Do human prototypicality ratings correlate with neural network categorization?

Akif Berber, Lisa Goerke, Arianne van de Griend, Germonda Mooij, Ralitsa Spasova, Kai Standvoß

Abstract

This study rocks

Introduction

Individual

Methods

Neural network study

500 Images of each category, 20 categories, 1000 categorylabels (ImageNet) Weights from [**?**] For all 10,000 images the percentage of each of the 1000 labels is computed.

Baseline study

Human study

For the human experiment, a questionnaire is designed to measure human prototypicality ratings. The questionnaire consists of an introduction informing the participants about their rights and asking for their consent. Following this, the participants' demographic data, including age, gender and home-country, is accessed and participants are asked, if their color vision is impaired. Two example questions are presented to familiarize participants with the task. The images shown do belong to a category (envelopes) which is not presented in the actual study. For the actual study, 11 categories are selected (TODO based on what) as shown in table 1. For each category, all images are sorted after the classifications made by the neural network, and 10 images are chosen evenly over this distribution. The resulting range of network classification percentages can also be found in table 1. The 110 images are then randomized. Participants are asked for every image to rate its typicality on a Likert scale. In a Likert Scale participants give a quantitative value to a question based on a certain dimension, mostly the level of agreement/disagreement to that statement or question. In this questionnaire, the level of typicality was accessed. There are 7 answer options from the number one, representing "less typical", to seven, representing "very typical". The questionnaire ends with the possibility to give a free text comment about the survey.

To avoid learning effects, three differently randomized versions of the questionnaire are created and evenly distributed among participants (randomization 1: 35 participants, randomization 2: 13 participants, randomization 3: 26 participants)

3*Category	Minimum percentage network	Maximum percentage network
Volcano	0.0001	100
House	0.1	81
Airplane	0.1	95
Car	1	93
Coffee mug	1	98
Teapot	3	97
Table	13	96
Church	15	78
Castle	31	100
Fruit	55	99
Dog	93	100

Table 1. **TODO**

Results

Neural network study

Baseline study

Human study

For the human experiment, the designed questionnaire is distributed among various target groups. In total, 74 participants (35 male, 38, female, 1 other) answer to the questionnaire. Participant's age ranges from 12 to 56 ($\mu = 27$, $\sigma^2 = 9.6$).

3*Category	Pearson's correlation coefficient	3*p-values
Fruit	0.6586	**0.0384**
Church	0.1018	0.7795
Dog	0.2868	0.4218
Houses	0.5783	0.0799
Teapot	0.3454	0.3284
Table	0.0652	0.858
Airplane	-0.2698	0.451
Coffee mug	0.3791	0.28
Volcano	-0.3819	0.2762
Castle	0.1159	0.7499
Car	0.3516	0.3190

Table 2. **TODO**

[width=]fruit$_h ist.pdf$
(a)
t

Figure 1. **TODO**

[width=]volcano$_h ist.pdf$
(a)
t

Figure 2. **TODO**

[width=]table$_h ist.pdf$
(a)
t

Figure 3. **TODO**

Figure 4. **TODO**

Discussion

Individual

Conclusions

Individual

References