

Language Model Performance Assessment Report

Summary of the findings.

- It has been observed that even the instance numbers are uneven (Majority NP | Minority P) the model may have the capability to perform well instead of being biased towards one prediction type.
- The condition seems to be – is the assessment of the model with same number of instances of data sequence which it was trained on.
- Even if the data sequence is unseen to the model it can perform well if instance numbers are same as mentioned in the above point.
- When instance numbers do not match (extended or under sampled) with the training data the model was trained on – it provides a biased result.
- In multiple instances this biased result contains the singular data instance with single label for all test instances.
- The reason towards this behavior can be the attention mechanism used in the model as observed in the dummy data experiments. (mentioned in point 8 - 15)
- The experiments were conducted with both dummy and real OM data to observe and summarize the findings.
- Below the experiment details are documented in two sections.

Section 1: Experiment with Dummy Data

Dataset:

https://github.com/rashedhasan090/OM-ML_Research/blob/main/January%202023%20Analysis/New%20Analysis/sampleset/sample_space2/dummy_data.csv

Notebook:

https://github.com/rashedhasan090/OM-ML_Research/blob/main/January%202023%20Analysis/New%20Analysis/sampleset/sample_space2/series_2_Language_Model_Performance_assessment_ipynb.ipynb

1. This experiment was conducted with Dummy Data. The data looks like below.

class1,table1,obj1,atr1	class1,table1,obj1,atr1,NP
class1,table2,obj1,atr1	class1,table2,obj1,atr1,NP
class1,table3,obj1,atr1	class1,table3,obj1,atr1,NP
class1,table4,obj1,atr1	class1,table4,obj1,atr1,NP
class1,table5,obj1,atr1	class1,table5,obj1,atr1,NP
class1,table6,obj1,atr1	class1,table6,obj1,atr1,NP
class1,table7,obj1,atr1	class1,table7,obj1,atr1,NP
class1,table8,obj1,atr1	class1,table8,obj1,atr1,NP
class1,table9,obj1,atr1	class1,table9,obj1,atr1,NP
class1,table10,obj1,atr1	class1,table10,obj1,atr1,NP
class1,table11,obj1,atr1	class1,table11,obj1,atr1,NP
class1,table12,obj1,atr1	class1,table12,obj1,atr1,NP
class1,table13,obj1,atr1	class1,table13,obj1,atr1,NP
class1,table14,obj1,atr1	class1,table14,obj1,atr1,NP
class1,table15,obj1,atr1	class1,table15,obj1,atr1,NP

2. The dataset had total 122 instances, among these 15 labeled as P and 107 labeled as NP.

3. The model was trained and classification report states below result for the same set of data **which it was trained on**.

precision	recall	f1-score	support	
0	0.96	1.00	0.98	107
1	1.00	0.73	0.85	15

accuracy			0.97	122
macro avg	0.98	0.87	0.91	122
weighted avg	0.97	0.97	0.96	122

4. Then a set of Dummy Unseen data was created which looks like below. This data is different than the original data the model was trained on.

The dataset can be found here:

https://github.com/rashedhasan090/OM-ML_Research/blob/main/January%202023%20Analysis/New%20Analysis/sampleset/sample_space2/dummy_data_unseen.csv

```
table121,class221,obj31,atr1,obj32,atr2
table122,class222,obj31,atr1,obj32,atr2
table123,class223,obj31,atr1,obj32,atr2
table124,class224,obj31,atr1,obj32,atr2
table125,class225,obj31,atr1,obj32,atr2
table126,class226,obj31,atr1,obj32,atr2
table127,class227,obj31,atr1,obj32,atr2
table128,class228,obj31,atr1,obj32,atr2
table129,class229,obj31,atr1,obj32,atr2
```

5. This dummy dataset had 122 instances. Which is same as the training dataset.

6. A classification report was obtained for this and it displays similar result as the original test data.

	precision	recall	f1-score	support
0	0.96	1.00	0.98	107
1	1.00	0.73	0.85	15

accuracy			0.97	122
macro avg	0.98	0.87	0.91	122
weighted avg	0.97	0.97	0.96	122

7. To further understand why the result was identical – A dataset with dummy unseen data was extended to 155 instances.

Dataset URL:

https://github.com/rashedhasan090/OM-ML_Research/blob/main/January%202023%20Analysis/New%20Analysis/sampleset/sample_space2/dummy_data_unseen_extended.csv

8. However when the translation function was utilized to obtain prediction, the same **problem was observed** ! which was also seen on the real experiments with OM data.

9. The problem was not only one single label was printed for all dummy instances – but only one single data sequence was predicted with that label.

10. In the real OM experiments this problem was seen where only single instance of data was predicted for all unseen instances with one single label.

11. No variation was seen in this process. Such as below.

```
class4,table6,obj2,atr2,np  
class4,table6,obj2,atr2,np  
class4,table6,obj2,atr2,np  
class4,table6,obj2,atr2,np  
class4,table6,obj2,atr2,np  
class4,table6,obj2,atr2,np  
class4,table6,obj2,atr2,np
```

12. As also the classification report in comparison with real data could not be obtained as because of instance mismatch. (122 on the training dataset and 155 on Dummy Testing dataset). It displayed below message.

```
ValueError: Found input variables with inconsistent numbers of  
samples: [122, 155]
```

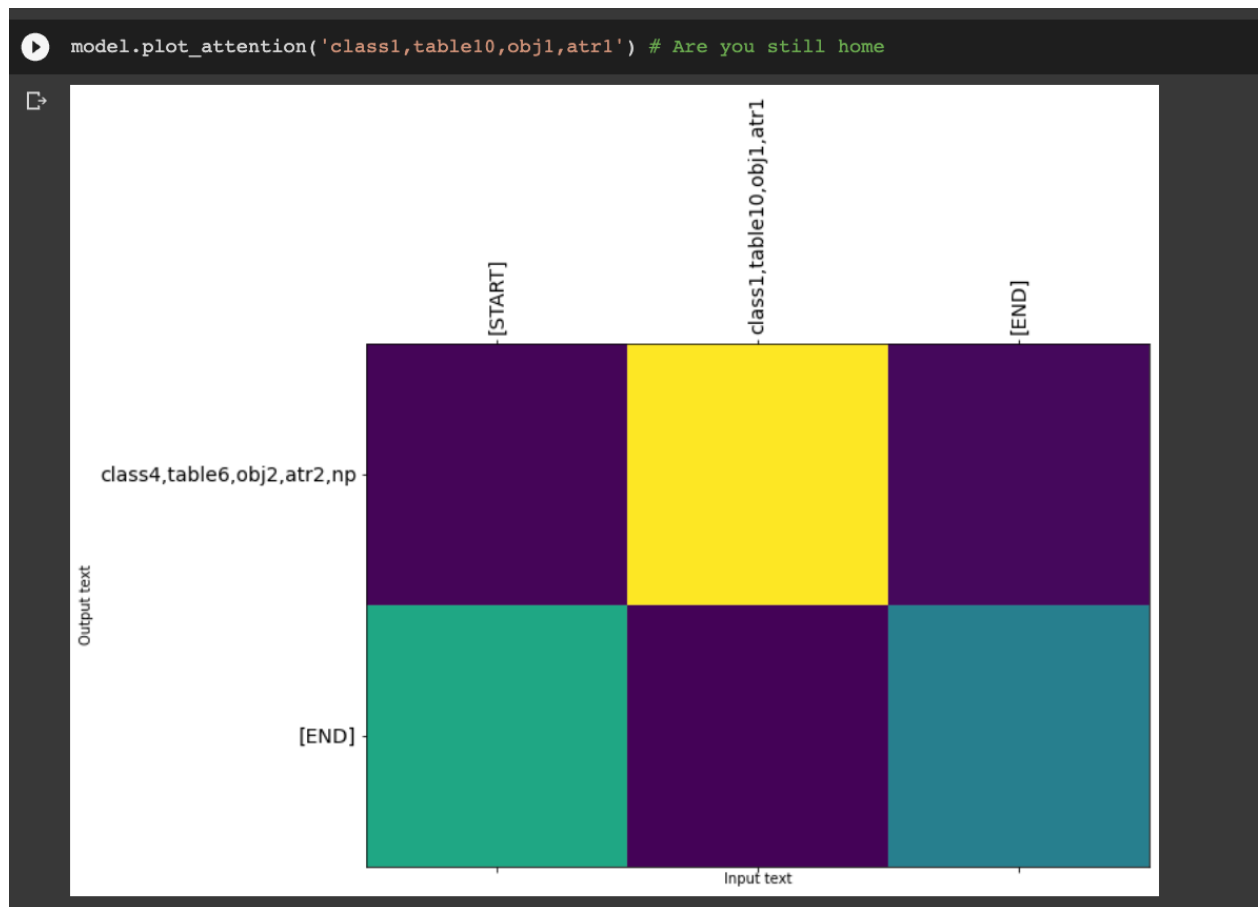
13. After this step the experiment was conducted with an under sampled dummy dataset which had 100 instances.

URL:

https://github.com/rashedhasan090/OM-ML_Research/blob/main/January%202023%20Analysis/New%20Analysis/sampleset/sample_space2/dummy_data_unseen_undersample.csv

14. The result was identical as described in step 8 – 11.

15. The data and the label which was repeatedly printed was seen in the attention plot for a given instance.



Section 2 : Experiment with original OM Data.

Dataset:

https://github.com/rashedhasan090/OM-ML_Research/blob/main/January%202023%20Analysis/New%20Analysis/samples/sample_space3/om-exp.csv

Notebook:

[https://github.com/rashedhasan090/OM-ML_Research/blob/main/January%202023%20Analysis/New%20Analysis/samples/sample_space3/series_3_Language_Model_Performance_assessment_\(Real_Data\).ipynb](https://github.com/rashedhasan090/OM-ML_Research/blob/main/January%202023%20Analysis/New%20Analysis/samples/sample_space3/series_3_Language_Model_Performance_assessment_(Real_Data).ipynb)

OM Data used :

Bank OM, Decider, Camping OM.

16. After the experiment with dummy data an experiment with real instances of OM data was conducted. Where 100 instances of data was integrated from 3 OM's.

(27 labeled as P | 73 labeled as NP)

17. The model was trained and classification report states below result for the same set of OM Data **which it was trained on.**

```
Confusion Matrix (Train Data):  
[[55 18]  
 [ 2 25]]
```

	precision	recall	f1-score	support
0	0.96	0.75	0.85	73
1	0.58	0.93	0.71	27
accuracy			0.80	100
macro avg	0.77	0.84	0.78	100
weighted avg	0.86	0.80	0.81	100

18. Afterwards I utilized unseen data to test the model. The unseen dataset can be found below.

https://github.com/rashedhasan090/OM-ML_Research/blob/main/January%202023%20Analysis/New%20Analysis/sampleset/sample_space3/om-exp-unseen.csv

19. These were also the original samples from Library Management OM.

20. Total 100 instances are present in this dataset where 15 marked as P and 85 marked as NP.

21. After this dataset was given out to the model and classification report was obtained , it was observed the **model actually performed relatively well on unseen data.**

Confusion Matrix (Test Data):

```
[[58 15]
 [ 1 26]]
```

	precision	recall	f1-score	support
0	0.98	0.79	0.88	73
1	0.63	0.96	0.76	27
accuracy			0.84	100
macro avg	0.81	0.88	0.82	100
weighted avg	0.89	0.84	0.85	100

22. As after this step the number of samples were extended in the unseen test dataset. These samples were obtained from Customer-Order OM. (9 extra samples to see the models behavior)

The total sample number was 108. The dataset can be found here.

https://github.com/rashedhasan090/OM-ML_Research/blob/main/January%202023%20Analysis/New%20Analysis/sampleset/sample_space3/om-exp-unseen-extended.csv

23. After the translation module was executed on these instances the same problem was encountered (one single label – same seq of data for all instances). **As mentioned in points 8-11.**

[illegible]

24. As also the classification report could also not be obtained due to instance mismatch with the original dataset issuing below error.

```
ValueError: Found input variables with inconsistent numbers of
samples: [100, 108]
```

25. This step was also conducted with a dataset which was under sampled having 71 instances. The result was same as above mentioned in step 23.

Dataset:

https://github.com/rashedhasan090/OM-ML_Research/blob/main/January%202023%20Analysis/New%20Analysis/sampleset/sample_space3/om-exp-unseen-undersample.csv