

Social Media Analysis using Optimized K-Means Clustering

Ahmed Alsayat, Doctoral Student
Department of Computer Science, Bowie State University

May 5, 2016 @ 2:30 - 3:30 PM, CSB - Room 210

Ahmed Alsayat is a D.Sc. candidate working under the supervision of Dr. El-Sayed. Mr. Alsayat graduated from Florida Institute of Technology with a Master's of Science in Computer Science. From the same school, he received his second master in Engineering Management. Prior to coming to Bowie State University, Ahmed worked as a lecturer in the department of computer science, Aljouf University, Aljouf, Saudi Arabia, who are sponsoring his post graduate studies. He was also working as the manager of the network and operating systems department in the IT of Aljouf University.

Abstract:

The increasing influence of social media and enormous participation of users creates new opportunities to study human social behavior along with the capability to analyze large amount of data streams. One of the interesting problems is to distinguish between different kinds of users, for example users who are leaders and introduce new issues and discussions on social media. Furthermore, positive or negative attitudes can also be inferred from those discussions. Such problems require a formal interpretation of social media logs and unit of information that can spread from person to person through the social network. Once the social media data such as user messages are parsed and network relationships are identified, data mining techniques can be applied to group different types of communities. However, the appropriate granularity of user communities and their behavior is hardly captured by existing methods. In this paper, we present a framework for the novel task of detecting communities by clustering messages from large streams of social data. Our framework uses K-Means clustering algorithm along with Genetic algorithm and Optimized Cluster Distance (OCD) method to cluster data. The goal of our proposed framework is *twofold* that is to overcome the problem of general K-Means for choosing best initial centroids using Genetic algorithm, as well as to maximize the distance between clusters by pairwise clustering using OCD to get an accurate clusters. We used various cluster validation metrics to evaluate the performance of our algorithm. The analysis shows that the proposed method gives better clustering results and provides a novel use-case of grouping user communities based on their activities. Our approach is optimized and scalable for real-time clustering of social media data.

Contact Dr. Soo-Yeon Ji (sji@bowiestate.edu) if you have any question.