

# **ITCS 6100 – Big Data Analytics for Competitive Advantage**

Project

## **MC-TACO: Temporal Comprehension**

Project Report submitted by:

**Hiten Changlani  
Swapn Shah  
Shimpli Kalse  
Amulya Chepyala**

*At*

***The University of North Carolina at Charlotte***

Current date: 05/05/2021

## Abstract:

Understanding the passing of time is important for comprehending events described in natural language. Since people rarely say what they mean, common sense awareness of various temporal aspects of events, such as length, frequency, and temporal order, is often needed. However, this critical issue has received only sporadic attention to date. As per research paper: 'Going on a vacation" takes longer than "Going for a walk"' A Study of Temporal Commonsense Understanding". The best current methods used on MCTACO are still far behind human performance, by about 20%" This project deals with training the model based on an available set of questions to indicate answers with common sense based on trained data.

The overall objective of this project is:

Our project is focused on temporal commonsense. We evaluate current AI systems on temporal commonsense tasks that measure both the knowledge that these systems possess as well as their ability to reason with and use that knowledge in context.

## 1. INTRODUCTION:

MC-TACO is a dataset of 13k question-answer pairs that require temporal commonsense comprehension. The dataset contains five temporal properties, (1) duration (how long an event takes), (2) temporal ordering (typical order of events), (3) typical time (when an event occurs), (4) frequency (how often an event occurs), and (5) stationarity (whether a state is maintained for a very long time or indefinitely). We hope that this dataset can promote the future exploration of this class of reasoning problems.

Look at the example given below for better understanding:

Statement	Question	Answer	Label	Category
Durer's father died in 1502, and his mother di...	How long was his mother ill?	six centuries	0	0
Durer's father died in 1502, and his mother di...	How long was his mother ill?	she was ill for 90 years	0	0
Durer's father died in 1502, and his mother di...	How long was his mother ill?	6 months	1	0
Durer's father died in 1502, and his mother di...	How long was his mother ill?	six minutes	0	0
Durer's father died in 1502, and his mother di...	How long was his mother ill?	she was ill for 30 years	0	0

It is common logic that no person can be ill for 6 centuries, 90 years, 30 years or 6 minutes. Thus, we see a label of 0 against these answers while a label 1 is assigned where the answer is 6 months as it is a realistic answer.

This is a work mainly done at **University of Pennsylvania** in collaboration with **AI2 Aristo**, which is the Allen Institute for Artificial Intelligence's research initiative on machine reasoning. The mission is to understand and develop capabilities for natural language understanding, by focusing on the reasoning aspect of human behavior.

Performance will be measured using two metrics:

- Exact Match -- the average number of questions for which all the candidate answers are predicted correctly.
- F1 -- which is slightly more relaxed than EM, which measures the overlap between one's predictions and the ground truth, by forming a geometric mean of Precision-Recall measures.

## **2. SECTION TITLE DESCRIBING YOUR PROBLEM**

The code for the neural models as reported in the EMNLP-2019 paper is available in the MC-TACO GitHub repository. This repository contains ESIM, BERT, RoBERTa predictions. MC-TACO to probe has been used to check the capability of systems on temporal commonsense understanding. As per research it is found that systems equipped with state-of-the-art language models such as ELMo and BERT are still far behind humans, thus motivating future research in this area.

## **3. SECTION TITLE DESCRIBING YOUR DATA**

### Data set Description:

In each file (dev\_3783.tsv and test\_9442.tsv), there are lines of tab-separated data, each line representing an instance of a question-answer pair.

Specifically, the format is as the following:

sentence \t question \t answer \t label \t category

\* sentence: a sentence where the question is based on.

\* question: a question querying some temporal knowledge.

\* answer: a potential answer to the question. all lowercase.

\* label: whether the answer is a correct (likely) answer. "yes" indicates the answer is likely, "no" otherwise.

\* category: the temporal category the question belongs to one of Event Ordering, Event Duration", Frequency, Stationarity, or Typical Time.

## **4. SECTION TITLE DESCRIBING YOUR APPROACH AND EXPERIMENTS**

We completed the BERT model code and all the errors and bugs in the code have been successfully solved. Also, we have improved accuracy to some extent. We ran 8 experiments and made sure to avoid overfitting the model.

### **4.1. Subsection motivating your approach**

The basic way to train the model for prediction in case of our dataset was using the TF-IDF model as there is a lot of information available on the internet to study and implement the model. Once we were able to obtain a solution in the TF-IDF model, we switched to the BERT model since it is more complicated than the previous one. We got rid of all the errors and bugs and were able to successfully make the BERT code working. We have improved accuracy of the code to some extent. We shall try to improve the accuracy even further and try to use higher end models once we are done. Once done, we will finally switch to the RoBERTa model as there is little information about the concepts and implementation available on the internet.

### **4.2. Subsection describing the details of your methods**

The methods used for our project are Naive Bayes, TF, TF-IDF for the initial project work. The data is read from the file and is preprocessed and cleaned. The next step involves vectorization of the data using the transformers module. The vectorized data is then merged and a TF-IDF model is built on that. We have eventually switched to the BERT model following similar Machine Learning concepts and all the errors and issues that we faced with this code have been solved. We shall now try to improve the accuracy even further before we move on to the higher end models like RoBERTa or RAG.

### **4.3. Subsection details of your experiments**

Naive Bayes code execution:

The screenshot shows the MC-taco JupyterLab interface. The left sidebar displays a file explorer with a directory structure including 'bin', 'boot', 'content', 'drive', 'sample\_data', 'dev\_3783.tsv', 'test\_9442.tsv', 'datalab', 'dev', 'etc', 'home', 'lib', 'lib32', 'lib64', 'media', 'mnt', 'opt', 'proc', and 'root'. The main area contains a code editor with the following Python code:

```
[15] df2 = data_merging(tfidf_statement, tfidf_question, tfidf_answer)

df2.shape

(9441, 6782)

[17] def model_building(df2, df):
    X_train, X_test, y_train, y_test = train_test_split(df2, df['Label'], test_size=0.3)
    NBClassifier = MultinomialNB()
    NBClassifier.fit(X_train, y_train)
    predictions = NBClassifier.predict(X_test)
    accuracy = NBClassifier.score(X_test, y_test)
    return predictions, accuracy, confusion_matrix(y_test, predictions), f1_score(y_test, predictions), NBClassifier

[18] predictions, accuracy, conf_matrix, f_score, NBClassifier = model_building(df2, df)

[19] print(accuracy)

0.7084362866219556

[20] print(f_score)

0.540890868596883
```

The status bar at the bottom indicates '0s completed at 9:40 AM'.

BERT model code execution:

The screenshot shows the MC TACO - BERT.ipynb JupyterLab interface. The left sidebar displays a file explorer with a directory structure including 'bin', 'boot', 'content', 'drive', 'sample\_data', 'dev\_3783.tsv', 'test\_9442.tsv', 'datalab', 'dev', 'etc', 'home', 'lib', 'lib32', 'lib64', 'media', 'mnt', 'opt', 'proc', and 'root'. The main area contains a code editor with the following Python code:

```
text_model.compile(loss="binary_crossentropy", optimizer="adam", metrics=["accuracy"])

text_model.fit(train_data, epochs=NB_EPOCHS)

Epoch 1/5
267/267 [=====] - 30s 108ms/step - loss: 0.6166 - accuracy: 0.6686
Epoch 2/5
267/267 [=====] - 29s 107ms/step - loss: 0.5237 - accuracy: 0.7395
Epoch 3/5
267/267 [=====] - 29s 107ms/step - loss: 0.4135 - accuracy: 0.8200
Epoch 4/5
267/267 [=====] - 29s 107ms/step - loss: 0.3697 - accuracy: 0.8408
Epoch 5/5
267/267 [=====] - 29s 108ms/step - loss: 0.3409 - accuracy: 0.8579
<tensorflow.python.keras.callbacks.History at 0x7fb82bd99610>

[90] results = text_model.evaluate(test_data)
print(results)

29/29 [=====] - 1s 19ms/step - loss: 0.5289 - accuracy: 0.7554
[0.5288770794868469, 0.7553879022598267]

[91] from sklearn.metrics import f1_score
```

The status bar at the bottom indicates '0s completed at 9:36 AM'.

Sr. No.	Model	Parameters	No. of experiments
1	TF-IDF model	Predictions, accuracy, conf_matrix, f_score were used	22
2	BERT model	Epochs, batch size and number of layers were experimented	26

## 5. DISCUSSION AND RELATED WORK

Hiten Changlani came up with the idea of using the RoBERTa/RAG model as it is the latest available model and has far more features than the traditional BERT model. It was difficult to obtain any information on the RoBERTa model over the internet as it is very new and not a lot of work has been done using that model. So in order to move to the RoBERTa model, we needed a base. Swapn Shah used naive Bayes theorem and TF-IDF model to train the data and predict the scenarios. It works as expected. Even the BERT model has been successfully implemented now with increased accuracy.

Shimpli Kalse provided us with the base information from the internet and has been working on the research since the start. She shares all the information available about the project to the group members which helps a lot in moving ahead with the code and the project overall. Amulya Chepyala has taken the Machine Learning course this semester and she provides us with the creative ideas of implementing the code. We all schedule a team meeting every week and discuss the next steps to be taken. We assign the tasks to each member of the group and a follow-up is done for the same. If we face problems, we contact either Prof. Z or TA and ask for their help. Once the task for the week has been achieved, all the members work on the report together.

### Error Analysis & Mitigation Steps:

- We faced error in converting dataset to Tensorflow compliant dataset
- There were issues in generating batches from the dataset
- We faced initial problems in data engineering & tokenization of text
- There were some errors in creating convolution layers
- We checked the shapes of the input and output layer of the dataset and found out the discrepancy in both. This is how we were able to find a solution to the problem
- We transformed the dataset to match the padded shape arguments
- We had the data in the .tsv format which was successfully converted to .csv format
- We chose to go on with 1D convolution for the time being instead of going for higher dimension convolutions which were generating errors

## 6. CONCLUSIONS AND OPEN PROBLEMS

The project is progressing at expected pace and we are able to tackle our problems with the help of Prof. Z and TA Venkata. We have solved all the errors and issues with the code in the BERT model and have worked on improving the accuracy. Once the accuracy of the BERT model is satisfactory, we then move ahead with the RoBERTa model as planned. It is a difficult project as none of us have prior experience with Big Data or Machine Learning but we are optimistically moving forward with our project and learning a lot while we do so.

## 7. CONTRIBUTIONS AND CONFLICT OF INTEREST DECLARATIONS

There was no conflict of interest within the members of the team with respect to any of the approaches to solving the problem statement. All the ideas were welcomed, tried out and the best among them were taken into the final solution. All the members of the team had their equal share of inputs. We used to connect twice a week to discuss the new ideas to be implemented in the coming week and to proceed with the progress updates a day before the presentation day. We contacted Prof. Wlodek Zadrozny and TA Venkata whenever we faced some issues and needed some help with the problems, we were unable to solve on our own.

### Progress summary:

Date	Measure you use	Value of the measure	Method	Number of experiments	Comments
3/10/2021	Exact match F1: geometric mean	N/A	We are yet to decide if we want to go for NLTK or TextBlob model	0	Since we have midterms in the coming week, we are looking into the dataset. We have also collected research papers related to the topic
3/17/2021	Exact Match & F1	NA	Checking for BERT and RoBERTa model to be used for language modelling	0	Code given on Allen Ai is old and throwing errors hence working on troubleshooting part
3/24/2021	Exact Match & F1	Accuracy: 69.35%	Naive Bayes theorem, tf, tf-idf concepts used for new code	4	Code is now working. However, we need to figure out implementation in RoBERTa
4/5/2021	Exact Match & F1	Accuracy: 69.80%	Naive Bayes theorem, tf, tf-idf concepts used for new code	10	Added a few features however accuracy is not improving.

4/11/2021	Exact Match & F1	Accuracy: 68.93%for TF-IDF  BERT model not working properly	BERT model	8	Tried using the BERT model. We are having some issues with the code. Have contacted Venkata for help.
4/25/2021	Exact Match & F1	Accuracy: 72.41%	BERT model	8	The BERT model is ready, and we have successfully fixed all the bugs.
5/5/2021	Exact Match & F1	Accuracy: 75.54%	BERT model	10	We changed and tried different numbers of epochs and batch sizes for experimentation and finally decided the optimal one in the BERT model.

**The problems we are facing are:**

- RoBERTa model is new and there is hardly any detailed information available on the internet, so we require more time to build the code

**We plan to address them as follows:**

- We are constantly contacting people who know about the model or take references from internet to work on the same in limited amount of time

**We need help with:**

- Understanding the tokenization in RoBERTa model and library requirements.

**Summary of individual contributions**

	<b>Hiten Changlani</b>	<b>Swapn Shah</b>	<b>Shimpli Kalse</b>	<b>Amulya Chepyala</b>
03/03/2021 - 03/10/2021	-Participated in initial discussion about the project.	-Participated in initial discussion about the project	-Participated in initial discussion about the project	-Participated in initial discussion about the project



	<ul style="list-style-type: none"> <li>-Created Account on the Allen AI website and downloaded the dataset</li> <li>-Dropped an email to the Allen AI team to help us with team registration problem</li> <li>-Read the research paper related to text analysis models</li> <li>-Worked on project report simultaneously</li> </ul>	<ul style="list-style-type: none"> <li>-Created Account on the Allen AI website, downloaded the dataset and viewed the nature of data</li> <li>-Analysed and worked on the dataset</li> <li>-Read the research paper related to text analysis models</li> <li>-Worked on project report simultaneously</li> </ul>	<ul style="list-style-type: none"> <li>-Created Account on the Allen AI website and downloaded the dataset</li> <li>-Read the research paper related to text analysis models</li> <li>-Worked on project report simultaneously</li> </ul>	<ul style="list-style-type: none"> <li>Created Account on the Allen AI website, downloaded the dataset and viewed the nature of data</li> <li>-Read the research paper related to text analysis models</li> <li>-Worked on project report simultaneously</li> </ul>
03/10/2021 - 03/17/2021	<ul style="list-style-type: none"> <li>-Researched about BERT, RoBERTa and RAG model and their usage and libraries</li> <li>-Worked on RoBERTa model code</li> <li>-Attended meeting with Dr. Z to clear queries with problems concerning code debugging and more information about RoBERTa model</li> <li>-Dropped an email to Seetha regarding help with RAG model</li> </ul>	<ul style="list-style-type: none"> <li>-Researched about BERT, RoBERTa and RAG model and their usage and libraries</li> <li>-Worked on BERT model code</li> <li>-Attended meeting with Dr. Z to clear queries with problems concerning code debugging and more information about RoBERTa model</li> </ul>	<ul style="list-style-type: none"> <li>-Researched about BERT, RoBERTa and RAG model and their usage and libraries</li> <li>-Attended meeting with Dr. Z to clear queries with problems concerning code debugging and more information about RoBERTa model</li> <li>-Worked on project report simultaneously</li> </ul>	<ul style="list-style-type: none"> <li>-Researched about BERT, RoBERTa and RAG model and their usage and libraries</li> <li>-Attended meeting with Dr. Z to clear queries with problems concerning code debugging and more information about RoBERTa model</li> <li>-Worked on project report simultaneously</li> </ul>
03/17/2021 - 03/24/2021	<ul style="list-style-type: none"> <li>-Attended team meeting to plan further steps</li> <li>-Worked on RoBERTa model</li> <li>-Worked on project report simultaneously</li> </ul>	<ul style="list-style-type: none"> <li>-Attended team meeting to plan further steps</li> <li>-Worked on naive bayes, tf, tf-idf concepts to build working code</li> <li>-Worked on training model using code</li> </ul>	<ul style="list-style-type: none"> <li>-Attended team meeting to plan further steps</li> <li>-Worked on BERT code</li> <li>-Worked on project report simultaneously</li> </ul>	<ul style="list-style-type: none"> <li>-Attended team meeting to plan further steps</li> <li>-Worked on BERT code</li> <li>-Worked on project report simultaneously</li> </ul>
03/24/2021 - 04/05/2021	<ul style="list-style-type: none"> <li>-Attended team meeting to plan further steps</li> <li>-Worked on RoBERTa model</li> </ul>	<ul style="list-style-type: none"> <li>-Attended team meeting to plan further steps</li> </ul>	<ul style="list-style-type: none"> <li>-Attended team meeting to plan further steps</li> </ul>	<ul style="list-style-type: none"> <li>-Attended team meeting to plan further steps</li> <li>-worked on experiments using existing code</li> </ul>

	<ul style="list-style-type: none"> <li>-worked on experiments using existing code</li> <li>-Worked on training model using code</li> </ul>	<ul style="list-style-type: none"> <li>-worked on adding new features to existing code</li> <li>-Worked on training model using code</li> </ul>	<ul style="list-style-type: none"> <li>-worked on experiments using existing code</li> <li>-Worked on project report simultaneously</li> </ul>	<ul style="list-style-type: none"> <li>-Worked on project report simultaneously</li> </ul>
04/05/2021 - 04/11/2021	<ul style="list-style-type: none"> <li>-Attended team meeting to plan further steps</li> <li>-Worked on RoBERTa model</li> <li>-Worked on project report simultaneously</li> </ul>	<ul style="list-style-type: none"> <li>-Attended team meeting to plan further steps</li> <li>-Converted tf-df code into BERT model code</li> <li>-Worked on project report simultaneously</li> </ul>	<ul style="list-style-type: none"> <li>-Attended team meeting to plan further steps</li> <li>-Dropped email to TA regarding the problems we are facing and seeked help for the same</li> <li>-Worked on project report simultaneously</li> </ul>	<ul style="list-style-type: none"> <li>-Attended team meeting to plan further steps</li> <li>-Studied about the concepts of BERT model on internet to help with the code</li> <li>-Worked on project report simultaneously</li> </ul>
04/11/2021 - 04/25/2021	<ul style="list-style-type: none"> <li>-Attended team meeting to plan further steps</li> <li>-Worked on RoBERTa model. Collected information on RAG model as well</li> <li>-Worked on project report simultaneously</li> </ul>	<ul style="list-style-type: none"> <li>-Attended team meeting to plan further steps</li> <li>-Fixed BERT code and worked on accuracy</li> <li>-Worked on project report simultaneously</li> </ul>	<ul style="list-style-type: none"> <li>-Attended team meeting to plan further steps</li> <li>-Handled communications with Venkata regarding the problem faced and the solutions obtained</li> <li>-Worked on project report simultaneously</li> </ul>	<ul style="list-style-type: none"> <li>-Attended team meeting to plan further steps</li> <li>-Helped in increasing accuracy of the code</li> <li>-Worked on project report simultaneously</li> </ul>
04/25/2021 - 05/05/2021	<ul style="list-style-type: none"> <li>-Attended team meeting to plan further steps</li> <li>-Worked on results optimization</li> <li>-Worked on project report simultaneously</li> </ul>	<ul style="list-style-type: none"> <li>-Attended team meeting to plan further steps</li> <li>-Increased accuracy of the code</li> <li>-Worked on project report simultaneously</li> </ul>	<ul style="list-style-type: none"> <li>-Attended team meeting to plan further steps</li> <li>-Created powerpoint presentation</li> <li>-Worked on project report simultaneously</li> </ul>	<ul style="list-style-type: none"> <li>-Attended team meeting to plan further steps</li> <li>-Created powerpoint presentation</li> <li>-Worked on project report simultaneously</li> </ul>

## 8. REFERENCES

1. <https://arxiv.org/abs/1909.03065>
2. <https://github.com/CogComp/MCTACO>
3. <https://www.tensorflow.org/datasets/catalog/mctaco>

4. <https://monkeylearn.com/blog/practical-explanation-naive-bayes-classifier/>
5. <https://tomroth.com.au/keras/>
6. <https://stackabuse.com/text-classification-with-bert-tokenizer-and-tf-2-0-in-python/>

## APPENDIX

The team of MC-TACO allows anyone skilled in the art to reproduce our results for technological advancement. All the team members agree to this clause.

Team members:

- Hiten Changlani
- Swapn Shah
- Shimpli Kalse
- Amulya Chepyala

*Our data is available at:* Data is included in the zip file submitted with this report.

*Our code is available at:*

TF-IDF model: <https://colab.research.google.com/drive/1iXHV1vdqzg5Kk-cK2TxP7Nn2mzvqzRq?usp=sharing>

BERT model: [https://colab.research.google.com/drive/13\\_CLxt-Z9sqTwfLCq0aZsqNXLfUbwrs?usp=sharing](https://colab.research.google.com/drive/13_CLxt-Z9sqTwfLCq0aZsqNXLfUbwrs?usp=sharing)