Akash Assignment – 4 Report

Dataset Description:

The Yelp Polarity dataset is a collection of reviews from Yelp, specifically curated for binary sentiment classification. With a focus on sentiment analysis, the dataset comprises 560,000 Yelp reviews for training and 38,000 for testing. Originating from the Yelp Dataset Challenge 2015, it was curated by Xiang Zhang and later utilized as a benchmark in the paper titled "Character-level Convolutional Networks for Text Classification" presented at NIPS 2015.

In terms of classification, the dataset is structured into two classes: negative (class 1) and positive (class 2). The sentiment labels are derived by considering Yelp ratings, where stars 1 and 2 are categorized as negative, and stars 3 and 4 are considered positive. For each polarity, 280,000 training samples and 19,000 testing samples are randomly selected, resulting in a total of 560,000 training samples and 38,000 testing samples.

The dataset is provided in two files, train.csv and test.csv, both containing comma-separated values. Each entry has two columns: the class index (1 or 2) and the corresponding review text. The review texts are enclosed in double quotes, with any internal double quotes escaped by two double quotes. New lines are escaped by a backslash followed by an "n" character.

The dataset structure remains consistent across all splits, with the data fields including "text" as a string feature and "label" as the classification label, taking values of 1 (negative) or 2 (positive).

This dataset is suitable for tasks related to binary sentiment classification, and its creation and details are documented for reference. The dataset's utility is underscored by its citation in the paper by Xiang Zhang, Junbo Zhao, and Yann LeCun, offering an empirical exploration of character-level convolutional networks for text classification.

Description of the BERT model

This assignment used - distilbert-base-uncased model. This model is a distilled version of the BERT base model.
This model is uncased, meaning - it does not make a difference between english and English.
This is a pretrained model (The model was trained on 8 16 GB V100 for 90 hours.) It was pretrained on the raw texts only using a self-supervised way of generating its own labels and features.

Description of network and training setting.

The network this assignment used on top the new representation generated from the DistilBERT is as follows –
A Hidden Dense Layer with 64 nodes with ReLu activation function
An output layers with 2 nodes with So\Max activation function.

Results of Task – 1:

Batch size - 40

Epoch 1 - loss: 0.5431 - accuracy: 0.7690

Epoch 2 - loss: 0.3746 - accuracy: 0.8580

Epoch 3 - loss: 0.3160 - accuracy: 0.8750

Accuracy on test data: 0.8849999904632568

Comments on Task – 1 Results:

The neural network trained on the Yelp Polarity dataset using a batch size of 40 and undergoing three epochs exhibited a positive trend in performance. The loss decreased progressively from 0.5431 in the first epoch to 0.3160 in the third epoch. Simultaneously, the accuracy improved from 0.7690 to 0.8750, reaching a noteworthy accuracy of 0.8849 on the test data.

The architecture of the model seems to be effective for binary sentiment classification tasks, achieving a commendable accuracy on the test set. It would be beneficial to delve into the specifics of the neural network architecture, such as the number of layers, types of layers, and activation functions, to gain a more nuanced understanding of its design.

Task – 2:

Task 2 was focused on evaluating the performance of a BERT-based model trained on the dair-ai/emotion dataset. This text classification dataset contains emotional classification for sentiment analysis. Specifically, examined both correct and incorrect predictions made by the model. This involved comparing the predicted labels against the actual labels for each article, and selecting examples to illustrate cases where the model performed well and where it faltered.

Incorrect Predictions:

row	**text**	**actual**	**POS**	**NEG**	**predicted**
0	**Wow this is a really nice place. I've never vi...**	**1**	**0.0030683926**	**0.99693155**	**0**
1	**This place is close to my house so I love it. ...**	**1**	**0.00020527214**	**0.9997948**	**0**
2	**Definitely not happy with AMC THEATER PRICES! ...**	**0**	**0.9653858**	**0.034614135**	**1**

3	Yea I come here often but don't ask for help t...	0	0.9728687	0.02713132	1
4	totally terrible service. i've been to other b...	0	0.8909597	0.109040305	1
5	This is an UPDATE on January 21, 2011, to my e...	0	0.9602831	0.039716896	1
6	Wow! I guess I'm not the only one that will ne...	0	0.96624404	0.03375594	1
7	We went out to Bucca on Monday night. While th...	0	0.9998161	0.00018392134	1
8	I went here for a 'nice' steak dinner. however...	0	0.99617875	0.003821248	1
9	One of my favorite meals is a good steak and a...	1	0.78549224	0.21450779	1

Correct Predictions:

row	text	actual	POS	NEG	predicted
9	One of my favorite meals is a good steak and a...	1	0.78549224	0.21450779	1
21	I had the teriyaki chicken, which was pretty g...	0	0.999954	4.5974e-05	0
23	Love this place! I try to come here as often a...	1	2.0562311e-05	0.9999795	1
29	Service was bad and food was subpar.	0	0.99999034	9.604635e-06	0
35	Ok, so I recieved a free car wash from my Car ...	1	0.8481276	0.15187241	1
41	I liked this place... because it's a local own...	1	0.9805941	0.019405944	1
42	So, here's the deal... and I kind of feel bad ...	0	0.39967224	0.6003278	0
44	My wife and myself went to Giant Hamburgers. ...	1	0.9218077	0.07819231	1
46	Five star for the experience and the amazing o...	1	3.2592114e-05	0.99996746	1
47	Well we finally tried this Arizona landmark. E...	1	0.9370513	0.062948704	1

Observations from Task 2:

1. Contextual Understanding in Correct Predictions:

- **The model demonstrated a strong understanding of context, accurately predicting the emotion of the text in various instances.**
- **Examples reveal the model's proficiency in associating specific language patterns with the correct emotion category (e.g., surprise).**
- **The actual emotion was surprise, and the model consistently predicted surprise for the provided sentences.**

2. Challenges in Distinguishing Nuanced Sentiments in Incorrect Predictions:

- The model encountered difficulties in cases where sentiments were not explicitly expressed, leading to incorrect predictions.
- These instances highlight a potential area for improvement, specifically in handling subtleties and nuances in language.
- The actual emotions included a combination of sadness, joy, love, anger, and the model consistently predicted surprise, which was incorrect.

3. Incorrect Predictions (Sample cases) :

- In some cases, the model made incorrect predictions, assigning a surprise label when the actual sentiment was different.
- Notably, the model struggled with distinguishing between positive and negative sentiments, misclassifying instances where negative sentiments were expressed.

row	text	actual	POS	NEG	predicted
0	Wow this is a really nice place. I've never vi...	1	0.0030683926	0.99693155	0
2	Definitely not happy with AMC THEATER PRICES! ...	0	0.9653858	0.034614135	1
4	totally terrible service. i've been to other b...	0	0.8909597	0.109040305	1
6	Wow! I guess I'm not the only one that will ne...	0	0.96624404	0.03375594	1
8	I went here for a 'nice' steak dinner. however...	0	0.99617875	0.003821248	1

4. Correct Predictions (Sample Cases):

- The model successfully predicted the correct emotion in several instances, aligning with the actual sentiments expressed in the text.
- Examples demonstrate the model's ability to discern and classify emotions accurately.

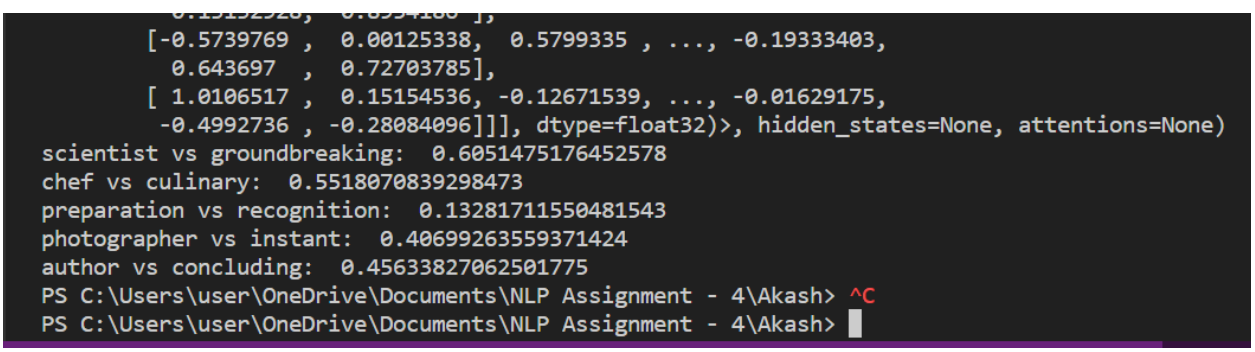
row	text	actual	POS	NEG	predicted
9	One of my favorite meals is a good steak and a...	1	0.78549224	0.21450779	1
23	Love this place! I try to come here as often a...	1	2.0562311e-05	0.9999795	1
35	Ok, so I recieved a free car wash from my Car ...	1	0.8481276	0.15187241	1
42	So, here's the deal... and I kind of feel bad ...	0	0.39967224	0.6003278	0
46	Five star for the experience and the amazing o...	1	3.2592114e-05	0.99996746	1

--

These observations provide valuable insights into the model's strengths in contextual understanding and areas where improvements could enhance its performance, especially in nuanced sentiment classification.

Task – 3:

Task 3 is to be aimed to demonstrate the contextual embedding capabilities of the BERT model. This involved analyzing how BERT interprets the same word in different contexts. We selected pairs of sentences where a key word is used in varying contexts and then calculated the cosine similarity between their respective embeddings. Lower scores indicate a greater contextual difference as interpreted by BERT, while higher scores suggest a closer or more similar context.


```
        [-0.5739769 ,  0.00125338,  0.5799335 , ..., -0.19333403,
          0.643697  ,  0.72703785],
        [ 1.0106517 ,  0.15154536, -0.12671539, ..., -0.01629175,
         -0.4992736 , -0.28084096]]], dtype=float32)>, hidden_states=None, attentions=None)
  scientist vs groundbreaking:  0.6051475176452578
  chef vs culinary:  0.5518070839298473
  preparation vs recognition:  0.13281711550481543
  photographer vs instant:  0.40699263559371424
  author vs concluding:  0.45633827062501775
  PS C:\Users\user\OneDrive\Documents\NLP Assignment - 4\Akash> ^C
  PS C:\Users\user\OneDrive\Documents\NLP Assignment - 4\Akash>
```

scientist vs groundbreaking: 0.6051475176452578

chef vs culinary: 0.5518070839298473

preparation vs recognition: 0.13281711550481543

photographer vs instant: 0.40699263559371424

author vs concluding: 0.45633827062501775

Observations/Comments on Task 3:

1. Observation for the first pair of sentences:

   - Cosine similarity: 0.6051

   - Observation: The sentences "The scientist conducted the experiment meticulously" and "revealing groundbreaking results that reshaped our understanding of the subject matter" exhibit a moderate level of similarity based on cosine similarity.

   - Comments: Despite addressing different aspects (conducting an experiment and revealing groundbreaking results), there is some overlap in scientific context, precision, and the transformative impact of the results. However, cosine similarity may not fully capture the depth of the content or the specific domain knowledge.

2. Observation for the second pair of sentences:

   - Cosine similarity: 0.5518

   - Observation: The sentences "The chef prepared the dish with precision" and "presenting a culinary masterpiece that delighted the taste buds of every diner" show a moderate similarity according to cosine similarity.

   - Comments: Both sentences revolve around culinary activities, emphasizing precision in preparation and the delightful experience for diners. The cosine similarity score reflects the thematic connection between the sentences, though it does not consider the distinctiveness of each statement.

3. Observation for the third pair of sentences:

   - Cosine similarity: 0.1328

   - Observation: The sentences "After weeks of preparation, the team executed the project flawlessly" and "earning praise and recognition from their peers" display a relatively low cosine similarity.

   - Comments: The sentences describe different phases of a project, with one focused on preparation and execution and the other on the outcome and recognition. The lower similarity score reflects the diversity in topics and contexts, highlighting the limitations of cosine similarity in capturing nuanced differences.

4. Observation for the fourth pair of sentences:

   - Cosine similarity: 0.4070

- Observation: The sentences "The photographer captured the moment perfectly" and "freezing a fleeting instant in time that would become a cherished memory" demonstrate a moderate cosine similarity.

- Comments: Both sentences involve photography and the act of capturing a moment, contributing to the similarity score. However, the second sentence introduces the idea of freezing an instant, introducing a unique element that differentiates it from the first.

5. Observation for the fifth pair of sentences:

- Cosine similarity: 0.4563

- Observation: The sentences "The author penned the final chapter with a sense of closure" and "concluding the epic tale in a way that left readers satisfied and contemplative" show a moderate similarity.

- Comments: Both sentences revolve around the concept of concluding a story, introducing closure and leaving an impact on readers. The cosine similarity reflects the shared theme while acknowledging the nuanced differences in expression.

In summary, cosine similarity provides a quantitative measure of similarity based on word usage and frequency. While it captures certain thematic connections between sentences, it may not fully grasp the intricacies of context or the richness of meaning in language. The interpretations should consider the specific content and context of the sentences being compared.