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

Chukwuemeka Uchenna Eneh, B. Sc.

March 11, 2019

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

ORIGINAL EXPOSE

Thesis Description

Recent advances in artificial intelligence with the success of neural network models have seen progress in conversational agents. With sufficient capacity and access to large datasets, researchers have been able to produce meaningful results in some chitchat settings, using recurrent neural network based models [8]. Nevertheless, building intelligent non-goal driven 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") [1], and not display a consistent personality [4, 8].

Existing dialogue datasets such as Twitter, Ubuntu, Reddit and OpenSubtitles corpora [5] contain a sequence of utterances and responses from different speakers without any explicit background knowledge linked to the speaker or the topic of discussion associated with them. These datasets have traditionally been used to train conversational models causing the models to treat a conversation as a sequence-to-sequence generation task resulting in a simplistic response [1] 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 [3].

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 which are encoded as multiple sentences of textual description, termed as profile. These profiles are then stored in a memory-augmented neural network and used to produce more personal, specific, consistent and engaging responses, giving the model a more human-like conversational attribute [9].

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

The first database, the PERSONA-CHAT dataset, is a crowd-sourced dataset created by Zhang et al. [9] and used for the ConvAI2 competition, is available open source on ParlAI (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.

Another advantage of using this crowd-sourced dataset is that, unlike the traditionally large-scale dataset (e.g. Twitter, Reddit) which are often very noisy as they are extracted from online forums that contain a lot of typographical errors and use shorthand expressions, the PERSONA-CHAT dataset is less noisy as the Turkers have been instructed to use clean sentences which although may not always be the case, makes the pre-processing easier and faster.

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. The project is divided into two main parts.

The first part for the thesis, I would be trying out new architectures on the PERSONA-CHAT dataset. Ranking models are good at selecting what part of the training set should be used as the response while generative models are good at generating new sentences as the reply based on dialogue history.

The second part of the work consists of implementing the Key Value Profile Memory Network [2] 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 [9], self-attention would also be added to the model's decoder [7, 6] in order to prevent the model from repeating previous replies as its output.

Evaluation

The PERSON-CHAT dataset comes with 3 automated metrics on its evaluation set [9], and human evaluation would also be considered. For this thesis, we aim to reach the state-of-the-art in the challenge Conversational AI challenge (http://convai.io).

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] 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.

[2] 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.

[3] Nikita Moghe, Siddhartha Arora, Suman Banerjee, and Mitesh M Khapra. Towards exploiting background knowledge for building conversation systems. *arXiv preprint arXiv:1809.08205*, 2018.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] Oriol Vinyals and Quoc Le. A neural conversational model. *arXiv preprint arXiv:1506.05869*, 2015.

[9] 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.