Neural Process Lett
DOI 10.1007/s11063-017-9597-3

# A New Evolving Tree-Based Model with Local Re-learning for Document Clustering and Visualization

**Wui Lee Chang**[1] · **Kai Meng Tay**[1] · **Chee Peng Lim**[2]



**Abstract** The Evolving tree (ETree) is a hierarchical clustering and visualization model that allows the number of clusters to grow and evolve with new data samples in an online learning manner. While many hierarchical clustering models are available in the literature, ETree stands out because of its visualization capability. It is an enhancement of the Self-Organizing Map, a famous and useful clustering and visualization model. ETree organises the trained data samples in the form of a tree structure for better presentation and visualization especially for high-dimensional data samples. Even though ETree has been used in a number of applications, its use in textual document clustering and visualization is limited. In this paper, ETree is modified and deployed as a useful model for undertaking textual documents clustering and visualization problems. We introduce a new local re-learning procedure that allows the tree structure to grow and adapt to new features, i.e., new words from new textual documents. The performance of the proposed ETree model is evaluated with two (one benchmark and one real) document data sets. A number of key aspects of the proposed ETree model, which include its topology representation, learning time, as well as recall and precision rates, are evaluated. The results show that the proposed local re-learning procedure is useful for handling increasing number of features incrementally. In summary, this study contributes towards a modified ETree model and its use in a new domain, i.e., textual document clustering and visualization.

**Keywords** Evolving tree · Textual documents · Clustering · Visualization · Local re-learning

## 1 Introduction

Clustering is a process of organising a set of data samples comprising multi-dimensional features into different appropriate groups based on data similarity [1]. The data samples within

✉ Kai Meng Tay
kmtay@unimas.my; tkaimeng@yahoo.com

1 Faculty of Engineering, Universiti Malaysia Sarawak, Kota Samarahan, Sarawak, Malaysia

2 Institute for Intelligent Systems Research and Innovation, Deakin University, Geelong, Australia

Published online: 06 February 2017