Coactive Learning

Coactive Learning: [Shivaswamy and Joachims, 2012]
- Given context \( x \), predict object \( y \) to optimize utility \( U(x, y) \).
- Models the interaction between user(s) and learning system.

Instability of the Preference Perceptron

Linear Utility: \( U(x, y) = w^\top \phi(x, y) \)
Preference Perceptron Algorithm:
(Proposed in [Shivaswamy and Joachims, 2012])
1. Initialize weight vector \( w_1 \leftarrow 0 \).
2. Given context \( x_t \) present
   \( \hat{y}_t \leftarrow \arg \max_{y} w^\top \phi(x_t, y) \).
3. Observe clicks. If clicked document is lower
   element of pair, move it up by one to get \( \hat{y}_t \).
4. \( w_{t+1} \leftarrow w_t + \phi(x_t, \hat{y}_t) - \phi(x_t, y_t) \).
5. Repeat from step 2.

User Study:
- Experimented using full-text search engine at arxiv.org.
- Goal: Learning a ranking function from implicit feedback i.e., user clicks.
- Interleaved evaluation against hand-tuned baseline ranker.
- Win ratio of 1 indicates being no better than the baseline. Higher win ratio is better.

Experimental Results
- Performed offline experiments on a search dataset (Yahoo! LTR) and two news recommendation datasets: RCV1 and News.
- Simulated user behavior with and without noise.
- NDCG@5 was the utility for all three datasets.

Comparison with other methods:
- PrefP[top]: Preference Perceptron with move-to-top feedback.
- PrefP[pair]: PrefP with pairwise feedback i.e., 3PR with \( p_1 = 0 \).
- SVM: Ranking SVM with move-to-top feedback.
- Perceptron which receives optimal is (rough) upper bound.

Fixed Probability 3PR

The regret of 3PR with fixed swap probability \( p \) (i.e., \( \forall t: p_t = p \)) is:
\[
\text{Regret} = \frac{\sum_{t=1}^{T} \alpha}{\alpha} + \frac{p(1-p^2)R\|w_0\|}{\alpha} \sqrt{2(4-p^2)(1-\frac{1}{2p})\|w_0\|} + \frac{\|w_T\|}{\alpha \sqrt{T}}.
\]

Dynamic 3PR

For any \( \Delta \geq 0 \), dynamically setting the swap probability \( p_t \) of 3PR has regret:
\[
\text{Regret} = \frac{\sum_{t=1}^{T} \alpha}{\alpha} + \frac{\|w_T\|}{\alpha \sqrt{T}} \left( \frac{4R^2 + 2\Delta + 2R}{T} \right).
\]

Theoretical Analysis

\( \alpha \)-Informative Feedback:
We characterize the utility of the feedback received \( \hat{y}_t \) as:
\[
E[U(x, \hat{y}_t)] \geq U(x, y) + \alpha (U(x, y') - U(x, y)) - \xi
\]
- where \( y' \) is the optimal and \( y \) is the presented object.
- Note that this is just a characterization (not an assumption).
- Used to prove regret bounds.

Regret:
We define the regret after \( T \) iterations as:
\[
\frac{1}{T} \sum_{t=1}^{T} (U(x_t, y'_t) - E[U(x_t, y'_t)])
\]

For \( n = 10 \) and \( 20\% \) user error rate, average rank of the relevant document for 3PR (with \( p = 0.5 \)) is 2.1 (compared to the 9.4 for PrefP).

Stable Coactive Learning via Perturbation
Karthik Raman, Thorsten Joachims (Cornell University)
Pannaga Shivaswamy (AT&T Research), Tobias Schnabel (University of Stuttgart)