

ネットワーク型データモデルを用いた問題点の可視化と 問題分析への応用例

Visualization of defects using Network Type Data Model and some examples of analyzing defects

株式会社日立ソリューションズ ソリューション品質保証本部
Solution Quality Assurance Support Division, Hitachi Solutions,Ltd.
角口 勝隆¹⁾
Yoshitaka Kadoguchi

Abstract We have conventionally been analyzing software quality by totalizing the typical classifications of defects (grouping by functions or programs), in order to grasp the trends of defects and the points to be improved. However, we cannot always analyze software quality or defects precisely by using typical classifications. Therefore, we have considered using "Network Type Data Model" in order to visualize atypical defects, not depending on classifications. Using this model, we have succeeded in grasping the trends of defects and analyzing the points to be improved. We report some examples and results of this analysis in this paper.

1. はじめに

大規模なソフトウェア開発プロジェクトの現場では、テストにより不具合が数千件摘出されることもある。従来は、不具合の発生傾向や改善が必要な箇所を把握するために、不具合情報へ「重要度」「現象分類」「原因分類」「作り込み工程」「摘出すべき工程」「未発見理由」などの定型的な分類情報^[1]を付与し、それらの分類情報を機能やプログラムごとに集計することで品質状況の分析を行ってきた(図1参照)。

しかし、ソフトウェア開発プロジェクトで扱う問題は多岐にわたり、分析する情報が定型的に分類されているとは限らない。また、分類されていたとしても、情報の欠落や属人的な分類判断により、しばしば精度に欠けた情報で分析を行わざるを得ないのが実状である。プロジェクトによって分析の目的/情報/期間はさまざま、必ずしも上述の方法で品質状況や問題を的確に分析できるとは言えない。

そこで、不定形な問題情報を可視化する方法として、「ネットワーク型データモデル」の採用を検討した。これにより、問題点の発生傾向の把握および、改善が必要な箇所を分析することができた。当論文では、その分析実施例と成果について報告する。

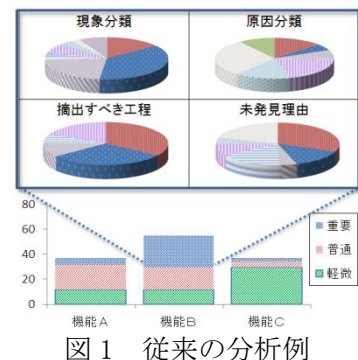


図1 従来の分析例

株式会社日立ソリューションズ ソリューション品質保証本部
Solution Quality Assurance Support Division, Hitachi Solutions,Ltd.

大阪府大阪市北区堂島浜 1-2-1 新ダイビル Tel:06-6147-5028
e-mail:yoshitaka.kadoguchi.uj@hitachi-solutions.com
1-2-1, Kita-ku Doujima-hama, Osaka City, Japan
1) Lead Engineer, Quality Assurance Department

2. 分析の目的と、従来の分析方法での問題点

2.1 分析の目的と、分析を実施する上での問題点

ソフトウェアの誤動作が社会に与える影響が年々高まっている中、金額不正などの処理結果不正による重要障害を撲滅させるために社内の「障害管理システム」に蓄積された情報を分析し、重要障害の再発防止／未然防止を図ることになった。この分析作業での目的を表1に示す。

表1 分析作業の目的

分析作業の目的	
①現状把握	問題の根本原因を対策するために、問題の背景にある要因を掘り下げて、正しく現状を把握すること
②傾向分析	効率的に対策効果を上げるために、問題の発生傾向や共通的な要因を分析し、問題の発生しやすい部分(ウィークポイント)を洗い出すこと
③改善施策	洗い出したウィークポイントから、再発防止／未然防止に対し実効性のある施策を提案すること

当論文の対象範囲

このため、従来実施してきたソフトウェア不具合の分析方法のように定型的な分類種別を集計／比較することで、障害内容の発生状況を定量的かつ定性的に把握することを試みた。しかし、障害管理システムはお客様状況の把握や対策完了までのステータス管理を主目的としており、ソフトウェア不具合票のように細分化された分類情報は付与されておらず、分析を実施するには以下の問題があった。

- (1) 文章(テキスト)として記述された「現象」「原因」等の情報からキーワードを抽出し、分類・整理をしなければ、ソフトウェア不具合と同様な分析を行うことができない。
- (2) 約1,200件もある大量データを分析するには、人手では結果を得られるまで時間を要する。

2.2 より効率的な分析方法の考察

上述の問題を解決するためには、分類作業の効率化が必要である。このことから、テキスト情報を自動的に分類・分析する「テキストマイニング^{[2][3]}」の活用を検討した。また、テキストマイニングツールとして著名なオープンソースの「KH Coder(作:立命館大学産業社会学部 樋口耕一准教授)^[4]」を採用することにした。

KH Coderでは、ExcelやCSVファイルの特定列に記述されたテキスト情報を元に、さまざまな分析を行うことができる。図2に、KH Coderの分析機能による出力結果の一例を示す。


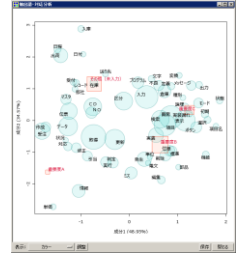
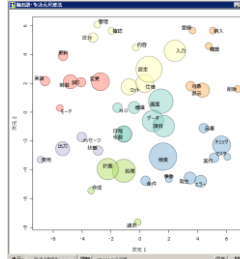
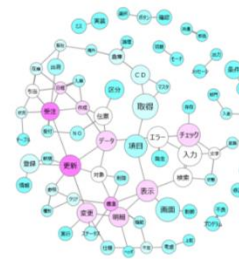
カテゴリズ分析	クラスター分析		ネットワーク分析
<p>抽出語リスト</p> 	<p>対応分析</p> 	<p>多次元尺度構成法</p> 	<p>共起ネットワーク</p> 
<p>語の出現頻度を、品詞別に一覧表示したもの。品詞ごとの特徴や傾向を確認することが可能。</p>	<p>似たような語を、関連する項目と組み合わせ配置した図。項目ごとに、語の出現傾向を確認することが可能。</p>	<p>似たような語同士をバブルプロットした図。類似する語の座標位置から、相関関係などを確認することが可能。</p>	<p>語と語の繋がりをネットワーク図化したもの。感覚的に、どのような事象が頻出しているのかを確認することが可能。^[3]</p>

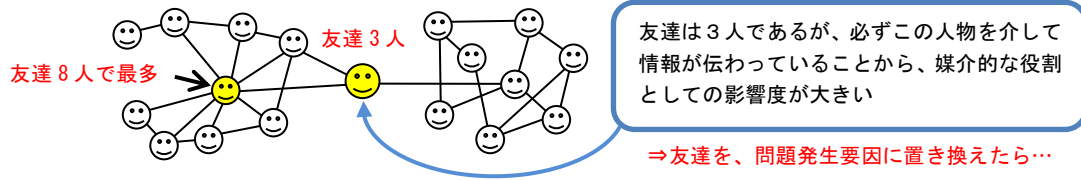
図2 KH Coderの分析機能による出力結果の一例

KH Coder の分析機能は多々あるが、問題情報(テキスト情報)を「ネットワークデータ型モデル」
として可視化できる「共起ネットワーク」が現状把握に最も適していると考えた。その理由は表
2に示すように、不定形な問題情報の見える化(モデリング)にネットワークデータ型モデルが
適していると判断したからである。

表2 主なデータモデルと特徴

データモデルの種類	特徴
①階層型データモデル	カテゴライズ(定型的に分類)した情報の整理に適したモデル ※従来のソフトウェア不具合分析
②リレーション型データモデル	
③クラスターモデル	類似性や相関関係、パターン ^① の可視化に適したモデル
④ネットワーク型データモデル	物事の関連性や相互作用のように、不定形な情報の表現に適したモデル

また、ネットワークデータ型モデルによる分析の活用例として「社会ネットワーク分析」「DNA
解析」「人工知能の意思決定」などが挙げられる。この分析方法を、問題分析へ応用できないか考
えた。例えば社会ネットワーク分析では、情報を効率良く拡散させたい場合は、媒介的な位置づ
けとなる人物を特定して情報を伝達している。問題分析においては、複数要因の中から関連性の
強い要因を特定し除去することで、全体的な問題発生頻度を抑止できるのではないかと考えた。
図3に、社会ネットワーク分析を例とした、問題分析への応用案を示す。



社会ネットワークでの分析例	問題分析での応用案(仮説)
組織の中で、媒介的なつながりが多い人は、社会的影響度 ^② が大きい。(インフルエンサーと呼ばれる)情報を効率良く拡散させたい場合、社会的影響度が大きい人に情報を与えるのが良い。	複数要因をパイプ的に連結させている箇所は、全体的に関連性の強い問題発生要因である。全体的な問題発生頻度を抑止させたい場合、関連性の強い要因を除去するのが良い。

図3 社会ネットワーク分析の、問題分析への応用案

3. ネットワーク型データモデルを用いた分析

3.1 ネットワーク型データモデルによる、問題の可視化

「共起ネットワーク図」の「共起」とは、任意の文書や文において“ある文字列と他のある文字列が同時に出現する(=共に起こる)こと”である。「共起ネットワーク図」とは、文字列間の共起性をリンクとして表したものである。図4に、共起ネットワーク図の作成概念を示す。

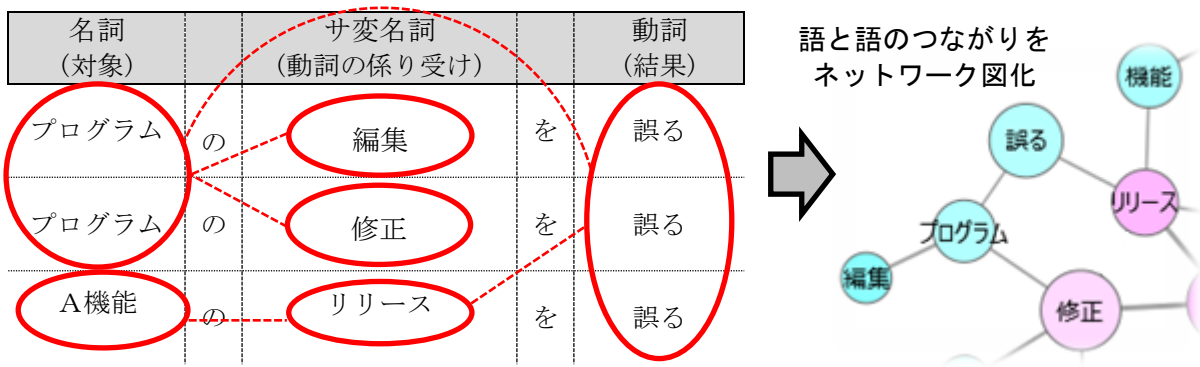


図4 共起ネットワーク図の作成概念

また、共起ネットワーク図から意味のある分析結果を得るためには、「入力対象の選択/よく使われる単語の辞書登録/ノイズ除去/品詞の選択」といったことを意識して作成する必要がある。共起ネットワーク図を作成する上で留意すべき事項を、表3に示す。

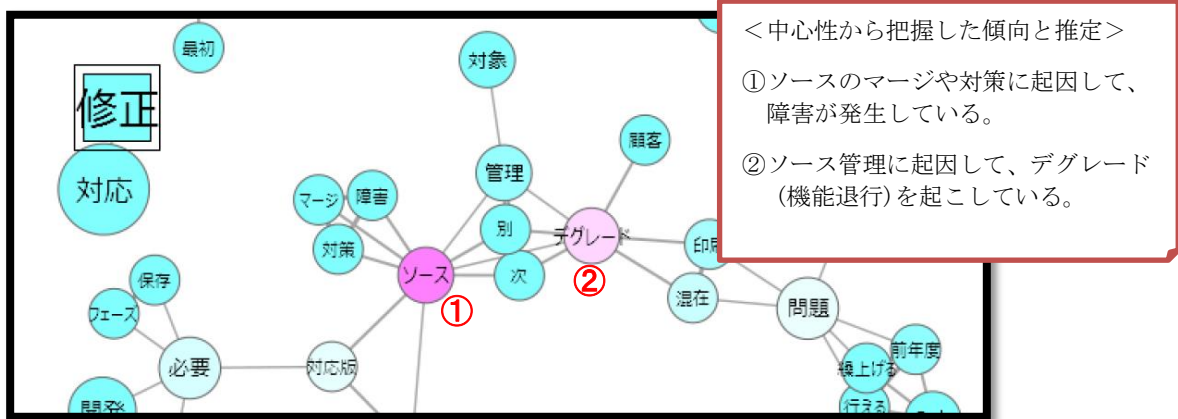


図7 「修正」に絡む語をネットワーク図化し、次数中心性を求めた結果

図7では「ソース」「デグレード」で次数中心性が高くなっている。その周辺の関連語から、ソースのマージや対策に起因して、障害が発生している可能性が推察される。また、ソース管理に起因して、デグレード(機能退行)が発生している可能性が推察される。

(3) 捉えた特徴の詳細を確認

KH Coder では、「KWIC コンコーダンス」機能で特定のキーワードに着目して、前後の文脈を確認することが可能である。図8に、「ソース」をキーワードとして、前後の文脈を確認した例を示す。これにより、対象を絞った上で要点を確認することが可能である。その結果、図7での推定と、詳細確認結果は一致していた。

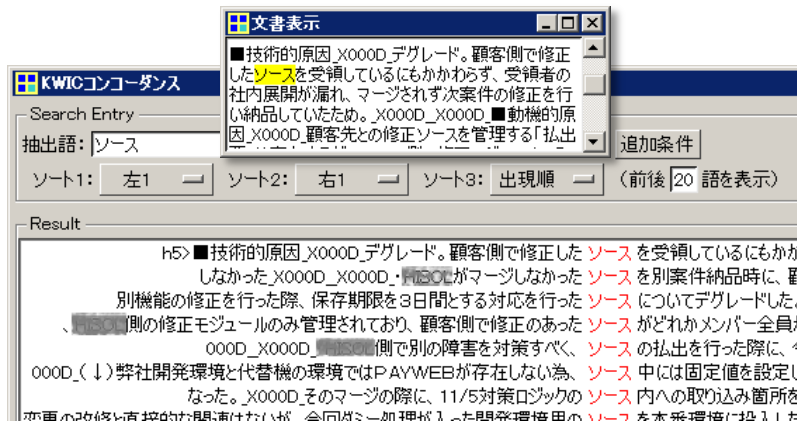


図8 捉えた特徴の詳細確認例

3.3 ネットワーク型データモデルによる、問題分析の結果

上述と同様の方法で、他の中心性についても確認を繰り返し、有効な情報がないか探索した。その結果、「修正」に関連した中心性から「機能修正/変更時の、影響範囲調査が不足」という特徴も掴むことができた。ネットワーク図と中心性指標により分析した結果を、表5に示す。

表5 ネットワーク図と中心性指標による分析結果

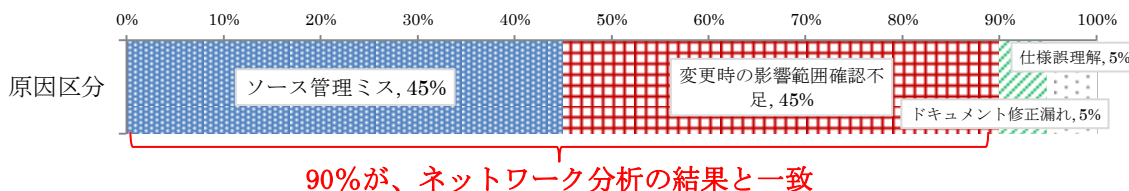
#	中心性から掴んだ原因の特徴	問題背景（仮説）
1	ソース管理や、ソースのマージ作業で障害が発生していた。	プロジェクトメンバーが入れ替わった際、ソース管理ルールや躰が引き継がれていないのではないかと推察される。
2	機能修正/変更時の、影響範囲調査が不足していた。	ドキュメント体系が複雑 もしくは 記載が不十分？ また、デグレード防止確認（機能退行テスト）の実施方法に問題があるのではないかと推察される。

仮説に基づき、
事実を確認

表5に示す仮説に基づき、事実を確認し、要因を掘り下げて根本原因を対策することで、有効な再発防止/未然防止策を図ることが可能となる。

3.4 分析結果の検証

「中心性の指標から影響の大きい要素を特定することで、問題の発生傾向や改善が必要な箇所を確認できる」という仮説に基づいて分析した結果を検証するために、サンプリングで無作為に約10%を抽出したデータ(120件)で、目視による分類整理を実施した。その結果を図9に示す。ネットワーク図の中心性から求めた原因傾向は、全体の90%に合致していた。このことから、ネットワーク図による分析は実用に耐え得るものとする。



90%が、ネットワーク分析の結果と一致

図9 目視による分類結果との比較

4. ネットワーク型データモデルによる分析の効果と、今後への期待

4.1 導入による効果

表1の「分析作業の目的」に対する達成状況を、表6に示す。
また、ネットワーク分析を採用したことによる効果を、表7に示す。

表6 分析作業の目的に対する達成状況

分析作業の目的		達成状況
①現状把握	問題の根本原因を対策するために、問題の背景にある要因を掘り下げて、正しく現状を把握すること	問題の背景を掘り下げるための仮説(ヒント)を得やすい
②傾向分析	効率的に対策効果を上げるために、問題の発生傾向や共通的な要因を分析し、問題の発生しやすい部分(ワークポイント)を洗い出すこと	ネットワークの中心性から、問題が発生しやすい部分の把握が可能

表7 ネットワーク分析による効果(メリット/デメリット)

メリット：期待できる効果	デメリット：期待できない事項
<p>(1)問題発生傾向の概要を素早く把握 不定形な問題情報(テキスト情報)を、ネットワーク型データモデルで可視化することにより、現状の問題発生傾向を素早く把握することが可能。</p> <p>(2)改善が必要な箇所を素早く把握 中心性指標により、ネットワーク図から改善が必要な箇所を素早く把握することが可能。</p> <p>(3)大量データを分析可能 人手による分類や集計作業は不要。大量データの分析で、特に効果を発揮する。(目安として120件の分析所要時間が、約4.5時間から1時間に短縮。件数が増大しても、ツールの処理時間はさほど変わらない)</p>	<p>(1)定量的な判断ができない 分析結果は相関関係であり、定量的な良し悪しは判断できない。定量データと併せた分析・評価が必要。</p> <p>(2)対象データが少ないと効果を得られない 目安で100件以上のデータが必要。</p> <p>(3)分析結果は分析者のスキルに依存 抽象化された特徴しか把握できず、特徴のない部分が情報落ちし、見逃してしまう可能性がある。 また、抽象化された情報となるため、具体的に物事を捉えるには、現場経験や、問題背景の理解が必要となる。</p>

ネットワーク分析は、大量の不定形な問題情報(テキスト情報)を分析するのに有効である。他に実績を得た活用事例として、ヘルプデスクでの問合せ傾向分析による、FAQ記載項目へのフィードバックにも活用した。また、従来の分類整理による分析と組み合わせ、相互の特徴を活かして利用することで、より多角的に問題を捉えることができると考える。

4.2 効果を上げるために必要な要素

ネットワーク型データモデルによる問題情報の可視化は、現状の問題発生傾向や、改善が必要な箇所を素早く把握する用途には有効である。ただし、この分析方法を活用する上で必要となる要素がすべて揃っていないと、効果を上げることができない。表8に、ネットワーク分析で必要となる要素を示す。

表8 ネットワーク分析で必要となる3つの要素

必要な要素	詳細
データ	<ul style="list-style-type: none"> ・入力するテキスト情報に、分析に必要な情報が記載されていること ・「現象」や「原因」のように意味のある内容でまとまっていること ・分析対象データが、ある程度存在すること（目安：100件以上）
ツール	<ul style="list-style-type: none"> ・ネットワーク図を生成できるツールが必要（KH Coder など）
分析スキル	<ul style="list-style-type: none"> ・テキストマイニングに関する知識が必要 ・ネットワークの中心性の理解が必要 ・得られる結果は、あくまで現状分析であり、改善施策の立案と推進については別途知識や経験が必要となる

4.3 今後への期待と応用

(1) ソフトウェア開発プロジェクトに限らず、あらゆる現場での問題分析に応用可能。

一例として、厚生労働省が医療事故情報収集等事業にて公開している「ヒヤリ・ハット事例データベース^[7]」から取得した情報のネットワーク分析結果と、公開されている統計レポートの問題傾向は一致していた。

(2) ツールの分析機能を活用した、より多角的な分析。

紙面の都合で割愛したが、KH Coder では分類整理された定型情報（機能や部署/担当者、作り込み工程/抽出すべき工程など）と組み合わせてネットワーク図を作成することが可能であり、より多角的に問題を捉えることができる。例えば、「作り込み工程（定型情報）」と「問題発生傾向（非定型情報）」を組み合わせて分析することで、その工程においてプロジェクトに潜在している問題点を把握し、進行中のプロジェクトへフィードバックするような活用方法が考えられる。

謝辞

本論文の執筆にあたり、有用な知識やツールを公開いただいている方々ならびに、弊社関係者へこの場をお借りして深謝申し上げます。ネットワーク分析がさまざまな現場で活用され、より研究が広まることを祈念します。

参考文献など

- [1] 奈良隆正、安田勝通、ソフトウェア品質保証入門、日科技連出版社、2008
- [2] 那須川哲也、テキストマイニングを使う技術／作る技術、東京電機大学出版局、2006
- [3] 大野弘祐、鈴木康雄、テキストマイニングを活用した SW プロジェクトの品質管理事例、ソフトウェア品質シンポジウム 2011、B4-3 セッション、2011
<http://www.juse.jp/sqip/symposium/2011/day2/files/B4-3P.pdf>
- [4] 樋口耕一、社会調査のための計量テキスト分析、ナカニシヤ出版、2014
- [5] Linton.C.Freeman、A Set of Measures of Centrality Based on Betweenness、American Sociological Association、1977
- [6] 神戸雅一、山本修一郎、企業内 SNS への社会ネットワーク分析手法の適用、人工知能学会、第三回知識流通ネットワーク研究会、2008
<http://www4.atpages.jp/sigksn/conf03/SIG-KSN-003-05.pdf>
- [7] 厚生労働省、ヒヤリ・ハット事例情報データベース、2001
<http://www.mhlw.go.jp/topics/bukyoku/isei/i-anzen/jiko/>