The Adaptive Complexity of Maximizing a Submodular Function

Abstract:

In this paper we study the adaptive complexity of submodular optimization. Informally, the adaptive complexity of a problem is the minimal number of sequential rounds required to achieve a constant factor approximation when polynomially-many queries can be executed in parallel at each round. Adaptivity is a fundamental concept that is heavily studied in computer science, largely due to the need for parallelizing computation. Somewhat surprisingly, very little is known about adaptivity in submodular optimization. For the canonical problem of maximizing a monotone submodular function under a cardinality constraint, to the best of our knowledge, all that is known to date is that the adaptive complexity is between 1 and $\Omega(n)$. Our main result in this paper is a tight characterization showing that the adaptive complexity of maximizing a monotone submodular function under a cardinality constraint is, up to lower order terms, $\Theta(\log n)$: We describe an algorithm which requires $O(\log n)$ sequential rounds and achieves an approximation that is arbitrarily close to $1/3$; We show that no algorithm can achieve an approximation better than $O(1/\log n)$ with fewer than $O(\log n / \log \log n)$ rounds. Thus, when allowing for parallelization, our algorithm achieves a constant factor approximation exponentially faster than any known existing algorithm for submodular maximization. Importantly, the approximation algorithm is achieved via adaptive sampling and complements a recent line of work on optimization of functions learned from data. In many cases, we do not know the functions we optimize and learn them from labeled samples. Recent results show that no algorithm can obtain a constant factor approximation guarantee using polynomially-many labeled samples as in the PAC and PMAC models, drawn from any distribution. Since learning with non-adaptive samples over any distribution results in a sharp impossibility, we consider learning with adaptive samples where the learner obtains $\text{poly}(n)$ samples drawn from a distribution of her choice in every round. Our result implies that in the realizable case, where there is a true underlying function generating the data, $\hat{1}(\log n)$ batches, up to lower order terms, of adaptive samples are necessary and sufficient to approximately "learn to optimize" a monotone submodular function under a cardinality constraint. This is joint work with Yaron Singer.