

Kalam Chat Bot

Shreya Sharma
2015096 / IIIT Delhi
shreya15096@iiitd.ac.in

Ashutosh Batabyal
2015017 / IIIT Delhi
ashutosh15017@iiitd.ac.in

Sarthak Jindal
2015169 / IIIT Delhi
sarthak15169@iiitd.ac.in

Abstract

Natural language processing and deep learning techniques have been applied to various real world problems liking building chat engines. Software chat engines are everywhere from E-commerce websites to social media. The idea of using chatbots to mimic certain personalities has also been used in the past by researchers. The main motivation is to be able to recreate the inspiration which is obtained from eminent personalities. In this work, we build a chatbot engine to mimic Dr. APJ Abdul Kalam . We use a sequence to sequence model to be able to generate our responses to queries put forward by the user. We have manually scraped query-response pairs from online sources such as interviews with Dr. Kalam. The responses of our trained chatbot are close to the speaking style of Dr. Kalam. This work is just an application of work already done to a specific purpose. We have used already existing ideas.

1 Introduction

Avul Pakir Jainulabdeen Abdul Kalam is popularly known as Dr APJ Abdul Kalam. He lives in Indian peoples heart as the Missile Man of India. He was the 11th President of India and is called People's President as he is the only president of India who was able to connect with people and impact their lives in all spheres. Kalam worked as a scientist at DRDO where he designed a hovercraft. He has worked with the legendary space scientist, Vikram Ambalal Sarabhai who is known as the father of India's Space Programme and Kalam's biggest inspiration. Later, Kalam was transferred to ISRO and became the project director in 1969

of Indias first indigenous Satellite Launch Vehicle, SLV-III. Time and again his words have motivated millions of hearts. Be it, his speeches, interviews or his novels, Kalam has always tried to spread courage and wisdom among the people of India. He had a vision for India and believed in the youth to achieve that vision. He enthusiastically participated in various interactive sessions with college and school students throughout his life, where he used to inspire young minds to achieve their dreams and follow their passion. After his term as the president of India, he chose his old way of life which included education, innovation and research. Unfortunately, He left us on 27 July 2015 at the age of 83 and even today his speeches are viral over social media and his books still carry his essence. Creating a chatbot that mimics Dr Kalam has many applications in today's world. It can be useful for the younger generation which is always looking for inspiration and motivation. It can help people in depression as the chatbot will act as a companion which can guide the patient with motivational messages just like Kalam. It can also help people to learn from his life in an interactive manner. He has tried to teach us lessons through short stories he has mentioned in his novels but the reader is left with many questions in mind and can not find a way to get the answers. His contribution to the scientific field and knowledge of space crafts makes him the first person that comes to our minds when we seek answers to problems related to the future of technology in India. By interacting with a Kalam chatbot, students interested in science and scientist working in spacecraft design can get answers to many scientific queries as well. In many interviews, he has given examples of leadership skills from his life and the chatbot can be of use to those who need leadership tips from time to time. Even today, India needs his words that made us visualize a bright future of

our country, to make progress technically as well as spiritually. Chatbots are computer programs that are capable of carrying out sensible conversations in textual format. There are many chat engines that help in dealing with customers for a company as customer-help-providers but these chatbots are usually trained on commonly asked questions asked by customers and are specific to the services of the company. These chatbots use Information Retrieval techniques which are used to answer factoid type questions. Conversational agents deviate a little from customer-help-provider chatbots in terms of application, usage and methodology. Apple's Siri, Microsoft's Cortana, the recent Google assistant and Amazon's Alexa are good examples of conversational agents today. These chatbots are still a work in progress because creating a conversational agent is a huge challenge in the industry. Though these chatbots talk just like humans, they don't have a personality or a speech style associated with them. Mimicking someone's speech style in a chatbot adds an additional challenge in designing a conversational chatbot. There have been many attempts to create chatbots with personality traits and Section 2 mentions related works in this domain. Section 3 describes the dataset used in this paper and Section 4 describes the model that is built to create a chatbot that mimics APJ Abdul Kalam. Section 5 is the evaluation section where all results are mentioned and section 6 concludes the paper.

2 Previous work

Creating chatbots is still an unsolved problem and hence many attempts have been made to create a bot that can chat just like humans. Chatbots are majorly used in E-commerce industry and are also deployed in medical industry. But the majority of the chatbots before were based on hand-crafted rules like Chatscript, a popular rule-based conversational model. In 2014, the sequence to sequence model being used for translation opened the possibility of phrasing dialogues as a translation problem: translating from an utterance to its response. The systems built using this principle, while conversing fairly fluently, are not very convincing because of their lack of personality and inconsistent persona. However the work done on chatbots with a personality is limited. One of the chatbots with a personality was built by a group of Stanford students. In that paper, they experimented

by building an open-domain response generator with personality and identity. They built chatbots that imitate characters in popular TV shows. A model like this could have a variety of applications like talking to your favourite personality which could be either used to help depression or just for fun. We could also potentially have an alter-ego or a clone bot if we provide it the data. The results of the translation model gave poor scores. In another paper they used two models - models, a single-speaker speaker model and a dyadic speaker-addressee model, within a sequence-to-sequence (SEQ2SEQ) framework. The Speaker Model integrates a speaker-level vector representation into the target part of the seq2seq model. Analogously, the Speaker-Addressee model encodes the interaction patterns of two interlocutors by constructing an interaction representation from their individual embeddings and incorporating it into the SEQ2SEQ model. These persona vectors are trained on human-human conversation data and used at test time to generate personalized responses.

3 Dataset

The data set contains 500 question-answer pairs. The data follows a particular format. Each question-answer pair is separated by a line and the line containing the question starts with Q# and the line containing the answer starts with A#. Figure 1 shows a sample data. The data was collected from various sources over the internet. The question-answer pairs are extracted from DR APJ Abdul Kalam's conversations with journalists and students. The interviews are publicly available on the sites of many newspapers, magazines and blogs. One of the major sources was his official website - abdulkalam.com

4 Methodology

4.1 Generative Model

We have decided to use a sequence to sequence model for generating the responses from our chatbot. There are a few basic sections to design for any such sequence to sequence model. These sections of our model are mentioned below.

4.1.1 Vocabulary

Firstly, we build a vocabulary. It serves two primary purposes. One is to get a unique id for each word in our training set. And the other is to get

Q#What are the job opportunities in Public and Private sector for us (Speech and Hearing Impaired Students)?
A#3 % reservation is allotted for the differently abled persons including speech and hearing impaired students. Certainly now the number of private industries are encouraging and appointing the speech and hearing impaired students such as Sakthi masala in Tamilnadu. Many more industries will identify the right jobs and every one of the students who are having speech and hearing impaired students.

Q#Though India is continuing with corruption, hike in price and economic crisis, can it become a developed country within 2020 as you said? What are the steps to be taken by the youth towards that development?
A#As a youth of the nation, first and foremost your action is to study well. That is your first responsibility. Regarding corruption and other issues, the reform has to start from our home. If our home is corruption free, our village and city will be corruption free, then our state will be corruption free and the nation will be corruption free. Certainly our leaders should have the wisdom to bring down the price rise and manage the economic crisis.

Q#What is the biggest challenge facing India today and what should we do to contribute positively towards this challenge?
A#The youth needs to develop an attitude : I can do it! We can do it !! India will do it!!!

Q#what is the secret of success in your life?
A#Have an aim in life, acquire knowledge through all possible sources, work hard and never accept defeat by the problem, always defeat the problem and succeed.

Figure 1: Sample Data

a reverse mapping from index to the corresponding word. This is useful while training the model since the model cannot take as input raw text. In order to get better results we perform several operations on words in our vocabulary. These involve removing punctuations or symbols, removing sentences which are greater in length than a pre-defined threshold, etc. One of the major problems in training chatbot engines is the slow speed of convergence during training. Since the amount of data required for training chatbots is large some common techniques have been used to aid in faster convergence. One of the best techniques is to remove the query-response pairs in our dataset which have infrequent words. Basically, we trim out any pairs which contain any word which has occurred less than a threshold number of times. Not only does this help in faster convergence during training but it also makes the generation job easy for the model. The model has to learn a slightly lesser complex function.

4.1.2 Using mini-batches

We could simply train our sequence to sequence model using only one data point at a time. However, once again this would lead to slower convergence during training. We want to be able to leverage the parallelization capabilities of our systems. But, in order to that we need to train on mini-batches instead of one sample at a time.

4.1.3 Sequence to Sequence Model

The name suggests everything about this model. Its input and output are both sequences. These sequences can be of variable length. However, the model is interesting in the sense that its own size is fixed.

5 Results

We obtained satisfactory results. The bot is generating sensible answers which are similar to Kalam's words as it can be seen in Figure 2 in which the sample chat provided.

Bot: the youth needs to develop an attitude i can do it ! we can do it ! ! india will do it ! ! ! !
> scientific development?
Bot: . scientific magnanimity . science should be a life time mission . research challenges . vision for the nation
> why do you like children of India?
Bot: because they are vibrant full of energy innovative and keen learners . the ignited minds of the youth is most powerful resource on the earth above the earth and under the earth .
> what to do?
Bot: we have to be a youth . work hard and persevere .
> why do you like youth of India?
Bot: because they are vibrant full of with persevere . i have always regarded aim in life .
> what inspires you
Bot: i am overwhelmed by the gesture . being the nation will bring knowledge back to the nation .
> your teacher?
Bot: my teacher task to have a vision and aim in life .
> should we aim for success?
Bot: definitely . all of the youth is always my priority .
> how to be good at science?
Bot: . scientific magnanimity . science should be a life time mission . hence india is taking steps to attract the scientific and thus created iiser in many regions . i have always a democratic country we have a fro and fitted to parliament . this has ofcourse to increase in

Figure 2: Sample Chat

References

We have used online chatbot tutorials to build our chatbot.