Browsing Visual Sentiment Datasets using Psycholinguistic Groundings

Marc A. Kastner[1][0000−0002−9193−5973], Ichiro Ide[1], Yasutomo Kawanishi[1], Takatsugu Hirayama[2], Daisuke Deguchi[3], and Hiroshi Murase[1]

[1] Graduate School of Informatics, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, 464-8601 Japan
kastnerm@murase.is.i.nagoya-u.ac.jp
{ide, kawanishi, murase}@i.nagoya-u.ac.jp

[2] Institute of Innovation for Future Society, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, 464-8601 Japan
takatsugu.hirayama@nagoya-u.jp

[3] Information Strategy Office, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, 464-8601 Japan
ddeguchi@nagoya-u.jp

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 demonstration, 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.

Keywords: Visual sentiment · Psycholinguistics · Visualization.

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 show a tool to visualize psycholinguistic groundings for a sentiment dataset. Using this, we analyze the relationship between texts and images, trying to get a better understanding of the groundings of human perception. For each image, individual psycholinguistic ratings are computed from the image's textual metadata. Combined with sentiment scores available from the used dataset, a sentiment-psycholinguistic spatial embedding is computed. It shows a distribution of sentiment images close 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. [6] analyzed the concreteness, imagery, and meaningfulness of nouns. The most recent database for psycholinguistics is published by Scott et al. [7], 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. Our previous research on the visual variety of images in datasets [3] shows a connection in how humans perceive semantics of text and images.

There is various research on sentiment and emotion in multimedia applications [4], spanning visualization, datasets [2] and recognition techniques [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 demonstration, we aim to present a means to analyze psycholinguistic groundings for sentiment image datasets. As a first step, a visual sentiment dataset having a large number of images annotated with adjective-noun pairs is retrieved. Using the textual 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.

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

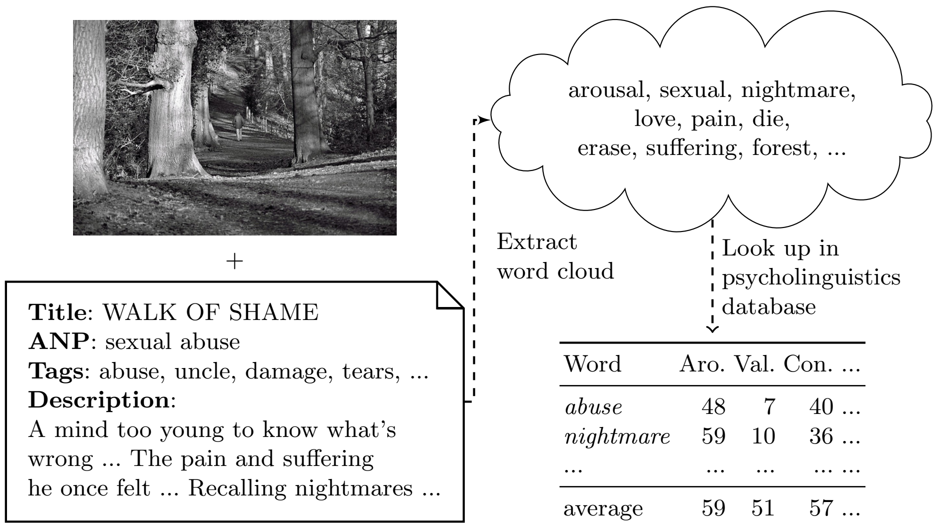


Fig. 1: The process of calculating per-image psycholinguistic scores. For each image, a word cloud is extracted from the textual metadata. Using a psycholinguistics database, an average score describing the image is computed for each psycholinguistic rating (Aro.=*Arousal*, Val.=*Valence*, Con.=*Concreteness*). The example image is courtesy of Flickr user despitestraightlines[4].

but for different adjectives, and vice versa, can be filtered. Each ANP comes with 21 sentiment scores (e.g., $joy = 0.6$, $ecstasy = 0.8$), but all images with the same ANP share the same sentiment score. Each image also comes with textual metadata containing a title, a description text, and tags. This metadata is used in the following section to compute an individual 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. [7] 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. Lastly, we compute nine psycholinguistic ratings by averaging

[4] http://flickr.com/photos/despitestraightlines/6677983565/

the corresponding scores for each word in the word-cloud. The process of calculating per-image psycholinguistic scores is shown in Fig. 1. Filtering out images where the number of words available in the psycholinguistics dictionary was not sufficient, this results in approximately 400,000 images with nine individual psycholinguistic ratings each.

For each noun, adjective, and ANP, we compute a spatial embedding using UMAP [5]. Additionally, we compute an embedding including all images, filtering for extreme cases with very high or very low scores for some psycholinguistic ratings. As input, we use a 30-dimensional vector for each image, composed of the 21 sentiment scores of its ANP as well as the nine psycholinguistic ratings calculated through the metadata.

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. Using this tool, it is possible to browse the dataset, filter it for different adjective or nouns, and see the scoring for different images. A three-dimensional view shows the sentiment-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 Figure 2.

The sentiment-psycholinguistic space is shown with an interactive 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 individual 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). The different color modes are shown in Figure 3.

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 introduce a tool to visualize sentiment image datasets regarding their psycholinguistic grounding. For each image, nine individual psycholinguistic scores are computed using textual metadata. A spatial embedding is computed to visualize their relationship of text and image. The interactive tool showcases the MVSO dataset, either wholly or by filtering it for nouns,

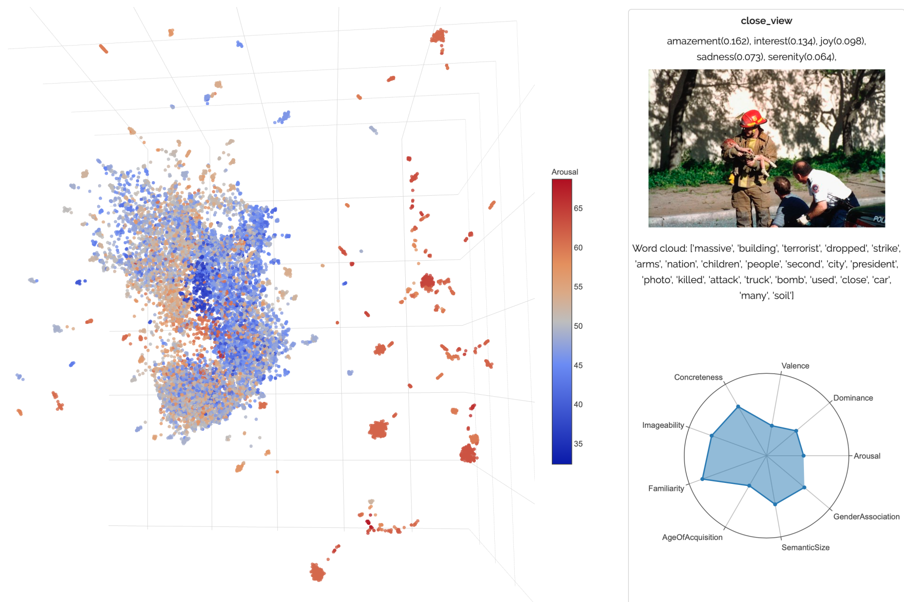


Fig. 2: The user interface of the tool. The upper left shows a three-dimensional embedding of the sentiment-psycholinguistics space. Detailed information for a selected image is shown on the right. The embedding can be filtered by noun, adjective, or ANP. Different color grading options are shown in Figure 3.

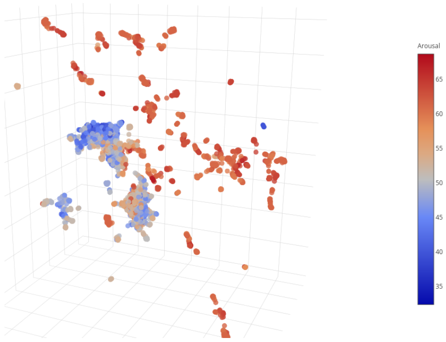
adjectives, or ANPs. The spatial embedding gives further insights on how images for the same noun form different clusters regarding their human perception. Different color modes can be used to either highlight a single sentiment or psycholinguistic rating, or visualize the ontology of the dataset. In future work, we plan to use this tool to compare the visual characteristics of different clusters.

Acknowledgments

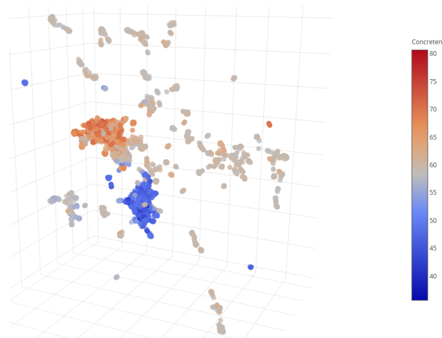
Parts of this research were supported by the MEXT, JSPS KAKENHI 16H02846, and a joint research project with NII, Japan.

References

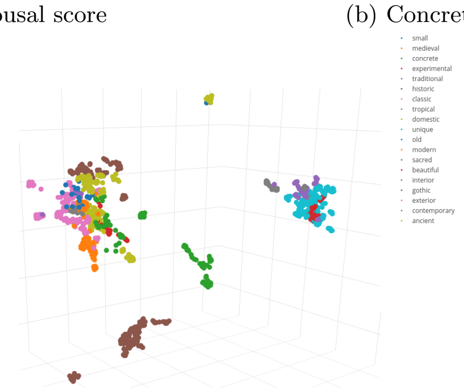
1. Jindal, S., Singh, S.: Image sentiment analysis using deep convolutional neural networks with domain specific fine tuning. In: Proc. 2015 IEEE Int. Conf. Image Process. pp. 447–451 (Dec 2015). https://doi.org/10.1109/INFOP.2015.7489424
2. Jou, B., Chen, T., Pappas, N., Redi, M., Topkara, M., Chang, S.: Visual affect around the world: A large-scale multilingual visual sentiment ontology. ArXiv e-prints 1508.03868 (2015)



(a) Arousal score



(b) Concreteness score



(c) Ontology visualization

Fig. 3: The spatial embedding can be colored in different ways. The psycholinguistic grounding can be shown in a heat map as shown in (a) and (b) highlighting each image's arousal and concreteness scores, respectively. Alternatively, the ontology of adjectives or nouns can be highlighted as shown in (c).

3. Kastner, M.A., Ide, I., Kawanishi, Y., Hirayama, T., Deguchi, D., Murase, H.: Estimating the visual variety of concepts by referring to Web popularity. Multimed Tools Appl **78**(7), 9463–9488 (Apr 2019). https://doi.org/10.1007/s11042-018-6528-x
4. Kim, E., Klinger, R.: A survey on sentiment and emotion analysis for computational literary studies. ArXiv e-prints 1808.03137 (2018)
5. McInnes, L., Healy, J.: UMAP: Uniform Manifold Approximation and Projection for dimension reduction. ArXiv e-prints 1802.03426 (Feb 2018)
6. Paivio, A., Yuille, J.C., Madigan, S.A.: Concreteness, imagery, and meaningfulness values for 925 nouns. J Exp Psychol **76**(1), 1–25 (1968)
7. Scott, G.G., Keitel, A., Becirspahic, M., Yao, B., Sereno, S.C.: The Glasgow norms: Ratings of 5,500 words on nine scales. Behav Res Meth **51**(3), 1258–1270 (Jun 2019). https://doi.org/10.3758/s13428-018-1099-3