Graph Attention Network on Istella22 dataset

Simone Boldrini
Alessandro Carella

Index

- Istella 22 Dataset
- Paper results recreation
- Graph Attention Network
- GAN Extensions
- Results (TensorBoard)

Istella 22 Dataset

Purpose

- Bridge the gap between feature-based and text-based Learning-to-Rank (LtR) models.
- Enable fair evaluation of both techniques on the same dataset.

Key Features

- 8.4M Web Documents: Multi-lingual corpus with preprocessed fields (e.g., text, title, URL).
- Queries & Judgments:
  - 2,198 test queries with 5-grade human relevance judgments.
  - Separate training, validation, and test sets.
- 220 Hand-Crafted Features: Cover query, document, and query-document interactions.

Impact

- Advances hybrid ranking techniques.
- Fills gaps by uniting text and feature vectors in a single resource.

Back to index

Paper results recreation

Lightgbm lambdamart

- We set the parameters to train the model to match the ones in the release model file and we achieved the same performances with a lower total number of trees.

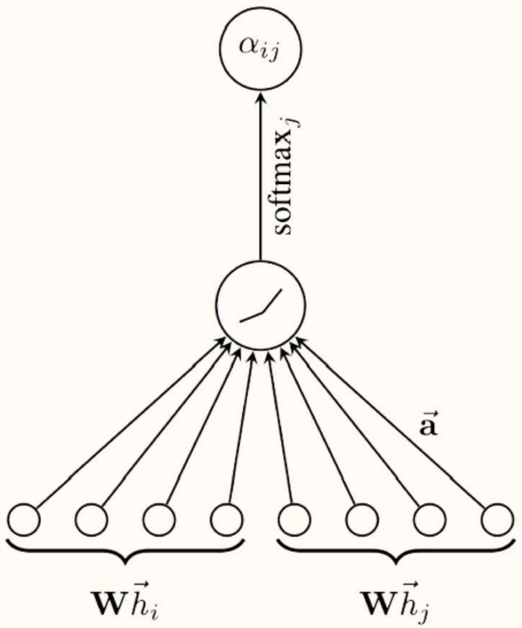
Number of trees: 1514

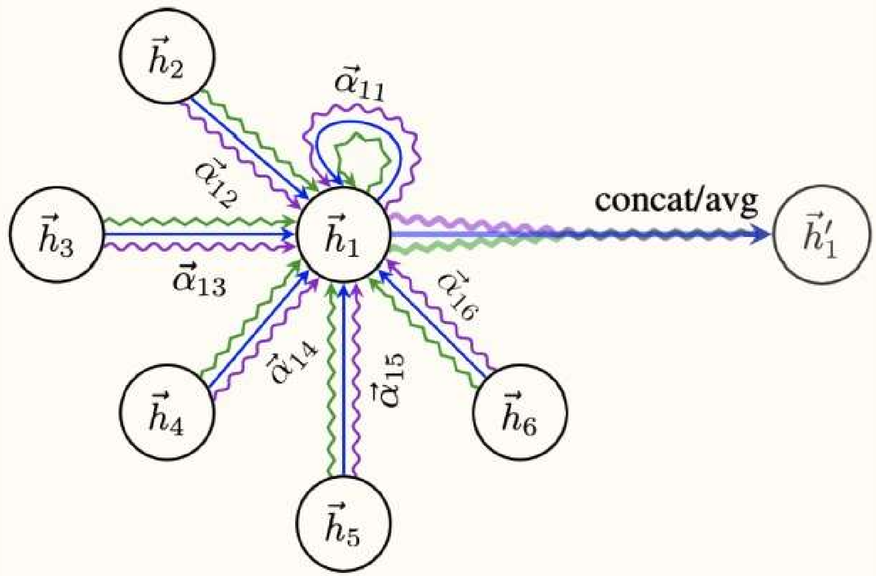
metric	Model Performance
MRR	0.972405
MAP	0.889125
P@1	0.955869
P@5	0.724530
P@10	0.460924
R@100	0.993639
R@1000	1.000000
NDCG@5	0.783217
NDCG@10	0.818759
NDCG@20	0.828635

Number of trees: 961

metric	Model Performance
MRR	0.970008
MAP	0.884904
P@1	0.951774
P@5	0.721436
P@10	0.460333
R@100	0.993072
R@1000	1.000000
NDCG@5	0.779751
NDCG@10	0.814866
NDCG@20	0.825039

Back to index



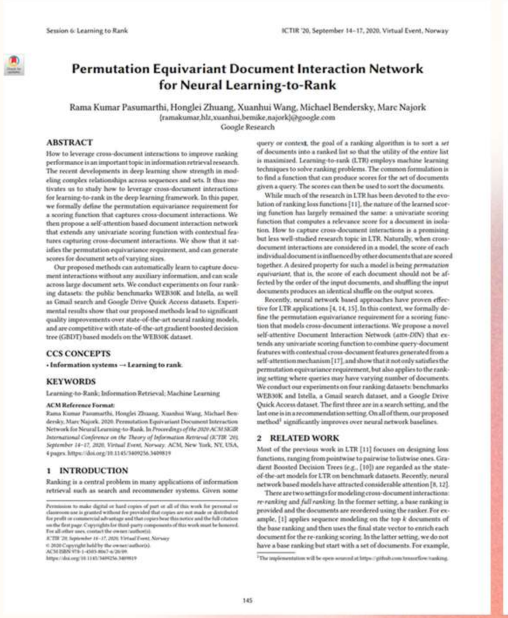


Graph Attention Network

- Key Features
  - Attention Mechanism:
    - Dynamically assigns weights to neighbors based on relevance for better feature aggregation.
  - Learnable Parameters:
    - Enhances representation by focusing on the most relevant nodes.
- Architecture
  - Nodes calculate attention scores for neighbors using multi-head attention.
  - Features are combined via weighted averaging.
- Advantages
  - Effectively handles diverse relationships.
  - Scalable for large graphs with sparse data.

Back to index

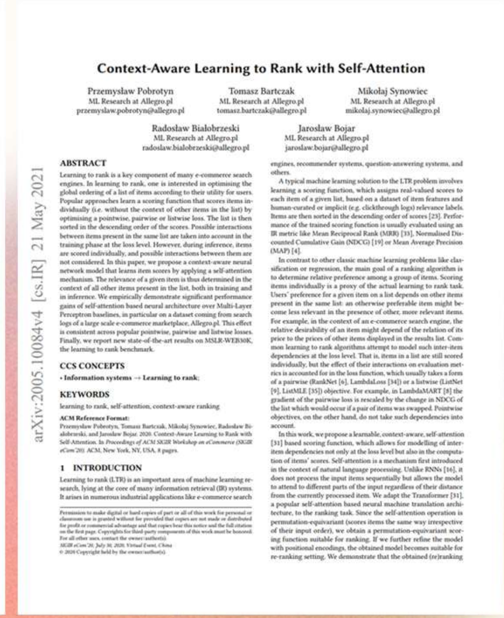
GAN Extensions



Permutation Equivariant Document Interaction Network for Neural Learning-to-Rank

Rama Kumar Pasumarthi, Honglei Zhuang, Xuanhui Wang, Michael Bendersky, Marc Najork

Google Research



Context-Aware Learning to Rank with Self-Attention

Przemysław Pobrotyn, Tomasz Bartczak, Mikołaj Synowiec, Radosław Białobrzeski, Jarosław Bojar

ML Research at Allegro.pl

arXiv:2005.10084v4 [cs.IR] 21 May 2021

Ordinal Loss

Structured Label Representation: Converts relevance labels into binary vectors for finer-grained learning.
Level-wise Predictions: Outputs separate relevance scores for each level.
Sigmoid Activation: Ensures smooth binary cross-entropy loss computation.

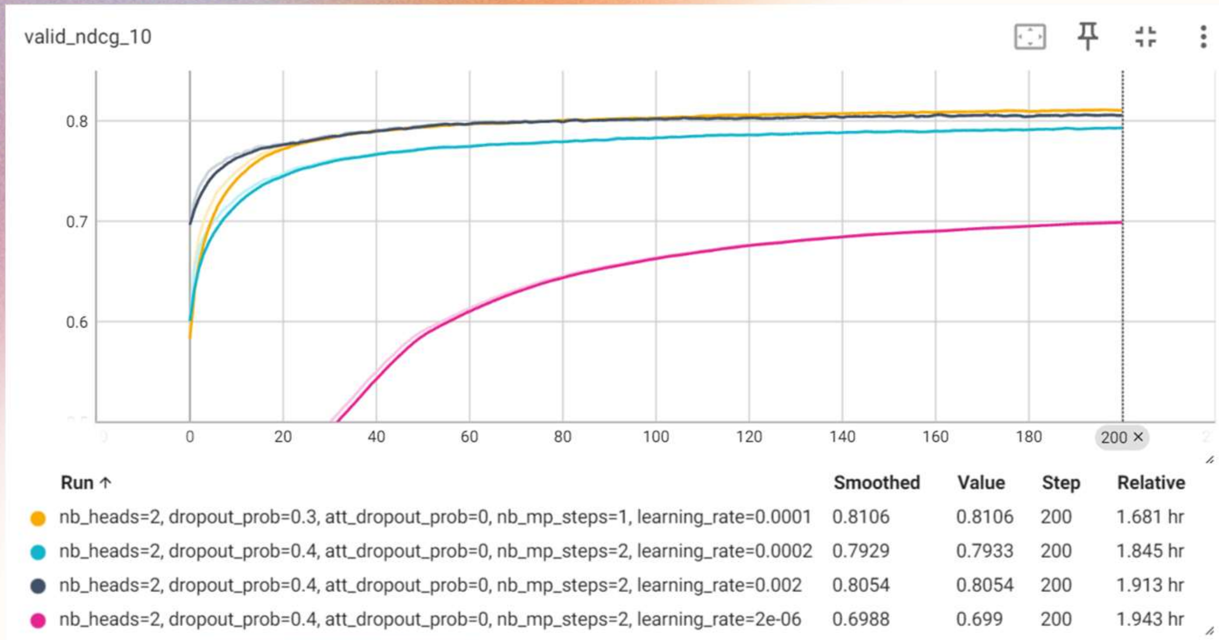
Multilayer perceptron

Multilayer Perception (MLP) in attn-DIN to enhance the Learning-to-Rank task by combining query-document features, given by the Istella 22 dataset, with contextual cross-document interactions derived from self-attention layers

Back to index

Results (TensorBoard)

- Interactive scalars dashboard for metrics.
- Graphs of computational models.
- Tools for comparing multiple runs.
- Easy integration with TensorFlow workflows.



[Back to index]

Results Comparison

Table 1: Comparison of NDCG[3] between various ranking models on the Web30K and Istella datasets. △/▽ indicate statistically significant increase/decrease of *attn-DIN* compared to best neural ranking baseline (p-value<0.05).

(a) WEB30K	NDCG@1	NDCG@5	NDCG@10
LambdaMART (RankLib)	0.4535	0.4459	0.4646
LambdaMART (lightGBM)	0.5057	0.4991	0.5183
LambdaMART + DLCM [1]	0.4630	0.4500	0.4690
GSF(m=64) with Softmax loss [2]	0.4421	0.4446	0.4677
FFNN with $\mathbb{E}$[ApproxNDCG] [3]	0.4951	0.4820	0.4996
SetRank with Softmax Loss [14]	0.4904	0.4885	0.5101
attn-DIN with Softmax Loss	0.5005△	0.5014△	0.5218△

(b) Istella	NDCG@1	NDCG@5	NDCG@10
LambdaMART (RankLib)	0.6571	0.6118	0.6591
LambdaMART (lightGBM)	0.7264	0.6883	0.7356
LambdaMART + DLCM [1]	0.6272	0.5848	0.6310
FFNN with Softmax Loss	0.6645	0.6422	0.6962
SetRank with Softmax Loss [14]	0.6702	0.6419	0.6958
attn-DIN with Softmax Loss	0.6747	0.6455△	0.6999△

Permutation Equivariant Document Interaction Network for Neural Learning-to-Rank paper results

Table 4: Performance of baseline retrieval and re-ranking systems. Re-ranking systems operate over the initial ranked list from Istella, and include LtR, neural-reranking, and hybrid LtR-using-neural-reranking systems.

Method	Feats.	Neural Text	P@1	P@5	P@10	NDCG@10	NDCG@20	MRR	MAP
Retrieval									
BM25 (default)	-	–	0.4331	0.2939	0.2055	0.2280	0.2447	0.5439	0.3649
BM25 (tuned)	-	–	0.4339	0.2947	0.2055	0.3854	0.4207	0.5494	0.3686
DPH	-	–	0.4408	0.2868	0.2020	0.2281	0.2443	0.5479	0.3618
Re-Ranking									
λ-MART	✓	–	0.9559	0.7245	0.4609	0.8188	0.8286	0.9724	0.8891
MONOT5-MSMARCO	–	Ttl+Txt	0.5568	0.3893	0.2699	0.2990	0.3157	0.6675	0.4889
MONOT5-mMARCO	–	Ttl+Txt	0.5868	0.4147	0.2829	0.3175	0.3338	0.6976	0.5203
MONOT5-tuned	–	Ttl+Txt	0.8407	0.5813	0.3792	0.4418	0.4482	0.9005	0.7262
MONOT5-tuned	–	Ttl+Url	0.8412	0.5990	0.3914	0.4402	0.4472	0.9025	0.7396
MONOT5-tuned	–	Ttl+Url+Txt	0.8581	0.5945	0.3910	0.4515	0.4586	0.9132	0.7462
λ-$\text{MART}_{\text{MONOT5}}$	✓	Ttl+Url	0.9550	0.7223	0.4597	0.8152	0.8258	0.9716	0.8859
λ-$\text{MART}_{\text{MONOT5}}$	✓	Ttl+Url+Txt	0.9509	0.7238	0.4602	0.8153	0.8258	0.9701	0.8849

The Istella22 Dataset: Bridging Traditional and Neural Learning to Rank Evaluation paper various methods results

Back to index

Results Comparison

Table 1: Comparison of NDCG[3] between various ranking models on the Web30K and Istella datasets. △/▽ indicate statistically significant increase/decrease of *attn-DIN* compared to best neural ranking baseline (p-value<0.05).

(a) WEB30K	NDCG@1	NDCG@5	NDCG@10
LambdaMART (RankLib)	0.4535	0.4459	0.4646
LambdaMART (lightGBM)	0.5057	0.4991	0.5183
LambdaMART + DLCM [1]	0.4630	0.4500	0.4690
GSF(m=64) with Softmax loss [2]	0.4421	0.4446	0.4677
FFNN with $\mathbb{E}$[ApproxNDCG] [3]	0.4951	0.4820	0.4996
SetRank with Softmax Loss [14]	0.4904	0.4885	0.5101
attn-DIN with Softmax Loss	0.5005$^{\triangle}$	0.5014$^{\triangle}$	0.5218$^{\triangle}$

(b) Istella	NDCG@1	NDCG@5	NDCG@10
LambdaMART (RankLib)	0.6571	0.6118	0.6591
LambdaMART (lightGBM)	0.7264	0.6883	0.7356
LambdaMART + DLCM [1]	0.6272	0.5848	0.6310
FFNN with Softmax Loss	0.6645	0.6422	0.6962
SetRank with Softmax Loss [14]	0.6702	0.6419	0.6958
attn-DIN with Softmax Loss	0.6747	0.6455$^{\triangle}$	0.6999$^{\triangle}$

Permutation Equivariant Document Interaction Network for Neural Learning-to-Rank paper results

Table 4: Performance of baseline retrieval and re-ranking systems. Re-ranking systems operate over the initial ranked list from Istella, and include LtR, neural-reranking, and hybrid LtR-using-neural-reranking systems.

Method	Feats.	Neural Text	P@1	P@5	P@10	NDCG@10	NDCG@20	MRR	MAP
Retrieval									
BM25 (default)	-	–	0.4331	0.2939	0.2055	0.2280	0.2447	0.5439	0.3649
BM25 (tuned)	-	–	0.4339	0.2947	0.2055	0.3854	0.4207	0.5494	0.3686
DPH	-	–	0.4408	0.2868	0.2020	0.2281	0.2443	0.5479	0.3618
Re-Ranking									
λ-MART	✓	–	0.9559	0.7245	0.4609	0.8188	0.8286	0.9724	0.8891
monoT5-MSMARCO	–	Ttl+Txt	0.5568	0.3893	0.2699	0.2990	0.3157	0.6675	0.4889
monoT5-mMARCO	–	Ttl+Txt	0.5868	0.4147	0.2829	0.3175	0.3338	0.6976	0.5203
monoT5-tuned	–	Ttl+Txt	0.8407	0.5813	0.3792	0.4418	0.4482	0.9005	0.7262
monoT5-tuned	–	Ttl+Url	0.8412	0.5990	0.3914	0.4402	0.4472	0.9025	0.7396
monoT5-tuned	–	Ttl+Url+Txt	0.8581	0.5945	0.3910	0.4515	0.4586	0.9132	0.7462
λ-MART$_{\text{monoT5}}$	✓	Ttl+Url	0.9550	0.7223	0.4597	0.8152	0.8258	0.9716	0.8859
λ-MART$_{\text{monoT5}}$	✓	Ttl+Url+Txt	0.9509	0.7238	0.4602	0.8153	0.8258	0.9701	0.8849

The Istella22 Dataset: Bridging Traditional and Neural Learning to Rank Evaluation paper various methods results

Metric	Value	Metric	Value
P@1	0.919472	R@100	0.988753
P@5	0.669700	R@1000	1.000000
P@10	0.427389	RR	0.952130
nDCG(dcg='exp-log2')@5	0.727369	AP	0.830362
nDCG(dcg='exp-log2')@10	0.770694	Judged@10	1.000000
nDCG(dcg='exp-log2')@20	0.785701		

Our model results

Back to index