Data Chat

Urjitkumar Patel Vincent Chabot
up276@nyu.edu vec241@nyu.edu

***Abstract*—AI based Chatbots, programs which are able to interact with humans in natural language and perform actions to meet user's request (like booking a flight), are considered as the future our interactions with the digital world in which we live. End-to-end chatbots are constituted of several building blocks and the Natural Language Processing block, that allow the program to understand language (and possibly generate answers), is one of them. The focus of our project will be on developing an advanced Question Answering NLP engine to be integrated as part of a end-to-end chatbot developed by Zelros (zelros.com) and aiming at solving conversational BI challenge.**

I. Artificial Intelligence Chatbot Context & Motivations

Artificial Intelligence, or our future way to interact with machines and objects

We used to interact with computers via command line, which was later on replaced by the graphical user interface and point-and-click mouse. Smartphones and touchscreens changed those interactions into swipes and taps, but yet it still takes time to have to switch from one application to another each time we want to do something different. Think about planning a weekend out of the city : you need first to check airbnb, open google to compare the flights, and you'll also have to juggle between Uber Yelp and other Trip Advisor to find what to do on site.

So what could be easier than just downloading an app and tapping a few virtual buttons with our fingers? What if interactions with the world around us were as easy as walking into our favorite bar and getting our favorite drink in hand before we even reach the bar. The bartender knows us, knows exactly what drink we like. Thats a lot of interaction, without any interaction", and that's exactly what AI could solve along with the tremendous amount of personal data we generate each day. Building a new user interaction model to enable devices to listen, feel and talk to us is exactly the promise of AI.

For the moment, the way towards which we are heading to use AI to interact with the world around us seems to be using texting to chatbots. Chatbots are like regular contacts in your messaging app but instead of talking to a real human, you talk to a bot (program). And why texting ? Because text messaging is natural to all of us and it is the most widely used smartphone feature. For instance, you could type "I'd like to go out of the city with my wife next weekend, what do you recommend me ?" Based on your data, it would recommend you few destinations you'd be likely to like, and once you'd have made your choice it would offer to book transportation and room for you. You'll ask the chatbot to order the uber at your landing and during your weekend it will make you all sorts of restaurants and activities recommendations.

What Chatbots can currently do

But although there is currently a big hype around chatbots and good reasons to believe that they will the way we'll interact with our world in the future, we are still very far from chatbots being able to handle full conversations as humans would do and perform all the tasks we would like them to perform.
Nevertheless, there is not one day without articles written on the subject and all major companies, starting with the GAFA, are currently spending millions to develop their own chatbots capabilities. The first we can think about is Facebook that enables developers to create their own chatbot in messenger (still in beta). For instance, mealou.co asks you what you'd like to eat and recommends you restaurants around you, instead of having to open yelp and select manually many filters. Second, Google is developing Alo, Gmail's chatbot that should be able to handle many tasks for you like answering certain emails for you (like scheduling appointments). Finally, more and more APIs become available such as IBM Bluemix, Microsoft Luis, wit.ai etc. on the market for developers to develop their own chatbots without needing particular skills in Artificial Intelligence and Language Understanding.
Even if chatbots are generating big hopes and are evolving quickly, they can still only perform limited task and need to be task specific for the moment in order to perform well.

II. Project context

Building the NLP Engine for a conversational BI chatbot

For our project, we've decided to work with the company Zelros (zelros.com) whose vision is to build Conversational BI, that is to say a natural way for people to interact with companies data. Indeed, data sharing and data accessing is a real pain point within companies. It usually requires to use SQL-like query, open often out of date source files, or even worse, email exchanges between business users and data owners.
Zelros believes that being able to access quickly available data in natural language should be the norm. A sales manager could then ask "what is the average sales amount per vendor for last month" and get the answer instantly without having to write in some SQL request or ask his data analyst and then have to wait for the answer.
Therefore, to solve this, the objective of Zelros is to build

a chatbot as a natural interface between users and data. To answer a question such as the previous one, the chatbot would first have to understand the question and figure out context information such as to which month ”last month” is referring to. Then it would have to find where in the data source file is the data corresponding to ’sales amount’ and ’vendors’ located. And finally, since the user asked for the average sales amount, it would have to perform the required aggregation on the data before delivering the result to the user.

Technically, end-to-end chatbots are constituted of several building blocks in order to perform all the necessary steps and answer the user’s question. For our project, we will focus on helping Zerlos to build the NLP (Natural Language Processing) part of the bot. The role of the NLP engine is to determine what the user is saying (the query or intent). It learns from sentences examples, and then is able to classify new sentences among the known types.

Challenges for building an NLP engine

- **Data Availability** Getting domain specific knowledge to build the chatbot itself is the biggest problem in current ongoing development of chatbots. In short, Training should never stop. With each question user asks, chatbot should learn whatever possible from this question.
- **Context Based Conversation** Right now, chatbots require us to master their syntax. If you ask them Whats the weather? youll usually get an answer. If you ask Could you check the weather? you might not. To get around this problem, we need to enable bots to understand all the ways we say something and how context changes what we mean
- **Finding sub information for user’s question** Rather than try to parse a single sentence like schedule a meeting with Rob for Friday to talk about Marketing, and the 100 different ways I could make that request, you might take it in steps and make your chatbot ask specific questions.

  Me: Schedule a meeting
  Chatbot: Who would you like to meet with?
  Me: Rob
  Chatbot: When would you like to meet with Rob?

- **Linking user questions to data** There are different ways we can think of to match a question in natural language to the accurate data. Converting a user question into a SQL query is one of them. Generating questions from data headers and then matching user’s question to the most likely data generated question is an other.As an example, you upload a csv file having Year and population for that year, your AI Chat agent can be able to find the data corresponding to ”What was the population in 2008?”.

III. Project Description

What are we targeting to solve?

Our goal for this project is to solve the challenge of understanding a user’s question by building an efficient Question Answering NLP engine that performs accurate question classifications and extract from the questions any context based information that will later on help the system to use these context information to perform better.

Strategy

To solve the problem we described, we will follow the steps as below,

- **Data Collection**: Data Collection is very important as for any ChatBot to work better we need good amount of data on which Chatbot can be trained on.
- **Data Augmentation and Cleaning**: Data augmentation adds value to base data by adding information derived from available data. Simple example of this would be creating 10 sentences having the same meaning out of 1 sentence. We will need to clean the data to make it consumable.
- **Developing Base QA Machine**: Once having the data, our first goal would be to develop a very basic Question Answering Machine which can understand the user question, and fetch the available information relative to the user question.
- **Relative Information Retrieval**: There may be many relative information for the user question. Interesting part would be to understand what data out of this all relative information should be shown to the user.
- **Adding Context**: Our next goal would be to make the conversation contextual and more natural for the user. We should be able to relate the new question with the context of the question even though not specified directly in the question.

Data

We went through many papers which in context of Natural Language Processing, and came to know about really good data source for Natural Language Understanding, **Yahoo Question Answer Data set**, which has approximately **4.5 Million questions with best answers**. We expect to get good results using this data source.

Evaluation

Zelros already has a benchmark evaluation system, which they named as ”Bunt”, which allows to compare their own solution with the other available market tools. We would use this tool on the top of other Data Science Evaluation Matrices. Our baseline would be the current available NLP engine built at Zelros.