Khiops CoViz: A Tool for Visual Exploratory Analysis of k-Coclustering Results

Bruno Guerraz, Marc Boullé, Dominique Gay, and Fabrice Clérot

Orange Labs
2, avenue Pierre Marzin, F-22307 Lannion Cedex, France
firstname.name@orange.com

Abstract. Identifying and visually analyzing interesting interactions between variables in large-scale data sets through k-coclustering is of high importance. We present Khiops CoViz [1], a tool for visual analysis of interesting relationships between two or more variables (categorical and/or numerical). The visualization of k variables coclustering takes the form of a grid/matrix whose dimensions are partitioned: categorical variables are grouped into clusters and numerical variables are discretized. The tool allows several kinds of visualization at various scales for grid representation of coclustering results by means of several criteria each of which providing different insights into the data. Hereafter, several screen shots describe the main visual components of the tool.

[1] http://www.khiops.com, an exhaustive user manual is available when downloading the tool.

T. Calders et al. (Eds.): ECML PKDD 2014, Part III, LNCS 8726, pp. 444–447, 2014.

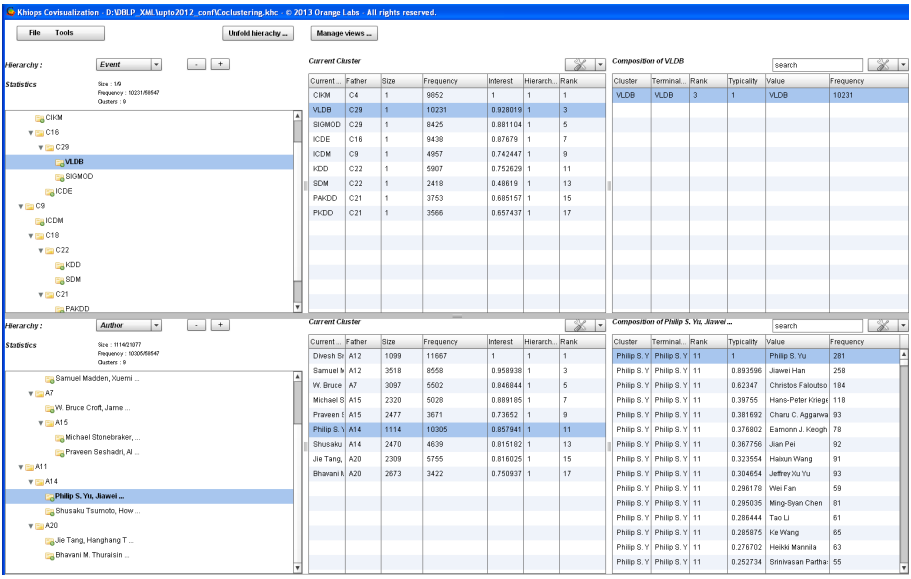

Fig. 1. Case study on a examplary sample of DBLP data set. Three-dimensional data ($Author \times Year \times Event$) is considered. A screen shot of Khiops CoViz: (from left to right) hierarchies of parts for two dimensions ($Author$ and $Event$), list of terminal parts of hierarchies and composition of selected parts.

Fig. 2. (Left): Grid/matrix visualization of contribution to mutual information for $Year \times Event$ dimensions. Other criteria for visualization are available: e.g., contrast, frequency, conditional probability and joint probability... The tool allows navigating along the partition of a selected dimension (e.g., *Author*) while the others ($Year, Event$) are fixed and dedicated to the visualization.

Fig. 3. Choosing the wanted granularity for visualizing the grid is available through the "Unfold Hierarchy" functionality. The user can control either the number of parts of the dimensions or the grid quality (w.r.t. to the optimal grid) by optimal merging or per-dimension customized non-optimal merging.