effect. Our future line of work will focus on two main areas. Firstly, the accuracy of our proposed multi-view clustering technique is directly related to the clustering algorithm used. Several algorithms have been proposed in the literature that uses more sophisticated algorithms, such as co-clustering, to improve the quality of the clustering. These algorithms can be incorporated into the proposed framework to increase the accuracy of the resultant clustering.Present methodology can be applied to a wide range of other practical problems including social network analysis and bioinformatics.

VI REFERENCES

[1] K. Jain, "Data clustering: 50 years beyond k-means", *Pattern Recognition Letters, vol. 31, no. 8, pp. 651–66, 2010.*

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

[3] A. Topchy, A. K. Jain, and W. Punch, "A mixture model of clustering ensembles", in *Proc. SIAM Intl. Conf. on Data Mining. Citeseer, 2004.*

[4] A. Goder and V. Filkov, "Consensus clustering algorithms: Comparison and refinement", in *ALENEX, vol. 8, 2008, pp. 109–117.*

[5] X. Wang, C. Yang, and J. Zhou, "Clustering aggregation by probability accumulation", Pattern Recognition, *vol. 42, no. 5, pp. 668–675, 2009.*

[6] S. Vega-Pons, J. Correa-Morris, and J. Ruiz-Shulcloper, "Weighted partition consensus via kernels", Pattern Recognition, *vol. 43, no. 8, pp. 2712–2724, 2010*

[7] S. Bickel and T. Scheffer, "Multi-view clustering", Proceedings of the 4th*IEEE International Conference on Data Mining, pp. 19-26, 2004*

[8] V.R. De Sa," Spectral clustering with two views", Proceedings of the 22th*IEEE International Conference on Machine Learning, pp. 20-27, 2005.*

[9] A. Kumar and H. Daume, "A co-training approach for multiview spectral clustering", Proceedings of the 28th*IEEE International Conference on Machine Learning, pp. 393-400, 2011*

[10] A. Kumar, P. Rai and H. Daume, "Co-regularized Multi- view Spectral Clustering", Proceedings of the 12th IEEE International Conference on Data Mining, pp. 675-684, 2012.

[11] G. Tzortzis and A. Likas, "Kernel-based Weighted Multi-view Clustering", Proceedings of the 12th IEEE International Conference on Data Mining, pp. 675-684, 2012.

[12] Frings, O., Alexeyenko, A., Sonnhammer, E.L. (2013), "MGclus: network clustering employing shared neighbours", Molecular BioSystems.

[13] Tang, L., Wang, X., Liu, H. (2010), "Community detection in multi-dimensional networks", *Technical Report, Defense Technical Information Center.*

[14] Janssens, F., Gl¨anzel,W., De Moor, B. (2007), "Dynamic hybrid clustering of bioinformatics by incorporating text mining and citation analysis", In Proceedings of the 13th *ACM SIGKDD international conference on knowledge discovery and data mining (pp. 360–369). New York.*

[15] Hong, Yi, Sam Kwong, Hanli Wang, and QingshengRen, "Resampling-based selective clustering ensembles", *Pattern recognition letters 30, no. 3 (2009): 298-305.*

[16] Azimi, J., & Fern, X. (2009, July), "Adaptive Cluster Ensemble Selection", *InIJCAI (Vol. 9, pp. 992-997)*

[17] A. Strehl and J. Ghosh, "Cluster Ensembles-A Knowledge Reuse Framework for Combining Multiple Partitions", Journal of Machine Learning Research, pp. 583- 617, 2002.

[18] Jia, Jianhua, Xuan Xiao, Bingxiang Liu, and Licheng Jiao, "Bagging-based spectral clustering ensemble selection", Pattern Recognition Letters 32, no. 10 (2011): 1456-1467.

[19] Strehl, A., Ghosh, J., Cardie, C. (2002), "Cluster ensembles—a knowledge reuse framework for combining multiple partitions*", Journal of Machine Learning Research, 3, 583–617*.

[20] Zheng, L., Li, T., Ding, C. (2010), " Hierarchical ensemble clustering", In Proceedings of the 2010 *IEEE international conference on data mining (pp. 1199–1204). Washington, DC.*

[21] Reed, J.W., Jiao, Y., Potok, T.E., Klump, B.A., Elmore, M.T., Hurson, A.R. (2006), "TF-ICF: a new term weighting scheme for clustering dynamic data streams", In Proceedings of the 5th *international conference on machine learning and applications (pp. 258–263). Washington, DC*.

[22] Johnson, D. H., &Sinanovic, S. *(2003). Symmetrizing the kullback-leibler distance.*

[23] Wang, Y., Wang, L., Li, Y., He, D., Chen, W., & Liu, T. Y. (2013, May), "A theoretical analysis of NDCG ranking measures", In Proceedings of the *26th Annual Conference on Learning Theory (COLT 2013)*.