

Portland Pattern Repositoryにおけるソフトウェアパターンに対するネットワーク分析

USOL東京

川村 健

takeshi.kawamura@usol.co.jp

開発における問題点

適用パターンがパターン参照関係の末端にある場合、特定領域の問題を解決することが多く、より広範囲な問題を解決するパターンとあわせて適用することを検討することが望ましい。しかし、選択したパターンがパターン参照関係のどこに位置するのか把握する方法はよくわかっていない。

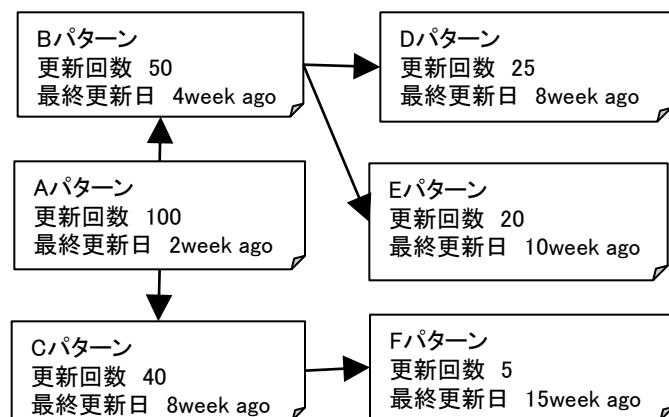
手法・ツールの適用による解決

パターン収集サイトPortland Pattern Repositoryへ下記の手順でデータ分析を行い、パターン参照関係の構造的特徴を明らかにした。
 ①WebクローラーwebSPHINXを用いてパターン参照関係を取得。②ネットワーク分析ツールPajekを用いてパターン参照関係に対してネットワーク分析を実施し、中心性を算出。③統計ソフトRを用いて得られたメトリクスに対してメトリクス分析を実施。

データ分析 全体概要

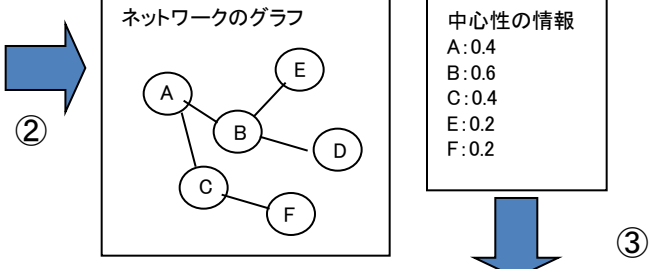
【分析手順】

①パターン参照関係取得⇒②ネットワーク分析⇒③メトリクス分析

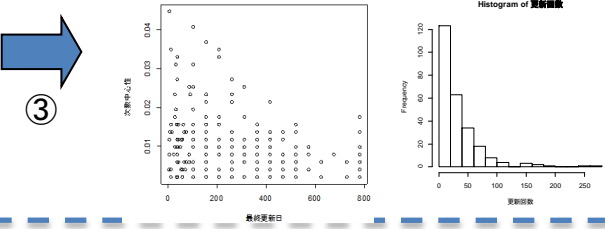


参照関係の情報
 A ⇒ B
 A ⇒ C
 B ⇒ D
 B ⇒ E
 C ⇒ F

履歴情報
 A:100回 2week ago
 B:50回 4week ago
 C:40回 8week ago
 D:25回 8week ago
 E:20回 10week ago
 F:5回 15week ago



中心性の情報
 A:0.4
 B:0.6
 C:0.4
 E:0.2
 F:0.2

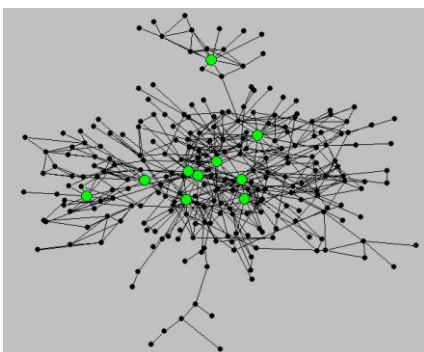


Portland Pattern RepositoryのWebページリンク関係

結果

【結果】

実際のソフトウェアに適用する頻度が高いと思われるパターンは次数中心性が高い傾向があり、参照関係において中心部に位置することがわかった。



パターン間の参照関係 (表のパターンを緑色で表示)

パターン名	次数 中心性
ModelViewController	0.0447
AdapterPattern	0.0408
HandleBodyPattern	0.0369
VisitorPattern	0.0350
SynchronizationStrategies	0.0350
ValueObject	0.0330
SceneGraph	0.0330
StrategyPattern	0.0311
CompositePattern	0.0311
ScapeGoat	0.0311

次数中心性上位10パターン

考察・今後の課題

【考察】

次数中心性の高いパターンは、パターンの参照関係の中心付近に存在する傾向がある。次数中心性の高いパターンは利用頻度が高いと思われるパターンが多かった。
 パターン適用時には、次数中心性の高いパターンから適用を検討したほうが、効率がよいと考えられる。

【今後の課題】

- 以下が今後の課題としてあげられる。
 - ・次数中心性が高いパターンの使用頻度の裏づけ
⇒OSSコミットログ、Googleキーワードランクで利用頻度を調査
 - ・パターンの話題性の調査
⇒インターネットアーカイブを利用しパターン登録日を取得
 - ・一般性の調査
⇒Portland Pattern Repository以外のパターン集に対してのデータの分析