SMART REACH
(S/W CREATORS & TRAINERS)

Ph: 9585554590, 9585554599
Email: support@salemsmartreach.com
URL: www.salemsmartreach.com

OCCT: A One-Class Clustering Tree for Implementing One-to-Many Data Linkage

Abstract:

One-to-many data linkage is an essential task in many domains, yet only a handful of prior publications have addressed this issue. Furthermore, while traditionally data linkage is performed among entities of the same type, it is extremely necessary to develop linkage techniques that link between matching entities of different types as well. In this paper, we propose a new one-to-many data linkage method that links between entities of different natures. The proposed method is based on a one-class clustering tree (OCCT) that characterizes the entities that should be linked together. The tree is built such that it is easy to understand and transform into association rules, i.e., the inner nodes consist only of features describing the first set of entities, while the leaves of the tree represent features of their matching entities from the second data set. We propose four splitting criteria and two different pruning methods which can be used for inducing the OCCT. The method was evaluated using data sets from three different domains. The results affirm the effectiveness of the proposed method and show that the OCCT yields better performance in terms of precision and recall (in most cases it is statistically significant) when compared to a C4.5 decision tree-based linkage method.