Abstract

Clustering on web usage data is useful to identify what users are exactly looking for on the world wide web, like user traversals, users behavior and their characteristics, which helps for Web personalization. Clustering web sessions is to group them based on similarity and consists of minimizing the Intra-cluster similarity and maximizing the Inter-group similarity. In the past there exist multiple similarity measures like Euclidean, Jaccard, Cosine, Manhattan, Minkowski, and many to measure similarity between web patterns. In this paper, we enhanced Incremental clustering algorithm (ICA) based on OPTICS. Experiments are performed on MSNBC.COM website (free online news channel), on sequential data streams in the context of clustering in the domain of Web usage mining. Specially, we present a detailed comparison of ICA and OPTICS and the results illustrate that ICA is much more suitable for clustering the dynamic datasets. The Inter-cluster and Intra-cluster distances are computed using Average Levensthen distance (ALD) to demonstrate the usefulness of the proposed approach in the context of web usage mining. This new enhanced (ICA algorithm) has good results when compared with existing OPTICS clustering technique, and provided good time requirements of
the newly developed algorithms.

References


Index Terms

Computer Science Information Systems

Keywords

Sequence Mining, Clustering, Density Based Clustering(optics). Data Mining, Clustering, similarity measures, Web Personalization,