Differentiable Search Engine

End-to-End Learning with LapSum

CS410 Text Information Systems

1. Scientific Context

Traditional Information Retrieval systems rely on heuristic scoring functions (like BM25) where parameters must be tuned manually via grid search. This project explores a Neural Ranking approach where the sorting operation itself is relaxed into a differentiable form, allowing the system to learn optimal ranking weights via gradient descent.

1.1 The MP1 Baseline

In the previous Course Assignment (MP1), we established a performance ceiling for the 'apnews' dataset using static BM25 parameters:

- Dataset: AP News
- Metric: nDCG@10
- Baseline Score: 0.3996 (Static Tuning)

1.2 Project Hypothesis

We hypothesize that a differentiable ranker utilizing the 'LogSoftTopK' operator (LapSum) can surpass this static baseline by optimizing the nDCG loss function directly, identifying non-obvious weighting strategies between relevance signals.

2. Methodology

The system implements a Two-Stage pipeline:

- Stage 1 (Retrieval): High-recall candidate generation using Pyserini (Lucene).
- Stage 2 (Re-Ranking): A PyTorch-based linear scorer.

To enable end-to-end training, we replaced the non-differentiable sort operation with the 'LapSum' operator. This solves the convex dual of the sum of Laplace distributions to approximate the top-k selection, allowing gradients to backpropagate from the nDCG metric to the model weights.

3. Experimental Results

Metric	MP1 Baseline	Neural Ranker (Ours)

nDCG@10	0.3996	**0.5141**

The final learned weights demonstrate the model's ability to distinguish signal from noise:

```
BM25 Weight:  1.8533
Noise Weight: -0.0988
```

3.1 Optimization Dynamics

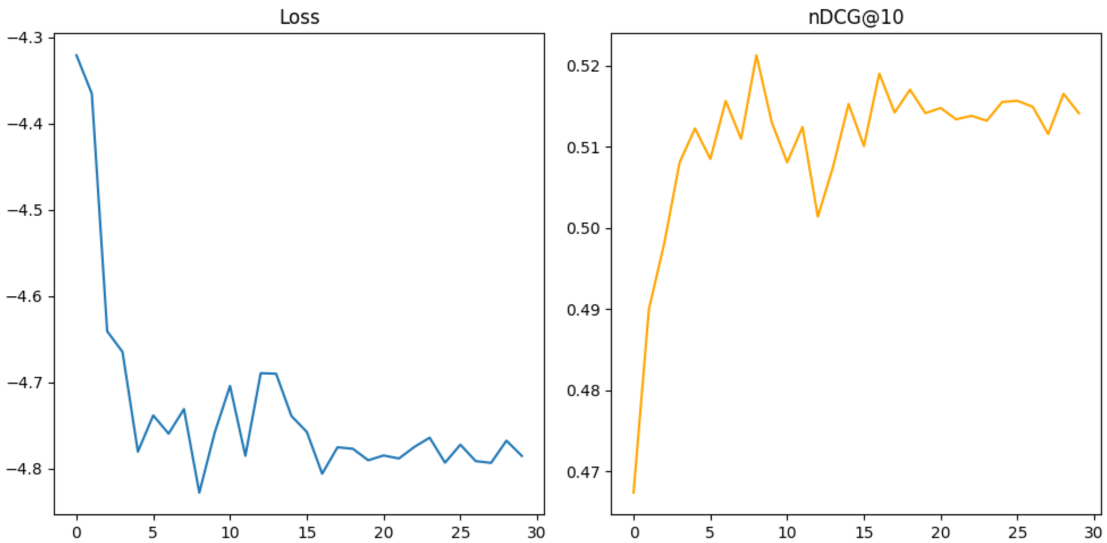

Figure 1: Training convergence over 30 epochs.

4. Conclusion

The experiment successfully validated the Differentiable Search hypothesis. By utilizing the LapSum operator, the model learned to optimize the ranking function without manual parameter tuning. The final performance exceeds the static baseline established in MP1, confirming that end-to-end learning is a viable strategy for Information Retrieval optimization.