Co-occurrence statistics from vision and language capture thematic relationships between objects

Elizabeth H Hall, Joy J Geng

Department of Psychology, UC Davis ; Center for Mind and Brain, UC Davis

Contact: ehhall@ucdavis.edu
jjgeng@ucdavis.edu

UC DAVIS Center for Mind and Brain

Introduction

How do people think about themes or events? How should we model the relationship between objects found in a theme?

People will automatically attend to *thematically related objects*, even if they are task irrelevant [1].

Semantic models have explained variance in response in fMRI [2] and MEG studies [3].

Methods

Group 1 N = 72

For a given prompt…

Imagine you're at a ***child's birthday party***. Name 10 unique objects you might see.

24 prompts x 20 raters

amazon mechanical turk Artificial Artificial Intelligence

Averaged to find 10 most frequent items

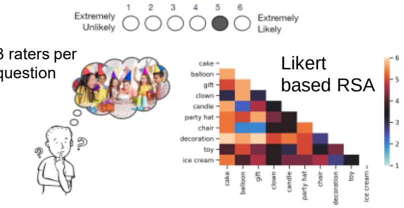
Group 2 N = 101

For each pair of the top 10 items…

How likely are you to see **cake** and **a balloon** together in the same place in the real-world?

3 raters per question

Likert based RSA

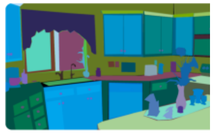


Co-occurrence models

Different training corpuses *Same unsupervised algorithm*

Newsgroup 20 [4]
18,847 articles
20 subcategories
143,802 unique nouns

CBOW
300 dimensions
1000 epochs
Negative sampling = 20



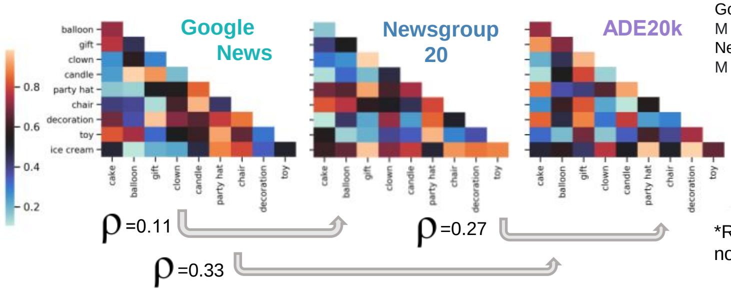
ADE20K [5]
17,820 pictures with all the objects labeled
572 scene categories
3,148 unique objects

Google News [7]
1 billion articles
~3 million unique words

What model best explains the relationship between objects?

Spearman correlation between models

For one theme: "*child's birthday party*"

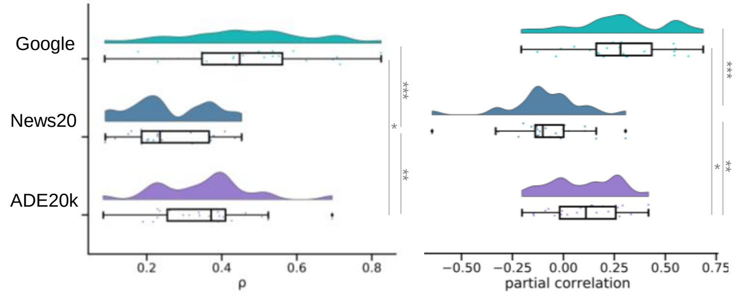


Across themes:
Google News → News 20
M = 0.27 , StDev = 0.11
Google News → ADE20k
M = 0.35 , StDev = 0.14
News 20 → ADE20k
M = 0.34 , StDev = 0.10

*RSAs are rank normalized

Model explained variance in human ratings

Spearman rank correlation *Partial correlation*



Google News model explains the most variance in human ratings followed by ADE20k

But ADE20K does significantly better than size-matched Newsgroup 20

Conclusions

We should be thinking about using pictures to model the co-occurrence of objects in scenes

but ….

larger corpuses still perform the best at capturing object relationships.

[1] Malcolm, Rattinger, and Shomstein, 2016
[2] Bankson et al., 2018
[3] Groen et al., 2018
[4] Lang, 1995
[5] Zhou et al., 2017
[6] Mikolov et al., 2013