Analyzing Psychological Similarity Spaces for Shapes

1 Background and Motivation

The cognitive framework of conceptual spaces [3] proposes to represent concepts and properties such as APPLE and ROUND as convex regions in perception-based similarity spaces. By doing so, the framework can provide a grounding for the nodes of a semantic network. Moreover, semantic similarity and concept hierarchies can be derived from the geometric notions of distance and subsethood, respectively. In order to use this framework in practice, one needs to know the structure of the underlying similarity space. In our study, we focus on the domain of shapes. We analyze similarity spaces of varying dimensionality which are based on human similarity ratings.

2 Data Collection

We used 60 standardized black-and-white line drawings of common objects (six visually consistent and six visually variable categories with five objects each) for our experiments. We collected 15 shape similarity ratings for all pairwise combinations of the images. Image pairs were presented one after another on the screen (in random order) and subjects were asked to judge the respective similarity on a Likert scale ranging from 1 (totally dissimilar) to 5 (very similar). The distribution of within-category similarities showed that the internal shape similarity was higher for visually consistent categories($M = 4.18$) than for visually variable categories($M = 2.56$; $p < .001$). For further processing, the shape similarity ratings were aggregated into a global matrix of dissimilarities by taking the mean over the individual responses and by inverting the scale (i.e., $dissimilarity(x, y) = 5 - similarity(x, y)$).

In the psychological literature, different types of perceptual features are discussed as determining the perception of complex objects, among others the line shape (LINES) and the global shape structure (FORM) [1]. We collected values for all images with respect to these two features in two experimental setups.

In a first line of experiments, we collected image-specific ratings which are based on *attentive (att)* image perception (9 ratings per image). Groups of four images were presented one after another on the screen (in random order) together with a continuous scale representing the respective feature (LINES: absolutely straight to strongly curved; FORM: elongated to blob-like). Subjects were asked to arrange the images on the respective scale such that the position of each image in the final configuration reflected their value on the respective feature scale. The resulting values were aggregated for each image by using the median.

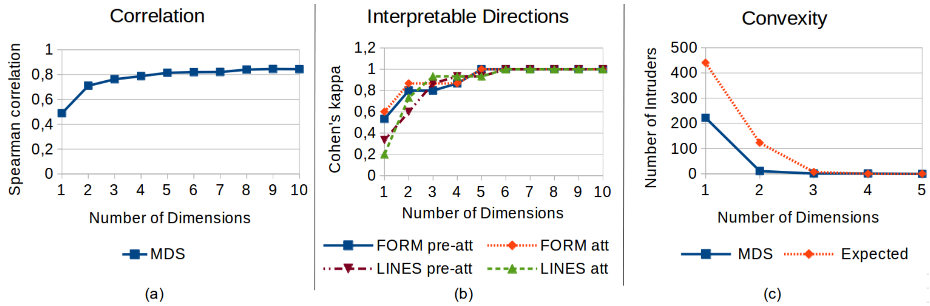

Fig. 1. Results of our analysis of the similarity spaces.

In a second line of experiments, we collected image-specific values which are based on *pre-attentive (pre-att)* image perception. Each image was presented for 50 ms on the screen; immediately before and after the image a pattern mask was shown for 50 ms in order to prevent conscious perception of the image. Subjects were asked to decide per button press as fast as possible which property of the respective feature pertained to the critical image mostly (Lines: straight vs. curved; Form: elongated vs. blob-like). The binary values (in total 18 per image for each feature) were transformed into graded values (percentage of curved and blob-like responses, respectively).

A comparison of the two types of image-specific values for the dimensions revealed a strong correlation between attentive and pre-attentive shape perception ($r_s = 0.83$ for Lines and $r_s = 0.85$ for Form). In both cases, the 15 images with the highest and lowest values were used as positive and negative examples for the respective feature.

3 Analysis

We used the SMACOF algorithm [4] for performing nonmetric multidimensional scaling (MDS) on the dissimilarity matrix. Given a desired number n of dimensions, MDS represents each stimulus as a point in an n-dimensional space and arranges these points in such a way that their pairwise distances correlate well with the pairwise dissimilarities of the stimuli they represent.

Figure 1a shows the Spearman correlation of dissimilarities and distances as a function of the number of dimensions. As we can see, a one-dimensional space is not sufficient for an accurate representation of the dissimilarities. We can furthermore observe that using more than five dimensions does not considerably improve the correlation to the dissimilarities. As a baseline, we have also computed the distances between the pixels of various downscaled versions of the images. These pixel-based distances reached only a Spearman correlation of $r_s = 0.40$ to the dissimilarities, indicating that shape similarity cannot easily be determined based on raw pixel information.

In order to identify interpretable directions in the similarity spaces, we trained a linear support vector machine to separate positive from negative examples for each of the candidate features. The normal vector of the separating hyperplane can be interpreted as the direction representing this feature [2]. Figure 1b shows the quality of this separation (measured with Cohen's kappa) as a function of the number of dimensions. While a one-dimensional space again gives poor results, increasing the number of dimensions improves the evaluation metric. Six dimensions are always sufficient for perfect classification. Moreover, it seems like the feature FORM is found slightly earlier than LINES. Finally, we do not observe considerable differences between pre-attentive and attentive ratings.

The framework of conceptual spaces proposes that conceptual regions in the similarity space should be convex. We have therefore also counted the number of intruders inside the convex hull for each of the categories. Figure 1c plots the overall number of intruders as a function of the number of dimensions. As we can see, the expected number of violations drops very fast with more dimensions and becomes zero in a five-dimensional space. However, the solution found by MDS produces clearly less violations of the convexity criterion. Overall, it seems that conceptual regions tend to be convex in our similarity spaces.

4 Conclusions

In our study, we found that similarity spaces with two to five dimensions seem to be good candidates for representing shapes: A single dimension does not seem to be sufficient while more than five dimensions do not improve the quality of the space. The shape features postulated in the literature were indeed detectable as interpretable directions in these similarity spaces. In order to understand the similarity space for shapes even better, additional features from the literature (such as ORIENTATION) will be investigated.

Moreover, the similarity spaces obtained through MDS can only be used for the given set of stimuli. In future work, we aim to overcome this limitation by training an artificial neural network on mapping images to points in the shape similarity spaces (cf. [5]).

References

1. Biederman, I.: Recognition-by-Components: A Theory of Human Image Understanding. Psychological Review **94**(2), 115–147 (1987)
2. Derrac, J., Schockaert, S.: Inducing Semantic Relations from Conceptual Spaces: A Data-Driven Approach to Plausible Reasoning. Artificial Intelligence **228**, 66–94 (Nov 2015). https://doi.org/10.1016/j.artint.2015.07.002
3. Gärdenfors, P.: Conceptual Spaces: The Geometry of Thought. MIT press (2000)
4. de Leeuw, J.: Recent Development in Statistics, chap. Applications of Convex Analysis to Multidimensional Scaling, pp. 133–146. North Holland Publishing (1977)
5. Sanders, C.A., Nosofsky, R.M.: Using Deep-Learning Representations of Complex Natural Stimuli as Input to Psychological Models of Classification. In: Proceedings of the 2018 Conference of the Cognitive Science Society, Madison. (2018)