A NEW APPROACH TO DATA CLUSTERING USING A COMPUTATIONAL VISUAL ATTENTION MODEL

Peilin Jiang[1,2], Fuji Ren[1,3] and Nanning Zheng[2]

[1]Graduate School of Advanced Science Technology Education
The University of Tokushima
Tokushima, 7708506, Japan
{ jiang; ren } @is.tokushima-u.ac.jp

[2]Institute of Artificial Intelligence and Robotics
Xian Jiaotong University
Xian, P. R. China
{ pljiang; nnzheng } @aiar.xjtu.edu.cn

[3]Beijing University of Posts and Telecommunications
Beijing, P. R. China

ABSTRACT. *Cluster analysis plays an important role in many respects such as knowledge discovery, data mining and information retrieval. In this paper, we propose a new approach inspired by the early vision system of the primate for data clustering. Human beings are able to locate key points that contains more important information in a complex scene. To realize this function, our approach uses a computational visual attention model that selects and extracts salient areas in visual field by local difference features. Then the extracted salient areas in original visual field can be regarded as the clusters in the data feature space. Without prior knowledge, this attention model based approach can identify data clusters with arbitrary shapes at different scales. Finally our algorithm has been tested in the evaluation experiments on the benchmark datasets to show its competitive performance.*
Keywords: Data clustering, Bio-inspired approach, Selective attention, Saliency map

1. **Introduction.** The data clustering methods are designed to discover the meaningful subclasses in a set of data. It is an unsupervised learning method that is used to explore the implicit relationships among the attributes of the data. Up to now a large quantity of clustering methods have been presented in the past years [1, 2, 3]. One problem of current clustering methods is that it is confusing how to interpret the clusters in the data [2]. In most situations, various clustering methods focus on different data sets and there is not any general method which can solve every issue. Another key problem is no prior labeled data available before unsupervised classification and we have no idea on the relationships and distribution of the attributes of the data. Some methods such as the well-known K-means algorithm and Fuzzy C Means clustering method [4, 5] need to preliminarily know the number of clusters. However there is little prior knowledge before we start to solve the clustering problems.

In general, the traditional clustering algorithms are composed by several groups: partitioning algorithm [9, 10], hierarchical algorithm [6], density based algorithm [7], grid based algorithm [8], model based algorithm and recently, some researches such as bio-inspired clustering algorithms have been presented and shown good performance, such as swarm-based [11], and flocks based algorithms [12] which are verified to be powerful. Each kind of methods shows its advantages towards different problems: initial condition