Learning to Rank and pubmedKB Phenotype to Gene

2023/09/22

Li Peng-Hsuan 李朋軒

Agenda

- Learning to Rank (LTR)
- LTR Approaches
- pubmedKB Phenotype to Gene

Agenda

- Learning to Rank (LTR)
- LTR Approaches
- pubmedKB Phenotype to Gene

Learning to Rank (LTR)

$LTR: (q, D) \rightarrow \pi$

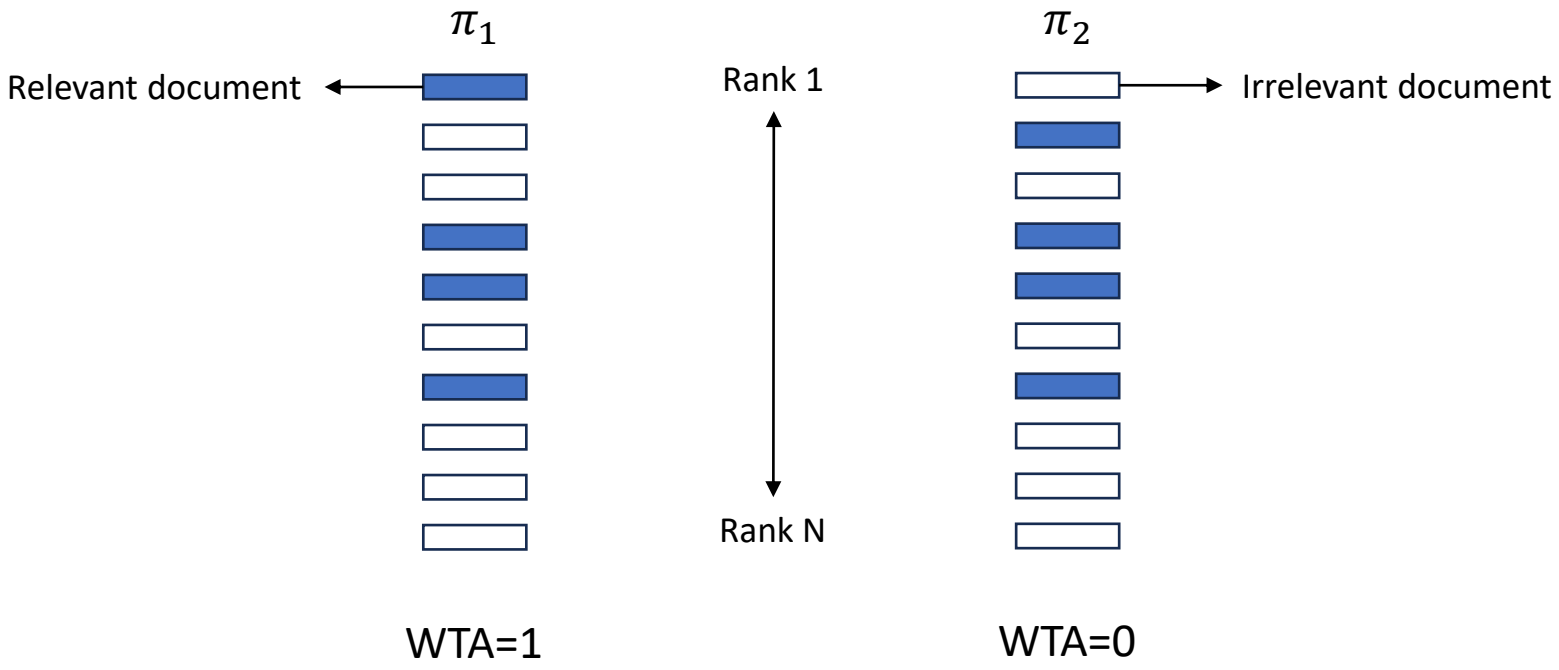
- q: a query
- $D = \{d_i\}$: a set of documents
- π: a permutation (ranked list) of D

Learning to Rank (LTR) — Applications

- Recommendation system
- Search engine
- Information retrieval
- …

Learning to Rank (LTR) — Metrics

- Winner-takes-all (WTA)



Learning to Rank (LTR) — Metrics

- Normalized Discounted Cumulative Gain (NDCG)
- Mean Reciprocal Rank (MRR)

Sum of (true) document relevance scores, each of which decayed by (predicted) ranking.

I.e., top-weighted relevance sum.

Learning to Rank (LTR) — Metrics

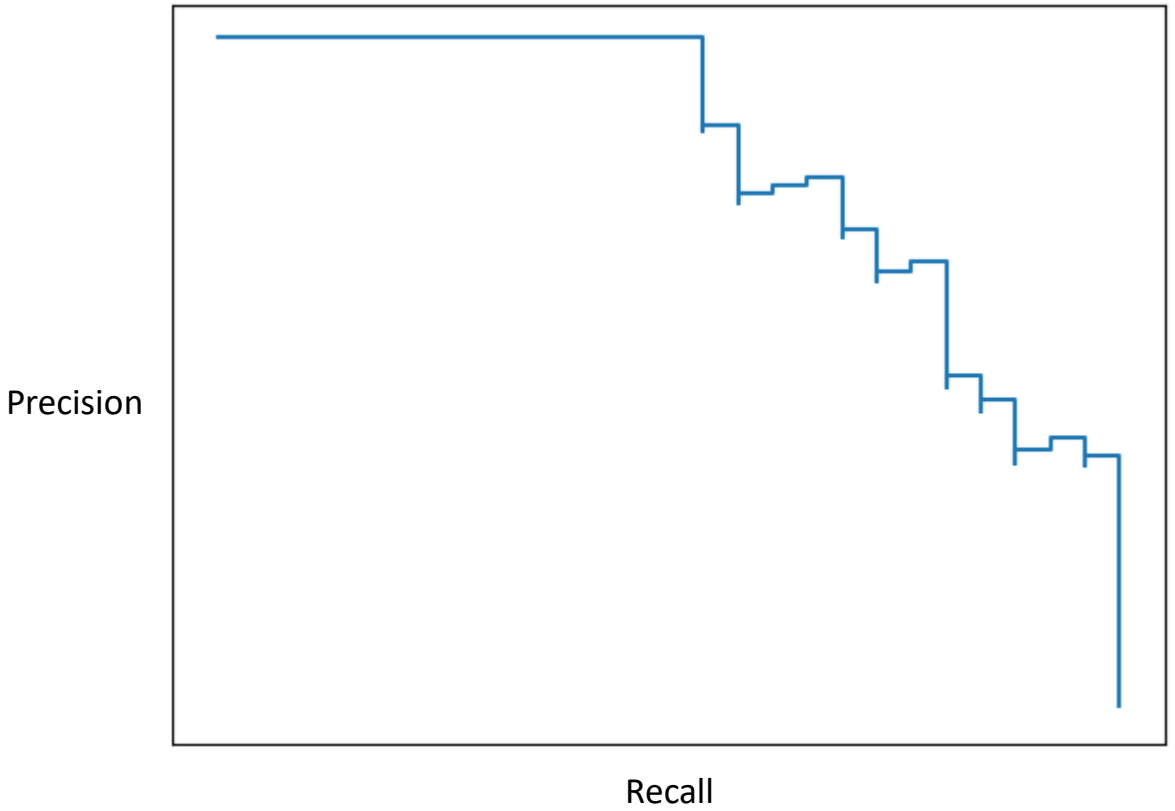
- Mean Average Precision (MAP)

Average Precision (AP)

- Area under precision recall curve
- One per ranked list
- Chance-level: true positive percentage

MAP

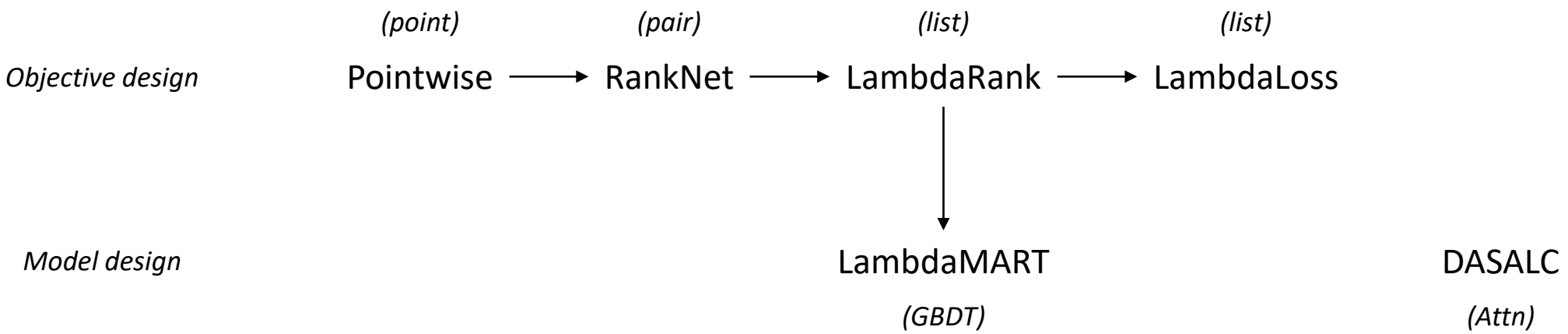
- Mean AP across all queries



Agenda

- Learning to Rank (LTR)
- LTR Approaches
- pubmedKB Phenotype to Gene

LTR Approaches



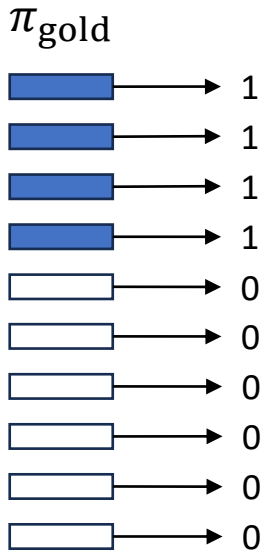
LTR Approaches — Pointwise

LTR: $(q, D) \rightarrow \pi$

Pointwise: $(q, d_i) \rightarrow s_i^q$

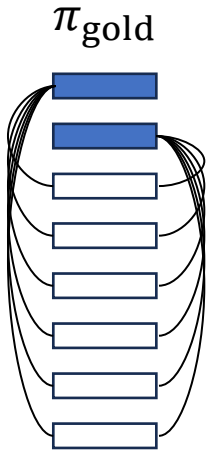
- s_i: absolute relevance score

(q omitted for brevity)



LTR Approaches — RankNet

RankNet: predict pair order



- Less sensitive to class imbalance
- Do not need absolute relevance score
- Pair orders can be partial ordering or even cyclic

(ICML-2005) Learning to Rank using Gradient Descent

LTR Approaches — RankNet

RankNet: predict relative rank probability

$$P(i \to j) \equiv \frac{1}{1 + e^{-\sigma(s_i - s_j)}}$$

$i \to j$: d_i is more relevant than d_j (I.e., d_i ranked higher in π_{gold})

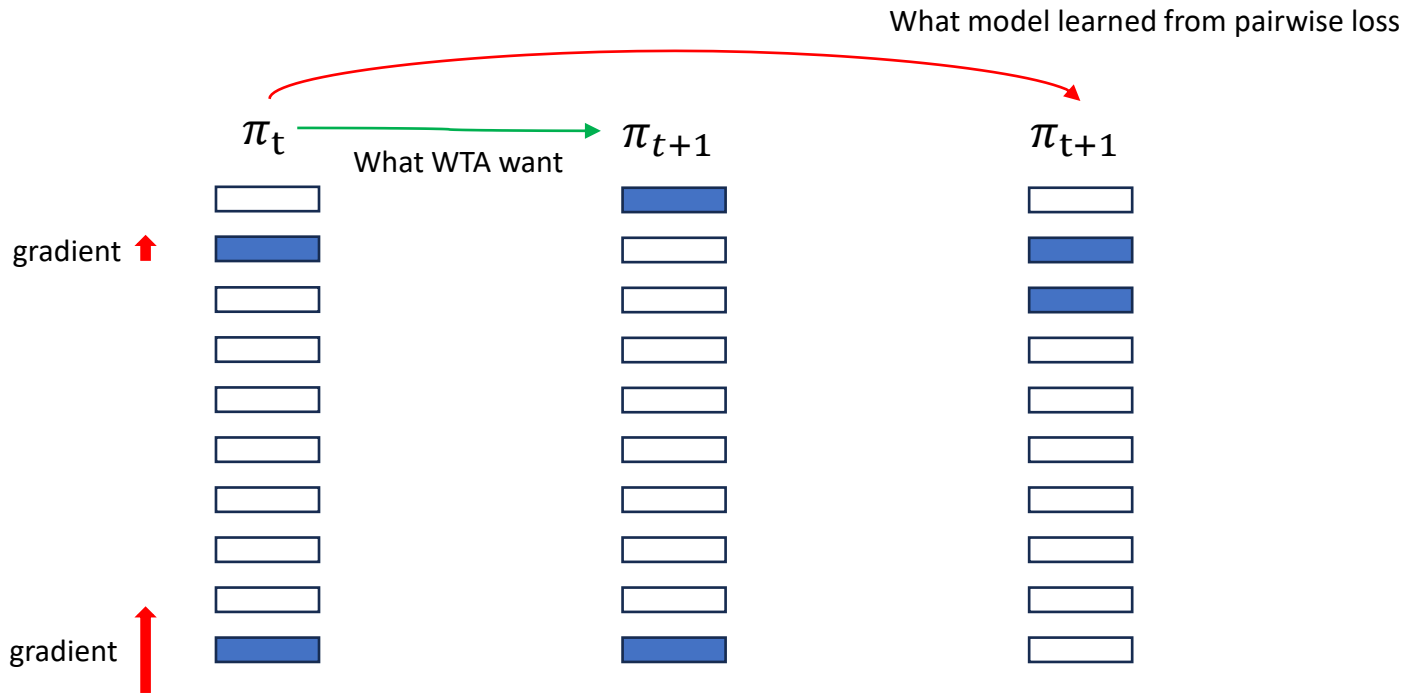
Such that the cross-entropy loss

$$L_{|i \to j} \equiv -logP(i \to j)$$

Is minimized

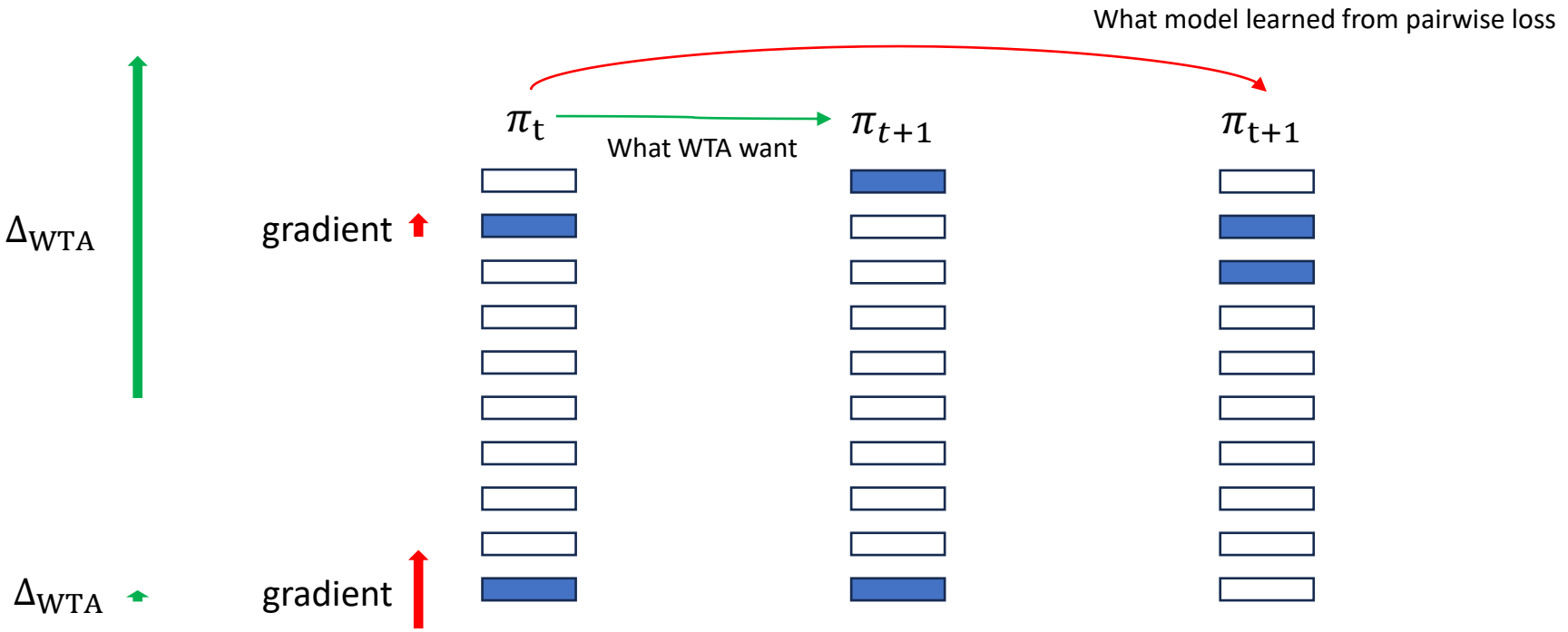
LTR Approaches — Remark

The problem with pairwise objectives



LTR Approaches — LambdaRank

LambdaRank: multiply gradient by metric change



(NeurIPS-2006) Learning to Rank with Non-Smooth Cost Functions

LTR Approaches — LambdaRank

LambdaRank: multiply gradient by metric change

$$\text{(RankNet) } L_{|i\to j} \equiv -logP(i \to j) = log\left(1 + e^{-\sigma(s_i - s_j)}\right)$$

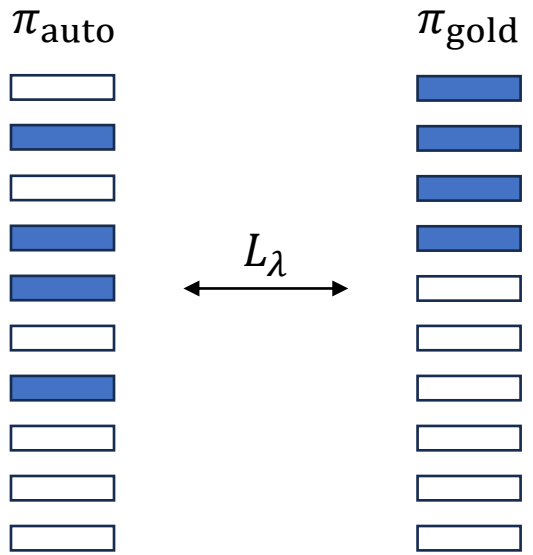
$$\text{(LambdaRank) } \lambda_{ij} \equiv \frac{\partial L}{\partial s_i} \cdot \Delta_{\text{metric}}(i,j) = \frac{-\sigma}{1+e^{\sigma(s_i - s_j)}} \cdot \Delta_{\text{metric}}(i,j)$$

Theorem. λ is the graident of some function L_λ and L_λ is convex

LTR Approaches — LambdaRank

LambdaRank: minimize an implicit global ranked list loss

- Align training objective with target metric



LTR Approaches — LambdaLoss

LambdaLoss: generalized listwise loss

$$l(\mathbf{y}, \mathbf{s}) = -\log_2 P(\mathbf{y}|\mathbf{s}) = -\log_2 \sum_{\pi \in \Pi} P(\mathbf{y}|\mathbf{s}, \pi) P(\pi|\mathbf{s})$$

Ground truth relevance scores

Predicted relevance scores

pointwise, pairwise, listwise

Hard, gaussian, …

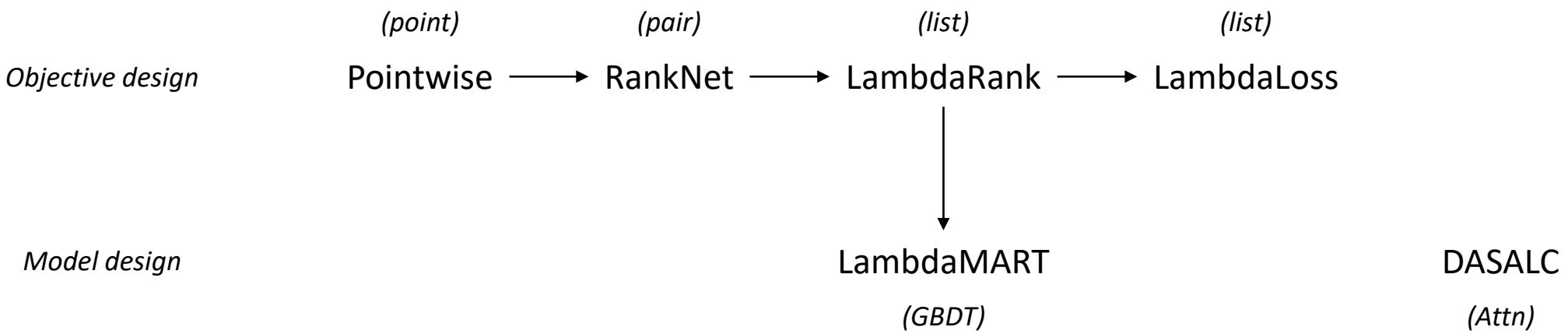
LTR Approaches — LambdaLoss

LambdaLoss: generalized listwise loss

$$l(\mathbf{y}, \mathbf{s}) = -\log_2 P(\mathbf{y}|\mathbf{s}) = -\log_2 \sum_{\pi \in \Pi} P(\mathbf{y}|\mathbf{s}, \pi) P(\pi|\mathbf{s})$$

- Explicit loss definition
- Optimized by Expectation-Maximization (EM)
- *Theorem.* $\mathrm{NDCG} < L_{\mathrm{LambdaLoss-NDCG}} < L_{\lambda}$

LTR Approaches — Recap



LTR Approaches — LambdaMART

Gradient Boosting (GB)

- Learning weak models and their linear ensemble by functional gradient descent

Gradient-Boosted Decision Tree (GBDT/MART)

- GB with decision trees as the weak models

LambdaMART

- GBDT with λ as the functional gradient

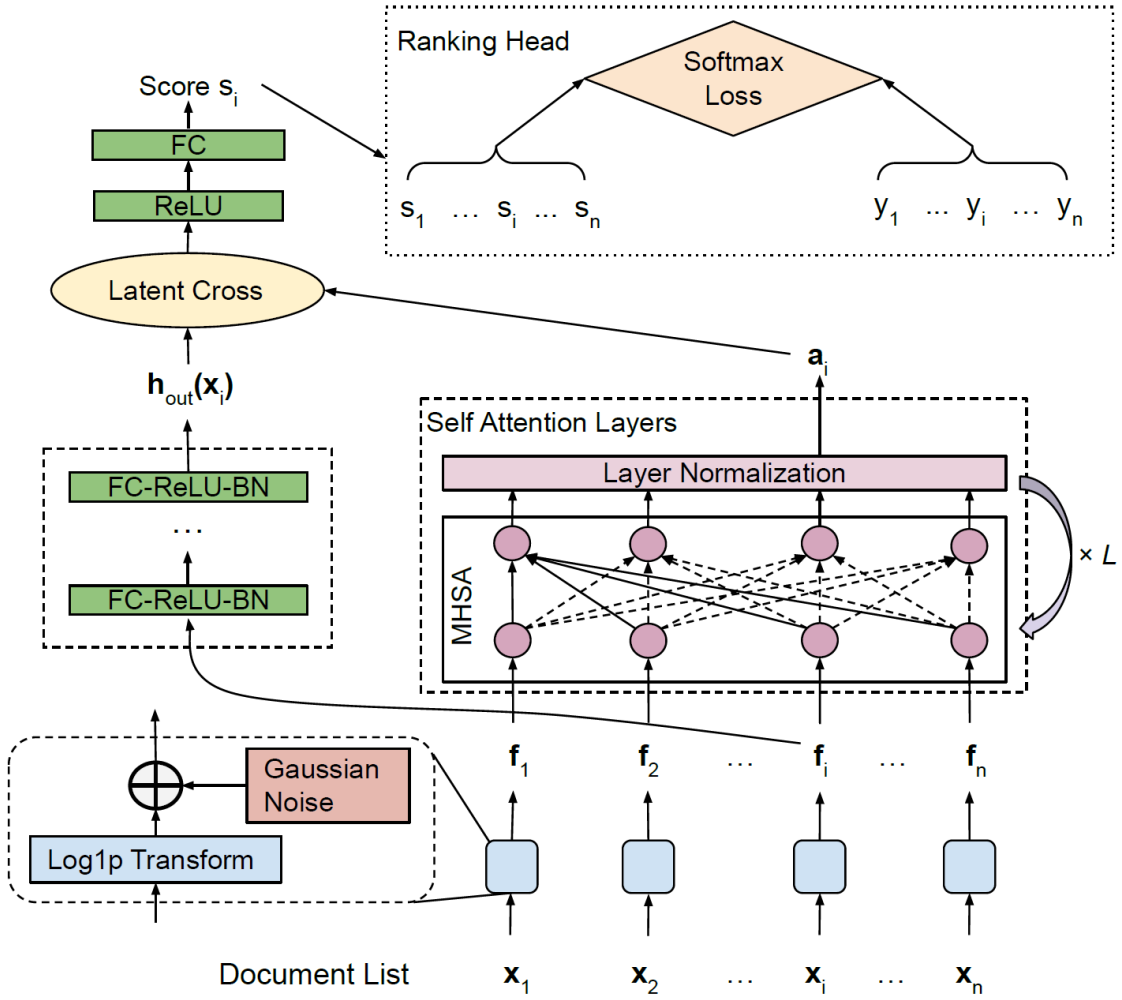
LTR Approaches — LambdaMART

LambdaMART: LambdaRank + GBDT

- Implementation
  - LightGBM (2016)
  - XGBoost (2014, 2.0: 2023/09/12)

  → LightGBM and XGBoost use different tree algorithms

LTR Approaches — DASALC



(ICLR-2021) Are Neural Rankers still Outperformed by Gradient Boosted Decision Trees?

Agenda

- Learning to Rank (LTR)
- LTR Approaches
- pubmedKB Phenotype to Gene

pubmedKB Phenotype to Gene

Using pubmedKB annotations to predict relevant genes per disease

- An *LTR*: $(q, D) \rightarrow \pi$

  q: a query → a disease

  $D = \{d_i\}$: a set of documents → genes

  π: a permutation (ranked list) of D → genes sorted by predicted relevance

- Evaluation

  Mean Average Precision (MAP)

pubmedKB Phenotype to Gene

Dataset: ClinVar 2023 disease-gene association

- Disease: MeSH diseases

  → Mapped from OMIM for ClinVar

  → Retain 2,482 MeSH diseases that are in both ClinVar and pubmedKB

- Gene: 20,670 human protein-coding genes

	# pathogenic MeSH-gene pairs	# MeSHs across pairs	# genes across pairs
ClinVar	4,311	3,175	2,416
pubmedKB	3,128,402	8,894	18,393

pubmedKB Phenotype to Gene

Features: pubmedKB annotation statistics

- Max-min normalized per disease

pubmedKB Phenotype to Gene

Evaluation

Method	MAP	W-MAP
#paper	61.5%	54.8%
Hand-crafted score	64.6%	57.6%
Ridge regression	66.4%	59.3%
XGBoost-LambdaMART-MAP	80.6%	73.5%

W-MAP: mean AP weighted by #pathogenic genes of each disease