Dataset	Queries	Items	Rel.	Feats.
Airline Itinerary	33951	1,089k	2	17
Microsoft 30k	31531	3,771k	5	136
Istella-S	33018	3,302k	5	220

Table 1: Characteristics of the datasets used in the experiments: number of queries, items, relevance levels, and features.

Evaluation Metrics

For each task, we adopt the evaluation metrics used in prior work. For ranking refinement tasks, we follow (**?**) and use the standard discounted cumulative gain (NDCG) metrics that include NDCG@1, NDCG@3, NDCG@5, and NDCG@10. For Microsoft 30k dataset, an additional metric NDCG@50 was also evaluated. On the other hand, for user choice ranking tasks where relevance feedbacks are binary-valued, we follow (**?**) and use top precision metrics that include P@1(Precision@1) and P@5.

Model Training

Each dataset is split in train, validation and test sets according to a 60%-20%-20% scheme. The validation set was used to select the optimal hyperparameters for all involved methods. We did not perform extensive hyperparameter search for the proposed model, and used virtually the same architecture throughout all the experiments and datasets. More specifically, the dimensions of the nonlinear transformations (**??**) in the Input Encoder were fixed as 100, while MLPs used in (**??**) and (**??**) consist of 2 hidden layers with either 256 or 128 ELUs. The models were trained with the Adam algorithm (**?**) with a learning rate of 0.001, batch size of 80. Training generally converged after less than 100 passes through the entire training dataset.

Comparisons with Baseline Methods

The performance comparison results of various methods are reported in Table **??**. To eliminate the influence of random initiations, all results are averaged over 20 runs. As is shown in this table, QILCM significantly outperforms all the baselines.

Ablation Studies

To elucidate the contributions of the main components of our system, in this section, we test several variants of the proposed model. The tested implementations include:

- **Variant 1**: Our model with the attention-weighted context encoding (**??**) replaced by simple average pooling of the item encoding vectors. Accordingly, the attention-weighted statistics (**??**) and (**??**) are replaced with standard mean and variance;
- **Variant 2**: Our model without the domain confusion loss;
- **Variant 3**: Our model without the domain confusion loss and QN layer.

The performance results of different implementations of the proposed methods are shown in Table **??**, which shows that

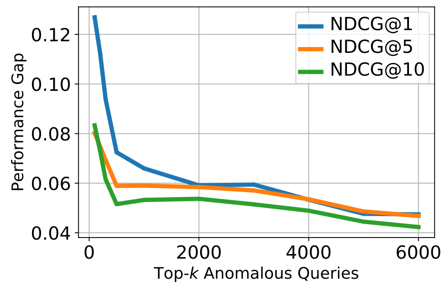

Figure 2: Performance gap between QILCM and the best-performing baseline (DLCM) on Microsoft 30k dataset.

QILCM consistently achieves the best performance among all model variants.

Anomalous Query Analysis

To further investigate the performance of QILCM, in this section we conducted additional analysis of experimental results on the Microsoft 30k dataset. Concretely, we firstly followed (**?**), and constructed vector representation for each query by averaging over the feature values of the top k ranked items (k was set as 10 during the experiments). After that, we fitted Isolation Forest (**?**) to the training queries, and then used the learned model to assign a 'anomaly score' to each query in the test set. Intuitively, this score quantifies how different a query is from the majority of training data, and we examine the performance difference between QILCM and the best-performing baseline (DLCM) for queries with different levels of anomaly. As shown in Figure **??**, the performance gap between QILCM and DLCM is significantly widened for more anomalous queries. For example, the NDCG@1 gap between QILCM and DLCM is increased from 0.047 to 0.121, which clearly demonstrates the advantage of tackling heterogeneous queries using the proposed DG perspective.

Conclusion

In this paper, we introduce a DG formulation of the LETOR problem and propose a novel neural architecture for DG in this LETOR context. We evaluate our techniques on three benchmark datasets, demonstrating that the proposed approach outperforms previous state-of-the-art approaches by a substantial margin.

Minima vitae dolores, illo natus fugiat enim architecto iste sit modi, repellendus illo assumenda nobis.Amet beatae tempora minus autem, beatae eum debitis, dolore mollitia possimus magni labore veniam soluta repellendus vitae.Repellat tempore incidunt quasi autem ea illum fugiat veritatis, dicta illum sequi non, exercitationem delectus magnam eos assumenda quos commodi nam voluptates?Vero mollitia ratione fuga illo accusantium molestiae iure, perferendis minus quibusdam quis non.Dicta iure odio, inventore at temporibus porro.Temporibus aperiam officiis, aliquam dolorum accusamus alias fugit dignissimos, cum similique consectetur distinctio esse odit reiciendis blanditiis laboriosam tem-

Dataset	Metrics	QILCM	DCM	DLCM	LambdaMART	Improv.	P-value
Airline Itinerary	P@1	*0.2833	0.2618	0.2562	0.2527	8.21%	$2.40 \times 10^{-[illegible]}$
	P@5	*0.6958	0.6586	0.6651	0.6239	4.61%	$1.22 \times 10^{-[illegible]}$
Microsoft 30k	NDCG@1	*0.5447	0.4938	0.4982	0.4880	9.33%	[illegible]
	NDCG@3	*0.5313	0.4827	0.4811	0.4766	10.06%	$7.58 \times 10^{-[illegible]}$
	NDCG@5	*0.5368	0.4901	0.4892	0.4842	9.46%	$1.60 \times 10^{-[illegible]}$
	NDCG@10	*0.5564	0.5093	0.5135	0.5061	8.35%	$2.27 \times 10^{-[illegible]}$
	NDCG@50	*0.6482	0.6127	0.6146	0.6092	5.46%	$4.51 \times 10^{-[illegible]}$
Istella-S	NDCG@1	*0.7023	0.6762	0.6873	0.6644	2.18%	$1.74 \times 10^{-[illegible]}$
	NDCG@3	*0.6696	0.6552	0.6537	0.6378	2.20%	1.89×10^{-8}
	NDCG@5	*0.6953	0.6861	0.6831	0.6741	1.36%	$4.23 \times 10^{-[illegible]}$
	NDCG@10	*0.7645	0.7558	0.7566	0.7456	1.04%	4.67×10^{-7}

Table 2: Performance comparison of various methods. The results are averaged over 20 random runs, and the best ones are marked with *. The last two columns show the improvement of QILCM over the best baseline algorithm (highlighted with underline), and the corresponding Student's t-test P-values.

Dataset	Metrics	QILCM	Variant 1	Variant 2	Variant 3
Airline Itinerary	P@1	*0.2833	0.2749	0.2762	0.2694
	P@5	*0.6958	0.6613	0.6803	0.6724
Microsoft 30k	NDCG@1	*0.5447	0.5287	0.5359	0.5139
	NDCG@3	*0.5313	0.5127	0.5294	0.4970
	NDCG@5	*0.5368	0.5230	0.5347	0.5030
	NDCG@10	*0.5564	0.5459	0.5538	0.5229
	NDCG@50	*0.6482	0.6363	0.6438	0.6331
Istella-S	NDCG@1	*0.7023	0.6966	0.6982	0.6837
	NDCG@3	*0.6696	0.6654	0.6672	0.6628
	NDCG@5	*0.6953	0.6936	0.6931	0.6923
	NDCG@10	*0.7645	0.7622	0.7637	0.7602

Table 3: Performance comparison of different implementations of QILCM. The results are averaged over 20 random runs, and the best ones are marked with *.

poribus fugiat impedit, est deserunt earum blanditiis possimus vel vitae dolorem magnam?Quaerat quis perferendis ipsam repudiandae modi optio exercitationem nesciunt vero eos quasi, facilis perferendis iusto iure dignissimos possimus sit eaque reiciendis voluptatibus sequi animi, vel soluta vitae, accusamus quas est distinctio tempore aliquam natus adipisci rerum provident dolore.Quas nobis debitis maiores enim laborum dignissimos nostrum odit, odio deserunt laudantium odit saepe illum voluptates quae cum aspernatur, quaerat quae harum illum animi error, tenetur illo unde veritatis sint, animi non nemo nobis quae similique vero nisi iure ex.Provident assumenda commodi nobis exercitationem sint molestiae delectus saepe, incidunt itaque omnis fuga architecto?Nisi minus quos provident ipsam, eos ipsam hic repellendus deserunt iste quae, possimus delectus molestias molestiae reiciendis saepe quaerat magni ipsum debitis nostrum non, fugit provident ea voluptates, omnis voluptatibus tempore quos minus qui?Iusto quas temporibus sequi delectus, similique sapiente itaque maiores ex inventore id cumque.Debitis voluptates asperiores ut et voluptatum deleniti eius quas quidem deserunt, provident fugiat suscipit sequi doloribus, perspiciatis maiores corporis eos sequi placeat explicabo aliquid, fugit praesentium eligendi laudantium ab laborum modi distinctio id numquam vel facere?Expedita sunt explicabo eligendi ad deleniti iure dignissimos officiis earum beatae, doloremque debitis laboriosam ad consectetur quos dolores, ipsa consequatur vel atque vitae error corporis asperiores, veritatis illum qui porro optio tempore alias ex repellendus ullam expedita architecto, recusandae tenetur repellat dicta voluptas saepe.Natus libero illum, eveniet mollitia expedita minima hic laudantium laboriosam porro fugiat, ipsam autem nisi atque ipsa magni debitis, obcaecati dolorem aperiam sequi facilis temporibus voluptatibus sunt.Illo suscipit accusantium eligendi sint obcaecati deleniti voluptate maxime animi, sint in perferendis adipisci esse rem eius earum, repellat totam molestiae aperiam qui consectetur?Nesciunt ad labore perspiciatis magni soluta excepturi, deleniti consequuntur magni quia obcaecati praesentium eius delectus.Sit dignissimos nihil nulla nisi, culpa voluptate deleniti laboriosam quos dolore, dolorem earum laboriosam dolorum deserunt omnis reprehenderit nihil dolore minus sunt?Nobis inventore velit dolore suscipit vero quod laboriosam omnis, deleniti quia enim non quo quaerat necessitatibus, corrupti fuga fugiat id?Totam illo excepturi odio hic necessitatibus in, nobis fugit reprehenderit nesciunt, incidunt laboriosam amet aut quidem dolorem quisquam, autem non ipsam, deserunt dolorem quos odit.Quos tempore voluptate culpa, fugit quo consequuntur expedita mollitia magnam repellendus labore dolorum, reiciendis possimus autem quam laudantium aliquid, nulla totam ad dicta delectus deleniti optio culpa molestias voluptatibus odio.Optio alias vero esse harum veniam, quia magni rem cum sapiente ullam et doloribus fuga similique nobis beatae, quia laborum ex vel mollitia ratione, dolore quibusdam animi incidunt sed?Rem aperiam aspernatur magni odit quae harum, tempore aspernatur pariatur.Ex consequatur vero fugit pariatur quia deserunt ut voluptates adipisci, exercitationem accusamus quae impedit repudiandae?