Lightning Talk - Ruuh: A Conversational Social Agent

Meghana Joshi
Microsoft, India
mejoshi@microsoft.com

Sonam Damani
Microsoft, India
sodamani@microsoft.com

Khyatti Gupta
Microsoft, India
khgupt@microsoft.com

Nitya Raviprakash
Microsoft, India
niravipr@microsoft.com

ABSTRACT

Dialogue systems and conversational agents are becoming increasingly popular in the modern society but building an agent capable of holding intelligent conversation with its users is a challenging problem for artificial intelligence. In this talk, we share challenges and learnings from our journey of building a deep learning based conversational social agent called "Ruuh" (m.me/Ruuh) developed by a team at Microsoft India to converse on a wide range of topics. The authors are co-creators of Ruuh and the original paper was presented in NeurIPS 2018 Demonstration Track by two of the authors. As a social agent, Ruuh needs to think beyond the utilitarian notion of merely generating "relevant" responses and meet a wider range of user social needs. The agent also needs to detect and respond to abusive language, sensitive topics and trolling behavior of users. Some of the above objectives pose significant research challenges in the areas of NLP, IR and AI. Our agent has interacted with over 2 million real world users till date which has generated over 150 million user conversations. We intend to walk the audience through our journey of overcoming several research challenges to become the most popular social agent in India.

1 TECHNICAL DETAILS

In this section we will discuss a brief overview of the technical aspects of our approach.

Retrieving relevant responses

We model the task of providing relevant chat responses as an Information Retrieval problem, where for a given user message M and context C, the system retrieves and ranks the candidates by relevance and outputs one of the highest scoring responses. We create an index of paired tweets and their responses offline. At runtime, we first use TF-IDF-based fetch to generate a candidate set, and then extract features using a convolutional deep structured semantic network, followed by a ranker trained to select the best response. [3].

Emotion detection

As humans, on reading "Why don't you ever text me!", we can either interpret it as a sad or an angry emotion and the same ambiguity exists for machines too. Lack of facial expressions and voice modulations make detecting emotions in text a challenging problem. However, the success of our agent depends on our ability to modulate responses based on user emotions. In [2], we propose a Deep Learning based approach to detect emotions in textual dialogues. The essence of our approach lies in combining both semantic and sentiment-based representations for more accurate emotion detection.

Detecting offensive conversations

Users often abuse and provoke the agents to elicit inappropriate or controversial responses. We actively identify "controversial topics" and make clever dodging techniques to avoid responding to such topics. This problem is wrought with challenges like handling natural language ambiguity, rampant use of spelling mistakes and variations for offensive terms, and disambiguation with entity names such as pop songs which usually have abusive terms in them. We use a neural Bi-directional LSTM based model for this task [4].

Human-like image commenting

Users often interact with social agents by sharing their personal pictures, other images and videos. In such scenarios, agents are not expected to routinely describe the facts within the image but to express some interesting emotions and opinions about it. We use a modified version of [1], where the model is learnt using millions of image-comment pairs mined from social network websites like Instagram, Twitter etc.

REFERENCES

[1] Hao Fang, Saurabh Gupta, Forrest Iandola, Rupesh K Srivastava, Li Deng, Piotr Dollár, Jianfeng Gao, Xiaodong He, Margaret Mitchell, John C Platt, et al. 2015. From captions to visual concepts and back. In *Proceedings of the IEEE conference on computer vision and pattern recognition*.

[2] Umang Gupta, Ankush Chatterjee, Radhakrishnan Srikanth, and Puneet Agrawal. 2017. A sentiment-and-semantics-based approach for emotion detection in textual conversations. *Neu-IR: The SIGIR 2017 Workshop on Neural Information Retrieval* (2017).

[3] Abhay Prakash, Chris Brockett, and Puneet Agrawal. 2016. Emulating human conversations using convolutional neural network-based IR. *Neu-IR: The SIGIR 2016 Workshop on Neural Information Retrieval* (2016).

[4] Harish Yenala, Manoj Chinnakotla, and Jay Goyal. 2017. Convolutional Bi-directional LSTM for Detecting Inappropriate Query Suggestions in Web Search. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer, 3–16.