

Question Answering + Score Calculation

July 16, 2024

```
[5]: import pandas as pd
import numpy as np
from transformers import pipeline
import pandas as pd
import ast
import re
import warnings
warnings.filterwarnings("ignore")
```

```
[6]: model_name = "distilbert-base-uncased-distilled-squad"
qa_pipeline = pipeline("question-answering", model=model_name)

context = "UCLA is an American University from Los Angeles, California"
question = "Where is UCLA?"

result = qa_pipeline(question=question, context=context)
print(result)
```

```
{'score': 0.8982033133506775, 'start': 36, 'end': 59, 'answer': 'Los Angeles, California'}
```

1 Obtaining Answers to all questions

The function below does the following:

- transform the context from a string to a list type
- each item in this list contains 2 elements: A title, and a text about that title
- the text is cleaned by first having all of its words joined and then by removing whitespaces
- the text is further cleaned by removing square brackets, and forward and backward dashes from it
- the title is the joined together with the text and added to the cell where it belongs

```
[7]: df = pd.read_csv('/Users/sandinatatu/Desktop/qa_dataset.csv', index_col = 0)
print(df.iloc[0]['context'])
```

```
['Radio City (Indian radio station)', ["Radio City is India's first private FM radio station and was started on 3 July 2001.", ' It broadcasts on 91.1 (earlier 91.0 in most cities) megahertz from Mumbai (where it was started in 2004),
```

Bengaluru (started first in 2001), Lucknow and New Delhi (since 2003).', ' It plays Hindi, English and regional songs.', ' It was launched in Hyderabad in March 2006, in Chennai on 7 July 2006 and in Visakhapatnam October 2007.', ' Radio City recently forayed into New Media in May 2008 with the launch of a music portal - PlanetRadiocity.com that offers music related news, videos, songs, and other music-related features.', ' The Radio station currently plays a mix of Hindi and Regional music.', ' Abraham Thomas is the CEO of the company.']], ['History of Albanian football', ['Football in Albania existed before the Albanian Football Federation (FSHF) was created.', " This was evidenced by the team's registration at the Balkan Cup tournament during 1929-1931, which started in 1929 (although Albania eventually had pressure from the teams because of competition, competition started first and was strong enough in the duels) .", ' Albanian National Team was founded on June 6, 1930, but Albania had to wait 16 years to play its first international match and then defeated Yugoslavia in 1946.', ' In 1932, Albania joined FIFA (during the 12-16 June convention) And in 1954 she was one of the founding members of UEFA.']], ['Echosmith', ['Echosmith is an American, Corporate indie pop band formed in February 2009 in Chino, California.', ' Originally formed as a quartet of siblings, the band currently consists of Sydney, Noah and Graham Sierota, following the departure of eldest sibling Jamie in late 2016.', ' Echosmith started first as "Ready Set Go!'", ' until they signed to Warner Bros.', ' Records in May 2012.', ' They are best known for their hit song "Cool Kids", which reached number 13 on the "Billboard" Hot 100 and was certified double platinum by the RIAA with over 1,200,000 sales in the United States and also double platinum by ARIA in Australia.', ' The song was Warner Bros.', " Records' fifth-biggest-selling-digital song of 2014, with 1.3 million downloads sold.", ' The band\'s debut album, "Talking Dreams", was released on October 8, 2013.']], ["Women's colleges in the Southern United States", ["Women's colleges in the Southern United States refers to undergraduate, bachelor's degree-granting institutions, often liberal arts colleges, whose student populations consist exclusively or almost exclusively of women, located in the Southern United States.", " Many started first as girls' seminaries or academies.", ' Salem College is the oldest female educational institution in the South and Wesleyan College is the first that was established specifically as a college for women.', ' Some schools, such as Mary Baldwin University and Salem College, offer coeducational courses at the graduate level.']], ['First Arthur County Courthouse and Jail', ['The First Arthur County Courthouse and Jail, was perhaps the smallest court house in the United States, and serves now as a museum.']], ["Arthur's Magazine", ["Arthur's Magazine (1844-1846) was an American literary periodical published in Philadelphia in the 19th century.", ' Edited by T.S. Arthur, it featured work by Edgar A. Poe, J.H. Ingraham, Sarah Josepha Hale, Thomas G. Spear, and others.', ' In May 1846 it was merged into "Godey\'s Lady\'s Book".']], ['2014-15 Ukrainian Hockey Championship', ['The 2014-15 Ukrainian Hockey Championship was the 23rd season of the Ukrainian Hockey Championship.', ' Only four teams participated in the league this season, because of the instability in Ukraine and that most of the clubs had economical issues.', ' Generals Kiev was the only team that participated in the league the previous season, and the season started first after the year-end of 2014.', ']

The regular season included just 12 rounds, where all the teams went to the semifinals.', ' In the final, ATEK Kiev defeated the regular season winner HK Kremenchuk.']], ['First for Women', ["First for Women is a woman's magazine published by Bauer Media Group in the USA.", ' The magazine was started in 1989.', ' It is based in Englewood Cliffs, New Jersey.', ' In 2011 the circulation of the magazine was 1,310,696 copies.']], ['Freeway Complex Fire', ['The Freeway Complex Fire was a 2008 wildfire in the Santa Ana Canyon area of Orange County, California.', ' The fire started as two separate fires on November 15, 2008.', ' The "Freeway Fire" started first shortly after 9am with the "Landfill Fire" igniting approximately 2 hours later.', ' These two separate fires merged a day later and ultimately destroyed 314 residences in Anaheim Hills and Yorba Linda.']], ['William Rast', ['William Rast is an American clothing line founded by Justin Timberlake and Trace Ayala.', ' It is most known for their premium jeans.', ' On October 17, 2006, Justin Timberlake and Trace Ayala put on their first fashion show to launch their new William Rast clothing line.', ' The label also produces other clothing items such as jackets and tops.', ' The company started first as a denim line, later evolving into a men's and women's clothing line.']]

```
[9]: def clean_data(serialized_list):
    try:
        data = ast.literal_eval(serialized_list)
    except ValueError as e:
        print(f"Error converting to list: {e}")
        return serialized_list

    result_string = ""
    for item in data:
        title = item[0]
        details = item[1]

        cleaned_text = " ".join(details)
        cleaned_text = re.sub(r'\s+', ' ', cleaned_text).strip()

        cleaned_text = re.sub(r'[\[\]\{\}\|\,]', '', cleaned_text)

        result_string += f"{title}: {cleaned_text}\n"

    return result_string.strip()

df['context'] = df['context'].apply(clean_data)
```

```
[10]: print(df.iloc[0]['context'])
```

Radio City (Indian radio station): Radio City is India's first private FM radio station and was started on 3 July 2001. It broadcasts on 91.1 (earlier 91.0 in most cities) megahertz from Mumbai (where it was started in 2004), Bengaluru (started first in 2001), Lucknow and New Delhi (since 2003). It plays Hindi,

English and regional songs. It was launched in Hyderabad in March 2006, in Chennai on 7 July 2006 and in Visakhapatnam October 2007. Radio City recently forayed into New Media in May 2008 with the launch of a music portal - PlanetRadiocity.com that offers music related news, videos, songs, and other music-related features. The Radio station currently plays a mix of Hindi and Regional music. Abraham Thomas is the CEO of the company.

History of Albanian football: Football in Albania existed before the Albanian Football Federation (FSHF) was created. This was evidenced by the team's registration at the Balkan Cup tournament during 1929-1931, which started in 1929 (although Albania eventually had pressure from the teams because of competition, competition started first and was strong enough in the duels) . Albanian National Team was founded on June 6, 1930, but Albania had to wait 16 years to play its first international match and then defeated Yugoslavia in 1946. In 1932, Albania joined FIFA (during the 12-16 June convention) And in 1954 she was one of the founding members of UEFA.

Echosmith: Echosmith is an American, Corporate indie pop band formed in February 2009 in Chino, California. Originally formed as a quartet of siblings, the band currently consists of Sydney, Noah and Graham Sierota, following the departure of eldest sibling Jamie in late 2016. Echosmith started first as "Ready Set Go!" until they signed to Warner Bros. Records in May 2012. They are best known for their hit song "Cool Kids", which reached number 13 on the "Billboard" Hot 100 and was certified double platinum by the RIAA with over 1,200,000 sales in the United States and also double platinum by ARIA in Australia. The song was Warner Bros. Records' fifth-biggest-selling-digital song of 2014, with 1.3 million downloads sold. The band's debut album, "Talking Dreams", was released on October 8, 2013.

Women's colleges in the Southern United States: Women's colleges in the Southern United States refers to undergraduate, bachelor's degree-granting institutions, often liberal arts colleges, whose student populations consist exclusively or almost exclusively of women, located in the Southern United States. Many started first as girls' seminaries or academies. Salem College is the oldest female educational institution in the South and Wesleyan College is the first that was established specifically as a college for women. Some schools, such as Mary Baldwin University and Salem College, offer coeducational courses at the graduate level.

First Arthur County Courthouse and Jail: The First Arthur County Courthouse and Jail, was perhaps the smallest court house in the United States, and serves now as a museum.

Arthur's Magazine: Arthur's Magazine (1844-1846) was an American literary periodical published in Philadelphia in the 19th century. Edited by T.S. Arthur, it featured work by Edgar A. Poe, J.H. Ingraham, Sarah Josepha Hale, Thomas G. Spear, and others. In May 1846 it was merged into "Godey's Lady's Book".

2014-15 Ukrainian Hockey Championship: The 2014-15 Ukrainian Hockey Championship was the 23rd season of the Ukrainian Hockey Championship. Only four teams participated in the league this season, because of the instability in Ukraine and that most of the clubs had economical issues. Generals Kiev was the only team that participated in the league the previous season, and the season started first after the year-end of 2014. The regular season included just 12 rounds,

where all the teams went to the semifinals. In the final, ATEK Kiev defeated the regular season winner HK Kremenchuk.

First for Women: First for Women is a woman's magazine published by Bauer Media Group in the USA. The magazine was started in 1989. It is based in Englewood Cliffs, New Jersey. In 2011 the circulation of the magazine was 1,310,696 copies.

Freeway Complex Fire: The Freeway Complex Fire was a 2008 wildfire in the Santa Ana Canyon area of Orange County, California. The fire started as two separate fires on November 15, 2008. The "Freeway Fire" started first shortly after 9am with the "Landfill Fire" igniting approximately 2 hours later. These two separate fires merged a day later and ultimately destroyed 314 residences in Anaheim Hills and Yorba Linda.

William Rast: William Rast is an American clothing line founded by Justin Timberlake and Trace Ayala. It is most known for their premium jeans. On October 17, 2006, Justin Timberlake and Trace Ayala put on their first fashion show to launch their new William Rast clothing line. The label also produces other clothing items such as jackets and tops. The company started first as a denim line, later evolving into a men's and women's clothing line.

Below we make the LLM answer all of the questions.

```
[ ]: results = []
i=0

for _, row in df.tail(40000).iterrows():

    i+=1
    if i%500 ==0:
        print(i)

    result = qa_pipeline(question=row['question'], context=row['context'])
    results.append({
        'question': row['question'],
        'predicted_answer': result['answer'],
        'actual_answer': row['answer']
    })

results_df = pd.DataFrame(results)
```

2 Calculating Scores

Prior to calculating the scores, we need to “clean” the answers. Below I clean both the predicted and actual answers by replacing dashes (, /, or -) with a space, removing non-alphanumeric characters and then transforming all words to lowercase.

```
[ ]: import re

def clean_ans(ans):
    ans = re.sub(r"\\\/-)", ' ', ans)
    ans = re.sub(r'^a-zA-Z0-9\s', '', ans)
    return ans.lower()

results_df['predicted_answer'] = results_df['predicted_answer'].apply(clean_ans)
results_df['actual_answer'] = results_df['actual_answer'].apply(clean_ans)

results_df = results_df[(results_df['predicted_answer'].str.len()>0) &
    ↪(results_df['actual_answer'].str.len()>0)]
```

2.0.1 BLEU SCORE

Below I write a function calculating the BLEU score. Then, I create a new column ('bleu_score') in the answers dataframe which contains the BLEU score for each answer.

```
[6]: #Note: you will need to do this:
```

```
#import nltk
#nltk.download('punkt')
```

```
[7]: from nltk.translate.bleu_score import sentence_bleu
from nltk.tokenize import word_tokenize
```

```
[8]: # word_tokenize transformes our cleaned predicted and actual answers to lists
    ↪of words
# Because of the way BLEU is computed, an answer shorter than 4 words always
    ↪results in a score of 0.
# I modified that by specifying the max number of n-grams BLEU should go up to
# then bleu score is calculated. for whatever reason, when using sentence_bleu,
# the first argument needs to be in square brackets
```

```
[9]: def calculate_bleu(predicted, actual):
    reference = word_tokenize(actual)
    candidate = word_tokenize(predicted)
    max_n = min(4, len(reference))
    weights = tuple((1/max_n,)*max_n)
    return sentence_bleu([reference], candidate, weights=weights)
```

```
[10]: results_df['bleu_score'] = results_df.apply(lambda row:
    ↪calculate_bleu(row['predicted_answer'], row['actual_answer']), axis = 1)
```

2.0.2 ROUGE-L SCORE

Below I write a function calculating the ROUGE-L score. Then, I create a new column ('rouge-l_score') in the answers dataframe which contains the ROUGE-L F1 score for each answer.

```
[11]: from rouge import Rouge

def calculate_rouge_l(predicted, actual):
    rouge = Rouge()
    scores = rouge.get_scores(predicted, actual)
    return scores[0]['rouge-l']['f']

results_df['rouge-l_score'] = results_df.apply(lambda row:
    ↪calculate_rouge_l(row['predicted_answer'], row['actual_answer']), axis=1)
```

2.0.3 METEOR Score

```
[12]: from nltk.translate.meteor_score import meteor_score, single_meteor_score
from nltk.tokenize import word_tokenize
```

```
[13]: def calc_METEOR(predicted, actual):
    reference_tokens = word_tokenize(predicted)
    hypothesis_tokens = word_tokenize(actual)
    return meteor_score([reference_tokens], hypothesis_tokens)
```

```
[14]: results_df['METEOR'] = results_df.apply(lambda row:
    ↪calc_METEOR(row['predicted_answer'], row['actual_answer']), axis=1)
```

2.0.4 Word Error Rate

```
[15]: from jiwer import wer
```

```
[16]: results_df['Word-Error-Rate'] = results_df.apply(lambda row:
    ↪wer(row['predicted_answer'], row['actual_answer']), axis=1)
```

2.0.5 Cosine Similarity

```
[17]: from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.metrics.pairwise import cosine_similarity
```

```
[18]: tfidf = TfidfVectorizer()
all_words = results_df['predicted_answer'].tolist() +
    ↪results_df['actual_answer'].tolist()
tfidf.fit(all_words)
```

```
[18]: TfidfVectorizer()
```

```
[19]: def calculate_cosine_similarity(predicted, actual, vectorizer):
```

```
    matrix = vectorizer.transform([predicted, actual])
```

```
    similarity = cosine_similarity(matrix[0:1], matrix[1:2])
```

```
    return similarity[0,0]
```

```
[20]: results_df['Cosine Similarity'] = results_df.apply(lambda row:
    ↪calculate_cosine_similarity(row['predicted_answer'], row['actual_answer'],
    ↪tfidf), axis=1)
```

```
[21]: results_df.head()
```

```
[21]: question \
```

```
0 Which magazine was started first Arthur's Maga...
```

```
1 The Oberoi family is part of a hotel company t...
```

```
2 Musician and satirist Allie Goertz wrote a son...
```

```
3 What nationality was James Henry Miller's wife?
```

```
4 Cadmium Chloride is slightly soluble in this c...
```

	predicted_answer	actual_answer	bleu_score \
0	first for women first for women	arthurs magazine	0.0
1	delhi	delhi	1.0
2	allison beth president richard nixon		0.0
3	british	american	0.0
4	methyl alcohol	alcohol	0.5

	rouge-1_score	METEOR	Word-Error-Rate	Cosine Similarity
0	0.000000	0.000000	1.0	0.000000
1	1.000000	0.500000	0.0	1.000000
2	0.000000	0.000000	1.5	0.000000
3	0.000000	0.000000	1.0	0.000000
4	0.666667	0.263158	0.5	0.680118