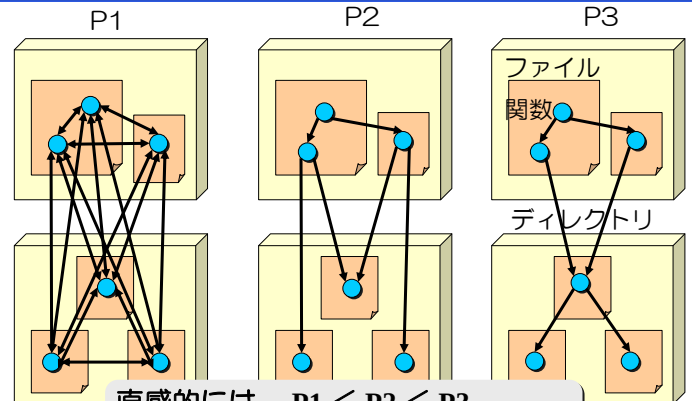


C言語プログラムソースコードの 再利用性測定法とその評価

鷲崎 弘宜 早稲田大学
小池 利和 ヤマハ株式会社
波木 理恵子 オージス総研
田邊 浩之 オージス総研

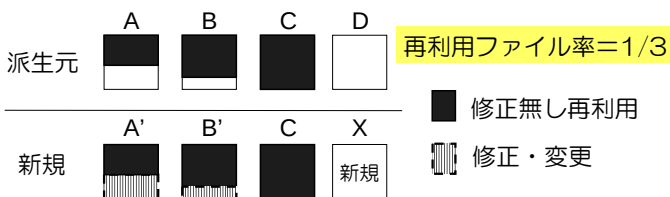
どれが、再利用により派生開発しやすいだろう？



直感的には、 $P1 < P2 < P3$ 。
その科学的根拠は？ 閾値は？

再利用、重要

- 再利用のしやすさを客観的に判断したい
 - 短納期・低コスト $\Leftrightarrow$ 高品質。2/3は派生開発。
 - 再利用性: 一部/全てを他の開発に利用できる能力
 - 再利用率: 再利用の実績
- 流用 (変更あり) $\Leftrightarrow$ 再利用 (変更なし)



再利用率は、再利用が終わって初めて分かる。
それを再利用の前に推測できないか？

問題と解決

- プログラムの特徴から再利用のしやすさを判定したい
 - 開発中に再利用性を「作りこみたい」。問題箇所を特定したい。
 - 経験的には、依存が複雑なほど再利用しにくい。本当に？
 - 様々な測定法 (メトリクス) はあるが、使われていない。

【問題A】測定法が現実
に有用なことが未検証。
閾値なし。

【解決1】測定法 (メトリクス)
を網羅的に識別
■ GQM法

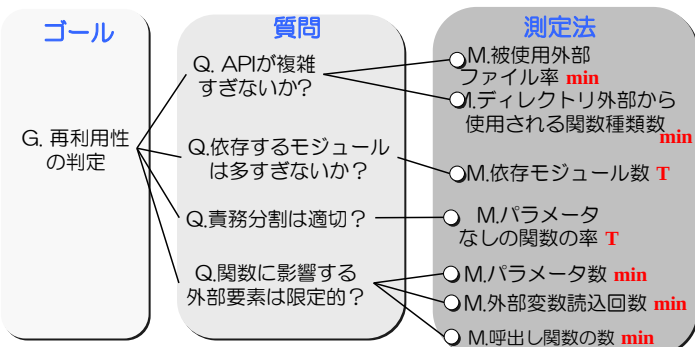
【解決2】再利用率と照らし合
わせて検証
■ 過去プロジェクトデータ
■ 評定水準 (閾値) の統計的導出

【問題B】複数あるとき、
どの測定法がどの程度
使えるのか不明

【解決3】重回帰分析による
個々の測定法の影響を導出

【解決1】GQMによる測定法の網羅的識別

- Goal-Question-Metricによる段階的マッピング
 - 専門家集団によるレビュー/修正の繰り返し (約1年)
 - 測定値はファイル $\rightarrow$ ディレクトリ $\rightarrow$ システムと集約可



【評定水準種別: T 閾値内、min 小さいほどよい】

【解決2, 3】有効性の実証実験

- (1) 10個の組込み開発プロジェクトデータ対
 - 全て過去1度の派生開発実績あり。データあり。
 - 統計的に有意となるよう、再利用率が異なるもの。

10プロジェクトの規模データ			
	ディレクトリ数	ファイル数	有効行数ELOC
総計	213	10,298	5,291,096
被再利用部の合計	173	7,940	3,734,614

- (2) 初期の単位・データ形式選定
 - 再利用率を目的変数、測定値を説明変数とし、重回帰分析
- 測定法の単位: システム単位で有意
測定値のデータ形式: 率データ (規模の影響を受けない)
再利用率の種類: 再利用関数率でもっとも有効

【解決2, 3】 (3)測定法の選定

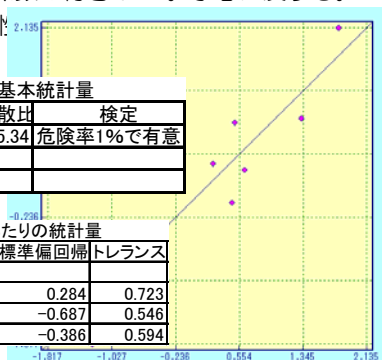
- 全測定法（説明変数）7個を採用した場合
 - 自由度調整済み寄与率 = 0.812
 - 5個について偏回帰係数の符号が +。予想に反する。
 - 相関性から、重共線性

表3: 説明変数を選択後の基本統計量

要因	平方和	自由度	分散	分散比	検定
回帰	10.713	3	3.571	15.34	危険率1%で有意
残差	1.397	6	0.233		
計	12.11	9			

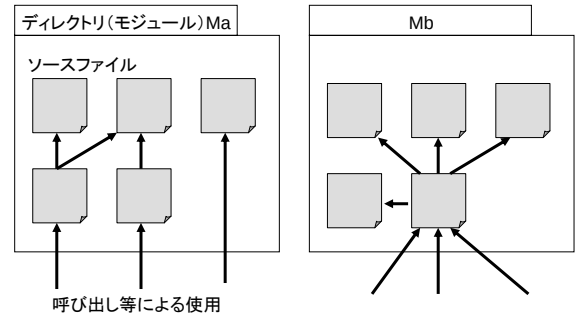
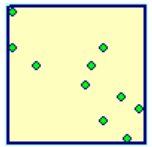
表4: 説明変数を選択後の変数あたりの統計量

変数名	偏回帰係数	標準誤差	t値	標準偏回帰	トレランス
定数項	3.446	1.769	1.948		
MMd042率	0.362	0.207	1.744	0.284	0.723
MMd032率	-8.453	2.31	-3.66	-0.687	0.546
MFn042率	-1.029	0.48	-2.14	-0.386	0.594



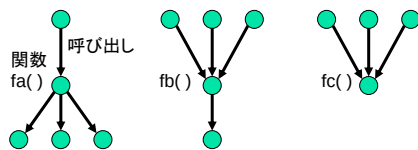
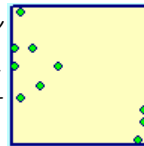
得られた知見

- 再利用は、関数単位で行われる
- 再利用のしやすさは、システム全体で把握できる
- 「依存関係が、再利用には重要！」
 - 外部から使われるファイルが制限されているほど再利用しやすい (MMd032)



得られた知見 (つづき)

- 他の関数の呼び出しが少ないほど再利用しやすい (MFn042)
- 再利用のしやすさには、被使用外部ファイル率が呼び出し関数数よりも大きく影響する



知見の補足

- MMd042 について、依存モジュール数が多いほど再利用率が下がると予想していた
- 実験の結果、ほぼ相関なし
- 理由: MMd042はモジュールの分け方に強く依存し、その分け方はプロジェクトにより様々であったため

まとめと展望

- C言語プログラムソースコードを対象として、再利用性の測定に有効と想定される複数の測定法を定性的に識別
- 複数の再利用実績データを用いた精密な分析を通じて、それらのうちで3つがそれぞれ異なる影響の度合いを持って実際の再利用性の大きさに関係している傾向にあることを統計的に明らかにした
- 「依存関係が、再利用には重要！」
- C言語プログラムの構成要素や全体の内部品質を測定するために構築した枠組みの一部
 - 47の質問、101の副質問、236の測定法
 - <http://www.ogis-ri.co.jp/solution/a-04-02.html>
- 今後の課題
 - プロジェクトデータの増加と分析の繰り返しに基づくさらなる測定法の見直し
 - 再利用率の定義を派生元と派生先が1対1に対応していない状況にまで拡大
 - 再利用回数といった他の再利用実績データとの関係の分析

有難うございました。