

TSPRank: Bridging Pairwise and Listwise Methods with a Bilinear Travelling Salesman Model

Waylon Li¹ Yftah Ziser² Yifei Xie¹ Shay B. Cohen¹ Tiejun Ma¹

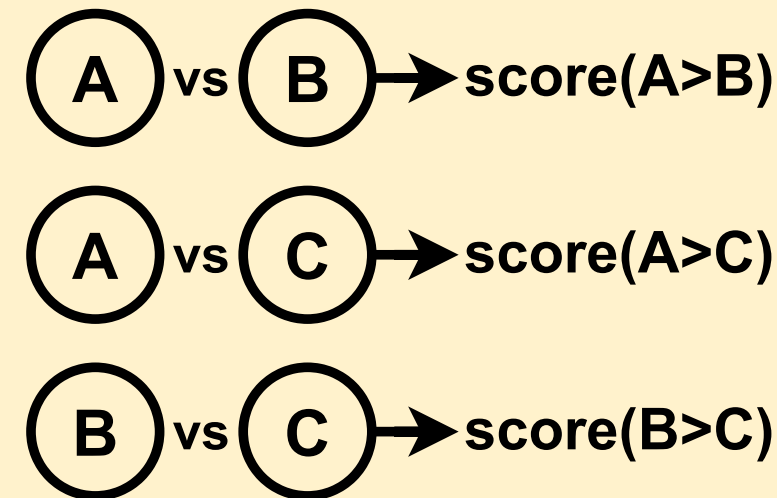
¹University of Edinburgh ²Nvidia Research

Overview

- We propose TSPRank, a novel ranking framework that **formulates ranking as a Travelling Salesman Problem**, enabling direct optimisation for globally coherent rankings.
- By bridging pairwise and listwise approaches, TSPRank leverages a bilinear scoring function and combinatorial inference to jointly capture local preferences and global order.
- TSPRank delivers consistently stronger global ranking metrics (e.g., Kendall's Tau, NDCG) and **maintains high performance even with noisy pairwise signals**.

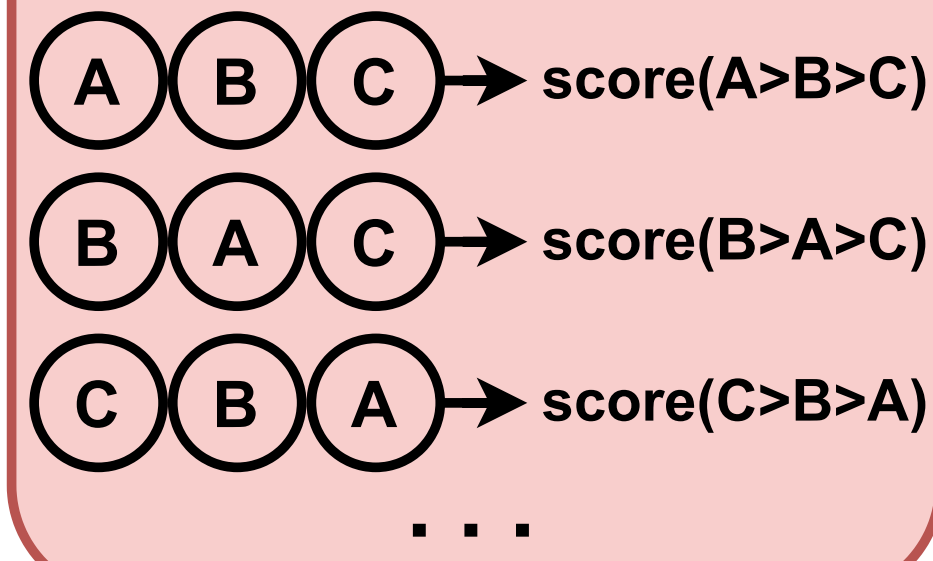
Gap Between Pairwise and Listwise Ranking

Pairwise Ranking



- ✓ Robust (usually GBDT-based).
 - ✗ Not optimised on list level, leading to sub-optimal results [3].
- Representative method: LambdaMART [1]

Listwise Ranking



- ✓ Capture the list-level information, optimised for listwise order.
 - ✗ Less robust and require complex tuning to achieve marginal gains over pairwise models like LambdaMART on information retrieval benchmarks [8].
- Representative method: deep learning based (SetRank [6], Rankformer [2])

Question: Is it possible to combine the advantages of both pairwise and listwise methods?



Predicting the order of a list is challenging because ranking N entities from 1 to N is complex. However, breaking it down into $N \times N$ pairwise comparisons simplifies the task, as each pairwise comparison is more straightforward than ranking the entire list.

Travelling Salesman Problem

Traveling Salesman Problem

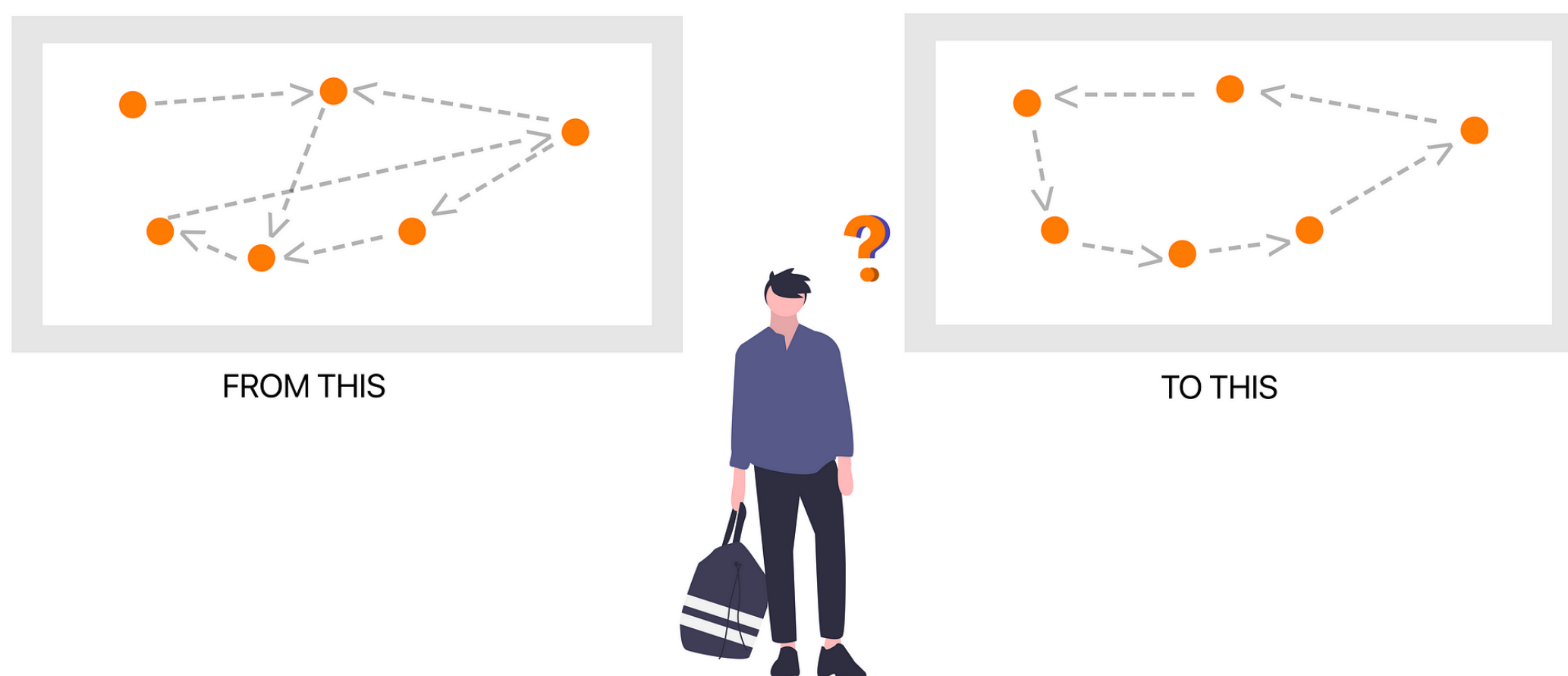
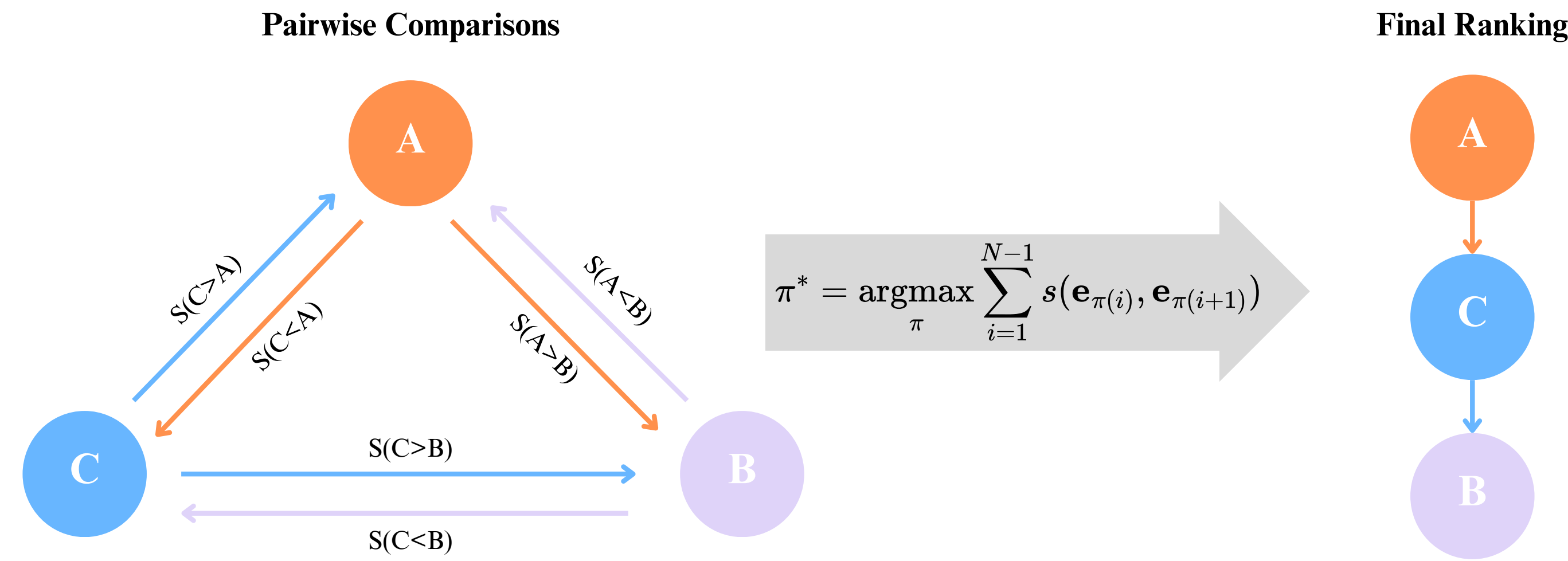


Figure 1. Illustration of the travelling salesman problem. (Adapted from <https://www.linkedin.com/pulse/traveling-salesman-problem-14-different-solutions-sandeep-kella/>)

Rethink Pairwise Ranking in a Graph ...



Doesn't it look similar to TSP?

So, we consider ranking as a TSP where the traveller does not go back to the start point at the end. It is also referred as the Open-Loop TSP.

TSPRank: A Generic Ranking Model for Existing Backbone Encoders

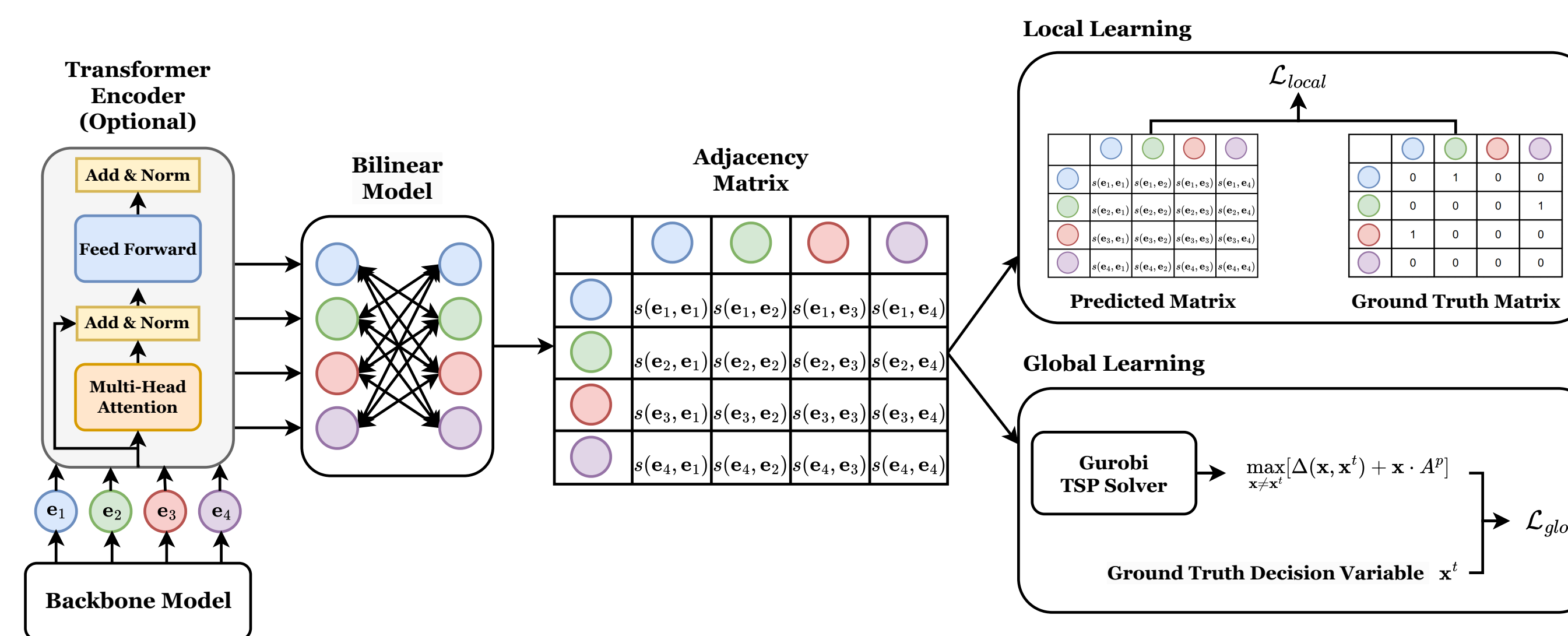


Figure 2. Illustration of TSPRank and the complete pipeline of local and global learning. The pipeline starts with a Transformer Encoder or any embeddings generated from another existing Backbone Model, followed by a Bilinear Model generating pairwise scores to form an Adjacency Matrix. Local learning compares the predicted matrix with the ground truth to calculate the local loss $\mathcal{L}_{local}$. Global learning uses the max-margin loss $\mathcal{L}_{glob}$ to incorporate the Gurobi TSP solver during training.

Experimental Setup

Dataset	Task
Stock Ranking: introduced by [4], which includes NASDAQ and NYSE.	Rank next day stocks in the same sector and historical trading data from 2013 to 2017 for choose the top-K to invest.
Information Retrieval: MQ2008-list [7] from Microsoft.	Rank a list of documents based on their technical indicators.
Event Ordering: "On This Day 2" (OTD2) [5]	Event Ordering: chronologically ordering historical events given text embeddings.

Benchmark Models:

- LambdaMART [1] (pairwise, GBDT-based)
- Rankformer [2] (listwise, transformer-based)
- TSPRank (hybrid, ours)

Backbone Encoders:

- Stock ranking: [4]
- Information retrieval: BM25, TF-IDF
- Event ordering: *text-embedding-3-small*

Results

Market	Model	τ	IRR@1	SR@1	MAP@1	IRR@3	SR@3	MAP@3	IRR@5	SR@5	MAP@5
NASDAQ	[4] + MLP (Original)	0.0093	0.1947	0.5341	0.1690	0.2366	0.9881	0.3253	0.1892	0.9682	0.5871
	[4] + LambdaMART	0.0071	0.0310	-0.0873	0.1539	0.0340	0.0445	0.3144	0.0505	0.2678	0.5858
	[4] + Rankformer	0.0110	0.2257	0.5464	0.1620	0.2857	1.1245	0.3216	0.2309	1.0943	0.5860
	[4] + TSPRank-Local	0.0291	0.5353	1.2858	0.1658	0.4416	1.7401	0.3297	0.2537	1.2623	0.5932
	[4] + TSPRank-Global	0.0447	0.7849	1.7471	0.1633	0.5224	2.0359	0.3364	0.2937	1.4331	0.5999
NYSE	[4] + MLP (Original)	0.0162	0.4170	1.0755	0.1791	0.2574	1.2367	0.2841	0.2257	1.3186	0.4649
	[4] + LambdaMART	0.0054	0.1005	0.1367	0.1307	0.0732	0.4192	0.2592	0.1063	0.6882	0.4574
	[4] + Rankformer	0.0181	0.2924	0.9113	0.1535	0.2701	1.2890	0.2758	0.2515	1.4200	0.4651
	[4] + TSPRank-Local	0.0313	0.5012	1.5710	0.1424	0.3974	1.9735	0.2756	0.2788	1.6662	0.4680
	[4] + TSPRank-Global	0.0422	0.4787	1.4552	0.1392	0.3889	1.9976	0.2756	0.2816	1.7350	0.4732

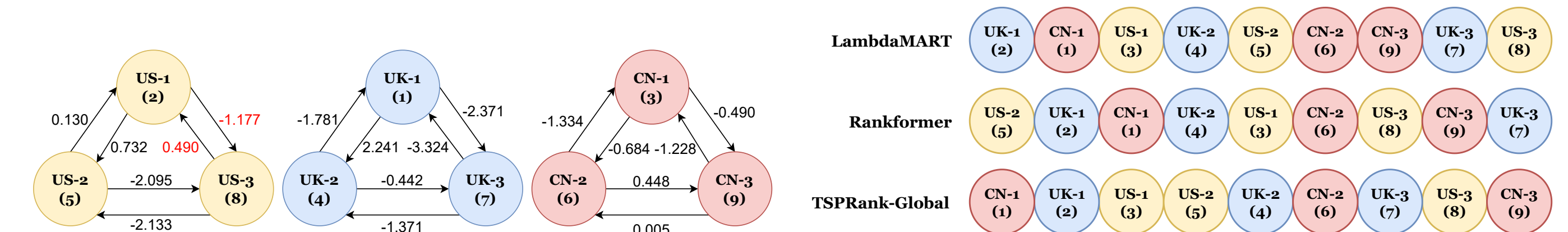
Table 1. Performance on the NASDAQ and NYSE stock ranking dataset, averaged across all filtered sectors.

		Top 10					Top 30				
Model	Type	NDCG@3	NDCG@5	NDCG@10	MRR	τ	NDCG@3	NDCG@5	NDCG@10	MRR	τ
LambdaMART	Pairwise	0.6833	0.7222	0.8707	0.4259	0.1474	0.7340	0.7298	0.7403	0.3617	0.2372
Rankformer	Listwise	0.7220	0.7565	0.8865	0.4661	0.2317	0.7486	0.7470	0.7596	0.3732	0.2834
TSPRank-Local	Pairwise-Listwise	0.6858	0.7213	0.8719	0.4266	0.1544	0.7189	0.7240	0.7362	0.3206	0.2054
TSPRank-Global	Pairwise-Listwise	0.7281	0.7585	0.8884	0.4861	0.2212	0.7582	0.7558	0.7631	0.3895	0.2647

Table 2. Evaluation on MQ2008-list information retrieval dataset for top 10 and top 30 documents.

Group Size		10				30				50			
Model	Type	$\tau \uparrow$	EM $\uparrow$	MRR $\uparrow$	RMSE $\downarrow$	$\tau \uparrow$	EM $\uparrow$	MRR $\uparrow$	RMSE $\downarrow$	$\tau \uparrow$	EM $\uparrow$	MRR $\uparrow$	RMSE $\downarrow$
te-3-small + LambdaMART	Pairwise	0.6297	0.3008	0.7554	1.993	0.5929	0.1064	0.6122	5.969	0.6000	0.0639	0.5596	9.618
te-3-small + Rankformer	Listwise	0.6190	0.2899	0.7361	1.998	0.5859	0.0921	0.4911	5.973	0.5724	0.0527	0.3526	10.069
te-3-small + TSPRank-Local	Pairwise-Listwise	0.5658	0.2856	0.7679	2.296	0.5095	0.0873	0.5739	6.930	0.4713	0.0460	0.3949	12.084
te-3-small + TSPRank-Global	Pairwise-Listwise	0.6301	0.3350	0.7936	2.057	0.6302	0.1384	0.7300	5.770	0.6207	0.0871	0.6618	9.602

Table 3. Evaluation on historical events ordering for group sizes of 10, 30, and 50. "te-3-small" stands for "text-embedding-3-small".



References

- Christopher Burges. From ranknet to lambdarank to lambdamart: An overview. *Learning*, 11, 01 2010.
- Maarten Buyt, Paul Missault, and Pierre-Antoine Sondag. Rankformer: Listwise learning-to-rank using listwise labels. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD '23*, page 3762–3773, New York, NY, USA, 2023. Association for Computing Machinery.
- Zhe Cao, Tao Qin, Tie-Yan Liu, Ming-Feng Tsai, and Hang Li. Learning to rank: from pairwise approach to listwise approach. In *Proceedings of the 24th International Conference on Machine Learning, ICLR '07*, page 129–136, New York, NY, USA, 2007. Association for Computing Machinery.
- Fuli Feng, Xiangnan He, Xiang Wang, Cheng Luo, Yiqun Liu, and Tat-Seng Chua. Temporal relational ranking for stock prediction. *ACM Trans. Inf. Syst.*, 37, 2019.
- Or Honovich, Lucas Torroba Hennigen, Omri Abend, and Shay B. Cohen. Machine reading of historical events. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7486–7497, 2020.
- Liang Pang, Jun Xu, Qingyao Ai, Yanyan Lan, Xueqi Cheng, and Jirong Wen. Setrank: Learning a permutation-invariant ranking model for information retrieval. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 499–508, 2020.
- Tao Qin and Tie-Yan Liu. Introducing letor 4.0 datasets. *ArXiv*, abs/1306.2597, 2013.
- Zhen Qin, Le Yan, Honglei Zhuang, Yi Tay, Rama Kumar Pasumarthi, Xuanhui Wang, Mike Bendersky, and Marc Najork. Are neural rankers still outperformed by gradient boosted decision trees? In *International Conference on Learning Representations (ICLR)*, 2021.

Paper



Code & Data

