COMP - 5014
Natural Language Processing

Commonsense Validation and Explanation with Natural Language Processing

Group Info :

Utsav Jivani

Baili Yin

Parth Dhameliya

Garima Chhaparwal

Jaydeep Joshi

Problem Statement

- Distinguish the statements of making sense or not from the given two sentences and Give reason.

Example

Statement 1	John put a turkey into a fridge	Correct
Statement 2	John put an elephant into the fridge	Incorrect
	Elephant can not fit into the fridge	Reason

Motivation

- Computers are becoming more intelligent.
- Computers can understand a language and generate another sentence.
- One of the applications of NLU is common sense validation.
- NLG helps generate reasons commonly
- System must have some common sense which can differentiate between sentences making any sense or not

Challenges

- Pre-processing task
- Training the model
- Combine the model
- Sentence generation

Data Set Used

- Train set: 10,000 instances
- Test set: 1000 instances

Resources Used

- Nvidia Tesla K80 GPU from Google Colab platform
- The pre-trained Transformer-based language models
- Jupyter notebook

Task A: Validation (Sentence Classification)

- Divide each row into two sentences and assign labels.

ID	Sentence	Labels
0	John put a turkey into a fridge	1 (Correct)
0	John put an elephant into a fridge	0 (Incorrect)

Task A: Validation (Sentence Classification)

Models

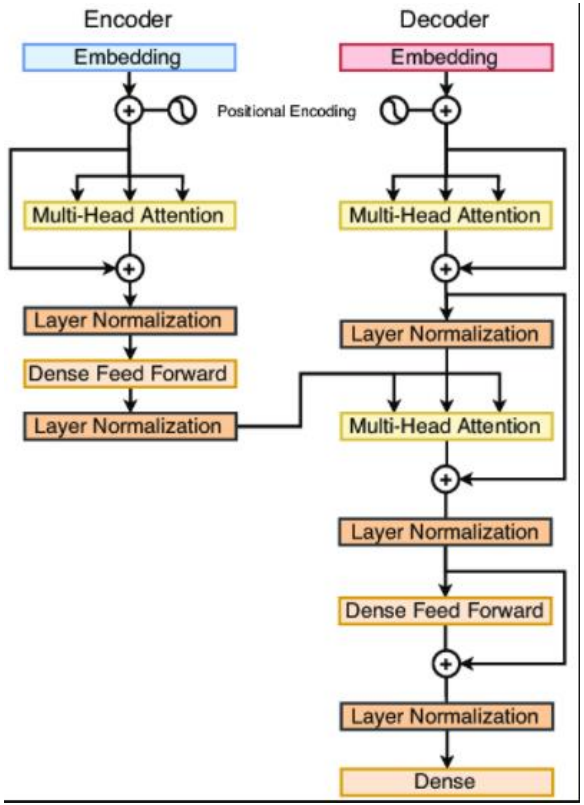
	BERT	DistilBert	GPT
Learning Rate	2e-5	5e-5	• Depends on perplexity score • Sense $\propto \frac{1}{Perplexity}$
Batch Size	32	8	
Epochs	8	4	

Result

	BERT	GPT	DistilBert	Overall
Accuracy	83.99	75	85.6	86.5

The evaluation metric for Subtasks A is accuracy

Transformers



Task B: Explanation (Multiple Choice)

- Choose the reason from the given 3 sentences.

Example

Statement: John put an elephant into the fridge.

Reason 1	An elephant is much bigger than fridge.	**Correct**
Reason 2	Elephants are usually white while fridges are usually white.	**Incorrect**
Reason 3	An elephant can not eat a fridge.	**Incorrect**

Task B: Explanation (Multiple Choice)

Models

	BERT	DistilBert	XLNET	Roberta	Albert
Learning Rate	2e-5	2e-5	2e-5	2e-5	2e-5
Batch Size	32	32	4	32	32
Epochs	4	4	4	4	8

Task B: Explanation (Multiple Choice)

Result

- Baseline Accuracy vs Single model accuracy (Experimental Results)

	Baseline Accuracy	Experimental Results
BERT	83.1	71.06
XLNET	81.6	71.96
DistlBert	79.3	68.17
Roberta	87.4	77.76
ALBert	82.6	72.76

Task B: Explanation (Multiple Choice)

Result

- Multi model accuracy

Merged Models	Accuracy
Xlnet + Albert + Bert + Distilbert + Roberta	89.90
Albert + Bert + Distilbert + Roberta	88.10
Bert + Distilbert + Roberta	85.30
Albert + Distilbert + Roberta	86.70
Albert + Bert + Roberta	87.10

* The evaluation metric for Subtasks B is accuracy *

Task C: Explanation (Text Generation)

- This subtask is similar to the second one since the input to both is a single natural language statement that contradicts common sense and the goal is to determine why is that.
- However, in this subtask, the system is expected to generate the justification from scratch.
- The generated text is evaluated against three correct reference justifications.

Task C: Explanation (Text Generation)

Example

False Statement : A bear was driving a car in Canada.

Referential Reasons:

A	Bear don't have car license and are unable to drive.
B	bears don't drive cars.
C	They do not drive cars.

Task C: Explanation (Text Generation)

Accomplishment

- We generated reasons consists of multiple sentences. The model's input is a concatenation of the statement and each explanation. The final answers are evaluated by BLUE.

Generated Reason Example

- A bear is not a vehicle.
- It is a wild animal.
- It cannot drive a car.
- It cannot be a person.

Task C: Explanation (Text Generation)

We utilised the pre-trained GPT-2, Openai-GPT and XLNET model for subtask C, 3 set of correct reasons for each false statement as input for fine-tuned training. Then load the trained models to generate reasons.

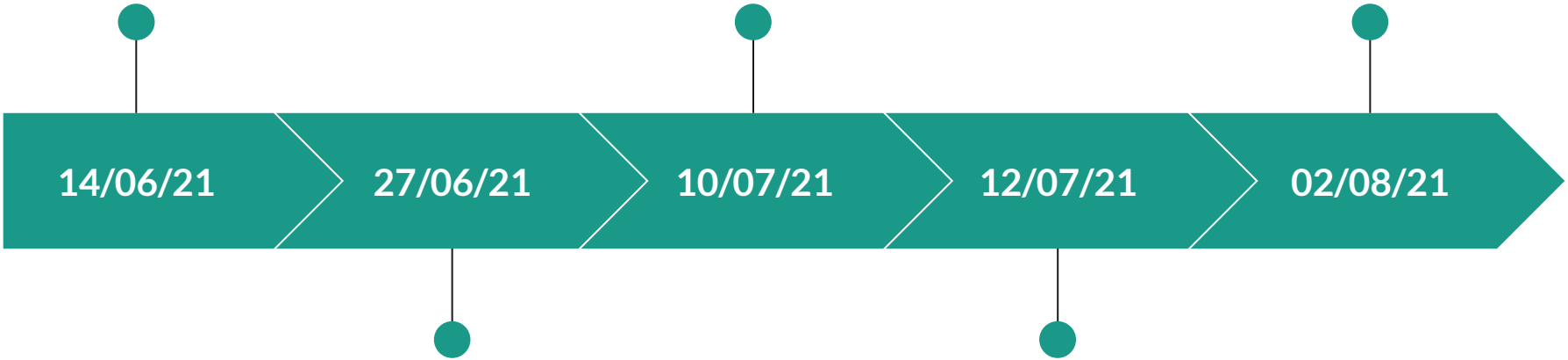
Model	Test BLEU Score
GPT-2	0.9808
OPENAI-GPT (GPT-1)	0.8679
XLNET	1.2267

The evaluation involves the BLEU score.

Project Proposal

Task B: Explanation (Multiple Choice) Implemented.

Expected Final Project and Report Submission



Task A: Validation (Sentence Classification) Implemented.

Project Progress Report

References :

- “SemEval-2020 Task 4: Common-sense Validation and Explanation”, Wang, Cunxiang and Liang, Shuai long and Jin, Yili and Wang, Yilong and Zhu, Xiaodan and Zhang, Yue ,”Proceedings of The 14th International Workshop on Semantic Evaluation”, ”Association for Computational Linguistics”, ”2020”
- “Common sense validation and reasoning using Natural Language Processing”, Shrenik Doshi, Praveen Joshi, Haithem Afli, ”2020”.
- Url : https://huggingface.co/ [online]
- Url : https://swatimeena989.medium.com/bert-text-classification-using-keras-903671e0207d [online]

Team work distribution :

Utsav Jivani (1138453) :
Worked on Subtask B using Roberta and XLNET language model

Beili Yin (1148875) :
Worked on Preprocessing task in Subtask B and used GPT2 to performSubtask C and Calculated BLEU score

Parth Dhameliya (1132450) :
Worked on Subtask B using Bert,Distilbert language models

Garima Chhaparwal (1131259) :
Worked on Subtask A using Bert, DistilBert language models

Jaydeep Joshi (1160015) :
Worked on Subtask A using GPT and did research related to Related works.
Worked on the power point presentation

THANK YOU