ANLY 580 Project Proposal

1. **Group Members:**

Jieqian Liu, Jieyi Sun, Zidong Xu, Minglei Cai

1. **Topic & Background:**

***Open-domain Dialogue Systems:***

Popular approaches for dialogue systems include Sequence-to-Sequence (Seq2Seq) response generation, knowledge graph-driven question answering (KG-QA) , context-sensitive response retrieval and RL methods.

In this project, we will focus on fine-tuning pre-trained seq2seq models. Depending on the time, we will decide whether or not to have trials on the other ones.

1. **Datasets:**

We primarily focus on the popular Ubuntu Dialogue Corpus.

Kaggle link: <https://www.kaggle.com/datasets/rtatman/ubuntu-dialogue-corpus>

In addition, we might use some extra datasets for model-validation in the following links:

<https://www.kaggle.com/code/luisjimenezquesada/nlp-with-stackoverflow>

https://www.kaggle.com/datasets/veeralakrishna/questionanswer-combination

1. **Methods & Main Procedures:**
2. Data Cleaning:

For schemas used for data cleaning of Ubuntu Dialogue Corpus, we will refer to the following paper:

https://arxiv.org/pdf/1506.08909.pdf

1. Loading Pre-trained Models and Fine-tuning

We basically use the pre-trained transformers from *huggingface*:

* DialoGPT-medium, DialoGPT-large
* Blenderbot-400M-distill, blenderbot-1B-distill, etc.

While there are ways of fine-tuning, we basically use the method of vanilla fine-tuning.

1. Model Evaluation

While human evaluations in certain metrics are quite common for evaluating the performance of dialogue systems, we don’t have access to the experts nor specific human resources for evaluations. We will do the human evaluation ourselves.

1. UI & Implementation

We will implement the UI/GUI and deploy to the io domain.

1. **Unknowns:**
2. Schemas used to extract question-response pairs from the original corpus might lead to loss of dataset volume.
3. Values of hyperparameters used for the training process.
4. Uncertainties existing in human evaluations.
5. **Presentation:**

Live demo in website + presentation.

1. **References:**
2. Zhang, Y., Sun, S., Galley, M., Chen, Y.-C., Brockett, C., Gao, X., Gao, J., Liu, J., & Dolan, B. (2020, May 2). *Dialogpt: Large-scale generative pre-training for conversational response generation*. arXiv.org. Retrieved November 8, 2022, from https://arxiv.org/abs/1911.00536
3. Roller, S., Dinan, E., Goyal, N., Ju, D., Williamson, M., Liu, Y., Xu, J., Ott, M., Shuster, K., Smith, E. M., Boureau, Y.-L., & Weston, J. (2020, April 30). *Recipes for building an open-domain chatbot*. arXiv.org. Retrieved November 8, 2022, from https://arxiv.org/abs/2004.13637
4. Lowe, R., Pow, N., Serban, I., & Pineau, J. (2016, February 4). *The ubuntu dialogue corpus: A large dataset for research in unstructured Multi-Turn Dialogue Systems*. arXiv.org. Retrieved November 8, 2022, from https://arxiv.org/abs/1506.08909
5. Tang, Z., Kulkarni, H., & Yang, G. H. (2021, June 9). *High-quality diversification for task-oriented dialogue systems*. arXiv.org. Retrieved November 8, 2022, from https://arxiv.org/abs/2106.00891
6. Adiwardana, D., Luong, M.-T., So, D. R., Hall, J., Fiedel, N., Thoppilan, R., Yang, Z., Kulshreshtha, A., Nemade, G., Lu, Y., & Le, Q. V. (2020, February 27). *Towards a human-like open-domain chatbot*. arXiv.org. Retrieved November 8, 2022, from https://arxiv.org/abs/2001.09977