

Stable Coactive Learning via Perturbation

Karthik Raman ¹ Thorsten Joachims ¹ Pannaga Shivaswamy ²
Tobias Schnabel ³

¹Cornell University {*karthik,tj*}@*cs.cornell.edu*

²AT&T Research *pannaga@research.att.com*

³Stuttgart University *tbs49@cornell.edu*

June 19, 2013

Learning model

Repeat forever:

- System receives context \mathbf{x}_t .

Coactive Learning

Learning model

Repeat forever:

- System receives context \mathbf{x}_t .

→ e.g. : Search Engine

Coactive Learning

Learning model

Repeat forever:

- System receives context \mathbf{x}_t .

e.g. : Search Engine

User Query

Coactive Learning

Learning model

Repeat forever:

- System receives context \mathbf{x}_t .
- System makes prediction \mathbf{y}_t .

→ e.g. : Search Engine

→ User Query

→ Ranking

Coactive Learning

Learning model

Repeat forever:

- System receives context \mathbf{x}_t .
- System makes prediction \mathbf{y}_t .
- Regret = Regret + $U(\mathbf{x}_t, \mathbf{y}_t^*) - U(\mathbf{x}_t, \mathbf{y}_t)$

e.g. : Search Engine

User Query

Ranking

User utility

Coactive Learning

Learning model

Repeat forever:

- System receives context \mathbf{x}_t .
- System makes prediction \mathbf{y}_t .
- Regret = Regret + $U(\mathbf{x}_t, \mathbf{y}_t^*) - U(\mathbf{x}_t, \mathbf{y}_t)$
- System gets feedback:
Full information: $U(\mathbf{x}_t, \mathbf{y}^{(1)}), U(\mathbf{x}_t, \mathbf{y}^{(2)}), \dots$

e.g. : Search Engine

User Query

Ranking

User utility

Coactive Learning

Learning model

Repeat forever:

- System receives context \mathbf{x}_t .
- System makes prediction \mathbf{y}_t .
- Regret = Regret + $U(\mathbf{x}_t, \mathbf{y}_t^*) - U(\mathbf{x}_t, \mathbf{y}_t)$
- System gets feedback:
Full information: $U(\mathbf{x}_t, \mathbf{y}^{(1)}), U(\mathbf{x}_t, \mathbf{y}^{(2)}), \dots$

e.g. : Search Engine

User Query

Ranking

User utility

Unrealistic for users to provide (e.g., implicit feedback).

Coactive Learning

Learning model

Repeat forever:

- System receives context \mathbf{x}_t .
- System makes prediction \mathbf{y}_t .
- Regret = Regret + $U(\mathbf{x}_t, \mathbf{y}_t^*) - U(\mathbf{x}_t, \mathbf{y}_t)$
- System gets feedback:
~~Full information: $U(\mathbf{x}_t, \mathbf{y}^{(1)}), U(\mathbf{x}_t, \mathbf{y}^{(2)}), \dots$~~
Bandit: $U(\mathbf{x}_t, \mathbf{y}_t)$

e.g. : Search Engine

User Query

Ranking

User utility

Unrealistic for users to provide (e.g., implicit feedback).

Coactive Learning

Learning model

Repeat forever:

- System receives context \mathbf{x}_t .
- System makes prediction \mathbf{y}_t .
- Regret = Regret + $U(\mathbf{x}_t, \mathbf{y}_t^*) - U(\mathbf{x}_t, \mathbf{y}_t)$
- System gets feedback:
~~Full information: $U(\mathbf{x}_t, \mathbf{y}^{(1)}), U(\mathbf{x}_t, \mathbf{y}^{(2)}), \dots$~~
~~Bandit: $U(\mathbf{x}_t, \mathbf{y}_t)$~~
Coactive: $U(\mathbf{x}_t, \bar{\mathbf{y}}_t) \geq_{\alpha} U(\mathbf{x}_t, \mathbf{y}_t)$

e.g. : Search Engine

User Query

Ranking

User utility

Coactive Learning

Learning model

Repeat forever:

- System receives context \mathbf{x}_t .
- System makes prediction \mathbf{y}_t .
- Regret = Regret + $U(\mathbf{x}_t, \mathbf{y}_t^*) - U(\mathbf{x}_t, \mathbf{y}_t)$
- System gets feedback:
~~Full information: $U(\mathbf{x}_t, \mathbf{y}^{(1)}), U(\mathbf{x}_t, \mathbf{y}^{(2)}), \dots$~~
~~Bandit: $U(\mathbf{x}_t, \mathbf{y}_t)$~~
Coactive: $U(\mathbf{x}_t, \bar{\mathbf{y}}_t) \geq_{\alpha} U(\mathbf{x}_t, \mathbf{y}_t)$

e.g. : Search Engine

User Query

Ranking

User utility

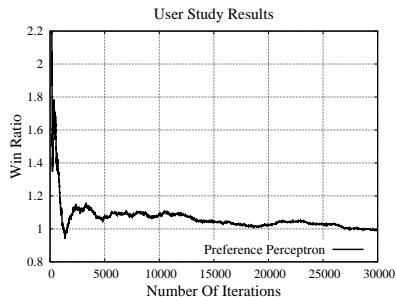
Perceptron has regret $O\left(\frac{1}{\alpha\sqrt{T}}\right)$ for linear utility ($U(\mathbf{x}, \mathbf{y}) = \mathbf{w}_*^T \phi(\mathbf{x}, \mathbf{y})$).

User Study: Learning Rankings using Perceptron

- On live search engine.
- Goal: Learn ranking function from user clicks.
- Interleaved comparison against hand-tuned baseline.

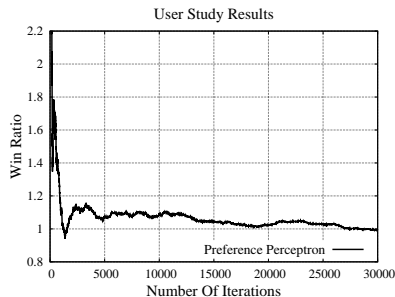
User Study: Learning Rankings using Perceptron

- On live search engine.
- Goal: Learn ranking function from user clicks.
- Interleaved comparison against hand-tuned baseline.
- Win ratio of 1 means no better than baseline (Higher = Better).



User Study: Learning Rankings using Perceptron

- On live search engine.
- Goal: Learn ranking function from user clicks.
- Interleaved comparison against hand-tuned baseline.
- Win ratio of 1 means no better than baseline (Higher = Better).



Perceptron performs poorly!

User Study: Learning Rankings using Perceptron

Preference Perceptron Algo:

- 1 Initialize weight vector $\mathbf{w}_1 \leftarrow \mathbf{0}$.
- 2 Given context \mathbf{x}_t present $\mathbf{y}_t \leftarrow \operatorname{argmax}_{\mathbf{y}} \mathbf{w}_t^T \phi(\mathbf{x}_t, \mathbf{y})$.

- On live search engine.
- Goal: Learn ranking function from user clicks.
- Interleaved comparison against hand-tuned baseline.
- Win ratio of 1 means no better than baseline (Higher = Better).

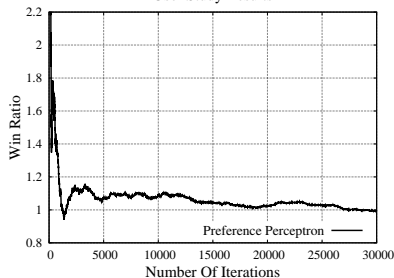
Presented Ranking (y)



The screenshot shows a list of search results with the following titles and snippets:

- Fed minutes to clarify extent of discord on easing** (3:11 PM EDT) - A number of top Federal Reserve officials likely saw a need for additional monetary easing at the central bank's meeting last month, although there are few signs the central bank will raise rates. [More »](#)
- Obama's budget offers short-term help for economy** (2:17 PM EDT) - The president will send Congress a budget that will provide short-term help to a struggling economy while offering a long-term plan to deal with soaring deficits. [More »](#)
- BlackBerry out at U.S. climate agency** (1:44 PM EDT) - Research In Motion's BlackBerry smartphones has dropped to win over U.S. consumers but the Canadian company has long been able to rely on the [More »](#)
- Rock bottom Novara stuns hapless Inter** (1:17 PM EDT) - Nov Novara dumped to a 0-1 successive game without a win as rock bottom Novara walked away from the San Siro on Sunday with a stunning 1-0 victory. [More »](#)

User Study Results



Perceptron performs poorly!

User Study: Learning Rankings using Perceptron

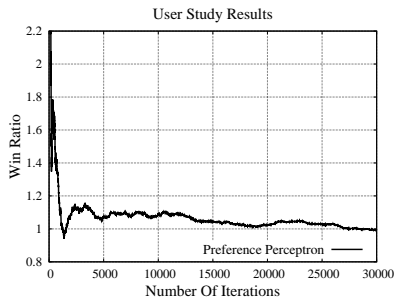
Preference Perceptron Algo:

- 1 Initialize weight vector $\mathbf{w}_1 \leftarrow \mathbf{0}$.
- 2 Given context \mathbf{x}_t present $\mathbf{y}_t \leftarrow \operatorname{argmax}_{\mathbf{y}} \mathbf{w}_t^T \phi(\mathbf{x}_t, \mathbf{y})$.
- 3 Observe clicks and construct feedback ranking $\bar{\mathbf{y}}_t$.

Presented Ranking (y)



- On live search engine.
- Goal: Learn ranking function from user clicks.
- Interleaved comparison against hand-tuned baseline.
- Win ratio of 1 means no better than baseline (Higher = Better).



Perceptron performs poorly!

User Study: Learning Rankings using Perceptron

Preference Perceptron Algo:

- 1 Initialize weight vector $\mathbf{w}_1 \leftarrow \mathbf{0}$.
- 2 Given context \mathbf{x}_t present $\mathbf{y}_t \leftarrow \operatorname{argmax}_y \mathbf{w}_t^\top \phi(\mathbf{x}_t, \mathbf{y})$.
- 3 Observe clicks and construct feedback ranking $\bar{\mathbf{y}}_t$.

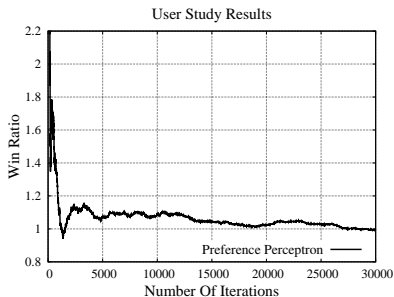
- On live search engine.
- Goal: Learn ranking function from user clicks.
- Interleaved comparison against hand-tuned baseline.
- Win ratio of 1 means no better than baseline (Higher = Better).

Presented Ranking (\mathbf{y})

Feedback Ranking ($\bar{\mathbf{y}}$)

Click!

Click!



Perceptron performs poorly!

User Study: Learning Rankings using Perceptron

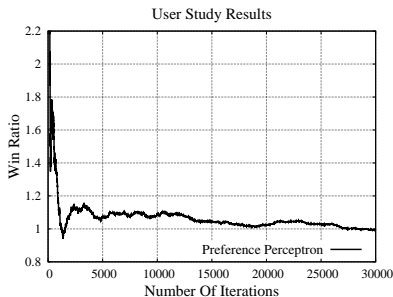
Preference Perceptron Algo:

- 1 Initialize weight vector $\mathbf{w}_1 \leftarrow \mathbf{0}$.
- 2 Given context \mathbf{x}_t present $\mathbf{y}_t \leftarrow \operatorname{argmax}_y \mathbf{w}_t^\top \phi(\mathbf{x}_t, \mathbf{y})$.
- 3 Observe clicks and construct feedback ranking $\bar{\mathbf{y}}_t$.
- 4 $\mathbf{w}_{t+1} \leftarrow \mathbf{w}_t + \phi(\mathbf{x}_t, \bar{\mathbf{y}}_t) - \phi(\mathbf{x}_t, \mathbf{y}_t)$.
- 5 Repeat from step 2.

Presented Ranking (\mathbf{y})

Feedback Ranking ($\bar{\mathbf{y}}$)

- On live search engine.
- Goal: Learn ranking function from user clicks.
- Interleaved comparison against hand-tuned baseline.
- Win ratio of 1 means no better than baseline (Higher = Better).

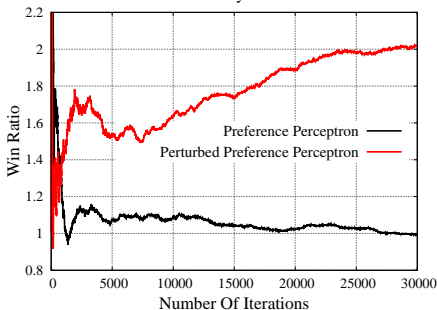


Perceptron performs poorly!

Perturbed Preference Perceptron

- 1 Initialize weight vector $\mathbf{w}_1 \leftarrow \mathbf{0}$.
- 2 Given context \mathbf{x}_t compute $\hat{\mathbf{y}}_t \leftarrow \operatorname{argmax}_{\mathbf{y}} \mathbf{w}_t^T \phi(\mathbf{x}_t, \mathbf{y})$.
- 3 Present $\mathbf{y}_t \leftarrow \text{Perturb}(\hat{\mathbf{y}}_t)$
(Randomly swap adjacent pairs).
- 4 Observe clicks and construct feedback ranking $\bar{\mathbf{y}}_t$.
- 5 $\mathbf{w}_{t+1} \leftarrow \mathbf{w}_t + \phi(\mathbf{x}_t, \bar{\mathbf{y}}_t) - \phi(\mathbf{x}_t, \mathbf{y}_t)$.
- 6 Repeat from step 2.

User Study Results



Predicted Ranking ($\hat{\mathbf{y}}$)

A list of four news items with small thumbnail images:

- Fed minutes to clarify extent of discord on easing** - 34 mins ago
WASHINGTON (Reuters) - A number of top Federal Reserve officials likely saw a need for additional monetary easing at the central bank's meeting last month, although there are few signals the central bank will move soon. Minutes [More](#)
- Obama's budget offers short-term help for economy** - 34 mins ago
The president will send Congress a budget that will provide short-term help to a struggling economy while offering a long-term plan to deal with soaring deficits, the ... [More](#)
- BlackBerry out at U.S. climate agency, iPhone in** - Alister Sharp - Reuters - Fri, Feb 10, 2012
TORONTO (Reuters) - Research In Motion's BlackBerry smartphone has struggled to win over U.S. consumers but the Canadian company has long been able to rely on the
- Rock bottom Novara stun hapless Inter** - AFP - 1 hr 21 mins ago
Inter Milan slumped to a fifth successive game without a win as rock-bottom Novara walked away from the San Siro on Sunday with a stunning 1-0 victory. [More](#)

PERTURB

Presented Ranking (\mathbf{y})

A list of four news items, identical to the predicted ranking, but with a blue arrow indicating a swap between the second and third items:

- Obama's budget offers short-term help for economy** - 34 mins ago
The president will send Congress a budget that will provide short-term help to a struggling economy while offering a long-term plan to deal with soaring deficits, the ... [More](#)
- Fed minutes to clarify extent of discord on easing** - 34 mins ago
WASHINGTON (Reuters) - A number of top Federal Reserve officials likely saw a need for additional monetary easing at the central bank's meeting last month, although there are few signals the central bank will move soon. Minutes [More](#)
- BlackBerry out at U.S. climate agency, iPhone in** - Alister Sharp - Reuters - Fri, Feb 10, 2012
TORONTO (Reuters) - Research In Motion's BlackBerry smartphone has struggled to win over U.S. consumers but the Canadian company has long been able to rely on the
- Rock bottom Novara stun hapless Inter** - AFP - 1 hr 21 mins ago
Inter Milan slumped to a fifth successive game without a win as rock-bottom Novara walked away from the San Siro on Sunday with a stunning 1-0 victory. [More](#)

I will tell you:

- Why the preference perceptron performs poorly?
- Why does perturbation fix the problem?
- What are the regret bounds for the algorithm?
- How do we do this more generally for non-ranking problems?