

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

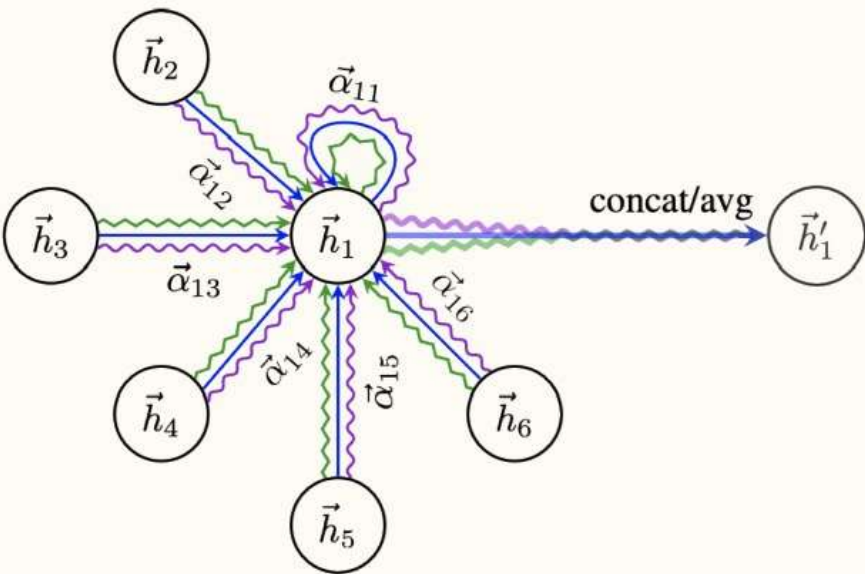
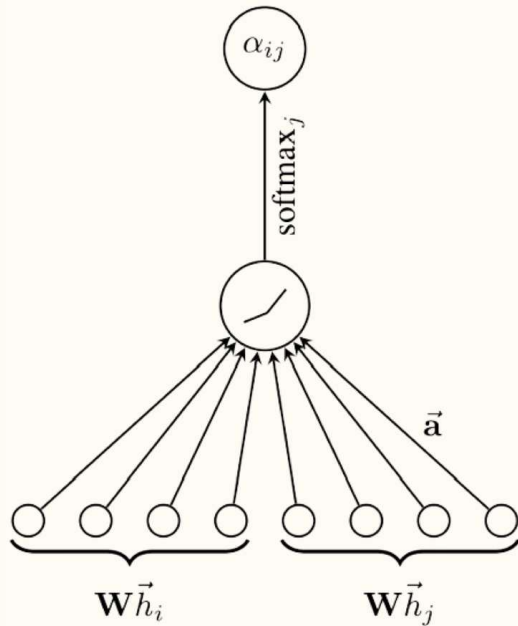
Lightgbm lambdamart

- 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		Number of trees: 961	
Model Performance		Model Performance	
metric		metric	
MRR	0.972405	MRR	0.970008
MAP	0.889125	MAP	0.884904
P@1	0.955869	P@1	0.951774
P@5	0.724530	P@5	0.721436
P@10	0.460924	P@10	0.460333
R@100	0.993639	R@100	0.993072
R@1000	1.000000	R@1000	1.000000
NDCG@5	0.783217	NDCG@5	0.779751
NDCG@10	0.818759	NDCG@10	0.814866
NDCG@20	0.828635	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.

GAN Extensions

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

Rama Kumar Pasumarthi, Honglei Zhuang, Xuanhui Wang, Michael Bendersky, Marc Najork
[ramakumar.kh, xuanhui, benderk, najork]@google.com
Google Research

ABSTRACT

How to leverage cross-document interactions to improve ranking performance is an important topic in information retrieval research. The recent developments in deep learning show strength in modeling complex relationships across sequences and sets. In this paper, we propose a self-attention based document interaction network that extends any univariate scoring function with contextual features capturing cross-document interactions. We show that it satisfies the permutation-equivariance requirement, and can generate scores for document sets of varying sizes.

The proposed methods can automatically learn to capture document interactions without any auxiliary information, and can scale across large document sets. We conduct experiments on four ranking datasets: the public benchmarks WEB3K and Istella, as well as Gmail search and Google Drive Quick Access datasets. Experimental results show that our proposed methods lead to significant quality improvements over state-of-the-art neural ranking models, and are competitive with state-of-the-art gradient boosted decision tree (GBDT) based models on the WEB3K dataset.

CCS CONCEPTS

Information systems → Learning to rank

KEYWORDS

Learning to Rank, Information Retrieval, Machine Learning

ACM Reference Format:

Rama Kumar Pasumarthi, Honglei Zhuang, Xuanhui Wang, Michael Bendersky, Marc Najork. 2020. Permutation Equivariant Document Interaction Network for Neural Learning-to-Rank. In *Proceedings of the 2020 ACM SIGIR International Conference on the Theory of Information Retrieval (ICTIR '20)*, September 14–17, 2020, Virtual Event, Norway. ACM, New York, NY, USA, 9 pages. <https://doi.org/10.1145/3392523.3392527>

1 INTRODUCTION

Ranking is a central problem in many applications of information retrieval such as search and recommendation systems. Given some information to rank, the goal is to learn a function that orders the items in the information set according to their relevance to the user. The relevance is generally defined as the probability that the user will click on the item. The goal is to learn a function that orders the items in the information set according to their relevance to the user. The relevance is generally defined as the probability that the user will click on the item. The goal is to learn a function that orders the items in the information set according to their relevance to the user. The relevance is generally defined as the probability that the user will click on the item.

Context-Aware Learning to Rank with Self-Attention

Premysław Poborszyn
ML Research at Allegro.pl
premyslawn.poborszyn@allegro.pl

Tomasz Bartczak
ML Research at Allegro.pl
tomasz.bartczak@allegro.pl

Mikołaj Synowiewicz
ML Research at Allegro.pl
mikołaj.synowiewicz@allegro.pl

Radosław Białobrzewski
ML Research at Allegro.pl
radoslaw.bialobrzewski@allegro.pl

Jarosław Bojar
ML Research at Allegro.pl
jaroslaw.bojar@allegro.pl

ABSTRACT

Learning to rank is a key component of many e-commerce search engines. In learning to rank, one is interested in optimizing the global ordering of a list of items according to their utility for users. Possible approaches learn a scoring function that scores items individually (i.e., without the context of other items in the list) by optimizing a pointwise, pairwise or listwise loss. The list is then sorted in the descending order of the scores. Possible interactions between items present in the same list are taken into account in the training phase at the list level. However, during inference, items are scored individually, and possible interactions between them are not considered. In this paper, we propose a context-aware neural network model that learns item scores by applying a self-attention mechanism. The relevance of a given item in the list depends on the context of all other items present in the list, both in training and in inference. We empirically demonstrate significant performance gains of self-attention based neural architecture over Multi-Layer Perceptron baselines, in particular on a dataset coming from search logs of a large scale e-commerce marketplace, Allegro.pl. This effect is consistent across popular pointwise, pairwise and listwise losses. Finally, we report new state-of-the-art results on MSUR, WEB3K, the learning to rank benchmarks.

CCS CONCEPTS

Information systems → Learning to rank

KEYWORDS

Learning to rank, self-attention, context-aware ranking

ACM Reference Format:

Premysław Poborszyn, Tomasz Bartczak, Mikołaj Synowiewicz, Radosław Białobrzewski, and Jarosław Bojar. 2020. Context-Aware Learning to Rank with Self-Attention. In *Proceedings of ACM SIGIR Workshop on e-commerce (eCom '20)*, ACM, New York, NY, USA, 9 pages.

1 INTRODUCTION

Learning to rank (LTR) is an important area of machine learning research, lying at the core of many information retrieval (IR) systems. It arises in numerous industrial applications like e-commerce search engines. In learning to rank, one is interested in optimizing the global ordering of a list of items according to their utility for users. Possible approaches learn a scoring function that scores items individually (i.e., without the context of other items in the list) by optimizing a pointwise, pairwise or listwise loss. The list is then sorted in the descending order of the scores. Possible interactions between items present in the same list are taken into account in the training phase at the list level. However, during inference, items are scored individually, and possible interactions between them are not considered. In this paper, we propose a context-aware neural network model that learns item scores by applying a self-attention mechanism. The relevance of a given item in the list depends on the context of all other items present in the list, both in training and in inference. We empirically demonstrate significant performance gains of self-attention based neural architecture over Multi-Layer Perceptron baselines, in particular on a dataset coming from search logs of a large scale e-commerce marketplace, Allegro.pl. This effect is consistent across popular pointwise, pairwise and listwise losses. Finally, we report new state-of-the-art results on MSUR, WEB3K, the learning to rank benchmarks.

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.

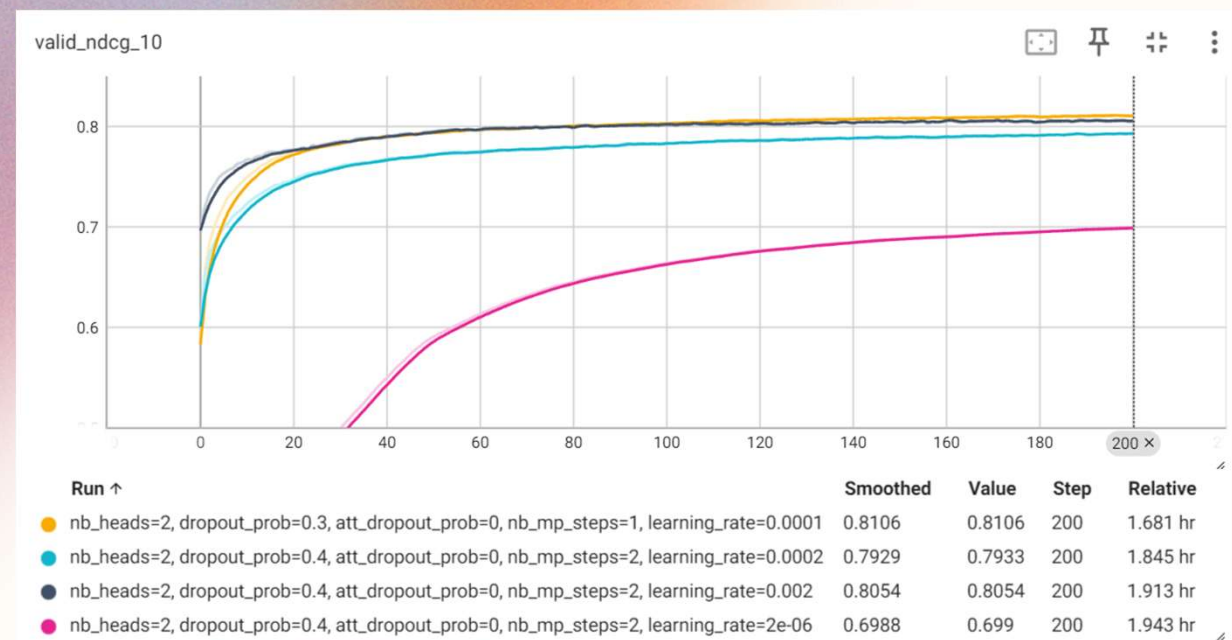
Multilayer perceptron

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.



Results Comparison

Table 1: Comparison of NDCG³ between various ranking models on the Web30K and Istella datasets. Δ/∇ 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 $\mathbb{E}[\text{ApproxNDCG}]$ [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 Δ	0.5014 Δ	0.5218 Δ

(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 Δ	0.6999 Δ

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
<i>Retrieval</i>									
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
<i>Re-Ranking</i>									
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 _{MONOT5}	✓	Ttl+Url	0.9550	0.7223	0.4597	0.8152	0.8258	0.9716	0.8859
λ -MART _{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 NDCG³ between various ranking models on the Web30K and Istella datasets. Δ/∇ 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 Δ	0.5014 Δ	0.5218 Δ
(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 Δ	0.6999 Δ

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
<i>Retrieval</i>									
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
<i>Re-Ranking</i>									
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 _{MONOT5}	✓	Ttl+Url	0.9550	0.7223	0.4597	0.8152	0.8258	0.9716	0.8859
λ -MART _{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](#)