Adopting Mental Similarity Notions of Categorical Data Objects to Algorithmic Similarity Functions

David Sessler†1 and Jürgen Bernard† and Arjan Kuijper‡ and Jörn Kohlhammer

Fraunhofer IGD, Germany 2TU Darmstadt, Germany

Abstract
The creation of similarity functions based on visual-interactive user feedback is a promising means to capture the Mental Similarity Notion (MSN) in the heads of domain experts. To this end, the user interface (UI) and the algorithmic feedback model (FM) of such interactive systems need to comply with the MSN. The user feedback for categorical attributes shows specific characteristics which we analyze in this work. To this end, we first conducted a user study to observe the users intuition and to identify challenges for the design of the algorithmic FM. Finally, we present a solution that overcomes these challenges for categorical attributes.

Introduction and Background We conducted a preliminary user study to observe how users align objects on a 2D data landscape to express their MSN. In particular, we were interested in the users intuition of arranging objects of different categories (see the attached poster). The observation showed that for different object categories most users prefer object arrangements with identical pairwise distances in terms of regular geometries. However, the intuition of pairwise distances poses challenges for both the UI and the FM since it is not possible to represent equal pairwise distances for more than three objects in 2D. As a consequence, for more than three objects the MSN of a domain expert diverges from the geometrical ‘feasibility’ in 2D. In a previous work, we presented a FM which utilizes pairwise distances between objects [BSR'14]. The result of the FM is a weighting of every attribute. In this work, we extend the FM with an additional improvement step with tackles the described challenge for categorical attributes.

Test Description We generalize the object arrangements observed in the user study and apply these as the basis for the test setup. The results are shown at the left of Figure 1. White dots indicate the tested geometries, colored boxplots show the attribute weightings calculated by the FM (see caption for more details). We indicate lower weights for object geometries with a higher cardinality (the described challenge). Stacking additional objects onto the arrangements, e.g., 8 objects onto 5 category positions (the nested x-axis of the diagram) slightly improves the attribute weighting. Further, different geometries with the same cardinality yield different weights.

Improvement Step Our goal is to provide a FM which complies with the MSN of the users. Therefore, the FM should generate optimal attribute weights for the preferred arrangement strategies. To achieve this, we adopt the median weights obtained in the test as the new optimal weight criteria. The results of the test repetition at the right of Figure 1 show the improvement of the attribute weightings.

Discussion In this work, we improved the weighting generation of the FM to cope with user-preferred arrangements for categorical objects. However, we tested the weighting generation on uniformly distributed attributes. In the future, we intend to investigate the dependency of generated attribute weightings on non-uniform distributions.

References

Figure 1: Results of the Initial Test and the Test Repetition Boxplots showing the functional dependencies between the object geometry (global x-axis), the number of objects used in the feedback (nested x-axis), and the attribute weight [0.0-1.0] (y-axis). The white dots represent the object geometry, the percentage values reflect the choice of the users. Left: an initial test shows that the FM did not generate optimal weights for perfect object geometries with a cardinality of four or higher. Right: a second test under the same conditions shows that the improved FM yields a better attribute weighting.

submit to Vision, Modeling, and Visualization (2014)