Cognitive Science and AI

Describing Visual Stimuli in Hindi using Learnt Semantic Representations of Brain Activity

Team: AlphaZ

1 Background

Many neuroscience studies have attempted to quantitatively analyze the semantic representation of what a human recalls using the fMRI data of brain activity evoked by visual stimuli, such as natural movies and images. These try to assign semantic labels to still pictures using natural language descriptions synchronized with the pictures and discuss the relationships between the visual stimuli evoked by the still pictures and brain activity. Based on these relationships, they construct a model to classify brain activity into semantic categories to reveal areas of the brain that deal with particular semantic categories. We in this project will try to extract these learn semantic representations and elucidate their language in-specificity using our proposed framework.

2 Problem Statement

The problem statement we are trying to address is to use brain activity data as input rather than images and attempt to generate natural language descriptions for the brain data in Hindi. The project broadly aims to generate natural language sentences that describe what a human being calls to mind using brain activity data observed by fMRI as input information (Matsuo et al., 2018). However, the interesting caveat that we want to exploit is that these recollections to the mind by the human are language inspecific, i.e, captions for multiple different languages can be generated using these semantic representations even though the subject whose fMRI is being captured is natively English speaking for example.

3 Project Description

3.1 Main goal

Designing a Machine Learning Framework to predict captions in Hindi using Brain activity data on exposure to Visual Stimuli.

3.2 What task/objective will you address

- First, we will use an existing Image captioning model in Hindi trained on MSCOCO images and Hindi captions by (Mishra et al., 2021). The image captioning model consists of a Convolutional Neural Network (eg VGGNet) to generate image features from an input image (Step 1 Part 1) and an LSTM/GRU model to predict captions using these image features (Step 1 Part 2).
- Next, we will use fMRI images obtained on visual stimuli to generate the image features. This will be done by extracting image features for the original image using the CNN model (Step 1 Part 1) and then training an ML model to predict these image features (Step 2).
- Finally, the two modules will be be combined, i.e, Brain fMRI -> Image Features (Step 2) and Image Features -> Hindi Captions (Step 1 Part 2) to finally get the desired output.

3.3 What data will you use? Already existing code?

- **Data:** MSCOCO dataset with Hindi Captions was used by (Mishra et al., 2021). It is available on request and contains 84000 images with corresponding captions. fMRI dataset for visual stimuli is available for 4500 images in (Matsuo et al., 2018).
- **Code:** Code for Hindi Image captioning (Step 1) is available on Github: https://github.com/santosh1821cs03/Image_Captioning_Hindi_Language. Code for Step 2 and Step 3 are not available.

3.4 What baseline(s) will you use? How will you evaluate your results?

There is no set metric for evaluation. Previous publications used the models BLEU (Papineni et al., 2002) and METEOR (Denkowski and Lavie, 2014), which are the most commonly automatic evaluation metrics in the caption generation literatur. However, we do not have available groundtruth sentences, i.e., human generated descriptions for the images that evoked brain activities, which are necessary for computing these metrics. Thus, we will use 10-best beam search on direct image captioning (Step 1) as approximate groundtruth data.

4 Timeframe + Work Distribution (For two months)

	Task	**Distribution**	**Start and End Dates**
Phase One	Acquisition of the MSCOCO Hindi Captions and Brain Activity Datasets	All	March 11 - March 18
Phase Two	Implementing Step 1 model Design, Evaluation, and Testing	All	March 18 - April 1
Phase Three	Designing Step 2 and Step 3 models, Evaluation, Testing and Analysis	All	April 1 - April 30

5 References

1. Matsuo, Eri, et al. "Describing semantic representations of brain activity evoked by visual stimuli." 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC). IEEE, 2018.

2. Mishra, Santosh Kumar, et al. "A Hindi image caption generation framework using deep learning." Transactions on Asian and Low-Resource Language Information Processing 20.2 (2021): 1-19.