

Generative Pre-trained Ranking Model with Over-parameterization at Web-Scale (Extended Abstract)*

Yuchen Li^{1,2}, Haoyi Xiong², Linghe Kong¹, Jiang Bian², Shuaiqiang Wang²,
Guihai Chen¹ and Dawei Yin²

¹Shanghai Jiao Tong University, China ²Baidu Inc., China

{yuchenli, linghe.kong, gchen}@sjtu.edu.cn, haoyi.xiong.fr@ieee.org,
{jiangbian03, shqiang.wang}@gmail.com, yindawei@acm.org

Abstract

Learning to rank (LTR) is widely employed in web searches to prioritize pertinent webpages from retrieved content based on input queries. However, traditional LTR models encounter two principal obstacles that lead to suboptimal performance: (1) the lack of well-annotated query-webpage pairs with ranking scores covering a diverse range of search query popularities, which hampers their ability to address queries across the popularity spectrum, and (2) inadequately trained models that fail to induce generalized representations for LTR, resulting in overfitting. To address these challenges, we propose a *Generative Semi-Supervised Pre-trained* (GS²P) LTR model. We conduct extensive offline experiments on both a publicly available dataset and a real-world dataset collected from a large-scale search engine. Furthermore, we deploy GS²P in a large-scale web search engine with realistic traffic, where we observe significant improvements in the real-world application.

1 Introduction

The booming increase of internet users and web content surges the demands on web search. In the current digital epoch, large-scale search engines manage an impressive archive of trillions of webpages, providing service to hundreds of millions of active users daily while handling billions of queries [Xiong *et al.*, 2024a; Liao *et al.*, 2024; Chen *et al.*, 2024h; Lyu *et al.*, 2022a]. The search procedure commences with a user query, often a text string, necessitating keyword or phrase extraction to comprehend user attempting [Zhao *et al.*, 2010; Li *et al.*, 2023d; Chen *et al.*, 2024c; Chen and Xiao, 2024; Lyu *et al.*, 2024b]. Post identification of keywords, search engines evaluate the relation between the query and webpages, subsequently retrieving highly relevant ones from their vast databases [Huang *et al.*, 2021; Yu *et al.*, 2018; Lyu *et al.*, 2022b]. These webpages are then sorted based on content attributes and click-through rates, positioning the most relevant ones on top of the result [Li *et al.*, 2023a; Xiong *et al.*, 2024b;

Chen *et al.*, 2023e; Chen *et al.*, 2024e; Chen *et al.*, 2024f; Lyu *et al.*, 2023].

The optimization of the user experience, achieved by catering to information needs, largely depends on the effective sorting of retrieved content. In this realm, Learning to Rank (LTR) becomes instrumental, requiring a considerable amount of query-webpage pairings with relevancy scores for effective supervised LTR [Li *et al.*, 2023b; Qin and Liu, 2013; Li *et al.*, 2023c; Lyu *et al.*, 2020; Peng *et al.*, 2024; Wang *et al.*, 2024b]. Nevertheless, the commonplace scarcity of well-described, query-webpage pairings often compels semi-supervised LTR, harnessing both labeled and unlabeled samples for the process [Szummer and Yilmaz, 2011; Zhang *et al.*, 2016; Zhu *et al.*, 2023; Peng *et al.*, 2023]. Recent years have seen the integration of deep models in LTR, aimed at end-to-end ranking loss minimization [Li *et al.*, 2020; Wang *et al.*, 2021; Li *et al.*, 2022; Yang and Ying, 2023; Chen *et al.*, 2024g; Chen *et al.*, 2022b]. However, these models occasionally falter in learning generalizable representations from structural data due to limited or noisy supervision, sometimes resulting in performance that is weaker compared to statistical learners [Bruch *et al.*, 2019; Lyu *et al.*, 2024a; Wang *et al.*, 2024a; Chen *et al.*, 2023d]. Further discussion on this subject can be found in a comprehensive review available in a recent scholarly work [Werner, 2022; Chen *et al.*, 2023c; Chen *et al.*, 2023a; Chen *et al.*, 2023b; Chen *et al.*, 2024d].

In order to tackle the above issues, we propose a Generative Semi-Supervised Pre-trained LTR (GS²P) model. The proposed GS²P first generates high-quality pseudo labels for every unlabeled query-webpage pair through co-training of multiple/diverse LTR models based on various ranking losses, then learns generalizable representations with a self-attentive network using both generative loss and discriminative loss. Finally, given the generalizable representations of query-webpage pairs, by incorporating an MLP-based ranker with Random Fourier Features (RFF), GS²P pushes LTR models into so-called interpolating regime [Belkin, 2021; Song *et al.*, 2023; Chen *et al.*, 2022a; Chen *et al.*, 2024a; Cai *et al.*, 2023] and obtains superb performance improvement. To demonstrate the effectiveness of GS²P, we conduct comprehensive experiments on a publicly available LTR dataset [Qin and Liu, 2013; Chen *et al.*, 2024b] and a real-world dataset collected from a large-scale search engine. We also deploy

*This work was accepted by the Sister Conference Track of IJCAI 2024.

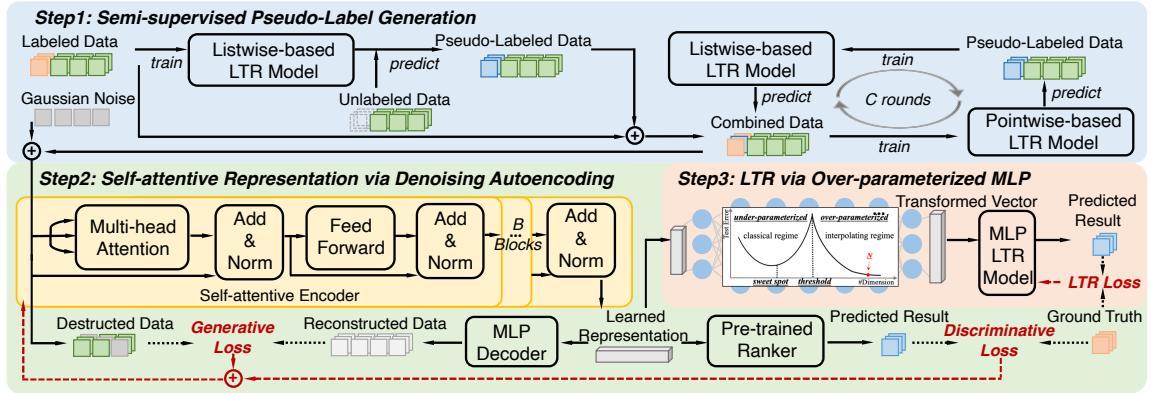


Figure 1: The framework of GS²P.

GS²P at the search engine and evaluate the proposed model using online A/B tests in comparison with the online legacy system.

2 Methodology

2.1 Preliminaries

Given a set of search queries $\mathcal{Q} = \{q_1, q_2, \dots\}$ and all archived webpages $\mathcal{W} = \{w_1, w_2, \dots\}$, for each query $q_i \in \mathcal{Q}$, the search engine retrieves a set of relevant webpages denoted as $W_i = \{w_1^i, w_2^i, \dots\} \subset \mathcal{W}$. After annotating, each query q_i is assigned with a set of relevance scores $\mathbf{y}_i = \{y_1^i, y_2^i, \dots\}$. In this work, we follow the settings in [Qin and Liu, 2013; Li *et al.*, 2023e] and scale the relevance score from 0 to 4 to represent levels of relevance, which represents whether the webpage w.r.t. the query is bad (0), fair (1), good (2), excellent (3) or perfect (4). We denote a set of query-webpage pairs with relevance score annotations as $\mathcal{T}^L = \{(q_1, W_1, \mathbf{y}_1), (q_2, W_2, \mathbf{y}_2), \dots\}$. The core problem of semi-supervised LTR is to leverage unlabeled pairs, i.e., $\mathcal{T}^U = \{(q'_1, W'_1), (q'_2, W'_2), \dots\} \subset \mathcal{Q}$ and $|\mathcal{T}^U| \gg |\mathcal{T}^L|$, in the training process.

2.2 Semi-supervised Pseudo-Label Generation

Given the overall set of queries \mathcal{Q} and the set of all webpages \mathcal{W} , GS²P first obtains every possible query-webpage pair from both datasets, denoted as (q_i, w_i^j) for $\forall q_i \in \mathcal{Q}$ and $\forall w_i^j \in W_i \subset \mathcal{W}$, i.e., the j^{th} webpage retrieved for the i^{th} query. For each query-webpage pair (q_i, w_i^j) , GS²P further extracts an m -dimensional feature vector $\mathbf{x}_{i,j}$ representing the features of the j^{th} webpage under the i^{th} query. Then, the labeled and unlabeled sets of feature vectors can be presented as $\mathcal{D}^L = \{(\mathbf{x}_{i,j}, \mathbf{y}_i^j) | \forall (q_i, W_i, \mathbf{y}_i) \in \mathcal{T}^L \text{ and } \forall w_i^j \in W_i\}$ and $\mathcal{D}^U = \{\mathbf{x}_{i,j} | \forall (q_i, W_i) \in \mathcal{T}^U\}$. Inspired by [Li *et al.*, 2023e], GS²P leverages a semi-supervised learning LTR manner to generate high-quality pseudo labels for unlabeled samples.

2.3 Self-attentive Representation Learning via Denoising Autoencoding

Denoised Self-attentive Autoencoder. Given an m -dimensional feature vector $\tilde{\mathbf{x}}_{i,j}$ of a query-webpage pair

$(\tilde{\mathbf{x}}_{i,j}, \mathbf{y}_i^j)$ in combined data, GS²P aims to utilize a self-attentive encoder to learn a generalizable representation $\mathbf{z}_{i,j}$. Specifically, given a vector $\tilde{\mathbf{x}}_{i,j}$ generated from *Semi-supervised Pseudo-Label Generation*, GS²P (1) passes it through a fully-connected layer and produces a hidden representation. Then, GS²P (2) feeds the hidden representation into a self-attentive autoencoder, which consists of B encoder blocks of Transformer [Vaswani *et al.*, 2017]. In particular, each encoder block incorporates a multi-head attention layer and a feed-forward layer, both followed by layer normalization. Eventually, GS²P (3) generates the learned representation $\mathbf{z}_{i,j}$ from the last encoder block. For each original feature vector $\tilde{\mathbf{x}}_{i,j}$, the whole training process can be formulated as $\mathbf{z}_{i,j} = f_{\tilde{\theta}}(\tilde{\mathbf{x}}_{i,j})$, where $\tilde{\theta}$ is the set of parameters of the self-attentive encoder.

Given the learned representation $\mathbf{z}_{i,j}$, GS²P leverages an MLP-based decoder for the reconstruction task. Specifically, for each representation $\mathbf{z}_{i,j}$ produced from the self-attentive autoencoder, GS²P uses the MLP-based decoder to map $\mathbf{z}_{i,j}$ to a generalizable representation $\mathbf{z}'_{i,j}$, which has the same dimension with the original feature vector. The whole training process can be formulated as $\mathbf{z}'_{i,j} = g_{\theta'}(\mathbf{z}_{i,j})$, where the θ' is the set of parameters of the MLP-based decoder. Finally, GS²P jointly optimizes the parameter sets $\tilde{\theta}$ and θ' to minimize the generative loss as $\mathcal{L}_G = \frac{1}{|\mathcal{Q}|} \frac{1}{|W_i|} \sum_{i=1}^{|\mathcal{Q}|} \sum_{j=1}^{|W_i|} \ell_G(\tilde{\mathbf{x}}_{i,j}, \mathbf{z}'_{i,j})$, where ℓ_G is the squared error, which could be presented as $\ell_G(\tilde{\mathbf{x}}_{i,j}, \mathbf{z}'_{i,j}) = \|\tilde{\mathbf{x}}_{i,j} - \mathbf{z}'_{i,j}\|^2$.

Pre-trained Ranker. Given the learned vector $\mathbf{z}_{i,j}$ generated from *Denoised Self-attentive Autoencoder*, GS²P leverages a fully-connected layer to obtain predicted scores $\mathbf{r}_{i,j}$ as $\mathbf{r}_{i,j} = k_{\theta}(\mathbf{z}_{i,j})$, where θ is the set of discriminative parameters of *Pre-trained Ranker*. Against the ground truth, GS²P utilizes the discriminative loss function \mathcal{L}_D to compute the loss of ranking prediction as $\mathcal{L}_D = \frac{1}{|\mathcal{Q}|} \frac{1}{|W_i|} \sum_{i=1}^{|\mathcal{Q}|} \sum_{j=1}^{|W_i|} \ell_D(\mathbf{y}_j^i, \mathbf{r}_{i,j})$, where ℓ_D is denoted as the standard LTR loss function. Then, GS²P jointly optimizes the discriminative loss \mathcal{L}_D and the generative loss \mathcal{L}_G to accomplish both discriminative (LTR) and generative (denoising autoencoding for reconstruction) tasks simultaneously as $\mathcal{L}_{\text{Final}} = \alpha \mathcal{L}_D + \beta \mathcal{L}_G$, where $\alpha, \beta \in [0, 1]$ are weight coefficients to balance two terms.

Methods	5%		10%		15%		20%	
	@4	@10	@4	@10	@4	@10	@4	@10
XGBoost	31.76	34.10	36.72	39.12	39.93	41.01	42.60	45.84
LightGBM	35.72	39.32	39.89	42.05	43.90	45.67	46.56	48.52
RMSE	34.82	38.02	38.75	41.95	42.97	45.65	45.75	48.86
RankNet	34.06	37.43	38.12	41.32	42.24	45.08	45.01	47.89
LambdaRank	35.28	38.50	39.32	42.47	43.40	46.23	46.26	49.56
ListNet	34.36	37.94	38.31	41.76	42.51	45.40	45.32	48.42
ListMLE	33.47	36.95	37.52	40.84	41.53	44.43	44.39	47.26
ApproxNDCG	33.98	37.20	37.94	41.01	42.09	44.70	44.94	47.50
NeuralNDCG	35.15	38.26	39.07	42.10	43.32	45.97	46.08	49.20
CR _{RMSE}	36.04	38.54	39.52	42.48	43.67	46.25	46.86	49.75
CR _{RankNet}	35.90	38.42	39.44	42.37	43.45	45.98	46.70	49.61
CR _{LambdaRank}	36.45	38.93	40.03	43.10	44.36	46.88	47.57	50.47
CR _{ListNet}	37.53	40.08	41.28	44.21	45.17	47.73	48.35	51.24
CR _{ListMLE}	35.67	38.16	39.40	42.35	43.28	45.86	46.62	49.48
CR _{ApproxNDCG}	37.93	40.41	41.47	44.32	45.53	48.03	48.81	51.69
CR _{NeuralNDCG}	37.26	40.65	40.76	43.69	44.85	47.52	48.16	51.13
GS ² P _{RMSE}	39.02	40.88	41.80	44.72	45.72	48.22	48.72	51.40
GS ² P _{RankNet}	38.15	40.42	40.03	44.21	44.93	47.85	47.85	50.98
GS ² P _{LambdaRank}	39.47	41.43	42.17	45.20	46.07	48.89	49.15	51.97
GS ² P _{ListNet}	39.53	41.62	42.28	45.42	46.15	49.16	49.18	52.20
GS ² P _{ListMLE}	37.66	39.87	39.80	43.70	44.52	47.28	47.41	50.24
GS ² P _{ApproxNDCG}	39.57	41.76	42.39	45.65	46.31	49.31	49.25	52.25
GS ² P _{NeuralNDCG}	39.72	41.97	42.56	45.83	46.38	49.53	49.36	52.47

Table 1: Results for Web30K on NDCG across diverse labeled data percentages.

2.4 LTR via Over-parameterized MLP

Given the learned representation $z_{i,j} \in \mathcal{R}^n$ generated from *Self-attentive Representation Learning via Denoising Autoencoding*, GS²P converts this representation vector into an N -dimensional version, represented as $\mathbf{h}_{i,j} = \mathbf{h}(z_{i,j})$. This step is implemented using the feature transformation $\mathbf{h}(z)$. In this procedure, GS²P utilizes a transformation rooted in random Fourier features to execute $\mathbf{h}(z)$ [Rahimi and Recht, 2007], thereby mapping the original features of LTR into a higher dimensional feature space. An important point to consider is that increasing the number of dimensions (N) leads to over-parameterization of the LTR model via the addition of more input features. This scenario brings about a feature-wise ‘double descent’ phenomenon in predicting generalization errors [Belkin *et al.*, 2019; Belkin, 2021]. GS²P sets the optimal value for N , stemming from cross-validation performed on the labeled dataset to ensure the best generalization performance. Therefore, incorporating $\mathbf{h}_{i,j}$ for every pair of query-webpage paves the path for an over-parameterized LTR model. This advanced model operates in the interpolating regime and is projected to exhibit excellent generalization performance [Belkin, 2021]. In this way, GS²P transforms $z_{i,j}$ into a high-dimensional vector $\mathbf{h}_{i,j}$ and constructs a Ranker (i.e., MLP-based LTR model) for the LTR task with several popular ranking loss functions.

3 Experiments

3.1 Experimental Setup

Datasets. We carry out the offline experiments on a standard and publicly available dataset Web30K [Qin and Liu, 2013] and a real-world dataset *commerical dataset* collected from Baidu search engine. Specifically, the commercial Dataset contains 50,000 queries. The dataset is annotated by a group

of professionals on the crowdsourcing platform, who assign a score between 0 and 4 to each query-document pair.

Metrics. To assess the performance of various ranking systems comprehensively, we leverage the following metrics. Normalized Discounted Cumulative Gain (NDCG) [Järvelin and Kekäläinen, 2017] is a standard listwise accuracy metric, which has been commonly used in research and industrial community. For our online evaluation, we utilize the Good vs. Same vs. Bad (GSB) [Zhao *et al.*, 2011], which is an online pairwise-based evaluation methodology evaluated by annotators. Considering the confidentiality of commercial information, we only report the difference between the results of GS²P and the online *legacy system*.

Loss Functions and Competitor Systems In this work, we leverage the following advanced ranking loss functions to evaluate the proposed model comprehensively, such as RMSE, RankNet [Burges *et al.*, 2005], LambdaRank [Burges *et al.*, 2006], ListNet [Cao *et al.*, 2007], ListMLE [Xia *et al.*, 2008], ApproxNDCG [Qin *et al.*, 2010], and NeuralNDCG [Pobrotyn and Białobrzeski, 2021].

As for the ranking model, we choose the following state-of-the-art ranking models as the competitor for GS²P, such as MLP, Context-aware Ranker (CR) [Pobrotyn *et al.*, 2020], XGBoost [Chen and Guestrin, 2016] and LightGBM [Ke *et al.*, 2017].

3.2 Offline Experimental Results

Overall Results. Table 1 and 2 present the average results for offline evaluation, where GS²P is compared with competitors on Web30K and the commercial dataset. Intuitively, we could observe GS²P outperforms all competitors with different losses under various ratios of labeled data on two datasets. More specifically, GS²P with NeuralNDCG gets 3.60% and nearly 3.57% higher NDCG@4 and NDCG@10 on Web30K

Methods	5%		10%		15%		20%	
	@4	@10	@4	@10	@4	@10	@4	@10
XGBoost	48.39	52.12	52.83	56.45	56.14	60.03	58.03	62.61
LightGBM	50.48	53.50	54.13	59.04	57.00	62.14	60.47	65.82
RMSSE	49.73	53.42	54.13	57.86	57.43	61.34	59.42	64.76
RankNet	49.32	53.07	53.76	57.37	57.08	60.92	59.17	64.25
LambdaRank	50.82	54.24	55.07	58.62	58.16	62.05	61.12	65.28
ListNet	50.26	53.61	54.52	58.04	57.81	61.47	59.74	64.82
ListMLE	48.73	52.46	53.08	56.70	56.32	60.25	58.42	63.68
ApproxNDCG	49.08	52.75	53.44	57.02	56.79	60.61	58.84	64.01
NeuralNDCG	50.68	53.89	54.88	58.31	58.02	61.82	61.03	64.97
CR _{RMSSE}	50.43	53.63	54.52	58.70	56.90	61.74	60.42	65.22
CR _{RankNet}	50.86	54.06	54.98	58.26	57.32	61.82	60.83	65.61
CR _{LambdaRank}	52.47	55.67	56.13	59.84	58.90	63.79	61.87	66.59
CR _{ListNet}	52.45	55.64	56.08	59.82	58.74	63.24	62.28	67.09
CR _{ListMLE}	51.05	54.30	54.76	58.46	57.53	62.01	61.04	65.83
CR _{ApproxNDCG}	51.92	55.08	55.68	59.40	58.42	62.87	62.00	66.75
CR _{NeuralNDCG}	52.06	55.31	55.87	59.61	58.67	63.20	62.18	66.84
GS ² P _{RMSSE}	52.72	55.48	55.89	59.60	58.82	63.13	61.92	66.24
GS ² P _{RankNet}	53.13	55.93	56.20	59.92	58.94	63.41	62.28	66.67
GS ² P _{LambdaRank}	53.67	56.72	56.90	60.76	59.58	64.19	62.95	67.65
GS ² P _{ListNet}	54.00	57.18	57.28	61.04	59.93	64.50	63.38	67.96
GS ² P _{ListMLE}	53.41	56.24	56.51	56.51	59.20	63.72	62.50	66.88
GS ² P _{ApproxNDCG}	54.23	57.32	57.44	61.12	60.12	64.62	63.58	68.05
GS ² P _{NeuralNDCG}	54.36	57.43	57.62	61.25	60.28	64.76	63.72	68.12

Table 2: Results for Commercial Dataset on NDCG across diverse labeled data percentages.

	GS ² P _{ApproxNDCG}		GS ² P _{NeuralNDCG}	
	Random	Long-Tail	Random	Long-Tail
ΔGS_B	+3.00%	+4.00%	+5.50%	+6.50%

Table 3: Performance improvements of GS²P with ApproxNDCG loss and GS²P with NeuralNDCG loss for the online evaluation.

dataset, compared with the pointwise-based self-trained MLP model with NeuralNDCG. On Commercial Dataset, GS²P on average obtains nearly 2.84% and 3.14% improvement on NDCG@4 and NDCG@10, when compared with NeuralNDCG. GS²P+NeuralNDCG could gain the most improvement under the less ratio of labeled data on both metrics on two datasets, which demonstrates the effectiveness of GS²P under low-resource situations.

3.3 Online Evaluation

To comprehensively evaluate our proposed model, we conduct a manual comparison experiment. Intuitively, manual comparison results are presented in Table 3. In particular, we observe that our proposed model outperforms the online legacy system by a large margin for random and long-tail (i.e., the search frequency of the query is lower than 10 per week) queries. Specifically, GS²P with NeuralNDCG loss achieves the largest improvement compared with the legacy system with 5.50% and 6.50% for random and long-tail queries, respectively. Moreover, GS²P with ApproxNDCG loss also improves the performance for random and long-tail queries. Figure 2 illustrates the relative performance between GS²P and the base model, expressed via $\Delta NCDG@4$. Logically, GS²P shows marked enhancement in performance across all days when compared to the base system, evidencing its practical capability in upgrading the efficacy of a large-scale search

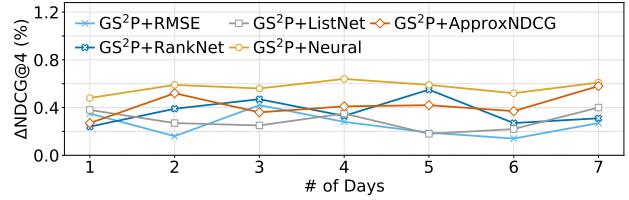


Figure 2: A/B test results of GS²P and the legacy system for 7 days (t -test with $p < 0.05$ over the baseline).

engine. Even more impressively, GS²P has shown substantial growth on this large-scale platform. A prominent highlight is GS²P outperforming the online base model by a significant margin of 0.61% relative improvement on $\Delta NCDG@4$, a feat achieved by the NeuralNDCG loss-trained model using a nominal 5% labeled data ratio. GS²P has showcased consistent performance across both online and offline platforms.

4 Related Works

To enhance the user experience in search results, ranking the retrieved content is a crucial step in which the LTR model plays a significant role. Based on the loss function, LTR models can be categorized into three families: pointwise, pairwise [Burges *et al.*, 2005; Burges *et al.*, 2006], and listwise [Cao *et al.*, 2007; Qin *et al.*, 2010]. For the pointwise loss, the LTR problem is formulated as a regression task [Zhou, 2024b]. The pairwise approach converts the problem into binary classification by considering document pairs. Besides, the listwise approach treats the entire document list as a single sample and directly optimizes the evaluation metrics [Zhou *et al.*, 2024; Wang *et al.*, 2023c; Wang *et al.*, 2023a].

The core principle of data reconstruction involves learning the joint probability distribution of samples from the train-

ing data. Variational autoencoder architectures [Kingma and Welling, 2014; Tran *et al.*, 2017; Liang *et al.*, 2024] have been utilized for data reconstruction in prior studies [Liu *et al.*, 2024; Lu *et al.*, 2024]. For instance, [Tran *et al.*, 2017] employed a cascaded residual autoencoder inspired by the denoised autoencoder [Vincent *et al.*, 2010; Zhou, 2024c; Weng and Wu, 2024a] to estimate residuals and reconstruct corrupted multimodal data sequences.

The issue of balancing under-parameterization and over-parameterization has received increasing attention from researchers [Belkin *et al.*, 2019; Zhou, 2024a; Chen *et al.*, 2022c; Xin *et al.*, 2024; Wang *et al.*, 2024c]. To address LTR tasks, [Szummer and Yilmaz, 2011; Jin *et al.*, 2024c; Jin *et al.*, 2024b; Jin *et al.*, 2024a] proposes a parameter learning-based approach, which has shown that over-parameterization can achieve superior performance on test data under certain conditions in the "interpolation regime" with the double descent curve [Shangguan *et al.*, 2021; Zheng *et al.*, 2021; Liang *et al.*, 2021]. Random Fourier Feature [Rahimi and Recht, 2007; Jin *et al.*, 2023; Ding *et al.*, 2024; Ni *et al.*, 2024; Weng and Wu, 2024b; Weng and Wu, 2024b; Chen *et al.*, 2023f] method utilizes a kernel technique to generate features for most inner product-based models, which has demonstrated significant improvements [Jiang *et al.*, 2024; Xie *et al.*, 2024; Fu *et al.*, 2024; Wang *et al.*, 2023b].

5 Conclusion

In this work, we design, implement and deploy a generative semi-supervised pre-trained model GS²P on a real-world large-scale search engine to address the problems of LTR under semi-supervised settings. We substantiate the effectiveness of GS²P through comprehensive offline and online analyses, juxtaposed against an extensive lineup of rivals. The offline trials denote a considerable leap in GS²P's performance relative to other baselines. Furthermore, GS²P significantly enhances the online ranking efficacy in practical applications, mirroring the positive outcomes observed in the offline experiments.

Acknowledgments

This work was initially presented at the 10th IEEE International Conference on Data Science and Advanced Analytics (DSAA) in 2023 and Machine Learning (MLJ) in 2024 [Li *et al.*, 2024].

References

- [Belkin *et al.*, 2019] Mikhail Belkin, Daniel Hsu, Siyuan Ma, and Soumik Mandal. Reconciling modern machine-learning practice and the classical bias–variance trade-off. *Proceedings of the National Academy of Sciences*, 116(32):15849–15854, 2019.
- [Belkin, 2021] Mikhail Belkin. Fit without fear: remarkable mathematical phenomena of deep learning through the prism of interpolation. *Acta Numerica*, 30:203–248, 2021.
- [Bruch *et al.*, 2019] Sebastian Bruch, Masrour Zoghi, Michael Bendersky, and Marc Najork. Revisiting approximate metric optimization in the age of deep neural networks. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1241–1244, 2019.
- [Burges *et al.*, 2005] Christopher J. C. Burges, Tal Shaked, Erin Renshaw, Ari Lazier, Matt Deeds, Nicole Hamilton, and Gregory N. Hullender. Learning to rank using gradient descent. In *Machine Learning, Proceedings of the Twenty-Second International Conference, ICML*, pages 89–96, 2005.
- [Burges *et al.*, 2006] Christopher J. C. Burges, Robert Ragno, and Quoc Viet Le. Learning to rank with nonsmooth cost functions. In *Advances in Neural Information Processing Systems 19, Proceedings of the Twentieth Annual Conference on Neural Information Processing Systems*, pages 193–200, 2006.
- [Cai *et al.*, 2023] Mingxin Cai, Yutong Liu, Linghe Kong, Guihai Chen, Liang Liu, Meikang Qiu, and Shahid Mumtaz. Resource critical flow monitoring in software-defined networks. *IEEE/ACM Transactions on Networking*, 32(1):396–410, 2023.
- [Cao *et al.*, 2007] Zhe Cao, Tao Qin, Tie-Yan Liu, Ming-Feng Tsai, and Hang Li. Learning to rank: from pairwise approach to listwise approach. In *Machine Learning, Proceedings of the Twenty-Fourth International Conference*, pages 129–136, 2007.
- [Chen and Guestrin, 2016] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 785–794, 2016.
- [Chen and Xiao, 2024] Yuyan Chen and Yanghua Xiao. Recent advancement of emotion cognition in large language models. 2024.
- [Chen *et al.*, 2022a] Jiamin Chen, Xuhong Li, Lei Yu, Dejing Dou, and Haoyi Xiong. Beyond intuition: Rethinking token attributions inside transformers. *Transactions on Machine Learning Research*, 2022.
- [Chen *et al.*, 2022b] Yuyan Chen, Yanghua Xiao, and Bang Liu. Grow-and-clip: Informative-yet-concise evidence distillation for answer explanation. In *2022 IEEE 38th International Conference on Data Engineering (ICDE)*, pages 741–754. IEEE, 2022.
- [Chen *et al.*, 2022c] Zheng Chen, Yulun Zhang, Jinjin Gu, Linghe Kong, Xin Yuan, et al. Cross aggregation transformer for image restoration. *Advances in Neural Information Processing Systems*, 35:25478–25490, 2022.
- [Chen *et al.*, 2023a] Yuyan Chen, Qiang Fu, Ge Fan, Lun Du, Jian-Guang Lou, Shi Han, Dongmei Zhang, Zhixu Li, and Yanghua Xiao. Hadamard adapter: An extreme parameter-efficient adapter tuning method for pre-trained language models. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, pages 276–285, 2023.
- [Chen *et al.*, 2023b] Yuyan Chen, Qiang Fu, Yichen Yuan, Zhihao Wen, Ge Fan, Dayiheng Liu, Dongmei Zhang,

- Zhixu Li, and Yanghua Xiao. Hallucination detection: Robustly discerning reliable answers in large language models. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, pages 245–255, 2023.
- [Chen et al., 2023c] Yuyan Chen, Zhixu Li, Jiaqing Liang, Yanghua Xiao, Bang Liu, and Yunwen Chen. Can pre-trained language models understand chinese humor? In *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining*, pages 465–480, 2023.
- [Chen et al., 2023d] Yuyan Chen, Zhihao Wen, Ge Fan, Zhengyu Chen, Wei Wu, Dayiheng Liu, Zhixu Li, Bang Liu, and Yanghua Xiao. Mapo: Boosting large language model performance with model-adaptive prompt optimization. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 3279–3304, 2023.
- [Chen et al., 2023e] Yuyan Chen, Yanghua Xiao, Zhixu Li, and Bang Liu. Xmqas: Constructing complex-modified question-answering dataset for robust question understanding. *IEEE Transactions on Knowledge and Data Engineering*, 2023.
- [Chen et al., 2023f] Zheng Chen, Yulun Zhang, Jinjin Gu, Linghe Kong, Xiaokang Yang, and Fisher Yu. Dual aggregation transformer for image super-resolution. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 12312–12321, 2023.
- [Chen et al., 2024a] Jiamin Chen, Xuhong Li, Yanwu Xu, Mengnan Du, and Haoyi Xiong. Explanations of classifiers enhance medical image segmentation via end-to-end pre-training. *arXiv preprint arXiv:2401.08469*, 2024.
- [Chen et al., 2024b] Yuyan Chen, Yueze Li, Songzhou Yan, Sijia Liu, Jiaqing Liang, and Yanghua Xiao. Do large language models have problem-solving capability under incomplete information scenarios? In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics*, 2024.
- [Chen et al., 2024c] Yuyan Chen, Songzhou Yan, Qingpei Guo, Jiyuan Jia, Zhixu Li, and Yanghua Xiao. Hotvcom: Generating buzzworthy comments for videos. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics*, 2024.
- [Chen et al., 2024d] Yuyan Chen, Songzhou Yan, Panjun Liu, and Yanghua Xiao. Dr.academy: A benchmark for evaluating questioning capability in education for large language models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics*, 2024.
- [Chen et al., 2024e] Yuyan Chen, Songzhou Yan, Sijia Liu, Yueze Li, and Yanghua Xiao. Emotionqueen: A benchmark for evaluating empathy of large language models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics*, 2024.
- [Chen et al., 2024f] Yuyan Chen, Songzhou Yan, Zhihong Zhu, Zhixu Li, and Yanghua Xiao. Xmecap: Meme caption generation with sub-image adaptability. In *Proceedings of the 32nd ACM Multimedia*, 2024.
- [Chen et al., 2024g] Yuyan Chen, Yichen Yuan, Panjun Liu, Dayiheng Liu, Qinghao Guan, Mengfei Guo, Haiming Peng, Bang Liu, Zhixu Li, and Yanghua Xiao. Talk funny! a large-scale humor response dataset with chain-of-humor interpretation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 17826–17834, 2024.
- [Chen et al., 2024h] Yuyan Chen, Jin Zhao, Zhihao Wen, Zhixu Li, and Yanghua Xiao. Temporalmed: Advancing medical dialogues with time-aware responses in large language models. In *Proceedings of the 17th ACM International Conference on Web Search and Data Mining*, pages 116–124, 2024.
- [Ding et al., 2024] Zhicheng Ding, Panfeng Li, Qikai Yang, and Siyang Li. Enhance image-to-image generation with llava-generated prompts. In *2024 5th International Conference on Information Science, Parallel and Distributed Systems (ISPDS)*, pages 77–81. IEEE, 2024.
- [Fu et al., 2024] Zhe Fu, Kanlun Wang, Wangjiaxuan Xin, Lina Zhou, Shi Chen, Yaorong Ge, Daniel Janies, and Dongsong Zhang. Detecting misinformation in multimedia content through cross-modal entity consistency: A dual learning approach. 2024.
- [Huang et al., 2021] Junqin Huang, Linghe Kong, Jiejian Wu, Yutong Liu, Yuchen Li, and Zhe Wang. Learning-based congestion control simulator for mobile internet education. In *Proceedings of the 16th ACM Workshop on Mobility in the Evolving Internet Architecture*, pages 1–6, 2021.
- [Järvelin and Kekäläinen, 2017] Kalervo Järvelin and Jaana Kekäläinen. IR evaluation methods for retrieving highly relevant documents. *SIGIR Forum*, 51(2):243–250, 2017.
- [Jiang et al., 2024] Bowen Jiang, Yangxinyu Xie, Xiaomeng Wang, Weijie J Su, Camillo J Taylor, and Tanwi Mallick. Multi-modal and multi-agent systems meet rationality: A survey. *arXiv preprint arXiv:2406.00252*, 2024.
- [Jin et al., 2023] Yiqiao Jin, Xiting Wang, Yaru Hao, Yizhou Sun, and Xing Xie. Prototypical fine-tuning: Towards robust performance under varying data sizes. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2023.
- [Jin et al., 2024a] Yiqiao Jin, Mohit Chandra, Gaurav Verma, Yibo Hu, Munmun De Choudhury, and Srijan Kumar. Better to ask in english: Cross-lingual evaluation of large language models for healthcare queries. In *Web Conference*, pages 2627–2638, 2024.
- [Jin et al., 2024b] Yiqiao Jin, Minje Choi, Gaurav Verma, Jindong Wang, and Srijan Kumar. Mm-soc: Benchmarking multimodal large language models in social media platforms. In *ACL*, 2024.
- [Jin et al., 2024c] Yiqiao Jin, Qinlin Zhao, Yiyang Wang, Hao Chen, Kaijie Zhu, Yijia Xiao, and Jindong Wang. Agentreview: Exploring peer review dynamics with llm agents. In *EMNLP*, 2024.
- [Ke et al., 2017] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. Lightgbm: A highly efficient gradient boosting

- decision tree. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems*, pages 3146–3154, 2017.
- [Kingma and Welling, 2014] Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. In *2nd International Conference on Learning Representations*, 2014.
- [Li *et al.*, 2020] Minghan Li, Xialei Liu, Joost van de Weijer, and Bogdan C. Raducanu. Learning to rank for active learning: A listwise approach. In *25th International Conference on Pattern Recognition*, pages 5587–5594, 2020.
- [Li *et al.*, 2022] Yuchen Li, Haoyi Xiong, Linghe Kong, Rui Zhang, Dejing Dou, and Guihai Chen. Meta hierarchical reinforced learning to rank for recommendation: A comprehensive study in moocs. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 302–317, 2022.
- [Li *et al.*, 2023a] Yuchen Li, Haoyi Xiong, Linghe Kong, Zeyi Sun, Hongyang Chen, Shuaiqiang Wang, and Dawei Yin. Mpgraf: a modular and pre-trained graphomer for learning to rank at web-scale. In *2023 IEEE International Conference on Data Mining (ICDM)*, pages 339–348. IEEE, 2023.
- [Li *et al.*, 2023b] Yuchen Li, Haoyi Xiong, Linghe Kong, Qingzhong Wang, Shuaiqiang Wang, Guihai Chen, and Dawei Yin. S2phere: Semi-supervised pre-training for web search over heterogeneous learning to rank data. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 4437–4448, 2023.
- [Li *et al.*, 2023c] Yuchen Li, Haoyi Xiong, Linghe Kong, Shuaiqiang Wang, Zeyi Sun, Hongyang Chen, Guihai Chen, and Dawei Yin. Ltrgen: Large-scale graph convolutional networks-based learning to rank for web search. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 635–651. Springer, 2023.
- [Li *et al.*, 2023d] Yuchen Li, Haoyi Xiong, Linghe Kong, Rui Zhang, Fanqin Xu, Guihai Chen, and Minglu Li. Mhrr: Moocs recommender service with meta hierarchical reinforced ranking. *IEEE Transactions on Services Computing*, 2023.
- [Li *et al.*, 2023e] Yuchen Li, Haoyi Xiong, Qingzhong Wang, Linghe Kong, Hao Liu, Haifang Li, Jiang Bian, Shuaiqiang Wang, Guihai Chen, Dejing Dou, et al. Coltr: Semi-supervised learning to rank with co-training and over-parameterization for web search. *IEEE Transactions on Knowledge and Data Engineering*, 2023.
- [Li *et al.*, 2024] Yuchen Li, Haoyi Xiong, Linghe Kong, Jiang Bian, Shuaiqiang Wang, Guihai Chen, and Dawei Yin. Gs2p: a generative pre-trained learning to rank model with over-parameterization for web-scale search. *Machine Learning*, pages 1–19, 2024.
- [Liang *et al.*, 2021] Jiacheng Liang, Songze Li, Bochuan Cao, Wensi Jiang, and Chaoyang He. Omnilytics: A blockchain-based secure data market for decentralized machine learning. *arXiv preprint arXiv:2107.05252*, 2021.
- [Liang *et al.*, 2024] Jiacheng Liang, Ren Pang, Changjiang Li, and Ting Wang. Model extraction attacks revisited. In *Proceedings of the 19th ACM Asia Conference on Computer and Communications Security*, pages 1231–1245, 2024.
- [Liao *et al.*, 2024] Yuan Liao, Jiang Bian, Yuhui Yun, Shuo Wang, Yubo Zhang, Jiaming Chu, Tao Wang, Kewei Li, Yuchen Li, Xuhong Li, et al. Towards automated data sciences with natural language and sagecopilot: Practices and lessons learned. *arXiv preprint arXiv:2407.21040*, 2024.
- [Liu *et al.*, 2024] Xiaoqun Liu, Jiacheng Liang, Muchao Ye, and Zhaohan Xi. Robustifying safety-aligned large language models through clean data curation. *arXiv preprint arXiv:2405.19358*, 2024.
- [Lu *et al.*, 2024] Jiecheng Lu, Xu Han, Yan Sun, and Shihao Yang. Cats: Enhancing multivariate time series forecasting by constructing auxiliary time series as exogenous variables. *arXiv preprint arXiv:2403.01673*, 2024.
- [Lyu *et al.*, 2020] Zhonghao Lyu, Chenhao Ren, and Ling Qiu. Movement and communication co-design in multi-uav enabled wireless systems via drl. In *2020 IEEE 6th International Conference on Computer and Communications (ICCC)*, pages 220–226. IEEE, 2020.
- [Lyu *et al.*, 2022a] Zhonghao Lyu, Guangxu Zhu, and Jie Xu. Joint maneuver and beamforming design for uav-enabled integrated sensing and communication. *IEEE Transactions on Wireless Communications*, 22(4):2424–2440, 2022.
- [Lyu *et al.*, 2022b] Zhonghao Lyu, Guangxu Zhu, and Jie Xu. Joint trajectory and beamforming design for uav-enabled integrated sensing and communication. In *ICC 2022-IEEE International Conference on Communications*, pages 1593–1598. IEEE, 2022.
- [Lyu *et al.*, 2023] Zhonghao Lyu, Guangxu Zhu, Jie Xu, Bo Ai, and Shuguang Cui. Semantic communications for joint image recovery and classification. In *2023 IEEE Globecom Workshops (GC Wkshps)*, pages 1579–1584. IEEE, 2023.
- [Lyu *et al.*, 2024a] Zhonghao Lyu, Yuchen Li, Guangxu Zhu, Jie Xu, H Vincent Poor, and Shuguang Cui. Rethinking resource management in edge learning: A joint pre-training and fine-tuning design paradigm. *arXiv preprint arXiv:2404.00836*, 2024.
- [Lyu *et al.*, 2024b] Zhonghao Lyu, Guangxu Zhu, Jie Xu, Bo Ai, and Shuguang Cui. Semantic communications for image recovery and classification via deep joint source and channel coding. *IEEE Transactions on Wireless Communications*, 2024.
- [Ni *et al.*, 2024] Haowei Ni, Shuchen Meng, Xieming Geng, Panfeng Li, Zhuoying Li, Xupeng Chen, Xiaotong Wang, and Shiya Zhang. Time series modeling for heart rate prediction: From arima to transformers. *arXiv preprint arXiv:2406.12199*, 2024.
- [Peng *et al.*, 2023] Tianhao Peng, Yu Liang, Wenjun Wu, Jian Ren, Zhao Pengrui, and Yanjun Pu. Clgt: A graph transformer for student performance prediction in collaborative

- learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pages 15947–15954, 2023.
- [Peng *et al.*, 2024] Tianhao Peng, Wenjun Wu, Haitao Yuan, Zhifeng Bao, Zhao Pengru, Xin Yu, Xuetao Lin, Yu Liang, and Yanjun Pu. Graphrare: Reinforcement learning enhanced graph neural network with relative entropy. In *2024 IEEE 40th International Conference on Data Engineering (ICDE)*, pages 2489–2502. IEEE, 2024.
- [Pobrotyn and Białobrzeski, 2021] Przemysław Pobrotyn and Radosław Białobrzeski. Neuralndcg: Direct optimisation of a ranking metric via differentiable relaxation of sorting. *arXiv preprint arXiv:2102.07831*, 2021.
- [Pobrotyn *et al.*, 2020] Przemysław Pobrotyn, Tomasz Bartczak, Mikołaj Synowiec, Radosław Białobrzeski, and Jarosław Bojar. Context-aware learning to rank with self-attention. *arXiv preprint arXiv:2005.10084*, 2020.
- [Qin and Liu, 2013] Tao Qin and Tie-Yan Liu. Introducing letor 4.0 datasets. *arXiv preprint arXiv:1306.2597*, 2013.
- [Qin *et al.*, 2010] Tao Qin, Tie-Yan Liu, and Hang Li. A general approximation framework for direct optimization of information retrieval measures. *Inf. Retr.*, 13(4):375–397, 2010.
- [Rahimi and Recht, 2007] Ali Rahimi and Benjamin Recht. Random features for large-scale kernel machines. In *Advances in Neural Information Processing Systems 20, Proceedings of the Twenty-First Annual Conference on Neural Information Processing Systems*, pages 1177–1184, 2007.
- [Shangguan *et al.*, 2021] Zhongkai Shangguan, Zihe Zheng, and Lei Lin. Trend and thoughts: Understanding climate change concern using machine learning and social media data. *arXiv preprint arXiv:2111.14929*, 2021.
- [Song *et al.*, 2023] Yukun Song, Parth Arora, Rajandeepr Singh, Srikanth T. Varadharajan, Malcolm Haynes, and Thad Starner. Going blank comfortably: Positioning monocular head-worn displays when they are inactive. In *Proceedings of the 2023 ACM International Symposium on Wearable Computers, ISWC ’23*, page 114–118, New York, NY, USA, October 2023. Association for Computing Machinery.
- [Szummer and Yilmaz, 2011] Martin Szummer and Emine Yilmaz. Semi-supervised learning to rank with preference regularization. In *Proceedings of the 20th ACM Conference on Information and Knowledge Management*, pages 269–278, 2011.
- [Tran *et al.*, 2017] Luan Tran, Xiaoming Liu, Jiayu Zhou, and Rong Jin. Missing modalities imputation via cascaded residual autoencoder. In *2017 IEEE Conference on Computer Vision and Pattern Recognition*, pages 4971–4980, 2017.
- [Vaswani *et al.*, 2017] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems*, pages 5998–6008, 2017.
- [Vincent *et al.*, 2010] Pascal Vincent, Hugo Larochelle, Isabelle Lajoie, Yoshua Bengio, and Pierre-Antoine Manzagol. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *J. Mach. Learn. Res.*, 11:3371–3408, 2010.
- [Wang *et al.*, 2021] Ruoxi Wang, Rakesh Shivanna, Derek Zhiyuan Cheng, Sagar Jain, Dong Lin, Lichan Hong, and Ed H. Chi. DCN V2: improved deep & cross network and practical lessons for web-scale learning to rank systems. In *WWW ’21: The Web Conference*, pages 1785–1797, 2021.
- [Wang *et al.*, 2023a] Zepu Wang, Yuqi Nie, Peng Sun, Nam H Nguyen, John Mulvey, and H Vincent Poor. St-mlp: A cascaded spatio-temporal linear framework with channel-independence strategy for traffic forecasting. *arXiv preprint arXiv:2308.07496*, 2023.
- [Wang *et al.*, 2023b] Zepu Wang, Peng Sun, Yulin Hu, and Azeddine Boukerche. A novel hybrid method for achieving accurate and timeliness vehicular traffic flow prediction in road networks. *Computer Communications*, 209:378–386, 2023.
- [Wang *et al.*, 2023c] Zepu Wang, Yifei Sun, Zhiyu Lei, Xincheng Zhu, and Peng Sun. Sst: A simplified swin transformer-based model for taxi destination prediction based on existing trajectory. In *2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC)*, pages 1404–1409. IEEE, 2023.
- [Wang *et al.*, 2024a] Ning Wang, Jiang Bian, Yuchen Li, Xuhong Li, Shahid Mumtaz, Linghe Kong, and Haoyi Xiong. Multi-purpose rna language modelling with motif-aware pretraining and type-guided fine-tuning. *Nature Machine Intelligence*, pages 1–10, 2024.
- [Wang *et al.*, 2024b] Qunbo Wang, Ruyi Ji, Tianhao Peng, Wenjun Wu, Zechao Li, and Jing Liu. Soft knowledge prompt: Help external knowledge become a better teacher to instruct llm in knowledge-based vqa. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6132–6143, 2024.
- [Wang *et al.*, 2024c] Zepu Wang, Xiaobo Ma, Huajie Yang, Weimin Lv, Peng Sun, and Sharath Chandra Guntuku. Uncertainty-aware crime prediction with spatial temporal multivariate graph neural networks. *arXiv preprint arXiv:2408.04193*, 2024.
- [Weng and Wu, 2024a] Yijie Weng and Jianhao Wu. Big data and machine learning in defence. *International Journal of Computer Science and Information Technology*, 16(2), 2024.
- [Weng and Wu, 2024b] Yijie Weng and Jianhao Wu. Leveraging artificial intelligence to enhance data security and combat cyber attacks. *Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023*, 5(1):392–399, 2024.
- [Werner, 2022] Tino Werner. A review on instance ranking problems in statistical learning. *Mach. Learn.*, 111(2):415–463, 2022.

- [Xia *et al.*, 2008] Fen Xia, Tie-Yan Liu, Jue Wang, Wensheng Zhang, and Hang Li. Listwise approach to learning to rank: theory and algorithm. In *Machine Learning, Proceedings of the Twenty-Fifth International Conference*, pages 1192–1199, 2008.
- [Xie *et al.*, 2024] Yangxinyu Xie, Tanwi Mallick, Joshua David Bergerson, John K Hutchison, Duane R Verner, Jordan Branham, M Ross Alexander, Robert B Ross, Yan Feng, Leslie-Anne Levy, et al. Wildfiregpt: Tailored large language model for wildfire analysis. *arXiv preprint arXiv:2402.07877*, 2024.
- [Xin *et al.*, 2024] Wangjiaxuan Xin, Kanlun Wang, Zhe Fu, and Lina Zhou. Let community rules be reflected in online content moderation. *arXiv preprint arXiv:2408.12035*, 2024.
- [Xiong *et al.*, 2024a] Haoyi Xiong, Jiang Bian, Yuchen Li, Xuhong Li, Mengnan Du, Shuaiqiang Wang, Dawei Yin, and Sumi Helal. When search engine services meet large language models: Visions and challenges. *IEEE Transactions on Services Computing*, 2024.
- [Xiong *et al.*, 2024b] Haoyi Xiong, Xiaofei Zhang, Jiamin Chen, Xinhao Sun, Yuchen Li, Zeyi Sun, Mengnan Du, et al. Towards explainable artificial intelligence (xai): A data mining perspective. *arXiv preprint arXiv:2401.04374*, 2024.
- [Yang and Ying, 2023] Tianbao Yang and Yiming Ying. AUC maximization in the era of big data and AI: A survey. *ACM Comput. Surv.*, 55(8):172:1–172:37, 2023.
- [Yu *et al.*, 2018] Chao Yu, Dongxu Wang, Tianpei Yang, Wenzuan Zhu, Yuchen Li, Hongwei Ge, and Jiankang Ren. Adaptively shaping reinforcement learning agents via human reward. In *PRICAI 2018: Trends in Artificial Intelligence: 15th Pacific Rim International Conference on Artificial Intelligence, Nanjing, China, August 28–31, 2018, Proceedings, Part I 15*, pages 85–97. Springer, 2018.
- [Zhang *et al.*, 2016] Xin Zhang, Ben He, and Tiejian Luo. Training query filtering for semi-supervised learning to rank with pseudo labels. *World Wide Web*, 19(5):833–864, 2016.
- [Zhao *et al.*, 2010] Shiqi Zhao, Haifeng Wang, and Ting Liu. Paraphrasing with search engine query logs. In *COLING 2010, 23rd International Conference on Computational Linguistics, Proceedings of the Conference*, pages 1317–1325, 2010.
- [Zhao *et al.*, 2011] Shiqi Zhao, Haifeng Wang, Chao Li, Ting Liu, and Yi Guan. Automatically generating questions from queries for community-based question answering. In *Proceedings of 5th international joint conference on natural language processing*, pages 929–937, 2011.
- [Zheng *et al.*, 2021] Zihe Zheng, Zhongkai Shangguan, and Jiebo Luo. What makes a turing award winner? In *Social, Cultural, and Behavioral Modeling: 14th International Conference, SBP-BRiMS 2021, Virtual Event, July 6–9, 2021, Proceedings 14*, pages 310–320. Springer, 2021.
- [Zhou *et al.*, 2024] Qihua Zhou, Song Guo, Jun Pan, Jiacheng Liang, Jingcai Guo, Zhenda Xu, and Jingren Zhou. Pass: Patch automatic skip scheme for efficient on-device video perception. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.
- [Zhou, 2024a] Qiqin Zhou. Application of black-litterman bayesian in statistical arbitrage. *arXiv preprint arXiv:2406.06706*, 2024.
- [Zhou, 2024b] Qiqin Zhou. Explainable ai in request-for-quote. *arXiv preprint arXiv:2407.15038*, 2024.
- [Zhou, 2024c] Qiqin Zhou. Portfolio optimization with robust covariance and conditional value-at-risk constraints. *arXiv preprint arXiv:2406.00610*, 2024.
- [Zhu *et al.*, 2023] Guangxu Zhu, Zhonghao Lyu, Xiang Jiao, Peixi Liu, Mingzhe Chen, Jie Xu, Shuguang Cui, and Ping Zhang. Pushing ai to wireless network edge: An overview on integrated sensing, communication, and computation towards 6g. *Science China Information Sciences*, 66(3):130301, 2023.