

Ziliang WU, Wei CHEN, Yuxin MA, Tong XU, Fan YAN, Lei LV, Zhonghao QIAN, Jiazhi XIA, 2023. Explainable data transformation recommendation for automatic visualization. *Frontiers of Information Technology & Electronic Engineering*, 24(7):1007-1027. <https://doi.org/10.1631/FITEE.2200409>

Explainable data transformation recommendation for automatic visualization

Key words: Data transformation; Data transformation recommendation; Automatic visualization; Explainability

Corresponding author: Wei CHEN

E-mail: chenvis@zju.edu.cn

 ORCID: <https://orcid.org/0000-0002-8365-4741>

Motivation

1. Automatic visualization seeks to support data analysis and pattern finding for novice or casual users, while the currently adopted limited data transformations fail to capture complex patterns such as clusters and correlations.
2. Recent advances in feature engineering provide the potential for more kinds of automatic data transformations, while the auto-generated transformations lack explainability concerning how patterns are connected with original features.
3. An explainable automatic approach for involving extended kinds of data transformations in automatic visualization is needed.

Main idea

1. To tackle these challenges, we propose a novel explainable recommendation approach.
2. The space of feasible data transformations is summarized and delineated.
3. The measures on explainability of transformation operations are defined.
4. A recommendation algorithm is designed to compute optimal transformations, which can reveal specified types of patterns and maintain explainability.

Method

1. The transformations are defined as operations that transform the original data into new forms. Feasible transformations are derived from research literature and summarized as three categories: data preparation (DP), data mining (DM), and data wrangling (DW).

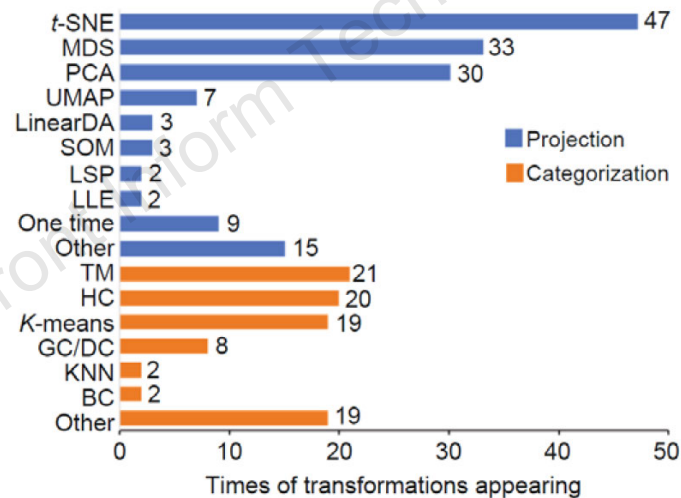


Fig. 1 Survey results of data mining (DM) transformation in IEEE Visualization Conference (VIS) papers

Method (Cont'd)

2. The explainability of data transformation is studied through user experiments. Three explainability measures and a pruning searching approach are proposed:

- Over-transformation
- Dimension matching
- Semantic information

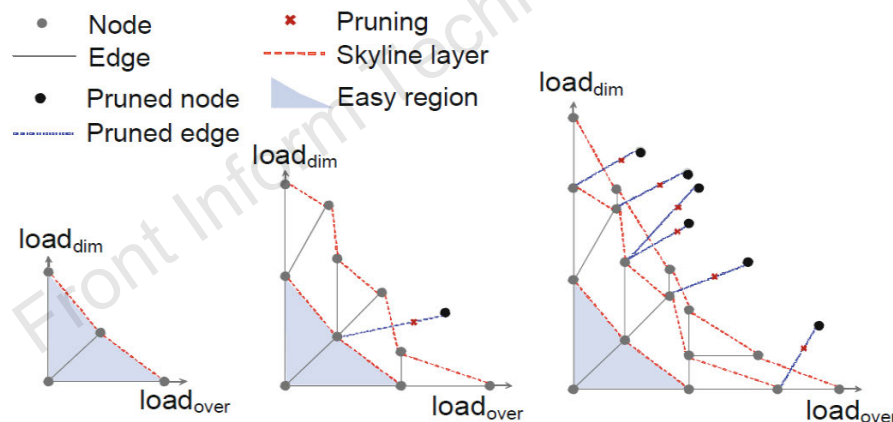


Fig. 5 The pruning process: (a) easy region; (b) layer-2; (c) layer-3. In layers out of the easy region, a skyline operator is executed to prune dominated points. The remaining points form the skyline layer and generate the next layer of points. References to color refer to the online version of this figure

Method (Cont'd)

3. In a complete recommendation workflow, the proposed method contains three progressive parts:

- space
- searching
- scoring

The method constructs a tree-structure search space, searches explainable transformation paths, and evaluates the resulting visualizations.

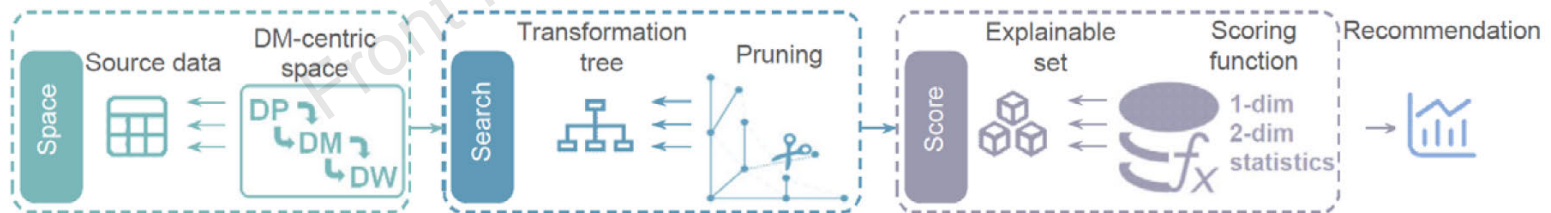


Fig. 4 Our approach consists of three parts, including searching space, pruning search, and scoring functions. A DM-centric searching space is first constructed based on the source data. As the transformation tree expands, nodes with low explainability are pruned to generate an explainable set. Insight scoring functions are applied on the set to rank candidates and decide recommendations. DP: data preparation; DM: data mining; DW: data wrangling

Major results

A prototype system was implemented to validate the proposed approach.

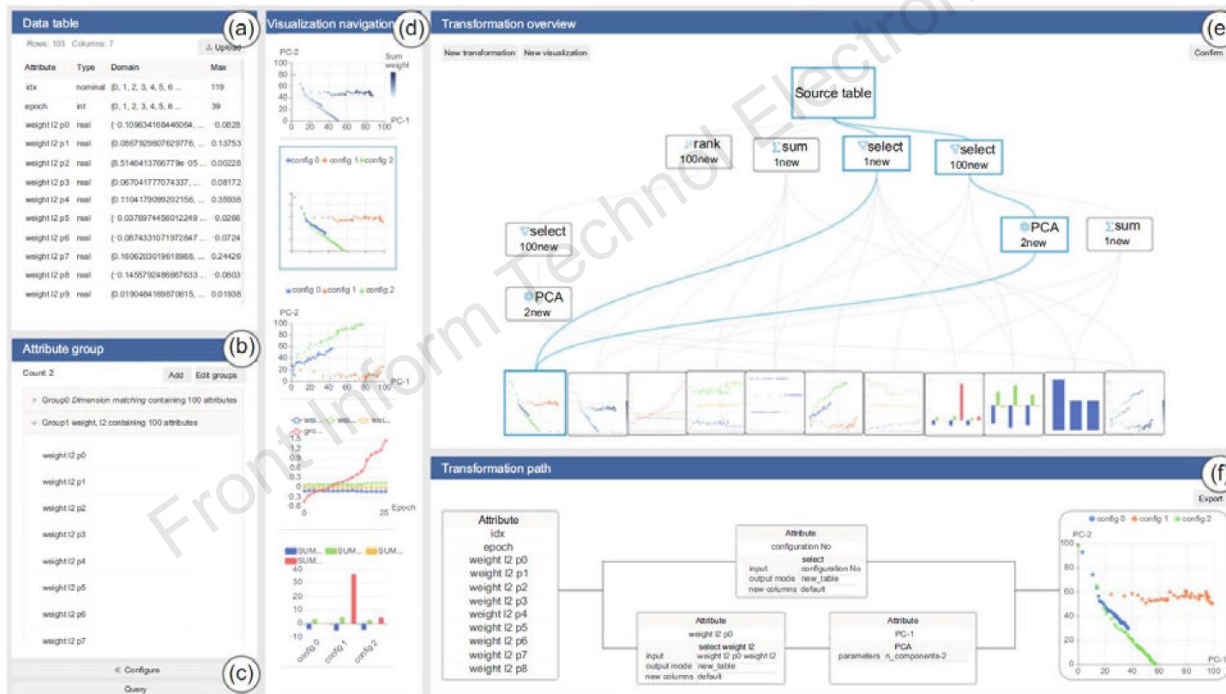


Fig. 6 Our visual interface. The data view (a) shows basic information. The attribute group view (b) presents groups automatically detected by the system and enables users to edit grouping situations. Two buttons (c) allow users to configure system parameters and trigger recommendations. A visualization navigation list (d) presents previews of recommendation. Overall transformation recommendations are presented in a tree structure with visualization thumbnails (e) and detailed information is shown below (f)

Major results (Cont'd)

Two usage cases of the system based on two real-world datasets are demonstrated.



Fig. 7 Two recommendation cases: (a-c) recommended based on an international trade dataset; (d-f) generated based on a model training dataset

Major results (Cont'd)

A quantitative expert comparative experiment and a qualitative user experiment were conducted.

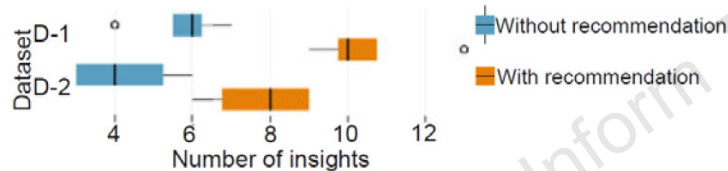


Fig. 8 Numbers of discovered insights in the user study. Each dataset is analyzed by six participants with recommendation and six participants without recommendation

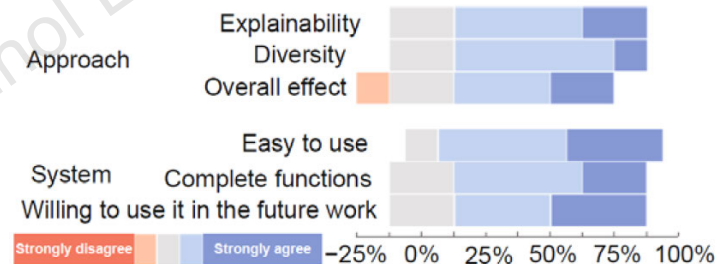


Fig. 9 User ratings for the recommendation approach and system on a five-point Likert scale (12 participants)

Conclusions

1. A novel data transformation recommendation approach was proposed for enhancing the pattern discovery ability of automatic visualization.
2. A space of feasible data transformations was summarized by surveying the literature, and a pilot study was conducted to locate three significant influencing factors and proposed the comprehension load to quantify them.
3. A prototype system was implemented to validate and evaluate the approach. Usage cases and user study demonstrated the effectiveness and efficiency of the approach.



Wei CHEN is a professor at the State Key Lab of CAD & CG, Zhejiang University. His research interests are on visualization and visual analysis. He has published more than 30 IEEE/ACM Transactions and IEEE VIS papers. He actively served as guest or associate editors of *IEEE Transactions on Visualization and Computer Graphics*, *IEEE Transactions on Intelligent Transportation Systems*, and *Journal of Visualization*. For more information, please refer to <http://www.cad.zju.edu.cn/home/chenwei>.

Front Inform Technol