Abstract:
This paper presents a new data structure called Voronoi tree to support the solution of proximity problems in general pseudo metric spaces with efficiently computable distance functions. We analyse some structural properties and report experimental results showing that Voronoi trees are a proper and very efficient tool for the representation of proximity properties and generation of suitable clusterings.

1. Introduction

Cluster analysis is an important pattern recognition technique. It may be characterized by the use of resemblance or dissemblance measures between objects to be identified. The objective of a cluster analysis is to uncover natural groupings, or types, of objects. [DS], [DJ]

In this paper we are concerned with the problem of how to support proximity and clustering problems in pseudo metric spaces. A set $E$ is called pseudo metric space if there is a distance function $d: E^2 \rightarrow \mathbb{R}_+$ with the following properties (for any $e, e', e'' \in E$):

1. $d(e,e) = 0$
2. $d(e,e') = d(e',e)$ (symmetry)
3. $d(e,e'') \leq d(e,e') + d(e',e'')$ (triangle inequality)

However, it may happen in pseudo metric spaces that there are two different elements $e, e' \in E$ with $d(e,e') = 0$.

Most of the literature on nearest neighbor problems is dealing with finite dimensional real spaces with some $l_p$ norm – very few

(1) School of Computer Science, Carleton Univ., Ottawa, Canada K1S5B6
(2) Lehrstuhl für Informatik I, Univ. of Würzburg, Würzburg, W.-Germany
are considering more general spaces. One of these few more general approaches is the recent work of I. Kalantari and G. McDonald [KMcD]. The data structure they propose is a straightforward generalization of binary search trees (which can support nearest neighbor search in $E^1$) and is applicable to any normed space if the norm is computable effectively. The next section will briefly summarize their approach but will also point out some disadvantages.

In section 3 we propose a slightly different structure (which we call 'Voronoi tree') which, however, will show significant structural advantages.

Some experimental results will be stated in section 4. Due to their structural advantages it turned out that Voronoi trees are a proper and very efficient tool for the representation of proximity properties and solution of clustering problems.

2. Bisector Trees

Let $E$ be an arbitrary space, $d: E^2 \rightarrow R_+$ a mapping which induces a metric on $E$, and $S=\{e_1, \ldots, e_n\}$ a finite point set in $E$. [KMcD] represents $S$ using a binary tree (called 'bisector tree' or 'bs-tree' for the remaining of this paper) as follows:

(a) each node of the actual tree $T_i$ (representing the actual set $S_i=\{e_1, \ldots, e_i\}$, $0 \leq i \leq n$) contains at least one ($p_L$) and at most two elements ($p_L$ and $p_R$) of $S_i$.

(β) to insert a new element $e_{i+1}$ into $T_i$ (yielding $T_{i+1}$) start at the root node of $T_i$ as follows:

- if the actual node $v$ contains only one element then insert $e_{i+1}$ into $v$

- if (otherwise) $v$ contains two elements $p_L$ and $p_R$ then recursively insert $e_{i+1}$ into the left subtree (which has root $p_L$) if $d(e_{i+1}, p_L) < d(e_{i+1}, p_R)$ or into the right subtree (which has root $p_R$) if $d(e_{i+1}, p_R) < d(e_{i+1}, p_L)$ and arbitrarily if the distances are equal.