Bandit Structured Prediction for Learning from Partial Feedback in Statistical Machine Translation

Changzhi Sun

- Expected Loss Minimization under Full Information

$$\mathbb{E}_{p(x,y)p_w(y'|x)}\left[\Delta_y(y')\right] = \sum_{x,y} p(x,y) \sum_{y' \in \mathcal{Y}(x)} \Delta_y(y') p_w(y'|x)$$

- minimum Bayes risk principle

$$\hat{y}_w(x) = \arg\min_{y \in \mathcal{Y}(x)} \sum_{y' \in \mathcal{Y}(x)} \Delta_y(y') p_w(y'|x).$$

- if $\Delta_y(y') = \mathbf{1}[y \neq y']$, MAP

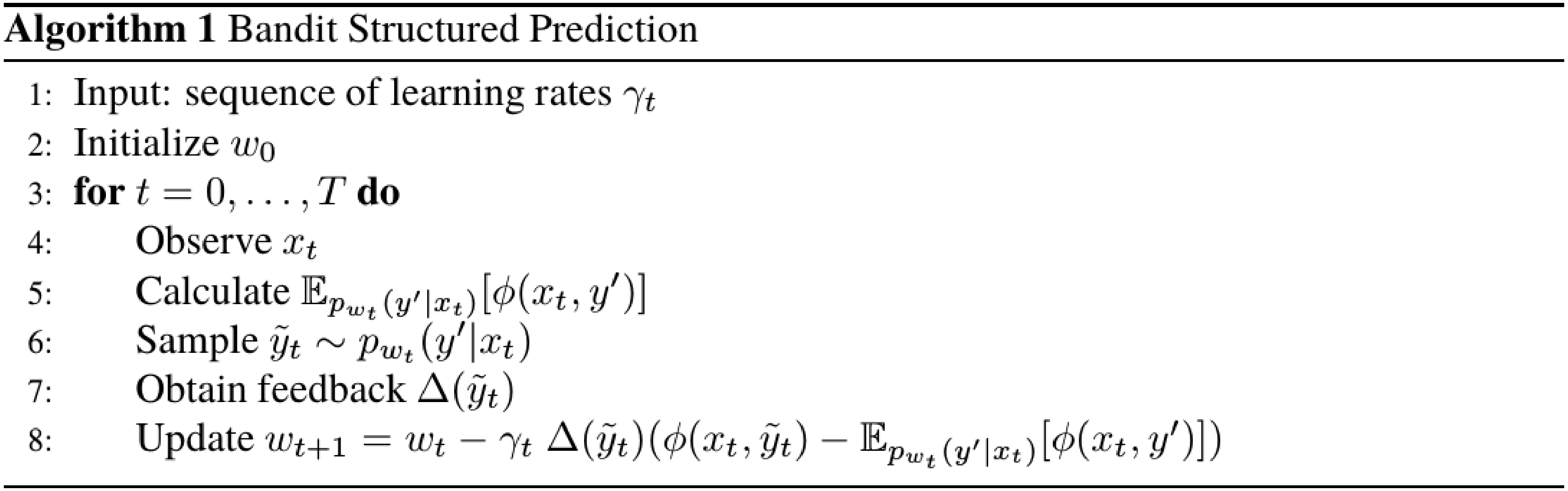
$$\begin{aligned}\hat{y}_w(x) &= \arg\max_{y \in \mathcal{Y}(x)} p_w(y|x) \\ &= \arg\max_{y \in \mathcal{Y}(x)} w^\top \phi(x,y).\end{aligned}$$

- $p(x, y)$ is approximated by $\tilde{p}(x, y) = \frac{1}{T} \sum_{t=0}^{T} \mathbf{1}[x = x_t]\mathbf{1}[y = y_t]$
- training data $\{(x_t, y_t)\}_{t=0}^{T}$

$$\mathbb{E}_{\tilde{p}(x,y)p_w(y'|x)}\left[\Delta_y(y')\right] = \frac{1}{T}\sum_{t=0}^{T}\sum_{y'\in\mathcal{Y}(x_t)} \Delta_{y_t}(y')p_w(y'|x_t).$$

$$\begin{aligned}
&\nabla\mathbb{E}_{\tilde{p}(x,y)p_w(y'|x)}\left[\Delta_y(y')\right] \\
&\quad = \mathbb{E}_{\tilde{p}(x,y)}\Big[\mathbb{E}_{p_w(y'|x)}[\Delta_y(y')\phi(x, y')] - \mathbb{E}_{p_w(y'|x)}[\Delta_y(y')]\ \mathbb{E}_{p_w(y'|x)}[\phi(x, y')]\Big] \\
&\quad = \mathbb{E}_{\tilde{p}(x,y)p_w(y'|x)}\Big[\Delta_y(y')(\phi(x, y') - \mathbb{E}_{p_w(y'|x)}[\phi(x, y')])\Big].
\end{aligned}$$

- Bandit Structured Prediction

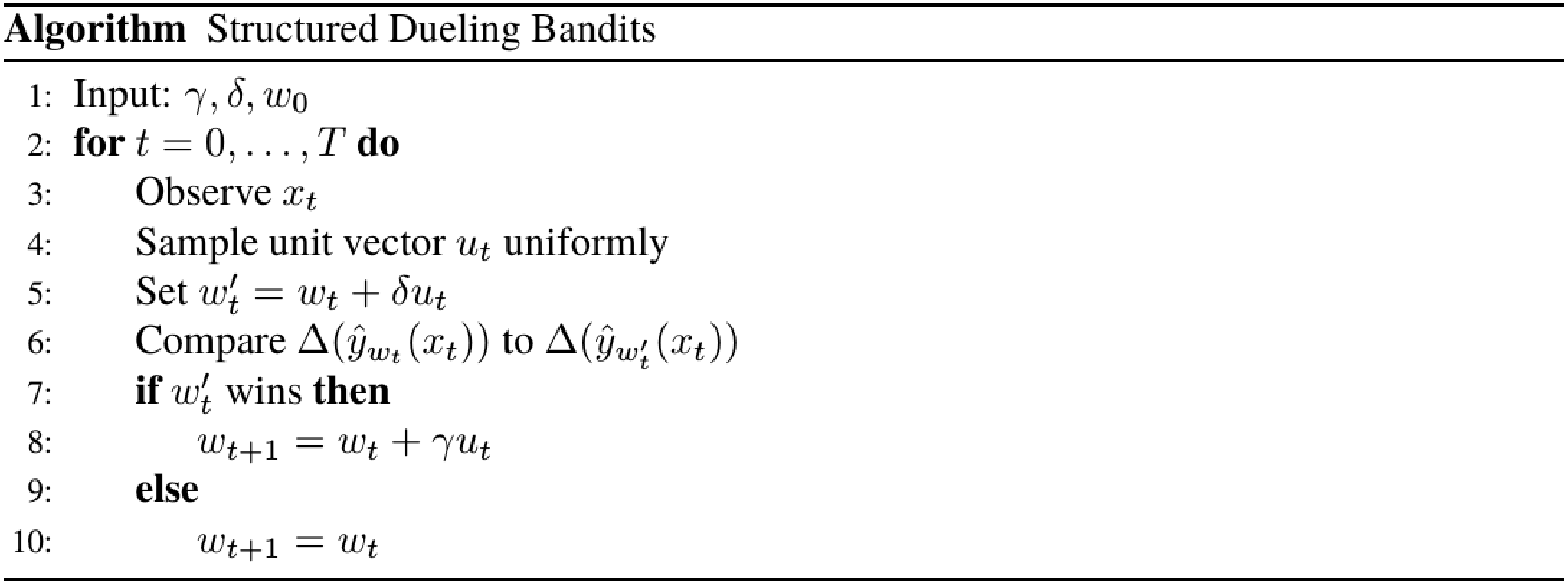


Algorithm 1 Bandit Structured Prediction

```
Input: sequence of learning rates $\gamma_t$
Initialize $w_0$
for $t = 0, \ldots, T$ do
    Observe $x_t$
    Calculate $\mathbb{E}_{p_{w_t}(y'|x_t)}[\phi(x_t, y')]$
    Sample $\tilde{y}_t \sim p_{w_t}(y'|x_t)$
    Obtain feedback $\Delta(\tilde{y}_t)$
    Update $w_{t+1} = w_t - \gamma_t \, \Delta(\tilde{y}_t)(\phi(x_t, \tilde{y}_t) - \mathbb{E}_{p_{w_t}(y'|x_t)}[\phi(x_t, y')])$
```

$$
\begin{aligned}
J(w) &= \mathbb{E}_{p(x)p_w(y'|x)} \left[\Delta(y')\right] \\
&= \sum_{x} p(x) \sum_{y' \in \mathcal{Y}(x)} \Delta(y') p_w(y'|x).
\end{aligned}
$$

- Structured Dueling Bandits



Algorithm Structured Dueling Bandits

```
Input: $\gamma, \delta, w_0$
for $t = 0, \dots, T$ do
    Observe $x_t$
    Sample unit vector $u_t$ uniformly
    Set $w'_t = w_t + \delta u_t$
    Compare $\Delta(\hat{y}_{w_t}(x_t))$ to $\Delta(\hat{y}_{w'_t}(x_t))$
    if $w'_t$ wins then
        $w_{t+1} = w_t + \gamma u_t$
    else
        $w_{t+1} = w_t$
```

Discussion

- first proposed **Bandit Structured Prediction**
- **single** point feedback