EagleBot: A Chatbot Based Question Answering System to Help University Students Using Deep Learning

Muhammad Rana, Dr. Mehdi Allahyari

Department of Computer Science, Georgia Southern University

Introduction

- One of the big challenges at every university or college is to answer vast amount of students' questions in a fast and efficient manner.
- Every day students' look for hundreds of questions regarding Courses, Admissions, Scholarships, Events, Jobs, Advisors etc. In the current system, they either need to surf the university website or meet with university staff to get precise answers. Both methods, especially the latter, can be very inefficient.
- In 2016, Georgia State used a chat bot called Pounce, developed by Admithub. The bot reduced their summer melting rate on that year by 22% and got a 3.9% increase in their enrollment! What about developing a chat bot which can answer more than enrollment issues and can suggest as an advisor.

Methodology

- **Data Extraction:** For evaluation we are primarily using Georgia Southern University. We extracted all the text data from over ***28,000*** websites from www.georgiasouthern.edu by using ***Beautifulsoup***.
- **Data Preprocessing:** Then, we cleaned up all the unnecessary texts, headings, images etc from these sites and divided the whole dataset into 3 types of data. I) Structured Tabular Data, II) FAQ Data, III) Unstructured passage data from www.georgiasouthern.edu and finally loaded them into ElasticSearch Data Storage.
- **Server Design:** Whenever a user submits a query, the system will first try to understand the query using our ***Dialogflow*** powered Natural Language Understanding (NLU) module. For faster answer retrieval in EagleBot we have designed a ***3 tier*** server architecture.
  - At first, the bot tries to fetch answer from Structured Tabular Data. This route is the fastest.
  - If not found in the 1st module, it tries to find the answer from EagleBot's FAQ knowledge base, where we have prestored answers of over 1000 FAQ Data.
  - Finally, the bot dives into the whole GSU site. It finds out the best *n* candidate links and from there it finds the best possible answer using both Supervised and Unsupervised approaches. For unsupervised retrieval we have used Facebook's sentence embedding model called ***Infersent***. As supervised approaches, We have trained our Deep Learning model using ***BiLSTM*** and ***BERT***.
- **User Interface Design:** For highest number of user accessibility we have published EagleBot in 2 Platforms. 1.) Facebook Messenger 2.) Web UI. So that, anyone with Internet access can use EagleBot.

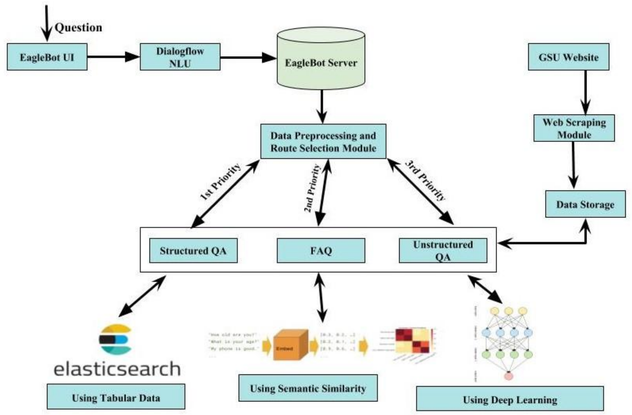

Figure 1. EagleBot high level architecture

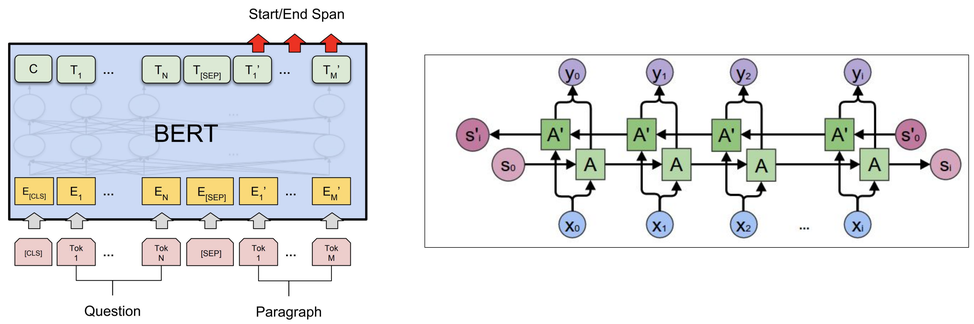

Figure 2-3. Deep Learning models used on EagleBot. (BERT and BiLSTM)

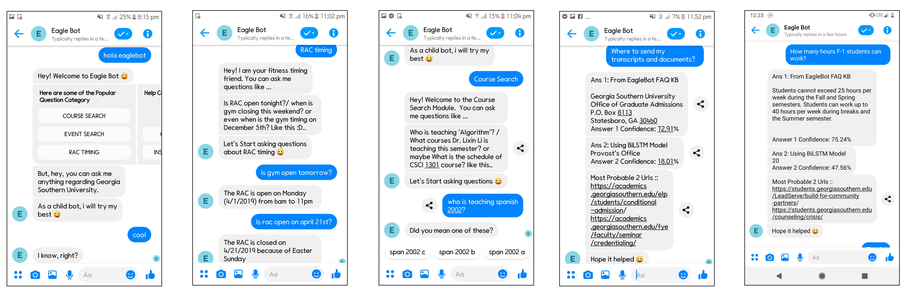

Figure 4-8. a) EagleBot introduces itself. b) Test user asks question about RAC timing. c) The bot asking back question to the user as the question was confusing. d,e) EagleBot fetching answer using it's FAQ KB and DL models.

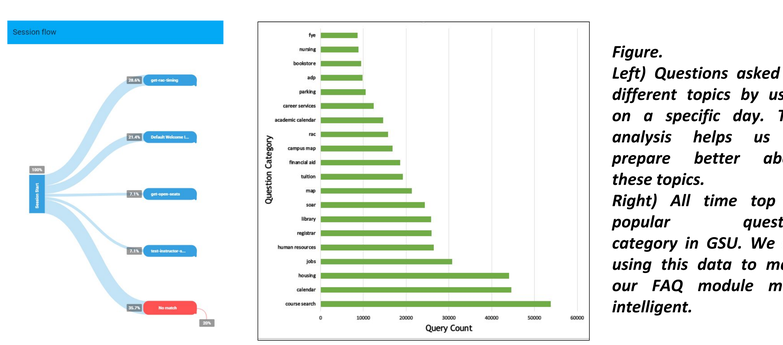

Figure.
Left) Questions asked on different topics by users on a specific day. This analysis helps us to prepare better about these topics.
Right) All time top 20 popular question category in GSU. We are using this data to make our FAQ module more intelligent.

Results

We tested the system with 20 volunteers from different student groups. In the next phase we are planning to test the system on all the students of Georgia Southern University.

- Our preliminary results show that, in more than *90%* of cases, EagleBot predicts the correct answer, if the answer is either on our FAQ KB or structured KB.
- We collected feedback from the users. In more than 70% cases, the users said the answer was helpful.
- For shorter length answer, both of the BiLSTM and BERT model works pretty well as both of these are trained on 100,000+ short QA from **SQuAD** dataset.

Findings

- We are now working on making the model understand larger answers as well. For achieving that we are preparing our custom dataset with mixture of short and long answers from university domain and we are going to finetune the models with this dataset gradually.
- From our analytics, we have started watching patterns from the questions asked by different student groups. For example the questions asked by the graduate students have a different pattern and domain set than the questions asked by freshman students. With more data we are expecting to have more distinguishable clusters which can be a valuable asset for research in this topic.
- We are logging every question, answer and user feedback about the reply in the system and using it for retraining the bot. So, every day it's becoming more matured
- The more curated and cleaner the data set is, the better performance we get from EagleBot. So, in the near future we expect to get far better results then now.

Applications

- EagleBot has a good possibility to be applied on several different domains. Given proper data, this system can be replicated for many institutions and for many domains where we need constant Question Answering assistance. For example: In ***Customer Care Service*** of any corporate house, where they need 24/7 assistance.
- Besides answering administrative questions, in the future we can use EagleBot as a ***Virtual Tutor***. EagleBot can help students on their academic tasks, such as their assignments, quizzes and exams. The bot can also monitor their performance by working as a personalized assistant.
- By utilizing EagleBot's Question-Answering capability, we can also use this system as a ***Virtual Doctor*** for Patients. It can answer their health-questions instantly. Additionally, this can be used to help counseling students with their psychological issues.
- How EagleBot is going to perform on different languages can be an interesting observation.