

Portland Pattern Repository 上のソフトウェアパターン群に対するネットワーク分析

角谷 将司† 鷺崎 弘宣† 川村 健‡ 深澤 良彰†

ソフトウェアパターンの収集サイト Portland Pattern Repository(以下 PPR)にネットワーク分析の手法を適用し、パターンの参照関係の位置を示す指標である中心性を算出し、パターン間の参照関係の構造的特徴を明らかにした。また、PPR に記載されている各パターンの履歴情報についてデータ分析を行った。分析の結果、使用頻度の高いと思われるパターンは次数中心性が高い傾向となった。このことは適用するパターンの検討や新たなパターンの抽出に有用であると考えられる。

Network analysis to Software Pattern on Portland Pattern Repository

Masashi Kadoya† Hironori Washizaki† Takeshi Kawamura‡ Yoshiaki Fukazawa†

We revealed construction of reference relationship of patterns by applying network analysis, which calculates centrality indicating where a pattern is in, to Portland Pattern Repository (PPR) which collects pattern. Then, we analyzed pattern's historical information written on PPR. As a result, a pattern which is used frequently tends to have high degree centrality. This fact is useful to extract new pattern and to consider a pattern applied accurately.

1. はじめに

ソフトウェアパターン(以下、パターン)とはソフトウェア開発の繰り返し現れる問題から得られる知識を、再利用できるように抽象化・形式化してまとめた解法である。パターンを選択する際の参考情報として、パターンの収集・整理を行っている Web サイト Portland Pattern Repository[1](以下、PPR と表記)があげられる。PPR では参照関係にあるパターンには、ページ間にハイパーリンクを設定し、参照関係を持たせている。これらの要素間を分析する手法としてネットワーク分析[2]があげられる。ソフトウェア開発者は、参照関係にあるパターンも適用することが多いが、参照関係の位置を把握する指標はあまりわかっていない[3]。

そこで本稿では、パターン間の参照関係の構造的特徴について明らかにした。

2. 背景

2.1. ネットワーク分析

ネットワーク分析は、社会学や通信ネットワークなど

の分野で多く用いられている分析手法で、数学のグラフ(Graph)理論に基礎を置いている。ここでいうネットワークとは、頂点(N: Node)と線(E: Edge)を基本構成要素としている。ネットワークは、線で頂点と頂点の関係を示しており、各頂点の持つ線の数を次数と呼ぶ。

ノードがネットワーク内でどこに位置するかを表す指標として、次数中心性、近接中心性、媒介中心性の3つ中心性が知られている。次数中心性とは「線が集まる点ほど中心性が高い」とする最もシンプルな中心性の計算法である。近接中心性とは「他の点と距離が近いほど中心性が高い」とする中心性の計算法である。媒介中心性とは「その点を通る経路が多いほど、中心性が高い」とする計算法である。以下に次数中心性 C_d 、近接中心性 C_c 、媒介中心性 $C_b(v_k)$ の式を示す。

$$C_d(v_i) = \frac{\text{deg}(v_i)}{n-1}$$

頂点 v_i 、頂点 v_i と隣接する頂点の次数 $\text{deg}(v_i)$ 、ネットワーク内の総頂点数 n

$$C_c(v_i) = \frac{n-1}{s(v_i)}$$

頂点 v_i 、他の頂点までの距離の総和 $s(v_i)$

$$C_b(v_k) = \frac{2BC(v_k)}{(n-1)(n-2)}$$

頂点の媒介値が理論的に最大となる場合(スターグラフ)の媒介値 $BC(v_k)$

2.2. パターン利用時の問題点

パターン利用者がパターンを適用する場合、特定領域の問題を解決するパターンのみを適用することは少

†早稲田大学

‡USOL 東京

なく、そのパターンと参照関係にある汎用性のあるパターンを組み合わせる適用することが多い。そのため、選択したパターンがパターン間の参照関係のどこに位置するのかを検討する必要がある。

3. Portland Pattern Repository の分析

パターン間の参照関係の構造的特徴について明らかにするため、次のような分析を行った。

- Step1 Web クローラーWebSPHINX を用いたツールを作成し、PPR に登録されているパターン間の参照関係と履歴情報をリンクのクローリングすることで収集した。
- Step2 その参照関係に対して、ネットワーク分析ツール Pajekc 使用して、各パターンの中心性を算出した。
- Step3 中心性と履歴情報のメトリクス間の関連を、統計ソフト R を使用して分析した。

3.1. 分析結果と考察

算出した3つの中心性、更新回数、最終更新日に対して、メトリクス単独でヒストグラムを作成し、パターン間の参照関係図をグラフ化した。それらのヒストグラムを図2から図6に示す。また、度数中心性、近接中心性、媒介中心性の高い上位6パターンをそれぞれ表1、表2、表3に示した。

各パターンの参照関係を図7に示す。各ノードがパターンを表しており、パターン間の参照関係を線で表示している。度数中心性の高い上位10パターンをグラフ化した。

図4から媒介中心性に関しては、スケールフリー性がみられ、ごく一部の限られたパターンがパターン間の関係を媒介しているという特徴がみられた。

表1、表2、表3から ModelViewController, VisitorPattern といった、実際によく利用されると思われるパターンは、度数中心性、近接中心性、媒介中心性のすべてが高い傾向があるということがわかった。

よって、度数中心性、近接中心性、媒介中心性の高いパターンを参照することで、ソフトウェア開発者が参照したいパターンを早期発見することができると思われる。

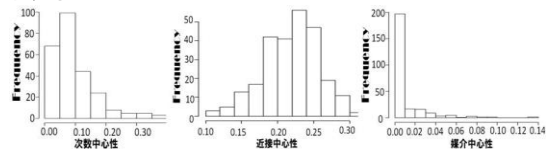


図1度数中心性 図2近接中心性 図3媒介中心性

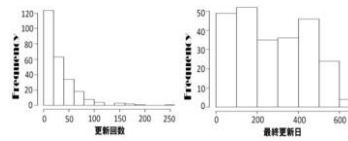


図4更新回数 図5最終更新日

表1 度数中心性の高い6ターン

パターン名	度数中心性
ModelViewController	0.0447
AdapterPattern	0.0408
HandleBodyPattern	0.0369
VisitorPattern	0.0350
SynchronizationStrategies	0.0350
ValueObject	0.0330

表2 近接中心性の高い6パターン

パターン名	近接中心性
VisitorPattern	0.0447
StrategyPattern	0.0408
CompositePattern	0.0369
NullObject	0.0350
ModelViewController	0.0350
UseClosuresNotEnumerations	0.0330

表3 媒介中心性の高い6パターン

パターン名	媒介中心性
VisitorPattern	0.1312
CompositePattern	0.1311
NullObject	0.0931
AdapterPattern	0.0929
StrategyPattern	0.0890
ModelViewController	0.0818

参考文献

- [1] <http://c2.com/cgi/wiki?PortlandPatternRepository>, Portland Pattern Repository, 2013年2月6日
- [2] 大平雅雄, ソーシャルネットワーク分析の基礎, http://se.naist.jp/~masao/lecture2008/sna/socialnetwork_analysis_1.pdf, 2013年1月8日
- [3] 井庭崇, 湯村 洋平, 若松 孝次, 古市 奏文 2007年, プロジェクト推進のパターン・ランゲージとその評価, 日本ソフトウェア科学会 ネットワークが創発する知能研究会&情報処理学会 数理モデル化と問題解決研究会合同ワークショップ