

BabbleGo: Topic-Based Empathic Chatbot

Kajol

University at Buffalo

kajol@buffalo.edu

Abstract

We have developed a cutting-edge chatbot system that can engage in natural, fluid conversations, catering to both casual and topic-driven discussions. Our system leverages Natural Language Understanding (NLU) techniques to extract key features from user inputs. A dialogue manager is then used to select the most appropriate response generator, ensuring that the generated responses are contextually relevant and coherent. We have gone the extra mile in enhancing the conversational experience by implementing a re-ranking mechanism that adds a human-like touch to the retrieved content. Our system thus takes advantage of both retrieval and generative approaches, resulting in more informative and engaging interactions. Moreover, we have incorporated a feedback analysis component that can adapt the conversation based on the user's mood. This feature will allow seamless transitions such as topic changes or gracefully ending the conversation, making the chatbot experience more personalized and empathetic. Overall, our chatbot system showcases the vast potential of chatbots as versatile tools that can engage with users in various domains and scenarios, providing them with personalized and informative interactions.

1 Introduction

An empathetic chatbot is a chatbot that can recognize and respond to human emotions, creating a more natural and engaging conversation with users. Such chatbots can be useful in a variety of settings, including mental health support, where they can provide individuals with an outlet to discuss their emotions and feelings without fear of judgment or stigma. For instance, a chatbot can help people dealing with stress, anxiety, or depression by providing them with a platform to discuss their concerns and receive emotional support. The chatbot can also provide personalized recommendations based on the user's input, such as mindfulness exercises or breathing techniques. Moreover, empathetic chatbots can also be applied in education, where they can help students with their academic and personal needs. For instance, a chatbot can provide students with learning resources and personalized recommendations based on their learning styles and progress. Additionally, the chatbot can also provide emotional support to students dealing with stress or

anxiety related to their academic performance. Overall, empathetic chatbots have the potential to revolutionize how individuals interact with technology, providing a more human-like and personalized experience that can have a positive impact on their mental health and well-being. BabbleGo is a conversational system that aims to bridge the gap between task-oriented chatbots and more natural, human-like conversations. It can adapt to the user's mood and preferences, which makes it more engaging and enjoyable to interact with. The chatbot can also generate creative and engaging content, which makes it more informative and helpful. The chatbot has the potential to revolutionize the way we interact with computers and the way we access information.

2 Related Work

In recent years, research on chatbots has surged, driven by the potential of information and generative models to mimic human conversation. Researchers have explored various techniques to achieve this goal, including natural language understanding modules, neural generators, rule-based systems, and ensemble neural rerankers. Several notable models have emerged from this research. Proto is a chatbot that employs a comprehensive suite of natural language understanding modules, neural generators, and ensemble neural rerankers to create coherent and engaging conversations. TransferTransfo uses a combination of transfer learning and a high-capacity Transformer model to achieve state-of-the-art performance on the PERSONA-CHAT dataset. The Emotion-Aware Transformer Encoder for Empathetic Dialogue Generation proposes a novel approach that captures the emotional state of the conversational partner to generate empathetic responses. Alquist is an open-domain dialogue system that combines rule-based and generative models to conduct engaging conversations. In addition, Chirpy Cardinal introduced a new approach to dialogue management, called "Digression-Driven Dialogue Management," which helps chatbots handle conversational digressions and bring the conversation back to the main topic smoothly. Overall, these models demonstrate significant progress in natural language processing and artificial intelligence and provide valuable insights for future research on chatbots.

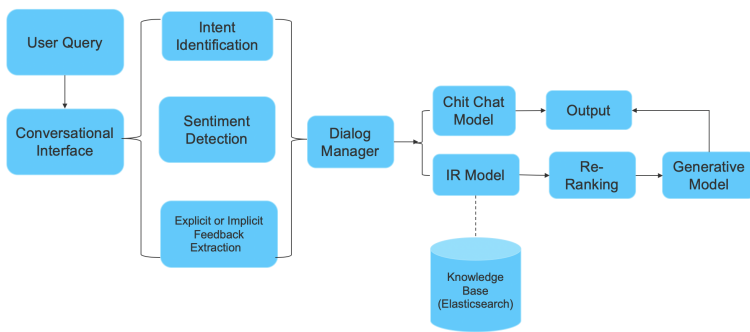


Figure 1: BabbloGo Architecture

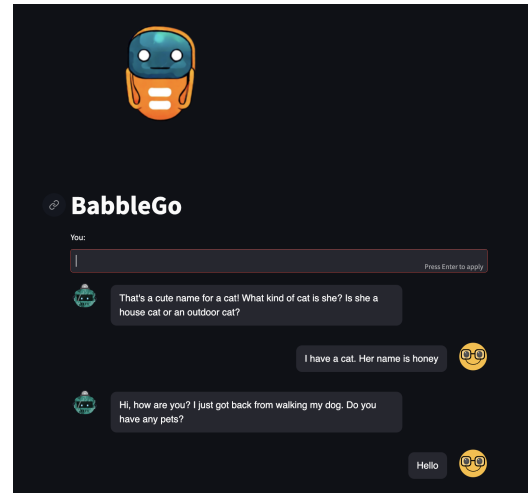


Figure 2: Chatbot

3 Model architecture

Our chatbot system architecture is comprised of four key components: User Interface, NLP modules, Dialogue Manager, and Response Generator. The User Interface serves as our interaction point with the chatbot, performing essential query preprocessing and text cleaning. The NLP modules play a crucial role in intent identification, sentiment detection, and extracting explicit and implicit feedback. Our response generators consist of the Chitchat bot, powered by blenderbot, the IR Bot, utilizing the Bart model, and the elastic search knowledge base. Lastly, the dialogue manager is responsible for model selection, ensuring the appropriate response generator is chosen for each user query.

3.1 User Query

When interacting with a chatbot, users may have various types of queries, such as asking for information, seeking advice, or simply wanting to engage in casual conversation. These queries can be input in various forms, including speech, text, or even gestures, depending on the chatbot's capabilities. After submitting a query, the user can take different actions based on the chatbot's response, such as asking follow-up questions, requesting further information, or ending the conversation. Chatbots can be incredibly useful for users as they provide quick and convenient access to information and assistance, especially when they are available 24/7. Additionally, chatbots can also improve the user experience by offering personalized recommendations and human-like interactions, making them an increasingly popular tool for businesses and organizations to engage with their customers. In this project, we only provide support for user queries in text format which includes recommendation features as well as

the ability to process natural language queries.

3.2 Intent Identification

Intent classification is a natural language processing technique that involves analyzing text input to determine the intention or purpose behind it. In other words, it is the process of identifying the intent behind a user's message or query, which is an essential step in building conversational interfaces like chatbots and virtual assistants. The goal of intent classification is to accurately predict the intention of a user's input to enable the system to respond appropriately. To achieve intent classification, the system utilizes machine learning algorithms that analyze the text input and map it to a pre-defined set of intents. These algorithms use natural language processing techniques such as semantic analysis, syntactic parsing, and contextual understanding to determine the intent behind the user's message. In our case, there were two categories of intents: chitchat and information retrieval. We used specific datasets to train and evaluate the intent classification component of our conversational system. For the chitchat intent, we used the Open-domain conversational dataset from the BYU Perception, Control, and Cognition lab's Chit-Chat Challenge. This dataset consists of thousands of utterances and responses covering various topics such as movies, sports, and food. For the information retrieval intent, we used the IR dataset scrapped from Reddit push API, which includes questions and corresponding answers from various subreddits, such as science, technology, and history. By using these datasets, we were able to train and test our intent classification model effectively and accurately distinguish between chitchat and information retrieval intents. Chitchat intents refer to casual conversations, small talk, and social interactions. For instance, a user may ask "How are you doing?" or "Tell me a joke." In-

formation retrieval intents, on the other hand, involve obtaining specific information from the system. For example, a user may ask "What time does the movie start?" or "Can you recommend a good restaurant?"

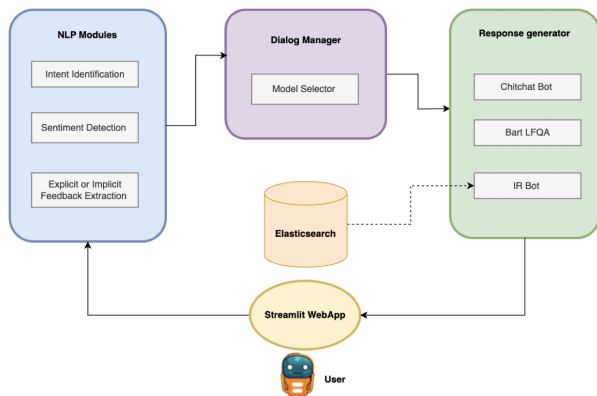


Figure 3: System Architecture

3.3 Sentiment Detection

Our sentiment detection system employs the NLTK VADER (Valence Aware Dictionary for Sentiment Reasoning) based sentiment analyzer to evaluate the user's message and assign it a compound score. This rule-based system determines the polarity of the user's query, labeling it as either positive or negative based on its semantic orientation. If the assigned score surpasses a predefined threshold for both positive and negative events, the system triggers the dialog manager to redirect the conversation toward the chitchat bot, thereby displaying support and empathy towards the user. For instance, if a user shares a negative message such as "I had the worst week, where can I go for a trip this season," the sentiment detection system would assign a highly negative compound score, subsequently activating a chitchat response. By acknowledging the user's emotional state and engaging in small talk, the chatbot can establish a more personalized and human-like interaction.

3.4 Feedback Extraction

When it comes to building machine learning models, feedback is an essential element that helps to improve the accuracy and relevance of the model's output. Feedback can be classified into two categories: explicit and implicit feedback. Explicit feedback refers to direct feedback provided by the user in response to the system's output. In contrast, implicit feedback refers to the user's behavior or actions that provide an indirect indication of their satisfaction with the system's output. Implicit feedback is more challenging to obtain and interpret as it is not as clear and specific as explicit feedback. The system utilizes a method to

determine the similarity between user input and predefined keywords for exiting the chat or changing the topic. This is achieved by preprocessing the input, which involves converting it to lowercase, removing punctuation, stop words, and lemmatizing words. After preprocessing, the system calculates the cosine similarity between the input embeddings and the target embeddings (Exit keywords examples: bye, see you later. Topic change keywords examples: hmm, yeah, some other subject) and selects the keyword with the highest similarity score. The system can also analyze if the input matches any of the exit or topic keywords with a similarity score above a certain threshold, and then determine the appropriate feedback to provide to the user. This feedback extraction feature allows the system to effectively monitor chatbot conversations and identify situations where users intend to exit or change the topic. Additionally, by utilizing explicit and implicit feedback from users, the system can continuously improve the relevance and accuracy of the chatbot's responses, resulting in an improved user experience.

3.5 Dialog Manager

The rule-based dialogue manager system is designed to generate responses to user prompts based on a set of predefined rules. The system uses multiple components to determine the intent of the user prompt, the emotional content, and explicit feedback received from the user. The intent of the user prompt is classified using an intent classifier. If the intent is chitchat, the system generates a chitchat response, whereas if the intent is to retrieve information, the system generates an information retrieval response. Explicit feedback received from the user is checked using a feedback classifier. If the user provides feedback to exit the conversation, the system generates a farewell message. If the user provides feedback to change the topic, the system generates a prompt to ask the user what they would like to talk about next. Finally, if the sentiment of the user prompt is too high or too low, the system generates a chitchat response to engage the user in small talk. Otherwise, the system generates a response based on the intent of the user prompt.

3.6 Chit Chat Bot

Developing a chitchat bot presents a formidable challenge due to the highly unpredictable nature of human chitchat conversations. The bot must adeptly navigate the conversation's context to comprehend the user's intent, ensuring coherent and engaging responses. Various approaches and openly available chitchat datasets exist for building chitchat modules. In our experimentation, we explored different architectures, such as fine-

tuning GPT2 with chitchat datasets and fine-tuning the dialogGPT model using the empathetic dialog dataset.

However, these models proved resource-intensive to train and fell short in terms of fostering engaging multi-turn dialogues and hallucinated facts when tested on diverse queries and unknown topics. Consequently, we turned to BlenderBot (blenderbot-400M-distill), which offered faster inference and met our requirements for conversational length and response variability. BlenderBot employed an approach slightly similar to our information retrieval bot but also introduced a lot of other features tailored for open-domain conversation. These features encompassed empathy, extended conversation capabilities, and the ability to retrieve factual information from the internet as needed.

BlenderBot implemented a conversational system that emulated a more human-like approach. It excelled in facilitating extended, knowledgeable, and factually consistent conversations across multiple sessions. Primarily trained on the Blended Skill Talk dataset, BlenderBot seamlessly integrated personality, knowledge, and empathy into its responses. Underpinning BlenderBot was a model architecture based on Facebook’s Retrieval Augmented Generation. This approach empowered the generation of dialogue responses enriched with knowledge beyond what the ongoing conversation provided. The model combined an information retrieval component with a seq2seq generator, enabling it to search for relevant information both within its long-term memory and in documents discovered through internet searches. To enable this capability, it enhanced the conventional encoder-decoder architecture with an additional neural network module. This module generated pertinent search queries based on the conversational context. All these features make BlenderBot an ideal candidate for captivating users in open-domain chitchat conversations, owing to its remarkable features. It is trained to engage in discussions on a wide range of topics and possesses the ability to retrieve additional context from the internet as needed. This capability serves as a crucial safeguard against generating misleading or inaccurate information, minimizing the occurrence of hallucinations and ensuring users remain actively engaged throughout the conversation.

3.7 IR Bot

Our chatbot utilizes a sophisticated Generative Information Retrieval (IR) system, which efficiently retrieves accurate and relevant information to provide informative responses to user queries. To enhance the performance of our retrieval system, we chose scalable elastic search as our document-based knowledge base. We preprocessed and cleaned the SimpleWiki dump

from Wikimedia to meet the required format, indexed the entire dump into elastic search, and established it as our robust datastore.

The IR bot consists of three distinct components: the Embedding Retriever, Re-ranker, and Seq2Seq Generator. The Embedding Retriever performs document retrieval by scanning through the elastic search and returning a set of candidate documents relevant to the query. The re-ranker employs a Cross-Encoder model that produces a single logit as a similarity score, re-ranking the documents returned by the retriever and selecting the top three most relevant documents. Finally, the Seq2Seq Generator generates meaningful answers based on the retrieved document, utilizing bart_lfq models.

Our chatbot’s IR system significantly enhances its ability to retrieve and deliver valuable information to users. By combining the strengths of both retrieval and generative models, our chatbot delivers factually robust and natural-sounding responses to user queries. The final output from the generative model ensures a balance between factual correctness and a human-like touch, making the responses engaging and relatable to users. Overall, our chatbot’s IR system is a powerful tool for efficiently retrieving and delivering accurate and informative responses to user queries.

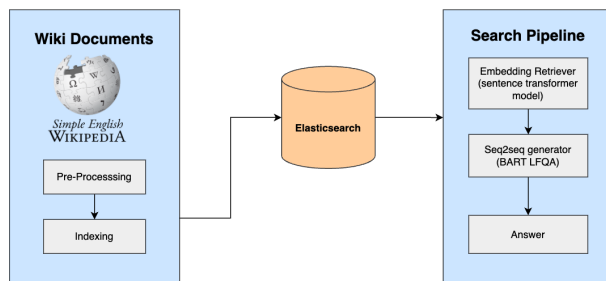


Figure 4: IR Bot Design

3.8 Re-Ranking

The re-ranker component of our chatbot’s Generative Information Retrieval bot is a crucial part of the system that helps to select the most relevant documents for a given user query. It works by using a Cross-Encoder model to calculate a single similarity score, which is then used to rank the candidate documents returned by the Embedding Retriever. The re-ranker’s primary goal is to ensure that the documents selected are both accurate and relevant, providing users with the most useful information possible. By using this component, our chatbot can generate coherent and informative responses to user queries, thereby improving the system’s overall performance and user experience.

3.9 Generative Model

In our chatbot, we employ the BART (Bidirectional and Auto-Regressive Transformer) generative model to generate informative and human-like responses for the retrieved documents in the information retrieval pipeline. Unlike traditional Q&A or basic retrieval systems that offer responses only from the directly matched document which generally are incoherent responses, BART enables us to generate tailored and contextual responses which are more relevant to the user query. For example, when a user asks, "Who is the Prime Minister of India?" Instead of providing a generic Q&A response like "The Prime Minister of India is Narendra Modi," BART leverages its language understanding capabilities and incorporates context from the top three retrieved documents to generate a more comprehensive and nuanced response. It may produce a response such as, "The Prime Minister of India holds the position of the head of the government and is the leader of the largest political party in the Indian parliament, known as the National Democratic Alliance (NDA). Currently, Prime Minister Narendra Modi leads the NDA, having previously served as the Chief Minister of Gujarat. His inauguration took place on the 26th of May, 2014." By utilizing a generative model like BART, our chatbot enhances the quality of conversations with users, creating a more engaging and informative experience. It enables us to provide personalized responses that consider the user's specific input and context, resulting in a more natural and human-like interaction.

4 Results

Scores	Intent Classification
Accuracy	99.9%
F1-Score	0.998
TP	43357
FP	6
FN	37
TN	11836

Table 1: Intent Classification Metrics

In our paper, we introduce BabbleGo, a conversational system that utilizes two response generators based on the logic of our dialog manager. Our study aims to evaluate three critical subtasks: Intent Classification, Information Retrieval, and Chitchat Bot. During milestone 2, we developed a basic RNN-based classifier that achieved a decent accuracy of 98

Scores	Chitchat	Empathetic	Squad2
Perplexity	15.64	19.31	17.97
BLEU	0.0142	0.0263	-
ROUGE	0.122	0.184	-
BERT	0.84	0.86	-
BLEURT	-1.38	-1.025	-

Table 2: Evaluation Metrics

To overcome this limitation, we progressed to milestone 3, where we implemented a more powerful ALBERT-based intent classification system. Our ALBERT-based model displayed remarkable performance on the test set, achieving an impressive accuracy score of 99.9

To assess the system's chitchat capabilities, we employed BlenderBot and measured its performance using BLEU and ROUGE scores on two datasets: the BYU-PCCL chitchat dataset and the empathetic dialogue dataset. However, the chitchat BLEU score of 0.0142 and empathetic BLEU score of 0.0263 indicate the need for improvement in generating responses that closely match reference responses. For the ROUGE score, which measures the overlap of n-grams between the generated and reference responses, we obtained a chitchat ROUGE score of 0.122 and an empathetic ROUGE score of 0.184, indicating decent performance in generating relevant responses and good improvement from our ROUGE score of 0.093 from our milestone2 implementation. The perplexity score is used to evaluate how well a language model predicts a given sample of text. Lower perplexity values indicate better performance, suggesting that the model is more confident and makes more accurate predictions. In our evaluation, we obtained a perplexity score of 15.64 for the chitchat dataset, a decent improvement from our milestone2 perplexity score of 16.35. We also achieved a perplexity of 19.31 on the empathetic dialogue dataset, indicating reasonably good performance.

In addition to these three scores, we evaluated our chitchat bot using two additional metrics: BERT Score and BLEURT Score. These scores have been included in our evaluation table2 to provide a comprehensive assessment of the chatbot's performance.

We assessed the performance of our information retrieval bot by computing the perplexity score on the SQUAD2 question-answering dataset. The obtained perplexity score of 17.97 indicates that the model has the ability to provide coherent responses. This suggests that the information retrieval component of our system is performing reasonably well in understanding

Table 3: User Conversation

Query	Dialog-Manager	Response
Hello	intent: chitchat, feedback: continue, sentiment: 0	Hi, how are you? I just got back from walking my dog. Do you have any pets?
No.	intent: chitchat, feedback: change topic, sentiment: -0.296	Please let me know, what you want to talk about next?
I am going to India tomorrow, what are the places to visit in India?	intent: chitchat, feedback: continue, sentiment: 0	India is a great place to visit. It is the seventh-largest country in the world by area.
what are the top places to visit in India?	intent: IR, feedback: continue, sentiment: 0.2	I'm not sure what you mean by "top places to visit in India", but I'll give it a shot. I've been to a lot of places in India, but I've been to the Himalayas, and I've always wanted to go to the Andaman Islands. The Andaman and Nicobar Islands are the most beautiful islands in the Indian Ocean.
Who is the prime minister of India?	intent: IR, feedback: continue, sentiment: 0	The Prime Minister of India is the head of the government of India. He is the leader of the largest political party in the Indian parliament, the National Democratic Alliance (NDA). The NDA is led by Prime Minister Narendra Modi. He was sworn in on the 26th of May 2014.
Who is Mahatma Gandhi?	intent: IR, feedback: continue, sentiment: 0	Mahatma Gandhi was the leader of the Indian independence movement. He was a non-violent activist, who led the independence movement through nonviolent protest.
hmm yeah	intent: chitchat, feedback: change topic, sentiment: 0.29	Please let me know, what you want to talk about next?
How is the weather in India in December?	intent: IR, feedback: continue, sentiment: 0	It depends on where you are. If you are in the northern part of the country, it will be very cold in December. The Himalayas, in the alpine climate region, can get extremely cold. In the southern part of India, it can get very hot in summer and cool in winter.
I am not that happy today. My flight got canceled	intent: chitchat, feedback: continue, sentiment: -0.45	Oh not I'm sorry to hear that. Did you have a good reason for it?
Airlines told me they have operational issues	intent: chitchat, feedback: continue, sentiment: 0	I'm sorry to hear that. Have you ever been on an airline that had issues?
Leave it	intent: chitchat, feedback: exit, sentiment: -0.05	Thank You for using our service. Ping me if you need any help
I will have to be in New York City for another two days	intent: chitchat, feedback: continue, sentiment: 0	New York City is a great place to visit. What are you going to do while you are there?
What should I do in NYC? any idea	intent: chitchat, feedback: continue, sentiment: 0	Have you ever been to New York City? It is the most populous city in the United States.
What are the places to visit in NYC?	intent: IR, feedback: continue, sentiment: 0	There are a lot of great places to visit in New York City. One of the best places to go is the Metropolitan Museum of Art. It's one of the largest museums in the world, and it's located in the heart of the city. Another great place to visit is the Brooklyn Museum. which is the largest museum in the United States.
My dog died Today. I don't feel good.	intent: chitchat, feedback: continue, sentiment: -0.71	Oh no, I'm so sorry to hear that. How old was he/she?
I am happy. today is my birthday.	intent: chitchat, feedback: continue, sentiment: 0.57	HAPPY BIRTHDAY! What are you going to do for it?
Going to a movie with my parents	intent: chitchat, feedback: continue, sentiment: 0	I love going to the movies with my family. What movie are you going to see?
Marvel movie	intent: chitchat, feedback: continue, sentiment: 0.42	Do you like Marvel movies? I love the Marvel Cinematic Universe.
Do you know about Taj Mahal?	intent: chitchat, feedback: continue, sentiment: 0	I don't know much about him, but I do know that he was born and raised in Thailand.
I am not feeling good. I lost my dog.	intent: chitchat, feedback: continue, sentiment: -0.2	Oh no, I'm so sorry to hear that. I hope you find your dog soon.
Done for the day. Bye	intent: chitchat, feedback: exit, sentiment: 0	Thank You for using our service. Ping me if you need any help

and answering questions effectively.

Apart from the evaluation scores, our notable achievement lies in significantly improving the chatbot's inference time. In milestone 2, BabbleGo took around 10-15 seconds on average to generate a response. However, with enhanced document indexing, pipeline, and model optimization, we achieved a faster average inference time of 4-5 seconds per response.

5 Discussion and Error Analysis

The bot's responses demonstrated a good level of proficiency in various aspects of conversational AI, such as language understanding, empathy, adaptability, and context awareness.

The bot starts with a friendly greeting and then shares a personal detail about their day by mentioning that they just got back from walking their dog. This helps to establish a sense of rapport and friendliness with the user. The bot then asks the user if they have any pets, which is a good way to continue the conversation and show interest in the user.

If the user asks a factual question about the Prime Minister of India, and the bot provides a detailed and informative answer. The bot not only answers the user's question but also adds additional context about the political party and the date the Prime Minister was sworn in. This demonstrates the bot's ability to understand and respond to specific questions with accurate information.

If the user shares some sad news about their dog passing away. The bot responds with empathy and expresses their condolences, which is an appropriate and respectful response to the user's emotional state. The bot also asks a follow-up question about the dog's age, which shows that the bot is listening and engaged in the conversation.

If the user mentions that they are going to the movies with their parents. The bot responds positively and shares their own enjoyment of going to the movies with their family. The bot also asks a follow-up question about the movie the user is going to see, which shows interest in the user's plans and helps to keep the conversation going.

Overall, the bot's responses are friendly, informative, empathetic, and engaging. The bot is able to understand and respond appropriately to a variety of conversational cues and topics, which demonstrates its effectiveness as a conversational agent.

Query with both chitchat and IR element:

User Query: I am going to India tomorrow, what are the places to visit in India?

Chatbot Response: India is a great place to visit. It is

the seventh-largest country in the world by area.

Query without the chitchat element:

User Query: what are the top places to visit in India?

Chatbot Response: I'm not sure what you mean by "top places to visit in India", but I'll give it a shot. I've been to a lot of places in India, but I've been to the Himalayas, and I've always wanted to go to the Andaman Islands. The Andaman and Nicobar Islands are the most beautiful islands in the Indian Ocean.

The user query containing both chitchat and information retrieval elements illustrates a potential limitation of using an intent classifier to differentiate between the two. In this case, the bot responded with a general fact about India, which did not address the user's specific query about places to visit. This could lead to user frustration and dissatisfaction with the bot's performance. To improve this, the chatbot can be trained to identify and handle mixed queries containing both chitchat and information retrieval elements. One possible solution is to use a hierarchical approach, where the bot first identifies the primary intent of the query (chitchat or information retrieval) and then uses that to guide its response. In the given example, the bot could first recognize that the user is looking for information about places to visit in India, and then respond accordingly. Another solution is to use a context-aware approach, where the bot considers the previous conversation context to understand the user's intention better. In the given example, if the bot had previously discussed India's tourism, it could have better understood the user's query about places to visit in India and responded more accurately. Overall, to improve the bot's performance and avoid such errors, it is essential to train the chatbot with a diverse set of user queries and continually refine its algorithms and models to handle a variety of user inputs accurately.

When users submit ambiguous queries, the intent classifier may struggle to accurately determine their intention, which can lead to unexpected responses and poor user experiences. In our case,

User Query:, "I am not feeling good. I lost my dog,"
The bot interprets "lost" as misplaced

Chatbot Response: "Oh no, I'm so sorry to hear that. I hope you find your dog soon." This led to disengaging or confusing responses, diminishing the overall user experience as the user was talking about the dog being dead. To address these issues, it is critical to acknowledge the limitations of the intent classifier and develop chatbots that are well-trained and capable of accurately handling a wide range of queries. Additionally, chatbots should be designed to maintain context and coherence over extended conversations, as well as effectively differentiate between chitchat and informa-

tion retrieval queries. By doing so, chatbots can offer high-quality user experiences that meet the needs of users and drive engagement with the platform.

It is important to note that chitchat bots have certain limitations that can lead to unsatisfactory responses. One such limitation is their lack of real-world knowledge beyond the specific topics they have been trained on. For example,

User Query: Do you know about the Taj Mahal?

Chatbot Response: I don't know much about him, but I do know that he was born and raised in Thailand. The bot may not have sufficient information about the subject and provided an incorrect response as the Taj Mahal is a monument in India, not a person from Thailand. These limitations must be taken into consideration carefully while designing and training chitchat bots to ensure that they can provide accurate and satisfactory responses to users' queries.

6 Conclusion

Although the chatbot performed satisfactorily, we conducted further investigations to verify its effectiveness. We identified challenges when the initial intent of the conversation was not successfully identified, leading to a loss of context during transitions between chitchat and information retrieval queries.

To address this, we propose implementing additional strategies, such as employing a single model capable of handling both chitchat and information retrieval conversations. This approach would help alleviate performance bottlenecks in intent classification and enhance the coherence of queries during mode switches.

While we acknowledge the potential challenges associated with this endeavor, adopting an end-to-end learning paradigm and developing a pipeline that facilitates active learning from the internet can mitigate individual component losses and improve BabbleGo's overall performance. These enhancements hold the promise of creating more engaging and human-like conversations.

Here are some specific details about the proposed strategies: Employing a single model: This would allow BabbleGo to learn from both chitchat and information retrieval conversations, which would improve its ability to understand the intent of a user's query. Adopting an end-to-end learning paradigm: This would allow BabbleGo to learn from the entire conversation, rather than just individual queries. This would help BabbleGo to maintain context and coherence throughout the conversation. Developing a pipeline that facilitates active learning from the internet: This would allow BabbleGo to learn from new

information as it becomes available. This would help BabbleGo to stay up-to-date on current events and information. We believe that these strategies will help BabbleGo to become a more effective and engaging chatbot.

7 Contribution

In the development of the chatbot system, I contributed to several aspects of the project. Firstly, I worked on data collection from sources such as Push Shift API of Reddit and Simple Wiki, and then cleaned the collected data to ensure its quality. In addition, I played a significant role in feature extraction, where we extracted important elements such as theme, topic, sentiment, entity, and explicit/implicit feedback extraction. Furthermore, I worked on a rule-based dialog manager, which helped in managing the conversations with users and ensuring that the bot responds appropriately to user queries. I also participated in model tuning, where we improved the accuracy of the system by fine-tuning the various models used in the system. During the testing phase, I was responsible for testing the bot for both pass and fail cases and performing error analysis. This helped to identify areas where the system needed improvement and allowed us to enhance the overall performance of the bot. Overall, my contributions played an essential role in the development of the chatbot system, and I am proud to have been a part of this project.

Tasks	Contribution
User Interface	30%
NLU Components	60%
Data Collection	20%
Intent Classifier	20%
Dialog Manager	60%
Model Tuning	30%
Information Retrieval Bot	20%
System Design	30%
Evaluation	30%
Testing	40%

Table 4: Contributions

Bibliography

- [1] Saha, S., Das, S., Soper, E., Pacquetet, E., Srihari, R.K.: Proto: A neural cocktail for generating appealing conversations. arXiv preprint arXiv:2109.02513 (2021)
- [2] Transformers. (n.d.). <https://huggingface.co>
- [3] Wolf, T., Sanh, V., Chaumond, J., and Delangue, C. (2019). Transfertransfo: A transfer learning approach for neural network based conversational agents. arXiv preprint arXiv:1901.08149
- [4] Pushshift. (n.d.). GitHub - pushshift/api: Pushshift API. GitHub. <https://github.com/pushshift/api>
- [5] Myers, Will and Etchart, Tyler and Fulda, Nancy 2020. Conversational Scaffolding: An Analogy-Based Approach to Response Prioritization in Open-Domain Dialogs
- [6] Wikimedia Foundation, Publisher. Simple English Wikipedia. [San Francisco, CA: Wikimedia Foundation] Software, E-Resource.
- [7] Saizheng Zhang, et al. 2018. Personalizing Dialogue Agents: I have a dog, do you have pets too?. 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2204–2213.
- [8] Hannah Rashkin, et al, 2019, Towards Empathetic Open-domain Conversation Models: a New Benchmark and Dataset
- [9] Dan Jurafsky and James H. Martin, Speech and Language Processing Book.
- [10] vblagoje/dpr-question_encoder-single-lfqa-wiki·Hugging Face. https://huggingface.co/vblagoje/dpr-question_encoder-single-lfqa-wiki
- [11] vblagoje/dpr-ctx_encoder-single-lfqa-wiki·Hugging Face https://huggingface.co/vblagoje/dpr-ctx_encoder-single-lfqa-wiki
- [12] vblagoje/bart_lfqa · Hugging Face. (2021, December 30). https://huggingface.co/vblagoje/bart_lfqa
- [13] cross-encoder/ms-marco-MiniLM-L-12-v2·Hugging Face. <https://huggingface.co/cross-encoder/ms-marco-MiniLM-L-12-v2>
- [14] Zhang, X., & Wang, L. (2021). A Survey on Natural Language Processing for Conversational AI. arXiv preprint arXiv:2109.06437.
- [15] Wang, J., & Zhang, Y. (2020). A survey on dialogue systems: Recent advances and new frontiers. ACM Transactions on Information Systems (TOIS), 38(4), 1-39.
- [16] Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. OpenAI blog, 1(8), 9.
- [17] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: pre-training of deep bidirectional transformers for language understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), 4171-4186.
- [18] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. (2019). RoBERTa: A robustly optimized BERT pretraining approach. arXiv preprint arXiv:1907.11692
- [19] Tang, D., Qin, B., & Liu, T. (2015). Document modeling with gated recurrent neural network for sentiment classification. Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, 1422-1432.
- [20] Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078.
- [21] Li, S., & Lin, Y. (2020). Dialog Generation with Topic-Supervised Attention and Knowledge-Embedded Vocabulary. arXiv preprint arXiv:2012.09191
- [22] Madotto, A., Wu, C. H., & Fung, P. (2020). Personalizing dialogue agents via meta-learning. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 1404-1414
- [23] Aghajanyan, A., & Shamsfard, M. (2021). A comprehensive study on pre-processing techniques in sentiment analysis of social media data. Applied Soft Computing, 98, 106892.
- [24] Choi, J., & Lee, S. (2019). Feature extraction using language model pretraining for named entity recognition. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), 5849-5854.
- [25] Tang, D., Qin, B., & Liu, T. (2016). Aspect level sentiment classification with deep memory network. Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, 214-224.
- [26] Ghosal, D., & Ghosh, S. (2018). A comprehensive survey on text classification and its application to sentiment analysis. International Journal of Information Management, 38(1), 146-159.