Supplementary Materials:
Histopathological Image Analysis via Active Learning

Yan Zhu\textsuperscript{1}, Wei Liu\textsuperscript{2}, Shaoting Zhang\textsuperscript{3}, and Dimitris N Metaxas\textsuperscript{1}

\textsuperscript{1} Department of Computer Science, Rutgers University, Piscataway, NJ, USA
\textsuperscript{2} IBM T. J. Watson Research Center, NY, USA
\textsuperscript{3} Department of Computer Science, University of North Carolina at Charlotte, NC, USA

1 Proof of Theorem 1

Lemma 1. (from Lemma 3 in \cite{1}) Let $V = \{1, ..., n\}$, $Y$ be finite sets; $f : 2^V \times Y \rightarrow \mathbb{N}$ monotonic and submodular, and $P(Y)$ such that $(f, P)$ is adaptive submodular. Let $A_1, A_2, ..., A_m \subseteq V$, and define for $i \in \{1, ..., m\}$, $Z_i = \{y_{j_1}, ..., y_{j_l}\}$ where $A_i = \{j_1, ..., j_l\}$, and $l$ is a constant integer. Let $W = \{1, ..., m\}$ and $Q(Z_W)$ be the distribution over $Z_1, Z_2, ..., Z_m$ induced by $P$. Let $Y' = \bigcup_{i \in W} \text{range}(Z_i)$. Define the function

$$\gamma : 2^W \times Y' \rightarrow 2^V \times Y, \gamma(\{(a_1, z_1), ..., (a_t, z_t)\}) = \bigcup_{j=1}^t \{(i, o) : i \in A_j, o = [z_j]_i\}$$

and define $g : 2^W \times Y' \rightarrow \mathbb{N}$ by $g(S) = f(\gamma(S))$. Then $g$ is submodular, and $(g, Q)$ is adaptive submodular.

Lemma 2. (from Theorem 7 in \cite{2}) For an adaptive monotonic submodular function $f : 2^E \times Y^E \rightarrow \mathbb{R}_{\leq 0}$ and a $p$-independent system $(E, I)$. Fix a policy $\pi$ which is $\alpha$-approximate greedy with respect to $f$ for constraint $I$. Then $\pi$ yields an $\frac{\alpha}{p+\alpha}$ approximation, meaning

$$f_{\text{avg}}(\pi) \leq \left(\frac{\alpha}{p+\alpha}\right) \max_{\text{feasible } \pi^*} f_{\text{avg}}(\pi^*)$$

where $\pi^*$ is feasible iff $E(\pi^*, \Phi) \in I$ for all $\Phi$.

Below is the proof of theorem 1. We adopt the similar proving technique as \cite{1}. Basically, we transfer from a batch mode policy for the original problem to a sequential policy to the superset of the original problem instance.

Proof of Theorem 1

Suppose we are given $f$, $V$, $Y$ and $P$ satisfying Lemma 1. Also we are given a set of disjoint ground sets $\mathcal{P}_1, \mathcal{P}_2, ..., \mathcal{P}_n$ partitioning $V$, therefore it gives a partition matroid constraint $\mathcal{M}$. Let $\{S_1, ..., S_M\}$ are the superset of all possible size $k$ subsets, where $M = \binom{n}{k}$. According to Lemma 1, an induced problem instance for $\{S_1, ..., S_M\}$ is $(g, Q)$, where $Q$ is the distribution of the observations for all possible size $k$ subsets $\{S_1, ..., S_M\}$. From Lemma 1, $(g, Q)$ is adaptive submodular. For every batch mode
policy for problem \( (f, P) \) subject to \( M \), there is a corresponding sequential policy for problem \( (g, Q) \) subject to \( M \).

According to Theorem 11 in [3], the greedy policy \( \pi \) satisfies

\[
cost_{\text{avg}}(\pi) \leq OPT_{\text{avg},k}(\ln(|H| - 1) + 1)
\]

where \(|H|\) is the size of the hypothesis space, and \( OPT_{\text{avg},k} \) is the optimal policy for size \( k \) batch selection. However, policy \( \pi \) is assuming that within each batch the selection is optimal. The proposed algorithm BGAL-PGM greedily select samples within each batch. Notice that a partition matroid constraint is a special case of \( p \)-independent system when \( p = 1 \). So According to Lemma 2, the policy adopting BGAL-PMC maximizes function \( g \) with a \( \frac{1}{2} \)-approximation to the optimal policy. Therefore, we prove that

\[
cost_{\text{avg}}(\pi_{\text{BGAL-PMC}}) \leq OPT_{\text{avg},k} \times 2 \times (\ln(|H| - 1) + 1)
\]

as stated in Theorem 1.

References