Developing an Intelligent Chatbot

Group 18: Alexander Vranic* - 100350592, James Hamilton - 100340006

May 29, 2024

Abstract

ABSTARACTIONS

1 Introduction

In recent years, chatbots have become increasingly popular in a variety of applications, as such, so has the technology surrounding them. It is not our mission to compete with the likes of OpenAI (2024*b*), Microsoft (2024) and GitHub (2024*a*) and their hugely successful Large Language Models (LLMS), GPT-4 OpenAI (2024*a*) and CoPilot GitHub (2024*b*) respectively. At the end of the day, the coursework is a learning experience, not a fully fledged product.

Our solution is a small clientside chatbot, integrated into an intuative graphical user interface and designed to handle prompts around a bespoke context. We utilise modern natural language processing (NLP) techniques, with a knowledge base and infernece engine, conjuncting machine learning and web-scraping, in the name of enabeling the user with concurrent information to make informed descicions on their train travel plans.

1.1 Background and Motivation

1.1.1 Compulsory Motivators (Assignment Brief)

As specified in the CMP6040/7028 assignment brief Wang (2018), tasks one and two are to implement an intelligent conversational system, designed to "*help their customers in finding the cheapest available ticket for their chosen journey*" covered by 1.2.1 and "*to improve customer service satisfaction by applying some appropriate AI techniques*" covered by section 1.2.2 respectively. Following course content and implicit suggestions from the modules authorative figures, our second task implements a delay prediction model, based on historical data (also provided in the course material), in the form of a KNN regressor (Fix & Hodges (1951)), embedded within the original chatbot system created in task one. In depth coverage of our interpretation seen in section 1.2.1.

Again, as stated in the brief, we are to provide some kind of user interface. Seeing that with our resources, any web based applications would be limited to local hosting anyway, we have decided to create a stand alone desktop application with a graphical user interface (GUI) to mimic the look and feel of modern chat applications. For further clarification, see section 1.2.2.

1.1.2 Chatbot History

Quite some time before the advanced development of AI chatbots, Alan Turing considered the idea of a hypothetical machines ability to think, proposing a method of benchmarking a machines intelligence, aptly named "*The Imitation Game*" Turing (1950), not to be confused with the Turing Test - being the broader concept of measuring a systems intelligence. In the paper, Turing proposed a system involving three parties: a human interrogator, a human respondent and a machine respondent. The core concept

being that the human interrogator must converse with bohth responding parties and determine which is the machine. The machine is deemed intelligent if it is not reliably distinguished from the human respondent. Turing states that "*at the end of the century*" - being the year 2000 - "*one will be able to speak of machines thinking without expecting to be contradicted*". Though the timing of his prediction can be argued either way, the concept of a machine being able to hold conversation with a human is now a reality, to the point where as a modern humans, we must be consciously question the *human-made* authenticity of the content and media we consume.

Aritificial intelligence, designed to mimic human conversation has come a long way since the days of Turing. From the first chatbot ELIZA Weizenbaum (1966) providing incoherent responses diverged from context, to today's CoPilot GitHub (2024*b*) baring the capability to generate code and explain it in any array of natural languages, the technology has become an integral part of our daily lives.

1.2 Aim and Objective

As I'm sure you're aware by this point in the report, we are to employ artificial intelligence techniques, in conjunction with webscraping to achieve the following two tasks (Wang (2018)), with data relevant to the current and/or user-specified timeframe(s).

1. Finding the cheapest train ticket
2. Improving Customer Service

The following subsections 1.2.1 and 1.2.2 outline our subjective interpretation of the task one and two respectively.

1.2.1 Finding the Cheapest Ticket

We've taken it upon ourselves to not just find the cheapest individual ticket, but to find the cheapest combination of tickets for a given journey. Once information is derived and tokenised, we achieve this by scraping the SplitMyFare website Fare (2024).

Seeing that their service does encour a small additional fee, initial intentions were to then compare the results from another source, such as Trainline Trainline (2024), to ensure the cheapest pricing of single ticket journeys. However, it would seem that these websites are intentionally or unintentionally difficult to scrape informatino from, as such, focus was shifted to the components relevant to the names sake of the module.

See conveyed evidence in sections 1.4 and 4, as well as the actual codebase itself *./webscrape/nationalrail.py* Hamilton & Vranic (2024).

References

Fare, S. M. (2024), 'Split my fare', `https://www.splitmyfare.com`.

Fix, E. & Hodges, J. L. (1951), 'Discriminatory analysis, nonparametric discrimination'.

GitHub (2024*a*), 'Github', `https://github.com`.

GitHub (2024*b*), 'Github copilot', https://copilot.github.com.

Hamilton, J. & Vranic, A. (2024), 'Implementation github repository', https://github.com/j-laze/Ai-CW2-Chatbot.

Microsoft (2024), 'Microsoft', https://www.microsoft.com.

OpenAI (2024*a*), 'Gpt-4', https://openai.com/gpt-4.

OpenAI (2024*b*), 'Openai', https://openai.com.

Trainline (2024), 'Trainline', https://www.thetrainline.com.

Turing, A. (1950), 'Computing machinery and intelligence', https://www.cs.colostate.edu/~howe/cs440/csroo/yr2015fa/more_assignments/turing.pdf.

Wang, W. (2018), 'Artificial intelligence modules (cmp6040, 7028) coursework specification'.

Weizenbaum, J. (1966), 'Eliza—a computer program for the study of natural language communication between man and machine', https://dl.acm.org/doi/pdf/10.1145/365153.365168.