

ホークス過程を用いたソフトウェア信頼性分析に関する考察

広島大学 ディペンダブルシステムズ研究室 平野 莉雄
 広島大学 ディペンダブルシステムズ研究室 土肥 正 教授
 広島大学 ディペンダブルシステムズ研究室 岡村 寛之 教授

1 はじめに

ソフトウェアの信頼性は、その品質と効率性を測る上で重要な指標であり、その予測と評価は産業界および学術界において広く研究されている。近年、ソフトウェアの複雑性と機能が増大するにつれて、そのフォールトやエラーを効果的に予測し管理することが、システム全体の安定性と信頼性を確保する上で不可欠となっている。本研究では、ホークス過程を用いたソフトウェア信頼性分析を探求し、特に指数カーネル [1] とガンマカーネル [2] を用いたモデルを展開する。また、非同次ポアソン過程 (NHPP) モデルを 11 種類の確率分布関数を用いて構築し、これらのモデルをソフトウェアフォールトの発生時刻データが含まれる 8 つのデータセットに適用する。各モデルの適合度と予測精度は、赤池情報量規準 (AIC) と予測平均二乗誤差 (PMSE) を用いて評価する。分析の結果、ホークス過程モデルは、特定の条件下で NHPP モデルと比較して適合度や予測精度が優れていることが示された。指数カーネルとガンマカーネルを用いたホークス過程は、ソフトウェアフォールトデータのダイナミクスを異なる視点から捉え、予測の精度を向上させるための新たな可能性を開く。本研究では、これらの結果を詳細に分析し、ソフトウェア信頼性の評価におけるモデル選択の指針を提供する。

2 確率過程モデル

この章では、本研究で比較する 3 種類モデルについて示す。NHPP モデルとは、事象が起きる発生率が時間とともに非齊次に変化するポアソン過程のこ

とである。最尤推定により未知のパラメータを推定し、AIC を計算する。累積イベント数を N_t とおくと、 N_t の確率関数は

$$Pr\{N(t) = n\} = \frac{\{\Lambda(t)\}^n}{n!} \exp -\Lambda(t) \quad (1)$$

であり、累積イベント数の期待値は

$$E[N(t)] = \Lambda(t) = \int_0^t \lambda(x) dx \quad (2)$$

$$\int_0^t \lambda(x) dx = \alpha F(t) \quad (3)$$

である。

ここで、 α はパラメータであり、 $F(t)$ は指数分布モデルの場合 $F(t) = 1 - e^{-bt}$ である。

一方 Hawkes 過程とはあるイベントが発生した場合に、イベントが発生する時刻が強度に依存し、次のイベントの発生に影響する自己励起性のある点過程のことである。本研究で扱う Hawkes 過程はカーネル関数 ϕ によって、2 つのモデルを扱う。まず、指数カーネルの Hawkes 過程は、

$$\phi = \alpha e^{-\beta t} \quad (4)$$

指数カーネルの Hawkes 過程モデルの強度関数は、

$$\lambda(t; t_1, t_2, \dots, t_m) = \mu e^{-bt} + \alpha \sum_{j=1}^m e^{-\beta(t-t_j)} \quad (5)$$

であり、指数カーネルの Hawkes 過程モデルの累積イベント数の期待値は、 $\alpha < \beta$ のとき

$$E[N(t)] = \frac{\mu}{\beta - \alpha} (1 - e^{-(\beta - \alpha)t}) \quad (6)$$

であり、 $\alpha > \beta$ のとき

$$E[N(t)] = \frac{\mu}{\alpha - \beta} (1 - e^{-(\alpha - \beta)t}) \quad (7)$$

となる。ここで、 α, β, μ は未知の非負実数値パラメータであり、 n は時刻 t までの累積イベント数である。時刻 t 時点での強度関数は時刻 t 以前のすべての注目時刻によって特徴づけられている。

また、もう 1 つのカーネル関数であるガンマカーネルの関数は、

$$\phi = \frac{\alpha t^{k_1 - 1} e^{-\frac{t}{k_2}}}{k_2^{k_1} \Gamma(k_1)} \quad (8)$$

である。 α はイベントの影響を表す性のパラメータであり、 k_1, k_2 はそれぞれ、カーネルの形状パラメータと、カーネルの減衰速度を表すレートパラメータである。 $\Gamma(k_1)$ はガンマ関数で、正規化に用いられる。また、 $k = 1$ の時、

$$\phi(t) = \frac{\alpha \exp\left(-\frac{t}{k_2}\right)}{k_2} \quad (9)$$

となる。この時の強度関数は、ホークス過程の条件付き強度関数 $\lambda^*(t)$ は基本強度 λ (背景率) と過去のイベントが与える影響の総和で構成される。ガンマカーネルを使用する場合、この強度関数は以下のように表される

$$\lambda^*(t) = \lambda + \sum_{k=1}^{N(t)} \frac{\alpha \exp\left(-\frac{t-t_k}{k_2}\right)}{k_2} \quad (10)$$

また、ガンマカーネルのホークス過程の累積期待値は

$$E(N(t)) = \frac{\lambda}{1 - \alpha} t - \frac{\alpha \lambda k_2}{(1 - \alpha)^2} \left[1 - \exp\left(-\frac{1 - \alpha}{k_2} t\right) \right] \quad (11)$$

となる。

3 数値実験

この章では、ホークス過程の指数カーネル、ガンマカーネル、および非同次ポアソン過程 (NHPP)

モデルの適合性と予測精度を比較する。評価指標については、適合性を赤池情報量規準 (AIC)、予測精度を予測平均二乗誤差 (PMSE) を用いて評価する。

表 1 実験結果 (8 つのうち 1 つのデータセットで、観測点 80 % を抜粋)

80%	NHPP	Hawkes(指数)	Hawkes(ガンマ)
AIC	679.56(lxvmax)	694.932	718.037
PMSE	0.763954(Pareto)	1.49994	3.04959

4 考察と今後の課題

本研究では、NHPP モデル、Hawkes 過程の指数カーネルとガンマカーネルのソフトウェアフォールトデータに対する適合度と予測性能の比較を示した。比較の結果、1 部のデータセットでは Hawkes 過程モデルが、ソフトウェア信頼性モデルとして一般的な NHPP モデルよりも良い適合度と評価性能を示した。今後の課題としては、新たなデータセットで実験を行い、モデルの統計的信頼性を向上させることと、各モデルがどのようなデータに対して良い性能を示すのかを明らかにしていきたい。

参考文献

- 1 P. Bao, H.-W. Shen, X. Jin, X.-Q. Cheng "Modeling and Predicting Popularity Dynamics of Microblogs using Self-Excited Hawkes Processes," Proceedings of the 24th International Conference on World Wide Web, pp.9-10, 2015.
- 2 Laurent Lesage, Madalina Deaconu, Antoine Lejay, Jorge Augusto Meira, Geoffrey Nichil, Radu State "Hawkes processes framework with a Gamma density as excitation function: application to natural disasters for insurance"