Selective Transfer Between Learning Tasks Using Task-Based Boosting
Supplementary Materials

Eric Eaton
Bryn Mawr College
Computer Science Department
Bryn Mawr, PA USA
eeaton@cs.brynmawr.edu

Marie desJardins
University of Maryland Baltimore County
Department of Computer Science and Electrical Engineering
Baltimore, MD USA
mariedj@umbc.edu

This supplement to the paper "Selective Transfer Between Learning Tasks Using Task-Based Boosting," which appeared in *Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence* (AAAI'11), provides the proof to Theorem 1 stated in the paper:

Theorem 1. *The training error* $\epsilon_T = \frac{1}{|T|}\left|\{j : H(x_j) \neq y_j\}\right|$ *on the target task for TransferBoost is bounded by*

$$\epsilon_T \leq \frac{|D|}{|T|} \prod_{t=1}^{K} Z_t \left(\sum_{j \in T} w_{K+1}(x_j) \right) \ .$$

Proof.

This proof generally follows Schapire and Singer's (1998) method for bounding AdaBoost's training error.

Let $f(x) = \sum_{t=1}^{K} \beta_t h_t(x)$, so that $H(x) = sign(f(x))$. The update rule (Algorithm 1, line 7) can be unraveled to determine the instance weights after the last boosting iteration K. Let $S_0 = T$ and $\alpha_t^0 = 0$. The update rule can be concisely rewritten for $x_j \in S_i$ as:

$$w_{t+1}(x_j) = \frac{w_t(x_j) \exp(-\beta_t y_j h_t(x_j) + \alpha_t^i)}{Z_t} \ . \quad (1)$$

Repeatedly applying the update rule for $t = 1, \ldots, K$ yields

$$\begin{aligned} w_{K+1}(x_j) &= w_0(x_j) \prod_{t=1}^{K} \frac{\exp\left(-\beta_t y_j h_t(x_j) + \alpha_t^i\right)}{Z_t} \\ &= \frac{\exp\left(-\sum_{t=1}^{K} \beta_t y_j h_t(x_j) + \sum_{t=1}^{K} \alpha_t^i\right)}{|D| \prod_{t=1}^{K} Z_t} \\ &= \frac{\exp\left(-y_j f(x_j)\right) \exp\left(\sum_{t=1}^{K} \alpha_t^i\right)}{|D| \prod_{t=1}^{K} Z_t} \ . \end{aligned}$$

It follows that

$$\begin{aligned} &\sum_{i=0}^{k} \sum_{j \in S_i} \exp\left(-y_j f(x_j)\right) \exp\left(\sum_{t=1}^{K} \alpha_t^i\right) \\ &\quad = \sum_{i=0}^{k} \sum_{j \in S_i} |D| w_{K+1}(x_j) \prod_{t=1}^{K} Z_t \\ &\sum_{i=0}^{k} \exp\left(\sum_{t=1}^{K} \alpha_t^i\right) \sum_{j \in S_i} \exp\left(-y_j f(x_j)\right) \\ &\quad = |D| \prod_{t=1}^{K} Z_t \left(\sum_{i=0}^{k} \sum_{j \in S_i} w_{K+1}(x_j) \right) \\ &\quad = |D| \prod_{t=1}^{K} Z_t(1) \ . \end{aligned}$$

Expanding the LHS and subtracting the portion due to the source tasks yields

$$\begin{aligned} \sum_{j \in T} \exp\left(-y_j f(x_j)\right) &= |D| \prod_{t=1}^{K} Z_t \\ &\quad - \sum_{i=1}^{k} \exp\left(\sum_{t=1}^{K} \alpha_t^i\right) \sum_{j \in S_i} \exp\left(-y_j f(x_j)\right) \ . \quad (2) \end{aligned}$$

Schapire and Singer (1998) note that $[\![H(x_j) \neq y_j]\!] \leq \exp(-y_j f(x_j))$, where $[\![\pi]\!]$ is 1 if predicate π holds and 0 otherwise (since $H(x_j) \neq y_j \Rightarrow y_j f(x_j) \leq 0 \Rightarrow exp(-y_j f(x_i)) \geq 1$). Since $\exp\left(\sum_{t=1}^{K} \alpha_t^i\right) \geq 0$, it follows that

$$\begin{aligned} [\![H(x_j) \neq y_j]\!] &\leq \exp(-y_j f(x_j)) \\ \sum_{j \in T} [\![H(x_j) \neq y_j]\!] &\leq \sum_{j \in T} \exp(-y_j f(x_j)) \ . \quad (3) \end{aligned}$$

By combining Equations 2 and 3,

$$\begin{aligned}
&\sum_{j \in T} [\![H(x_j) \neq y_j]\!] \\
&\quad \leq |D| \prod_{t=1}^{K} Z_t - \sum_{i=1}^{K} \exp\left(\sum_{t=1}^{K} \alpha_t^i \right) \sum_{j \in S_i} \exp\left(-y_j f(x_j)\right) \\
&\quad \leq |D| \prod_{t=1}^{K} Z_t - \sum_{i=1}^{K} \sum_{j \in S_i} \left(|D| w_{K+1}(x_j) \prod_{t=1}^{K} Z_t \right) \\
&\quad \leq |D| \prod_{t=1}^{K} Z_t - |D| \prod_{t=1}^{K} Z_t \left(\sum_{i=1}^{K} \sum_{j \in S_i} w_{K+1}(x_j) \right) \\
&\quad \leq |D| \prod_{t=1}^{K} Z_t \left(1 - \sum_{i=1}^{K} \sum_{j \in S_i} w_{K+1}(x_j) \right)
\end{aligned}$$

$$\frac{1}{|T|} \sum_{j \in T} [\![H(x_j) \neq y_j]\!] \leq \frac{|D|}{|T|} \prod_{t=1}^{K} Z_t \left(\sum_{j \in T} w_{K+1}(x_j) \right) \ .$$

The theorem follows directly. □

References

Schapire, R., and Singer, Y. 1998. Improved boosting algorithms using confidence-rated predictions. *COLT*, 80–91.