The Communication Complexity of Set Intersection under Product Distributions

Wednesday, July 12, 2023 11:20 AM (20 minutes)

Rotem Oshman and Tal Roth

Abstract: We consider a multiparty setting where k parties have private inputs $X_1, \ldots, X_k \subseteq [n]$ and wish to compute the intersection $\bigcap_{\ell=1}^k X_\ell$ of their sets, using as little communication as possible. This task generalizes the well-known problem of set disjointness, where the parties are required only to determine whether the intersection is empty or not.

In the worst-case, it is known that the communication complexity of finding the intersection is the same as that of solving set disjointness, regardless of the size of the intersection: the cost of both problems is $\Omega(n \log k + k)$ bits in the shared blackboard model, and $\Omega(nk)$ bits in the coordinator model.

In this work we consider a realistic setting where the parties' inputs are independent of one another, that is, the input is drawn from a product distribution. We show that this makes finding the intersection significantly easier than in the worst-case: only $\tilde{\Theta}((n^{1-1/k} (H(S) + 1)^{1/k}) + k))$ bits of communication are required, where H(S) is the Shannon entropy of the intersection S. We also show that the parties do not need to exactly know the underlying input distribution; if we are given in advance $O(n^{1/k})$

samples from the underlying distribution μ , we can learn enough about μ to allow us to compute the intersection of an input drawn from μ using expected communication $\tilde{\Theta}((n^{1-1/k}E[|S|]^{1/k}) + k)$, where |S| is the size of the intersection.

Presenter: ROTH, Tal

Session Classification: Track A-4