D. Hierarchical Clustering

Hierarchical clustering is a very useful method of clustering. The output is best visualized using a dendogram. The dendogram is a tree that illustrates the hierarchy of clusters. The output for clusters of contents can be seen in Figure 6. The same output for clusters of countries can be seen in Figure 7. By comparing Figure 5 and 7, it can be seen that there exists some geographic pattern in popularity of contents.

IV. Discussion and Future Work

Clustering is one of the initial challenges of any machine learning problem. With the advent of big data, performing unsupervised cluster analysis is becoming difficult. This analysis serves as an example of how to find geographic variation in popularity using unsupervised machine learning. The analysis could serve applications in recommendation systems, search engines and models for many other systems that require an insight on geographic variation in popularity of contents. Further statistical models can be built by making use of the hierarchical structure in the data, which will be able to predict popularity of contents in different geographic regions.

V. Conclusion

By applying unsupervised machine learning techniques and comparing the output, it is observed that there exists some variation in popularity of contents across Europe. The analysis clearly defines sequence of methods to find patterns of geographic variation in popularity of contents using unsupervised machine learning techniques. The clusters of similar objects are displayed is Figures 4, 5, 6 and 7.

VI. Acknowledgement

I am thankful for the guidance and support of my mentor Dr. P. S. Rao, Assistant Professor, Department of Applied Mathematics.

References

[1] (2017) Information courtesy of box office mojo. [Online]. Available: http://www.boxofficemojo.com

[2] Y. Sun, "Conversational recommendation system with unsupervised learning," *ACM Digital Library*, 2005.

[3] T. Liu, "Clustering billions of images with large scale nearest neighbor search," *IEEE Workshop on Applications of Computer Vision*, 2007.

[4] X. Y. Wang and J. M. Garibaldi, "A comparison of fuzzy and nonfuzzy clustering techniques in cancer diagnosis," *The University of Nottingham*, 2006.

[5] M. C. Pham, "A clustering approach for collaborative filtering recommendation using social network analysis," *Journal of Universal Computer Science, vol. 17, no. 4*, 2011.

[6] M. Dash, "Dimensionality reduction," *Nanyang Technological University, Singapore*, 2006.

[7] I. Jollife, *Principal Component Analysis*, 2nd ed. Springer, 2002.

[8] X. H. Chris Ding, "K-means clustering via principal component analysis," 2004.

[9] I. Naim and D. Gildea, "Convergence of the em algorithm for gaussian mixtures with unbalanced mixing coefficients," *Department of Computer Science, University of Rochester Rochester, NY 14627, USA*, 2012.

[10] C. M. Bishop, *Pattern Recognition and Machine Learning*, 1st ed. Springer, 2006.

[11] C. Fraley and A. E. Raftery, "Model-based clustering, discriminant analysis and density estimation," *Journal of the American Statistical Association 97:611-631*, 2002.

[12] G. H. Laurens van der Maaten, "Visualizing data using t-sne," *Journal of Machine Learning Research 9 (2008) 2579-2605*, 2008.

[13] G. Schwarz, "Estimating the dimension of a model," *The Annals of Statistics, Vol. 6, No. 2.*, 1978.