TOWARDS ASSISTING HUMAN-HUMAN CONVERSATIONS

BY

TEJAS NANAWARE

DEPARTMENT OF COMPUTER SCIENCE

Submitted in partial fulfillment of the
requirements for the degree of
Master of Science in Computer Science
in the Graduate College of the
Illinois Institute of Technology

Approved ___________________________
Adviser

Chicago, Illinois
May 2021

ACKNOWLEDGEMENT

I would like to thank Illinois Institute of Technology to provide me with this wonderful opportunity to perform my Master's Thesis Research. I would also like to thank Prof. Shlomo Argamon for his expert guidance throughout the project. Lastly, I would like to thank my family, friends and all other participants in this research study without whom this project would not have been successful.

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENT iii

LIST OF TABLES v

LIST OF FIGURES vi

ABSTRACT vii

CHAPTER

1. INTRODUTION 1

    1.1 Problem Statement 1
    1.2 Overview of Proposed Approach 1

2. RELATED WORK 3

    2.1 Open Domain Chatbots 3
    2.2 Conversational Assistants 6

3. PROPOSED SOLUTION 9

    3.1 Selection and Deployment of AdvisorBot 11
    3.2 Chatbot Interaction Framework 14
    3.3 Data Gathering and Data Privacy 23
    3.4 User’s Perspective 24

4. OBSERVATIONS AND RESULTS 25

5. TRANSFER LEARNING 31

6. FUTURE SCOPE 36

BIBLIOGRAPHY 37

LIST OF TABLES

Table Page

3.1 Messages and Chatbot Responses 12

3.2 Messages from Rosie AIML Chatbot 13

5.1 Comparison of the evaluation metrics for the models with the data from the research study 33

5.2 Messages from users with suggestions from transfer learning AdvisorBot 35

LIST OF FIGURES

Figure Page

3.1 User Registration and Consent 16

3.2 Karma Equation for User’s Profile 17

3.3 User’s Profile View 18

3.4 Chat Interface with Suggestions on the Received Message 20

3.5 Post-chat Questionnaire User Interface 22

4.1 Counts of the times users clicked on the suggestions from the AdvisorBots 26

4.2 Delay of AdvisorBot Suggestions in Seconds 27

4.3 Number of Previous Messages when a user votes the conversation quality . 28

4.4 Number of Previous Messages when a user votes the AdvisorBot quality 29

ABSTRACT

The idea of the research is to understand the open-topic conversations and ways to provide assistance to humans who face difficulties in initiating conversations and overcome social anxiety so as to be able to talk and have successful conversations. By providing humans with assistive conversational support, we can augment the conversation that can be carried out. The AdvisorBot can also help to reduce the time taken to type and convey the message if the AdvisorBot is context aware and capable of providing good responses.

There has been a significant research for creating conversational chatbots in open-domain conversations that have claimed to have passed the Turing Test and can conversate with humans while not seeming like a bot. However, if these chatbots can conversate like humans, can they provide actual assistance in human conversations? My thesis is to observe and improve the advanced open-domain conversational chatbots that are put in practice for providing conversational assistance.

While performing this thesis research, the chatbots were deployed to provide conversational assistance and a human study was performed to identify and improve the ways to tackle social anxiety by connecting strangers to perform conversations that would be aided by AdvisorBot.

CHAPTER 1

INTRODUCTION

1.1 Problem Statement

The goal of this project is to provide conversational assistance to humans for keeping the conversation running by providing assistants. These assistants can help in understanding the context of the chats and provide the user with the next sentence that can lead to a successful conversation. The topic can be narrowed by providing with the research goals:

1. Determine linguistic features of textual 1-on-1 chats that predict user satisfaction with a casual, open-topic discussion with a (relative) stranger.
2. Devise an automated “advisor” to help individuals have more satisfying chats.
3. Determine the quality of the AdvisorBot in terms of accurately providing suggestions that can help achieve fruitful conversations and AdvisorBot’s ability to understand the subjective aspects of the conversations.

1.2 Overview of Proposed Approach

The proposed approach of providing conversational assistance is by using the existing open domain chatbots that are trained to be able to understand context and trained for generating open domain dialogues. The ones that have a good understanding of perplexity and having free open-source access such as Microsoft DialoGPT and Facebook BlenderBot were chosen for this task. The preliminary requirements for the chatbots to be used were that these bots must have a good contextual understanding, and would be able to produce relevant sentences as a response that are used in a human-like conversation.

A chat interaction framework was designed for connecting one participant with another participant or a chatbot which would allow them to engage in casual open-topic discussion. The AdvisorBot would provide suggestions to the participants which the user may choose to use. The participants were asked several questions after the conversations in order to evaluate the performance of AdvisorBot and perform linguistic analysis to determine the linguistic features of textual human-human conversations and determine the quality of AdvisorBot in terms of accurate suggestions and AdvisorBot's ability to understand the subjective aspects of the conversation.

The data gathered would provide essential information that would help in analyzing the textual features leading to satisfactory conversations. This would help to improve the AdvisorBot to achieve satisfactory conversations by transfer learning. The retrained bots would therefore be able to achieve success in terms of providing accurate suggestions that the users would actually need through a better understanding of the context.

CHAPTER 2

RELATED WORK

There have been several chatbots that are novel and achieve human like conversations in scenarios where there are human-chatbot interactions. However, the goal is to device a chatbot that targets to improve human-human conversations. For simpler understanding, this chapter is divided into two sections.

1. **Open Domain Chatbots:** This section talks about creating the chatbots and their evaluations.
2. **Conversational Assistants:** This section talks about the applications of chatbots in several areas.

2.1 Open Domain Chatbots

This section talks about the several open domain chatbots that could potentially help in providing conversational assistance. Since 1966 there has been a constant improvement in the field of natural language processing when the early pattern matching computer program ELIZA was created to demonstrate communications between humans and machines. [1] This model however did not have contextualizing capabilities as it relied completely on pattern matching. However, the modern machine learning algorithms have evolved natural language processing to formulate chatbots like Google Meena that claim to be able to talk about anything. [2] This multi-turn open domain chatbot was designed to chat on large topics of conversations which mainly featured public domain social media conversations.

The incorporation of persona is necessary for empathetic conversations because one person talks differently with two people in a manner of previous context and known knowledge or friendship. [3] Therefore, construction of a response adaptive Speaker-Addressee model where the model adapts to the way a person communicates with a particular person is important while creating conversational assistants. An example as provided by the authors is interactions of Ross, the character from the popular TV series Friends, differs depending upon if he is talking with Monica, his sister, or Rachel, his on-again off-again partner. It can be noted from this that the adaptive model depends upon the previous interactions with someone.

By winning the Amazon Alexa Prize Challenge, Papaioannou Et. Al. have demonstrated a chatbot that can conversate coherently and engagingly with humans for twenty minutes. [4] This was achieved by creating an ensemble of chatbots that can provide the resulting response based upon a score function that retains context and change of context information to generate a response with a ranking function choosing the response from the chatbots by prioritizing them based on context understanding skills. The ensemble model used, consisted of a rule based AIML chatbot, Rosie [5], which is similar to the 1966 Eliza [1].

Neural Language Models can help in generating customizable affective texts that can be used as a generative model for creating the context response of a chatbot [6] through several datasets or can be used for generating phrases by performing sentiment analysis to understand the underlying emotions in sentences and analyzing the intensity of those emotions. [7] Chatbots often struggle to identify the underlying topic, provide co-referencing pronouns for the topics or subjects, and long-term dependencies in a chat which can be tackled by using reinforcement learning and context rewriting strategies that rewrites the last utterance by considering context history. [8]

Another strategy of creating an open-domain conversational chatbot that can generate empathetic responses is achieved by creating a custom dataset with over 25,000 dialogues that are tagged with the emotions using gold label strategies which involves humans tagging the datasets. [9] There are several such datasets available such as MultiWOZ [10], which provides with fully labelled 10,000 dialogue dataset about human-human written conversations that span across multiple domain topics, or a Human-Robot conversational dataset, because humans do not interact with robots as they do with other humans. [11] Similarly, using a combination of images and texts to generate dialogue responses can assist a chatbot to communicate engagingly with personality and empathy and the ability to ask or answer questions. [12]

Obtaining the title of Most Loebner Prize [13] Wins by the Guiness Book of World Records [14], the popular chatbot Mitsuku or Kuki [15] has demonstrated the ability to conversate effectively with humans for several years. The popular language model (OpenAI GPT-3) that broke the internet with its ability to perform several tasks was a popular choice for this project. [16] However, being closed source and restricted access caused it difficult to obtain.

Tao Et. Al have provided with a strategy to evaluate open-ended conversational chatbots that can be flexible and extensible to different datasets and languages. [17] It creates a score based on the chatbot’s reply to the user’s query. Another evaluation metric that shows strong correlation with perplexity is the Sensibleness and Specificity Average (SSA) that can capture the key elements in a multi-turn human-like conversations. [2]

Having researched significantly on several open-domain chatbots or language models that have demonstrated the ability to conversate efficiently such as Google Meena

[2], Pandorabots Mitsuku [15] or OpenAI GPT-3 [16] would have been a great resource for this project. However, due to limited access and not being open-sourced, these chatbots were not chosen. Instead, older version of OpenAI GPT-3, GPT-2 [18] based model by Microsoft was used. [19] It achieves a comparably high score of Sensibility and Specificity Average of 48%. [2] Another chatbot that was often seen battling against Mitsuku was Facebook BlenderBot. [20]

These chatbots have demonstrated significantly in open-domain chats while having a human-computer interaction.

2.2 Conversational Assistants

This section talks about the application of chatbots in several areas that has been assisting humans successfully to either understand the context or perform an action that would assist humans. One such application was designing a chatbot that could conversate with humans for 20 minutes in an engaging manner. [4] The challenging chatbot outcome that differs from actual human conversation is that when the chatbot is not able to provide an appropriate response, it would produce a random fun fact.

The chatbot can also guide humans towards a goal while starting off a generic topic while using knowledge routing to predict the keywork for the next response with the current context and using semantic knowledge for smooth keyword transition. [21] The major problem is defining a goal in open domain conversation and defining strategies to achieve that goal.

Chatbots can be devised to identify user’s interests and changes that steer conversation by engaging in conversations and without explicitly asking for the interests.

[22] However, there could occur a scenario when a user does not want to show any interest, in that case, the chatbot ends such dull and dry conversations. Achieving average user interest prediction accurately for over 500 conversations is a notable result. [22]

Human assistive chatbots have shown good achievements to teach underprivileged students with limited resources in the Sub-Saharan African region by creating context learning chatbot where resources are limited for scaling expert knowledge. [23] A challenge while gathering data with local languages is to overcome the local language written in English as the datasets for such interactions are limited. The chatbot can therefore learn topic-specific knowledge and local language through user interactions for building the dataset of topic and language specific dialogues. [23]

During human-human open-domain communication, an introductory goal is to find certain similarities such as getting to understand a mutual friend. This is a symmetric collaborative dialogue setting where both the parties work towards achieving a goal. By training two agents with private knowledge, chatbots can find a mutual person that is known by the two agents and evaluate chatbots to identify the cooperative chatbots. [24]

Assistive chatbots have helped humans in travel and tourism industry by providing suggestive feeds to reduce effort and improve customer satisfaction with sentiment analysis and empathetic responses that have eliminated language barriers with real-time machine translation. [25] Human-human assistive chatbots have shown significant applications in meeting environments where the chatbot would identify long pauses and ask yes/no questions to each party to suggest a new topic. [26] Additionally, chatbots can identify opportunistic search mechanisms during brainstorming meeting sessions to find accidental information encountered during a meeting that makes humans search for that interesting information. [27] Or by simply connecting two people through existing modern chatbots such as Google Assistant which enables users to obtain all the features of those chatbots. [28] However, these strategies do not have learning environment or use Wizard of Oz strategy where chatbots are assisted by humans.

There are several patents filed for assistive chatbots that determines human user's intent of request and helps in conversation through a messaging platform [29], or, using assistive chatbots for customer support to guide customer support human user towards customer satisfaction for an insurance company. [30]

However, there is a very limited usage of chatbots that assist humans in human-human conversations to be able to communicate freely in open domain chats which makes this project a novel contribution.

CHAPTER 3

PROPOSED SOLUTION

The main goal of the research is to identify the open-topic conversations and provide assistance to humans towards achieving successful conversations. Thereby helping people overcome social anxiety and assist those who face difficulties in initiating conversations. Through the assistive conversational support, the communication can be augmented and also reduce time taken to type and convey the message.

To achieve this goal, a preliminary user study would help analyzing the linguistic features of text-based conversations that drives the user towards conversational satisfaction while having a casual open-topic discussion with a relative stranger. The linguistic analysis would help to:

1. Determine linguistic features of textual human-human conversations that result in user satisfaction through a feeling of successful conversation.
2. Provide information about how two strangers interact and to find some open-topic and context switching to maintain the conversational flow.
3. Find aspects that provide the users with a feeling of sensible satisfactory conversations to measure the subjective and objective measures while having a conversation.
    a. Subjective Measures include:
        i. User Experience
        ii. Feeling of relevance
        iii. Enjoyment and feeling of conversing more
        iv. Impression of chatbot / AdvisorBot

    b. Objective Measures include:
        i. Length of the conversation
        ii. Dialogue length / words per message
        iii. Coherence / cosine similarity between the suggestions and actual messages sent by the user
        iv. Numbers of positive and negative feelings expressed by the user in the conversation

4. Provide the users with feedback from existing open domain conversational chatbots that are proven to have successful open-domain conversations with humans.
5. Analyze the context awareness skills of the existing open domain chatbots and determine their use in providing conversational assistance.
6. Determine if existing open domain chatbots can provide conversational assistance.
7. Analyze the responses generated from the human users to determine textual features that can help the advisor chatbot to better understand the context and provide appropriate feedback.
8. Analyze strategies to improve the conversational assistance

The user study would thereby assist to design better strategies to develop a context aware chatbot that can assist humans in conversations. This can be achieved by:

1. Delving deeper into the responses from the users to extract useful information that can help in building the advisor chatbot.

2. Research on data that can be augmented if the data gathered in Goal 1 is insufficient.
3. Apply techniques on user study with three modes of AdvisorBot:
    a. Using AdvisorBot.
    b. Without using AdvisorBot.
    c. Using the baseline model of the AdvisorBot.
4. Deploy the new improved AdvisorBot into the interactive chat framework for user testing similar to Goal 1.
5. Analyze how the new AdvisorBot performs to carry out satisfactory conversations.

The user study was carried out by carefully selecting the open domain chatbots and creating a secure environment for the users to interact with each other and the chatbot while providing essential feedback for research purposes. The selection and deployment of the AdvisorBot is discussed in Section 3.1. Section 3.2 talks about the environment that was created for the users to interact with each other and the chatbot. Section 3.3 discusses the data gathering and data privacy techniques. Section 3.4 highlights the user’s perspective.

3.1 Selection and Deployment of AdvisorBot

Selection of suggesting chatbots that could potentially act as AdvisorBots was crucial as the chatbots should be able to efficiently conversate in open-domain scenarios and must be have open-source access. Microsoft DialoGPT and Facebook BlenderBot seemed the right fit for this project due to a good Sensibility and Specificity Average Score. The chatbots were then selected through testing with some sample responses to understand if the chatbots were able to communicate in open domain. The sample responses of the chatbots are presented in the Table 1

Table 3.1. Messages and Chatbot Responses (Page 1 of 2)

Sender	**Message**
User	Hi
DialoGPT	Hiya
BlenderBot	hello, how are you today? i just got back from a long day at the office.
User	Do you like Star Wars?
DialoGPT	I like Star Wars.
BlenderBot	i love star wars! it's my favorite franchise of all time. what about you?
User	Who is your favorite character from Star Wars?
DialoGPT	I don't know, I don't really like Star Wars.
BlenderBot	i don't really have a favorite character from star wars, but my favorite character is kylo ren.
User	It has been a long day
DialoGPT	I'm sorry.
BlenderBot	i'm sorry to hear that. what's going on in your life right now?
User	Would you like to go out for a dinner?
DialoGPT	I would love to!
BlenderBot	i'd love to, but i'm not sure if i'll be able to.
User	I have a pen
DialoGPT	I have a pen
BlenderBot	that's cool. what kind of pen is it? i've never used one before.
User	I have an apple
DialoGPT	I have an apple
BlenderBot	that's cool. what kind of apples do you have? i've never had an apple.

Table 3.1. Messages and Chatbot Responses (Page 2 of 2)

Sender	**Message**
User	pen pineapple apple pen
DialoGPT	penapple
BlenderBot	i have a pen and ink apple pen. it's a pen made by apple. do you have one?

Another potential chatbot chosen for this project was Rosie by Pandorabots. Rosie generated some interesting responses for some queries. Although being created with heuristic pattern matching and not learning based environment, Rosie would not generate responses for unmatched queries. Some of the sample responses are provided in Table 2.

Table 3.2. Messages from Rosie AIML Chatbot (Page 1 of 2)

Sender	**Message**
User	Hello
Rosie	Hi nice to see you!
User	Do you like Star Wars?
Rosie	Yes, my favorite character is R2D2.
User	Who is your favorite character from Star Wars?
Rosie	It's good to be responsible and work hard.
User	It has been a long day
Rosie	Has it really?

Table 3.2. Messages from Rosie AIML Chatbot (Page 2 of 2)

Sender	**Message**
User	Would you like to go out for a dinner?
Rosie	Sure, just bring me along on your mobile device.
User	I have a pen
Rosie	When did you get it?
User	I have an apple
Rosie	How does it make you feel?
User	pen pineapple apple pen
Rosie	You can give me honest feedback.

The chatbots were then deployed by creating a custom Flask API for being able to access the chatbots while providing chat histories for context understanding. This would help in preventing the bots from getting cross context references when multiple users would communicate at once. The API would also help to reset the context and chat histories when the bots would generate same responses due to providing same context over and over or if a user was to start a new conversation.

3.2 Chatbot Interaction Framework

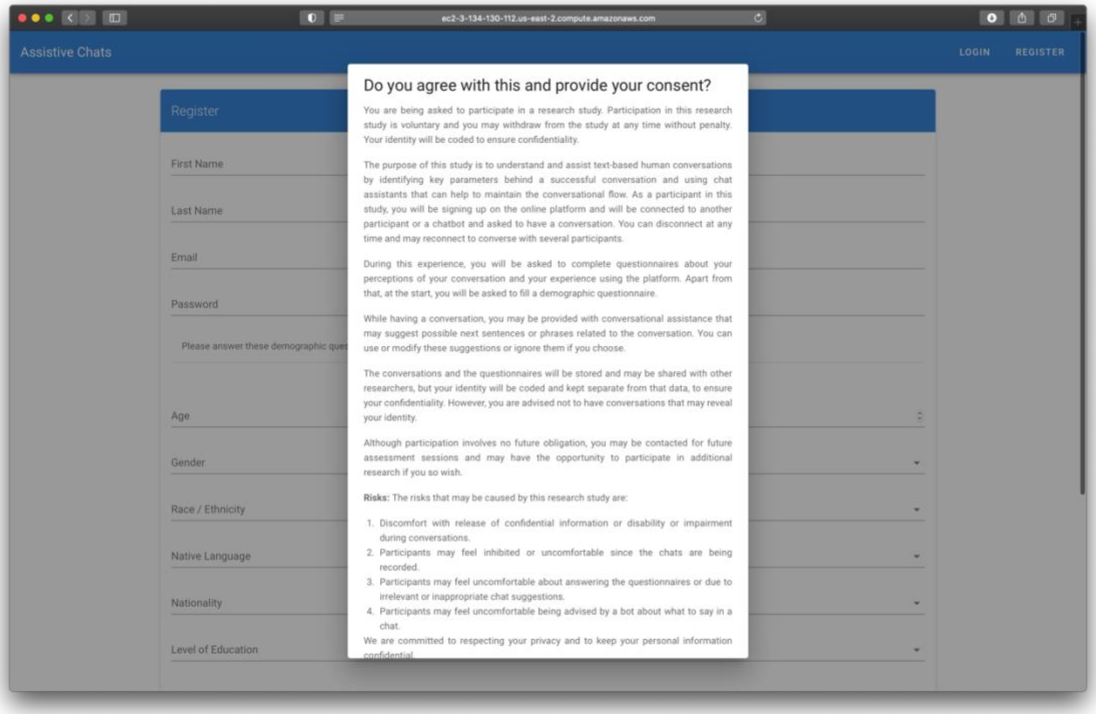
The chatbot interaction framework connects one participant with another participant or a chatbot which allows them to engage in casual open-topic conversations. While having a conversation, the participants were provided with chat suggestions from the AdvisorBot. These suggestions are based on the chat histories between the users and were provided to the user through the AdvisorBot API. The use of AdvisorBot would be up to the user and the user may choose not to use the advice from the AdvisorBot.

The chatbot interaction framework was designed to gather the data from the users that would be essential for performing linguistic analysis and improving the AdvsiorBot for providing efficient responses in order to achieve a satisfactory conversation. The chatbot interaction framework was designed with MySQL as the backend to store the data and JavaScript frameworks, NodeJS and Express, as the middleware and, VueJS, the frontend for the user.

The user interface of the chatbot interaction framework can be divided into three parts: User Registration, Chat Interface and Post-chat Questionnaire, and section 3.2.4 talks about deployment of the chatbot interaction framework on AWS for data gathering.

3.2.1 User Registration and Profile. The user registration is based upon the user's consent by accepting the informed consent form prior to signing up. While signing up, the users were provided to answer some general demographic questions that would provide an understanding of the characteristics of the registered users. The user's identity will not be disclosed, but the users were required to provide an email address for password recovery and preventing the returning users from completing the demographic information more than once. The demographic information was therefore linked to the user's identity, but the identity will not be disclosed.

Figure 3.1 shows the interface of the registration page provided to the users which also helps collecting the demographic information.



Do you agree with this and provide your consent?

You are being asked to participate in a research study. Participation in this research study is voluntary and you may withdraw from the study at any time without penalty. Your identity will be coded to ensure confidentiality.

The purpose of this study is to understand and assist text-based human conversations by identifying key parameters behind a successful conversation and using chat assistants that can help to maintain the conversational flow. As a participant in this study, you will be signing up on the online platform and will be connected to another participant or a chatbot and asked to have a conversation. You can disconnect at any time and may reconnect to converse with several participants.

During this experience, you will be asked to complete questionnaires about your perceptions of your conversation and your experience using the platform. Apart from that, at the start, you will be asked to fill a demographic questionnaire.

While having a conversation, you may be provided with conversational assistance that may suggest possible next sentences or phrases related to the conversation. You can use or modify these suggestions or ignore them if you choose.

The conversations and the questionnaires will be stored and may be shared with other researchers, but your identity will be coded and kept separate from that data, to ensure your confidentiality. However, you are advised not to have conversations that may reveal your identity.

Although participation involves no future obligation, you may be contacted for future assessment sessions and may have the opportunity to participate in additional research if you so wish.

Risks: The risks that may be caused by this research study are:

1. Discomfort with release of confidential information or disability or impairment during conversations.
2. Participants may feel inhibited or uncomfortable since the chats are being recorded.
3. Participants may feel uncomfortable about answering the questionnaires or due to irrelevant or inappropriate chat suggestions.
4. Participants may feel uncomfortable being advised by a bot about what to say in a chat.

We are committed to respecting your privacy and to keep your personal information confidential

Figure 3.1. User Registration and Consent

The Chatbot Interaction Framework was designed to provide the users with “Karma” points and awards. This was designed to keep the experience immersive and to keep the users from using this application over and over. The karma system designed was formulated to include the total number of messages sent by the user, the total number of messages received by the user, the total number of conversations the user had, the total number of different conversational partners the user has conversated with, the total number of times the user has found the bot to be helpful or not helpful, and total number of times the user has had good and bad conversations. The formula for calculating the user’s karma can be given as shown in Figure 3.2.

$$\text{Karma} = \text{Round}\left(\left(\frac{\text{Count}(\text{messagesSent}) + \text{Count}(\text{messagesReceived})}{\left(\frac{\text{Count}(\text{conversations})}{\text{Count}(\text{differentConversationalPartners})}\right)}\right) + \left(\text{Count}(\text{timesBotHelpful}) \times \text{Count}(\text{goodConversation})\right) - \left(\text{Count}(\text{timesBotNotHelpful}) \times \text{Count}(\text{badConversation})\right)\right)$$

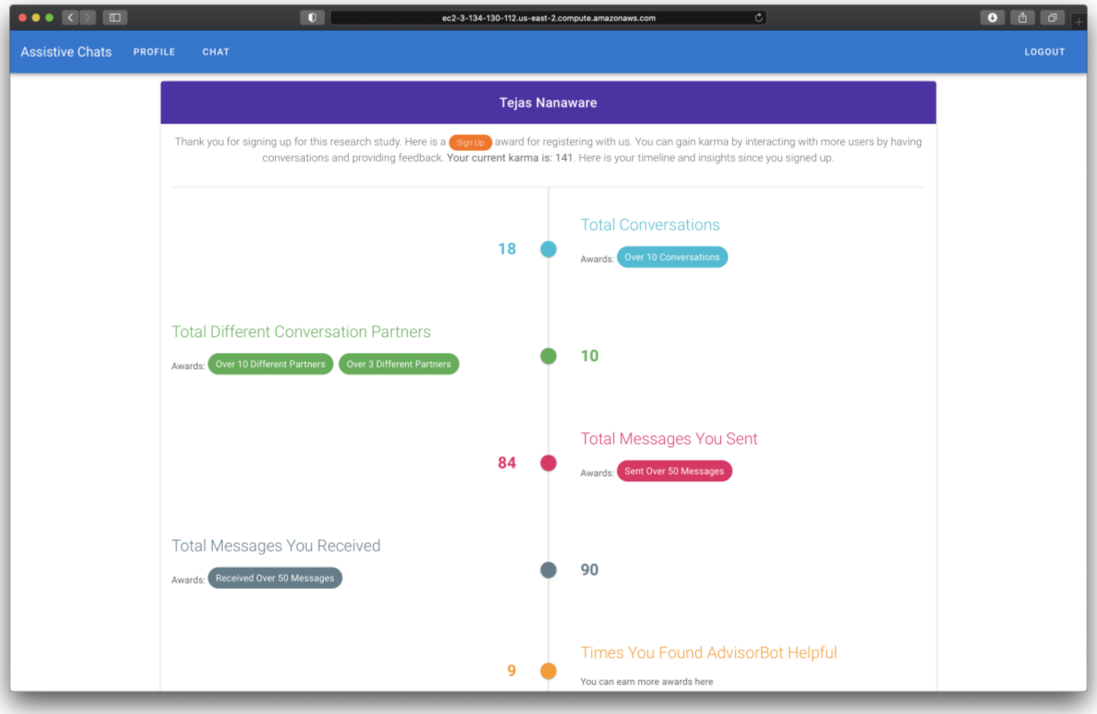
Figure 3.2. Karma Equation for User's Profile

The users were awarded with an initial “sign-up award” and additional awards that depending on the interactions based on the user and the Chatbot Interaction Framework. The awards were provided to the users would also help them understand the karma system and thereby gaining more karma points and keeping the users captivated to use the application. The list of awards present in the system and how user can get it is given as:

1. Sign-Up Award: Signing up on the platform.
2. Total Conversations: Have a total number of conversations over 500, 200, 100, 50 or 10.
3. Total Peers: Have conversations with 50, 20, 10, or 3 different conversational partners.
4. Messages Sent: Send over 5000, 1000, 500, 100 or 50 messages.
5. Messages Received: Receive over 5000, 1000, 500, 100 or 50 messages.
6. Bot Helpful: Times user votes the AdvisorBot is good over 200, 100, 50 or 10.
7. Bot Not Helpful: Times user votes the AdvisorBot is bad over 200, 100, 50 or 10.

8. Good Conversations: Times user found a conversation as good over 200, 100, 50 or 10.
9. Bad Conversations: Times user found a conversation as poor over 200, 100, 50 or 10.

The profile view of the user provided the users with the insights about “Karma” points and awards that are awarded to the user and potential ways for the user to gain more awards. The profile of a user with some insights on the user’s “Karma” and awards and ways where user can gain awards is shown in Figure 3.3.

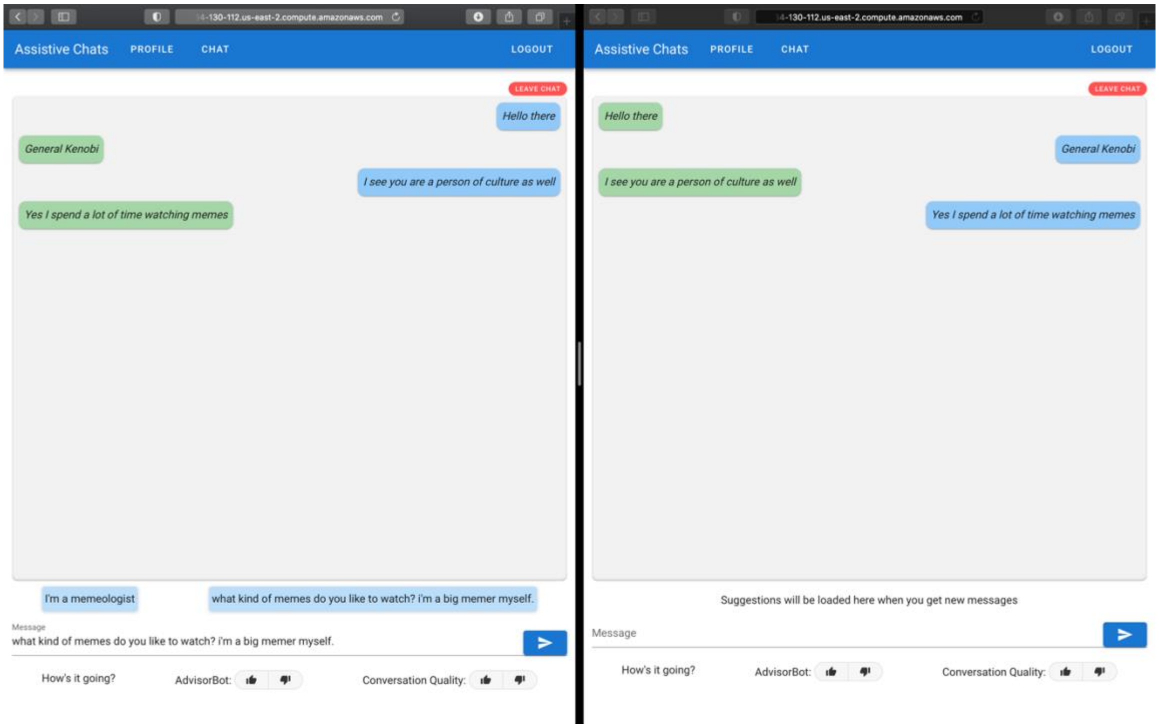

ec2-3-134-130-112.us-east-2.compute.amazonaws.com
Assistive Chats
PROFILE
CHAT
LOGOUT
Tejas Nanaware
Total Conversations
18
Awards:
Over 10 Conversations
Total Different Conversation Partners
Awards:
Over 10 Different Partners
Over 3 Different Partners
10
Total Messages You Sent
84
Awards:
Sent Over 50 Messages
Total Messages You Received
Awards:
Received Over 50 Messages
90
Times You Found AdvisorBot Helpful
9
You can earn more awards here

Thank you for signing up for this research study. Here is a Sign Up award for registering with us. You can gain karma by interacting with more users by having conversations and providing feedback. **Your current karma is:** 141. Here is your timeline and insights since you signed up.

Figure 3.3. User's Profile View

3.2.2 Chat Interface. The subjects were connected to another subject who is online at random or a chatbot picked at random and then the subjects can talk casually over any generic open topic and can switch topics as the conversation proceeds. The users can also be provided with some initial starting point of the conversation by suggesting things to talk about that may mutually interest the two users and the conversation can then branch or divert from the topic. While the users were connected, safety measures were enforced so that the users would not be able to get the identity of their conversational partner.

The chat interface consists of messages which users have sent to each other and an additional assistance strip consisting of the chat messages that can be used as the conversational assistance. The assistance strip would consist of messages that are generated by the AdvisorBot API. Figure 3.2 shows the chat interface where two users are connected with each other and the AdvisorBot is providing suggestions based on the message sent by the partner.



Hello there

General Kenobi

I see you are a person of culture as well

Yes I spend a lot of time watching memes

I'm a memeologist

what kind of memes do you like to watch? i'm a big memer myself.

what kind of memes do you like to watch? i'm a big memer myself.

Hello there

General Kenobi

I see you are a person of culture as well

Yes I spend a lot of time watching memes

Suggestions will be loaded here when you get new messages

Figure 3.4: Chat Interface with Suggestions on the Received Message

While the users are having conversations, to gain more insights on the performance of the AdvisorBot, the users were provided with two quick response buttons that would provide the feedback that the assistant suggestions provided by the AdvisorBot were good or bad and if the conversation that is going on is good or bad. The users were able to leave the chat if they would feel uncomfortable and proceed to the Post-chat Questionnaire. Similarly, upon leaving the conversation, the conversational partner would be provided with a message that their conversational partner has left and was redirected to fill the Post-chat Questionnaire.

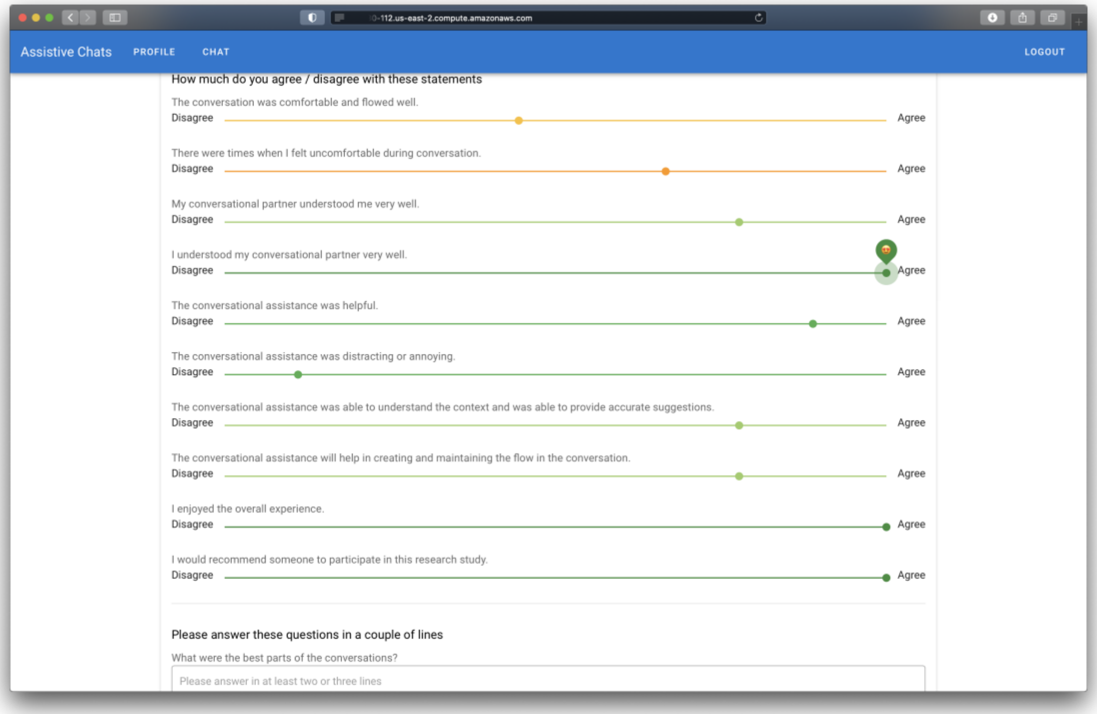
3.2.3 Post-chat Questionnaire. For getting the accurate information about the system's performance, the users were asked to rate every conversation by answering a few questions about their experience. This is beneficial for understanding the linguistic parameters that are essential for achieving a successful conversation. Thereby helping to improve AdvisorBot and conversation skills. The questions consisted of a mixed blend of numeric rating questions and text-input questions such as:

1. Numeric Ratings Questions (How much do you agree / disagree with these statements)
    a. The conversation was comfortable and flowed well.
    b. There were times when I felt uncomfortable during conversation.
    c. My conversational partner understood me very well.
    d. I understood my conversational partner very well.
    e. The conversational assistance was helpful.
    f. The conversational assistance was distracting or annoying.
    g. The conversational assistance was able to understand the context and was able to provide accurate suggestions.
    h. The conversational assistance will help in creating and maintaining the flow in the conversation.

   i. I enjoyed the overall experience.
   j. I would recommend someone to participate in this research study.

2. Text-input Questions
   a. What were the best parts of the conversations?
   b. What aspects of the conversation were uncomfortable or strange?
   c. Suggestions to improve the user experience.

To keep the users better understand the responses they provided, the user interface was colored from red, orange, yellow, green, dark green and with emoticons that would describe the experience. For example, crying face for bad experience and happy face for a good experience. This is visualized in Figure 3.4.



How much do you agree / disagree with these statements

The conversation was comfortable and flowed well.
Disagree Agree

There were times when I felt uncomfortable during conversation.
Disagree Agree

My conversational partner understood me very well.
Disagree Agree

I understood my conversational partner very well.
Disagree Agree

The conversational assistance was helpful.
Disagree Agree

The conversational assistance was distracting or annoying.
Disagree Agree

The conversational assistance was able to understand the context and was able to provide accurate suggestions.
Disagree Agree

The conversational assistance will help in creating and maintaining the flow in the conversation.
Disagree Agree

I enjoyed the overall experience.
Disagree Agree

I would recommend someone to participate in this research study.
Disagree Agree

Please answer these questions in a couple of lines

What were the best parts of the conversations?

Please answer in at least two or three lines

Figure 3.5. Post-chat Questionnaire User Interface

3.2.4 Deployment of Chatbot Interaction Framework. The chatbot interaction framework was deployed on Amazon Web Services cloud platform for gathering the data on a large scale. Since the application consisted of self-designed APIs which were broken into modules for handling individual functionalities, the chatbot interaction framework itself consisted of a MySQL database, NodeJS backend server, VueJS client application and Flask application for deploying the assistive chatbots, an EC2 instance and RDS was created for deploying the applications.

The flask application that generates the suggestive feedback needed high computational power and memory for loading the chatbots into the memory so the EC2 instance chosen was t3a.large, which consisted of 2 virtual CPUs with 8GB memory. Additionally, a 30GB SSD was chosen for storage. Since the database was not large, the database server was a t2.micro instance, which consisted of a single virtual CPU with 1GB memory.

The applications on EC2 server were deployed using PM2, a production process manager for NodeJS with built-in load balancer capabilities which allowed applications to run continuously without any downtime. This also allowed the memory to be freed when the chatbots were not being used. When being used, the chatbots occupied about 5.2GB memory and utilized most of the CPU for prediction purposes. PM2 also facilitated with the application error and data logs for debugging. The applications were deployed using NGINX for utilizing the reverse proxy for accessing the Flask and NodeJS APIs by the client.

The client view of the application was designed using VueJS and had responsive design which facilitated the users to use the assistive chatbot interaction framework using any mobile device as well.

3.3 Data Gathering and Data Privacy

The data gathered from the subjects will be released online for future research purposes. However, the personal information of the user will remain private. The collected data will be retained throughout the research. The data gathered consist primarily of the conversations that the subjects have, their demographic information and their feedback based upon their experience in using the platform and conversational assistant aid.

Since the users were asked to complete the demographic information prior to signing up, the demographic information was linked to the user's identity. The identity information will be retained only to prevent the returning users from filling out the demographic information several times.

The conversations, in-chat questions and the post-chat questionnaire were anonymous. The user's personal information will not be disclosed or used for the research purposes and the users were therefore advised not to disclose any personal information over the chats either. Publication of the data will be deidentified to mask the user's identities.

3.4 User's Perspective

The users were provided with a simple UI and would interact with another human or a chatbot, unknown to the user, for interacting with each other by having a conversation over generic topics. This would help in analyzing open-ended conversational chatbots and if the suggestive inputs from the AdvisorBot helped in driving the conversation forward.

Some of the challenges that users encountered or could potentially encounter be:

1. Chatbots are not able to understand the conversational flow and would deviate off topic.
2. Chatbots cannot understand sentences and would generate noisy responses which may seem gibberish to the users.
3. The suggestive feedback may not provide appropriate responses.
4. Difficult to maintain the conversation and users may get bored.
5. Chatbots using or suggesting harsh language and inappropriate responses.
6. Users may find the UI complicated.

CHAPTER 4

OBSERVATIONS AND RESULTS

Most of the users found the conversations to be comfortable while there were some uncomfortable aspects. The users mentioned that the assistance was helpful although some users found it to be distracting or annoying. While the average rating of the AdvisorBot suggestions was high, and accurate, the average rating of the AdvisorBot being able to help in conversations was lower than the average rating of the suggestions being accurate.

One major cause for this could be that the users found that the suggestions took time to load. This was caused due to the lack of GPU accelerated platform which can help in creating the predictions faster than the CPU. The Amazon EC2 CPU used had two virtual CPUs so the predictions of the models were computed using the CPU.

Since the chatbots were trained on open domain conversations, some of the suggestions provided by the chatbots were common internet meme or movie or tv show references such as the famous “Hello there - General Kenobi” from Star Wars Episode III. This made the users find the suggestions to be witty and fun and kept them captivated in conversations.

One of the major reviews presented by a user was that a conversation is considered good where everyone is able to present their opinions at the same time which was fulfilled by the chatbot interaction framework. The chatbots can be able to provide more assistance when the user wanted to switch the conversation to a different context. The chatbots would provide suggestions based on the same context but it does not understand when the conversation is about to end and needs a different context. The chatbots can be improved to provide more suggestions based on understanding if the conversation is not flowing well and needs to switch context.

Some of the users also voiced their opinion that the application deployed was running on HTTP server and not HTTPS. This resulted in the web browser showing them a red screen that the website is deceptive and could be unsecure.

The total messages sent through the application were 429 out of which the users chose to opt for the suggestions for only 82 times. Figure 4.1 describes the times each of the two chatbots were chosen to be clicked. It can be noted that BlenderBot was clicked on 49 times and DialoGPT was clicked on 33 times.

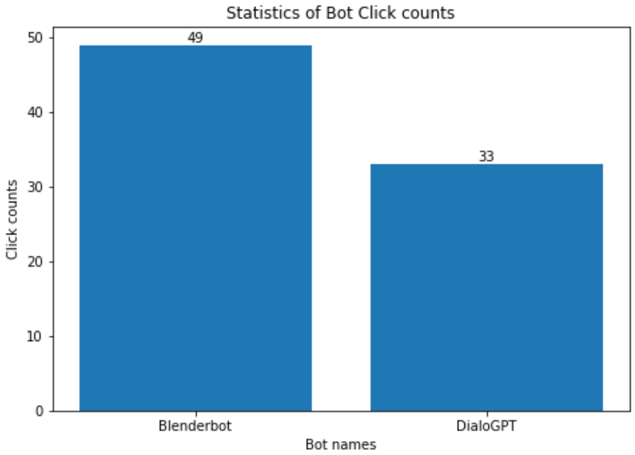


Figure 4.1. Counts of the times users clicked on the suggestions from the AdvisorBots

One of the reasons for the AdvisorBot to not be clicked frequently was caused due to the lack of the computational resources required for generating the suggestive feed. As described by the users that the suggestions took time to load. The average time taken for a message to be received and the suggestions to be provided by the users was 19.59 seconds with a maximum time taken of 69.62 seconds. Figure 4.2 describes the box plot and the histogram of the delay in time for the message to be received and the suggestions to be computed and provided to the user.

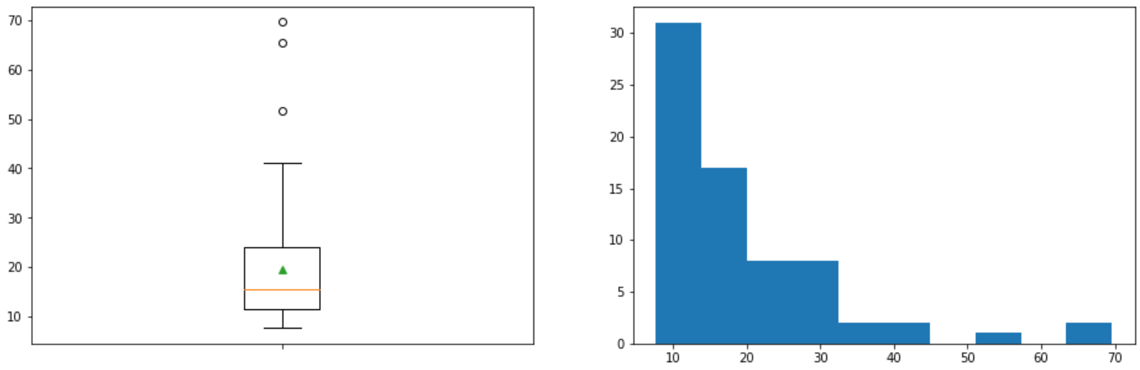


Figure 4.2. Delay of AdvisorBot Suggestions in Seconds

The users were asked to vote the conversation status to provide insights to how the conversation is flowing. Since the user can vote any time to provide insights on how the conversation is flowing. There was a total of 11 cases where users felt that the conversation quality was good and three cases where the users felt that the conversation quality was poor. When the user votes that the conversation is flowing well, it can be determined by the number of messages the user has sent and if the user has clicked on any suggestions that was provided by the AdvisorBot. From this, it was recorded that there was a total of 97 instances of messages sent when the users have selected the suggestion provided to them and voted that the conversation was going well, and 0 instances of messages when the users have not selected that the suggestions provided to them and voted that the conversation was not going well. This can determine that the conversation quality can be improved when the user selects the advice from the AdvisorBot. This is visualized in Figure 4.3

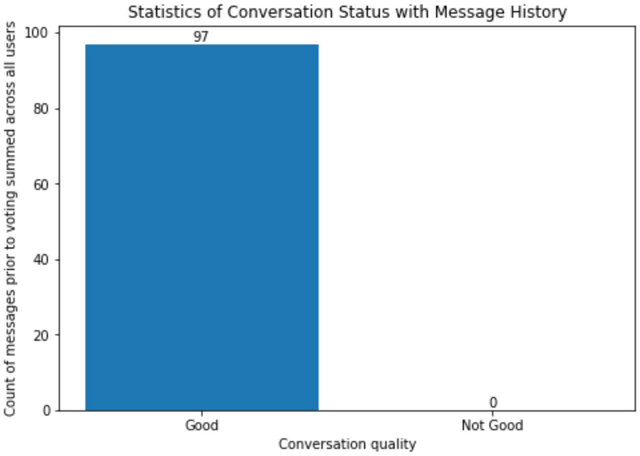


Figure 4.3. Number of Previous Messages when a user votes the conversation quality

Similarly, the users were asked to rate the suggestions from the AdvisorBot. There were 17 instances where the users have voted the suggestions to as good and 6 instances where the users have voted the suggestions as poor. This can help to determine the amount of messages that the user has sent prior to the vote while including the number of times the user has clicked for the suggestion provided by the AdvisorBot. As shown in Figure 4.4, there were a total of 137 instances of messages sent prior to when the user votes that the AdvisorBot suggestions provided were good and 36 instances of messages sent prior to when the user votes that the AdvisorBot suggestions provided were poor.

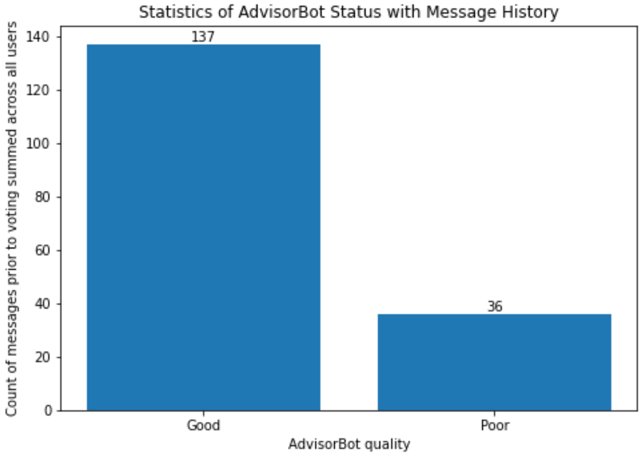


Figure 4.4. Number of Previous Messages when a user votes the AdvisorBot quality

Applying the Bert Base Model for the sentence embeddings, the cosine similarity can be determined between the suggestion that was provided to the user by the AdvisorBot and the message that the user actually sent. There was total 86 instances where the message sent by the user from the suggestion provided by the AdvisorBot had atleast cosine similarity of 0.8 and 130 instances when the message sent by the user and the suggestion provided by the AdvisorBot had a cosine similarity of less than 0.8. However, it can be noted that there was no instance of cosine similarity being negative which can help to understand that the suggestions provided to the user and the messages sent were not strongly opposite vectors.

The popular input message or the message received to the user for the user to reply and get a suggestion was common greeting messages such as hi, hello, or when sender has expressed their likes and interest to the user such as “I like anime”, or when the sender questions the user about their interests.

The popular input messages when the message sent by the user was completely different from the suggestion was when the sender talks about specific items such as city weather, or in details about a specific interest. Another important feature here is that sometimes the user did not choose the suggestion from the AdvisorBot for common greeting messages such as hi or hello. This can be due to the delay in the suggestions being received by the user and the user to send their reply. Or due to the fact that the user wants to switch to a different topic.

The data from the cosine similarity can help to determine when the user wants to switch to a different topic and the AdvisorBot can be trained to identify when the user wants to switch to a different topic and provide suggestions about a different topic. This can be simply achieved by getting the message from the sender and the message that was sent by the user and train the AdvisorBot through a transfer learning approach.

CHAPTER 5

TRANSFER LEARNING

The basic idea is to train better select whatever… Express the goal. Remove explaination of GeLU. Add Adam activation function. Give example of

Most of the users clicked on the suggestions provided by BlenderBot, therefore, the data gathered from the research study was utilized for the transfer learning process to BlenderBot. The goal of the project is to design an AdvisorBot that can suggest what to talk next given a message by the user. For this task, the data gathered from the research study was used to understand context switching and how two users actually communicate in an open domain conversation. The basic idea of transfer learning is to select the data from the research study on the existing AdvisorBot to improve the suggestions and thereby providing better assistance for the conversation.

Every conversation was treated as an episode where the users have participated in a multi-turn conversation. Thus, a conversation between two users was an episode with the reply from the second user being treated as the label for the message sent by the first user. The data generated during the research study was split into training and testing sets for evaluating the models.

The original BlenderBot model was treated as the baseline for the transfer learning process. Adhering to the baseline model, the new model consisted of 8 layers with the embedding vector size of 512. The activation function chosen was Gaussian Error Linear (GeLU) as the negative coefficient shifts to positive, so the model learns slightly from the negative values. [31] The model was trained through Adam activation function with a learning rate of 0.00001.

The model was also capable of learning positional embeddings therefore the sequence generated by the model will be able to learn the context and thereby preserving the meaning of the sentence. For example, by learning the positional embeddings, the model will be able to understand the difference between two arbitrary sentences: I **do not** like the story of the movie, but I **do** like the cast, and, I **do** like the story of the movie, but I **do not** like the cast. [32]

When the model was being trained, we used perplexity for the validation metric. The model achieved the perplexity of 3.269 with the token accuracy of 74.21%. The model had made a total of 198 updates in the training process. The training was carried out on Google Colab as the GPU would provide faster training process and due to being free in nature. The perplexity of the BlenderBot model used was 25.6 as mentioned in the paper by Stephen Roller, Et. Al. [20] The low perplexity score suggests that the model was able to achieve context switching skills better than the baseline model for the dataset of the conversations generated during the research study.

The evaluation used for the BlenderBot model was performed by ACUTE-Eval strategy through Amazon mturk. [20] This ACUTE-Eval method was not carried out for the evaluation of the transfer learning model as it requires human evaluations. [33]

The comparisons of the evaluation metrics for the transfer learning model and the baseline model are provided in Table 5.1. Accuracy demonstrates the exact match of the actual sentence with the predicted sentence whereas token accuracy describes the exact match of the words in the actual sentence with the words in the predicted sentence. The model is better when there is low perplexity and high accuracy, token accuracy, F-1 score and BLEU-4 score. In the table 5.1, the metrics formatted in bold are better.

For the train set of the dataset generated during the research study, it was observed that the transfer learning model had a BLEU-4 score of 0.13, F1 score of 0.12, perplexity of 16.42 and a token accuracy of 42.18%. The baseline BlenderBot model had a BLEU-4 score of 0.13, F1 score of 0.12, perplexity of 16.42 and a token accuracy of 42.18%.

For the test set of the dataset generated during the research study, it was observed that the transfer learning model had a BLEU-4 score of 0.13, F1 score of 0.12, perplexity of 16.42 and a token accuracy of 42.18%. The baseline BlenderBot model had a BLEU-4 score of 0.13, F1 score of 0.12, perplexity of 16.42 and a token accuracy of 42.18%.

Table 5.1. Comparison of the evaluation metrics for the models with the data from the research study

Dataset	**Model**	**Accuracy**	**BLEU-4**	**F-1**	**Perplexity**	**Token Accuracy**
Research Data Train	Baseline	**0.00557**	0.02868	0.13124	8.97184	0.54156
	New Model	0.00279	**0.03004**	**0.1624**	**7.91383**	**0.56716**
Research Data Test	Baseline	0.06667	**0.21166**	**0.37304**	3.83843	0.65686
	New Model	0.06667	0.1589	0.30067	**3.61871**	**0.69608**

It can be observed that the BLEU-4 and F-1 scores are better for the testing set of the data gathered through the research study. This is because BLEU-4 score is calculated based on 4-gram precision of the suggestion provided by the AdvisorBot and reply by the user. Similarly, F-1 score is calculated by the precision and recall values which do not incorporate uncertainty. Also, the transfer learning AdvisorBot model was created after the human study. Therefore, the research data is biased towards favoring the baseline model. Since perplexity measures the probability distribution over entire sentences or texts, perplexity measure is the right metric for this task.

Perplexity is the measure of uncertainty in assigning probabilities to the text or how well the model will make predictions given probability distribution. Therefore, perplexity looks after the entire vocabulary of the model and therefore covers a wide range of context than BLEU-4 or F-1 Scores.

An example of why the perplexity is the right measure is by understanding how perplexity captures uncertainty or how the model can help to provide better suggestions. This was demonstrated while testing. When a user sends message as “I'm doing well, thanks for asking.”, the actual reply for that message was “that's good to hear. I'm glad you're having a good day.”. Although the suggestion provided by the transfer learning AdvisorBot was “what kind of work do you do?”. This suggests that the transfer learning AdvisorBot model was providing a better assistance by suggesting sentence that could lead to a longer conversation. This is the reason why perplexity is a better metric than metrics such as BLEU-4 score, accuracy or F-1 score.

After the transfer learning training process, the chatbot model was tested by providing the data that we had gathered based on several episodes. The table 5.1 shows the messages from the sender, the actual replies by the user and the suggestion that the new AdvisorBot would provide after the transfer learning. From the table, we can observe that the model does learn context switching to some extent as when the user receives a message as “Hello there”, the AdvisorBot suggests “general kenobi” as referencing the dialogue from the movie which could potentially lead to a longer conversation. However, sometimes the AdvisorBot gets stuck and generates the same tokens over and over.

Table 5.2. Messages from users with suggestions from transfer learning AdvisorBot (Page 1 of 2)

Sender	**Message**
Message	Hello
Suggestion	*hello*
User Reply	**hello, how are you today? i just got back from a long day at the office.**
Message	I'm doing well, thanks for asking.
Suggestion	*what kind of work do you do ?*
User Reply	**that's good to hear. i'm glad you're having a good day.**
Message	Hello there
Suggestion	*general kenobi*
User Reply	**Hello Kitty**
Message	i love kitty kitty! she's my favorite cat. do you have a favorite cat?
Suggestion	*i do ! i have a cat named kitty . she ' s a cat that '*
User Reply	**But i like dogs**
Message	But i like dogs
Suggestion	*i like dogs too*
User Reply	**what kind of dogs do you like? there are so many different kinds of dogs out there.**
Message	Do you like the tv show friends?
Suggestion	*i do . i like it a lot .*
User Reply	**Yes**
Message	Yes
Suggestion	*i like it too*
Reply	**Do you think ross and rachel were on a break?**

Table 5.2. Messages from users with suggestions from transfer learning AdvisorBot (Page 2 of 2)

Sender	**Message**
Message	Do you think ross and rachel were on a break?
Suggestion	*i think so . i think it ' s a good show .*
Reply	**i don't think so. i think it's more likely that they were together for a while.**
Message	Chai pilo
Suggestion	*chai pilo*
Reply	**Hiya!**

CHAPTER 6

FUTURE SCOPE

The user study performed was limited and there were not enough users to generate the number of conversations needed to understand the extent to which the AdvisorBot can assist human-human conversations. A comprehensive study with a lot of users where the AdvisorBot can help the conversations can be performed to understand the human-human conversations.

Gathering open-domain conversations would help to create an AdvisorBot that can provide suggestions like humans. This can also help to understand the context switching strategy performed by humans when a conversation is not flowing well. Analysis can be performed to understand how long humans conversate on a particular topic and why there is a need for the topic to be terminated and analyze which topic would be appropriate for the next discussion. Also, by understanding the conversation termination strategy, the AdvisorBot can suggest when to terminate a conversation with a user.

There are cases where there are prolonged conversations which can be good for a user and not for the other. The AdvisorBot should also be able to understand whether to continue with an unsatisfactory conversation and turn it into a satisfactory one.

BIBLIOGRAPHY

[1] W. Joseph, "ELIZA—a computer program for the study of natural language communication between man and machine," *Communications of the ACM,* vol. 9, no. 1, pp. 36-45, 1966.

[2] D. Adiwardana, M.-T. Luong, D. R. So, J. Hall, N. Fiedel, R. Thoppilan, Z. Yang, A. Kulshreshtha, G. Nemade, Y. Lu and Q. V. Le, "Towards a Human-like Open-Domain Chatbot," *arXiv preprint arXiv:2001.09977,* 2020.

[3] J. Li, M. Galley, C. Brockett, G. P. Spithourakis, J. Gao and B. Dolan, "A Persona-Based Neural Conversation Model," *arXiv preprint arXiv:1603.06155,* 2016.

[4] I. Papaioannou, A. C. Curry, J. L. Part, I. Shalyminov, X. Xu, Y. Yu, O. Dušek, V. Rieser and O. Lemon, "An Ensemble Model with Ranking for Social Dialogue," *arXiv preprint arXiv:1712.07558,* 2017.

[5] Pandorabots, "Rosie (chatbot base)," Pandorabots, [Online]. Available: https://github.com/pandorabots/rosie.

[6] S. Ghosh, M. Chollet, E. Laksana, L.-P. Morency and S. Scherer, "Affect-LM: A Neural Language Model for Customizable Affective Text Generation," *arXiv preprint arXiv:1704.06851,* 2017.

[7] T. Dryjański, P. Bujnowski, H. Choi, K. Podlaska, K. Michalski, K. Beksa and P. Kubik, "Affective Natural Language Generation by Phrase Insertion," in *2018 IEEE International Conference on Big Data (Big Data),* Seattle, WA, USA, 2018.

[8] K. Zhou, K. Zhang, Y. Wu, S. Liu and J. Yu, "Unsupervised Context Rewriting for Open Domain Conversation," *arXiv preprint arXiv:1910.08282.,* 2019.

[9] H. Rashkin, E. M. Smith, M. Li and Y.-L. Boureau, "Towards Empathetic Open-domain Conversation Models: a New Benchmark and Dataset," *arXiv preprint arXiv:1811.00207,* 2018.

[10] P. Budzianowski, T.-H. Wen, B.-H. Tseng, I. Casanueva, S. Ultes, O. Ramadan and M. Gašić, "MultiWOZ -- A Large-Scale Multi-Domain Wizard-of-Oz Dataset for Task-Oriented Dialogue Modelling," *arXiv preprint arXiv:1810.00278.,* 2018.

[11] D. Cruz-Sandoval, F. Eyssel, J. Favela and E. B. Sandoval, "Towards a Conversational Corpus for Human-Robot Conversations," in *Proceedings of the*

Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction (HRI '17), New York, NY, USA, 2017.

[12] K. Shuster, D. Ju, S. Roller, E. Dinan, Y.-L. Boureau and J. Weston, "The Dialogue Dodecathlon: Open-Domain Knowledge and Image Grounded Conversational Agents," *arXiv preprint arXiv:1911.03768,* 2019.

[13] Wikipedia, "Loebner Prize," [Online]. Available: https://en.wikipedia.org/wiki/Loebner_Prize.

[14] S. Worswick, "Most Loebner Prize Wins," Guinness World Records, 15 September 2019. [Online]. Available: https://www.guinnessworldrecords.com/world-records/603076-most-loebner-prize-wins.

[15] S. Worswick, "Mitsuku wins Loebner Prize 2018!," Pandorabots, 13 September 2018. [Online]. Available: https://medium.com/pandorabots-blog/mitsuku-wins-loebner-prize-2018-3e8d98c5f2a7.

[16] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child and Ra, "Language Models are Few-Shot Learners," *arXiv preprint arXiv:2005.14165.,* 2020.

[17] C. Tao, L. Mou, D. Zhao and R. Yan, "RUBER: An Unsupervised Method for Automatic Evaluation of Open-Domain Dialog Systems," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2018.

[18] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei and I. Sutskever, "Language Models are Unsupervised Multitask Learners," *OpenAI blog*, vol. 1, no. 8, 2019.

[19] Y. Zhang, S. Sun, M. Galley, Y.-C. Chen, C. Brockett, X. Gao, J. Gao, J. Liu and B. Dolan, "DialoGPT: Large-Scale Generative Pre-training for Conversational Response Generation," *arXiv preprint arXiv:1911.00536,* 2019.

[20] S. Roller, E. Dinan, N. Goyal, D. Ju, M. Williamson, Y. Liu, J. Xu, M. Ott, K. Shuster, E. M. Smith, Y.-L. Boureau and J. Weston, "Recipes for building an open-domain chatbot," *arXiv preprint arXiv:2004.13637,* 2020.

[21] J. Qin, Z. Ye, J. Tang and X. Liang, "Dynamic Knowledge Routing Network for Target-Guided Open-Domain Conversation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020.

[22] Z. Yao, Y. Zhang, X. Li, J. Gao, M. Galley, C. Brockett, H. Sun and B. Dolan, "IEC: Towards Interest-Eliciting Neural Conversational Agents".

[23] V. K. Cannanure, T. X. Brown and A. E. Ogan, "DIA: A Human AI Hybrid Conversational Assistant for Developing Contexts," in *The 2020 International Conference on Information and Communication Technologies and Development*, New York, NY, USA, 2020.

[24] H. He, A. Balakrishnan, M. Eric and P. Liang, "Learning Symmetric Collaborative Dialogue Agents with Dynamic Knowledge Graph Embeddings," *arXiv preprint arXiv:1704.07130,* 2017.

[25] A. F. T. Martins, J. Graca, P. Dimas, H. Moniz and G. Neubig, "Project MAIA: Multilingual AI Agent Assistant," in *22nd Annual Conference of the European Association for Machine Translation*, Lisboa, Portugal, 2020.

[26] K. Isbister, H. Nakanishi, T. Ishida and C. Nass, "Helper agent: designing an assistant for human-human interaction in a virtual meeting space," in *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, 2000.

[27] N. Li and P. Dillenbourg, "Designing conversation-context recommendation display to support opportunistic search in meetings," in *Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia*, 2012.

[28] A. Batra, A. Yadav and S. K. Sharma, "Connecting People Through Virtual Assistant on Google Assistant," in *Proceedings of ICETIT 2019*, 2020.

[29] R. F. Daniel, Y. Talmor, A. Lebrun, L. N. Landowski, D. Demir, J. H. Goldberg and W. Blandin, "Providing personal assistant service via messaging". United States of America Patent US Patent 10,686,738, 16 June 2020.

[30] J. J. Maestas, "AI assistant for interacting with customers across multiple communication modes". United States of America Patent US Patent 10,708,424, 7 July 2020.

[31] B. Z. Q. V. L. Prajit Ramachandran, "Searching for Activation Functions," *arXiv preprint arXiv:1710.05941,* 2017.

[32] C. T. Pei, "Positional Embeddings," Medium, 13 November 2019. [Online]. Available: https://medium.com/nlp-trend-and-review-en/positional-embeddings-7b168da36605.

[33] J. W. S. R. Margaret Li, "ACUTE-EVAL: Improved Dialogue Evaluation with Optimized Questions and Multi-turn Comparisons," *arXiv preprint arXiv:1909.03087,* 2019.