



# **An Approach to an Emerging Classification Method for Large Dataset in Clustering**

**Kathiresan V<sup>1</sup>, Dr. P Sumathi<sup>2</sup>**

Research Scholar, Chikkana Government Arts College, Coimbatore<sup>1</sup>

Assistant Professor, Government Arts College, Coimbatore<sup>2</sup>

kathirsujith@gmail.com, sumathirajes@hotmail.com

## **Abstract**

Clustering analysis is used to explore the classification for large dataset and Canberra distance is generalized so that it can process the data with categorical attributes. Based on the generalized Canberra distance definition, an instance of constraint-based clustering is introduced [1]. Meanwhile, the nearest neighbor classification is improved. Class-labeled clusters are regarded as classifying models used for classifying data. The proposed classification method can discover the data of big difference from the instances in training data, which may mean a new data type. The generalize Canberra distance for continuous numerical attributes data to mixed attributes data, and use clustering analysis technique to squash existing instances, improve the classical nearest neighbor classification method.

**Keywords** – ID3; C4.5; Canberra Distance; Clustering; Improved Nearest Neighbour

Full Text: <http://www.ijcsmc.com/docs/papers/January2014/V3I1201468.pdf>