A Natural Language Processing Framework for Training a Neural Network Chatbot

Kamila Michel

Maciej Majchrzak

B.Sc.(Hons) in Software Development

March 25, 2019

Final Year Project

Advised by: Gerard Harrison

Department of Computer Science and Applied Physics
Galway-Mayo Institute of Technology (GMIT)

Contents

List of Figures

About this project

Abstract A brief description of what the project is, in about two-hundred and fifty words.

Authors This project was developed by Kamila Michel and Maciej Majchrzak, final year students of Software Development at Galway Mayo Institute of Technology.

Acknowledgements

We would like to take this opportunity to acknowledge and thank everyone who helped me during the creation of this project.

In our opinion, the person who deserves the most thanks for help at the key moment of this project is our supervisor Mr. Gerard Harrison. Thank you for all your help and support during the very difficult moments of the development phase.

We would also like to thank to Mr.John Healy, Head Lecturer of the Applied Project and Minor Dissertation module.

We would also like to thank the lecturers and employees of Software Development for their dedication, support and unforgettable four years during which We met many wonderful people and learned new skills. It was not always easy but sacrifice and hard work paid off and it was worth it. We have experienced many interesting moments here, and most importantly, We gained knowledge that will allow us to find a job We always dreamed about.

Chapter 1

Introduction

The constant process we are dealing with in the universe is evolution, thanks to which, over the years, as citizens of our planet, we have been able to manage this process. Evolution concerns not only the biological aspect but also the technological aspect. The 21st century has brought us great development of technology, practically every kind. Progress in recent years cannot be overlooked in information technologies and, above all, in the evolution of the Internet. Information to which we have access, and the form of their presentation exceed our imaginations from a few years ago. The reason for this is exponential growth in the quantity, quality and information presentation technology. During four years of our studies in software development, we had opportunities to learn new technologies and skills needed in our further career. However, there is a topic that still has many unresolved questions and is still growing. It is Artificial Intelligence. Our goal was to obtain information in this field and implement it with our project. We were looking for an idea where we could implement the knowledge we have already acquired and combine them with new problems that were not discussed in our course. The growing importance of computers in the daily life of people and their rapid development has meant that for years, attempts have been made to give them human characteristics. Implementation of something that is an extremely popular topic in science fiction literature, from year to year, it seems to be more and more likely. Regardless of whether these attempts are undertaken for entertainment or educational purposes, or to simplify and increase cooperation between the machine and human, an important element is the dialogue. The simplest form is the implementation of given commands. Humans in their nature personifies the objects, animals and phenomena. When working with an object for a long time, he begins to give it more and more human traits. It is not unusual, therefore, to talk to the device, and it is often encountered. Our first idea was to build a smart

speaker using the Google AIY Voice Kit [1]. After thorough analysis and implementation, we concluded that using the Google Project will be associated with a small amount of programming. Code was entirely accessible, and we could only modify it. We decided to build our own chatbot which will use the neural needed to train the conversation data. In the further part of dissertation, we will describe how neurons works and how Natural Language Processing was used in our project. The main purpose of our chatbot is to make a conversation with the user who can choose the conversation topic from given list.

1.1 A Brief Introduction to Artificial Intelligence

Artificial intelligence is now a vast and progressively developing field of computer science. The number of new products, tools and services using that mechanisms has been growing rapidly. AI technologies are becoming increasingly available as practical solutions in the real world. According to Wikipedia, the term ”artificial intelligence” was first used by John McCarthy in 1956 at a scientific conference in Dartmouth, where was defined as the ability of machines to exhibit manifestations of human-like intelligence. The basis of AI are algorithms that allow computers to learn and make decisions. The most advanced programs have been inspired by the human understanding of how our brain works - the connections between neurons. One of the techniques used to build AI systems is machine learning. ML algorithms are not based on rules manually entered into the system by humans, but they automatically build models based on training data and a defined goal. Another mechanism is Deep Learning. DL models are extremely complex structures containing millions of parameters and capable of analysing complicated tasks such as human speech, images, video streams or non-trivial patterns in business data. Deep Learning models usually require a large amount of training data and computational power necessary to process them - that's why we have been observing their effects only for several years. But it is the deep models behind the recent successes of AI and machine learning.

1.2 Chatbot

Chatbot is a computer program used to communicate between the machine and a human. The degree of advancement such interactions can be very diverse. The simplest chatbots are simply block diagrams which, while asking

more narrowly asked questions, help to reach specific answers. More complicated technologies are dialogues that detect certain words and sentences of a human being. They are able to match the closest retorts and responses. The next stage of advancement is the artificial intelligence module AI (Artificial Intelligence). Thanks to that, the computer can remember the answers we have given, as well learn and improve. The peak of advancing chatbots is the ability of software to read human speech in natural language, supported by the AI module and fluent answer to human interaction.

Chatbots are not an invention of recent years. In 1950, Alan Turing came up with the idea of creating a test as part of research on the development of artificial intelligence. He relied on human dialogues in natural language with other people or machines. Wanting to prove the intelligence of machine, which was supposed to rely on learning abilities in the 1950s. He created the Turing test, which is still used to evaluate the effectiveness of chatbots.

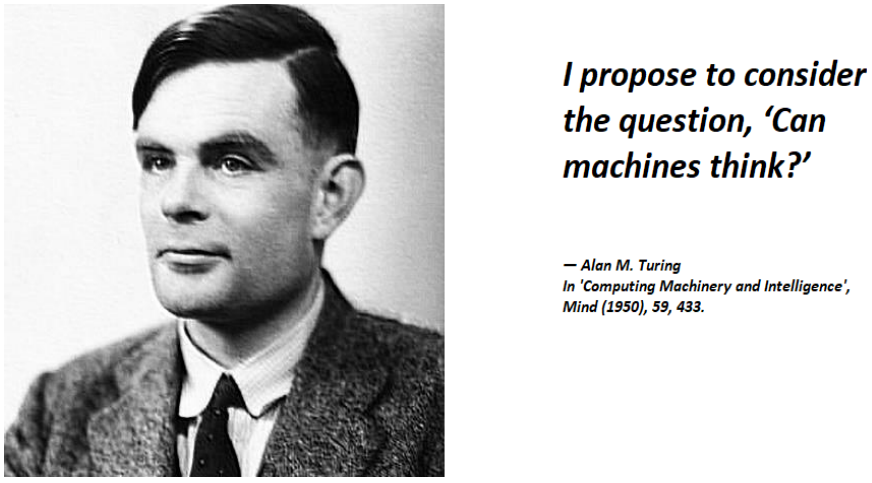

Figure 1.1: Alan Turing

The first program, which can be called a chatbot, was created in 1966 by Joseph Weizenbaum. It was named ELIZA and was quite primitive. Its operation was based on transforming the user's statements into questions, which actually gave the impression of contact with a human being, but it did not have practical application. It was necessary to wait for the creation of a breakthrough, intelligent chatbot until 1995. Then, the ALICE program appeared, which in terms of the accuracy of the response significantly departed from its predecessors. In the 21st century, the development of chatbots progresses very fast, programs cope well with machine learning, better

algorithms are created, and the systems' efficiency and server wealth allow chatbots to collect and use huge amounts of data.

Chatbot as a computer application usually consists of three parts. User is available to access the first part, a program that is a desktop or web application. The interface usually contains a window where is displayed conversation between chatbot and user, also a field for entering text and often a graphic representation of a chatbot. The other part of the chatbot is its engine, the heart of the program that processes the user's instructions. It implements all algorithms that allow to recognize the entered sentences, their processing and responses. The last but the most important part of the program is its knowledge base. It contains all the phrases, sentences on which the program can operate and what it understands. This database can be any implementation and depends mainly on the programmer. This knowledge is usually stored in text files or in a relational database, which often facilitates and speeds up the search and other data operations.

Chapter 2

Context

- Provide a context for your project.
- Set out the objectives of the project
- Briefly list each chapter / section and provide a 1-2 line description of what each section contains.
- List the resource URL (GitHub address) for the project and provide a brief list of the main elements at the URL.

2.0.1 Methodology

2.0.2 Technology Review

2.0.3 System Design

2.0.4 System Evaluation

2.0.5 Conclusion

2.1 Background Research

Chapter 3

Methodology

3.1 Agile Development

3.2 Version Control

3.3 Sprints

3.3.1 Sprint 1

3.3.2 Sprint 2

3.3.3 Sprint 3

3.3.4 Sprint 4

3.3.5 Sprint 5

3.3.6 Sprint 6

3.3.7 Sprint 7

3.3.8 Sprint 8

3.3.9 Sprint 9

3.3.10 Sprint 10

3.3.11 Sprint 11

3.3.12 Sprint 12

3.4 Testing

About one to two pages. Describe the way you went about your project:

Chapter 4

Technology Review

About seven to ten pages.

- Describe each of the technologies you used at a conceptual level. Standards, Database Model (e.g. MongoDB, CouchDB), XMl, WSDL, JSON, JAXP.
- Use references (IEEE format, e.g. [1]), Books, Papers, URLs (timestamp) – sources should be authoritative.

4.1 Visual Studio Code

4.2 GitHub

4.3 Python

4.4 JSON

4.5 Google Cloud Platform

4.6 AIY Voice Kit from Google

4.7 Natural Language Processing

4.8 TensorFlow

4.9 SQLite3

4.10 MySQL

4.11 Tkinter

4.12 CSV

4.13 LaTex

4.14 TeXstudio

Chapter 5

System Design

5.1 Architecture

5.2 Data Design

5.2.1 Dataset Generation

5.2.2 JSON

5.2.3 CSV

5.3 Component Design

5.3.1 Artificial Neural Networks

5.3.2 Pattern Matchers

5.3.3 NLP

5.4 GUI

As many pages as needed.

- Architecture, UML etc. An overview of the different components of the system. Diagrams etc... Screen shots etc.

Chapter 6

System Evaluation

As many pages as needed.

- Prove that your software is robust. How? Testing etc.
- Use performance benchmarks (space and time) if algorithmic.
- Measure the outcomes / outputs of your system / software against the objectives from the Introduction.
- Highlight any limitations or opportuni-ties in your approach or technologies used.

Chapter 7

Conclusion

About three pages.

- Briefly summarise your context and ob-jectives (a few lines).
- Highlight your findings from the evalua-tion section / chapter and any opportuni-ties identified.

Bibliography