TaCo: Comparative Visualization of Large Tabular Data

Reem Hourieh, Holger Stitz, Nils Gehlenborg and Marc Streit

Tabular data plays a vital role in many different domains. In the course of a project, changes to the structure and content of tables can result in multiple instances of a table. TaCo (Table Comparison) is an interactive comparison tool that effectively visualizes the differences between multiple tables at various levels of granularity:

1. Calculating the difference between tables
2. Visualizing the difference in an effective and scalable way

Existing table comparison tools (e.g., DiffKit [1] and Daff [2]) generate visualizations that lack the ability to perform simultaneous row-wise, column-wise, and cell-wise comparison of tables [5,6].

Comparing large tabular data requires a two-part solution:

1. Calculating the difference between tables
2. Visualizing the difference in an effective and scalable way

Figure 1: The multi-view interface of TaCo showing multiple instances of an artificially generated table. The overview on the left side [1] lets the user select a collection of tables that are plotted using (2) Multidimensional Scaling (MDS) based on the calculated similarity among the tables. (3) The user compares one selected reference table to all other tables in LineUp. The middle view on the right side shows (4) the reference table as a heatmap and (5) the aggregated differences to a selected group of tables for both row and column changes.

Figure 2: The four-stage visualization approach allows users to reduce the number of comparable tables from one stage to the next stage, while increasing the details shown.

Figure 3: The difference between two tables is visualized as (1) a difference table. Changes are summarized on a per cell basis and visualized as a (2) ratios bar plot. The difference table can be aggregated for (3) row and (4) column directions separately. (5, 6) Further aggregation for one direction is shown as a histogram. Summarizing changes for rows and columns results as a (7) 2D-ratio visualization.

Figure 4: The aggregation result of four TCGA tables is shown in the middle view. (2) One table was selected for a detail comparison and is visualized side-by-side with (3) the reference table as heatmap and (4) the diff table. The color scale of content changes can be manipulated by the user with (5) the color-slider.

Acknowledgements

This work was supported by:
- Austrian Science Fund (FWF P27975-NBL)
- State of Upper Austria (FFG 051466)

Contact
- holger.stitz@jku.at
- calydo.org
- github.com/Caleydo/taco

References