ANUNCIADO, CHRISTIAN JAY

TABASA, PRINCE JOSEPH

VIILLASO, VHANNE BRIAN

# **Teknoy INquiry Assistant: TINA 2.0**

1. **Abstract**

**The transition to traditional education to blended setup has caused the institution to address dozens of questions everyday regarding to its services offered. Dealing with a lot of queries can be tedious and repetitive at times, this problem can be automated by using chatbot. Teknoy Inquiry Assistant also known as TINA is the university chatbot that answers all the questions raised by the students or the inquirers. This chatbot addresses all the types of queries using natural language processing (NLP) with use of intent classification. To further know the intent behind the data, the Bidirectional Recurrent Neural Network containing attention layers is used, so that input with large number of tokens can be replied with more appropriate conversation. The dataset used for the training were provided by the university.** After, there is the training and evaluation of the models. The key metrics that were measured are the evaluation loss which calculates the error of the model, evaluation accuracy, which describes what percentage of data are classified correctly, predictions that show how many queries are labeled correctly, and the success rate per category. In the end, TINA can provide the answers to queries based on the models used for this chatbot.

1. **Introduction**

Currently, universities seek to improve their customer service by adapting to students’ queries. This is due to the transition of traditional education to blended setup education. Hence, it creates a lot of queries, from tuition prices, uniforms, programs offered by the university, requests for documents, and many other questions. Handling multiple queries can be tedious and repetitive and these queries can be labeled and sorted based on their type.

[1] Chatbots can be defined as software that helps humans to make coherent conversations with a machine using natural language like English, etc. Conversation can be engaging at times with large queries, topics, and vocabulary which makes chatbots one of the trends in the industry.

Intent Classification, also known as intent recognition, is an approach used in natural language processing and intent classification also uses machine learning. [2] Intent recognition works through the process of providing examples of text alongside their intents to a machine learning (ML) model. Intent recognition is critical for ensuring that Digital Employees intelligently and accurately respond to user inquiries.

The goal of this research is to explore and define the intents of data queries using attention and bidirectional RNN models that can be utilized in making the chatbot.

The scope of this study is only to discuss the process of cleaning data, enumerate the categories of the data, and show the result of the models used. Dataset was provided by the university based on the queries they have received since the transition to blended learning.

1. **Significance of the Study**

This research would provide significant improvements to the previous version of TINA. TINA would be able to answer questions and queries more efficient and accurate, depending on the student’s intentions (e.g., payment, FAQ, events., etc.). Through this research, students and other schools will further realize the importance of chatbots especially when acquiring school-related information. Furthermore, the analysis and results presented in this study could be used as a tool for future research to improve the effectiveness of TINA and other chatbots that use Bidirectional RNN and attention models.

1. **Theoretical Background**

This study is based on Mensio and Rizzo (2021) study on A RNN Contextual Approach to Intent Classification for Goal-oriented Systems, in which the researchers stated that the origins of chatbots can be traced back to chit-chat agents. These systems were created to overcome the Turing Test's initial challenge and to simulate a natural language conversation between humans.

This section explains and elaborates on the split of chatbots, with one focusing on rich interrogation and the other on more structured dialogues, emphasizing the differences and common points in the understanding section. This section will also explain why the use of Bidirectional RNN and Intent Classification on their chatbots is relevant to our study.

**4.1. Question and Answer System**

[3] A Question and Answer (QA) system is designed to provide a single answer to a specific question (so-called single turn), and modern systems achieve robust performance in such a scenario. However, most interactions with QA systems are based on multiple handshakes of question/answer pairs, in which the human refines the questions while the system collects the necessary information and generates a compelling final answer through multiple turns. The information in this type of system is typically stored in Knowledge Bases (KBs), which can take the form of structured tables, graphs, or textual documents written in natural language. For unstructured QA systems, two major paradigms have emerged: knowledge-based and information retrieval-based. The goal for both is to provide answers by performing three steps: i) understanding the question (extracting all relevant parameters), ii) retrieving the information (against the KB via queries or more unstructured forms), and iii) combining the results and presenting them in textual form.

**4.2. Goal-oriented agents**

[3] Instead of being a natural language interface to perform advanced queries, goal-oriented agents enter the Question Answering landscape as domain-specific systems that allow users to interact with some services to perform different tasks. As a QA branch, it provides users with domain-specific information. However, it can also be used to perform some actions, in the form of a digital personal assistant, in various specific domains, such as booking and travel services. The most common strategy for developing these systems, excluding button-based flows in which the agent takes complete control of the dialog, is to map sentences expressed by a user onto a fixed set of intents (sentence types) and slots (entities mentioned that are used as parameters in conjunction with a role). Approaches for these two tasks, namely intent classification and slot filling, typically employ Recurrent Neural Networks (RNN) that can work on word sequences. The current state of the art is achieved through a collaborative approach that employs an encoder-decoder structure originally designed for translations to perform both sequence tagging (slots tagging) and sentence categorization (intent). This method works well in single-turn cases because each sentence is processed independently. While these systems are similar to both structured and unstructured QA in terms of language understanding, there are some significant differences:

**main focus:** It is essential to provide access to available operations and manage interactive guided procedures that require conversation state management. The comprehension of complex interrogations is not emphasized;

**limited search capabilities:** unlike QA systems, access to a knowledge base may be restricted by a limited set of available operations, as a result of limited remote APIs or pre-existing application logic. These can correspond to a finite set of question types and parameters;

**bidirectional QA:** the system may require some missing parameters from the user, increasing the complexity of the interaction.

**Interaction with dynamic data:** because of resource availability, the information stored can change frequently over time (for example, when providing information about transportation or another dynamic domain). Furthermore, some actions can actually modify the stored information (for example, booking a table at a restaurant).

**4.3. Recurrent Neural Networks and Intent classification**

Other studies on the problem of sentence classification within an interaction context have also been done in the literature. The approach proposed in [8] for domain classification, which is equivalent to the high-level class in a hierarchical intent classification, concatenates the previous model prediction at the word level with each word vector. The approach proposed in this work differs both in the specific point at which the previous classification is used (not with the input words but on the sentence level, using the high-level RNN) and in the way the word-level features are summarized in sentence-level features and considered for next interactions by the learning network.

Three major challenges are addressed by these [3]:

**detect the change of intent in a multi-turn environment:** in other words, to know when a session (a sequence of messages related to a single intent) ends and a new one begins. This corresponds to deciding whether to keep the value of the previous intent or to consider some evidence on the current input for each input sentence. The first scenario occurs when the input sentence is from a previous session and the user is simply continuing the interaction with the same initial intent. The second case is when a new intent is expressed in the current sentence, indicating a change in intent.

**capture intent dependencies using the RNN:** capturing the sequences of intent values allows for a more accurate prediction of the sentence knowing the preceding intents. This is especially useful with sentences that aren't very expressive because they're referring to some context of the interaction.

**consider the current agent turn words:** knowing what has been replied to the user can help contextualize a new sentence that may lack obvious indicators of intent.

1. **Review of the Related Literature**

A chatbot is one of the simplest and most common forms of Intelligent Human Computer Interaction (HCI) and a classic example of an AI system. It is a computer program that responds intelligently to text or voice communications and uses natural language processing to comprehend one or more human languages by NLP or Natural Language Processing. A chatbot is described in the lexicon as “A computer program designed to stimulate conversation with human users, particularly over the Internet”. Smart bots, interactive agents, digital assistants, or artificial conversation entities are other names for chatbots (Adamopoulou, E., & Moussiades, L. (2020). An Overview of Chatbot Technology).

The goal of chatbots, a new technology, is to mimic human conversation utilizing natural language, audio, video, and media processing. These technological advancements are parts of artificial intelligence methods that use neural networks to offer improved information delivery with a layer of contextual knowledge. Chatbots can execute tasks as well as disseminate information to users. As users interact with chatbots more frequently, they can learn their patterns, improving both the consumer and business experience (Chatbots vs Search Engines—What’s the Difference? (2021, January 26)). Organizations can utilize chatbots to better understand their consumers’ needs, increase the precision of their sales or service messaging, and use copywriting language that works. By using appropriate conversational design techniques, you can make sure that your chatbot is answering any queries it may have.

To evaluate the effectiveness of the chatbot, when determining a chatbot’s activity volume, it is necessary to count all the interactions that occur between the moment a user asks a straightforward query and the start of a productive conversation. The number of user sessions that do not lead to the planned "specialized" usage of your chatbot is known as the "bounce rate." A high rate suggests that your bot isn't getting questions that are pertinent to its area of expertise. This should motivate you to change its content, reconsider where it fits within the user experience, or do both. It is an indicator that needs to be carefully monitored. Next is the use rate by open sessions, this is the total number of active chatbot sessions at any one time. To acquire a meaningful measurement, this rate must be weighted with the average number of open sessions over a given period. Target audience session volume this signal is crucial for confirming whether your aims are being met. If you want to make sure that the individuals you are trying to reach are using your chatbot effectively, you may check the penetration rate for the audience you are trying to reach. If not, you must reevaluate your customer experience or change management techniques to win over your users. The chatbot response volume specified will let you know how many inquiries your chatbot has handled. Interaction rates the user should pay close attention to this signal if you want to gauge user involvement during chats with your chatbot. You'll be able to gauge the typical volume of messages sent during a chat. Last is the most frequently asked question. Does it show what questions are asked to the chatbot the most frequently? With the aid of this information, you can modify your chatbot to focus on the topics that are discussed the most frequently and boost its functionality. Your corrective work will be guided by an analysis of repeated queries, allowing you to concentrate on the subjects that interest your users the most and the techniques that will help you raise the understanding levels and quality of your bot's responses (Measuring Chatbot Effectiveness: 15 KPIs to Track. (2018, November 27).

TensorFlow and deep learning NMT models have been used to create a chatbot. The BRNN and attention mechanism were the foundation of the chatbot's architecture. The Reddit dataset is used by the chatbot's open-source knowledge base, and it provides some sincere responses. The model will eventually be rewarded for providing responses that are both pertinent and sensitive. The Deep Reinforcement Learning (DRL) approach will be used in this. The methods used to deploy and train the chatbot can also be used to train chatbots for certain domains, such as the scientific, healthcare, banking, security, and educational ones. This method will make it simpler to create chatbots in any domain and can enhance chatbots that are already in use that are based on simple RNN architecture or other neural networks by employing the attention mechanism described above. One can download a specialized Subreddit for a given domain to create a chatbot tailored to that particular industry (such as healthcare, education, etc.). Building a healthcare chatbot will also be part of future work. This chatbot will help patients with diseases like COVID-19 (pandemic), Diabetes, High Blood Pressure, and Heart, among others, by answering questions about the diseases, suggesting foods to eat, and explaining how to handle various emergencies. It will also use a recommender system to power this chatbot. The novel concept in this study was to evaluate the MacBook Air as a platform for studying and refining deep neural network models. We consider the MacBook Air to be a subpar and entry-level deep learning machine. This finding can assist professionals or students at the beginner level in making an informed system choice before beginning deep learning (Dhyani, M., & Kumar, R. (2021). An intelligent Chatbot using deep learning with Bidirectional RNN and attention model).

**Definition of Terms**

* Chatbot - a computer program that is to stimulate conversation with human users, especially over the internet.
* Human-Computer Interaction - is a multidisciplinary field of study focusing on the design of computer technology and the interaction between humans and computers.
* Natural Language Processing (NLP) - is a branch of computer science concerned with giving computers the ability to understand the text and spoken words in much the same way as human beings can.
* Artificial conversation - refers to technology like chatbots and voice assistants that stimulate a human conversation.
* Artificial intelligence - is the theory and development of computer systems able to perform tasks that normally require human intelligence.
* Bidirectional RNN - duplicates the RNN processing chain so that inputs are processed in both forward and reverse time order. This allows a BRNN to look at future context as well.
* Machine Learning -   is a branch of [artificial intelligence (AI)](https://www.ibm.com/cloud/learn/what-is-artificial-intelligence) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy.

**REFERENCES**

**[1]** Dhyani, M., & Kumar, R. (2021). An intelligent Chatbot using deep learning with Bidirectional RNN and attention model. *Materials Today: Proceedings*, *34*, 817–824. <https://doi.org/10.1016/j.matpr.2020.05.450>

**[2]** Marshall, C. (2020, September 28). What is intent recognition and how can I use it? Super. AI. https://medium.com/mysuperai/what-is-intent-recognition-and-how-can-i-use-it-9ceb35055c4f

**[3]** Mensio, M., & Rizzo, G. (2018). Multi-turn QA: A RNN Contextual Approach to Intent Classification for Goal-oriented Systems. 1075–1080. <https://doi.org/10.1145/3184558.3191539>

**[4]** Adamopoulou, E., & Moussiades, L. (2020). An Overview of Chatbot Technology. In I. Maglogiannis, L. Iliadis, & E. Pimenidis (Eds.), *Artificial Intelligence Applications and Innovations* (pp. 373–383). Springer International Publishing. <https://doi.org/10.1007/978-3-030-49186-4_31>

**[5]** *Chatbots vs Search Engines—What’s the Difference?* (2021, January 26). The Chatbot Business Framework. <https://chatbotbusinessframework.com/search-engines-vs-chatbots-whats-the-difference/>

**[6]** Measuring Chatbot Effectiveness: 15 KPIs to Track. (2018, November 27). *Living Actor*. <https://corporate.livingactor.com/en/measuring-chatbot-effectiveness-kpi/>

**[7]** Dhyani, M., & Kumar, R. (2021). An intelligent Chatbot using deep learning with Bidirectional RNN and attention model. *Materials Today: Proceedings*, *34*, 817–824. <https://doi.org/10.1016/j.matpr.2020.05.450>

**[8]** Puyang Xu and Ruhi Sarikaya. 2014. Contextual domain classification in spoken language understanding systems using recurrent neural network. In 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 136–140