

平成 20 年 6 月 1 日現在

研究種目：若手研究 (B)
 研究期間：2007～2008
 課題番号：19700028
 研究課題名 (和文) オープンソース開発におけるソースコードの安定性予測法の開発
 研究課題名 (英文) A method for predicting source code stability
 in open source development
 研究代表者
 阿萬 裕久 (AMAN HIROHISA)
 愛媛大学・大学院理工学研究科・講師
 研究者番号：50333513

研究成果の概要：近年注目を集めているオープンソース開発に対し、ソースコードの特徴を定量的尺度によって定量化し、どういったコードであれば安定した開発が行われているのかについて分析並びに統計解析を行った。その結果、コード中のコメント文記述について「コメント文の記述密度は概ね40%以上が望ましいが、記述頻度は低くすべき。」という基準を見いだすことができた。また、安定時期の予測に向けて、成長曲線モデルの適用可能性も検討した。

交付額

(金額単位：円)

	直接経費	間接経費	合計
2007年度	1,100,000	0	1,100,000
2008年度	700,000	210,000	910,000
年度			
年度			
年度			
総計	1,800,000	210,000	2,010,000

研究分野：情報システム工学

科研費の分科・細目：情報学・ソフトウェア

キーワード：ソフトウェア工学

1. 研究開始当初の背景

(1) 当初、ソフトウェア開発の手法としてオープンソース開発が注目を浴びてきていた（現在もそうである）。オープンソース開発の場合、高品質なソフトウェアを低コストで開発できる場合が多く、大手のIT企業でも自社のソフトウェアの一部をオープンソース化するケースが増えてきている。

(2) オープンソース開発では開発に携わる人材や工数といった管理データの収集は難しく、必然的にコスト見積りが困難となってしまう。そこで応募者らは、オープンソース

開発の主たる成果物であるソースコードとその変更履歴の定量的な観測・分析を通じて（ソースコードに変更が施されなくなる）安定状態に達するまでの期間や変更量を予測するための手法開発が今後有用になると考えた。

2. 研究の目的

(1) ソースコード安定性を定量的に表現し、それに関する統計データをまとめる。

(2) メトリクスを使ってソースコードの品質属性を定量化し、ソースコードの安定性と

の因果関係を統計解析によって見いだす。

3. 研究の方法

(1) オープンソース開発サイトから多数のソースコードとその変更履歴情報を収集し、データベース化する。

(2) 収集したソースコードに対してソフトウェアメトリクスによる測定を行い、メトリクスデータ（品質測定値）をデータベース化する。

(3) データの分析と統計解析を通じて、どのような特性のソースコード（メトリクス値で特徴化）であれば安定性が高いのかを見いだす。また、それに関連して変更履歴の推移に関する観測・分析を行う。

4. 研究成果

(1) 国内外のオープンソース開発サイトから多数のソフトウェア（Eclipse, Apache Tomcat 等）の入手可能なすべてのバージョンのソースコードを入手しデータベース化した。また、そこでの変更履歴情報（変更量, 日時等）も整理し、データベース化した。（表 1, 表 2）

表 1. 収集したソースファイル数及び変更履歴件数（一部）

ソフトウェア	ファイル数	変更履歴
jEdit	548	2113
HSQLDB	298	1439
JFreeChart	760	2407
Eclipse	49312	15825

表 2. 収集したバグ情報数（一部）

ソフトウェア	バグ情報
Bugzilla	7086
Calendar	6827
Tomcat	28392

(2) 収集したソースコードに対して、コードの規模、コメント文記述の量、コメント文記述の頻度を測定し、データベース化した。なお、測定に関してはツールを自作した。

(3) ①統計解析の結果、コメント文記述のソースコード中での密度が40%以上ならば、概ね変更は起こりにくい（安定性が高い）という基準を見いだすことができた（図 1, 図 2）。

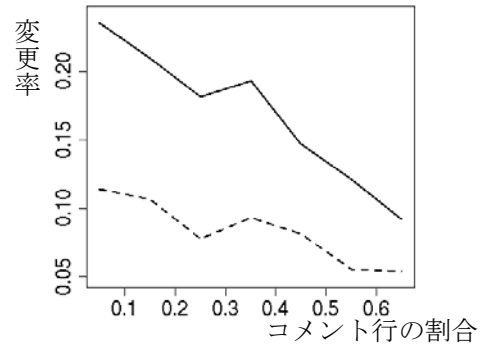


図 1. コメント行数の割合とソースコードの変更率の関係 (Eclipse); 実線は平均値, 破線は中央値を表わす。

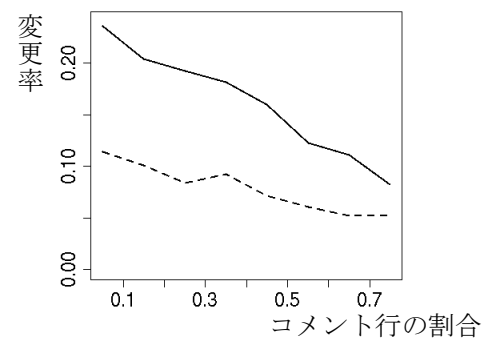


図 2. コメント文字数の割合とソースコードの変更率の関係 (Eclipse); 実線は平均値, 破線は中央値を表わす。

ただし、コメント文記述の頻度が高くなると、その傾向は弱まることも実験を通じて確認できた（図 3, 図 4）。

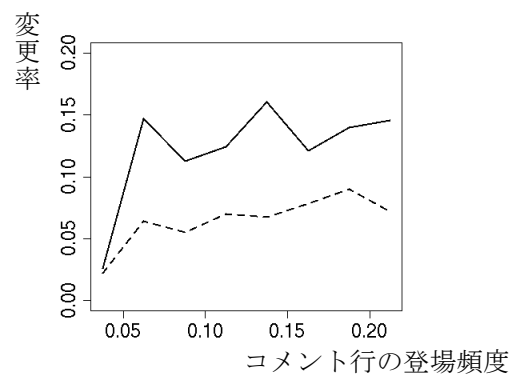


図 3. コメント文記述の頻度（行単位）とソースコードの変更率の関係 (Eclipse); 実線は平均値, 破線は中央値を表わす。

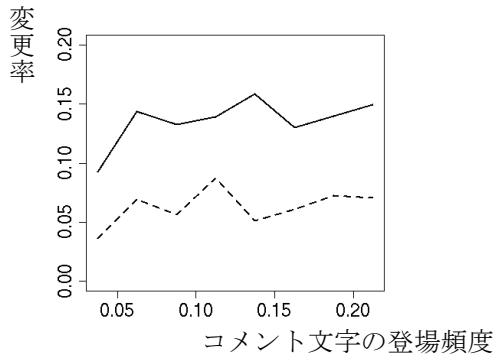


図4. コメント文記述の頻度（文字単位）とソースコードの変更率の関係（Eclipse）；実線は平均値，破線は中央値を表わす。

②ソースコードの変更そのものの時間的推移にも着目し，成長曲線モデルによる当てはめ実験も行った．その結果，ソースコードの変更量，規模ともに成長曲線モデルで比較的良好なモデル化が可能であることが分かった（図5，表3，図6，表4）．これに関してはまだ十分な考察を得るに至っていないが，今後の課題として引き続き研究を行っていく．

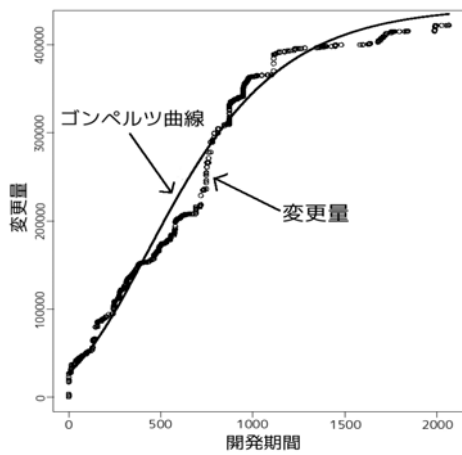


図5. ソースコードの変更量の推移とゴンペルツ曲線による当てはめ（HSQLDB）

表3. 三種類のソフトウェアの変更量に対する当てはめ誤差

	jEdit	HSQLDB	JFreeChart
平均値	6.2%	4.3%	7.4%
中央値	6.1%	4.1%	7.5%

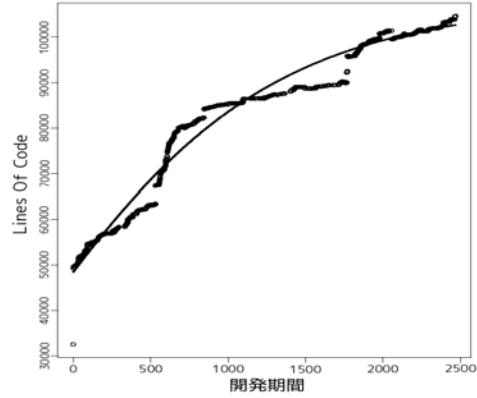


図6. ソースコードの規模の推移とゴンペルツ曲線による当てはめ（jEdit）

表4. 三種類のソフトウェアの規模に対する当てはめ誤差

	jEdit	HSQLDB	JFreeChart
平均値	2.2%	0.9%	1.5%
中央値	2.2%	0.9%	0.6%

現在，障害報告件数の時間的推移についても同様の傾向を確認できている（図7～9）．

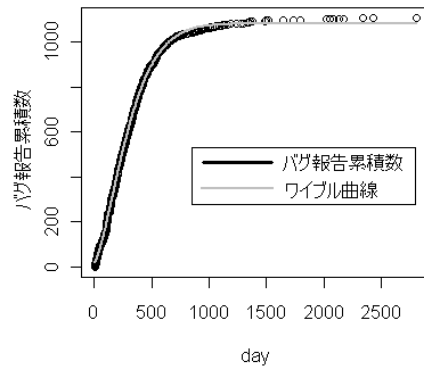


図7. 障害報告累積数とワイブル曲線による当てはめ（Tomcat Ver. 3）

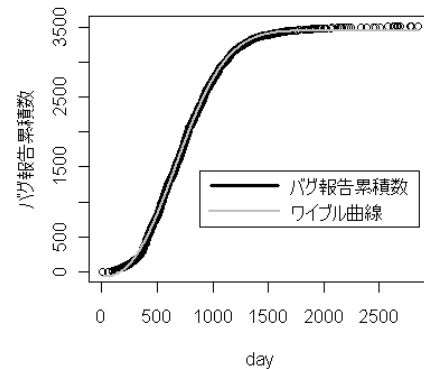


図8. 障害報告累積数とワイブル曲線による当てはめ（Tomcat Ver. 4）

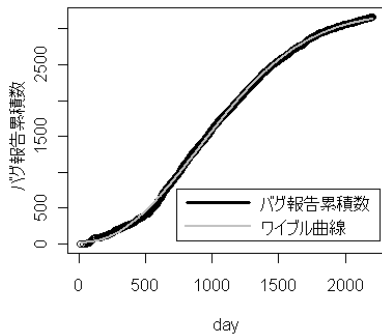


図9. 障害報告累積数とワイブル曲線による当てはめ (Tomcat Ver.5)

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計1件)

① Hirohisa AMAN, Hirokazu OKAZAKI, Impact of Comment Statement on Code Stability in Open Source Development, Knowledge-Based Software Engineering, IOS Press, 査読有, 2008, pp.415-419.

[学会発表] (計9件)

① 阿萬裕久, 成長曲線モデルを用いたオープンソース開発の管理, 情報処理学会ウインターワークショップ 2009・イン・宮崎論文集, pp.73-74, 2009年1月24日, 宮崎.

② 谷藤圭太, 阿萬裕久, Pareto 原理によるオープンソース開発の定常状態解析, 情報処理学会ウインターワークショップ 2009・イン・宮崎論文集, pp.77-78, 2009年1月23日, 宮崎.

③ 黒川耕平, 阿萬裕久, オープンソース開発におけるバグ報告累積数の成長曲線モデルを用いた分析, 電子情報通信学会技術報告, vol.108, no.326, pp.61-66, 2008年11月28日, 福岡.

④ 久保田寛, 阿萬裕久, オープンソースソフトウェアにおけるコードの安定性予測に向けたゴンペルツ曲線の適用, 電子情報通信学会技術報告, vol.108, no.157, pp.1-6, 2008年7月24日, 札幌.

⑤ 谷藤圭太, 阿萬裕久, ソースコードの変更量における Pareto 原理とそれを用いたプロジェクトの定常状態解析, 電子情報通信学会技術報告, vol.108, no.65, pp.53-58, 2008年5月29日, 宮崎.

⑥ 阿萬裕久, オープンソース開発における定量的品質管理について, 情報処理学会ウインターワークショップ 2008・イン・道後論文集, pp.103-104, 2008年1月24日, 松山.

⑦ 久保田寛, 阿萬裕久, オープンソースソフトウェアの安定性予測に向けたデータ解析, 電子情報通信学会技術報告, vol.107, no.212, pp.61-66, 2007年9月11日, マレーシア.

⑧ 森崎潤, 阿萬裕久, ソースコードの保守工程における変更連鎖と安定性の関係解析, 電子情報通信学会技術報告, vol.107, no.159, pp.7-12, 2007年7月24日, 札幌.

⑨ 岡崎博和, 阿萬裕久, ソースコードにおけるコメントの頻度と保守性の関係解析, 電子情報通信学会技術報告, vol.107, no.159, pp.1-6, 2007年7月24日, 札幌.

[図書] (計0件)

[産業財産権]

○出願状況 (計0件)

○取得状況 (計0件)

[その他]

① 阿萬裕久, ソフトウェア品質の定量的な測定と統計解析に基づいた品質管理・予測技術, 月刊愛媛ジャーナル, 第22巻, 7号, 2009, pp.82-85.

6. 研究組織

(1) 研究代表者

阿萬 裕久 (AMAN HIROHISA)

愛媛大学・大学院理工学研究科・講師

研究者番号: 50333513

(2) 研究分担者

なし

(3) 連携研究者

なし