KDD Cup ’22 Workshop

Predicting Query-Item Relationship using Adversarial Training and Robust Modeling Techniques

10th Place Solution in Product Substitute Classification task

Min Seok Kim (Marko)

Recommendation Team, Kakao Corp.

kakao

1. Problem
2. Solution
   a) Validation Strategy
   b) Model Structure
   c) Training Techniques
   d) Diversity-based Ensemble
3. Conclusion

Pretrained Transformers

LSTM Head with Different LR

Multi-Sampled Dropout

Adversarial Training

Exponential Moving Average

Learning Rate Scheduling

kakao

Problem

- Predicting relationship between a query and an item
- Query is never seen before during inference

Query

White Vans

Item Meta Data (ex. Title)

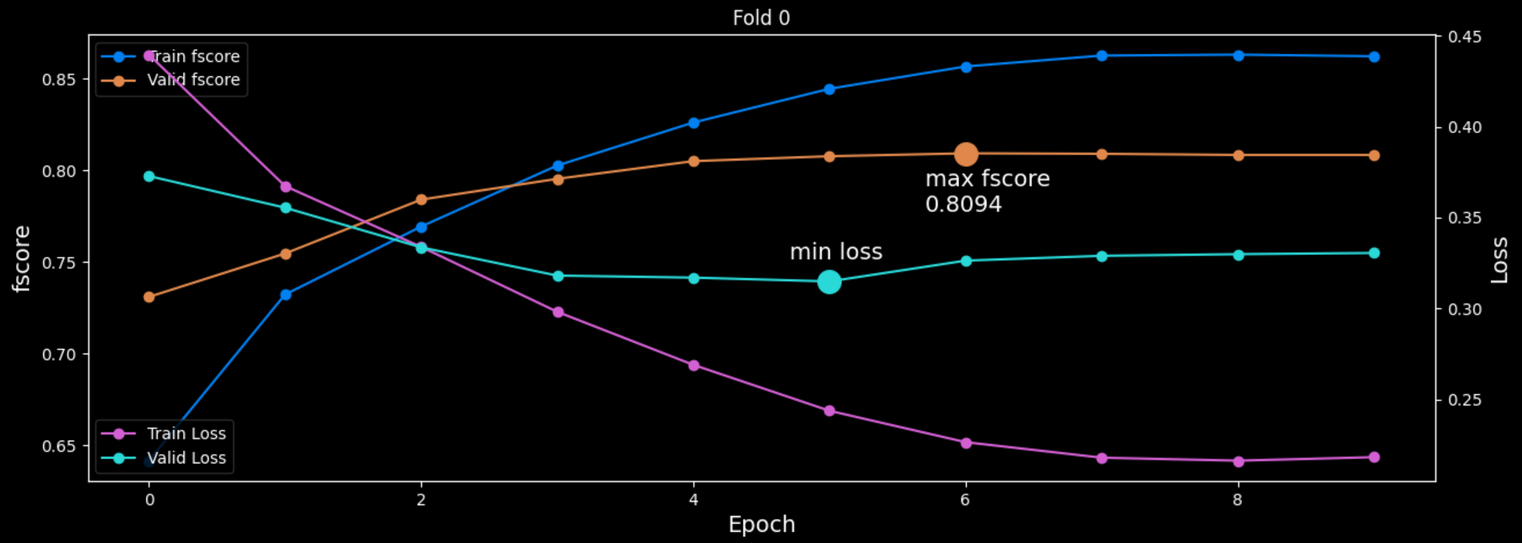
Vans Sneakers, Navy

Substitute?

kakao

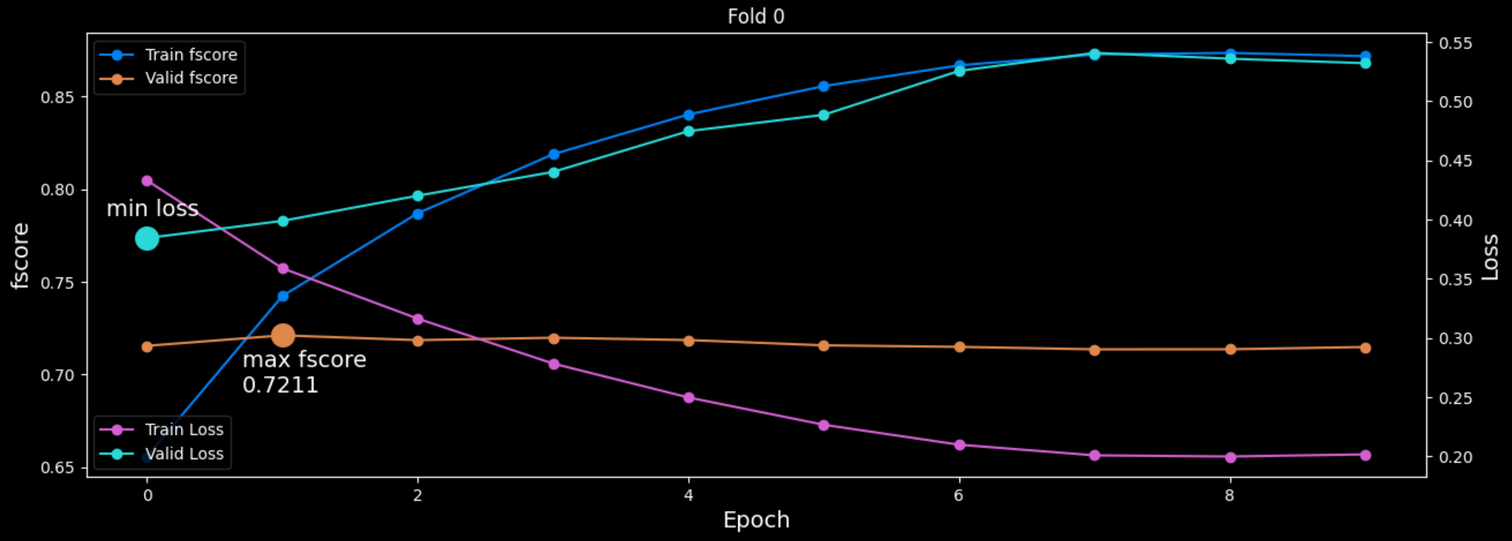
Validation Strategy

KFold



Public LB: 0.850

GroupKFold

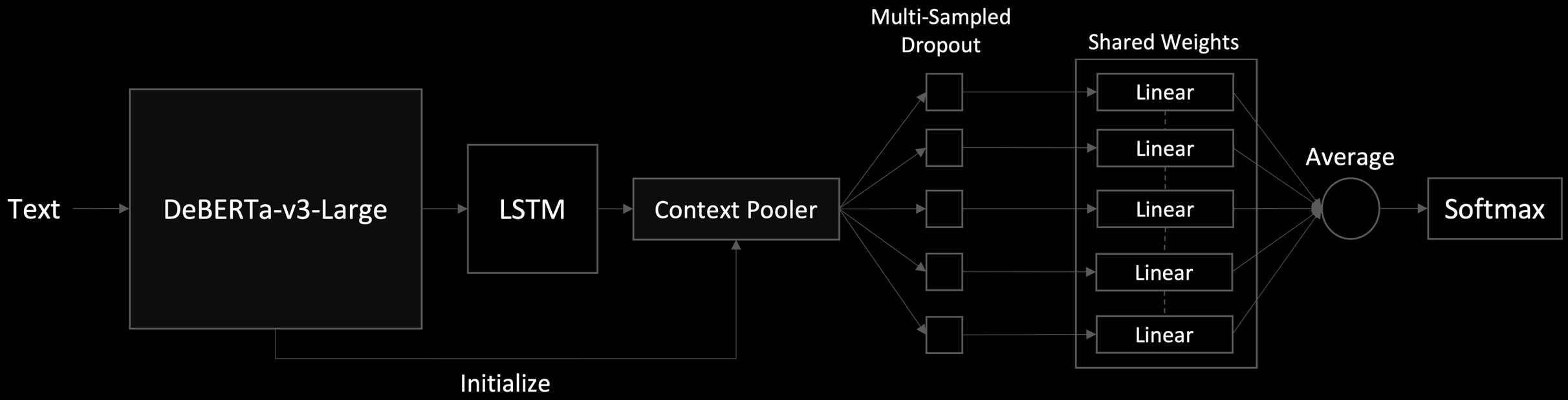


Public LB: 0.855

kakao

Model Structure

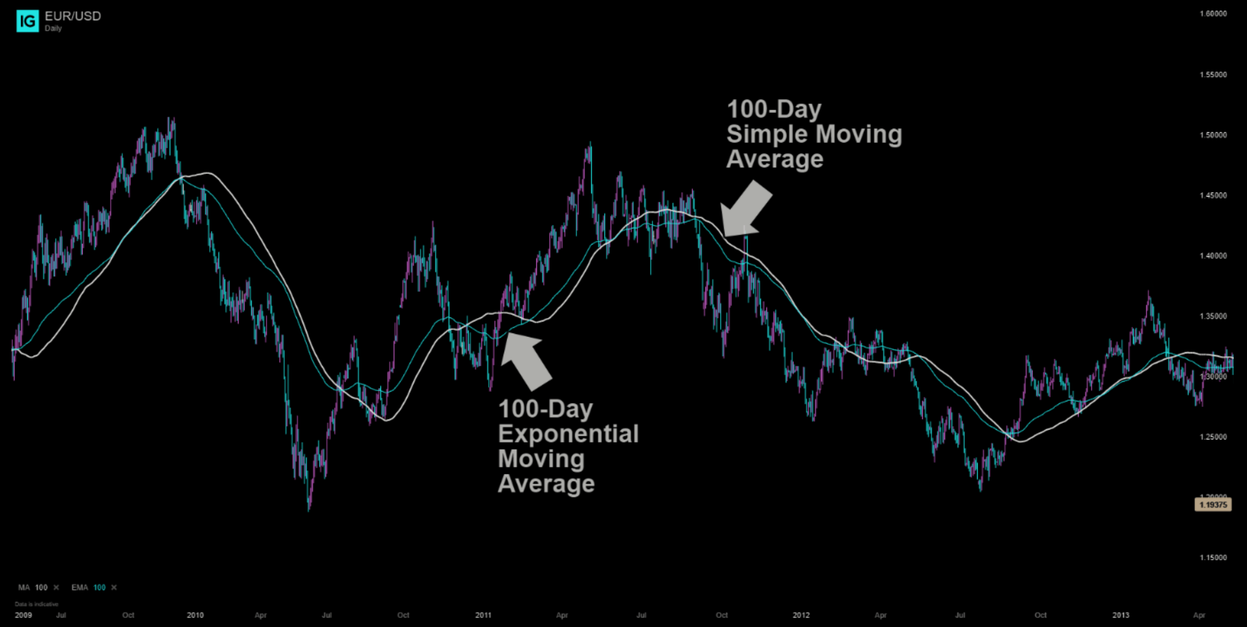
- Pre-trained Transformers
- LSTM Head with Different Learning Rate
- Multi-Sampled Dropout
- Adversarial Weight Perturbation (AWP)
- Exponential Moving Average of Weights
- Learning Rate Scheduling



kakao

Training Techniques

- Adversarial Weight Perturbation (AWP)
- Exponential Moving Average of Weights
- Learning Rate Scheduling



[1] https://www.dailyfx.com/education/moving-averages/simple-vs-exponential-moving-average.html

Diversity-based Model Ensemble

- Model Diversity
  - DeBERTa-v3-Large + RemBERT + XLM-RoBERTa-Large
  - Adding a different backbone with relatively lower cross validation score helped increase leaderboard score
- Data Diversity
  - DeBERTa-v3-Large validated on Fold 0, RemBERT validated on Fold 2…

kakao

Accelerated Inference

- Inference Times on 277044 samples using DeBERTa-v3-Large Model

Sequence Length	Float32 Precision	Float16 Precision
78	23 minutes	10 minutes
256	82 minutes	36 minutes

kakao

Summary of All Techniques

Technique	**Cross Validation Micro F1 Score**	**Relative Gain/Loss vs. Control Model**
DeBERTa-v3-Large (Baseline)	0.8229	+0.0000
Add LSTM Head	0.8226	-0.0003
Add LSTM Head with higher LR	0.8234	+0.0005
Adversarial Training	0.8262	+0.0028
Exponential Moving Average	0.8265	+0.0003
Multi-Sampled Dropout	0.8267	+0.0002
Cosine LR Schedule → Step LR Schedule	0.8267	+0.0005
Diversity-based Model Ensemble	-	-

Conclusion

- Built upon a solid validation strategy, we used a simple model structure of pretrained transformers, LSTM and Multi-Sampled Dropout to effectively prediction query-item relationship.
- All our strategies focus on increasing robustness of deep learning models and can be expanded to any other task that uses deep learning.

kakao

E.O.D

kakao