

---

---

# NLP TERM PROJECT

## 2019

TOPIC : Biomedical Question Answering

Team Name : Mame

---



# CONTENTS

1. PROBLEM STATEMENT
2. INTRODUCTION
3. PREVIOUS WORKS
4. MODEL AND ITS ARCHITECTURE
5. CONCLUSIONS

---

# PROBLEM STATEMENT

- Our task is focussed on Biomedical Question Answering.
  - We are given a question and a set of relevant articles independent of the question and from that we are required to extract the relevant article along with the word-span of the text relevant for answering the particular question.
  - In this task we would like to answer the factoid and list type questions.
  - BioASQ Task 7b in SQUAD format is used in this task.
-

---

# INTRODUCTION

For each observation in the training set we have a **context**, **question** & **text**.

**'context':** 'Beyoncé Giselle Knowles-Carter (/bi:ˈjɒnsər/ bee-YON-say) (born September 4, 1981) is an American singer, songwriter, record producer and actress. Born and raised in Houston, Texas, she performed in various singing and dancing competitions as a child, and rose to fame in the late 1990s as lead singer of R&B girl-group Destiny's Child. Managed by her father, Mathew Knowles, the group became one of the world's best-selling girl groups of all time. Their hiatus saw the release of Beyoncé's debut album, *Dangerously in Love* (2003), which established her as a solo artist worldwide, earned five Grammy Awards and featured the Billboard Hot 100 number-one singles "Crazy in Love" and "Baby Boy".',  
**'text':** 'in the late 1990s'  
**'question':** 'When did Beyonce start becoming popular?'

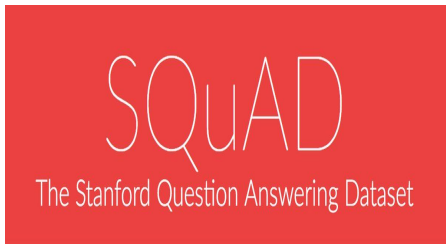
Sentence having the right answer

Exact Answer

- 
- The goal is to find the text for any new question and context provided.
  - Getting the sentence having the right answer.
  - Once the sentence is finalized, getting the correct answer from the sentence .
  - The datasets containing factoid and list type questions were converted into the format of the SQuAD datasets .
  - The dataset in the SQuAD format consists of passages and their respective question-answer sets.
-

---

# SQuAD Datasets



- A passage is an article which contains answers or clues for answers and is denoted as the context in the dataset.
  - The length of a passage varies from a sentence to a paragraph. An exact answer may or may not exist in the passage, depending on the task.
  - According to the rules of the BioASQ Challenge, all the factoid and list type questions should be answerable with the given passages. An exact answer and its starting position are provided in the answers field.
-

---

# Previous Works

It is worth noting that most successful models have used long short term memory units (LSTMs) to capture interactions over long sequences of words. Successful models also use some sort of attention mechanism, to capture relevance of words in the question and context paragraph as an intermediate step to answering the question. Examples :-

- Dynamic Coattention Networks
  - Dynamic Chunk Readers
-

---

# BIDAF Model

- The main model used in solving this task was **Bidirectional Attention Flow (BIDAF)** model.
  - We first used a word embedding layer that maps each word to an embedding vector (using pre-trained GloVe embeddings to represent words), for both the question and context paragraph.
  - We then used a sentence based embedding layer known as Infersent.
-



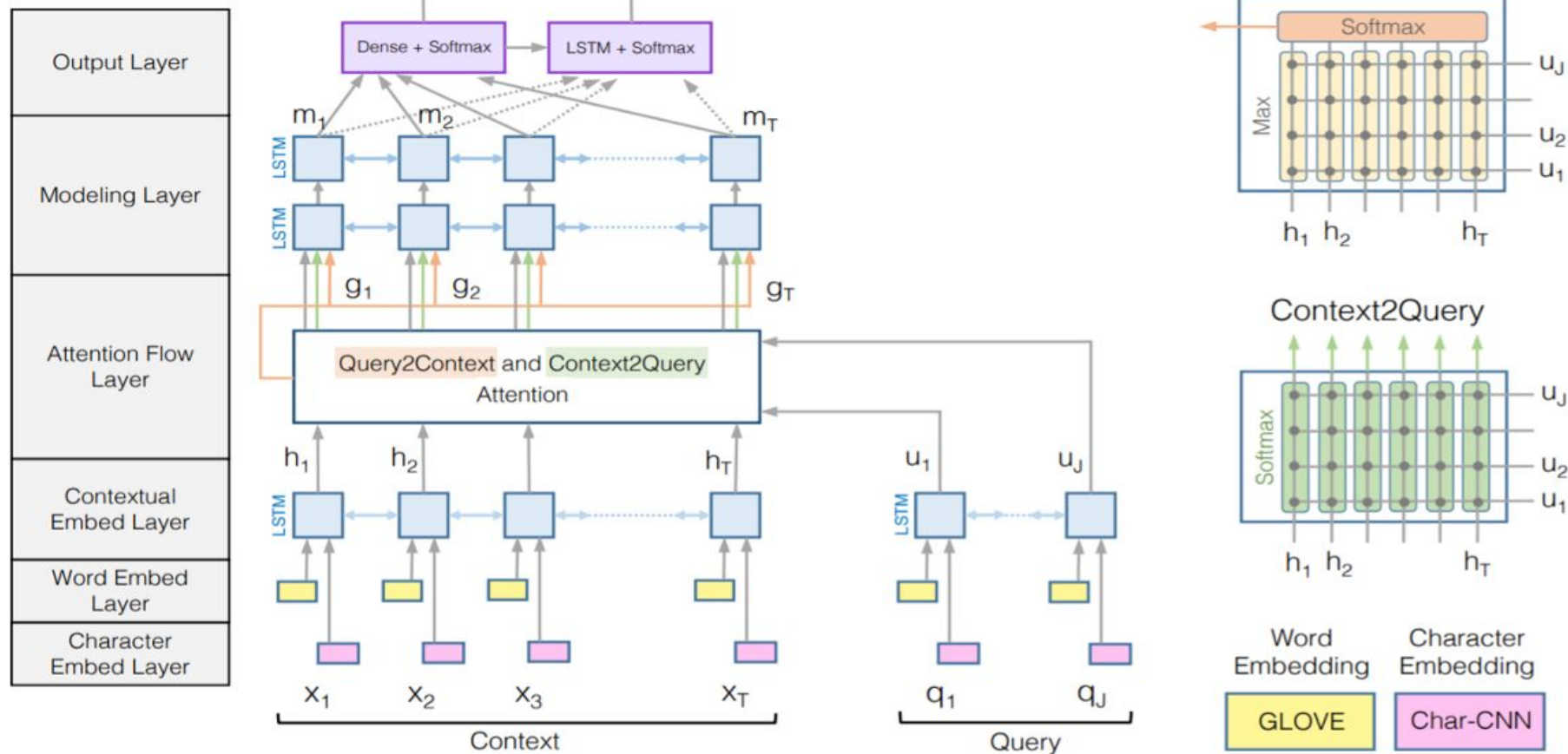
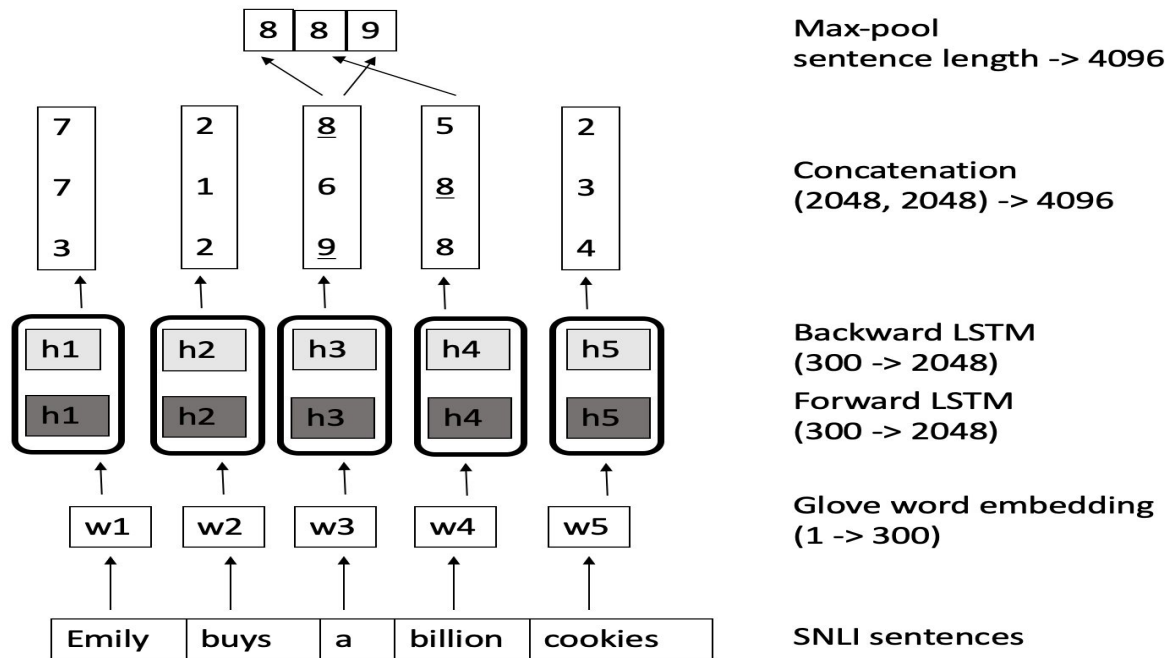


Figure 1: BiDirectional Attention Flow Model (best viewed in color)

# Sentence Embedding



---

# CONCLUSIONS

- We implemented a Bidirectional Attention Flow model.
  - We tested the model on simple baseline models.
  - The results were satisfactory and better results can be obtained with access to processing power.
  - As the code for yes/no based questions had some errors so we were not able to get results for that part of the task.
-

---

---

**THANK YOU**

---