

枝の重みが確率的なグラフにおける最長路の長さの分布

九州大学大学院システム情報科学研究科
福岡教育大学教育学部
九州大学大学院システム情報科学研究院

今林 裕 (Hiroshi IMAHAYASHI)
中田 寿夫 (Toshio NAKATA)
山下 雅史 (Masafumi YAMASHITA)

1 はじめに

有向グラフ (N, A) の各枝に重みとして実数値が与えられたとき、有向グラフの入口と出口を結ぶ路のうちで、路を構成する枝の重みの総和(その路の長さと呼ぶ)が最大の路を最長路(またはクリティカルパス)という。

最長路の長さを求める問題において、従来、枝の重みは定数で与えられる。その場合は最長路は1つ決まり、その長さも定数となる。しかし多くの例の中で、枝の重みはその枝を移動するか対応する作業を完了させるのに必要なコストや時間の大きさを表し、そのような時間やコストをある確率分布に従う確率変数にすることによってうまくモデル化できる場合が多い。

回路設計では、要求された性能をもつ回路を歩留まり良く作成するために、予め製造プロセスのばらつきを考慮し、適切な設計マージンを設定される。そのとき、各素子の遅延がある分布に従ったばらつきによるクリティカルパスの遅延時間の分布を高速に見積り、遅延制約を満たす回路の歩留まりを評価する必要がある。

クリティカルパスの遅延時間は回路の処理速度を決定する。例えば、順序回路のフリップフロップを駆動する回路のクリティカルパスの遅延時間が予想以上に大きいと所定のクロック周期で回路を動作させることができない。

溝口 [1] は1本のクリティカルパスの遅延時間の分布を求める手法を提案した。彼の手法は2入力素子までに限定し、ハザードが生じる場合や遅い方と早い方のどちらの入力が出力の遷移

を決定するかまで考慮している。

築山ら [2] は複数のクリティカルパスの最大の遅延時間の分布を近似的に求める手法を提案した。彼らは Clark [3] の計算方法を利用している。それは確率変数 X_1, X_2 が正規分布に従うとき $X = \max\{X_1, X_2\}$ が2変量正規分布に従うと仮定して、相関係数まで考慮して X の平均と分散を計算する方法である。

入口と出口を1個ずつ持つ、閉路を含まない連結な有向グラフ (N, A) を考える。そして、各枝は重みを持ち、それは与えられた確率分布をもつ確率変数で与えられる。さらに、全てのこれらの確率変数は互いに独立であるとする。

枝の重みが確率変数であるとき、入口から出口までの最長路の長さの分布を正確に計算することは非常に難しい。なぜならば、各路の長さはもはや確率変数であり、全ての路の集合上でそれらの変数の最大値を見つけなければならないからである。この問題は次の2つの理由により手に負えない。

- 全ての路の集合が非常に大きい
- たとえ枝の重みが互いに独立な確率変数であっても、路の長さを表す確率変数は共有する枝のために互いに独立ではない

Martin [4] は積分を繰り返して、最長路の長さの分布を正確に計算する方法を提案した。特に、直並列グラフにおいて、重みの確率密度関数が多項式の場合について述べている。

Alexopoulos [5] は枝の重みが離散的な分布に従うと仮定して、最長路の長さの分布を求める

手法を提案した。彼はグラフの状態空間の分割の繰り返しによって計算している。

本論文では、グラフを変形することによって、最長路の長さの分布(上側確率)の下限と上限を求める手法を提案する。これらの手法は最短路の長さの分布の上限と下限を求めることにも用いることができる。

2 準備

任意の有向グラフは (N, A) によって表わすことができる。ここで、 N は節点の集合であり A は枝の集合である。節点 $n_i \in N$ から節点 $n_j \in N$ への枝は $a_{ij} \in A$ で表される。時折、枝の集合の i 番目の要素を表すために記号 a_i を使う。

定義 1 有向グラフ (N, A) を、入口と出口を1個ずつ持つ、閉路を含まないグラフとする。

入口 $s \in N$ はそれに入る枝を持っていて、出口 $t \in N$ はそれから出る枝を持っていない節点である。 $s = n_0, t = n_{|N|-1}$ とする。

各枝は重みを持ち、それはある確率分布をもつ確率変数で与えられる。さらに、それらの確率変数は互いに独立であるとする。枝 e_i の重み X_i の分布を表す分布関数を $F_i(x)$ 、確率密度関数を $f_i(x)$ 、平均を τ_i 、分散を σ_i^2 で表す。

定義 2 . 確率変数 X, Y が任意の実数 x に対して $P(X > x) \geq P(Y > x)$ を満たすならば、 X は Y の上限であるという。反対に、 $P(X > x) \leq P(Y > x)$ を満たすならば、 X は Y の下限であるという。

3 最長路の長さの分布の計算

直並列グラフに対して、最長路の長さの分布を計算する方法を述べる。まず、2本の枝から成るグラフに対する計算方法から始める。

確率密度関数がそれぞれ $f_1(x)$ および $f_2(x)$ で与えられる枝 a_1 および a_2 が直列に接続しているとき、 $X = X_1 + X_2$ の分布関数 $F(x)$ と

確率密度関数 $f(x)$ は、これらの確率変数が互いに独立であれば、次式により得られる。

$$F(x) = \int_{-\infty}^{\infty} f_1(y)F_2(x-y) dy$$

$$f(x) = \int_{-\infty}^{\infty} f_1(y)f_2(x-y) dy$$

確率変数 X の平均 τ と分散 σ^2 は $\tau = \tau_1 + \tau_2, \sigma^2 = \sigma_1^2 + \sigma_2^2$ で与えられる。

枝 a_1 および a_2 が並列に接続しているとき、 $X = \max\{X_1, X_2\}$ の分布関数 $F(x)$ と確率密度関数 $f(x)$ は、これらの確率変数が互いに独立であれば、次式により得られる。

$$F(x) = F_1(x)F_2(x)$$

$$f(x) = f_1(x) \int_{-\infty}^x f_2(y) dy + f_2(x) \int_{-\infty}^x f_1(y) dy$$

このように、直並列グラフの最長路の長さは、各枝の重みを表す確率変数が互いに独立ならば上式の計算を再帰的に繰り返すことで計算することができる。

しかし、枝を共有する路が存在する有向グラフ (N, A) の最長路の長さの分布関数と確率密度関数を正確に計算することは、それらの路の長さを表す確率変数が互いに独立ではないので、非常に難しい。

4 最長路の長さの分布の下限

有向グラフから枝を除去したグラフの最長路の長さの分布は、有向グラフの最長路の長さの分布の下限である。除去した枝を含む路が最長路の候補から取り除かれるからである。

よって、グラフから枝を除去して、最長路の長さの分布を求めることによって、最長路の長さの分布の下限を求めることができる。

同様に、グラフから枝を除去して、最短路の長さの分布を求めることによって、最短路の長さの分布の上限を求めることができる。

下限と正確な値との誤差について述べる。シミュレーションの結果から、最長路の長さは大きな長さをもつ路によってほとんど決まることが分かっている。よって、大きな長さをもつ路ができるだけ残るようにグラフから枝を除去して、最長路の長さの分布を求めることによって、正確な値に近い最長路の長さの分布の下限を求めることができると期待できる。

5 最長路の長さの分布の上限

最初に、有向グラフ (N, A) を入口と出口を結ぶ全ての路が枝を共有しないグラフに変形する方法を与える。その方法はまず、有向グラフ (N, A) を根付き木に変形する。

1. 有向グラフ (N, A) の入口 $s = n_0$ を木の根とする。深さ 1 の節点の集合は $a_{0j} \in A$ であるような全ての節点 n_j から成り、それらの節点には入口 s から枝が出ている。もし節点 n_i が深さ $(k - 1)$ の節点の集合に含まれるならば、深さ k の節点の集合は $a_{ij} \in A$ であるような全ての節点 n_j を含む。ここで、もし $a_{ij} \in A$ であるような節点 n_j がいくつか存在するならば、 n_j をその個数だけ複製する。全ての節点 n_j には対応する節点 n_i から枝が出ている。
2. 葉から根へ向かって節点を調べて、節点 n_i から 2 本以上の枝が出ている場合は、入口 s から節点 n_i までの路上の節点と枝を n_i から出ている枝の数だけ複製して、各 n_i から 1 本ずつ枝を出す。
3. 木に含まれる全ての葉を 1 個に縮約する。

変形後のグラフには、入口と出口の節点を除いた他の全ての節点に対して、正確に 1 本のそれに入る枝とそれから出る枝がある。複製した枝はもとの枝と同じ分布の重みをもつとする。さらに、木に含まれる全ての枝の重みを表す確率変数は互いに独立であるとする。

定理 1. 有向グラフに対して、枝の重みを表す確率変数が互いに独立ならば、入口と出口を結ぶ全ての路が枝を共有しないように変形したグラフの最長路の長さの分布は、有向グラフの最長路の長さの分布の上限である。

この変形は、図 1 の変形の繰り返しになっている。ここで、確率変数 X_1, X'_1, X_2, X_3 は互いに独立で、 X_1, X'_1 は同じ分布であるとする。

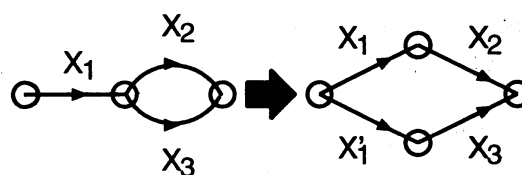


図 1: 枝を複製する変形

枝を複製する変形はステップ 1 と 2 に現れる。ステップ 1 では 2 本以上の枝が入る節点 n_i が複製される。 n_i から出口までの路の長さが X_1 であり、2 本以上の枝が出る節点から n_i までの路の長さが X_2, X_3 である。入口に近い所から変形するので、 n_i より上に 2 本以上の枝が入る節点はない。よって、枝を共有しないように再帰的に X_2, X_3 に対応する路をとることができる。また、閉路がないことから、 X_2, X_3 に対応する路は X_1 に対応する路と枝を共有しない。よって、 X_1, X_2, X_3 は互いに独立である。

ステップ 2 では入口 s から 2 本以上の枝が出る節点 n_i までの路の長さが X_1 であり、 n_i から出口 t までの路の長さが X_2, X_3 である。ステップ 1 により、2 本以上の枝が入る節点はない。また葉から変形するので、 n_i より下に 2 本以上の枝が出る節点はない。よって、枝を共有しないに X_2, X_3 に対応する路をとることができる。また、閉路がないことから、 X_1 に対応する路は X_2, X_3 に対応する路と枝を共有しない。よって、 X_1, X_2, X_3 は互いに独立である。

よって、図 1 の右のグラフの最長路の長さの分布が、左のグラフの上限になっていることを言えば定理 1 は証明される。

補題 1. 確率変数 X, Y が $X \geq Y$ を満たすならば任意の実数 x について,

$$P(X > x) \geq P(Y > x). \quad (1)$$

証明. $\omega \in \{Y(\omega) > x\}$ ならば条件より $X(\omega) \geq Y(\omega) > x$ であるので, $\omega \in \{X(\omega) > x\}$ となり, $\{X(\omega) > x\} \supset \{Y(\omega) > x\}$ がいえるので, 式 (1) がいえる. \square

補題 2. X_1, X'_1, X_2, X_3 は互いに独立で, X_1, X'_1 は同じ分布であるとする. このとき,

$$\begin{aligned} & P(\max\{X_1 + X_2, X'_1 + X_3\} > x) \\ & \geq P(\max\{X_1 + X_2, X_1 + X_3\} > x). \end{aligned} \quad (2)$$

証明. 確率変数 X_2, X_3 の分布関数をそれぞれ F_2, F_3 とする. このとき, (2) の右辺は

$$\int \int_{\{(x_2, x_3)\}} P(\max\{X_1 + x_2, X_1 + x_3\} > x | X_2 = x_2, X_3 = x_3) dF_2(x_2) dF_3(x_3)$$

となる. X_1, X_2, X_3 は互いに独立であるから

$$\int \int_{\{(x_2, x_3)\}} P(\max\{X_1 + x_2, X_1 + x_3\} > x) dF_2(x_2) dF_3(x_3)$$

であり, これを I とおく. 同様に, (2) の左辺は, X_1, X'_1, X_2, X_3 が互いに独立であるから

$$\int \int_{\{(x_2, x_3)\}} P(\max\{X_1 + x_2, X'_1 + x_3\} > x) dF_2(x_2) dF_3(x_3)$$

であり, これを I' とおく. ここで, 積分範囲を

$$E_1 = \{(x_2, x_3) : x_2 > x_3\}$$

$$E_2 = \{(x_2, x_3) : x_2 \leq x_3\}$$

に分け, I, I' の被積分関数を E_1, E_2 上で積分したものをそれぞれ I_1, I_2, I'_1, I'_2 とすると, $I = I_1 + I_2, I' = I'_1 + I'_2$ である. よって式 (2) を示すには $I'_1 \geq I_1$ かつ $I'_2 \geq I_2$ を示せばよい.

i) I_1, I'_1 については

$$I_1 = \int \int_{E_1} P(X_1 + x_2 > x) dF_2(x_2) dF_3(x_3)$$

$$I'_1 = \int \int_{E_1} P(\max\{X_1 + x_2, X'_1 + x_3\} > x) dF_2(x_2) dF_3(x_3)$$

であるが, $\max\{X_1 + x_2, X'_1 + x_3\} \geq X_1 + x_2$ であることより補題 1 より $P(\max\{X_1 + x_2, X'_1 + x_3\} > x) \geq P(X_1 + x_2 > x)$ である. 従って $I'_1 \geq I_1$ である.

ii) I_2, I'_2 については

$$I_2 = \int \int_{E_2} P(X_1 + x_3 > x) dF_2(x_2) dF_3(x_3)$$

$$I'_2 = \int \int_{E_2} P(\max\{X_1 + x_2, X'_1 + x_3\} > x) dF_2(x_2) dF_3(x_3)$$

であるが, X_1 と X'_1 は同じ分布であるので $P(X_1 + x_3 > x) = P(X'_1 + x_3 > x)$ が成り立つので

$$I_2 = \int \int_{E_2} P(X'_1 + x_3 > x) dF_2(x_2) dF_3(x_3)$$

である. 従って, I_1, I'_1 の議論と同様に $\max\{X_1 + x_2, X'_1 + x_3\} \geq X'_1 + x_3$ であるから補題 1 が使えて $P(\max\{X_1 + x_2, X'_1 + x_3\} > x) \geq P(X'_1 + x_3 > x)$ が成り立つので, $I'_2 \geq I_2$ である.

よって $I' \geq I$ となり証明がおわる. \square

定理 1 より, 入口と出口を結ぶ全ての路が枝を共有しないグラフに変形して最長路の長さの分布を求めることによって, 最長路の長さの分布の上限を求めることができる.

同様に, 入口と出口を結ぶ全ての路が枝を共有しないグラフに変形して最短路の長さの分布

を求めることによって、最短路の長さの分布の下限を求めることができる。

上限と下限のグラフを比べると、上限となるグラフは下限となるグラフに何本か路を並列に接続した形をしている。同じような分布をもつ路が並列に接続している場合、路の本数が大きくなるにつれて、路が1本増えることの影響はだんだん小さくなる。このことから、下限と上限の差は小さいと期待できる。

6 回路の遅延時間

4ビット算術論理演算器の遅延時間に対して行った実験について述べる。本論文で提案する入口と出口を結ぶ全ての路が枝を共有しないグラフに変形して最長路と最短路の長さの分布を求める手法と従来用いられている最大最小遅延モデルとの比較を行なう。また、提案手法の誤差を調べるために、モンテカルロシミュレーションの結果との比較も行なう。

6.1 実験手法

最大最小遅延 (min-max delay) モデルは、各素子の遅延時間のばらつきを考慮するモデルであり、広く用いられている。回路中の各基本素子の遅延時間に、最大値と最小値が指定される。遅延時間は必ず最大値と最小値の間の値を取ると仮定されている。

回路全体の遅延時間の最大値は各枝の重みを最大値にして最長路の長さを求める。同様に、最小値は各枝の重みを最小値にして最短路の長さを求める。

工業製品のばらつきは、一般的に正規分布をとることが知られている。この事実をふまえて、提案手法における各素子の遅延時間のばらつきには正規分布を用いる。配線遅延は存在せず、素子遅延だけ存在すると仮定する。また素子の立ち上がりも立ち下がりも遅延時間は同じであると仮定する。

最大最小遅延モデルと比較するために、各素

子の遅延時間の分布を次のような正規分布とする。最大最小遅延モデルにおける遅延時間の最大値より大きな遅延時間となる確率が1%であり、最小値より小さな遅延時間となる確率が1%であるような正規分布である。

回路全体の遅延時間を最大最小遅延モデルと比較するために、次のような遅延時間の最大値と最小値を求める。最長路の長さの上限が最大値より大きな遅延時間となる確率が1%であり、最短路の長さの下限が最小値より小さな遅延時間となる確率が1%である。

提案手法の誤差を調べるために、モンテカルロシミュレーションとの比較も行なった。シミュレーションは3万回行なう。各枝に対して、提案手法と同じ正規分布に従う乱数を発生させて、最長路と最短路の分布を計算する。遅延時間の最大値は最長路の長さとその値より大きいシミュレーションの結果の個数が全体の1%の値とする。同様に、最小値は最短路の長さとその値より小さいシミュレーションの結果の個数が全体の1%の値とする。

回路を次のような有向グラフによって表現する。各枝は素子または配線に対応する。素子に対応する枝の重みはその素子の遅延時間であり、配線に対応する枝の重みは0とする。さらに、入口からは各入力に対応する節点に、出口には各出力に対応する節点から、それぞれ重み0の枝を付ける。ただし、重み0の多重枝は1本の枝とする。このグラフは回路に閉路がなければ有向グラフ (N, A) の条件を満たしている。

このように回路を有向グラフ (N, A) によって表現することで、回路の最大、最小の遅延時間を求める問題は最長路、最短路の長さを求める問題に置き換えられる。

6.2 実験結果

4ビット算術論理演算器に対して行なった実験の結果を示す。この回路は74181ALUとして市販されている。

実験で用いた各素子の遅延時間の最小値と最

大値は、表 1 のとおりである。

表 1: 各素子の遅延時間の最小値と最大値

	最小値	最大値
2 入力 NAND	0.08	0.12
3 入力 NAND	0.08	0.26
4 入力 NAND	0.08	0.32
5 入力 NAND	0.39	0.57
NOT	0.05	0.07
BUFFER	0.17	0.19
2 入力 XOR	0.11	0.37
2 入力 NOR	0.06	0.18
3 入力 NOR	0.07	0.41
4 入力 NOR	0.07	0.63
2 入力 AND	0.13	0.17
3 入力 AND	0.15	0.31
4 入力 AND	0.15	0.37
5 入力 AND	0.44	0.62

表 2: 実験結果

	最小値	最大値
最大最小遅延モデル	0.24	2.39
提案手法	0.24	2.00
シミュレーション	0.25	1.95

6.3 考察

提案手法の方が最大最小遅延モデルよりも小さく、シミュレーションの結果に近い遅延時間の最大値を得ることができた。

ここで、最大値は小さければ良いというわけではない。最大最小遅延モデルでは同じ最大値をもつ枝が何本並列に接続していても 1 本の場合と変わらない。この場合には最大値の大きさは最大最小遅延モデルよりも大きくなければならない。最大最小遅延モデルより小さいことよりも、シミュレーションの値に近いことの方が大事である。

遅延の幅を比べると、シミュレーションは最大最小遅延モデルよりも遅延の幅が小さくなっ

ている。このことから、最大最小遅延モデルでは遅延の幅を取り過ぎていることがわかる。提案手法の遅延の幅のシミュレーションに対する誤差は 3.53 % である。

7 おわりに

提案手法は、最大最小遅延モデルよりも小さく、シミュレーションの結果に近い遅延時間の最大値を与える。つまり、提案手法の方が最大最小遅延モデルよりも正確な値に近い回路の遅延時間の最大値を求めることができる。

本手法は入口と出口を結ぶ全ての路が枝を共有しないグラフに変形して計算することによって、最長路の長さの分布の計算を簡単にしている。特に、枝の重みが正規分布に従うならば、各路の分布は平均と分散の和をとるだけで求めることができる。今後の課題は最大値の誤差を厳密に解析することである。

参考文献

- [1] 溝口大介. ゲート遅延分布を用いた性能テスト手法. 九州大学大学院システム情報科学研究科修士論文, 2000 年.
- [2] 築山修治, 田中正和, 福井正博. 組合せ回路におけるクリティカルパス遅延のばらつき見積り手法. 第 13 回回路とシステムワークショップ原稿, 2000 年.
- [3] C. E. Clark. The Greatest of a Finite Set of Random Variables. *Opns. Res.* Vol. 9, No. 2, pp.145-152, 1961.
- [4] J. J. Martin. Distribution of the Time through a Directed Acyclic Network. *Opns. Res.* Vol. 13, No. 1, pp.46-66, 1965.
- [5] C. Alexopoulos. State Space Partitioning Methods for Stochastic Shortest Path Problems. *Networks* Vol. 30, No. 1, pp.9-21, 1997.