

Memory Augmented Network for Non-goal Driven Conversation using Dialogue Dataset with improved Background Knowledge

Chukwuemeka Uchenna Eneh, B. Sc.

February 14, 2019

Academic Supervisor: Prof. Dr.-Ing. Elmar Noeth (to be decided)

Industrial Supervisor: Fabian Galetzka

Thesis Description

OVERVIEW

Recent advances in artificial intelligence with the success of neural network models have seen progress in producing meaningful results in non-goal driven conversational settings [9]. Nevertheless, building intelligent conversational agents remain a problem as these models have been limited in producing satisfying results. Conversation with them even for a short period have shown their weakness as they are not very captivating to the user due to the model's tendency to lack specificity, give uninformative replies (e.g. "I don't know") [2], and display an inconsistent personality [5, 9].

Existing dialogue datasets such as Twitter, Ubuntu, Reddit and OpenSubtitles corpora [6] have traditionally been used to train conversational models. These datasets contain a sequence of utterances and responses from different speakers without any explicit background knowledge linked to the speaker or the topic of discussion, thereby causing the trained models to treat a conversation as a sequence-to-sequence generation task resulting in a simplistic response [2] which is different to the way humans converse by relying on some background information, as opposed to simply relying on the previous sequence of utterances [4].

Current trends in chitchat modeling have shifted towards making more engaging conversational agents by training them with better datasets which integrate structured background knowledge. This background knowledge gives the model a configurable but persistent personality, termed as profile [10]. These profiles are stored in a memory-augmented neural network [3, 1] and used to produce more personal, consistent and engaging responses, giving the model a more human-like conversational attribute.

DATASET

This thesis has the option of using any of the following two knowledge based datasets, depending on if the second dataset to be created is satisfactory.

The first dataset, the PERSONA-CHAT dataset, is a crowd-sourced dataset created by Zhang et al. [10] and used in the ConvAI2 competition (<http://www.parl.ai/static/docs/tasks.html>). This dataset is collected using Amazon Mechanical Turk where two Turkers are paired to converse on a particular topic and each Turker is provided a random profile from a pool of profiles on which they have to structure their dialogue around.

The second dataset (currently being created) is inspired by the PERSONA-CHAT dataset and would also be obtained using Amazon's Mechanical Turk, but the conversation would be focused on popular movies, where each speaker is given a set of facts from the movie plot and trivia (extracted from the IMDb Database), and a persona (profile) which they are to structure their dialogue around. For example, information from the plot of the movie Pulp Fiction is given, and one speaker is instructed to use a persona that likes the movie, while the second speaker uses the persona that they dislike the movie. The speakers are then tasked to make a conversation on the movie using information on the movie given to them, while staying on track with their assigned personalities.

TASK

The thesis would focus on trying out new architectures on these Knowledge-based datasets such as:

1. Perform Transfer Learning on Ranking models: Wolf et al. have performed Transfer Learning on a generative Transformer model. I would implement on Key Value Memory Network, which is a ranking model
2. Combining Ranking and Generative Models such as Key Value Memory and LSTM Network.

Pre-training would be done on the Key Value Memory Network and the weights would be used to train a generative model

3.

The second part of the work consists of implementing the Key Value Profile Memory Network [3] on the improved dataset and comparing with the current state-of-the-art. Key Value Memory Network was proposed as an improvement to the memory network by performing attention over keys and outputting the values (instead of the same keys as in the original work). This method has been shown to outperform memory networks depending on the task and the definition of the key-value pairs.

For this thesis, during training, the keys (input) would be considered as a concatenation of the dialogue histories (e.g the previous two utterances and the current utterance from the training set) and the background knowledge (such as the speaker profiles and information from the movie), while the values (output) would be the next utterance (i.e.

the reply from the speaking partner). This allows for the model to have a memory of not just the previous utterances but also information about the speakers and about the topic of conversation (e.g. the movie facts) that it can directly use to help influence its prediction of the next response. As a variation to the currently existing work by Zhang et. al [10], self-attention would also be added to the model’s decoder [8, 7] in order to prevent the model from repeating previous replies as its output.

The project will be developed in Python script using the open-source software library Tensorflow for building the neural network models, and cloud computing will be done on Microsoft Azure.

EVALUATION

The PERSON-CHAT dataset comes with 3 automated metrics on its evaluation set [10], and human evaluation would also be considered. The best model in comparison to the current state-of-the art (<http://convai.io>) will be selected and evaluated as the final goal for the project.

This thesis would be a joint work between Volkswagen AG and Friedrich-Alexander University, Erlangen-Nrnberg, and is structured to run from April, 2019 to September 2019.

References

- [1] Ankit Kumar, Ozan Irsoy, Peter Ondruska, Mohit Iyyer, James Bradbury, Ishaan Gulrajani, Victor Zhong, Romain Paulus, and Richard Socher. Ask me anything: Dynamic memory networks for natural language processing. In *International Conference on Machine Learning*, pages 1378–1387, 2016.
- [2] Jiwei Li, Michel Galley, Chris Brockett, Jianfeng Gao, and Bill Dolan. A diversity-promoting objective function for neural conversation models. *arXiv preprint arXiv:1510.03055*, 2015.
- [3] Alexander Miller, Adam Fisch, Jesse Dodge, Amir-Hossein Karimi, Antoine Bordes, and Jason Weston. Key-value memory networks for directly reading documents. *arXiv preprint arXiv:1606.03126*, 2016.
- [4] Nikita Moghe, Siddhartha Arora, Suman Banerjee, and Mitesh M Khapra. Towards exploiting background knowledge for building conversation systems. *arXiv preprint arXiv:1809.08205*, 2018.
- [5] Iulian Vlad Serban, Ryan Lowe, Laurent Charlin, and Joelle Pineau. Generative deep neural networks for dialogue: A short review. *arXiv preprint arXiv:1611.06216*, 2016.

- [6] Iulian Vlad Serban, Ryan Lowe, Peter Henderson, Laurent Charlin, and Joelle Pineau. A survey of available corpora for building data-driven dialogue systems. *arXiv preprint arXiv:1512.05742*, 2015.
- [7] Louis Shao, Stephan Gouws, Denny Britz, Anna Goldie, Brian Strope, and Ray Kurzweil. Generating high-quality and informative conversation responses with sequence-to-sequence models. *arXiv preprint arXiv:1701.03185*, 2017.
- [8] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, pages 5998–6008, 2017.
- [9] Oriol Vinyals and Quoc Le. A neural conversational model. *arXiv preprint arXiv:1506.05869*, 2015.
- [10] Saizheng Zhang, Emily Dinan, Jack Urbanek, Arthur Szlam, Douwe Kiela, and Jason Weston. Personalizing dialogue agents: I have a dog, do you have pets too? *arXiv preprint arXiv:1801.07243*, 2018.