

CSCI 5622 - Machine Learning Project 1: Group Formation

Group Members:

- 1) Jay Patel
- 2) Mohan Dwarampudi
- 3) Saptarshi Mitra

Project:

Quest Quencher: A question answering model. This model will be trained with long passages of text and it will be able to answer questions which are specific to the domain we have trained the model in.

Motivation and use of Machine Learning:

During the last Black Friday sale, we purchased an Alexa and were really enthralled by its question answering skills. Also, during the last semester, we had taken the Natural Language Processing (CSCI 5832) class, taught by Professor James Martin and found this topic of **Question Answering** interesting, especially the part where he showed us a video in which IBM's Watson defeated a contestant on Jeopardy. So, we wanted to really delve into the topic to develop this model using Machine Learning and Neural Networks such as **LSTM, Bidirectional Attention Flow for Machine**(BiDAF model), etc. and learn more about the concepts of Computational Linguistics like Dependency Parsing, Word Embeddings(Representation of Words as Vectors in Space)

Applications:

It can be useful to auto grade reading comprehensions in competitive exams like GRE and SAT. Local businesses can use this to make Chatbots which can answer the basic FAQs of the customers.

Dataset:

We're going to use the SQUAD Dataset (Stanford Question Answering Dataset). Stanford Question Answering Dataset (SQUAD) is a reading comprehension dataset, consisting of questions posed by crowd workers on a set of Wikipedia articles, where the answer to every question is a segment of text, or span, from the corresponding reading passage, or the question might be unanswerable. The dataset is comprised of multiple paragraphs with context as the text and multiple questions and answers based on a single paragraph. Example of a datapoint from the dataset is as follows:

- a. **Context:** Architecturally, the school has a Catholic character. Atop the Main Building's gold dome is a golden statue of the Virgin Mary. Immediately in front of the Main Building and facing it, is a copper statue of Christ with arms upraised with the legend \"Venite Ad Me Omnes\". Next to the Main Building is the Basilica of the Sacred Heart. Immediately behind the basilica is the Grotto, a

Marian place of prayer and reflection. It is a replica of the grotto at Lourdes, France where the Virgin Mary reputedly appeared to Saint Bernadette Soubirous in 1858. At the end of the main drive (and in a direct line that connects through 3 statues and the Gold Dome), is a simple, modern stone statue of Mary.

b. Question: "To whom did the Virgin Mary allegedly appear in 1858 in Lourdes France?"

Answer Start: 515 (The position of first character- "S" of the answer in the paragraph)

Answer: "Saint Bernadette Soubirous"