

Dataset Description for “Metric Learning for Estimating Psychological Similarities”

Jun-Ming Xu, Xiaojin Zhu, Timothy T. Rogers
University of Wisconsin-Madison

February 23, 2012

1 Real Datasets

1.1 Participants

Participants were undergraduate students in an Introductory Psychology course at the University of Wisconsin-Madison who volunteered to participate in exchange for extra credits in the course.

1.2 Procedure

Up to six participants were run at one time on Dell desktop computers using ViewSonic 17 inch CRT monitors running in 16bit color at 1024 x 768 resolution and a refresh rate of 75 hertz. Participants sat at a testing carrel in a quiet room and were allowed to adjust their chairs and positions to be at comfortable viewing distances. The program presented participants with stimuli in a two-alternative forced-choice (2AFC) match-to-sample procedure. The following instructions (Figure 1) were given to the participants.

In this task you will be shown one picture, followed by two pictures, one of which is the same as the original picture. Your task is to identify which picture is the same as the original picture you saw by pressing Q, if it is the picture on the left, or P if it is the picture on the right.

Figure 1: Instructions

As these instructions indicate, on each trial the participant first viewed an *sample* image X (Figure 2), which was then replaced by two *option* images Q and P (Figure 3), one of which was identical to X. Their task was to decide, as quickly and accurately as possible, which option image was identical to the sample, by pressing Q if the match was on the left and P if it was on the right.

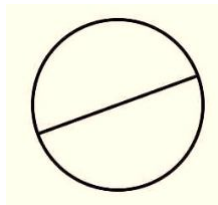


Figure 2: Example of a sample image from the Screws dataset.

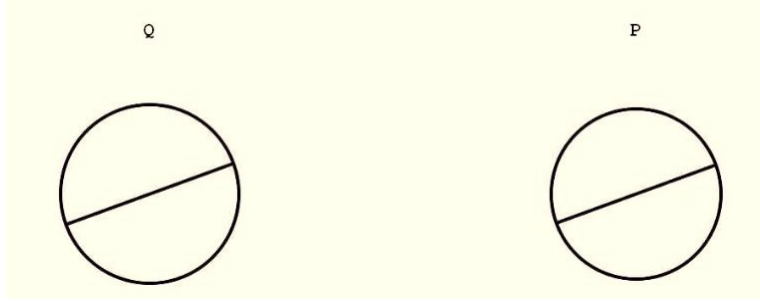


Figure 3: Example of option images for the Screws dataset.

Each trial began with a fixation point (+) in the middle of the screen presented for 1 second. This was followed by the stimulus for 1 second, followed by a mask consisting of white hash marks on a black background for 700ms. Finally, the mask was replaced by the two option images, one appearing to the left and one to the right of center, which remained onscreen until a P or Q response was detected. Following the response, screen would go blank for 700 milliseconds, after which the next trial would immediately begin. With each response the program recorded the parameters of the trial, response latency from the onset of the choice images, and accuracy. The program selected stimuli randomly without replacement from a set of predetermined pairs. For each pair, the target item and its location (left or right) were determined at random. When the participants made their judgments for all pairs, they were alerted that they had finished the task and should notify the proctor.

1.3 Data Format

There are three real datasets which encode data for three different kinds of stimuli. For the *Line* dataset, the images were horizontally-oriented lines varying in length. In the *Blobs* dataset, stimuli were renderings of the visually complex 3D hypershapes described in [2], which varied along a single dimension in a multidimensional perceptual space. In the *Screws* dataset, stimuli were bisected circles varying in radius and orientation of the bisecting line. The txt file in each folder records the experiment results. Each file includes seven columns:

1. *Subject*: an arbitrary code indicating the participant who completed the study.
2. *Pic1* (and *Pic2* below): the picture file names of the two stimuli involved in a XAB trail. The files are named with the following conventions:
 - (a) *Lines*: the lengths of line segments in pixels. (Picture files are not included as the lines were drawn by the program.)
 - (b) *Blobs*: w_PARAMETER.jpg, where PARAMETER indicates where along the 1D manifold the stimulus lies and is one of $\{-2, -1.75, -1.5, \dots, 2\}$
 - (c) *Screws*: zImg_DIAMETER_ANGLE.bmp. There are 135 pairs tested in the experiment as indicated in /stimuli/pair_list.txt)
3. *Pic2*: see above
4. *Accuracy*: 1 if the participant chose the correct answer, 0 otherwise.
5. *CorrectResponse*: the correct response for the current trial.
6. *Response*: the response chosen by the participant on the current trial.
7. *Reaction time*: Time elapsing between the onset of the option images and the detection of a response measured in ms.

2 Synthetic Datasets

There are six synthetic datasets: 1D-RBF, 1D-cos, 1D-exp, 2D-RBF, 2D-tanh, 2D-poly. Please refer the Section 4.1.1 of our paper [1] for the detailed generation process. Each dataset is saved in a .mat file. There are five matrices in each file:

1. X : column i is the coordinates of the i -th item in the original space
2. Y : column i is the coordinates of the i -th item in the “subjective” space
3. a : a_{ij} is the number of times that the participants were incorrect on the pair (x_i, x_j) simulated by a random sample from a Binomial distribution.
4. b : b_{ij} is the number of times that the participants were correct on the pair (x_i, x_j) simulated by a random sample from a Binomial distribution.
5. d : d_{ij} is the subjective distance between the pair (x_i, x_j) . Computed by their coordinates in the subjective space. Note: if $a_{ij} + b_{ij} = 0$, it means that the pair (x_i, x_j) was not experimented with.

References

- [1] Jun-Ming Xu, Xiaojin Zhu, and Timothy T. Rogers. Metric learning for estimating psychological similarities. *ACM Transactions on Intelligent Systems and Technology (ACM TIST)*, to appear, 2012.
- [2] Xiaojin Zhu, Timothy Rogers, Ruichen Qian, and Chuck Kalish. Humans perform semi-supervised classification too. In *Proceedings of the Twenty-Second Conference on Artificial Intelligence (AAAI-07)*, volume 22, page 864, 2007.