International Journal of Engineering & Technology, *7 (4) (2018) 2710-2714*

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET
doi: 10.14419/ijet.v7i4.15229
Research paper

Framework for novel subspace clustering using search optimization methodology

Radhika K. R [1]*, Pushpa C. N [1], Thriveni J [1], Venugopal K. R [1]

[1] *Department of Computer Science and Engineering, University Visvesvaraya College of Engineering, Bangalore, India*
**Corresponding author E-mail:radhika@bmsit.in*

Abstract

Improving the yield as well as the perform of subspace clustering is one of the less-investigated topics in high-dimensional data. After reviewing existing approaches, it seriously felt that there is a need for classification of data points retrieved from a different number of subspace. The proposed study has presented a novel framework that targets to improve the accuracy of subspace clustering by addressing the problem associated with the exist of occlusion noise and dimensional complexity. An analytical approach as been proposed to design this framework with more emphasis on outlier minimization followed by obtaining optimal clusters. The technique also introduces a simple search optimization method, which is less iterative and is more productive for identifying the élite outcomes in each iterative step. The study outcome shows superior accuracy with a low rate of error when compared with the conventional approach.

Keywords: Accuracy;Elite outcomes; *High-dimensional Data; Optimal Cluster, Subspace clustering;*

1. Introduction

The emergence of high-dimensional data may observe in most of the trending domains that pose issues over techniques of data mining for throughput and effectiveness. As the reason of sparsity continuously increases in this data type, accommodating clusters is a demanding task [1]. The approaches of cluster ensembling are acquiring the massive amount of attention due to its usefulness in applications such as bioinformatics, data mining techniques and pattern identification.

In comparison with the conventional approaches towards clustering, the cluster ensembling method enables the integrate of many clustering solutions attained via various sources of data and joined into a unified solution to ultimately give a stabilized robust outcome [2]. The application point of view of high-dimensional data is not constrained to one single field, such as face images is a set of high-dimensional data as the pixel number is typically large and the image set for a given face lies about in a linear subspace of 9-dimensions [3]. Other application areas include the image segmentation and representation, disease detection, computer vision, unsupervised learning and motion segmentation [4].

Another example of high-dimensional data occurrence is in the scenario wherein the technology of DNA microarray produces enormous amounts of data involving probes of micrometer scale dimension. In the process of analyzing text files, the dimension number can be equated to the vector of word-frequency [5]. High-dimensional spaces have peculiar characteristics essential for clustering. Clustering is a tool to analyze the data aiming to bind data into multiple homogenous groups [6]. The primary task in data mining and data analysis of high-dimensional data is data clustering. It targets to uncover the structure, which is latently inherent in the data set and is applicable for domains such as image processing, bioinformatics, pattern recognition [7].

Most often, high-dimensional data reside in the low-dimensional structure rather than being uniformly distributed over the ambient space provision. The issue of data separation depending on subspaces, which are underlying and encounter multiple applications in the field of computer vision, image processing, temporal video segmentation and motion segmentation. As data is distributed arbitrarily in a subspace and there are no surrounding centroids, the methods of standard clustering take the benefit of spatial data proximity on individual clusters, which are not permissible in subspace clustering [8].

Subspace clustering is widely applicable for pattern identification and computer vision related applications.It is an extremely challenging role to know that how subspace structures of low-dimensional data exist in the high-dimensional data in the presence of complex noise.The statistical structures of complex noise are highly complicated and are not included in the group of Laplace or Gaussian noise. It is a technique used to perform the segmentation operation on the high-dimensional data that are taken up from many subspaces union.

Subspace clustering initiates the task of finding a subspace belonging to the low-dimensional class in which the data from the individual groups can simultaneously accommodate its subspace structure. The classifications of subspace clustering methods are algebraic methodologies, iterative methods, spectral clustering derived methods and statistical technique of clustering [9]. An extension to the conventional method of clustering is the subspace clustering technique. Apart from this, another reason due to which the struggle for the high-dimensional data continues is the dimensionality curse. In a dataset as the number of dimensions tends to increase, measuring distances, in this case, would be pointless. Hence, an algorithm that can satisfy the need of high-dimensional datasets with the increasing number of sets is required. This manuscript presents one such solution. Section 2 discusses the existing literature where different techniques are discussed for detection schemes used in power transmission lines followed by the discussion of research problems and proposed solution Section 2. Section 3 presents algorithm implementation followed by the discussion of result analysis in Section 4. Finally, the conclusive remarks are provided in Section 5

2. Background

The study of Yu et al. [10] introduced a framework named Adaptive Semi-Supervised Clustering Ensemble (A-RSEMICE) to support high dimensional clustering. Li and Vidal [11] proposed an optimization framework of Structured Sparse plus Structure Low-Rank (S^3LR) to cluster and complete the data withdrawn from the subspace of low-dimensional union. Kim et al. [12] proposed a representation for subspace as an elastic-net, a new kind of the scheme that would imply the use of singular values of elastic-net regularization.Tang et al. [13] used the search method of k Nearest Neighbors (k-NN) algorithms, being important regarding the implementation of machine learning and computer vision applications.

Khachatryan et al. [14] showed the significant improvement in the self-tuning technique by initializing the configuration. Further to enhance the robustness and accuracy factor in self-tuning the clusters of dense subspaces were proposed in data projections. Peng et al. [15] proposed an Inductive spectral clustering algorithm named inductive SSC (iSSC) which allows Sparse Subspace Clustering (SSC) to cluster data out. Jing et al. [16] studied the Dictionary Learning (DL) technique to recognize the low-dimensional subspace representation acquired from high-dimensional data.

Chen et al. [17] solved the issue of high-dimensional data clustering by the method of project clustering in data space having subsets of dimensions. Aldroubi and Sekmen [18] proposed an algorithm for representing high-dimensional data from a subgroup achieved from a low-dimension union. Hou et al. [19] studied the problem of clustering in data mining and solved it by the method of joint dimensionality and clustering. Muja and Lowe [20] proposed an improvised version of nearest neighbor algorithms. Elhamifar and Vidal [21] suggested and studied the algorithm named Sparse Subspace Clustering (SSC) to set up the data point clustering process

Feldman et al. [22] used the technique of projective clustering and k-means algorithms to handle high-dimensional data. Wang et al. [23] propose a novel algorithm for k-means approximation to decrease the computational integrity of the system. Dezeure et al. [24] reviewed the dimensional interferences techniques for estimating the general linear scheme. Dyer et al. [25] introduced a greedy method to recover the sparse signal called as Orthogonal Matching Pursuit (OMP). Nie et al. [26] proposed a new model that is capable of learning the similarity between the matrix and structure of cluster simultaneously. The study of Song et al. [27] involved the estimation of efficiency and effectiveness using the FAST algorithm. It incorporates the fast choice of features based on clustering. Tang et al. [28] addresses the issue of clustering in high-dimensional data. With an aid of an independent dataset for training, the following work has accomplished an efficient clustering mechanism.

The study has also considered a problem formulation of subspace clustering with a secondary target to lower the overhead while processing high-dimensional data. Hence, it can be seen that there are diverse direction of the work being carried out towards addressing the problems of subspace clustering. Each study has their own advantage point about the problem statement.

This trend of the existing research work exhibits that there is still a scope of significant improvement while performing subspace clustering on high dimensional data. the next section highlights the research problems that has been extracted after reviewing existing techniques.

a) Research Problem

The significant research problems are as follows:

- Existing studies toward subspace clustering doesn't emphasize on the dimensional factor of the data points that adversely affect the outcome with non-scalability.
- Non-inclusion of occlusion noise is another problem that has been identified in the existing system without which the robustness is quite challenging to decide.
- Outliers and their possible effect in the subspace clustering is a significant problem in high dimensional data that will pose the challenge of false positives.
- Uses of the computational cost-effective technique are quite a few to find in the existing system where accuracy is obtained at the cost of computational complexity.

Therefore, the problem statement of the proposed study can be stated as "Designing a cost-effective methodology for performing subspace clustering in the presence of critical issues and error-prone environment poses to degrade data quality in high-dimensional data."

b) Proposed Solution

The proposed system is in continuation of our prior research work [29] [30]. The present research work emphasizes on improving the accuracy factor while performing subspace clustering for high-dimensional data. The proposed system applies the analytical methodology to implement this mechanism.

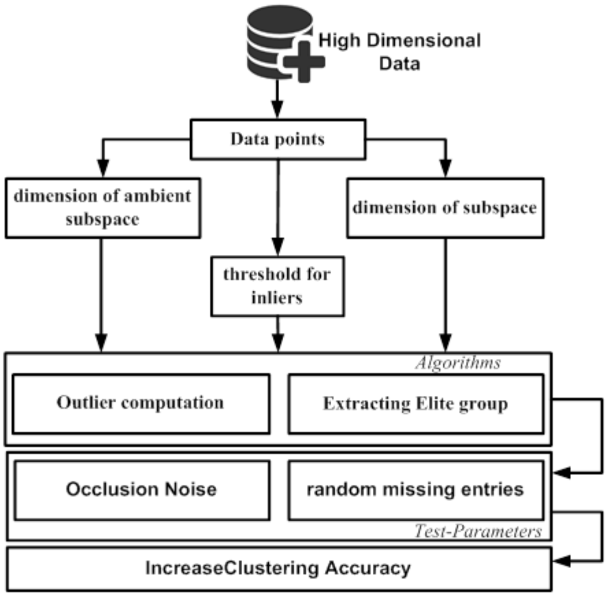


Fig. 1: Proposed Subspace Clustering Methodology.

The architecture presented in Figure. 1 follows the top-down approach where the discrete data points are extracted from given high dimensional data. This process leads to consideration of three sub-process i.e.dimensions for ambient subspace, the threshold for inliers, and dimensions for sub-spaces to be subjected to proposed algorithm. The proposed system constructs an algorithm that performs the computation of the outliers to ensure that more accuracy in the clustering process is ensured at the end. It is also followed by using a search optimization technique for obtaining best clusters that are required to offer better outcomes while performing clustering. The proposed solution also considers investigating the framework in the presence of occlusion noise to analyze its effectiveness. Further enhancement of the proposed system is still carried out by ensuring that the algorithm doesn't have any form of missed entries during the clustering process for the given high dimensional data.

3. Algorithm implementation

The complete design principle of the proposed algorithm depends on how accurate it can compute the outliers when it attempts performing the subspace clustering. At the same time, it also focuses on obtaining precise information about the superior clusters while performing subspace clustering. This section briefs the algorithm. The algorithm also deals in introducing a learning-based approach that further enriches the characteristics of the superior clusters.

Algorithm for Estimating Outliers

The prime role of this algorithm is to precisely estimate the number of outliers being present in clusters of high dimensional data. This operation mainly intends to enrich the quality of the information being retained in clusters. The input to the algorithms is η (subspace cardinality), δ (dimension of subspace to be estimated), α (Data points), and c (threshold) that leads to the generation of Ω (estimated outlier) as an output.

Table 1: Algorithm for Estimating Outliers

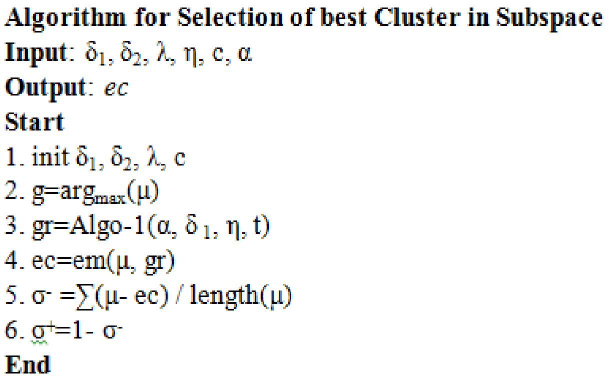
```
Input: α, δ, η, c
Output: Ω
Start
    Init α, δ, η, c
    For (i=1: η)
    If (size(Ω)<4)
    For (j=1:i-1)
    H=I(α)-β2
    d=∑ (H.α)2
    End
    dmin =argmin(d)[v, o] → sort(dmin)
    Ω → order(1:4)
    End
End
```

The initial step of the above algorithm performs computation of the sizes of the matrix that stores the significant data points and the technique constructs a matrix of size K x M. For a given set of real-time data, there is a possibility of infinite number of noises as well as outliers, which is quite computationally challenging to measure. Therefore, for the ease of computing, we consider the presence of outlier attribute within the range of [1 N], where N is a positive integer. The process also constructs a three dimensional matrix β where each dimensions of this matrix corresponds with δ, K, and some random test value. After this multi-dimensional matrix is formed, the algorithm proceeds to perform a check for presence of outliers. The best way to perform this checking for cost-effective computational operation is to use a certain limiting factor n.This operation is implemented by comparing the empirical value of outliers with specific limiting factor. We have considered the limiting factor as 4 hypothetically that can be altered. The algorithm then performs an iterative operation to assess the presence of significant sub-spaces bearing potential information. This operation assists in estimating the matrix H that offers more insight to the trend of the data in three dimensional matrix β.

The computation of the new matrix H is carried out by multiplying I with β_2 where β_2 represents dot product of β and β^{-1}. The matrix I represents an identity matrix. This operation leads to disclosure of significant information within the high-dimensional data thereby narrowing the search process towards the specific amount of outliers that could be present in a scattered formed in the entire clusters. To further optimize the search process computationally, the proposed system makes use of priorly computed matrix H along with the data points for better exploration towards outliers. Finally, the algorithm also computes the distance factor that is involved in every iterative steps of the outlier exploration process in high dimensional data using the matrix d_{min}. Further, the algorithm performs sorting operation of the d_{min} obtained. In the final step of algorithm implementation, the proposed system performs computation of the outliers that is retained in a new matrix Ω. The significant contribution of this algorithm is that, it performs dual task in parallel. The first task is primarily related to outlier computation while the secondary task is to minimize the entropy error to highest feasible extent. This occurrence of dual-task processing results in cost-effective algorithm implementation with an optimal enhancement to the accuracy of extracting the outlier information optimizing the process of subspace clustering.

Algorithm for Selection of best Cluster in Subspace

The successful implementation of the prior algorithm results obtaining precise information of outliers that assists to identify the quality information of clusters while performing subspace clustering. This operation results in large number of potential clusters where it is essential to select the best out of them for computational efficiency. This algorithm is mainly a continuation of the prior algorithm where the present focus is mainly emphasized on selection of the best cluster. The algorithm is executed by considering feed of δ_1 (dimension of subspace), δ_2 (dimension of ambient subspace), λ (number of point in each group), η (number of clusters), c (threshold for inliers), and α (data point) that after processing results in ec (elite cluster). The significant steps of the algorithm are shown in the Table 2.



```
Algorithm for Selection of best Cluster in Subspace
Input: δ1, δ2, λ, η, c, α
Output: ec
Start
    1. init δ1, δ2, λ, c
    2. g=argmax(μ)
    3. gr=Algo-1(α, δ1, η, t)
    4. ec=em(μ, gr)
    5. σ- =∑(μ- ec) / length(μ)
    6. σ+=1- σ-
End
```

As this algorithm is more inclined towards exploring the best cluster, consideration of the dimensional attributes plays a critical role. The execution of this algorithm begins by initialization of the subspace dimensions along with dimensional attribute relating to ambient space. The initialization steps also consider data points and threshold for inliers. The proposed system constructs a label matrix μ that is utilized for organizing the data points where the process performs transformation operation to obtain a single variable from multiple variables. This phenomenon is strategically formed in order to narrow down the search optimization process on subspace clustering with challenges existing on high dimensional data. The proposed algorithm considers that the empirical value of label μ is nearly equivalent to cumulative clusters obtained from the prior implementation of algorithm. For better study formulation, the value of threshold for inliers is considered to be residing with probability limit of [0-1]. Once the computation of the outlier is over in the first algorithm, the obtained value of the outlier is subjected to present an algorithm for extracting the information associated with the subspace parameter η. The algorithm further performs processing in order to extract the value of an efficient map of cluster i.e. em. For obtaining enhanced order of subspaces, the proposed system uses a method em that performs significant permutation of the label matrix μ. A simplified process of obtaining matrix using squared cost-based approach is performed in order to construct the method more effectively.

Further, the system computes the cost involved in both assignment operation so that overall computational cost involved in the process can be significantly minimized considering all the probability of the data points that are existing within the obtained subspaces. This operation ultimately results in highly enhanced form of the subspaces that further upgrades the information retained in high dimensional data. Although, the clustering operation is carried out in effective rate of computation, there are also chances that certain extracted data of cluster may be skipped owing to random iterative process in different domains of high dimensional data. This problem is addressed by estimating the possible missing rate during subspace clustering and inclusion of this step would rather incorporate more accuracy of obtaining better clusters. The label matrix μ is utilized for this reason and these results in estimation of missing rate i.e. σ^- obtained during subspace clustering process in iterative manner. This process finally performs estimation of accuracy σ^+ that is required to infer the quality of the obtained clusters at the end.

The distinct contribution of the proposed algorithm is its utilization of search optimization technique that narrow downs the search process during subspace clustering in order to finally extract the ultimate clusters from the complex high dimensional data.

Another unique property of the proposed algorithm is its faster computation time irrespective of the increasing data points present in high-dimensional data. This results in minimization of the entropy error from every generated cluster during the extraction of the best cluster process and hence results in superior version of the cluster. The final contribution of the proposed algorithm is that complete process of extracting best cluster doesn't have any form of memory dependencies as proposed optimization stages performs disposal of the memory once iteration step is incremented. Hence, using same memory size, the computational efficiency can be obtained in the process of subspace clustering.

4. Performance

This section outlines the outcomes obtained after implementing the algorithms discussed in prior section. As the proposed system focuses on subspace clustering of high-dimensional data; signal-based data repository for testifying the algorithm is considered. The analysis is carried out considering the image database which is characterized by 16,128 significant images with multiple complex states of illumination. The performance parameter considered to assess the effectiveness of the proposed system is clustering accuracy and error rate, where the results are captured for different states of occlusion. The study outcome of the proposed system is compared with the most related work presented by Song et al. [31] who have performed Sparse Subspace Clustering (SSC) using Hilbert space. The study outcome of proposed system has been also compared with out prior model RMSC [32]. The performance Analysis of the existing algorithms RMSC, SSC and proposed algorithm SBCS is shown in Fig. 2 & 3.

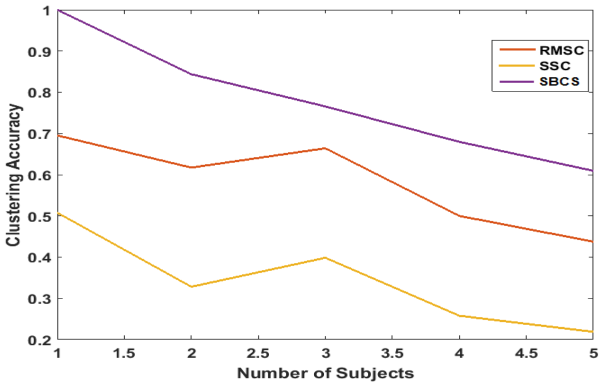


Fig. 2: Clustering Accuracy in the Absence of Occlusion.

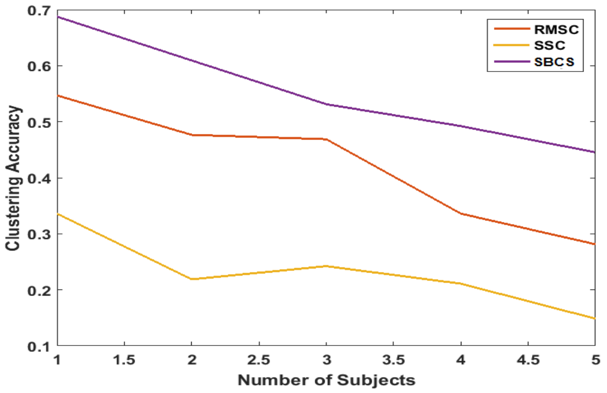


Fig. 3: Clustering Accuracy in the Presence of Occlusion.

Fig. 2 and Fig. 3 highlight the performance of clustering accuracy concerning occlusion. Although the trend of clustering accuracy diminishes with increase of subjects during the analysis, the outcome shows that proposed system offers comparatively more clustering accuracy as compared to the approach of Song et al. [31]. An interesting observation, in this case, is that irrespective of any proportion of occlusion level proposed system offers better sustainability against the noise level too. The approach presented by Song et al. [31] uses multiple steps of classification with an assistance of joint coding approach. This mechanism increases accuracy only in the absence of occlusion; however, when the occlusion is increased slowly, there is sharp fall of accuracy observed in Song et al. [31] work. On the other hand, the proposed system has focused on outlier minimization in the preliminary stage itself leading to increased accuracy.

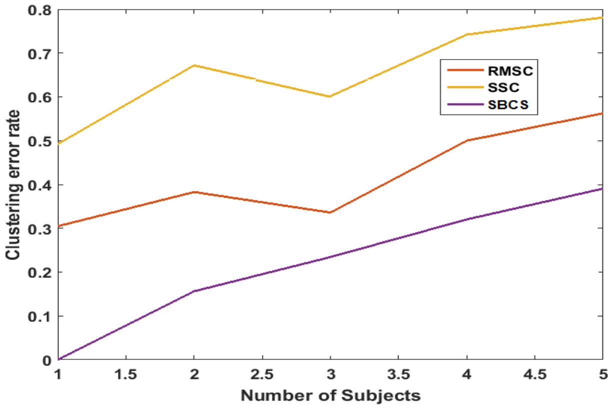


Fig. 4: Analysis of Error Rate in the Absence of Occlusion.

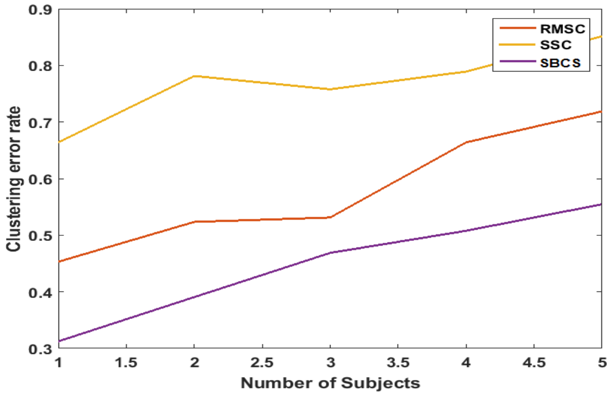


Fig. 5: Analysis of Error Rate in the Presence of Occlusion.

Fig. 4 and Fig. 5 present analysis of the error rate which shows the consistency of the proposed system towards error rate performance to be lower than the existing system of Song et al. [31]. The approach of Song et al. [32] have presented unsupervised classification approach with prime intension is to extract the hidden structure of cluster. However, it doesn't emphasize on accuracy parameters much whereas the proposed system mainly presents a set of an algorithm which is explicitly capable of addressing the non-Gaussian noise along with more emphasis on identification of outliers followed by computation of inliers. This process results in an efficient error recovery methodology which is highly suitable for high-dimensional data.

Table 1: Summary of Time Complexity Analysis

	Processing Time
SSC	0.8766
RMSC	0.991
SBSC	0.115

It can be seen that proposed system offers approximately 76% and 87%of improvement in processing time in comparison to existing RMSC and SSC. Hence, the proposed system of SBSC takes approximately 85% comparatively lower algorithm processing time to offer a faster response in contrast to the existing technique of subspace clustering.

5. Conclusion

Subspace clustering is one of the essential operations that enhance the quality of the information present in high-dimensional data.

However, the massiveness and vulnerability of such data towards errors and noise are so high that the result of applying analytical operation significantly fails. Therefore, subspace clustering offers better solution towards this process. We reviewed the existing system to find that there is still a large scope for improvement as problems associated with subspace clustering is not yet solved. Therefore, the proposed system offers a discussion of a novel framework with core intention of enhancing the operation of subspace clustering. The outcomes tested on multiple scenarios of noise and occlusion shows that proposed system yield better outcome of accuracy in comparison to exiting subspace-clustering methodology.

References

[1] Tomasev, Nenad, et al. "The role of hubness in clustering high-dimensional data." IEEE Transactions on Knowledge and Data Engineering 26.3 (2014): 739-751. https://doi.org/10.1109/TKDE.2013.25.

[2] Yu, Zhiwen, et al. "Incremental semi-supervised clustering ensemble for high dimensional data clustering." IEEE Transactions on Knowledge and Data Engineering 28.3 (2016):701-714. https://doi.org/10.1109/TKDE.2015.2499200.

[3] Heckel, Reinhard, and Helmut Bölcskei. "Robust subspace clustering via thresholding." IEEE Transactions on Information Theory 61.11 (2015): 6320-6342. https://doi.org/10.1109/TIT.2015.2472520.

[4] Wu, Tong, and Waheed U. Bajwa. "Learning the nonlinear geometry of high-dimensional data: Models and algorithms." IEEE transactions on signal processing 63.23 (2015): 6229-6244. https://doi.org/10.1109/TSP.2015.2469637.

[5] Yuan, Xiaoru, et al. "Dimension projection matrix/tree: Interactive subspace visual exploration and analysis of high dimensional data." IEEE Transactions on Visualization and Computer Graphics 19.12 (2013): 2625-2633. https://doi.org/10.1109/TVCG.2013.150.

[6] Bouveyron, Charles, and Camille Brunet-Saumard. "Model-based clustering of high-dimensional data: A review." Computational Statistics & Data Analysis 71 (2014): 52-78. https://doi.org/10.1016/j.csda.2012.12.008.

[7] Tian, Jinyu, et al. "Learning the Distribution Preserving Semantic Subspace for Clustering." IEEE Transactions on Image Processing 26.12 (2017): 5950-5965. https://doi.org/10.1109/TIP.2017.2748885.

[8] Elhamifar, Ehsan, and Rene Vidal. "Sparse subspace clustering: Algorithm, theory, and applications." IEEE transactions on pattern analysis and machine intelligence 35.11 (2013): 2765-2781. https://doi.org/10.1109/TPAMI.2013.57.

[9] He, Ran, et al. "Robust subspace clustering with complex noise." IEEE Transactions on Image Processing 24.11 (2015): 4001-4013. https://doi.org/10.1109/TIP.2015.2456504.

[10] Yu, Zhiwen, et al. "Adaptive ensembling of semi-supervised clustering solutions." IEEE Transactions on Knowledge and Data Engineering (2017). https://doi.org/10.1109/TKDE.2017.2695615.

[11] Li, Chun-Guang, and René Vidal. "A Structured Sparse Plus Structured Low-Rank Framework for Subspace Clustering and Completion." IEEE Trans. Signal Processing 64.24 (2016): 6557-6570. https://doi.org/10.1109/TSP.2016.2613070.

[12] Kim, Eunwoo, Minsik Lee, and Songhwai Oh. "Robust Elastic-Net Subspace Representation." IEEE Transactions on Image Processing 25.9 (2016): 4245-4259.

[13] Tang, Xiaoxin, et al. "Scalable multicore k-nn search via subspace clustering for filtering." IEEE Transactions on Parallel and Distributed Systems 26.12 (2015): 3449-3460. https://doi.org/10.1109/TPDS.2014.2372755.

[14] Khachatryan, Andranik, et al. "Improving accuracy and robustness of self-tuning histograms by subspace clustering." IEEE Transactions on Knowledge and Data Engineering 27.9 (2015): 2377-2389. https://doi.org/10.1109/TKDE.2015.2416725.

[15] Peng, Xi, Lei Zhang, and Zhang Yi. "Inductive sparse subspace clustering." Electronics Letters 49.19 (2013): 1222-1224. https://doi.org/10.1049/el.2013.1789.

[16] Jing, Liping, Michael K. Ng, and Tieyong Zeng. "Dictionary learning-based subspace structure identification in spectral clustering." IEEE transactions on neural networks and learning systems 24.8 (2013): 1188-1199. https://doi.org/10.1109/TNNLS.2013.2253123.

[17] Chen, Lifei, Qingshan Jiang, and Shengrui Wang. "Model-based method for projective clustering." IEEE Transactions on Knowledge and Data Engineering 24.7 (2012): 1291-1305. https://doi.org/10.1109/TKDE.2010.256.

[18] Aldroubi, Akram, and Ali Sekmen. "Nearness to local subspace algorithm for subspace and motion segmentation." IEEE Signal Processing Letters 19.10 (2012): 704-707. https://doi.org/10.1109/LSP.2012.2214211.

[19] Hou, Chenping, et al. "Discriminative embedded clustering: A framework for grouping high-dimensional data." IEEE transactions on neural networks and learning systems 26.6 (2015): 1287-1299. https://doi.org/10.1109/TNNLS.2014.2337335.

[20] Muja, Marius, and David G. Lowe. "Scalable nearest neighbor algorithms for high dimensional data." IEEE Transactions on Pattern Analysis and Machine Intelligence 36.11 (2014): 2227-2240. https://doi.org/10.1109/TPAMI.2014.2321376.

[21] Elhamifar, Ehsan, and Rene Vidal. "Sparse subspace clustering: Algorithm, theory, and applications." IEEE transactions on pattern analysis and machine intelligence 35.11 (2013): 2765-2781. https://doi.org/10.1109/TPAMI.2013.57.

[22] Feldman, Dan, Melanie Schmidt, and Christian Sohler. "Turning big data into tiny data: Constant-size coresets for k-means, PCA, and projective clustering." Proceedings of the twenty-fourth annual ACM-SIAM symposium on discrete algorithms. Society for Industrial and Applied Mathematics, 2013. https://doi.org/10.1137/1.9781611973105.103.

[23] Wang, Jingdong, et al. "Fast approximate k-means via cluster closures." Multimedia Data Mining and Analytics. Springer International Publishing, 2015. 373-395.

[24] Dezeure, Ruben, et al. "High-Dimensional Inference: Confidence Intervals, $ p $-Values and R-Software hdi." Statistical science 30.4 (2015): 533-558. https://doi.org/10.1214/15-STS527.

[25] Dyer, Eva L., Aswin C. Sankaranarayanan, and Richard G. Baraniuk. "Greedy feature selection for subspace clustering." The Journal of Machine Learning Research 14.1 (2013): 2487-2517.

[26] Nie, Feiping, Xiaoqian Wang, and Heng Huang. "Clustering and projected clustering with adaptive neighbors." Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2014.

[27] Song, Qinbao, Jingjie Ni, and Guangtao Wang. "A fast clustering-based feature subset selection algorithm for high-dimensional data." IEEE transactions on knowledge and data engineering 25.1 (2013): 1-14. https://doi.org/10.1109/TKDE.2011.181.

[28] Tang, Hao, et al. "Partially supervised speaker clustering." IEEE transactions on pattern analysis and machine intelligence 34.5 (2012): 959-971. https://doi.org/10.1109/TPAMI.2011.174.

[29] Radhika K R, Pushpa C N, Thriveni J, Venugopal K R, "Insights to Existing Techniques of Subspace Clustering in High-Dimensional Data," International Journal of Scientific and Engineering Research, 2016.

[30] Radhika, K.R., and Pushpa, C.N. and Thriveni, J. and Venugopal, K.R. (2016) EDSC: Efficient document subspace clustering technique for high-dimensional data. In: 2016 International Conference on Computational Techniques in Information and Communication Technologies (ICCTICT), 11-13 March 2016, Bangalore.

[31] H. Song, W. Yang, N. Zhong and X. Xu, "Unsupervised Classification of PolSAR Imagery via Kernel Sparse Subspace Clustering," in IEEE Geoscience and Remote Sensing Letters, vol. 13, no. 10, pp. 1487-1491, Oct. 2016. https://doi.org/10.1109/LGRS.2016.2593098.

[32] K. R. Radhika, C. N. Pushpa, J. Thriveni, K. R. Venugopal, "RMSC: Robust modeling of subspace clustering for high dimensional data", International Conference on Advances in Computing, Communications and Informatics (ICACCI), pp.1535-1539, 2017.