Chatbot Built with Deep Learning: Fannie Mae Internship

CS4991 Capstone Report, 2023

Sooren Ghodsi
Computer Science
The University of Virginia
School of Engineering and Applied Science
Charlottesville, Virginia USA
sg7ytn@virginia.edu

ABSTRACT
At the Federal National Mortgage Association, also known as Fannie Mae, the amount of time and business knowledge required to run certain database queries limited the efficiency of work. To address this challenge, I developed a deep learning chatbot to automate these processes. This made it easier for a larger number of associates to retrieve and interpret crucial data. I used the TensorFlow framework for natural language processing (NLP) and wrote Python scripts to connect to and extract data from Oracle databases. I defined use cases for the chatbot through collaboration with business associates, and the entire development followed the Kanban framework of Agile. As a result, we anticipate a notable decrease in the time needed to generate reports, leading to a smoother workflow. In the upcoming phases, Fannie Mae will transition the chatbot from the development environment to a full-scale production setting. Additionally, the chatbot will be integrated into most internal system user interfaces (UI) and refined to offer more human-like interactions.

1. INTRODUCTION

In today's digital age, the financial industry is driven by an abundance of data, serving as the backbone for decision-making. This explosion of data means that the way organizations interact with, analyze, and interpret their data repositories becomes crucial. Fannie Mae, a pivotal player in the U.S. housing market, has witnessed the difficulties that come with managing large volumes of data associated with mortgages and housing trends. Efficiently accessing this data is vital for daily operations, strategic decision-making, and maintaining market relevance.

The issue I had to resolve was not just about data access but also its manipulation and interpretation. A significant challenge came up: only a select group of associates with the required business knowledge could effectively run certain database queries and interpret their results. This created a bottleneck, slowing processes and leading to potential inefficiencies. The introduction of a deep learning chatbot was more than a solution. It was an innovation aimed to democratize data access, ensuring that a larger segment of associates could not only retrieve but also act on the data.

2. RELATED WORKS

In the development of the chatbot, I drew from existing knowledge and methods within the field. The chatbot uses TensorFlow, as detailed by Singh, et. al. (2018), to create a neural network model that can understand and maintain the context within a conversation. This framework is essential for ensuring that conversations with the chatbot are user-friendly and can handle the specific requests associated with the mortgage workers' queries. This article is referenced more in sections 3.2-3.3 along with more details about our chatbot's implementation of TensorFlow.

Additionally, the survey by Adamopoulou and Moussiades (2020) provides an overview of chatbot technology, including their history, underlying technologies, and the current types of applications. Their discussion about chatbot design, especially when it comes to user interaction and the importance of NLP, align with the goals of my project. They highlight the role of chatbots as digital helpers and state that their potential will greatly improve efficiency in data-heavy industries like mortgage finance.

Both these works support the concepts and methodologies driving my project. I have used TensorFlow as a viable and adaptable tool, demonstrated by Singh, et. al. (2018), to create my responsive chatbot. Adamopoulou and Moussiades (2020) also give the context for advancements in artificial intelligence (AI) and the broad applications of chatbots. Their work can be used as a lens to judge my project's contributions to the field of mortgage AI.

3. PROJECT DESIGN

3.1 Problem Identification and Objectives

First, my team and I needed to collect data that the chatbot could use for training. Business-specific queries were identified as the most important type of query to be addressed, mainly because they were only feasible for the business associates with the specific knowledge and expertise for it. These associates had to run several of these queries throughout the day by the request of other employees, which took up hours of their time that could be spent doing more valuable work.

We came up with a system where those associates would start logging the most common queries that were requested of them onto a shared Confluence page over time. Confluence, in this context, serves as a collaboration software used for project management and issue tracking. Alongside this process, I had several one-on-one meetings with the associates to better understand the complexities of the queries that they were running. These meetings required me to learn about the internal architecture of Fannie Mae systems and the purpose of these business queries. I also asked about any specific features or characteristics that they would like to see in the implementation of the chatbot.

The consensus was that the chatbot should be able to talk in a user-friendly manner with all mortgage employees regardless of their expertise and assist them with their commands. These commands would fall into the following categories:

- Company information
- Mortgage calculation
- Data retrieval
- Data manipulation
- Report generation

Thus, we had identified the initial objectives and use cases, meaning that we could move onto choosing the appropriate tools to address them.

3.2 Choice of Technology

We chose TensorFlow as the AI framework for this chatbot due to its compatibility with our technology stack, its computational power, and its ability to train neural networks. After acquiring the necessary permissions to use TensorFlow in our project, integrating it into Fannie Mae's existing Python codebase was straightforward. This ease of integration is due to Python being the most supported language for using TensorFlow, backed by extensive library support. After installing TensorFlow on the company systems, using it was as simple as importing the necessary "tensorflow" and "tflearn" libraries.

3.3 Chatbot Design

Once the objectives and use cases had been determined, I transformed them into a list of intents placed in the "intent file," formatted in JSON. This file is the learning foundation for the chatbot, comprised of tags, patterns, responses, and sometimes actions. This structure, shown in Figure 1, enables the chatbot to associate user commands with appropriate responses.

```
{"tag": "dus_info",
 "patterns": ["I have a question about the dus
lenders program", "I want to learn about dus",
"enquiry about dus"],
 "responses": ["Okay, what do you want to
know?", "what are you curious about?", "Okay,
what is your question"]
},
{"tag": "query_xyz",
 "patterns": ["Run business query xyz", "query
xyz please", "can you do query xyz?"],
 "responses": ["Sure, here are the results: ",
"These are the results: ", "your results: ",
"Sorry, I couldn't process your query"],
 "actions": ["query xyz"]
}
```

Figure 1: Sample Snippet from an Intent File

In addition to an intent file, I needed a neural network model for interpreting user inputs and keeping the context of the conversation. The neural network structure we used in the chatbot development involves a feed-forward network created with TensorFlow, having four layers (see Figure 2): one input layer, two hidden layers, and one output layer. During training, the intent file is used to create a "bag of words" array, representing all unique words. As described by Singh, et. al. (2018), this "bag of words" technique is used for feature extraction, where binary indicators in a vector represent the words' presence or absence in the input. When a user command is received, it goes through a three-part NLP pipeline to convert it into this "bag of words" format: Tokenization turns sentences into individual words, lemmatization reduces these words to their base forms, and part-of-speech tagging helps with the understanding of grammar structure. The model then uses this data to predict and generate responses, picking the best choice based on the probability of different tags.

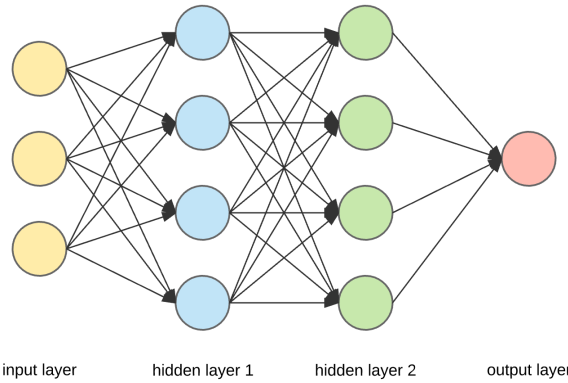

Figure 2: Neural Network Design [3]

The system also includes a feedback mechanism, where employee ratings on responses are stored and used to improve and improve interactions in the future.

3.4 Integration with Oracle Databases

Complementing this, Fannie Mae utilizes Oracle Database as one of its primary database management systems, which we used for the chatbot's query functionality. Given that Oracle Database was already well-established across internal systems at Fannie Mae and contained a large portion of the mortgage data, it was an ideal choice for the chatbot.

3.5 User Interaction Design

The chatbot, powered by TensorFlow, streamlined user interaction, making database queries seamless. Users no longer needed external software to interact with complex databases; they simply talked with the chatbot, which intelligently processed their requests.

3.6 Agile Development Process

My team and I adopted an Agile development methodology with daily stand-ups to report

progress. Using a shared Kanban board, we created and tracked stories related to the chatbot's development. For version control, we used Bitbucket, working on separate branches for each feature and merging them to the main branch upon completion. This CI/CD pipeline ensured continuous integration and delivery of features.

4. ANTICIPATED RESULTS

My internship time was not sufficient to finish the development and integration of the chatbot. However, I anticipate the chatbot to streamline operations, improve workflow, and reduce the time for completing queries. Tasks that would typically take up hours of time from business associates will be completely automated and delegated to the chatbot.

The chatbot is also anticipated to level the playing field at Fannie Mae. It democratizes access to important data, allowing all employees, regardless of their prior technical knowledge, to retrieve and learn about mortgage-related information. I expect that the chatbot's interface and human-like communication will greatly improve the user experience for employees across various departments.

5. CONCLUSION

In conclusion, the development of the deep learning chatbot at Fannie Mae represents a step forward in using AI to enhance data management in the mortgage industry. This project is set to upgrade the way Fannie Mae operates by automating data queries and bringing information access to all employees. Its integration demonstrates a major shift towards efficiency and inclusivity within the organization. Personally, this project has been a great learning experience, improving my skills in AI development and teaching me about the practical applications of technology. The chatbot's potential to streamline operations and bring about more informed decision-making marks it as an asset to Fannie Mae. This potential highlights the power of AI in modern business practices.

6. FUTURE WORK

In the future, the primary focus will be on transitioning the deep learning chatbot from the development environment to full-scale production at Fannie Mae. This transition involves rigorous testing to make sure the chatbot can be integrated with the company's existing systems. Additionally, the creation of a user interface is planned to improve the overall user experience, making the chatbot more intuitive and user-friendly. The goal is to fully integrate the chatbot into Fannie Mae's internal systems, which in turn maximizes its reach and effectiveness when it comes to automating and streamlining data-related tasks across the company.

7. ACKNOWLEDGMENTS

I want to thank my coworkers on the Fannie Mae Multifamily Securitization team for guiding me along this process.

REFERENCES

[1] Singh, R., Paste, M., Shinde, N., Patel, H., & Mishra, N. (2018). "Chatbot using TensorFlow for small Businesses." 2018 *Second International Conference on Inventive Communication and Computational Technologies (ICICCT),* 1614-1619. https://doi.org/10.1109/ICICCT.2018.8472998

[2] Adamopoulou, E., & Moussiades, L. (2020). "An Overview of Chatbot Technology." In I. Maglogiannis, L. Iliadis, & E. Pimenidis (Eds.), *Artificial Intelligence Applications and Innovations.* AIAI 2020. IFIP Advances in Information and Communication Technology, 584. Springer, Cham. https://doi.org/10.1007/978-3-030-49186-4_31

[3] Ognjanovski, G. (2020). “Everything you need to know about Neural Networks and Backpropagation—Machine Learning Easy and Fun.” Medium. https://towardsdatascience.com/everything-you-need-to-know-about-neural-networks-and-backpropagation-machine-learning-made-easy-e5285bc2be3a