Is Top-k Sufficient for Ranking?

Yanyan Lan, Shuzi Niu, Jiafeng Guo, Xueqi Cheng
Institute of Computing Technology,
Chinese Academy of Sciences
Outlines

• Motivation
• Problem Definition
• Empirical Analysis
• Theoretical Results
• Conclusions and Future Work
Outlines

• Motivation
• Problem Definition
• Empirical Analysis
• Theoretical Results
• Conclusions and Future Work
Traditional Learning to Rank

- Learning to Rank has become an important means to tackle ranking problem in many application!

Training data are not reliable!

(1) Difficulty in choosing gradations;
(2) High assessing burden;
(3) High level of disagreement.

From Tie-Yan Liu’s Tutorial on WWW’08
Top-k Learning to Rank

- Revisit the training of learning to rank:
  - Full-Order Ranking Lists
  - User mainly care about top results!
  - Ideal
  - Top-k Ground-truth
  - Surrogate

- Top-k labeling strategy based on pairwise preference judgment:
  - Assumption: top-k ground-truth is sufficient for ranking!
  - The training data are proven to be more reliable!
  - [SIGIR2012, CIKM2012]
  - Best Student Paper Award

- HeapSort

Motivation

Preferences Order

$\begin{pmatrix}
    x_{i_1} \\
    x_{i_2} \\
    \vdots \\
    x_{i_{n-1}} \\
    x_{i_n}
\end{pmatrix}$
Outlines

• Motivation
• Problem Definition
• Empirical Analysis
• Theoretical Results
• Conclusions and Future Work
Problem Definition

Assumption: top-k ground-truth is sufficient for ranking!

Training on top-k setting is as good as that in full-order setting.

Top-k ground-truth are utilized for training.

Full-order ranking lists are adopted as ground-truth.
Full-Order Setting

• Training Data
  \[
  \{(q_i, x_i, y_i)\}_{i=1}^{N}
  \]
  \[\begin{aligned}
  (x_1^{(i)}, \ldots, x_{n_i}^{(i)}) & : (y_1^{(i)}, \ldots, y_{n_i}^{(i)}) \\
  \text{Documents} & \quad \text{full-order ranking lists}
  \end{aligned}\]
  The index of the item ranked in corresponding position

• Training Loss
  – Pairwise Algorithm
    • Ranking SVM (*hinge loss*)
    • RankBoost (*exponential loss*)
    • RankNet (*logistic loss*)
  – Listwise Algorithm
    • ListMLE (*likelihood loss*)

\[
L^p(f; x_{y_j}, x_{y_l}) = \sum_{i=1}^{N} \sum_{j=1}^{n_i} \sum_{l=j+1}^{n_i} L^p(f; x_{y_j}, x_{y_l}),
\]

\[
L^l(f; x, y) = - \log P(y|x, f), \quad P(y|x, f) = \prod_{j=1}^{n-1} \frac{\exp\{f(x_{y_j})\}}{\sum_{l=j}^{n} \exp\{f(x_{y_l})\}}.
\]
Top-k Setting

• Training Data
  \[(q_i, x_i, Y_k^{(i)})\]
  \(\{x_1^{(i)}, \ldots, x_{n_i}^{(i)}\}\)
  Query \rightarrow \text{Documents} \rightarrow \text{A set of full-order ranking lists}

  – example: \(\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}\) \(\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}\)

• Training Loss
  – Pairwise Algorithm
  \[\sum_{i=1}^{N} \min_{y \in Y_k^{(i)}} \sum_{j=1}^{k} \sum_{l=j+1}^{n_i} L^p(f; x_{y_j}^{(i)}, x_{y_l}^{(i)})\]

  – Listwise Algorithm
  \[= \sum_{i=1}^{N} \sum_{j=1}^{k} \sum_{l=j+1}^{n_i} L^p(f; x_{y_j}^{(i)}, x_{y_l}^{(i)}), \forall y^{(i)} \in Y_k^{(i)}\]

• ListMLE \rightarrow \text{Top-k ListMLE (Xia et al. NIPS’09)}
Outlines

• Motivation
• Problem Definition
• Empirical Analysis
• Theoretical Results
• Conclusions and Future Work
Empirical Study

Assumption: top-k ground-truth is sufficient for ranking!

Training on top-k setting is as good as that in full-order setting.

Ranking function f1

Ranking function f2

Test Performance Comparison
Experimental Setting

• Datasets
  – LETOR 4.0(MQ2007-list, MQ2008-list)
    • Ground-truth: full order
    • Top-k ground-truth are constructed by just preserving the total order of top k items

• Algorithms
  – Pairwise: Ranking SVM, RankBoost, RankNet
  – Listwise: ListMLE

• Experiments
  – Study how the test performances of ranking algorithms change w.r.t. k in the training data of top-k setting.
Figure 1: Performance variations of different ranking algorithms in top-k setting on MQ2007-list with the increase of k.
Figure 2: Performance variation of different ranking algorithms in top-$k$ setting on MQ2008-list with the increase of $k$
Experimental Results

(1) Overall, the test performance of ranking algorithms in top-k setting increase to a stable value with the growth of k.

(2) However, when k keeps increasing, the performances will decrease.

(3) The test performances of the four algorithms increase quickly to a stable value with the increase of k.

• Empirically, top-k ground-truth is sufficient for ranking!
Outlines

• Motivation
• Problem Definition
• Empirical Analysis
• Theoretical Results
• Conclusions and Future Work
Theoretical Problem Formalization

Assumption: top-k ground-truth is sufficient for ranking!

Training on top-k setting is as good as that in full-order setting.

Relationships between losses in top-k setting and full-order setting.

We can prove that:
(1) Pairwise losses in full-order setting are upper bounds of that in top-k setting.
(2) The loss of ListMLE in full-order setting is an upper bound of top-k ListMLE.

What we really care about is the opposite of the coin!

Relationships among losses in top-k setting, losses in full-order setting and IR evaluation measures!
Theoretical Results

**Losses in Top-k Setting** \(\leq\) **Losses in Full-Order Setting**

\[ L^p(f; x, Y_k) = \min_{y \in Y_k} \sum_{j=1}^k \sum_{i=j+1}^n L^p(f; x_{y_j}, x_{y_l}). \]

\[ L^i(f; x, Y_k) = \min_{y \in Y_k} \sum_{j=1}^k \{ -f(x_{y_j}) + \log(\sum_{l=j}^n \exp\{f(x_{y_l})\}) \}. \]

**Weighted Kendall’s Tau**

\[ L_\alpha(f; x, Y_k) \leq \frac{1}{\ln 2} (\max_{1 \leq i \leq k} \alpha(i)) L^i(f; x, Y_k); \]

**IR Evaluation Measures (NDCG)**

\[ 1 - \text{NDCG}@k(f; x, y) \leq \frac{1}{N_k} L_\alpha(f; x, Y_k); \]

\[ \text{NDCG}@k(f; x, y) = \frac{1}{N_k} \sum_{j=1}^k g(l(y_j)) D(r_j), \]

**Conclusion:** Losses in top-k setting are tighter bounds of 1-NDCG, compared with those in full-order setting!
Conclusion & Future Work

• We address the problem of whether the assumption of top-k ranking holds.
  – Empirically, the test performance of four algorithms (pairwise and listwise) quickly increase to a stable value with the growth of k.
  – Theoretically, we prove that loss functions in top-k settings are tighter lower bounds of 1-NDCG, as compared to that in full-order setting.

• Our analysis from both empirical and theoretical aspects show that top-k ground-truth is sufficient for ranking.

• Future work: theoretically study the relationship between different objects from other aspect such as statistical consistency.
Thanks for your attention!

Q&A : lanyanyan@ict.ac.cn