Position-Aware ListMLE: A Sequential Learning Process for Ranking

Yanyan Lan, Yadong Zhu, Jiafeng Guo, Shuzi Niu, Xueqi Cheng
Institute of Computing Technology, Chinese Academy of Sciences
{lanyanyan, guojiafeng, cxq}@ict.ac.cn, {zhuyadong, niushuzi}@software.ict.ac.cn

1. Backgrounds

- Notations
  - $x = \{x_1, \ldots, x_n\} \in X$ is a set of objects to be ranked;
  - $y = \{y_1, \ldots, y_n\} \in Y$ is the ground-truth, $y_i$ is the position of $x_i$;
  - $y^{-1}(i)$ stands for the index of items in the $i$-th position of $y$;
  - $\mathbf{f} : X \rightarrow \mathbb{R}^n$ is a ranking function.
- Training Data $S = \{\mathbf{x}, \mathbf{y}\}^{N}_{i=1}$, where $\mathbf{x} = \{x_1^{(i)}, \ldots, x_n^{(i)}\}$ and $\mathbf{y} = \{y_1^{(i)}, \ldots, y_n^{(i)}\}$.
- Expected true risk

$$R_0(h) = \int_{x \in X} L_0(f(x, y)) dP_{XY}(x, y).$$
- $L_0$ is a true loss, e.g., permutation level 0-1 loss $L_{0-1}(f(x,y))=1_{f(x)\neq y}$.
- Empirical surrogate risk

$$\hat{R}(h) = \frac{1}{N} \sum_{i=1}^{N} L_0(h(x_i, y_i)).$$
- $L_0$ is a surrogate loss, usually convex.

2. Motivation

- ListMLE does not consider position importance, which is a key factor in ranking.
  - ListMLE utilizes a likelihood loss as the surrogate loss

$$L(f; x, y) = -\log P(y|x; f),$$

$$P(y|x; f) = P(y^{-1}(1)|x; f) \prod_{i=2}^{n} P(y^{-1}(i)|x; f, y^{-1}(1), \ldots, y^{-1}(i-1); f)$$

$$P(y^{-1}(i)|x; f) = \frac{\exp(f(x, y^{-1}(i)))}{\sum_{j=1}^{i} \exp(f(x, y^{-1}(j)))},$$

$$P(y^{-1}(i)|x; f, y^{-1}(1), \ldots, y^{-1}(i-1); f) = \frac{\exp(f(x, y^{-1}(i)))}{\sum_{j=i-1}^{n} \exp(f(x, y^{-1}(j)))}, \forall i = 2, \ldots, n$$
- Theoretically, ListMLE has been proven to be consistent with permutation level 0-1 loss, which ignores position importance.
- The chain rule of probability is the essential reason

$$P(A_1, \ldots, A_n) = P(A_1)P(A_1|A_2)\cdots P(A_1|A_2|\cdots A_n),$$

where $(i_1, \ldots, i_n)$ is any permutation of $(1, \ldots, n)$.

3. Sequential Ranking Approach

Step 1: Maximizing the probability that the top 1 object is selected with mathematical description as follows.

$$\max_{f \in \mathcal{F}} P(y^{-1}(1)|x; f);$$

Step i: For $i = 2, \ldots, n$, we denote the subset of ranking functions that reach the maximum in Step $i-1$ as $S_{i-1}$. The task of step $i$ is to maximize the probability that the object with position $i$ in ground-truth permutation is selected given the top $(i-1)$ objects ranked correctly. The mathematical formulation is described as follows.

$$\max_{f \in S_{i-1}} P(y^{-1}(i)|x; y^{-1}(1), \ldots, y^{-1}(i-1); f);$$

Step n+1: The learning process ends, and the ranking function $f$ is randomly selected from $S_{n-1}$ as the output ranking function.

4. P-ListMLE

- Optimization Problem

$$\min_{f \in \mathcal{F}} \Phi(f),$$

$$\Phi(f) = \sum_{i=2}^{n} \alpha(i) \log P(y^{-1}(i)|x; y^{-1}(1), \ldots, y^{-1}(i-1); f)$$

$$- \alpha(i) \log P(y^{-1}(i)|x; f),$$

where $\alpha(\cdot)$ is a decreasing function, i.e. $\alpha(i) > \alpha(i+1)$.
- Loss function

$$L_p(f; x, y) = \sum_{i=1}^{n} \alpha(i)(-f(x, y^{-1}(i)) + \log(\sum_{j=1}^{n} \exp(f(x, y^{-1}(j))))).$$

- With the Rank-Differentiable assumption (Lan 2012), we can prove that p-ListMLE is consistent with WPDL (Duchi 2010), which a position-aware loss function.

5. Experiments

- Datasets: MQ2007 and MQ2008 in benchmark data sets LETOR4.0, where ground-truth is a full order ranking list.
- Both ListMLE and p-ListMLE are implemented by Stochastic Gradient Descent (SGD).
- We set $\alpha(i)$ as $2^{m-i} - 1$, as guided in the discussion section.
- NDCG is used for evaluation, where label is set as $\mathcal{L}(x_i) = n - y_i$, where $y_i$ is the rank of item $i$ in the listwise ground-truth.
- Baselines also include state-of-the-art listwise ranking algorithms RankCosine and ListNet.

6. Conclusion & Future Work

- We find that ListMLE highly ignore the position importance in ranking.
- We propose a new listwise ranking algorithm, named position-aware ListMLE to address this problem.
- We provide both theoretical and empirical analysis to show that p-ListMLE is better than ListMLE.
- For future work, we plan to further investigate the problem of how to set the position factor $\alpha$ in real application, or how to guide the settings from other theoretical aspects.