

Abstract

Data modeling using graph has been implemented by many applications and large-scale systems in various fields. By using graph, a real object represented as a node and edge represents relationship between two objects. To understand the characteristics of graph, it needs summarization techniques.

This final project use CANAL (Categorization of Attributes with Numerical Values based on Attribute Values and Link Structures of Nodes) method to summarize graph. This method is an enhancement of Graph-Based Aggregation method which doing summarization by classifying and merging node into a super node to gain knowledge of the data to find the cutoff used in the grouping process automatically. CANAL method improves previous work named SNAP and k-SNAP whose still have deficiency in handling numerical node attribute. Both of them only handle categorical attribute node so when it faced with a numerical attribute, users still have to categorize manually based on their knowledge of the data.

Result of this system is a graph summary that represents the relationship pattern between groups in summary. The pattern could be used to help understand the information that is hidden in the original graph. Then summary produced by CANAL will be compared with summary using manual cutoff. This comparison shows that summary produced by CANAL which almost have same quality as good as the manual cutoff.

Keywords: graph summarization, Aggregation-Based Graph summarization, node attribute, link structure, interestingness measure