Chapter 5
Self–Organizing Tree Using Artificial Ants

Hanene Azzag
Université Paris 13, France

Mustapha Lebbah
Université Paris 13, France

ABSTRACT

In this paper, the authors propose a new approach for topological hierarchical tree clustering inspired from the self-assembly behavior of artificial ants. The method, called SoTree (Self-organizing Tree), builds, autonomously and simultaneously, a topological and hierarchical partitioning of data. Each ‘‘cluster’’ associated to one cell of a 2D grid is modeled by a tree. The artificial ants similarly build a tree where each ant represents a node/data. The benefit of this approach is the intuitive representation of hierarchical relations in the data. This is especially appealing in explorative data mining applications, allowing the inherent structure of the data to unfold in a highly intuitive fashion.

INTRODUCTION

The quantity of information that is stored in databases is rapidly growing. Additionally, databases are not only composed of numeric or symbolic attributes but may also be enriched, for instance, by sounds, images, videos, texts or Web sites. Finding the relevant hidden information in structured datasets is a difficult task. Clustering and visualization methods are one of the basic techniques which are often applied in analyzing the structured data sets. The main focus of this paper is on helping the domain expert to intuitively analyze such data sets. In our opinion, intuitive analysis of data implies the use of hierarchical clustering and visualization techniques (Jain & Dubes, 1988; Jain & Murty, 1999; Handl, Knowles, & Dorigo, 2003). These methods use both heuristic and mathematics principles.
Concerning heuristics methods, the numerous abilities of ants have inspired researchers for more than ten years regarding designing new clustering algorithms (Goss & Deneubourg, 1991; Lumer & Faieta, 1994). The initial and pioneering work in this area is due to Deneubourg and his colleagues (Goss & Deneubourg, 1991). These researchers have been interested in the way real ants sort objects in their nest by carrying them and dropping them in appropriate locations without a centralized control policy (Deneubourg, Goss, Franks, Sendova-Franks, Detrain, & Chretien, 1990). They have proposed the following principles: artificial ants move in a 2D grid where all objects to be sorted are initially scattered. An ant has a local perception of objects around its location and does not communicate directly with other ants. Instead, ants influence themselves using the configurations of objects on the grid. An ant has a variable probability of picking an encountered object. This probability is high when the frequency of encountering an object is low in the recent history of the ant. Similarly, an ant will drop a carried object with a high probability when similar objects have been perceived in ant’s neighborhood. Using this simple and distributed principle, ants are able to collectively sort the objects.

The next step toward data clustering has been performed in (Lumer & Faieta, 1994). These researchers have adapted the previous algorithm by assuming that an object is a datum and by tuning the picking/dropping probabilities according to the similarities between data. These ants-based algorithms inherit interesting properties from real ants, such as the local/global optimization of the clustering, the absence of need of a priori information on an initial partitioning or number of classes, or parallelism. Furthermore, the results are presented as visualization: a property which is coherent with an important actual trend in data mining called ‘visual data mining’ where results are presented in a visual and interactive way to the expert.

Many authors have used and extended this model especially in the case of real and more difficult data with very interesting results, such as in graph partitioning for VSLI (Very-large-scale integration) technology (Kuntz, Layzell, & Snyers, 1997; Kuntz, Snyers, Layzell, & Paul, 1998), in Web usage mining (Abraham & Ramos, 2003) or in document clustering (Handl, Knowles, & Dorigo, 2003). Other models of ‘clustering ants’ have been developed (Venturini, Labroche, & Guinot, 2004), but the model which has been studied the most is the way ants sort objects in their nest. In this work we introduce a new method named SoTree: Self-Organizing Tree which uses in the same way: a hierarchical and a topological clustering. Data moves autonomously toward a 2D Grid respecting bio-inspired rules, where each cell represents a tree structured data. Thus we will obtain in one pass: a Horizontal topological clustering and a vertical hierarchical clustering (tree in each cell). The topological function of the proposed algorithm is based on Kohonen approach (Kohonen, 2001) and the rules for building tree are based on artificial ants method named AntTree (Azzag, Guinot, Oliver, & Venturini, 2006; Azzag, Guinot, & Venturini, 2006).

The remaining of this paper is organized as follows: in the next section, we present the artificial ants model, followed by the topological model. We layout the main principles of the combined model and propose rules to build tree clustering. We also present the methodology and experimental results. The final section concludes this work and proposes some perspectives.

FROM REAL TO ARTIFICIAL ANTS: ANTTREE ALGORITHM

Biological Inspiration

The self-assembly behavior of individuals can be observed in several insects like bees or ants for instance (see a survey with spectacular photographs