A preliminary study on estimating word imageability labels using Web image data mining

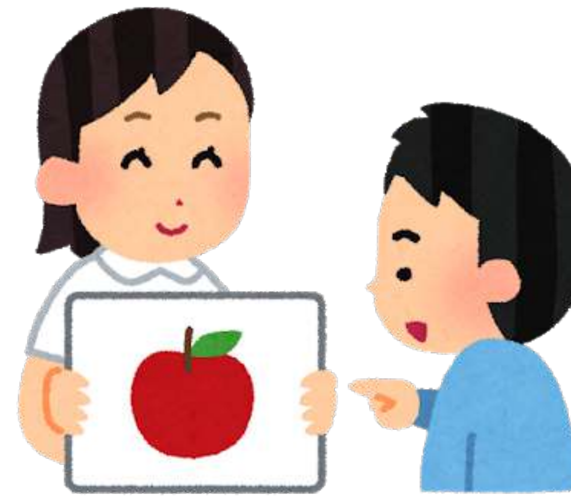
Marc A. Kastner, Ichiro Ide, Yasutomo Kawanishi
Takatsugu Hirayama, Daisuke Deguchi, Hiroshi Murase

Nagoya University

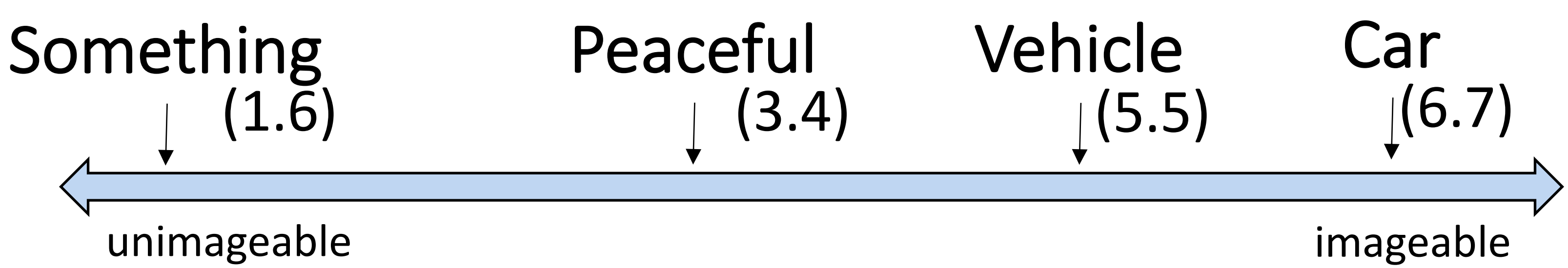
March 14, 2019

心像性

Imageability of words



- Concept from Psycholinguistics [1]
  - Quantize the perception of words
  - Often described on Likert scales
    - *Unimageable <-> Imageable* or *Abstract <-> Concrete*
  - Is a concept imageable? Do you have a mental image when thinking of a concept?

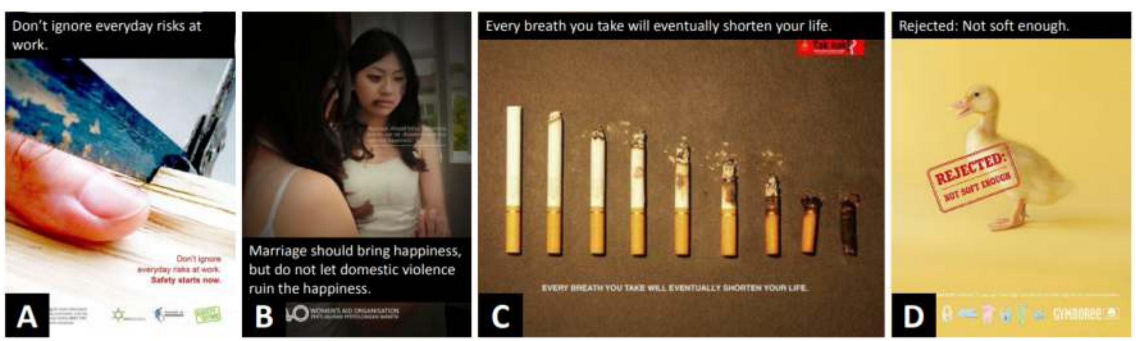


1: Pavio et al. Concreteness, imagery, and meaningfulness values for 925 nouns. J Exp Psych 1968

Applications of imageability

- Imageability is used in Psycholinguistics research
  - Influences how children learn and use grammar [2]

- Multi-modal approaches using text + image
  - Analyzing relationship of slogan and image for advertisements [3]



2: Smolik et al. The power of imageability: How the acquisition of inflected forms is facilitated in highly imageable verbs and nouns in Czech children. J First Lang 2015

3: Zhang et al. Equal But Not The Same: Understanding the implicit relationship between persuasive images and text. BMVC 2018

Motivation

- There are existing imageability dictionaries for English, Japanese and some other languages
  - Datasets are small, only for a few thousand words
  - Most dictionaries are created by hand
    - Extension is very labor intensive
    - Data often republished or reshuffled, but rarely increased

- Idea: **Estimate the imageability scores** to extend existing dictionaries by data-mining

Why use images?

- Imageability: How an average person imagines concepts (mental image)
- Social media: How common people self-annotate their perceived world by uploading images

- Core assumption
  - Relationship between imageability of words and visual characteristics of crowd-sourced images from social media

Crowd-sourced images from people
=> Average mental image

Purpose

- Estimate an imageability score for a word based on its visual characteristics
  - Mine image-data crawled for each word
  - Train regression model to estimate score based on visual features

Input:
Images for “leaf”

Data mining

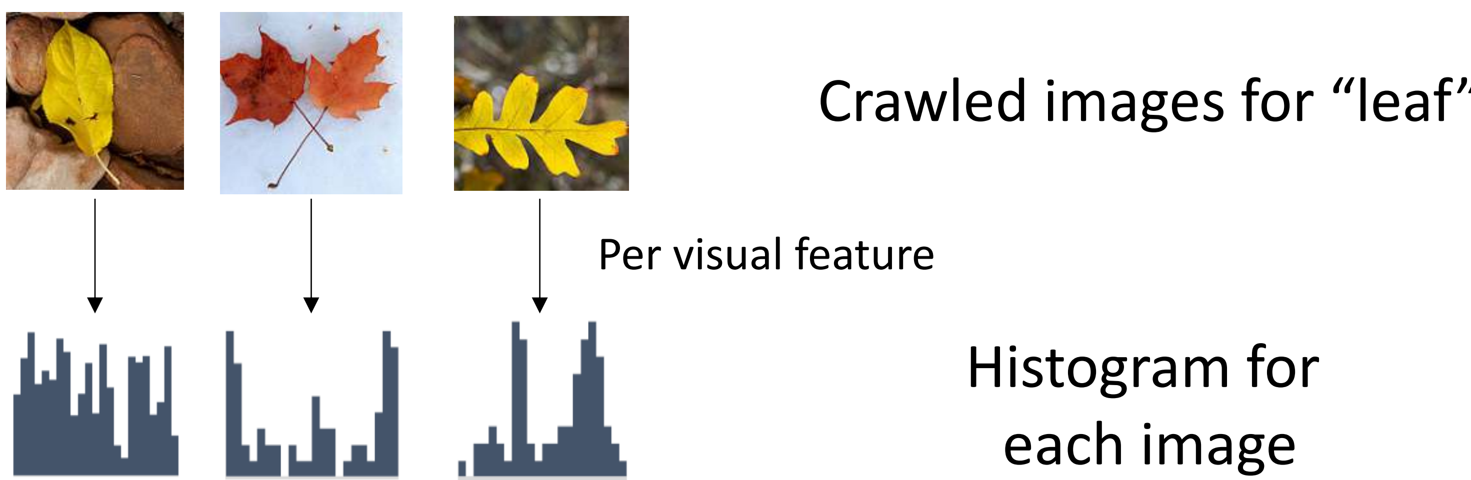
$I_{\mathrm{leaf}} \in [1,7]$

Output:
Imageability score for “leaf”

Approach

Extracting visual features

- For each word, crawl image data from social media
- Then, extract visual features from each image
  - E.g. Color histograms, Bag-of-Visual-Words histograms, ...



Approach

Cross comparison of images

- Cross-compare all images of same word
- Create similarity matrix containing similarity between all image pairs

Using visual feature histograms

Cross comparison between all images for “leaf” (using histogram similarity)

$$\begin{bmatrix} 1.0 & 0.3 & \cdots \\ \vdots & \ddots & \vdots \\ 0.7 & \cdots & 1.0 \end{bmatrix}$$

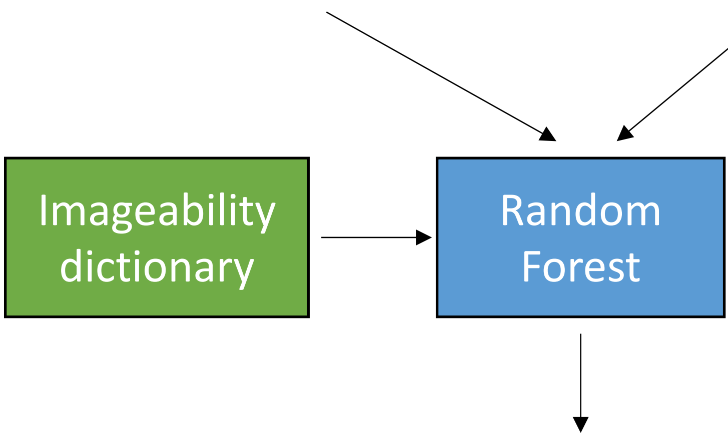
Similarity matrix per visual feature

Approach
Regression model

- Random forest based on visual characteristics
  - Train model on eigenvalues of similarity matrix
  - Use imageability dictionary as ground-truth

$$\begin{bmatrix} 1.0 & 0.5 & \cdots \\ \vdots & \ddots & \vdots \\ 0.1 & \cdots & 1.0 \end{bmatrix} \qquad \begin{bmatrix} 1.0 & 0.3 & \cdots \\ \vdots & \ddots & \vdots \\ 0.7 & \cdots & 1.0 \end{bmatrix}$$

Similarity matrix
for each visual feature



$$I_{\text{leaf}} \in [1,7]$$

Output:
Imageability for “leaf”

Input:
Images for “leaf”

For each visual feature

Histograms

Cross comparison between all images for “leaf”

$$\begin{bmatrix} 1.0 & 0.3 & \cdots \\ \vdots & \ddots & \vdots \\ 0.7 & \cdots & 1.0 \end{bmatrix}$$

Similarity matrix

Train on eigenvalues

Imageability dictionary → Random Forest

Regressor

$I_{\text{leaf}} \in [1,7]$

Output:
Imageability for “leaf”

Experiment

- Objective: Predict imageability for a set of words

- Using dataset of 577 words and 5,000 images each
  - Training: 462 words, Testing: 115 words

- Evaluation metrics
  - Mean Average Error
  - Pearson Correlation

Experiment: Datasets

- Imageability dictionary for ground-truth
  - Merged from [4] + [5]
  - Score from 1.0 (unimageable) to 7.0 (imageable)

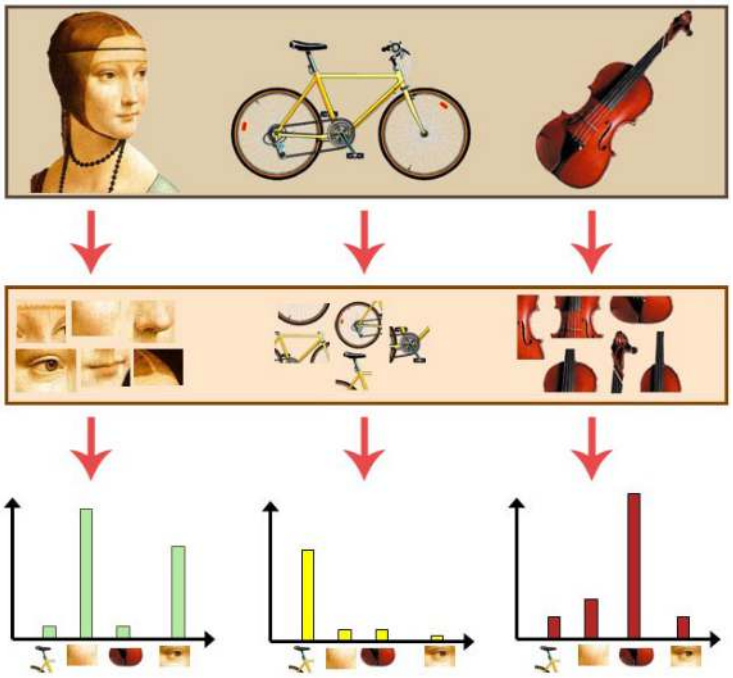
- Image dataset
  - Using YFCC100M [6] data (based on social media Flickr)
  - Crawled all images where a word from dictionary appears in meta data (such as title, description, tags)

4: Cortese et al. Imageability ratings for 3,000 monosyllabic words. Behav Res Method 2004
5: Reilly et al. Formal distinctiveness of high- and low-imageability nouns: analyses and theoretical implications. Cogn Sci 2007
6: Thomee et al. YFCC100M: The new data in multimedia research. CACM 2016

Experiment: Visual features

1. Color histogram
   - Overall color distribution based on HSV color space

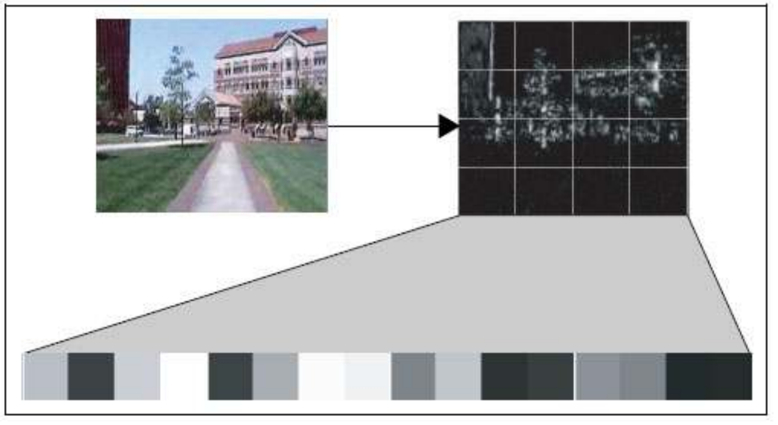
2. Bag-of-Visual-Words using SURF descriptors
   - Local feature transformation, used for object detection
   - Encoding shared patterns



https://towardsdatascience.com/bag-of-visual-words-in-a-nutshell-9ceea97ce0fb

Experiment: Visual features

3. GIST descriptors
    - Global gradients within images
    - Often used for scene analyses



http://ilab.usc.edu/siagian/Research/Gist/Gist.html

Evaluation: Results

- Estimating imageability for test data
  - Normalized to [0,100] for understandability

Feature	MAE	Correlation
(1) Color	11.78	0.56
(2) BoVW/SURF	12.53	0.55
(3) GIST	13.09	0.45
Combined	**11.68**	**0.62**

Evaluation: Examples

- A selection of low- and high- imageability words
  - Interval [100,700] for comparison with ground-truth

Type	Word	Predicted value (Ground-truth)
High imageability	coast	**5.78** (5.88)
	dusk	**5.85** (5.75)
Low imageability	doing	**3.07** (2.50)
	review	**4.22** (3.20)
Outliers	fauna	**5.35** (2.70)
	e-mail	**4.44** (6.70)

Discussion

- Tendency of imageability is correct for majority of words
- Features can complement each other to improve overall performance
- Method works better for high-imageability words
    - More abstract concepts are harder to grasp
    - More visual features are needed

Conclusion

- Proposed a method to estimate the imageability of words
  - By analyzing the visual characteristics of Web-crawled images from social media
  - Estimated imageability with an error of 11.68%

- Future work
  - Increase size of dataset
  - Use high-level features in addition to low-level features

https://www.marc-kastner.com **@mkasu**