

Report VIII: Turkish Speech to Text

Model I

AAI's DS2 implementation is trained on the METUbet Turkish read speech dataset.

Satisfying performance.

Hyperparameter tuning started.



Figure 1: Model Performance on METUbet

evaluating...

Target: hiç bir kimyasal gübre ve zehirli böcek ya da mantar ilacı kullanılmadı.
Predicted: hiç bir kimirason gübre vezehirli böcei kiyada manterleçı kuağnı mazı.
Target: onun sanatsal başarıları sosyal bir birlikteliğe girmesini garantiledi.
Predicted: onu saonan san başarıların sosyan bir birlekeniye gilmisiniderentiledi.

Figure 2: Best Metrics on Validation Set

Model II

Cleaned the Mozilla Dataset:

- Silences in the head and tail of the recordings are cut,
- Click sounds of the user (Could slow training due to random noise characteristic)

Trained further the pretrained model on this 14h set.

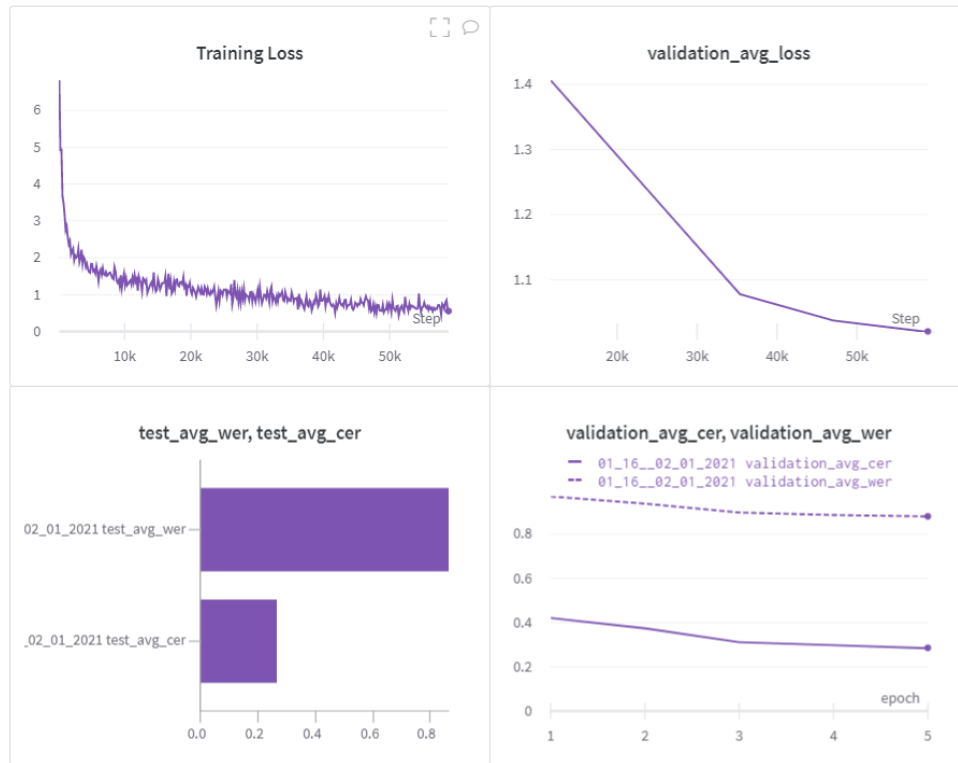


Figure 3: METUbet+ Mozilla* Trained Model Performance

```
evaluating...
-----
Target: bazılarına ilgi diğerlerine göre daha büyük.
Predicted: bazılarına ii diğeleme göre daha büyük.

Target: gözden geçirme sekiz ocak iki bin üçte başlatıldı.
Predicted: gözden geçirmen sekiz acak iki bin iş da başlatıldı.
Target: güreşçiler altmış kilo kategorisinde güreştiler.
Predicted: düreşiler htmışkiliokata bolisinde güreştiler.
-----
Test Set: Average loss: 0.9415, Average CER: 0.265451 Average WER: 0.8632
```

Figure 4: Evaluation on the Mozilla* Test Set

Tested the METUbet+Mozilla* trained model, back on the METUbet. Observed no significant performance change.

Questions:

Which metric to compare model performance with?

- Validation loss, average CER, average WER?
 - Validation loss is objective
 - No strong correlation between Loss and CER or WER
 - CER, WER depends on decoding algorithm
 - Currently: Argmax Decode
 - Possible: Beam Search decoding or Prefix Beam Search (Language Model integrated)

Possible Directions:

1. Training on merged dataset (Gathering more data not possible)
2. Language Model integration
 - Independent
 - Beam Search Integrated

To Do:

1. Hyperparameter Search until best model found for a task
 - Define a task
2. Test the METUbet trained model on METUbet+Mozilla* and compare with METUbet+Mozilla* trained model performance on METUbet+Mozilla. (Investigate the effect of poor speaking and noisy recordings)