The 22nd Meeting on Image Recognition and Understanding

On Visualizing Psycholinguistic Groundings for Sentiment Image Datasets

Marc A. KASTNER[1,a)] Ichiro IDE[1,b)] Yasutomo KAWANISHI[1,c)]
Takatsugu HIRAYAMA[1,d)] Daisuke DEGUCHI[1,e)] Hiroshi MURASE[1,f)]

Abstract

Recent multimedia applications commonly use text and imagery from Social Media for tasks related to sentiment research. As such, there are various image datasets for sentiment research for popular classification tasks. However, there has been little research regarding the relationship between the sentiment of images and its annotations from a multi-modal standpoint. In this research, we built a tool to visualize psycholinguistic groundings for a sentiment dataset. For each image, individual psycholinguistic ratings are computed from the image's metadata. A sentiment-psycholinguistic spatial embedding is computed to show a clustering of images across different classes close to human perception. Our interactive browsing tool can visualize the data in various ways, highlighting different psycholinguistic groundings with heatmaps.

1. Introduction

The use of text and imagery from Social Media for tasks related to sentiment and emotion research became ubiquitous in recent research. However, there has been little research regarding the multi-modal implications of images and its annotations related to human perception. In this demonstration, we build a tool to visualize psycholinguistic groundings for a sentiment dataset. Using this, we want to analyze the relationship between texts and images, trying to understand the groundings of human perception.

For each image, individual psycholinguistic ratings are computed from the image's textual metadata. Combined with sentiment scores, a sentiment-psycholinguistic spatial embedding is computed. It shows a clustering of sentiment images closer to human perception. Based on this, we create an interactive browsing tool, which can visualize the data in various ways. The tool allows to highlight different psycholinguistic ratings in heatmaps separately, as well as understand the structure of different datasets based on their ontology.

In Section 2 we briefly overview related research. Section 3 then discusses the idea of combining the sentiment scores of a given dataset with psycholinguistic groundings from the image metadata to compute individual scores for each image. Lastly, Section 4 showcases the interactive dataset browser we built to visualize embeddings of the sentiment-psycholinguistic space, which can be filtered across different nouns and adjectives. Various color modes allow for highlighting the different sentiment and psycholinguistic ratings.

2. Related Work

The human perception of natural language is part of the field of Psycholinguistics. In the 1960s, Paivio et al. [7] analyzed the concreteness, imagery, and meaningfulness of nouns. The most recent database for psycholinguistics is published by Scott et al. [8], which provides nine psycholinguistic ratings for 5,500 words. As these values describe human perception, the scores in such databases are typically obtained through psychological experiments. However, in our previous research [3], we also successfully used data-mining techniques on visual features to estimate imageability scores for unknown words. This and further research on the visual variety of images in datasets [4] shows a connection in how humans perceive semantics of text and images.

There is various research on sentiment and emotion in multimedia applications [5], spanning visualization, datasets [2] and recognition techniquess [1]. The connection between psycholinguistics features of text and visual features in sentiment images has not been researched to the best of our knowledge.

3. Approach

In this research, we aim to build a means to analyze psycholinguistic groundings for sentiment image datasets. As a first step, a sentiment dataset having a large number of images annotated with sentiment scores and adjective-noun pairs is retrieved. Using the metadata attached to an individual image, nine psycholinguistic scores are computed for each image. Lastly, a set of spatial embeddings based on each individual images' sentiment-psycholinguistic scores are computed for each noun, adjective and adjective-noun pair, respectively.

1 Nagoya University
a) kastnerm@murase.is.i.nagoya-u.ac.jp
b) ide@i.nagoya-u.ac.jp
c) kawanishi@i.nagoya-u.ac.jp
d) takatsugu.hirayama@nagoya-u.jp
e) ddeguchi@nagoya-u.jp
f) murase@i.nagoya-u.ac.jp

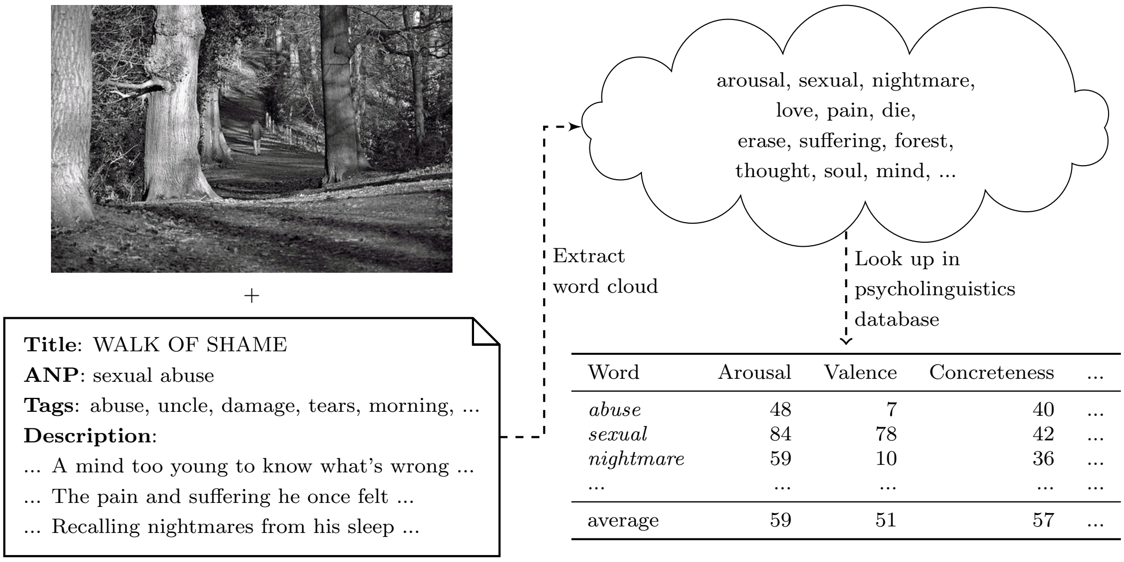


Fig. 1: The process of calculating per-image psycholinguistic scores. For each image, a word cloud is extracted from the textual metadata. For each word, the corresponding scores from the psycholinguistics database are gathered. Then, an average score weighted for word frequency is computed for each psycholinguistic rating separately. The image and metadata are courtesy of Flickr user despitestraightlines[*1].

3.1 MVSO dataset

As the baseline for the visualization tool, we use the MVSO dataset [2]. The dataset consists of seven million images, their textual metadata, and sentiment scores, collected through Flickr and crowd-sourcing. Each image is annotated with a single adjective-noun pair (ANP,) e.g. *abandonned_city* or *old_dog*, describing its sentiment. We split the ANP into two labels, *noun* and *adjective* to create a flat ontology-like structure. Using this, images related to the same noun but for different adjectives, and vice versa, can be filtered. Each ANP comes with 21 sentiment scores like *joy*, *ecstasy*, and others, which were verified by crowd-sourcing. All images with the same ANP share the same sentiment scores.

Furthermore, each image comes with textual metadata containing a title, a description text, and tags. This metadata is used in the following section to compute a psycholinguistic grounding for each image.

3.2 Per-image psycholinguistic scores

To create an embedding with a meaningful spatial distribution per image, individual scores for each image are needed. We compute a psycholinguistic grounding of the textual metadata for each image. Scott et al. [8] provide a psycholinguistics dataset with nine ratings each for 5,500 words. The nine ratings available are: arousal, dominance, valence, imageability, concreteness, familiarity, semantic size, age of acquisition, and gender association.

For each image, we extract the title, description, and tags from the MVSO dataset. All these data are provided by the image uploader, which makes them noisy. We generate a word cloud from all words used in the metadata, stripping grammatical affixes through lemmatization. Furthermore, all words not contained in the psycholinguistics database are filtered out. Note, that stop-words are implicitly skipped, as they are not part of the database. Lastly, we compute nine psycholinguistic ratings by averaging the corresponding scores for each word in the word cloud. The process of calculating per-image psycholinguistic scores is shown in Fig. 1.

To filter out some of the noise, and also implicitly remove most images with short identical titles and such, all images with less than twenty words in their word cloud are filtered out. This results in approximately 400,000 images with nine individual psycholinguistic ratings each.

3.3 Spatial embedding

For each noun, adjective, and ANP, we compute a spatial sentiment-psycholinguistic embedding using UMAP [6]. As input, we use a 31-dimensional vector for each image, using the 22 sentiment scores of its ANP as well as the nine psycholinguistic ratings calculated through the metadata.

Lastly, an embedding using a sample of images from the whole MVSO dataset is computed. In this sample, we focus on the images with the most extreme scorings for each rating.

4. Visualization

To visualize the relationship between human sentiment ratings of an image and the psycholinguistic characteristics of words used in the image metadata, we built a dataset browser with an interactive interface. Using this tool, it is possible to browse the MVSO dataset, filter it for different adjective or nouns, and see the scoring for different images. A three-dimensional view shows the sentiment-

[*1] http://flickr.com/photos/despitestraightlines/6677983565/

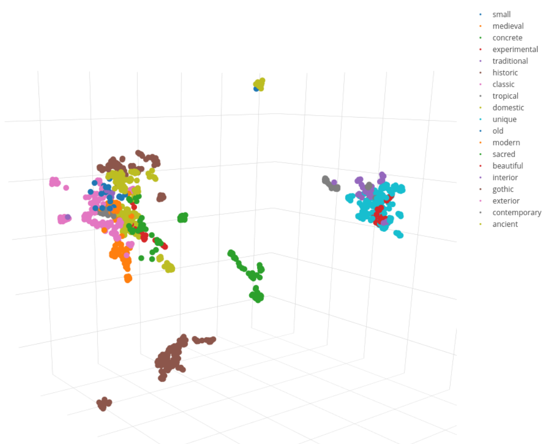


Fig. 2: Ontology-based colors. Using the ontology-based color option, classes inside the currently shown data are color-graded. Here, the spatial embedding for the noun *architecture* is shown, where each color describes a different ANP like *small_architecture* or *medieval_architecture*.

psycholinguistic spatial embedding of the selected dataset. Different color modes allow for analyzing the dataset regarding its ontology and human perception scores established in Section 3. The full user interface of the proposed tool is shown in Fig. 3.

4.1 Spatial embedding view

The sentiment-psycholinguistic space is shown with an interactive three-dimensional interface allowing for zooming and panning. Each data point represents one image from the MVSO dataset plotted on a three-dimensional embedding based on its sentiment and psycholinguistic scores. The user can switch between sampling a selection of images across the whole dataset, or showing all images of a selected noun, adjective, or ANP.

The color displayed in the spatial embedding can be selected to either show scores related to human perception as heatmap-based color gradings, or highlight the ontology-based class labels (e.g., *different adjectives* for a filtered *noun* dataset.)

Figure 2 shows an example of per-class color mode for the dataset *architecture*. If the selected dataset is an adjective, the images for different nouns can be displayed in separate colors, and vice-versa.

Alternatively, the user can switch between highlighting one specific rating as a heatmap-like grading from images with low values (blue) to high values (red.) This is available for each psycholinguistic and sentiment rating. In this mode, it is possible to see which cluster in the spatial embedding is influenced by which rating. Figure 4 shows the visualized spatial embedding for images of the whole dataset highlighting two different psycholinguistic ratings — arousal and concreteness.

4.2 Detailed view

When selecting a data sample in the spatial embedding, a detailed view opens on the right. Here, one can see the actual image behind the sample, as well as some of its metadata related to the sentiment score. A table shows the computed psycholinguistic values, as well as its highest and lowest significant words for each rating.

5. Conclusion

In this demonstation, we introduced a tool to visualize sentiment image datasets regarding their psycholinguistic grounding. For each image, nine psycholinguistic scores are computed using image metadata like title, description, and tags. Using a combination of sentiment scores and psycholinguistic scores, a spatial embedding is computed to visualize their relationship. The interactive tool can browse the MVSO dataset, either wholly or by filtering it for nouns, adjectives, or ANPs. The three-dimensional plotting of the spatial embedding gives further insights on how images for the same noun or images form different clusters regarding their human perception. Different color modes can be used to either highlight a single sentiment or psycholinguistic rating, or visualize the ontology of the dataset.

In future work, we plan to use this tool to analyze the clustering shown in the tool and compare the visual characteristics of different clusters. Furthermore, the use of visual information to detect per-image sentiment scores could give additional insights on the perception of individual images. Lastly, as the MVSO datasets includes data in multiple languages, we want to extend the browser to work across multiple languages.

Acknowledgments

Parts of this research were supported by the MEXT, JSPS KAKENHI 16H02846, and a joint research project with NII, Japan.

References

[1] Jindal, S. and Singh, S.: Image sentiment analysis using deep convolutional neural networks with domain specific fine tuning, *ICIP 2015*, pp. 447–451 (2015).

[2] Jou, B., Chen, T., Pappas, N., Redi, M., Topkara, M. and Chang, S.: Visual affect around the world: A large-scale multilingual visual sentiment ontology, *ArXiv e-prints 1508.03868* (2015).

[3] Kastner, M. A., Ide, I., Kawanishi, Y., Hirayama, T., Deguchi, D. and Murase, H.: A preliminary study on estimating word imageability labels using Web image data mining, *ANLP, Japan*, No. A4–7, pp. 747–750 (2019).

[4] Kastner, M. A., Ide, I., Kawanishi, Y., Hirayama, T., Deguchi, D. and Murase, H.: Estimating the visual variety of concepts by referring to Web popularity, *Multimed Tools Appl*, Vol. 78, No. 7, pp. 9463–9488 (2019).

[5] Kim, E. and Klinger, R.: A survey on sentiment and emotion analysis for computational literary studies, *ArXiv e-prints 1808.03137* (2018).

[6] McInnes, L. and Healy, J.: UMAP: Uniform Manifold Approximation and Projection for dimension reduction, *ArXiv e-prints 1802.03426* (2018).

[7] Paivio, A., Yuille, J. C. and Madigan, S. A.: Concreteness, imagery, and meaningfulness values for 925 nouns, *J Exp Psychol*, Vol. 76, No. 1, pp. 1–25 (1968).

[8] Scott, G. G., Keitel, A., Becirspahic, M., Yao, B. and Sereno, S. C.: The Glasgow Norms: Ratings of 5,500 words on nine scales, *Behavior Research Methods*, Vol. 51, No. 3, pp. 1258–1270 (2019).

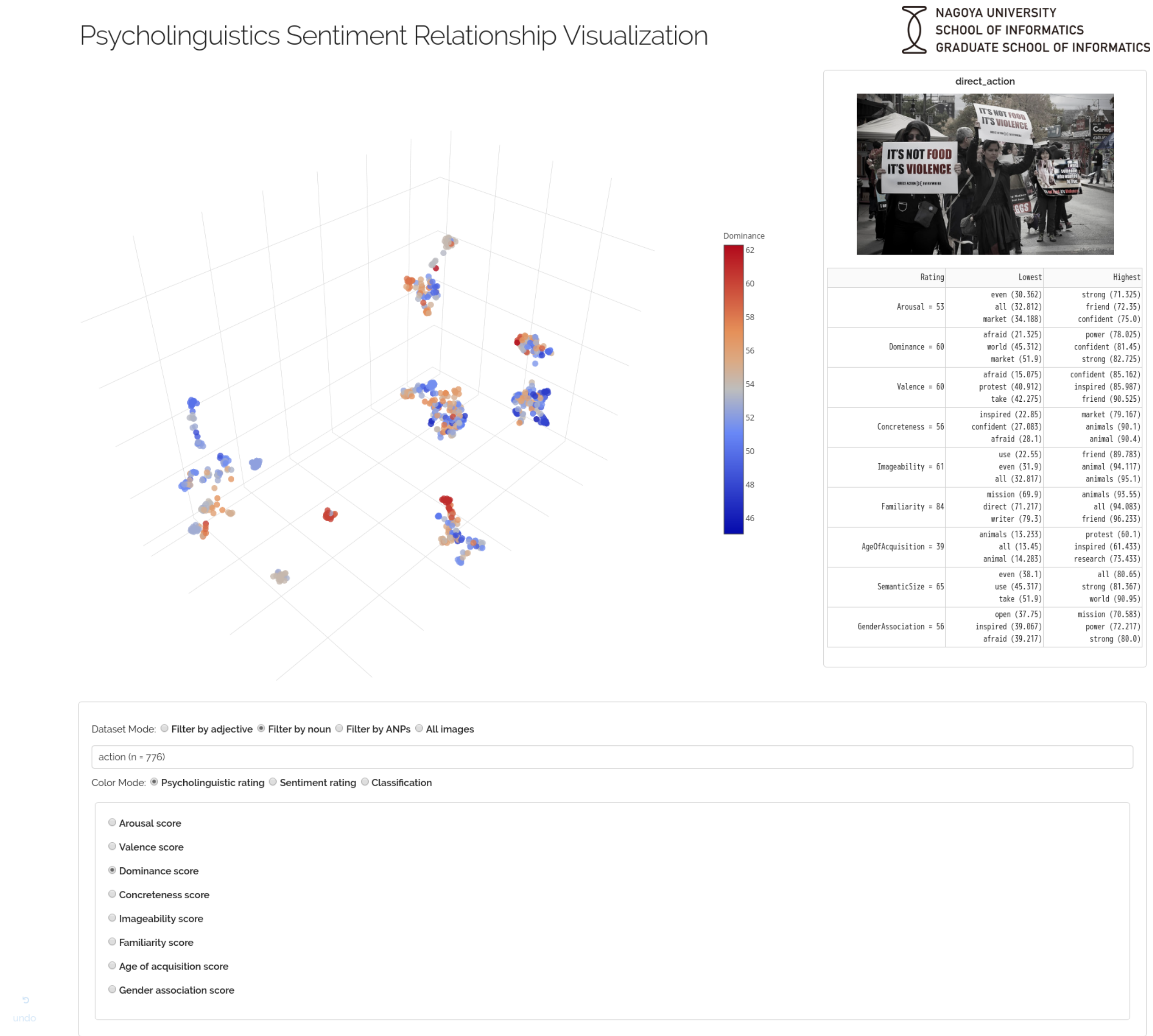
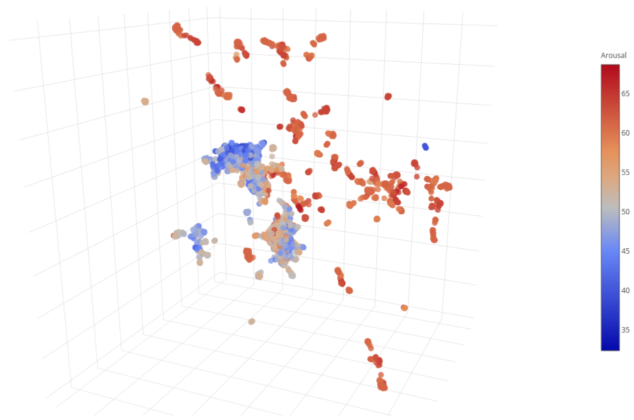
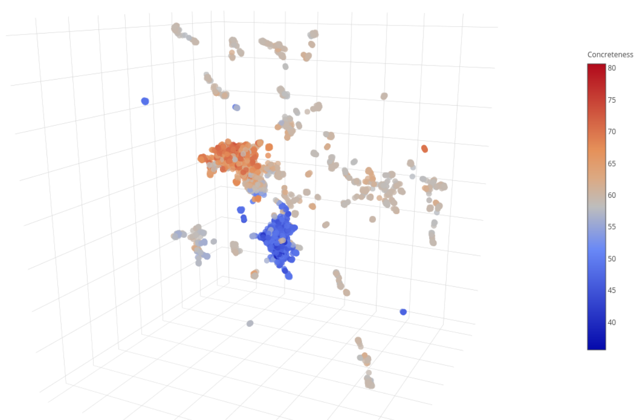


Fig. 3: The user interface of the dataset browser. The upper left screen shows a three-dimensional spatial embedding of the sentiment-psycholinguistics space for a set of images. It can be browsed interactively. Clicking on a data point shows detailed information on this image on the right. Using the settings in the bottom, the user can filter the shown images per noun, per adjective, or per ANP. The color grading can be switched between either highlighting a single psycholinguistic or sentiment score as a heatmap, or displaying ontology-based sub-classes of the shown dataset.



(a) Arousal score



(b) Concreteness score

Fig. 4: The sentiment-psycholinguistics spatial embedding for different psycholinguistics ratings. In Fig. (a), the arousal score is highlighted using a heatmap from blue (low) to red (high,) while Fig. (b) shows the corresponding concreteness score.