Supplementary Materials: Histopathological Image Analysis via Active Learning

Yan Zhu[1], Wei Liu[2], Shaoting Zhang[3], and Dimitris N Metaxas[1]

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

1 Proof of Theorem 1

Lemma 1. *(from Lemma 3 in [1]) Let $\mathcal{V} = \{1, ..., n\}$, $\mathcal{Y}$ be finite sets; $f : 2^{\mathcal{V}\times\mathcal{Y}} \rightarrow \mathbb{N}$ monotonic and submodular, and $P(\mathbf{Y}_{\mathcal{V}})$ such that (f, P) is adaptive submodular. Let $\mathcal{A}_1, \mathcal{A}_2, ..., \mathcal{A}_m \subseteq \mathcal{V}$, and define for $i \in \{1, ..., m\}$, $Z_i = [\mathbf{Y}_{j_1,...,j_l}]$ where $\mathcal{A}_i = \{j_1, ..., j_l\}$, and l is a constant integer. Let $\mathcal{W} = \{1, ..., m\}$ and $Q(\mathbf{Z_W})$ be the distribution over $Z_1, Z_2, ..., Z_m$ induced by P. Let $\mathcal{Y}' = \bigcup_{i\in\mathcal{W}} range(Z_i)$. Define the function*

$$\gamma : 2^{\mathcal{W}\times\mathcal{Y}'} \rightarrow 2^{\mathcal{V}\times\mathcal{Y}}, \gamma(\{(a_1, \mathbf{z}_1), ..., (a_t, \mathbf{z}_t)\}) = \bigcup_{j=1}^{t} \{(i, o) : i \in A_j, o = [\mathbf{z}_j]_i\} \quad (1)$$

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

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

$$f_{avg}(\pi) \leq (\frac{\alpha}{p+\alpha}) \max_{feasible \pi^*} f_{avg}(\pi^*) \quad (2)$$

where π^* is feasible iff $E(\pi^*, \Phi) \in \mathcal{I}$ for all Φ.

Below is the proof of theorem 1. We adopt the similar proving technique as [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 $f, \mathcal{V}, \mathcal{Y}$ and P satisifying Lemma 1. Also we are given a set of disjoint ground sets $\mathcal{P}_1, \mathcal{P}_2, ..., \mathcal{P}_n$ partitioning $\mathcal{V}$, 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 $\mathcal{M}$, there is a corresponding sequential policy for problem (g, Q) subject to $\mathcal{M}$.

According to Theorem 11 in [3], the greedy policy π satisfies

$$cost_{avg}(\pi) \leq OPT_{avg,k}(ln(|\mathcal{H}| - 1) + 1) \tag{3}$$

where $|\mathcal{H}|$ is the size of the hypothesis space, and $OPT_{avg,k}$ is the optimal policy for size k batch selection. However, policy π is assuming that within each batch the seelection 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 systemm 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

$$cost_{avg}(\pi_{BGAL-PMC}) \leq OPT_{avg,k} \times 2 \times (ln(|\mathcal{H}| - 1) + 1) \tag{4}$$

as stated in Theorem 1. ■

References

1. Y. Chen and A. Krause. Near-optimal batch mode active learning and adaptive submodular optimization. In *Proceedings of The 30th International Conference on Machine Learning*, pages 160–168, 2013.
2. D. Golovin and A. Krause. Adaptive submodular optimization under matroid constraints. *arXiv preprint arXiv:1101.4450*, 2011.
3. D. Golovin and A. Krause. Adaptive submodularity: Theory and applications in active learning and stochastic optimization. *Journal of Artificial Intelligence Research*, 42(1):427–486, 2011.