Proc. 20th Int. Conf. on Data Engineering (ICDE'04), Boston, MA, 2004

# BOSS: Browsing OPTICS-Plots for Similarity Search

Stefan Brecheisen, Hans-Peter Kriegel, Peer Kröger, Martin Pfeifle, Maximilian Viermetz
Institute for Computer Science, University of Munich
{brecheis,kriegel,kroegerp,pfeifle,viermetz}@dbs.informatik.uni-muenchen.de

Marco Pötke
sd&m AG software design & management
marco.poetke@sdm.de

## 1. Introduction

In the last ten years, an increasing number of database applications has emerged for which efficient and effective support for similarity search is substantial. Particularly, the task of finding similar shapes in 2-D and 3-D becomes more and more important. Examples for new applications that require the retrieval of similar 3-D objects include databases for molecular biology, medical imaging and computer aided design. Hierarchical clustering was shown to be effective for evaluating similarity models [3]. Furthermore, visually analyzing cluster hierarchies helps the user, e.g. an engineer, to find and group similar objects. This demonstration presents an interactive browsing tool called *BOSS* (*B*rowsing *O*PTICS-Plots for *S*imilarity *S*earch), which utilizes solid automatic cluster recognition and extraction of meaningful cluster representatives in order to provide the user with significant and quick information.

## 2. System Overview

The basis of our interactive browsing tool are reachability plots generated by OPTICS [1] which is a density-based hierarchical clustering algorithm. The key idea of density-based clustering is that for each object of a cluster the neighborhood of a given radius $\varepsilon$ has to contain at least a minimum number *MinPts* of objects. OPTICS is relatively insensitive to its input parameters [1] and being a hierarchical clustering method it yields more information about the cluster structure than a method that computes a flat partitioning of the data. BOSS is a comprehensive, scalable and distributed computing solution designed to make the advantageous properties of OPTICS and the analytical capabilities of similarity search available to a broader audience. It is a client/server system allowing users to provide their own data locally, along with an appropriate similarity model.

## 3. Demonstration Overview

Based on reachability plots, BOSS allows the user to inspect the cluster structure and the cluster contents inherent in a data set with respect to a given similarity model. The user can achieve this either by manually selecting clusters or objects contained therein or by using the automatic approaches for cluster extraction and cluster representation implemented in BOSS. These techniques can be used as a basis for visual data mining. Currently, three different algorithms for automatic cluster recognition are implemented [2]. Furthermore, for finding cluster representatives, two new approaches [2] are implemented along with the commonly known medoid approach. The user may select any of these algorithms at run time. The data provided by the user consists of the feature vectors of the objects to be clustered, as well as of a set of media files to visualize these objects, e.g. VRML files for CAD data or JPEG images of proteins. By means of suitable plug-ins, multimedia data such as images or 3-D models can easily be displayed in a web browser. By externalizing the visualization procedure we can resort to approved software components, which have been developed for displaying the particular file format.

## References

[1] M. Ankerst, M. M. Breunig, H.-P. Kriegel, and J. Sander. "OPTICS: Ordering Points to Identify the Clustering Structure". In *Proc. ACM SIGMOD Int. Conf. on Management of Data (SIGMOD'99)*, pages 49–60, 1999.

[2] S. Brecheisen, H.-P. Kriegel, P. Kröger, M. Pfeifle, and M. Viermetz. "Representatives for Visually Analyzing Cluster Hierarchies". In *Proc. 4th Int. Workshop on Multimedia Data Mining (MDM/KDD'03)*, pages 64–71, 2003.

[3] H.-P. Kriegel, P. Kröger, Z. Mashael, M. Pfeifle, M. Pötke, and T. Seidl. "Effective Similarity Search on Voxelized CAD Objects". In *Proc. 8th Int. Conf. on Database Systems for Advanced Applications (DASFAA'03)*, pages 27–36, 2003.