Abstract

One of the most important undeciphered scripts of the ancient world is the Indus script. Earlier studies had focused on the correlations between signs in the Indus texts using various statistical and computational techniques such as N-grams or Markov chains. In the present study, K-means clustering, an unsupervised machine learning technique is used to identify clusters of similar texts without making any assumptions about its content. The technique is effective in extracting significant clusters and patterns in the script. Nine clusters are extracted from this study. The texts in each cluster share a common set of structural elements and are more similar to each other than the texts in other clusters. The clusters, as extracted from the study, reveal inherent patterns due to adjacent and non-adjacent dependencies between signs in the Indus texts. These clusters have definitive patterns in the usage of the signs but are only weakly associated to any archaeological site or medium of writing. The characteristic signature features of each cluster are identified in the study. The study provides a good handle to extract the logic of writing in the Indus script.
Clustering Indus Texts using K-means

References

vol. 36, pp. 795-805.

**Index Terms**

Computer Science  
Information Sciences

**Keywords**

Indus texts, ancient script, undeciphered script