Abstract

Data mining and Information extraction is an emerging research area that attempts to find out the hidden and relevant knowledge from the overwhelming amount of information available in structured, semi structured and unstructured forms. Power Iteration Clustering is a powerful clustering algorithm that is proved to be efficient for the clustering of structured data. Large amount of data available in Gmail, micro-blogs etc doesn’t have proper structure renowned as unstructured data which covers remarkable percentage of data available on the internet. Due to the unordered structure of the data, the extraction of relevant information from this huge collection is a complex task. This work used the predominant features of Power
Iteration Clustering algorithm for the extraction of information and visualization of unstructured data from micro blogging sites.

References

- David N. Reshef et al. Detecting Novel Associations in Large Data Sets. Science AAAS. 2011
- L. Xinwu. Research on Text Clustering Algorithm Based on k-means and SOM. International Symposium on Intelligent Information Technology Application Workshops 2008
- Fiedler, Miroslav. An Algebraic Approach to Connectivity of Graphs. Recent advances in graph theory: proceedings of the Symposium held in Prague, June 1975

Index Terms

Computer Science

Power Systems
Keywords
Information Extraction  Micro Blogs  Power Iteration Clustering  Unstructured Data