Name: Sushruth Danivasa Sridhar

Assignment – 2

Among all the comparisons, such as different RNN layers, different number of nodes, different number of layers, bidirectional or not, pre-trained embeddings or not, and different training set sizes, the comparison that I have choose for task 1 to perform is "Comparison between Simple RNN and Long Short-Term Memory (LSTM)".

1. **Description of your two meaningful comparisons from Task 1:**

   Simple RNN:

   For our first comparison, we employed a model utilizing Simple RNN layers for sequence processing. Simple RNNs are characterized by their straightforward architecture, allowing for ease of training and interpretation. However, they suffer from the vanishing gradient problem, especially in tasks involving long sequences. This problem arises when the gradients diminish exponentially as they are back-propagated through time, making it challenging for the network to capture dependencies over long distances. In the context of named entity recognition (NER), where understanding the entire context of a sentence is crucial, Simple RNNs might face limitations in effectively capturing relationships between distant words.

   LSTM:

   In contrast, we assessed a model with Long Short-Term Memory (LSTM) layers. LSTMs are specifically designed to address the vanishing gradient problem by introducing memory cells that can retain information over extended sequences. This architecture is well-suited for tasks requiring the understanding of context and dependencies across various parts of a sentence. In NLP tasks like NER, where recognizing entities often depends on the context of the entire sentence, LSTMs are expected to outperform Simple RNNs. The ability of LSTMs to capture long-term dependencies makes them a popular choice in natural language processing applications.

2. **The results in terms of precision/recall/F-measure for all entities and overall, in the form of tables or graphs.**

   - Entity type: geo

	Simple RNN	LSTM
Total Entities	6194	6194
Total predicted	6515	6541
Correctly extracted	5314	5313
Precision	81.57 %	81.23 %

Recall	85.79 %	85.78 %
F-measure	83.63 %	83.44 %

- Entity type: gpe

	Simple RNN	LSTM
Total Entities	2757	2757
Total predicted	2694	2703
Correctly extracted	2555	2561
Precision	94.84 %	94.75 %
Recall	92.67 %	92.89 %
F-measure	93.74 %	93.81 %

- Entity type: per

	Simple RNN	LSTM
Total Entities	2784	2784
Total predicted	2682	2520
Correctly extracted	1913	1894
Precision	71.33 %	75.16 %
Recall	68.71 %	68.03 %
F-measure	70.0 %	71.42 %

- Entity type: org

	Simple RNN	LSTM
Total Entities	3400	3400
Total predicted	2831	3034
Correctly extracted	1971	2100
Precision	69.62 %	69.22 %
Recall	57.97 %	61.76 %
F-measure	63.26 %	65.28 %

- Entity type: tim

	Simple RNN	LSTM
Total Entities	3431	3431
Total predicted	3114	3139
Correctly extracted	2637	2655
Precision	84.68 %	84.58 %
Recall	76.86 %	77.38 %
F-measure	80.58 %	80.82 %

- Entity type: art

	Simple RNN	LSTM
Total Entities	75	75
Total predicted	0	0
Correctly extracted	0	0
Precision	cannot be computed	cannot be computed

Recall	0.0 %	0.0 %
F-measure	cannot be computed	cannot be computed

- Entity type: nat

	Simple RNN	LSTM
Total Entities	36	36
Total predicted	0	0
Correctly extracted	0	0
Precision	cannot be computed	cannot be computed
Recall	0.0 %	0.0 %
F-measure	cannot be computed	cannot be computed

- Entity type: eve

	Simple RNN	LSTM
Total Entities	41	41
Total predicted	0	0
Correctly extracted	0	0
Precision	cannot be computed	cannot be computed
Recall	0.0 %	0.0 %
F-measure	cannot be computed	cannot be computed

- All Entities combined:

	Simple RNN	LSTM
Total Entities	18718	18718
Total predicted	17836	17937
Correctly extracted	14390	14523
Precision	80.68 %	80.97 %
Recall	76.88 %	77.59 %
F-measure	78.73 %	79.24 %

3. Your comments on the results of Task 1.

LSTM is the better model for NER, with an F-measure of 79.24% compared to the Simple RNN's F-measure of 78.73%. LSTM models are better at learning long-term dependencies between words in a sentence, which is important for NER tasks.

In addition to the F-measure, we can also look at the precision, recall, and total entities for each model. The LSTM model has a slightly higher precision and recall than the Simple RNN model, and it correctly extracted more entities.

LSTM model is better than Simple RNN model on all metrics:

Precision: LSTM (80.97%) > Simple RNN (80.68%)
Recall: LSTM (77.59%) > Simple RNN (76.88%)
Total entities correctly extracted: LSTM (14523) > Simple RNN (14390)

4. Error analysis from Task 2

Sentence And Target	Simple RNN	LSTM
Test example: 3499 Sentence: ['If', 'the', 'National', 'Air', 'and', 'Space', 'Administration', 'is', 'not', 'able', 'to', 'launch', 'Atlantis', 'next', 'week', ',', 'the', 'mission', 'to', 'the', 'International', 'Space', 'Station', 'likely', 'will', 'be', 'postponed', 'until', 'October', '.'] Target: ['O', 'O', 'B-org', 'I-org', 'I-org', 'I-org', 'I-org', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'B-geo', 'I-geo', 'I-geo', 'O', 'O', 'O', 'O', 'O', 'B-tim', 'O']	Predicted: ['O', 'O', 'B-org', 'I-org', 'I-org', 'I-org', 'I-org', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'B-org', 'I-geo', 'I-geo', 'O', 'O', 'O', 'O', 'O', 'B-tim', 'O'] At 2 ('org', 'National Air and Space Administration') Extracted. At 20 ('geo', 'International Space Station') Missed. At 28 ('tim', 'October') Extracted. At 20 ('org', 'International') Incorrectly extracted.	Predicted: ['O', 'O', 'B-org', 'I-org', 'I-org', 'I-org', 'I-org', 'O', 'O', 'O', 'O', 'O', 'B-tim', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'B-org', 'I-geo', 'I-geo', 'O', 'O', 'O', 'O', 'O', 'B-tim', 'O'] At 2 ('org', 'National Air and Space Administration') Extracted. At 20 ('geo', 'International Space Station') Missed. At 28 ('tim', 'October') Extracted. At 12 ('tim', 'Atlantis') Incorrectly extracted. At 20 ('org', 'International') Incorrectly extracted.
Test example: 3504 Sentence: ['Interim', 'Prime', 'Minister', 'Ehud', 'Olmert', "'s", 'centrist', 'Kadima', 'Party', 'and', 'the', 'Dovish', 'Labor', 'party', 'have', 'signed', 'a', 'coalition', 'agreement', '.'] Target: ['O', 'B-per', 'I-per', 'I-per', 'I-per', 'O', 'O', 'B-per', 'I-per', 'O', 'O', 'B-org', 'I-org', 'O', 'O', 'O', 'O', 'O', 'O', 'O']	Predicted: ['O', 'B-per', 'O', 'B-per', 'I-per', 'O', 'O', 'B-geo', 'I-org', 'O', 'O', 'B-org', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O'] At 1 ('per', 'Prime Minister Ehud Olmert') Missed. At 7 ('per', 'Kadima Party') Missed. At 11 ('org', 'Dovish Labor') Missed. At 1 ('per', 'Prime') Incorrectly extracted. At 3 ('per', 'Ehud Olmert') Incorrectly extracted. At 7 ('geo', 'Kadima') Incorrectly extracted. At 11 ('org', 'Dovish') Incorrectly extracted.	Predicted: ['O', 'B-per', 'I-per', 'B-per', 'I-per', 'O', 'O', 'B-geo', 'I-org', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O'] At 1 ('per', 'Prime Minister Ehud Olmert') Missed. At 7 ('per', 'Kadima Party') Missed. At 11 ('org', 'Dovish Labor') Missed. At 1 ('per', 'Prime Minister') Incorrectly extracted. At 3 ('per', 'Ehud Olmert') Incorrectly extracted. At 7 ('geo', 'Kadima') Incorrectly extracted.

Test example: 3509 Sentence: ['But', 'he', 'believes', 'that', 'is', 'impossible', 'because', 'the', 'Islamic', 'militant', 'group', 'Hamas', ',', 'which', 'seeks', 'Israel', "'s", 'destruction', ',', 'now', 'heads', 'the', 'Palestinian', 'Authority', '.'] Target: ['O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'B-org', 'I-org', 'I-org', 'I-org', 'O', 'O', 'O', 'B-geo', 'O', 'O', 'O', 'O', 'O', 'O', 'B-gpe', 'O', 'O']	Predicted: ['O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'B-org', 'O', 'O', 'O', 'B-geo', 'O', 'O', 'O', 'O', 'O', 'O', 'B-gpe', 'O', 'O'] At 8 ('org', 'Islamic militant group Hamas') Missed. At 15 ('geo', 'Israel') Extracted. At 22 ('gpe', 'Palestinian') Extracted. At 11 ('org', 'Hamas') Incorrectly extracted.	Predicted: ['O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'B-org', 'O', 'O', 'I-org', 'O', 'O', 'O', 'B-geo', 'O', 'O', 'O', 'O', 'O', 'O', 'B-gpe', 'O', 'O'] At 8 ('org', 'Islamic militant group Hamas') Missed. At 15 ('geo', 'Israel') Extracted. At 22 ('gpe', 'Palestinian') Extracted. At 8 ('org', 'Islamic Hamas') Incorrectly extracted.
Test example: 3516 Sentence: ['His', 'coalition', 'with', 'Labor', 'does', 'not', 'give', 'him', 'a', 'majority', 'in', 'the', '120-member', 'Knesset', 'or', 'parliament', '.'] Target: ['O', 'O', 'O', 'B-org', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O']	Predicted: ['O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O'] At 3 ('org', 'Labor') Missed.	Predicted: ['O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'B-org', 'O', 'O', 'O'] At 3 ('org', 'Labor') Missed. At 13 ('org', 'Knesset') Incorrectly extracted.
Test example: 3518 Sentence: ['Analysts', 'say', 'the', 'government', 'could', 'collapse', 'in', 'two', 'or', 'three', 'years', ',', 'when', 'the', 'time', 'comes', 'to', 'remove', 'tens', 'of', 'thousands', 'settlers', 'from', 'their', 'homes', '.'] Target: ['O', 'O', 'O', 'O', 'O', 'O', 'O', 'B-tim', 'I-tim', 'I-tim', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O']	Predicted: ['O', 'O', 'O', 'O', 'O', 'O', 'O', 'B-tim', 'I-tim', 'I-tim', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O'] At 7 ('tim', 'two or three') Extracted.	Predicted: ['O', 'O', 'O', 'O', 'O', 'O', 'O', 'B-tim', 'O', 'B-tim', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O'] At 7 ('tim', 'two or three') Missed. At 7 ('tim', 'two') Incorrectly extracted. At 9 ('tim', 'three') Incorrectly extracted.

Test example: 5344 Sentence: ['Two', 'British', 'attempts', 'at', 'establishing', 'the', 'island', 'as', 'a', 'penal', 'colony', '(', '1788', '-', '1814', 'and', '1825', '-', '55', ')', 'were', 'ultimately', 'abandoned', '.'] Target: ['O', 'B-gpe', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'B-tim', 'I-tim', 'I-tim', 'O', 'B-tim', 'I-tim', 'I-tim', 'O', 'O', 'O', 'O', 'O']	Predicted: ['O', 'B-gpe', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'I-tim', 'B-tim', 'O', 'B-tim', 'O', 'I-tim', 'O', 'O', 'O', 'O', 'O'] At 1 ('gpe', 'British') Extracted. At 12 ('tim', '1788 - 1814') Missed. At 16 ('tim', '1825 - 55') Missed. At 14 ('tim', '1814') Incorrectly extracted. At 16 ('tim', '1825 55') Incorrectly extracted.	Predicted: ['O', 'B-gpe', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'B-tim', 'B-tim', 'B-tim', 'O', 'B-tim', 'I-tim', 'O', 'O', 'O', 'O', 'O', 'O'] At 1 ('gpe', 'British') Extracted. At 12 ('tim', '1788 - 1814') Missed. At 16 ('tim', '1825 - 55') Missed. At 12 ('tim', '1788') Incorrectly extracted. At 13 ('tim', '-') Incorrectly extracted. At 14 ('tim', '1814') Incorrectly extracted. At 16 ('tim', '1825 -') Incorrectly extracted.

"In the comparison between Simple RNN and LSTM models, there were instances where the LSTM outperformed the Simple RNN in correctly extracting certain word pairs that were previously missing or incorrectly extracted by the Simple RNN. However, it's worth noting that there were situations where the LSTM missed extractions or incorrectly identified entities that the Simple RNN handled better.

These performance disparities could potentially be addressed by modifying the model architecture. Fine-tuning the number of nodes and adding more dense layers during the creation of both the Simple RNN and LSTM models, for example, may improve their ability to correctly extract complex relationships and overall entity recognition.

5. Results of Task 3

- Entity type: geo

	LSTM
Total Entities	7
Total predicted	7
Correctly extracted	7
Precision	100.0 %
Recall	100.0 %
F-measure	100.0 %

- Entity type: gpe

	LSTM
Total Entities	3
Total predicted	2
Correctly extracted	2
Precision	100 %
Recall	66.67 %
F-measure	80.0 %

- Entity type: per

	LSTM
Total Entities	0
Total predicted	0
Correctly extracted	0
Precision	cannot be computed
Recall	cannot be computed
F-measure	cannot be computed

- Entity type: org

	LSTM
Total Entities	2
Total predicted	2
Correctly extracted	1
Precision	50.0 %
Recall	50.0 %
F-measure	50.0 %

- Entity type: tim

	LSTM
Total Entities	1
Total predicted	1
Correctly extracted	1
Precision	100.0 %
Recall	100.0 %
F-measure	100.0 %

- Entity type: art

	LSTM
Total Entities	0
Total predicted	0
Correctly extracted	0
Precision	cannot be computed
Recall	cannot be computed
F-measure	cannot be computed

- Entity type: nat

	LSTM
Total Entities	0
Total predicted	0
Correctly extracted	0
Precision	cannot be computed
Recall	cannot be computed
F-measure	cannot be computed

- Entity type: eve

	LSTM
Total Entities	0
Total predicted	0
Correctly extracted	0
Precision	cannot be computed
Recall	cannot be computed
F-measure	cannot be computed

- All entities combined:

	LSTM
Total Entities	13
Total predicted	12
Correctly extracted	11
Precision	91.67 %
Recall	84.62 %
F-measure	88.0 %

6. Your comments on the results of Task 3.

The results of Task 3 show that the LSTM model excels in certain entity types (e.g., geo, tim) but may require adjustments to improve performance in others (e.g., gpe, org, per). Further tuning or modifications to the model architecture could enhance overall performance.

To improve the model's performance on these entity types,

- Increase the size and diversity of the training set.
- Use a different model architecture like Simple RNN for small dataset or small sentences.