Unsupervised and Self-adaptive Algorithm for Particle Size Distribution Clustering

Hossein Izadi[1], Vahidoddin Fattahpour[2], Morteza Roostaei[2], Mahdi Mahmoudi[2], Noel Devere-Bennett[3]
1: University of Tehran, 2: RGL Reservoir Management, 3: Nexen CNOOC

Summary

Particle size distributions (PSDs) plays an important role in designing sand control screens. Using different techniques (Dry Sieving, LPSA, and Dynamic Image Analysis (DIA)), large number of PSDs could be measured for core samples in a certain project. Moreover, large-scale sand retention tests are becoming common practice in recent years. These tests usually use duplicated sand mixtures of representative PSDs. Therefore, clustering the PSD data is essential for sand control design and sand retention tests.
Supervised and unsupervised machine learning algorithms are getting more attention in computational petroleum engineering. Usually there is no clear idea that how many clusters are supposed to be detected in each PSD database. Therefore, due to the limitation for setting the number of clusters, PSD clustering could not be accomplished using conventional clustering algorithms such as k-means (Likas et al., 2003) or artificial neural networks (Hagan et al., 1996). As a new approach, PSD clustering based on an incremental clustering algorithm is used here. The proposed algorithm has online incremental learning capability and it is based on adaptive resonance theory (ART) (Sadri et al., 2006). Besides, the number of clusters is not needed to be assigned as an input parameter in the algorithm. The algorithm, based on a self-adaptation approach, tries to minimize the number of clusters. Accordingly, it is appropriate for PSD clustering of big databases. The proposed algorithm can be used in industrial applications such as sand control design and sand control evaluation testing.

Theory / Method / Workflow

The steps of the proposed clustering algorithm are presented hereunder (Sadri et al., 2006):

1. Adjust the minimum similarity threshold (δ in the range of 0 and 1).
2. Create cluster center list and consider the first PSD as the center of the first cluster.
3. Read the next PSD.
4. Find the center of clusters with a similarity greater than δ to the inputted PSD.
   If found: assign the PSD to those clusters, and re-adjust the cluster center, and combine those clusters which are overlapped (have at least one common PSD).
   If not found: create a new cluster and insert the inputted PSD to cluster center list as a new cluster center.
5. Repeat steps in 3-4 for all the PSDs.

Results, Observations, Conclusions

We used 475 PSD samples from the McMurray Formation derived from core in Nexen CNOOC's Long Lake area. We represent the results with value of δ=0.90 here; however, the algorithm could find the optimized value of δ based on provided conditions for the optimized clustering.

The results of the clustering for top four clusters is shown in Fig. 1. These clusters cover about 95% of all PSDs. The center of each cluster is depicted in Fig. 1f.

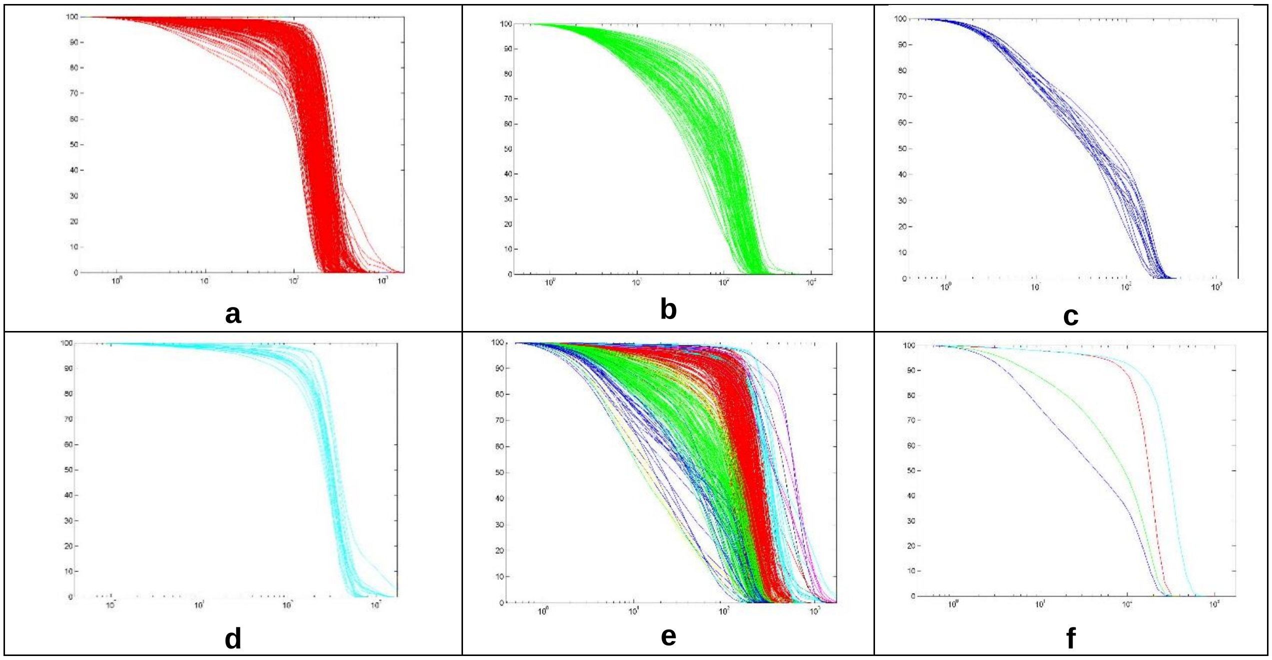

Fig. 1. The top four clusters (a to d) which includes 95% of the data base, (e) the whole data base, and (f) the center of the top four clusters. The x and y axis represent the size (µm) and cumulative passed percentage, respectively.

Novel/Additive Information

The method is an online algorithm, therefore the algorithm does not need to cluster previous PSDs, if new PSD data are introduced. Moreover, the algorithm does not need the number of clusters as an input data, as it minimizes the number of clusters using a self-adaptation approach. Additionally, unlike common practice in sand control design, the algorithm considers the whole PSD curve rather than single points (D-values) on each curve. Consequently, the proposed approach is an efficient and reliable algorithm for PSD clustering, especially for big databases.

Acknowledgements

The authors would like to acknowledge RGL Reservoir Management Inc. and Nexen CNOOC for the permission to publish this paper, and the financial support provided by NRC through their IRAP program, Natural Sciences and Engineering Research Council of Canada (NSERC) and Canada Foundation for Innovation

References

Hagan, M.T., Demuth, H.B., Beale, M.H., 1996. Neural Network Design. Pws Pub, Boston.

Likas, A., Vlassis, N., & Verbeek, J. J. (2003). The global k-means clustering algorithm. Pattern recognition, 36(2), 451-461.

Sadri, J., Ching. Y.S., Bui.T.D. (2006) A New Clustering Method for Improving Plasticity and Stability in Handwritten Character Recognition Systems, 18th International Conference on Pattern Recognition.