DS-03

On Quantizing the Mental Image of Concepts for Visual Semantic Analyses

Marc A. Kastner (Nagoya University) kastnerm@murase.is.i.nagoya-u.ac.jp
https://www.marc-kastner.com/

Background

Semantic gap problems

- Missing information between computer representation and human perception
- Often an issue in word choice problems and resulting in *unnatural* results

Psycholinguistics looks at perception of words[1]

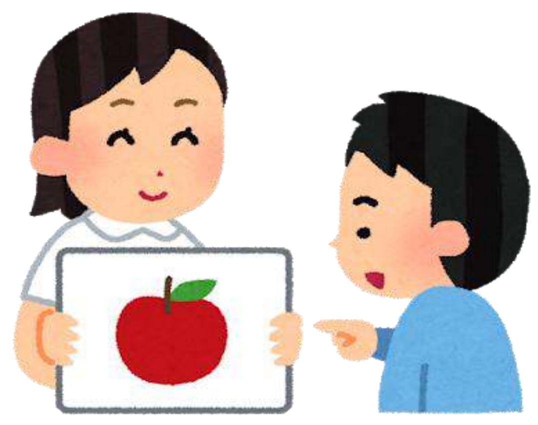
- Up to nine different measures per word …
- … but dataset creation is manual and labor intensive

In my doctoral studies I use the mental image of concepts for multimedia modeling.

Core ideas

Try to quantize semantic gap before solving it

- Use visual data mining to estimate variety differences across different datasets
- Estimate perception of concepts without manual labor needed



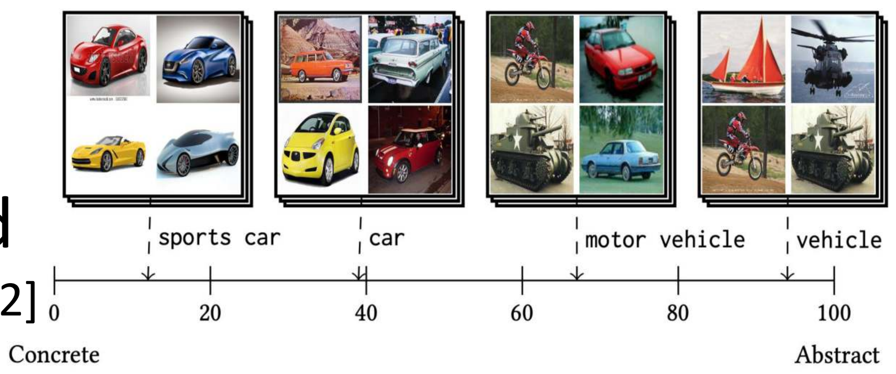
Applications

- Word choice problems like retrieval or tagging
- Increase vocabulary of psycholinguistics dictionaries

Visual variety (Topic 1)

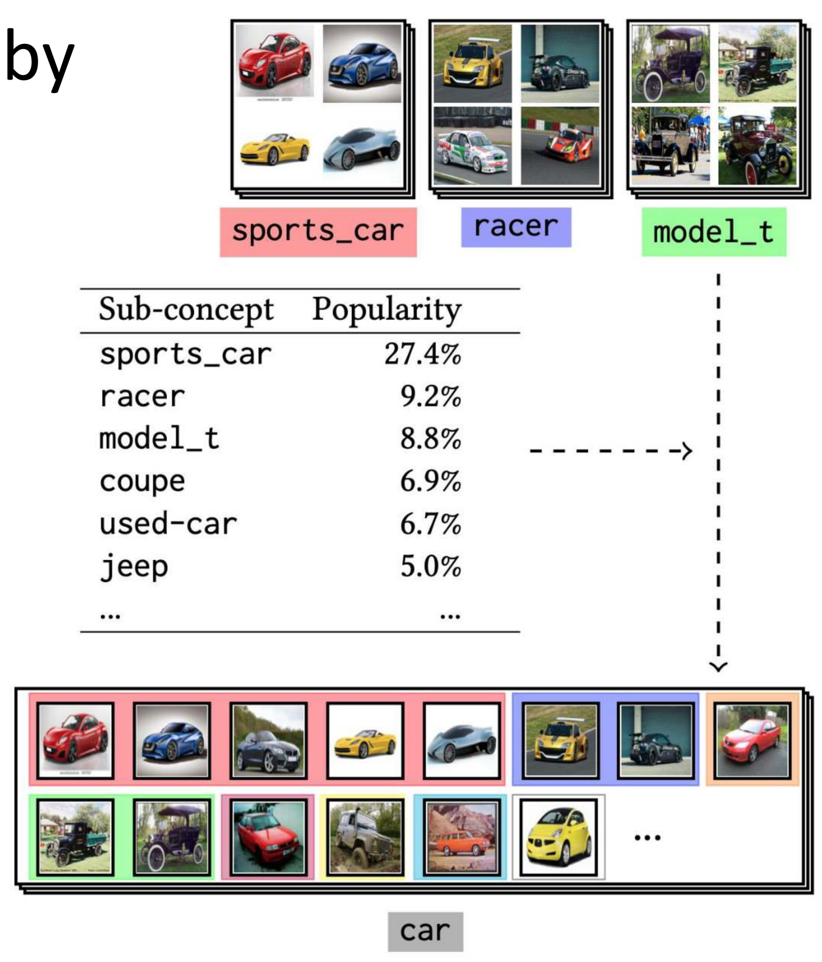
Idea: Data mine visual features to quantize feature variety across related words

- E.g. Compare variety of *car* vs. *sports car*
- Analyses quickly showed bias in existing datasets[2]



Proposed method: Improve dataset by recomposing existing datasets[2]

- Create hypernym datasets by combining its hyponyms
- Use *popularity* measure to determine ratio
- Popularity: #results for Google Image Search



Sub-concept	Popularity
sports_car	27.4%
racer	9.2%
model_t	8.8%
coupe	6.9%
used-car	6.7%
jeep	5.0%
…	…

Lastly, cluster visual features across datasets using Mean-Shift

- Re-composition removes bias!

Corpus	Correlation (1 = best)	MSE (0 = best)
Plain ImageNet (Baseline)	0.25	10.54
Equal weighting (Comparative)	0.62	9.23
Popularity weighting (Proposed)	**0.73**	**9.01**

Imageability (Topic 2)

Idea: Apply idea of visual variety on the concept of Imageability

- Concept coming from Psycholinguistics[1]
- Score words from 1 (unimageable) to 7 (imageable)

Regress imageability scores for words using visual data analysis similar to visual variety

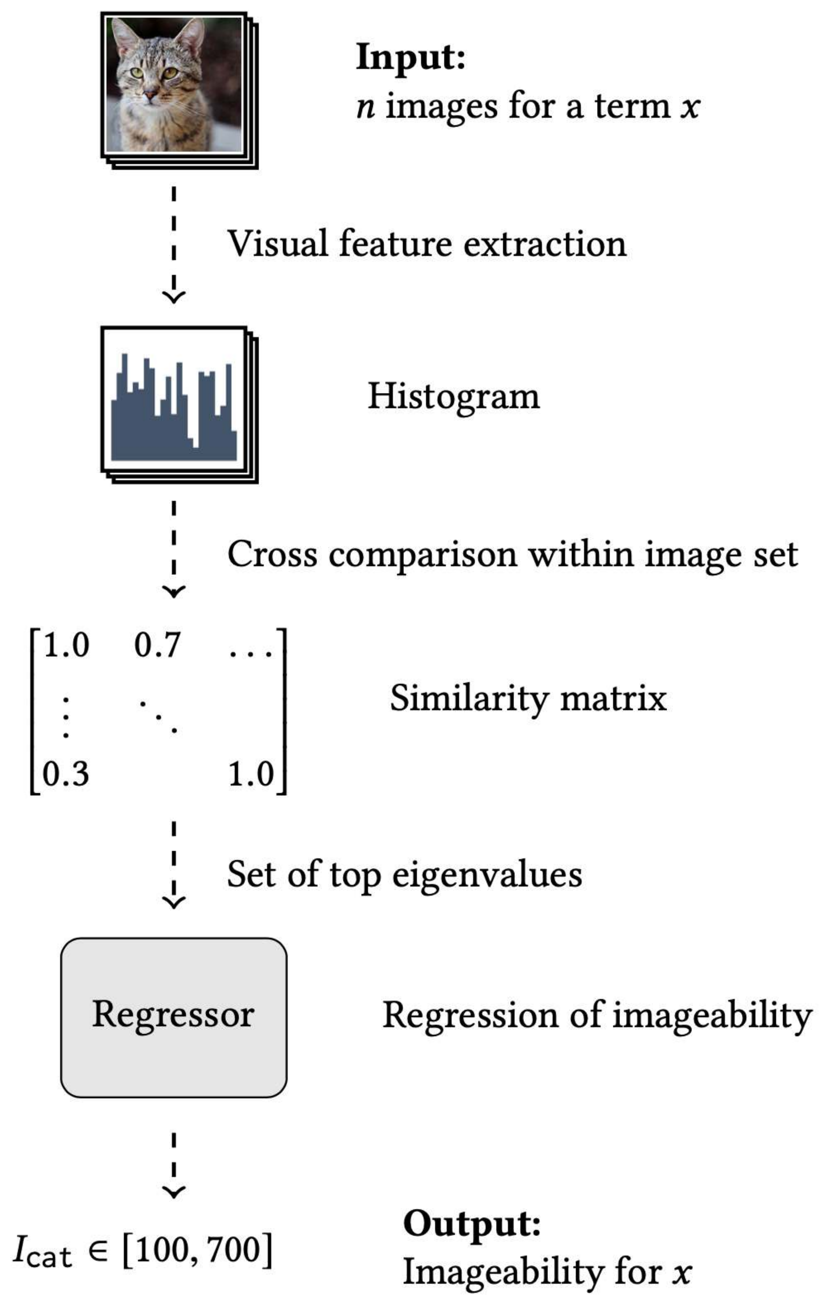
Proposed method: Gain visual information from mixture of low- and high-level features

- **Low**: Patterns, Shapes, Colors
- **High**: Objects, Concepts
- Train network based on these

Datasets

- 586 words with ground-truth imageability scores[3]
- 5,000 images per word crawled from Flickr[4]

Feature	Correlation (1 = best)	MAE (0 = best)
L1: Color histograms	0.53	11.30
L2: SURF + Bag of Words	0.54	11.48
L3: GIST	0.42	12.05
H1: Image theme (YFCC100M-based)	0.62	10.19
H2: Image content (YOLO-based)	0.43	12.55
H3: Image composition (YOLO-based)	0.25	13.98
Combined (Proposed method)	**0.63**	**10.14**
Local visual variety approach [3]	-0.01	67.31

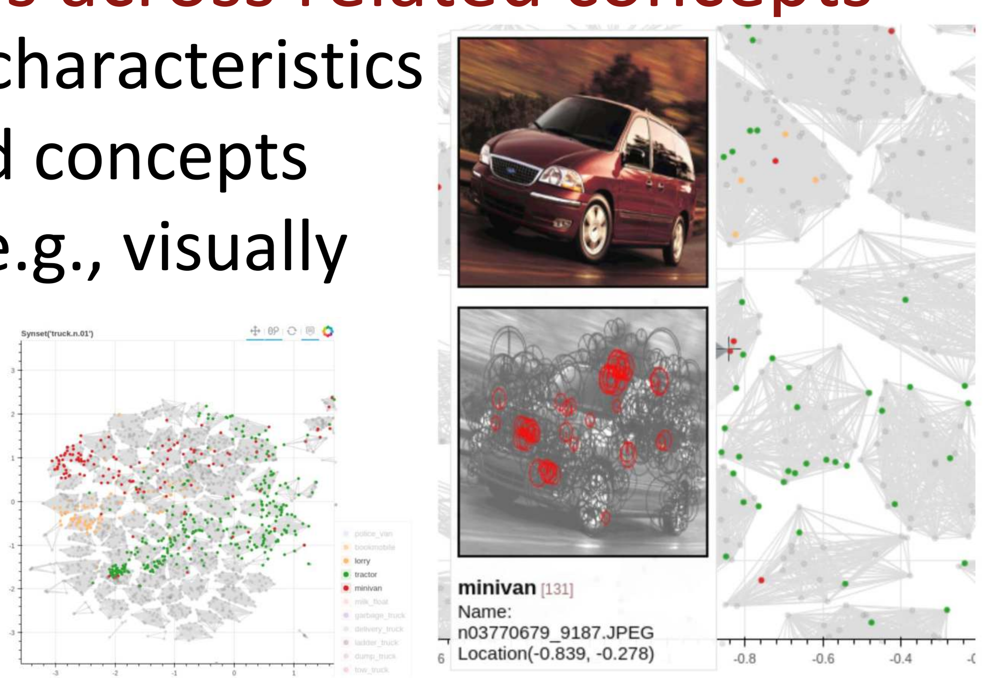


Visualizations (Topic 3)

Side projects to visualize datasets in Topics 1 & 2

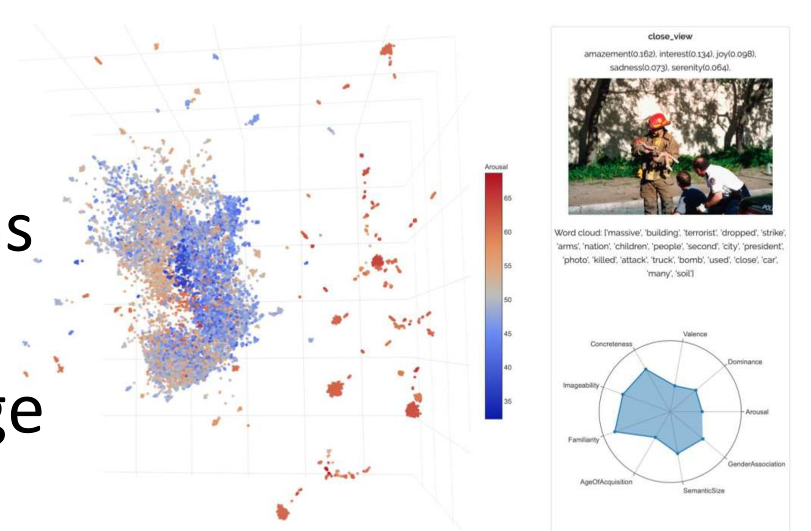
Visualize BoVW models across related concepts

- Highlight shared visual characteristics across images of related concepts
- Find out which region, e.g., visually *"makes a truck a truck"*



Browsing Visual Sentiment Datasets using Psycholinguistic Groundings[5]

- Show relationship between psycholinguistics features in textual annotations and sentiment annotations
- Use text to calculate per-image sentiment ratings



[1] Paivio et al. Concreteness, imageability, and meaningfulness values for 925 nouns. Behav Res Meth 1968
[2] Deng et al. ImageNet: A Large-Scale Hierarchical Image Database. CVPR 2009
[3] Cortese. Imageability ratings for 3,000 monosyllabic words. Behav Res Meth 2004
[4] Thomee et al. YFCC100M: The New Data in Multimedia Research. Commun ACM 2016
[5] Kastner et al. Browsing Visual Sentiment Datasets using Psycholinguistic Groundings. MMM 2020