On Multilabel Classification and Ranking with Partial Feedback

Claudio Gentile and Francesco Orabona

- Some positions are more important
- What to put in each banner?
- The user will click only on one of them
- No ranking information from user

Poster Th65
Cost sensitive, hamming, and ranking loss

10 classes, 4 predictions at most

\[
\hat{Y}_t = (4 \ 3 \ 6 \ 1) \quad Y_t = \{1 \ 3 \ 7 \ 8\} \quad \text{loss}(Y_t, \hat{Y}_t)
\]

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>V</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>V</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[
\text{Hamming loss}(Y_t, \hat{Y}_t) = |Y_t \setminus \hat{Y}_t| + |\hat{Y}_t \setminus Y_t| = 2 + 2 = 4
\]

\[
\text{Ranking loss}(Y_t, \hat{Y}_t) = 4 + 2 = 6
\]

\(Y_t\ is\ NOT\ ordered\)  
# “flipped” couples in \(\hat{Y}_t\)  
(1-4), (1-6), (3-4), (3-6)  
false negative
The probability to click on a banner is given by a Generalized Linear Model

\[ p_{it} = \mathbb{P}(y_{it} = 1 \mid \mathbf{x}_t) = p(u_i^\top \mathbf{x}_t) \quad i = 1 \ldots K \]

Learn approximate probability and uncertainty with a second-order algorithm

\[ \hat{p}_{it} = p( w_{it}^\top \mathbf{x}_t + \epsilon_{it} ) \]

Online regret analysis: with \( K \) classes and \( S \) predictions per round, with probability at least 1-\( \delta \)

\[ R_T = O \left( \sqrt{SKT \log(T/\delta)} \right) \]
Experiments: performance comparable to full-information