

Homework Four

Rob Phillips

May 13, 2019

Paper One

Title: The Communication Complexity of Distributed ϵ -Approximations

Abstract: Data summarization is an effective approach to dealing with the “big data” problem. While data summarization problems traditionally have been studied in the streaming model, the focus is starting to shift to distributed models, as distributed/parallel computation seems to be the only viable way to handle today’s massive data sets. In this paper, we study epsilon-approximations, a classical data summary that, intuitively speaking, preserves approximately the density of the underlying data set over a certain range space. We consider the problem of computing epsilon-approximations for a data set which is held jointly by k players, and give general communication upper and lower bounds that hold for any range space whose discrepancy is known.

Violations found:

- Violates Knuth’s rule in the title – no symbols should be used.
- “While data summarization problems traditionally have been studied in the streaming model” is grammatically incorrect.
- Uses the symbol for epsilon in title, but spells it out in the abstract – inconsistent usage.
- “preserves approximately” might sound better as “approximately preserves.”
- “which is held jointly by” could be simply “held jointly by.”
- “intuitively speaking” should be cut out. Might be personal preference, but it sounds presumptuous and choppy.
- Italicize the k in “ k players.”

Suggested revision:

Title: The Communication Complexity of Distributed Epsilon-Approximations

Abstract: Data summarization is an effective approach to dealing with the “big data” problem. While data summarization problems traditionally have been studied under the streaming model,

the focus is starting to shift to distributed models, as distributed and parallel computation seems to be the only viable way to handle today’s massive data sets. In this paper, we study epsilon-approximations, a classical data summary that approximately preserves the density of the underlying data set over a certain range space. We consider the problem of computing epsilon-approximations for a data set held jointly by k players, and give general communication upper and lower bounds that hold for any range space whose discrepancy is known.

Paper Two

Title: $(2 + \epsilon)$ -SAT is NP-hard

Abstract: We prove the following hardness result for a natural promise variant of the classical CNF-satisfiability problem: Given a CNF-formula where each clause has width w and the guarantee that there exists an assignment satisfying at least $g = \lceil \frac{w}{2} \rceil - 1$ literals in each clause, it is NP-hard to find a satisfying assignment to the formula (that sets at least one literal to true in each clause). On the other hand, when $g = \lceil \frac{w}{2} \rceil$, it is easy to find a satisfying assignment via simple generalizations of the algorithms for 2-SAT.

Viewing 2-SAT \in P as easiness of SAT when 1-in-2 literals are true in every clause, and NP-hardness of 3-SAT as intractability of SAT when 1-in-3 literals are true, our result shows, for any fixed $\epsilon > 0$, the hardness of finding a satisfying assignment to instances of “ $(2 + \epsilon)$ -SAT” where the density of satisfied literals in each clause is promised to exceed $\frac{1}{2+\epsilon}$.

We also strengthen the results to prove that given a $(2k + 1)$ -uniform hypergraph that can be 2-colored such that each edge has perfect balance (at most $k + 1$ vertices of either color), it is NP-hard to find a 2-coloring that avoids a monochromatic edge. In other words, a set system with discrepancy 1 is hard to distinguish from a set system with worst possible discrepancy.

Finally, we prove a general result showing intractability of promise CSPs based on the paucity of certain “weak polymorphisms.” The core of the above hardness results is the claim that the only weak polymorphisms in these particular cases are juntas depending on few variables.

Violations found:

- Violates Knuth’s rule in the title – no symbols should be used.
- Certainly violates the “short and self-contained” guideline, and it is outside the range of one to three paragraphs.
- Contains numerous formulae throughout the abstract, overusing the “only when needed” rule.
- Last paragraph contains a statement that does not suggest what the conclusions are.

Suggested revision:

Title: Variant of classical CNF-satisfiability problem is NP-hard

Abstract: We prove the following hardness result for a natural promise variant of the classical CNF-satisfiability problem: Given a CNF-formula where each clause has width w and the guarantee that there exists an assignment satisfying at least half of the width w in each clause, it is NP-hard to find a satisfying assignment to the formula (that sets at least one literal to true in each clause). On the other hand, when there exists an assignment that is exactly half the width, it is easy to find a satisfying assignment via simple generalizations of the algorithms for 2-SAT.

We also strengthen the results to prove that given a uniform hypergraph that can be two-colored such that each edge has perfect balance, it is NP-hard to find a two-coloring that avoids a monochromatic edge. In other words, a set system with a discrepancy of one is hard to distinguish from a set system with worst possible discrepancy.