RA-003

A Counting-Based Approximation the Distribution Function of the Longest Path Length in Directed Acyclic Graphs

Ei Ando, Hirotaka Ono, Kunihiko Sadakane and Masafumi Yamashita

Dept. Computer Sci. and Communication Eng., Kyushu University, 744 Motooka, Nishi-ku, Fukuoka, Fukuoka 819-0395, Japan.

Abstract. We consider a stochastic version of the longest path problem in DAGs and propose an algorithm that approximates the distribution function of the longest path length by utilizing (approximate) counting. In general, the stochastic longest path problem has several difficulties to solve; (1) we have no general method to calculate the exact distribution function $F_{\mathrm{MAX}}(x)$ of the longest path length, and even if we could do that, (2) the distribution function often takes a complicated form. For this problem, it is natural to consider calculating a simpler approximation of $F_{\mathrm{MAX}}(x)$ and give a certain approximation guarantee. We show an algorithm that gives a bound on $F_{\mathrm{MAX}}(x)$ by approximating $\ln|\mathcal{P}|$, where $\mathcal{P}$ is the family of paths. The output of our algorithm $A(a)$ satisfies that $A(a) \geq F_{\mathrm{MAX}}^{-1}(a)$ for $a \geq 2^{-|\mathcal{P}|}$, and the approximation factor is $O((\sigma/\mu)\sqrt{C/h})$ for $a \geq 1/2$, where μ and σ^2 are, respectively, the minimum mean and the maximum variance of the edge lengths, C is an upper bound on $\ln|\mathcal{P}|$ and h is the maximum number of edges in $P \in \mathcal{P}$. The running time of our algorithm is $O(|E|)$, where $|E|$ is the number of edges in a DAG.

1 Introduction

1.1 Problem and Results

We consider a stochastic version of the longest path problem in DAGs and propose an algorithm that approximates the distribution function of the longest path length by utilizing approximate counting. Let $G = (V, E)$ be a directed graph and each edge $e \in E$ is associated with an edge length X_e. In the longest path problem, we are to obtain $\max_{P\in\mathcal{P}} \{\sum_{e\in P} X_e\}$ where $\mathcal{P} \subseteq 2^E$ is the family of paths in G. In a stochastic version of the longest path problem, each edge length is given as a random variable. Then the longest path length of the problem is also a random variable. We consider the problem to find the distribution function of $\max_{P\in\mathcal{P}}\{\sum_{e\in P} X_e\}$. In this paper, we assume that the edge lengths are mutually independent and normally distributed. Although we deal with the longest path problem in this paper for simplicity, we can construct a similar algorithm for the shortest path problem by calculating the complementary distribution functions instead of the distribution functions.

Let $F_{\mathrm{MAX}}(x)$ be the distribution function of the longest path length. Then a stochastic version of the longest path problem is to calculate distribution function $F_{\mathrm{MAX}}(x)$. In general, the stochastic longest path problem has several difficulties to solve; (1) we have no general method to calculate the exact distribution function $F_{\mathrm{MAX}}(x)$, and even if we could do that, (2) the distribution function often takes a complicated form. For these problems, it is natural to consider calculating a simpler approximation of $F_{\mathrm{MAX}}(x)$ and give a certain approximation guarantee. There are roughly two standpoints of approximating a function by another function: (1) fitting: find a function that minimizes "errors" from the original function by adjusting its parameters and (2) bounding: find upper and lower bounds on the original function such that each of them is as close to the other as possible. In this paper, we take the latter standpoint to figure out the behaviour of $F_{\mathrm{MAX}}(x)$. A half of the problem, obtaining an upper bound on $F_{\mathrm{MAX}}(x)$, can be easily solved, because the distribution function $F_P(x)$ of $\sum_{e\in P} X_e$ for any $P \in \mathcal{P}$ is an upper bound on $F_{\mathrm{MAX}}(x)$. However the other half of the problem, obtaining a lower bound on $F_{\mathrm{MAX}}(x)$, is not trivial. Hence we concentrate on obtaining a lower bound on $F_{\mathrm{MAX}}(x)$.

In this paper, we show an approximation algorithm that gives a function by utilizing an upper bound C on $\ln|\mathcal{P}|$ and the function bounds $F_{\mathrm{MAX}}(x)$ from below. The output $A(a)$ of our algorithm satisfies $A(a) \geq F_{\mathrm{MAX}}^{-1}(a)$ for $a \geq 2^{-|P|}$, and the approximation factor is $O((\sigma/\mu)\sqrt{C/h})$ for $a \geq 1/2$, where μ and σ^2 are, respectively, the minimum mean and the maximum variance of the edge lengths, and h is the maximum number of edges in a path in G. The running time of our algorithm is $O(|E|)$.

1.2 Related Work

The stochastic longest path problem has been extensively investigated, since there are many useful applications. The (deterministic) longest path problem on DAGs are known as PERT [8] in the fields of scheduling and Operations Research, and their stochastic versions are considered from 1960s [4, 7, 6]. The problem is also intensively studied from the viewpoint of circuit delay analysis [2, 3, 5, 9]. They approximate the distribution function of the longest path length in heuristic ways.They run practically fast, but have no guarantee about the approximation. Then authors of this paper proposed an algorithm that approximates the longest path length from below with approximation guarantee [1].

2 Preliminaries

Let X be a random variable with mean μ and variance σ^2. A random variable X *obeys* a normal distribution $N(\mu, \sigma^2)$ if its distribution function is given as,

$$P(X \leq x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right) dt.$$

3 Approximation Algorithm

Let μ_e and σ_e^2 be, respectively, mean and variance of X_e, the length of edge $e \in E$. Suppose that $\mu_{\max} = \max_{P\in\mathcal{P}}\{\sum_{e\in P}\mu_e\}$ and $\sigma_{\max}^2 = \max_{P\in\mathcal{P}}\{\sum_{e\in P}\sigma_e^2\}$ are known. Then consider the distribution function $F_B(x)$ of $N(\mu_{\max}, \sigma_{\max}^2)$. We can prove that $(F_B(x))^{|\mathcal{P}|}$ is a lower bound on $F_{\text{MAX}}(x)$ for $x \geq \mu_{\max}$. In order to minimize the running time of the algorithm, we propose to approximate the inverse function of $(F_B(x))^{|\mathcal{P}|}$. Let

$$L(x;\mu,\sigma^2) = \exp\left(-\exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 + \ln\ln 2\right)\right).$$

Then we can prove that $L(x;\mu,\sigma^2) \leq F(x)$ for $x \geq \mu$ where $F(x)$ is the distribution function of $N(\mu,\sigma^2)$. Since equation $L(x;\mu,\sigma^2) = a$ has two real solution for $1/2 \leq a < 1$, we denote the larger solution by $L^{-1}(a;\mu,\sigma^2)$. Then we have

$$L^{-1}(a;\mu,\sigma^2) = \mu + \sigma\sqrt{-2\ln(-\ln a) + 2\ln\ln 2}.$$

An upper bound on the inverse function of $(F_B(x))^{|\mathcal{P}|}$ is given as

$$\begin{aligned} &F_B^{-1}(a^{1/|\mathcal{P}|}) \leq L^{-1}(a^{1/|\mathcal{P}|};\mu,\sigma^2) \\ &= \mu_{\max} + \sigma_{\max}\sqrt{2(\ln|\mathcal{P}| - \ln(-\ln a)) + 2\ln\ln 2}, \quad (1) \end{aligned}$$

for $2^{-|\mathcal{P}|} \leq a < 1$. The approximation algorithm is to calculate

$$A(a) = \mu_{\max} + \sigma_{\max}\sqrt{2(C - \ln(-\ln a)) + 2\ln\ln 2}, \quad (2)$$

where C is an upper bound on $|\mathcal{P}|$. If $C = \ln|\mathcal{P}|$, we have $A(a) = L(a^{1/|\mathcal{P}|};\mu_{\max},\sigma_{\max}^2)$.

The time complexity of our algorithm is $O(|E|)$. It is well known that we can calculate $\mu_{\max}$ and $\sigma_{\max}$ in $O(|E|)$ time using the technique of PERT [8]. Since we can easily show that $2^{|V|}$ is an upper bound on the number of paths in a DAG, one can take $\ln 2^{|V|}$ as the value of C. This takes only $O(\log|V|)$ time.

As for the approximation factor, we can prove the following theorem.

Theorem 1. *The approximation factor of the algorithm is given by*

$$\frac{A(a)}{F_{MAX}^{-1}(a)} = O\left(\frac{\sigma}{\mu}\sqrt{\frac{C}{h}}\right),$$

where $\mu = \min_{e\in E}\{\mu_e\}$, $\sigma = \max_{e\in E}\{\sigma_e\}$, *and* h *is the maximum number of edges in* $P \in \mathcal{P}$. □

This theorem shows a trade-off between the approximation factor and the computational time; we can have better approximation factor by taking the exact $\ln|\mathcal{P}|$ as the value of C, which may increase the computational time.

References

1. E.Ando, T.Nakata, M.Yamashita, Approximating the Longest Path Length of a Stochastic DAG by a Normal Distribution in Linear Time. *Submitted for publication.*
2. E.Ando, M.Yamashita, T.Nakata, Y.Matsunaga, The Statistical Longest Path Problem and Its Application to Delay Analysis of Logical Circuits. *Proceedings of the International Workshop on Timing Analysis (TAU2002)*, pp.134-139, 2002.
3. M.Berkelaar, Statistical delay calculation, a linear time method. *Proceedings of the International Workshop on Timing Analysis (TAU'97)*, pp.15-24, 1997.
4. C.E.Clark, The PERT model for the distribution of an activity time. *Operations Research* 10, pp.405-406, 1962.
5. M.Hashimoto and H.Onodera, A performance optimization method by gate sizing using statistical static timing analysis. *IEICE Trans. Fundamentals*, E83-A, 12, pp.2558-2568, 2000.
6. V.G.Kulkarni and V.G.Adlakha, Markov and Markov-Regenerative PERT Networks, *Operations Research*, Vol.34, No.5, pp.769-781, 1986.
7. J.J.Martin, Distribution of the time through a directed, acyclic network. *Operations Research* 13, pp.46-66, 1965.
8. D.G.Malcolm, J.H.Roseboom, C.E.Clark and W.Fazar, Application of a technique for research and development program evaluation. *Operations Research* 7, pp.646-669, 1959.
9. S.Tsukiyama, M.Tanaka and M.Fukui, An Estimation Algorithm of the Critical Path Delay for a Combinatorial Circuit. *IEICE Proceedings of 13th Workshop on Circuits and Systems*, pp.131-135, 2000 (in Japanese).