LT2318 - AI: Cognitive Systems
Project Description
Kate Viloria

Evaluation of Generated Captions on VizWiz Dataset

Problem:
How do captions generated by different attention mechanisms in image captioning models compare to captions written by humans for a domain-specific task such as caption generation for the visually impaired?

Key References:
I am hoping to use the code of models we have read and talked about in class. My initial idea is to compare the two models that differ in how attention is implemented—within the hidden state to predict image attention and a "visual sentinel" that decides whether the attention should be applied to the image or the language model (Xu et al. and Lu et al.).

Existing Resources (datasets, code, etc.):
VizWiz - A dataset comprised of 39,181 images taken by people who are visually impaired each with 5 human-made captions.

Tasks to be done:
*preliminary
- pre-process data for both models
- run existing code for both models
- evaluate through examining spatial relations in both the given and generated captions
- write paper
- report and present results

Expected result:
Given the domain-specific task and examining some of the images in the dataset, the model that is able to choose which parts of the image should be attended to might perform better. As these images have been taken by visually impaired persons, they are at times unclear and unfocused. The ability to choose what parts of the image the model should focus on would be greatly advantageous for this task. However, an argument could also be made that the second model has the ability to generate more descriptive captions since it can decide whether to pay more attention to the image or the language model.

Project Evaluation:
By extracting and classifying the spatial relations from the captions, I am hoping to evaluate the models through count and accuracy. (Birmingham & Muscat)

Expected challenges:
Finding the code and pre-processing the data for the models to successfully run it will be a challenge. It will also be difficult figuring out the best way to evaluate the data and implementing it.

Ethical issues:
This dataset was created through crowdsourcing using Amazon's Mechanical Turk. Apart from the exploitation of workers, crowdsourcing can also present a multitude of problems in terms of the data contributions of the workers. From skimming through random parts of the dataset, some captions and answers to visual questions provided by the annotators are sometimes irrelevant or do not suffice.