### CHAPTER 7

**SUMMARY, CONCLUSION, AND RECOMMENDATIONS**

<table>
<thead>
<tr>
<th>S.No.</th>
<th>Name of the Sub-Title</th>
<th>Page No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.1</td>
<td>Summary and Conclusion</td>
<td>104-105</td>
</tr>
<tr>
<td>7.2</td>
<td>Recommendations</td>
<td>105-107</td>
</tr>
<tr>
<td></td>
<td>7.2.1 Applications of the Thesis</td>
<td>105-106</td>
</tr>
<tr>
<td></td>
<td>7.2.2 Future Scope</td>
<td>106-107</td>
</tr>
</tbody>
</table>
CHAPTER 7

SUMMARY, CONCLUSION, AND RECOMMENDATIONS

Summary, conclusion, and recommendations of the thesis are presented in this chapter.

7.1 SUMMARY AND CONCLUSION

The effectiveness of clustering methods depends on the number of clusters. For this reason, several methods are investigated for automatic clusters detection. Among these methods, the Visual Access Tendency (VAT) is an optimal choice for detecting the clustering tendency (or number of clusters). This thesis presents the enhanced data visualization methods for effective assessment of clustering tendency for unlabeled datasets. The traditional clustering methods i.e., k-means and MST-based-clustering cannot detect the prior clustering tendency. Therefore, the VAT is used in both k-means and MST-based clustering methods for determining the clustering results along with prior clustering tendency. These proposed approaches are known as VAT-based-k-means, and VAT-based-MST-clustering methods. It is also possible to find the clustering tendency from sample data instead of original dataset. This is the motivation for enhancing the VAT by progressive random sampling technique (PS) and it is known as PSVAT. Other hybrid approaches are proposed using PSVAT and these are known as PSVAT-based-k-means clustering and PSVAT-based-MST-clustering methods. These hybrid methods are faster than VAT based clustering methods. However, the hybrid methods are expensive. Therefore, the present work extends the VAT for determining the explicit clustering results instead of using hybrid approaches. This proposed clustering method is known as a Visualized Clustering Approach (VCA). The VAT uses the Euclidean based dissimilarity matrix for detection of clusters. The cosine distance metric is more robust. By using cosine distance metric, the thesis presents other enhanced data visualization methods (enhanced VAT methods) namely cVAT, and cSpecVAT. The cVAT and cSpecVAT performs the assessment of clustering tendency effectively for synthetic, real,
image, and gene datasets. However, the thesis presents another data visualization method for speech datasets, which is GMMVS-VAT. It uses the two concepts: one is Gaussian Mixture Model (GMM), for designing the model of speech data, and the other is multi-viewpoints based cosine similarity (MVS-cosine metric) for finding the best dissimilarity matrix. The effectiveness of MVS is more, since it uses more than one viewpoint in similarity (or dissimilarity) features computation. The different VCA methods are established using VAT, cVAT, cSpecVAT, and GMMVS-VAT for performing both assessment of clustering tendency and discovering the clustering results. The proposed data visualization clustering methods are known as VCA-VAT, VCA-cVAT, VCA-cSpecVAT, and VCA-GMMVS-VAT. Various performance measures such as goodness by OTSU, clustering accuracy, and Normalized Mutual Information are used in the experimental work for evaluating the effectiveness of proposed VCA methods. In GMMVS-VAT, the MVS demands high-computational cost for high-dimensional space of GMM mean supervectors. Hence, the thesis uses the linear subspace learning (LSL) techniques for transforming of high-dimensional space of GMM mean supervectors into a low-dimensional manifold. Therefore, the LSL-based VCA methods are more efficient than VCA-GMMVS-VAT for speech datasets. The LSL based VCA methods, VCA-GMM-VAT-PCA (GMVP), VCA-GMM-VAT-LPP (GMLP), VCA-GMMVS-VAT-NPE (GMPE), and VCA-GMMVS-VAT-LDA (GMVLDA) are proposed for addressing the the dimensionality reduction problem in speech clustering.

7.2 RECOMMENDATIONS

Recommended applications and future scope of the thesis are presented as follows.

7.2.1 Applications of the Thesis

The thesis presented various visualized clustering approaches for assessment of clustering tendency and it works efficiently in many clustering related applications. The proposed techniques can be applied in many real time applications. The assessment of
clustering tendency in a visual approach provides excellent readability of clusters from user perspective and achieving prominent clustering results. Assessing the clustering tendency is an essential step in any clustering method.

The recommended applications of the proposed work are outlined as follows:

1. To discover the hidden structures of clusters for complex datasets
2. To determine the number of clusters for unlabeled data
3. To discover the proper data partitioning results for various real-time datasets
4. To learn the clustering results from proposed methods
5. To use the proposed methods in image clustering [91] (or segmentation) for detecting the number of different objects of an image.
6. The Visualized Clustering Approach (VCA) prominently groups the similar objects or characters, so it can be used as an effective clustering tool for object or character recognition.
7. The VCA-cVAT, VCA-cSpecVAT, and VCA-GMMVSVAT are efficient clustering methods and these methods can be applied for a broad range of real-life applications.
8. The proposed VCA is an effective clustering method and it is also prominently used for text mining.
9. A data visualization method poses the well-visualized analytics for end users in real-life applications.

**7.2.2 Future Scope**

The present thesis dedicated on the development of visualized clustering approaches for variety of datasets (synthetic, real, face image, gene, and speech datasets) other than the video data also. Hence, the present work can be experimented on dynamic
video data. Further, there is a scope to address many challenging issues such as scalability, dimensionality problem, and representation of the dynamic data in video clustering.