Evaluation of DNF Formulas

Piotr Wojciechowski
LCSEE,
West Virginia University,
Morgantown, WV
{pwojciec@mix.wvu.edu}

Zola Donovan
Department of Mathematics,
West Virginia University,
Morgantown, WV
{zdonovan@mix.wvu.edu}

Abstract

While the Stochastic Boolean Function Evaluation (SBFE) problem is inapproximable for general DNF formulas, Allen et al. give exact and approximate algorithms for monotone k-DNF and monotone k-term DNF formulas. In the SBFE problem for DNF formulas, the value of a DNF formula on an initially unknown input, $x = (x_1, ..., x_n)$, must be determined when: there is a cost, c_i, associated with obtaining the value of x_i; each x_i is equal to 1 with known probability, p_i; and, the x_i are independent. The goal is to minimize expected cost. Allen et al. also prove a lower bound result for evaluation of monotone CDNF formulas.

1 Introduction

This paper studies the complexity of the SBFE problem, and considers both exact and approximate versions of the problem for k-DNF and k-term DNF formulas. It is known that the general SBFE problem is NP-hard for arbitrary DNF formulas since satisfiability is NP-hard (Greiner et al. 2006). To show that the SBFE problem for k-DNF is NP-hard, even for $k = 2$, a simple reduction is used. A $\frac{4}{\rho}^k$ factor algorithm for evaluating monotone k-DNF is presented, along with a proof showing that the SBFE problem for monotone k-term DNF is solvable in polynomial time for some constant k.

Additionally, an approximation algorithm solving the SBFE problem for CDNF formulas (and decision trees) for the special case of unit costs, the uniform distribution, and monotone CDNF formulas is given (Kaplan et al. 2005). For k terms of the DNF, and d clauses in the CNF, the algorithm achieves an approximation guarantee of $O(\log kd)$ on the value of the expected certificate cost–a lower bound on the optimal solution. The algorithm alternates between two processes; one process attempts to achieve a 0-certificate, while the other attempts to achieve a 1-certificate. This ”round robin” technique is modified to handle arbitrary costs with no change in the approximation factor, as the original algorithm handles only unit costs.

Finally, it is shown that the approximation guarantee of $O(\log kd)$, is close to optimal. The approximation factor must be $\Omega((\log kd)^{\varepsilon})$, for $0 < \varepsilon < 1$; which also implies that the (optimal) average depth of a decision tree computing a Boolean function can be exponentially larger than the average certificate size for that function, while the depth complexity of a decision tree for a function, (worst-case) is at most quadratic. (Buhrman and Wolf 1999)

2 Statement of Problem

The Stochastic Boolean Function Evaluation (SBFE) problem is defined as follows. You are given a boolen formula $\phi(x_1, \ldots, x_n)$, the costs $c_1, \ldots, c_n \geq 0$ of determining the value of each x_i, and the probabilities $0 < p_1, \ldots, p_n < 1$ that each x_i is **True**. The goal of the problem is to determine, at minimum cost, the value of $\phi(\mathbf{x})$ for an unknown value of $\mathbf{x}$. This $\mathbf{x}$ is randomly generated so that each x_i is independent and $P[x_i = \mathbf{True}] = p_i$. While $\mathbf{x}$ remains initially unknown its value can be revealed at a cost of c_i.

3 Motivation and Related Work

Stochastic Boolean Function Evaluation is applicable in the field of medicine. In this case the x_is correspond to medical tests performed on a given patient, and the boolean function $\phi(\mathbf{x})$ evaluates to **True** is the patient has a certain disease. SBFE can also be applied to query optimization in databases, where all stored values of $\mathbf{x}$ that satisfy $\phi(\mathbf{x})$ need to be located.

Certain special cases of the SBFE problem can be solved exactly in polynomial time. These include including read-once DNF formulas and k-of-n formulas [Ünl04]. An approximation factor of n can be obtained for arbitrary boolean formulas by testing the variables in increasing order of their costs. [Kap05]

Deshpande et al. explored a generic approach to developing approximation algorithms for SBFE problems, called the Q-value approach. It involves reducing the problem to an instance of Stochastic Submodular Set Cover. They proved that the Q-value approach does not yield a sublinear approximation bound for evaluating k-DNF formulas, even for $k = 2$. They also developed a new algorithm for solving Stochastic Submodular Set Cover, called Adaptive Dual Greedy, and used it to obtain a 3-approximation algorithm solving the SBFE problem for linear threshold formulas. [Des13]

4 Critique

5 Conclusions

Thorough reduction to tautology of DNF formulas the general case of the SBFE problem is inapproximable unless $\mathbf{P} = \mathbf{NP}$. However when dealing with monotone k-DNF formulas the SBFE problem can be approximated in polynomial time with an approximation bound of $\frac{4}{\rho}^k$. Monotone k-term DNF formulas also have a polynomial time approximation algorithm, however, this algorithm has an approximation factor of $\max\{2 \cdot k, \frac{2}{\rho} \cdot (1 + \ln k)\}$.

References

[All13] Allen, S.; Hellerstein, L.; Kletenik, D.; and Ünlüyurt, T. 2013. Evaluation of DNF formulas. http://arxiv.org/abs/1310.3673

[Buh02] Buhrman, H., and Wolf, R. D. 1999. Complexity measures and decision tree complexity: A survey. *Theoretical Computer Science* 288:2002.

[Des13] Deshpande, A.; Hellerstein, L.; and Kletenik, D. 2013. Approximation algorithms for stochastic boolean function evaluation and stochastic submodular set cover. http://arxiv.org/abs/1303.0726

[Gre06] Greiner, R.; Hayward, R.; Jankowska, M.; and Molloy, M. 2006. Finding optimal satisficing strategies for and-or trees. *Artif. Intell.* 170(1):19-58.

[Kap05] Kaplan, H.; Kushilevitz, E.; and Mansour, Y. 2005. Learning with attribute costs. *STOC*, 356365.

[Ünl04] Ünlüyurt, T. 2004. Sequential testing of complex systems: a review. *Discrete Applied Mathematics* 142(1-3):189205.