Improving Communication in Enterprise Solutions: Challenges and opportunities

Vitaliy Gorbachov and Olga Cherednichenko

National Technical University "Kharkiv Polytechnic Institute",
Kirpichova str., 2, Kharkiv, Ukraine

`raikerian@me.com`, `olha.cherednichenko@gmail.com`

This study constitutes of current state in enterprise solutions and massive possibilities of integrating chatbots in real world products.

The current level of development of artificial intelligence allows you to develop not only programs that solve the same type of applied problems. This class of software serves to support decision making [1]. For example, to implement a virtual interlocutor, who could answer certain questions. This is called a chatbot.

The relevance of the development of chatbot is justified by the widespread popularity of such systems. Thanks to the chatbots, you can reduce the number of calls to support users, as the chatbot can answer most of the simple questions put to it.

Chatbots models and domains are being looked into. The main discussion is a possibility to integrate chatbots into a commercial network and choose appropriate model and domain.

An enterprise network is an enterprise's communications backbone that helps connect computers and related devices across departments and workgroup networks, facilitating insight and data accessibility. An enterprise network reduces communication protocols, facilitating system and device interoperability, as well as improved internal and external enterprise data management [2].

A chatbot is a service, powered by rules and sometimes artificial intelligence, that you interact with via a chat interface. The service could be any number of things, ranging from functional to fun, and it could live in any major chat product.

Chatbots have two different models: retrieval-based and generative models.

Retrieval-based models use a repository of predefined responses and some kind of heuristic to pick an appropriate response based on the input and context. The heuristic could be as simple as a rule-based expression match, or as complex as an ensemble of Machine Learning classifiers. These systems don't generate any new text, they just pick a response from a fixed set.

Generative models don't rely on pre-defined responses. They generate new responses from scratch. Generative models are typically based on Machine Translation techniques, but instead of translating from one language to another, we translate" from an input to an output (response).

Chatbots can be based on either open or closed domain.

In an open domain setting the user can take the conversation anywhere. There isn't necessarily have a well-defined goal or intention. Conversations on social media sites like Twitter and Reddit are typically open domain – they can go into all kinds of directions. The infinite number of topics and the fact that a certain amount of world

knowledge is required to create reasonable responses makes this a hard problem.

In a closed domain setting the space of possible inputs and outputs is somewhat limited because the system is trying to achieve a very specific goal. Technical Customer Support or Shopping Assistants are examples of closed domain problems. These systems don't need to be able to talk about politics, they just need to fulfill their specific task as efficiently as possible. Sure, users can still take the conversation anywhere they want, but the system isn't required to handle all these cases – and the users don't expect it to.

Chatbots can improve communication in most of today's enterprise solutions. As an example, we look into integrating chat bots into commercial network. This is closed domain solution, but the model is still needs to be determined. Chatbots potentially a huge business opportunity for anyone willing to jump headfirst and build something people want.

Generative model is the hardest one, but is more rewarding. It can learn in time and its optimizing itself for each use-case. But there is a common problem with generative systems is that they tend to produce generic responses like ―That's great!" or ―I don't know" that work for a lot of input cases.

While retrieval-based model is easiest one, it can only answer using pre-determine checks. It cannot be optimized on-run.

To produce sensible responses systems may need to incorporate both linguistic context and physical context [3]. In long dialogs people keep track of what has been said and what information has been exchanged.

The main purpose is to develop chatbot on closed domain, using generative model for commercial network.

Considering all the models we have researched, an open domain based retrieval model is impossible, because you can never collect enough responses to cover all cases. The generative system of the open area is almost artificial intelligence (AI), since it must handle all possible scenarios. We are very far from such systems (but there is a lot of research that is going on in this area).

Chatbot code will be hosted using AWS Lambda technology with a specific API gateway.

References

1. D. Bahdanau, K. Cho, and Y. Bengio. Neural machine translation by jointly learning to align and translate.arXiv preprint arXiv:1409.0473, 2014.
2. S. Hochreiter, Y. Bengio, P. Frasconi, and J. Schmidhuber. Gradient flow in recurrent nets: the difficultyof learning long-term dependencies, 2001.
3. K. Cho, B. Merrienboer, C. Gulcehre, F. Bougares, H. Schwenk, and Y. Bengio. Learning phrase representations using RNN encoder-decoder for statistical machine translation. In Arxiv preprint arXiv:1406.1078, 2014.