2001 年度日本オペレーションズ・リサーチ学会
秋季研究発表会

位相型ソフトウェア信頼性モデル

岡村寛之 (01013754), 渡部保博, 土肥正 (01307065)
広島大学大学院工学研究科情報工学専攻

1. はじめに

現在までに，ソフトウェアの信頼性を定量的に評価する方法として様々なソフトウェア信頼性モデルが提案されている．最も一般的に知られている手法は，テスト段階あるいは運用段階におけるフォールト発見事象を確率過程で表現する手法である．確率過程によるモデル化は構築されたモデルにおけるパラメータとソフトウェア開発工程における物理的な現象を対応づけることを容易にするという長所をもつ．しかしながら，実務面においてソフトウェア信頼性を評価する場合，複数のモデルの中から評価対象となるプロジェクトに対して最適なモデルを選択しなければならないという短所も同時に合わせ持つ．

そこで本研究では，確率過程によるソフトウェア信頼性モデルを統合する一般化順序統計量モデル [1] の枠組みにおいて，代表的な既存のソフトウェア信頼性モデル（指数形信頼度成長モデル [2]，遅延 S 字形信頼度成長モデル [3]，超指数形信頼度成長モデル [4] 等）を包括する位相型ソフトウェア信頼性モデルの提案を行う．また，実際のフォールト発見データからモデルのパラメータを推定するための EM アルゴリズムを構築し，推定手続きも含めた既存モデルに対する統一的な取扱いを可能にする．

2. 位相型ソフトウェア信頼性モデル

位相型ソフトウェア信頼性モデルの基本となる一般化順序統計量モデルについて概説する．ソフトウェア信頼性モデルにおいて，NHPP によるソフトウェア信頼性モデルを統一的に取扱う枠組みとして，一般化順序統計量モデルが提案されている．一般化順序統計量モデルは，次のような仮定の下で議論される [1].

(仮定 A) ソフトウェアのテスト中にソフトウェア故障が発生した場合，原因となるフォールトは瞬間的に発見・除去される．

(仮定 B) プログラム中に含まれる初期フォールト数 N は平均 $\omega\ (>0)$ のポアソン分布に従う．

(仮定 C) ソフトウェア故障はおのおの独立かつ時間に関してランダムに発生し，ソフトウェアフォールト 1 個当りのフォールト発見時刻は確率分布 $F(t)$，密度関数 $f(t)$, $t \ge 0$ に従って分布する．

このとき，時刻 t までに発見されるフォールト数の確率分布は

$$\Pr\{N(t)=m\} = \frac{\{\omega F(t)\}^m}{m!}\exp\{-\omega F(t)\}$$

と表現され，平均値関数 $\omega F(t)$ をもつ NHPP と等価である．つまり，単一フォールトの発見時間分布 $F(t)$ に依存して種々のソフトウェア信頼度成長モデルを表現することが出来る．

本稿では，単一フォールト発見時刻の確率分布関数に以下で定義される位相型分布を適用する．

$$F(t) = 1 - \boldsymbol{\alpha}\exp(\boldsymbol{Q}t)\boldsymbol{e}.$$

ここで,

$$\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_k),$$

$$\boldsymbol{e} = (1, \cdots, 1)^T,$$

$$\boldsymbol{Q} = \begin{pmatrix} -t_{11} & t_{12} & \cdots & t_{1k} \\ t_{21} & -t_{22} & \cdots & t_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ t_{k1} & t_{k2} & \cdots & -t_{kk} \end{pmatrix}$$

位相型分布 $F(t)$ は推移率 $t_{ij}, (i,j=1,\cdots,k)$ をもつ k 個の位相から構成され，時刻 $t=0$ での位相は $\boldsymbol{\alpha}=(\alpha_1,\ldots,\alpha_k)$ の初期確率ベクトルによって決定される．

位相型分布は指数族と呼ばれる確率分布関数を包括する分布であり，位相の数やパラメータの設定によって指数分布，アーラン分布，超指数分布，Cox 分布等を表現することが可能である．一般化順序統計量モデルにおいて，$F(t)$ に適当な確率分布関数を仮定することで種々のモデルを導出できることを鑑みれば，位相型ソフトウェア信頼性モデルにより既存の指数形信頼度成長モデル [2]，遅延 S 字形信頼度成長モデル [3]，超指数形ソフトウェア信頼度成長モデル [4] 等を包括出来ることが容易に推察できる．

次節では，ソフトウェア開発工程において実際のフォールト発見データが与えられたとき，位相型ソフトウェア信頼性モデルのパラメータを推定する手法について議論する．

3. EM アルゴリズムによるパラメータ推定

本研究では，EM アルゴリズム [5] を用いた推定アルゴリズムを提案する．総数が N 個の確率変数 $X_1,\ldots,X_N$ (N は確率変数) の順序統計量 $X_{[1]} < \cdots < X_{[N]}$ を考える．順序統計量に対する完全データ $\boldsymbol{x}=(x_{[1]},\ldots,x_{[n]})$ が得られたときの尤度関数は，$X_1,\ldots,X_n$ が互いに独立で同一な確率変

数であり，確率分布および密度関数がそれぞれ $F(x)$ および $f(x)$ であるとき

$$L(\boldsymbol{\theta}) = n!P(N=n)\prod_{i=1}^{n} f(x_{[i]};\boldsymbol{\theta})$$

と表される．ここで，$\boldsymbol{\theta}$ は確率分布に含まれるパラメータの集合である．特に N がパラメータ ω のポアソン分布に従う場合，

$$L(\omega,\boldsymbol{\theta}) = \omega^n \exp(-\omega)\prod_{i=1}^{n} f(x_{[i]};\boldsymbol{\theta})$$

となる．よって，上式を最大にするパラメータは，

$$\hat{\omega} = n, \qquad \hat{\boldsymbol{\theta}} = \mathrm{argmax}\{\log L_0(\boldsymbol{\theta})\}$$

によって定まる．ここで $L_0(\theta) = \prod_{i=1}^{n} f(x_{[i]})$ であり，$\mathrm{argmax}\{\log L_0(\boldsymbol{\theta})\}$ は $\log L_0(\boldsymbol{\theta})$ を最大にする $\boldsymbol{\theta}$ を表す．

$m < n$ に対して，不完全な観測データ $\boldsymbol{x} = (x_{[1]},\cdots,x_{[m]})$ が与えられたとき，EM アルゴリズムの n ステップにおけるパラメータの推定値 $\hat{\omega}^{(n)}$ および $\hat{\boldsymbol{\theta}}^{(n)}$ は期待対数尤度を最大にするパラメータによって決定される．すなわち，$E_{\boldsymbol{\theta}}[\cdot]$ をパラメータ集合 $\boldsymbol{\theta}$ を用いたときの期待演算子と定義れば，$n+1$ ステップにおけるパラメータの推定値 $\hat{\omega}^{(n+1)}$ と $\hat{\boldsymbol{\theta}}^{(n+1)}$ は，

$$\hat{\omega}^{(n+1)} = E_{(\omega^{(n)},\boldsymbol{\theta}^{(n)})}[N],$$

$$\hat{\boldsymbol{\theta}}^{(n+1)} = \mathrm{argmax}\{E_{(\omega^{(n)},\boldsymbol{\theta}^{(n)})}[\log L_0(\boldsymbol{\theta})]\}$$

によって算出される．

位相型ソフトウェア信頼性モデルにおいて，単一フォールトの発見時間分布 $F(t)$ は位相数 k の位相型分布に従うものとする．$(\tau_1,\cdots,\tau_k)^T = -\boldsymbol{Qe}$ とすると，通常の最尤法から得られるパラメータの推定値は，すべての $i,j=1,\ldots,k$ に対して，

$$\hat{\alpha}_i = \frac{B_i}{n}, \qquad \hat{t}_{ij} = \frac{N_{ij}}{Z_i},$$

$$\hat{\tau}_i = \frac{N_{i0}}{Z_i}, \qquad \hat{\omega} = n,$$

$$\hat{t}_{ii} = -\left(\hat{\tau}_i + \sum_{j=1, j\neq i}^{k} \hat{t}_{ij}\right)$$

となる．ここで，

B_i：初期位相として i が選択される回数

Z_i：位相 i の滞在時間

N_{ij}：位相 i から位相 j へ推移する回数（ただし位相 0 はフォールトの発見事象を表す）

である．上記の B_i, Z_i, N_{ij} および初期フォールト数に関連した確率変数 N を用いて，EM アルゴリズムは以下のように記述される．

E-Step: 観測されたデータ $\boldsymbol{x}_{\mathrm{obs}} = (x_{[i]}; i=1,\ldots,m)$ と n 番目のステップで推定されたパラメータ $\boldsymbol{\alpha}^{(n)}, \boldsymbol{Q}^{(n)}, \omega^{(n)}$ から B_i, Z_i, N_{ij}, N の条件付き期待値を求める．

M-Step: E-Step で求めた条件付き期待値から，以下の更新式を用いてパラメータの推定値を算出する．

$$\hat{\omega}^{(n+1)} = E_{(\boldsymbol{\alpha}^{(n)},\boldsymbol{Q}^{(n)},\omega^{(n)})}[N|\boldsymbol{x}_{\mathrm{obs}}],$$

$$\hat{\alpha}_i^{(n+1)} = \frac{E_{(\boldsymbol{\alpha}^{(n)},\boldsymbol{Q}^{(n)},\omega^{(n)})}[B_i|\boldsymbol{x}_{\mathrm{obs}}]}{E_{(\boldsymbol{\alpha}^{(n)},\boldsymbol{Q}^{(n)},\omega^{(n)})}[K|\boldsymbol{x}_{\mathrm{obs}}]},$$

$$\hat{t}_{ij}^{(n+1)} = \frac{E_{(\boldsymbol{\alpha}^{(n)},\boldsymbol{Q}^{(n)},\omega^{(n)})}[N_{ij}|\boldsymbol{x}_{\mathrm{obs}}]}{E_{(\boldsymbol{\alpha}^{(n)},\boldsymbol{Q}^{(n)},\omega^{(n)})}[Z_i|\boldsymbol{x}_{\mathrm{obs}}]},$$

$$\hat{\tau}_i^{(n+1)} = \frac{E_{(\boldsymbol{\alpha}^{(n)},\boldsymbol{Q}^{(n)},\omega^{(n)})}[N_{i0}|\boldsymbol{x}_{\mathrm{obs}}]}{E_{(\boldsymbol{\alpha}^{(n)},\boldsymbol{Q}^{(n)},\omega^{(n)})}[Z_i|\boldsymbol{x}_{\mathrm{obs}}]}.$$

上記のアルゴリズムでは初期パラメータを与える必要がある．このとき，与えられた初期パラメータに応じて位相型ソフトウェア信頼性モデルは多種のモデルを表現する．例として，既存のソフトウェア信頼性モデルとの対応関係を以下に示す．

(i) 指数形信頼度成長モデル [2]：

$$\boldsymbol{\alpha}^{(0)} = 1, \qquad \boldsymbol{Q}^{(0)} = -\beta$$

(ii) 遅延 S 字形信頼度成長モデル [3]：

$$\boldsymbol{\alpha}^{(0)} = (1,0), \qquad \boldsymbol{Q}^{(0)} = \begin{pmatrix} -\beta & \beta \\ 0 & -\beta \end{pmatrix}$$

(iii) 超指数形信頼度成長モデル [4]：

$$\boldsymbol{\alpha}^{(0)} = (\alpha_1,\ldots,\alpha_k),$$

$$\boldsymbol{Q}^{(0)} = \begin{pmatrix} -\beta_1 & 0 & \cdots & 0 \\ 0 & -\beta_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -\beta_k \end{pmatrix}$$

ただし，$\beta_1 > \cdots > \beta_k$.

参考文献

[1] N. Langberg and N. D. Singpurwalla, "Unification of some software reliability models," *SIAM J. Sci. Comput.*, **6**, 781-790 (1985).

[2] A. L. Goel and K. Okumoto, "Time-dependent error-detection rate model for software reliablity and other performance measures," *IEEE Trans. Reliab.*, **R-28**, 206-211 (1979).

[3] S. Yamada and S. Osaki, "Software reliability growth modeling: models and applications," *IEEE Trans. Software Eng.*, **SE-11**, 1431-1437 (1985).

[4] M. Ohba, "Software reliability analysis models," *IBM J. Research and Development*, **28**, 428-443 (1984).

[5] 岡村，渡部，土肥，尾崎，"EM アルゴリズムに基づいたソフトウェア信頼性モデルの推定，" 信学技報，**101**, 35-40 (2001).