

セマンティックグラフモデルに基づくデータ駆動要求獲得方法の提案とステークホルダ分析への適用評価

藤本 玲子^{1,†1} 青山 幹雄^{2,a)}

受付日 2017年8月2日, 採録日 2018年1月15日

概要: データ駆動要求工学 D2RE (Data-Driven Requirements Engineering) を実現するためにセマンティックグラフモデルに基づく情報モデルを提案し, それを用いて仮説設定, データ収集, 分析, 評価を繰り返す要求獲得プロセスモデル A* プロセスを提案する. D2RE の具体化として, データ駆動ステークホルダ分析のための A* プロセスの詳細を定義し, あわせて, その支援環境 D2RA (D2RE Analyzer) のプロトタイプを開発した. 提案方法と D2RA プロトタイプを実際の公共サービス開発会議における発話データへ適用し, ステークホルダの特定と構造化を行い, ステークホルダマトリクスの自動生成を行った. さらに, 対象の公共サービス開発会議の参加者へアンケート調査を行い, D2RE の分析結果と比較し, 提案方法の妥当性, 有効性を示した.

キーワード: 要求工学, 要求獲得, ステークホルダ分析, データ駆動, セマンティックグラフモデル, ビジネスアナリティクス, グラフデータベース

A Data-driven Requirements Elicitation Method Based on Semantic Graph Model and Its Evaluation with Stakeholder Analysis

REIKO FUJIMOTO^{1,†1} MIKIO AOYAMA^{2,a)}

Received: August 2, 2017, Accepted: January 15, 2018

Abstract: We propose a system of D2RE (Data-Driven Requirements Engineering). Based on the D2RE principles, we define the A* process, an iterative process consisting of setting the goal, creating the assumptions, and analyzing the data toward the goal. As a data model of D2RE, we define a semantic graph model for representing the data and their semantic relationships. To support D2RE, we propose D2RA (D2RE Analyzer). As a concrete system of D2RE, we derived D2RE for stakeholder analysis by refining the A* process, data model and D2RA prototype to stakeholder analysis. We applied the D2RE for stakeholder analysis to a set of meeting minutes of a project of public information service system, and identified the stakeholders and their structure. By conducting a survey to the managers involving in the project on the stakeholders, we compare the results of the D2RE with the survey, and demonstrate the validity and effectiveness of the proposed method.

Keywords: requirements engineering, requirements elicitation, stakeholder analysis, data-driven, semantic graph model, business analytics, graph data base

1. はじめに

情報システム開発の源流である要求獲得の重要性が高まっている. しかし, 現行の要求獲得はインタビューなど個人の意見に頼っており, かつ, 獲得を行う要求アナリストのスキルに依存し, 獲得した要求の合理性も明らかとはいえない. さらに, インタビューの対象者が増大するなど

¹ 南山大学大学院理工学研究科ソフトウェア工学専攻
Graduate School of Science and Engineering, Nanzan University, Nagoya, Aichi 466-8673, Japan

² 南山大学理工学部ソフトウェア工学科
Department of Software Engineering, Nanzan University, Nagoya, Aichi 466-8673, Japan

^{†1} 現在, 中部電力株式会社勤務
Presently with CHUBU Electric Power Co., Inc.

^{a)} mikio.aoyama@nifty.com

の要求獲得のスケールアップへの対応が困難である。一方、近年、企業活動や製品に関するステークホルダの意見や利用履歴などのデータを利用してビジネス価値を高めるビジネスデータ分析などの活用が研究、実践されている [8], [10]。このため、要求獲得におけるデータ活用の研究が提案されている [14], [21], [24]。また、ソーシャルネットワークの適用 [20] などの提案もある。しかし、これらの提案は要求獲得の特定のプロセスや対象に限定されており、要求獲得におけるデータ活用の方法が確立されているとはいえない。

本稿では、ビジネスデータ分析に基づく要求獲得方法 D2RE (Data-Driven Requirements Engineering) を実現するために、セマンティックグラフモデル (SGM) [30], [32] に基づく情報モデルを提案し、それを用いた要求獲得プロセスモデルとして獲得目的に基づき仮説設定、データ収集、分析、評価を繰り返す A* プロセスを提案する。D2RE の具体化として、データ駆動ステークホルダ分析のための A* プロセスの詳細を定義し、あわせて、支援環境 D2RA (D2RE Analyzer) のプロトタイプを開発した。提案方法と D2RA プロトタイプを実際の公共サービス開発の議事録データへ適用し、ステークホルダの特定と構造化を行った。さらに、対象の公共サービス開発に参画している関係者へステークホルダについてアンケート調査を行い、D2RE の分析結果と比較し、提案方法の妥当性、有効性を示す。

D2RE によって、ビジネスデータ分析に基づく要求工学の新たなアプローチと技術を拓くことが期待できる。

2. 研究課題

本研究では、以下の 3 点を研究課題とする。

研究課題 1: ビジネスアナリティクス技術を応用した要求獲得のためのモデルとしてセマンティックグラフモデル (SGM) の有効性を明らかにする。

研究課題 2: SGM を用いた要求獲得プロセスを明らかにする。

研究課題 3: 提案方法の有効性、妥当性を実データへ適用して明らかにする。

3. 関連研究

3.1 要求獲得とステークホルダ分析

要求獲得においてステークホルダ分析の重要性が高まっており、様々な研究と実践がある [1], [17], [18], [33]。ステークホルダ分析では、ステークホルダとその利害関係の構造をステークホルダマトリクスなどでモデル化する [3], [17], [18]。しかし、その分析は組織での役割やインタビューなどの人からの情報を用いて、かつ、その実行も人手に頼っており、合理性は明らかとはいえない。これに対し、ソーシャルネットワーク分析 [35], [38] を応用したステークホルダ分析方法 [20] やステークホルダ間の関係の視

覚化 [34] が提案されている。しかし、ステークホルダの識別は影響度に基づくこととされ、その合理的な識別方法は提案されていない。

3.2 BI と BA の要求工学への応用

BI (Business Intelligence) [15] や BA (Business Analytics) [10] などのデータ分析技術をビジネスの新たな価値や顧客の発見へ応用する方法が提案されている [28]。本稿では、これらのデータ分析方法をビジネスデータ分析と呼ぶこととする。ビジネスデータ分析のアプローチを要求工学へ応用する提案がある。たとえば、BI の分析から i* と BIM (BI Modeling) を用いて企業システムをモデル化する方法の提案がある [24]。しかし、提案方法は人手に頼っており、複雑な組織では対処しきれない課題がある。

また、i* モデルを用いてメッセージログを分析し、タスク間の依存関係の分析を行う提案がある [14]。しかし、要求獲得までには至っていない。

筆者らは、D2RE の概念に基づきユーザストーリーをユーザの声形式に変換し、ステークホルダ分析を行う方法を提案した [11]。しかし、ユーザストーリーは要求そのものに近いという問題があり、生のデータを扱える方法が必要である。

3.3 要求工学における議事録や発話などのデータ分析

要求工学において会議の発話履歴や議事録のテキストを分析する方法の研究がある [22]。しかし、多くの会議の発話は、問題と解決策の対などの良い構造であるとは限らず多様な話題が含まれている。これらの方法の適用は限定的である。

3.4 セマンティックグラフの要求工学への応用

ソーシャルネットワークなどの多くの組織的活動はグラフ構造によりモデル化される [35], [38]。グラフのノードとエッジに意味定義を可能にするグラフはセマンティックグラフと呼ばれる [32]。セマンティックグラフのクラスとして、エッジラベルグラフ (Edge-Labelled Graph) とプロパティグラフ (Property Graph) がある。

エッジラベルグラフはエッジにノード間の関係をラベルとして付与できる。たとえば、セマンティック Web の基礎である RDF (Resource Description Framework) [37] は主語から目的語へのノードに述語をラベルとして付与した三つ組みを基礎単位として、グラフを構成する。これに対し、プロパティグラフでは、ノードとエッジにプロパティと呼ぶ属性の集合を付与できることから、エッジラベルグラフに比べ高い表現能力を持つ [30]。そのため、本稿では、プロパティグラフをセマンティックグラフモデル (SGM: Semantic Graph Model) の基礎として導入する。さらに、近年、プロパティグラフを格納し、グラフ構造に基づくク

エリ言語を定義して処置できる GDB (Graph Data Base) が開発されている。例として, Neo4j, JanusGraph がある。本稿では, この中で, 広く利用され安定している Neo4j [25] を利用する。

このような GDB の開発にともないソフトウェア工学における SGM を利用した研究の萌芽が見られる。しかし, SGM を活用した開発方法が確立されるまでには至っていない。

4. アプローチ

4.1 D2RE が解決しようとする課題と解決へのアプローチ

D2RE のアプローチを図 1 に示す。D2RE の実現には次の課題があることを前提とする。

- (1) 明文化されていない情報が多く存在する中で実施するため, 人手に頼らざるをえない場合がある。
- (2) データから得られる要求に関する情報は完全とはいえない可能性がある。
- (3) データから獲得した要求の意味づけにはステークホルダの意見などが必要であると考えられる。

これらの前提条件の下で, D2RE はデータから現状システムをモデル化し, 洞察や気づきを得ることを目的とする。このため, 記述的モデリング [32] のアプローチをとる。

さらに, D2RE は, 従来の人手による要求獲得とコンピュータによるビジネスデータ分析を統合するアプローチをとる。これは, ビジネスデータ分析のみではデータやその結果の意味づけが十分でないためである。また, ビジネスデータ分析を活用することによって, 事実に基づく, 合理的かつ客観的な要求獲得が期待できる。さらに, 対象システムのスケールアップへの対応や獲得効率の向上, 獲得した要求品質の向上が期待できる。

4.2 D2RE における SGM とデータ分析のアプローチ

D2RE では, 要求工学が多様なステークホルダや既存システムなどに関する文書を要求の源泉とする点に着目し, この源泉をデータとして, かつ, 適切な構造化によって要求定義に貢献する分析が可能となるというアプローチをとる。さらに, データで表される要求の源泉の要素間が多様で, かつ, 構造化されているとは限らない点から, データモデルとしてセマンティックグラフモデル (SGM) [32] を採用することとした。

代表的な SGM として関連研究で述べたエッジラベルグラフとプロパティグラフがある。これまで, エッジラベルグラフの 1 つである RDF [37] がセマンティック Web などのデータの意味表現に応用されてきた。また, SNS 分析においてもエッジラベルグラフが利用されてきた [35], [38]。しかし, 要求定義では多様な情報とその間の多様な構造を分析する必要から, 図 2 に主語, 動詞, 目的語からなるトリプルの例で示すようにノードとエッジにプロパティが定

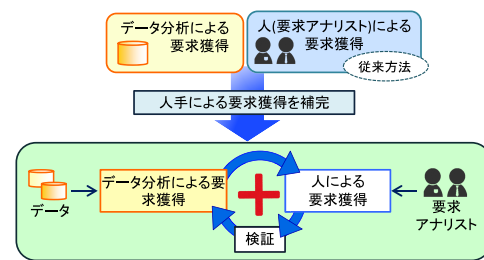


図 1 D2RE のアプローチ
Fig. 1 Approach of D2RE.

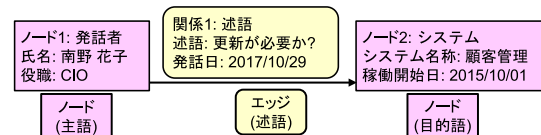


図 2 プロパティグラフに基づく SGM のトリプルの例
Fig. 2 An example of triple in property graph.

義できるプロパティグラフを基礎とすることとした。これによって, ノードに要求の源泉を定義したプロパティグラフに対して, 後述するように, グラフ理論に基づく閉包の概念, 連結性, 中心性, ノード間の測地線距離などのグラフ処理技術を応用して要求の獲得, 分析などが支援可能となる [30], [32], [35], [38]。

これは, 近年プロパティグラフをそのまま格納し, クエリ言語で様々な分析が可能で, かつ, スケーラブルな Neo4j などのグラフ DB の実用化 [30] に着目し, 要求工学への応用を試みるアプローチでもある。

5. D2RE システム

5.1 D2RE システムのフレームワーク

本稿では, D2RE のアプローチによる要求獲得の方法と支援環境をまとめて D2RE システムと呼ぶこととする。D2RE システムのフレームワークを図 3 に示す。

要求アナリストは, 対象システムのデータに対して D2RE システムを利用してデータを分析し, 洞察を得ることが期待できる。

D2RE システムを構成する各要素を次のように定義する。

- (1) D2RE システム原則
D2RE システムを構成, 実現するための原則
- (2) D2RE プロセス (A* プロセス)
D2RE を遂行するプロセス
- (3) D2RE データモデル
D2RE の対象データを表現するデータモデル (SGM)
- (4) D2RA (D2RE Analyzer)
D2RE の遂行を支援する環境

5.2 D2RE システム原則

D2RE システムを構築するために, 次の原則を定めた。このような原則を設定したのは, D2RE システムがデータ

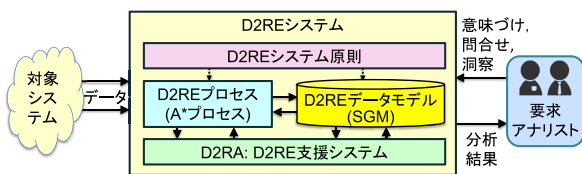


図 3 D2RE システムのフレームワーク
Fig. 3 Framework of D2RE system.

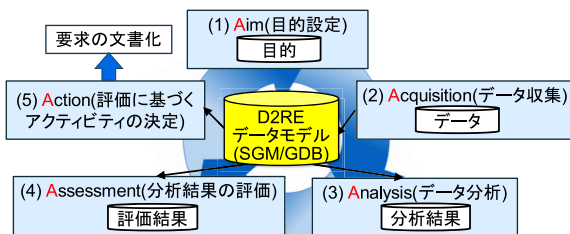


図 4 A* プロセス
Fig. 4 A* Process.

に基づく発見的な活動を中核としており、その結果が予測困難な場合があるため、システム全体としてガバナンスを維持するために必要と判断したためである。

(1) 原則 1：ゴール原則

D2RE は対象システムに関する多様なデータからそのシステムを理解することを目標とする。

(2) 原則 2：プロセス原則

D2RE のゴールを達成するため、対象システムに関する多様なデータに対する仮説と検証を発見的に繰り返すことを可能とすることがプロセスの原則となる。そのため、価値駆動に基づく反復型プロセスを基礎とする。

(3) 原則 3：データモデル原則

D2RE を用いて対象データから洞察を得るためにはデータとその関係の意味定義が必要である。そのため、D2RE のデータモデルとして SGM を基礎とする。

(4) 原則 4：D2RE と要求アナリストとの協調の原則

D2RE は、アプローチで述べたように、機械による自動化ではなく、機械と要求アナリストとの協調を支援することにより優れた洞察を得ることを原則とする。

5.3 D2RE の A* プロセス

5.3.1 A* プロセス構成の原則

D2RE の A* (A Star) プロセスモデルを図 4 に示す。

A* プロセスは、Aim (目的設定)、Acquisition (データ収集)、Analysis (データ分析)、Assessment (分析結果の評価)、Action (アクション) の 5 つのアクティビティを繰り返す構造をとり、これらの頭文字をとって A* と呼ぶ。各アクティビティでは D2RE の共通データモデルである SGM とその実装である GDB を繰り返し利用し、更新する。したがって、A* の各アクティビティにより得られたモデルは SGM 内で連続的に拡張できる。

A* プロセスは、システム原則 (2) で定めたように発見的プロセスである。このようなプロセスでは成果の不確実性が高いことから、不確実性を低減させるために適切にフィードバックできる繰り返し構造が必要となる。近年のアジャイル要求 [19] やリーンスタートアップの基礎である BML (Build Learn Measure) ループ [16], [29] などの経験から短期繰り返し型のプロセス構造の有効性が確認されつつある。特に、要求獲得では不確実性が高いことから、A* においても短期繰り返し型プロセスをその基本構造とし、そのうえで仮説検証型の要求獲得に適したプロセスを定義することとした。

5.3.2 A* プロセス繰り返しの原則

A* が基礎とする短期繰り返し型要求獲得プロセスを適切に繰り返すための原則は、対象システムに関する不確実性の低減にある。これによって、対象システムの適切なモデル化とそれによる知見、洞察を得ることができる。

A* の繰り返し構造は著者らが開発したビジネスモデルデザインのための BML ループを拡張した繰り返し型プロセス [16] に基づいている。この繰り返し型にビジネス分析におけるモデル化で提案されているプロセス [2], [27] を参考に、以下の 3 段階の目的を設定し、これに基づき繰り返すことを原則としている。

- (1) 対象システムのコンテキストの明確化
- (2) 対象システムのモデル化と視覚化
- (3) 対象システムの分析による理解

5.3.3 A* プロセスのアクティビティ

A* の各アクティビティの内容を以下に示す。

(1) Aim (目的設定)

対象システムとその解決すべき問題によって、何を明らかにしたいのかを、目的として設定する。5.3.2 項で定義した 3 つの目的を指針とする。

(2) Acquisition (データ収集)

目的に応じて、分析に必要なデータとその収集方法を検討し、適切なデータを適切な方法で収集し、データベースに格納する。データベースはデータの種類と分析方法に応じて選択する必要がある。本稿では収集した半構造化データをグラフモデルにマッピングし、GDB に格納する。収集するデータの分類とその収集方法は 5.5 節で述べる。

(3) Analysis (データ分析)

収集したデータを分析する。A* では、アプローチで述べたように、SGM を用いてグラフ理論に基づく構造的な分析を提案する。分析方法は目的と対象システムに応じて選択する必要がある。この分析方法については 6 章で例題を用いて説明する。

また、分析方法は 1 つとは限らず、複数の方法を組み合わせる必要がある場合もある。設定した目的と収集データに応じて、目標を満たすまでアナリストが分析を行う。

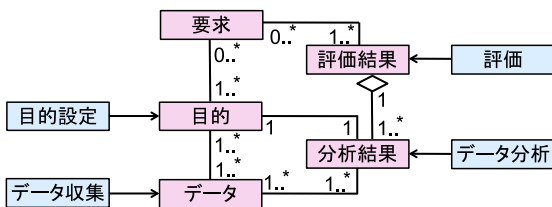


図 5 D2RE のメタモデル
Fig. 5 A metamodel of D2RE.

(4) Assessment (分析結果の評価)

分析結果が設定した目的が達成されているか、またその結果の妥当性や有効性を評価する。結果の解釈や意味づけは、機械的な分析では得られない場合があるので、A*では、必要に応じて人手による結果の補強や、人手による分析結果との比較を行い、結果の有効性や妥当性を高めることも推奨している。

(5) Action (評価に基づくアクティビティの決定)

評価結果から目的が達成されたかどうかを判断する。アクティビティ (1) で設定した目的が達成されていない場合や、さらに新たな目的を設定する場合は、必要なアクティビティに戻り、本プロセスを繰り返す。要求が獲得できたと判断した場合は、獲得した要求を文書化する。

5.4 D2RE データモデル

5.4.1 D2RE メタモデル

D2RE が対象とするデータのメタモデルを図 5 に示す。1つの目的に対して収集するデータは複数ある。一方、データ 1種類に対して複数の目的を達成する必要がある場合がある。1つの目的に対して分析結果は 1つであり、1つの分析結果に対して目的は 1つである。

1種類のデータに対して様々な分析が行えるため分析結果は複数ある。逆に、1つの分析結果を抽出する際に複数のデータを分析する場合も考えられる。

1つの分析結果に対してその評価は 1つであるが、複数の分析結果から 1つの評価を得ることもある。評価結果から、設定した目的と照合し、得たい要求の獲得可否を評価する。ここで、目的は必ず設定されるが、その目的を満たす要求が必ずしも獲得されるとは限らないことから、要求の多重度として 0 も含めている。

5.4.2 D2RE データモデル

D2RE の目的を達成するためのデータモデルとして、プロパティグラフを基礎とする SGM を導入する D2RE では次の理由から SGM を導入する。

(1) 半構造化データの意味表現と解析の必要性

D2RE で対象とするデータは半構造化データ、あるいは、非構造化データである。このようなデータを分析するためには、データの意味定義を表現でき、解析できる必要がある。特に、メールや発話などの文章は半構造化データとし

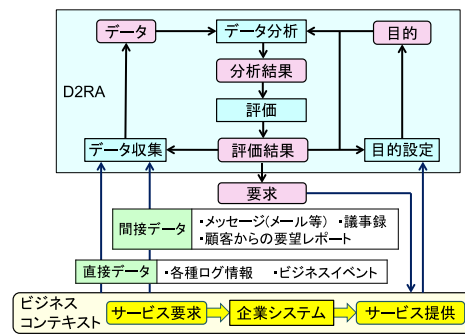


図 6 D2RA の参照アーキテクチャ
Fig. 6 A reference architecture of D2RA.

て解析対象となることから、トリプルグラフはこのようなデータのモデルとして適切である。

(2) 連続的拡張可能性

多様なデータを実行時に追加して、モデルを連続的に拡張できる必要があるため、データ構造が実行時に拡張できる必要がある。SGM とその実装である GDB は実行時に拡張可能なデータモデルである。

(3) スケーラビリティ

大量の相互に依存したデータを扱え、クエリにより分析できる必要がある。

5.5 D2RA (D2RE Analyzer) と分析対象データ

D2RE の A* プロセスとデータモデルに基づき、ビジネスデータ分析を実行する支援システムを D2RA (D2RE Analyzer) と呼ぶ。図 6 に D2RA の参照アーキテクチャを示す。A* プロセスの中で、目標設定を除く、D2RE データモデルを利用しながら行う 4 つのアクティビティをそれぞれ支援する 4 つのコンポーネントを統合している。

本稿では、D2RA で解析するデータを、次のように直接データと間接データに分類して扱う。

(1) 直接データ

ビジネス活動やそれを支援する情報システムから直接得られ、ビジネス活動や情報システムの状況や現象を表現するデータである。たとえば、工場の生産管理システムなどから送られる生産状況のイベントやデータ、サプライチェーンや流通における物流の移動状況のログデータ、業務プロセスのモニタリングから得られるビジネスイベントやデータがある [9], [15]。組込み機器などのプロダクトでは、ユーザの操作ログや故障などのログデータがある [36]。

これらのデータは一般に抽象度が低いことから、BI などを利用して業務や情報システムの現状の理解やボトルネックなどの問題点の特定などに利用されている。そのため、直接データを利用して将来システムの要求を獲得するためには、D2RA 上でデータ分析とその結果の意味付け、問題の特定といった段階的に抽象度を高めるなどの目的設定を行い、A* プロセスを繰り返す。

(2) 間接データ

情報システムや業務の遂行から直接得られる直接データに対し、その利用者やステークホルダなどから対象の情報システムや業務に関する意見やレポートなどの形態で間接的に得られるデータである。たとえば、情報システムや製品のユーザからの要望や操作性などのレポート [7], [13], 業務や情報システムへの要求に関するステークホルダへのインタビューや会議の発話、議事録などのデータ [4], [5] がある。

これらのデータは、一般に抽象度が高く、意味情報を含む反面、発話や自然言語で記述された文書であることが多いことから自然言語処理が求められる。また、レポートや議事録などでは、文書や発話が半構造的であることから、その構造分析も必要となる。

しかし、このようなデータは業務や情報システムの要求に関する豊富な情報を含むことが期待できる。このため、本稿では間接データを分析対象とする。

6. データ駆動ステークホルダ分析

6.1 D2RE に基づくデータ駆動ステークホルダ分析の実現と適用

本稿では D2RE の具体化として、データ分析の目的をステークホルダの識別と分析として A* プロセスを詳細化し、データ駆動ステークホルダ分析の A* プロセス、データモデル、D2RA プロトタイプを実現した。

さらに、この詳細化した A* プロセスを実際の公共サービス開発の議事録に適用し、ステークホルダを分析した。本議事録は、管理職が中心となり、「データカタログサイトの記載比率向上」システムの要求定義に関する 5 回の会議で話し合われた内容を文書化した、約 3,000 文字の発話データである。これを議事録データと呼ぶこととする。

以後、この適用事例を用いてデータ駆動ステークホルダ分析を示す。

6.1.1 A* プロセスのステークホルダ分析プロセスへの詳細化

ステークホルダ分析を分析全体の目的として、それに対して 5.3.2 項で述べた繰返しの原則に基づき、繰返しごとの 3 つの目的を次のように設定した。この目的に対して、詳細化した A* プロセスを次の 3 回にわたり繰り返した。

- (1) イテレーション 1：目的 1 = 発話構造のモデル化と視覚化
- (2) イテレーション 2：目的 2 = 発話者ステークホルダ分析
- (3) イテレーション 3：目的 3 = 話題ステークホルダ分析

ここで、目的 1 は 5.3.2 項の (2) に対応し、目的 2, 3 は同じく (3) に対応している。5.3.2 項 (1) の「対象システムのコンテキストの明確化」は、本適用事例では対象システムのコンテキストがあらかじめ呈示されていたことから

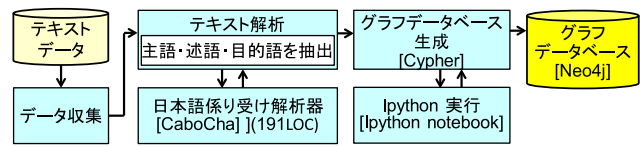


図 7 テキストデータのグラフデータベース変換システム
Fig. 7 A prototype for generating graph data base from text data.

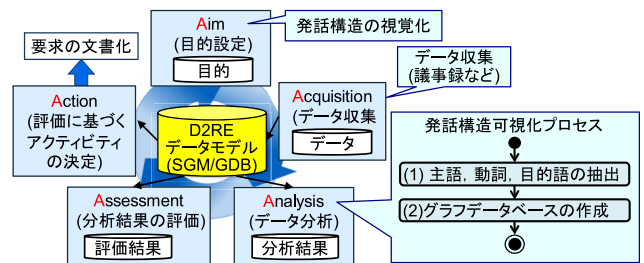


図 8 発話記録の視覚化プロセス
Fig. 8 Visualization process of speech record.

省略可能であると判断した。一方、5.3.2 項 (3) の「対象システムの分析による理解」に対しては、ステークホルダを 2 つのタイプに分けて分析することとした。このため、それぞれ目的を設定し、2 回の繰返しに分けて分析することにより、データ収集、分析とその結果の評価が簡略となり、かつ、明確になることを狙いとした。

なお、アプローチで示したように、本事例でも人手による分析も行っているが、提案するデータ駆動ステークホルダ分析方法を評価するベンチマークとして利用することとしたため、イテレーションは 3 回までとした。

6.1.2 D2RA プロトタイプ

図 6 に示した D2RA のアーキテクチャに基づき、収集した日本語テキストデータを解析し、GDB を生成するプロトタイプの構成を図 7 に示す。

テキストデータの解析には、日本語係り受け解析器である CaboCha [6] モジュールを Ruby で拡張して実装した。これを用いて、テキストから主語、述語、目的語のトリプルを抽出する。それをもとに Neo4j [25] を用いてグラフデータベースを生成する。Neo4jRestClient を用いて iPython notebook ライブラリで Python から Neo4j に接続し、解析結果を Neo4j GDB へ格納する。このために、Neo4j のクエリ言語である Cypher を用いている。

6.2 イテレーション 1：発話構造のモデル化と視覚化

6.2.1 A* に基づく分析プロセスと目標設定

イテレーション 1 において、日本語テキストデータから発話構造を SGM に変換し、ステークホルダとその発話意図、ならびに、発話意図間の構造を視覚化するよう詳細化した A* プロセスを図 8 に示す。

ここで、目的を「発話構造のモデル化と視覚化」とした。

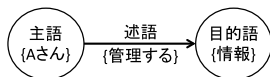


図 9 トリプルグラフ
Fig. 9 A triple graph.

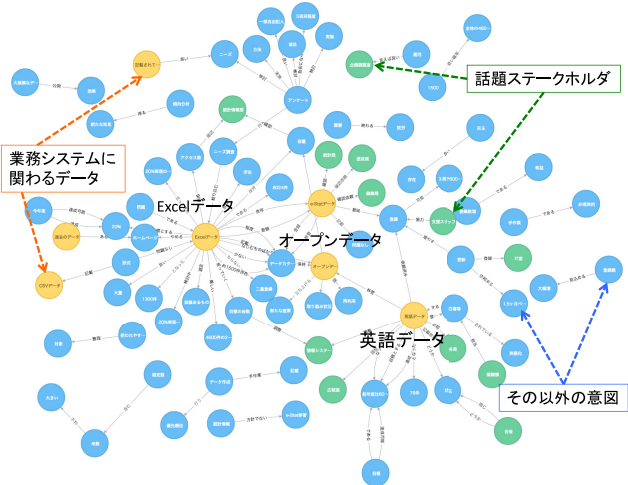


図 10 発話グラフ
Fig. 10 Speech graph.

6.2.2 データ収集：日本語発話データから GDB の生成

分析対象データを間接データの中のステークホルダの発話などが記述された日本語テキストデータとする。

本事例では、発話データそのものは収集済みであった。しかし、それを分析するために、以下のプロセスと方法によってグラフモデルにマッピングし、GDB に格納した。

(1) 発話からトリプルの抽出

1つの発話データから図 9 に示す「主語」、「動詞」、「目的語」のトリプルグラフ（以下、トリプルと略記）を基礎単位として発話内容を表現する。トリプルを集約して SGM を生成し、その構造をグラフとして視覚化する。そのため、日本語係り受け解析器 CaboCha [6] を用いて、「主語」、「述語」、「目的語」のトリプルを抽出する。

(2) GDB の作成

アクティビティ (1) で抽出したトリプルから、発話構造のグラフデータベースを生成する。たとえば、「Aさんは情報を管理する」という文があり、(1) で主語として「Aさん」、述語として「管理する」、目的語として「情報」のトリプルを抽出した場合、図 9 に示すトリプルを生成できる。

6.2.3 データ分析：発話構造の GDM によるモデル化と視覚化の評価

議事録の発話データから抽出したトリプルを用いて生成したグラフを発話グラフと呼ぶこととする。生成した発話グラフを図 10 に示す。「データカタログサイト」に関するテーマについて話し合われているため、様々な種類のデータが話題になっていることが分かる。黄色ノードがそのような対象業務システムに関わるデータ、緑色ノードが発話

記録に記載されたステークホルダ、青色ノードがその他の発話者の関心事や意図を表している。

本グラフから、オープンデータと隣接する黄色ノードは Excel データと英語データであったことから、Excel データと英語データをオープンデータとして記載したいのではないかという意図を推定できる。また、Excel データに隣接するノードが多いことから、Excel データに関する議論が活発に行われていることが分かる。このことから、Excel データがステークホルダの関心の焦点であることが推定できる。そのため、さらに、Excel データを中心とするサブグラフをクエリによって抽出した。このサブグラフを分析すると、「問題である」、「厳しい」といったネガティブな表現が多いことが分かった。

6.2.4 分析結果の評価と評価に基づくアクティビティの決定

分析結果から、Excel データをオープンデータとしてデータカタログサイトに記載するにあたって、何か問題が生じている、あるいは生じる可能性があるかと推定できる。このことから、発話データに基づき発話者をステークホルダとして分析することで要求獲得の手がかりが得られることが期待できる。したがって、新たに目的を設定し、A* プロセスを繰り返すこととする。

6.3 イテレーション 2：発話者ステークホルダ分析

6.3.1 A* に基づく分析プロセスと目的の設定

イテレーション 2 では、「発話者ステークホルダ分析」を目的とし、各発話者の発話意図に着目することで、発話者とその間の関係を構造化して分析する。

6.3.2 データ収集：発話データの GDB から発話意図の抽出

発話データの GDB からステークホルダの発話意図を特定するために、以下の手順で発話内容を分類する。これを GDB のプロパティに付与し、GDB を拡張する。

すべての発話者の中でステークホルダの候補と考えられる発話者を、発話者ステークホルダと呼ぶこととする。

(1) 発話内容の分類

テキスト解析された発話データのトリプルから、意図を表す述語に基づき分類することで、発話意図を特定する。日本語の一般的な発話の解析は、日本語の口語表現の多様性のために困難である。しかし、議事録などの発話データは一定の規則に則り記述されていることから、本稿では、一般の発話に比べ表現に規則性が見られることに着目している。さらに、日本語文章の構造化解析結果によれば、新聞記事のような一定の規則に基づき記述された日本語文章の主張文は、問い掛け、断定、推量、要望、判断、意見、理由、義務の 8 つに分類できることが指摘されている [12]。

このような前提に基づき、対象とする議事録を含む約 200 件の発話データを文法とその用例の分類 [23] に基づき

表 1 発話意図の分類規則と用例

Table 1 Classification example of speech intention.

発話意図	用例	品詞と日本語の文法用法
報告	である, ~した	断定の助動詞, 接続助詞, 補助動詞, 動詞, 助動詞. 過去, 完了.
返答	その通り	連体詞, 名詞. 相手の見解への同意.
受入	検討する	動詞サ行変格活用終止形. 言い切り.
問い	か	終助詞. 疑問, 質問.
要望	たい	助動詞. 希望.
示唆	良い, ~と考える, ~と思われる	形容詞, 終止形. 接続助詞, 動詞.

表 2 発話意図の重み

Table 2 Weighted of speech intention.

発話意図	報告	返答	受入	問い	要望	示唆
語尾例	である	その通り	検討する	か	たい	良い
重み	5	5	15	15	20	40

分析した発話意図の分類規則と対応する用例を表 1 に示す。要求に関する会議の発話であることから、「報告」、「示唆」、「要望」、「問い」、「受入」、「返答」の 6 つに分類した。この分類を本稿では発話意図と呼ぶこととする。これに基づき、発話ごとに分類をプロパティとして付加し、SGM を拡張し、以後の構造分析の手がかりとした。

(2) 各発話内容への重み付け

発話の種類、すなわち、発話意図によって、プロジェクトの計画や実行の変更への影響度が異なると考えられる。このため、分類した各発話の影響度を分析し、表 2 に示す重みを付与することとした。各発話意図の重みは、6 種の意図の合計が 100 となるように配分した。これは、要求の優先順位づけで用いられている「100 ドルテスト」の方法を応用した。たとえば、「報告」は現時点での事実報告を行っており、今後に対する新たな意見は含まれないと判断し、低い重みを付けた。これに対し、「示唆」はある事項に対する意見や、新たな視点で発話しているため、プロジェクトに対して大きな影響を及ぼす可能性があるとして判断した。

6.3.3 データ分析：発話者ステークホルダ分析

(1) 発話者ステークホルダ分析の方法

発話者ステークホルダの意図の構造に基づく、発話者ステークホルダ分析を行う。ステークホルダ分析では、一般にステークホルダマトリクスを用いて分析結果を表現する。このため、本事例でもステークホルダマトリクスの生成を分析結果の成果とする。

次の手順でステークホルダマトリクスを生成する。

1) 影響度と関与度の評価

各発話者に対して影響度と関与度を評価する。

一般に、ステークホルダ分析では、ステークホルダを重要度と影響で評価している。重要度とは、要求獲得への貢献の高さを意味している。しかし、このままでは評価が困難であるので、本事例では重要度の代わりに関与度を用い

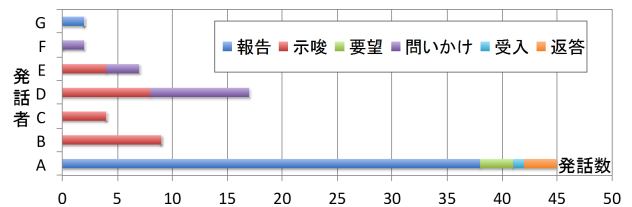


図 11 発話者ごとの発話意図

Fig. 11 Speech intention per speaker.

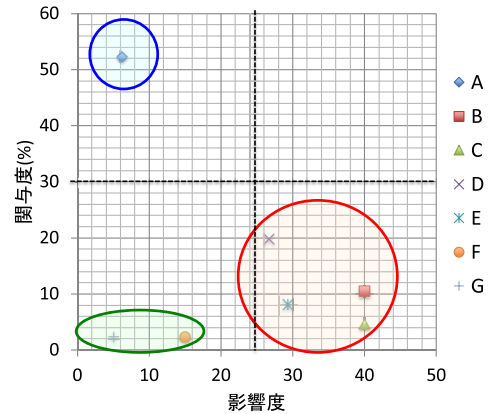


図 12 関与度と影響度

Fig. 12 Involvement rate and influence rate.

ることとした。

ここで、関与度とは要求獲得への関与の高さを意味し、発話者の発話数多いほど関与が大きいといえることから、式 (1) で定義する。

一方、影響度とは発話意図が要求獲得の結果に及ぼす影響の大きさを表す。このため、発話に発話意図の重みづけを行い、式 (2) で定義する。

$$\text{関与度} = \frac{\text{特定発話者の全発話数}}{\text{全発話数}} \times 100 \quad (1)$$

$$\text{影響度} = \frac{\sum(\text{各重み} \times \text{特定発話者における各発話内容の全発話数})}{\text{特定発話者の全発話数}} \quad (2)$$

2) 発話者ステークホルダマトリクスの生成

算出した関与度と影響度を用いて発話者ステークホルダマトリクスを生成する。

(2) 発話者ステークホルダ分析の事例への適用

図 11 に発話者ごとの発話内容の頻度分布を示す。横軸が発話数、縦軸が発話者である。B と C、D と E の役割がそれぞれ類似していることが分かる。

関与度と影響度を基準として生成したステークホルダマトリクスを図 12 に示す。

左上の枠内に位置する A の影響度は低いため意思決定権は低い、関与度が高いことから開発に大きく関与していることが分かる。よって、A は開発を進めるうえで多くの関与を果たしている人物であると判断できる。右下枠内に位置する、赤丸内の 4 人の発話者の関与度は高くはない

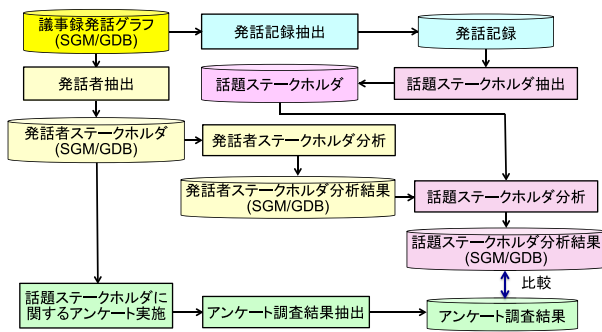


図 13 ステークホルダ分析のプロセス
Fig. 13 Process of stakeholders analysis.

め、プロジェクトに大きな関与はしていないが、影響度が高く、計画や実行変更の決定に大きく影響を及ぼす集団であると判断できた。このような関与度が低いにもかかわらず影響度が高いステークホルダは要求獲得で見落とされがちであり、このようなステークホルダの特定が重要であることが指摘されてきた [18]。

左下枠内に位置する G と F は、他の発話者と比べて関与度、影響度ともに低いため、積極的にプロジェクトに参加しているとはいえない。

6.3.4 分析結果の評価と評価に基づくアクティビティの決定

ステークホルダ分析の結果、発話者のデータから発話者を対象としてステークホルダマトリクスが生成できた。しかし、発話データには発話者以外のステークホルダも含まれている。ステークホルダ分析においてステークホルダの網羅性は重要であることから、発話データに含まれる発話者以外のステークホルダも含めて、A*プロセスを繰り返すこととした。

6.4 イテレーション 3：話題ステークホルダ分析

6.4.1 A*に基づく分析プロセスと目的設定

イテレーション 3 では、「話題ステークホルダ分析」を目的として設定し、話題ステークホルダとその関係を構造化する。

話題ステークホルダとは発話者ステークホルダの発話データに記載されている人や組織である。発話者以外のステークホルダも含まれる可能性があるため、発話者ステークホルダと区別する。

ステークホルダ分析プロセスを図 13 に示す。

発話者ステークホルダの分析結果を用いて、発話データから抽出した話題ステークホルダを分析する。さらに、発話者ステークホルダを対象とした、話題ステークホルダに関するアンケートを実施し、データ分析で明らかになった結果とアンケート結果を比較し、ステークホルダの特定や分析結果の妥当性、正当性を評価する。

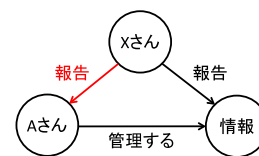


図 14 3者閉包
Fig. 14 Triadic closure.

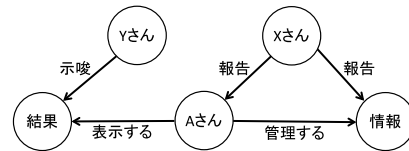


図 15 グラフの測地線距離
Fig. 15 Geodesic distance.

6.4.2 データ収集：GDB への発話者データの付加

6.1 節で作成した GDB で発話者を表現するように発話者情報を GDB へ付加する。各ノードを発話者の意図として、主語と目的語の各ノードに、発話意図の情報を関係とする有向グラフを付加する。ここで、SGM の 3 者閉包 (Triadic Closure) [30] の原理を用いている。図 14 に示す例では、「A さんは情報を管理する」と X さんが発言したとする。この文の発話意図は報告と判断できるので、X さんから A さんへ「報告」のリンクがグラフの 3 者閉包として付加される。

これによって、発話者ステークホルダと話題ステークホルダの両方を含むすべてのステークホルダを網羅した SGM が構成される。さらに、この SGM は単一の GDB に格納されていることから、単一のクエリで分析可能である。

6.4.3 データ分析：話題ステークホルダ分析

前項の手順で構成された SGM を用いて、次の手順で話題ステークホルダを分析する。

- (1) SGM の構造分析による話題ステークホルダ分析方法
 - 1) ステークホルダの構造分析
 ステークホルダ間の関係を SGM の構造分析により定量的に評価する。SGM のノード i からノード j までの最短経路の距離である測地線距離 [35] を式 (3) に示す。

$$\text{測地線距離} = d(i, j) \tag{3}$$

ステップ (1) で情報を付与した拡張された GDB から、各話題ステークホルダと各発話者ステークホルダの意味的な距離を測地線距離 $d(i, j)$ により評価する。

図 15 に示す例では、話題ステークホルダである A さんと発話者ステークホルダである X さんは隣接しているため、測地線距離 $d(A, X) = 1$ である。それに対して A さんと Y さんは「結果」ノードを媒介要素としており、測地線距離 $d(A, Y) = 2$ となる。

2) 話題ステークホルダの重要度と期待度の評価

上記で求めた測地線距離と、6.3.3 項で求めた発話者ステークホルダの関与度と影響度を用いて、話題ステークホ

ルダの重要度と期待度を求める。

発話者ステークホルダにおいて、関与度が高い人は会議で積極的に発言している人物であり、プロジェクトに大きく関与していることから、現状報告の発話が増えると考えられる。この発話の度合いを発話者ステークホルダ m の関与度を用いて、 m による話題ステークホルダ n の重要度を $i(n, m)$ として式 (4) で定義した。

一方、発話から得られた話題ステークホルダ n の重要度は、発話から得られた各話題ステークホルダ関与の大きさを表していると考えられるので、 $i(n)$ を話題ステークホルダ n の関与度の平均として式 (5) で定義した。関与度が高い発話者ステークホルダと近い関係にあるほど、現時点でプロジェクトに深く関わっていることを意味している。

$$\begin{aligned}
 i(n, m) &= \frac{\text{発話者ステークホルダ } m \text{ による話題ステークホルダ } n \text{ の重要度要素}}{\text{ } m \text{ との測地線距離}} \times m \text{ の関与度} \\
 &= \frac{d(n, 1) + d(n, 2) + d(n, 3) + \dots + d(n, g)}{d(n, m)} \\
 &\quad \times m \text{ の関与度} \\
 &= \frac{\sum_{j=1}^g d(n, j)}{d(n, m)} \times m \text{ の関与度} \quad (4)
 \end{aligned}$$

$$\begin{aligned}
 i(n) &= \frac{\text{話題ステークホルダ } n \text{ の関与度要素の総和}}{\text{発話者ステークホルダ数}} \\
 &= \frac{i(n, 1) + i(n, 2) + i(n, 3) + \dots + i(n, g)}{g} \\
 &= \frac{\sum_{j=1}^g i(n, j)}{g} \quad (5)
 \end{aligned}$$

発話者ステークホルダの影響度とは、プロジェクトの計画や実行に対する意思決定に及ぼす影響度である。それを用いて、重要度と同様に発話者ステークホルダ m による話題ステークホルダ n の期待度 $e(n, m)$ を式 (6) で定義した。

さらに、重要度と同様に話題ステークホルダ n の期待度 $e(n)$ を式 (7) で定義する。期待度は、影響度が高い発話者ステークホルダと近い関係にあるほど、今後プロジェクトに深く関わっていく可能性を示している。

$$\begin{aligned}
 e(n, m) &= \frac{\text{発話者ステークホルダ } m \text{ による話題ステークホルダ } n \text{ の期待度要素}}{\text{ } m \text{ との測地線距離}} \times m \text{ の影響度} \\
 &= \frac{d(n, 1) + d(n, 2) + d(n, 3) + \dots + d(n, g)}{d(n, m)} \\
 &\quad \times m \text{ の影響度} \\
 &= \frac{\sum_{j=1}^g d(n, j)}{d(n, m)} \times m \text{ の影響度} \quad (6)
 \end{aligned}$$

$$e(n) = \frac{\text{話題ステークホルダ } n \text{ の期待度要素の総和}}{\text{発話者ステークホルダ数}}$$

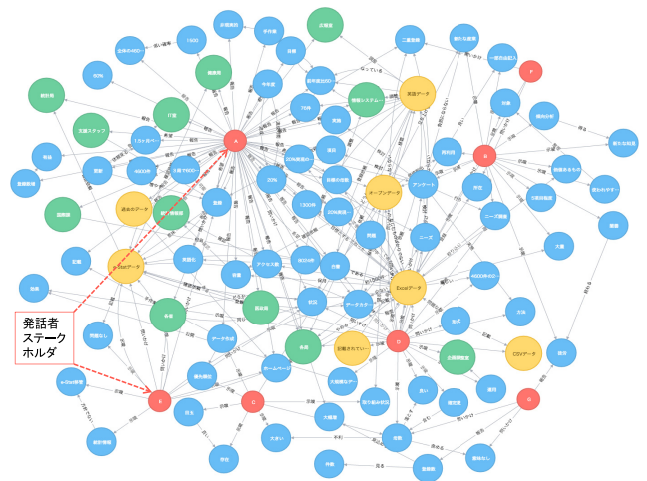


図 16 発話者付き発話グラフ

Fig. 16 Speech graph with speaker.

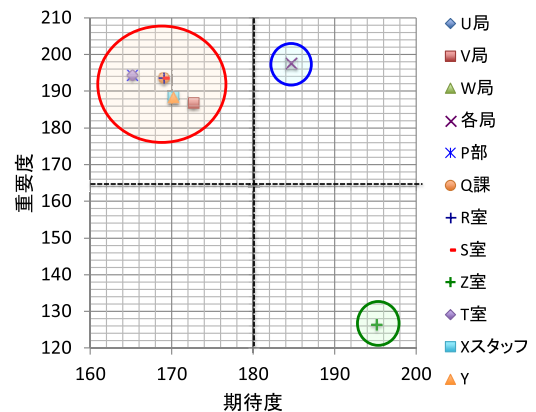


図 17 重要度と期待度

Fig. 17 Importance rate and expectation rate.

$$\begin{aligned}
 e(n) &= \frac{e(n, 1) + e(n, 2) + e(n, 3) + \dots + e(n, g)}{g} \\
 &= \frac{\sum_{j=1}^g e(n, j)}{g} \quad (7)
 \end{aligned}$$

3) 発話者ステークホルダマトリクス作成

議事録データから得られた重要度と期待度を用いて話題ステークホルダマトリクスを作成する。

(2) 話題ステークホルダ分析の事例への適用

図 10 に示した発話者を赤色ノード、発話内容をラベルとして付与した結果を図 16 に示す。Cypher クエリを用いて、各発話者ステークホルダと各話題ステークホルダとの測地線距離を求め、さらに、発話者ステークホルダの関与度と影響度を用いて、話題ステークホルダの重要度と期待度を求めた。その結果を図 17 に示す。

図 17 の右上 (第 1 象限) に位置する「各局」は、重要度、期待度ともに高く、現在も今後もプロジェクトに大きく関わることが推測できることから、プロジェクトを進めるうえで重要な組織であることが分かる。左上 (第 2 象限) に位置する「U局」、「V局」、「W局」、「P部」、「Q課」、「R

室], [S 室], [T 室], [X スタッフ], [Y] は, 重要度が高いため現段階においてプロジェクトで大きな役割を果たしていると判断できる. しかし, 期待度は高くないため, 今後その役割は低下する可能性がある. 右下 (第 4 象限) に位置する [Z 室] の重要度は低く, プロジェクトに大きな役割を果たしているとはいえないが, 期待度が大きいため, 今後プロジェクトに大きく関わり, 重要な役割を果たす可能性があると考えられる. また, 各局には [U 局] や [V 局], [W 局] が含まれている. 本プロジェクトでは各局が重要なステークホルダであるが, 特に, 前述した 3 つの局が注目されていることが明らかになった.

6.4.4 分析結果の評価と評価に基づくアクティビティの決定

発話者ステークホルダ分析では, 図 11, 図 12 に示すように, 7 人の発話者をステークホルダとして分析した. これに対し, 話題ステークホルダ分析では, 図 17 に示すように, 12 の組織をステークホルダとして分析し, 組織の利害を明らかにできた. このことから, ステークホルダ分析として一定の成果を得られたといえることから, A* プロセスを終了した. さらに, 本事例では次章で述べるように, 発話ステークホルダにアンケート調査を行い, データ分析に基づく結果の妥当性などを評価している.

7. 提案方法の評価

実際に会議に参加した発話者ステークホルダを対象に話題ステークホルダの重要度に関するアンケート調査を行い, 提案したデータ駆動ステークホルダ分析方法を評価する基準とした.

アンケート調査は, 発話者ステークホルダとして特定した 7 人のうち 6 人の協力を得た. 発話者ごとの影響度を考慮し, 影響度による重み付け (表 3) を行い, 加重平均を求めた. アンケート調査結果はプロジェクト全体を通じた評価であると判断し, データ分析結果は重要度と期待度を 5 段階で評価し直した結果 (表 4) と, その平均を示している.

アンケート調査結果とデータ分析結果を比較するため, 図 18 に散布図として示す.

縦軸がアンケート調査結果から得た加重平均値, 横軸がデータ分析結果から得た平均値である.

$y < x$ 領域に位置する R 室は, アンケート調査結果よりもデータ分析結果で重要であると判断されている. 逆に, $y > x$ 領域に位置する 6 つの部門は, アンケート調査結果では, データ分析結果より重要であると判断された.

このようにデータ分析の結果とアンケート調査結果とは, 必ずしも一致しないと考えられる. このような差異が発見できることが D2RE の意義であるとともに, アンケート調査を併用する意義でもある.

表 3 発話者ごとの重み

Table 3 Weighted per speaker.

	B	E	D	F	A	G
影響度	40	30	27	15	6	5
影響度による重み	8	6	5	3	1	1

表 4 5 段階評価

Table 4 Five-grade evaluation.

重要度と影響度	~159	160~	170~	180~	190~
5 段階評価	1	2	3	4	5

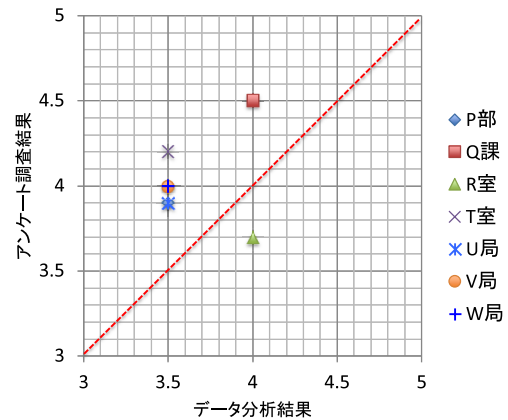


図 18 データ分析結果とアンケート調査結果の比較

Fig. 18 Comparison of survey result with analysis result.

8. 考察

ビジネスアナリティクスなどのデータ分析技術を要求獲得へ導入する D2RE のフレームワークを実現するために, SGM を用いたデータモデルとそれに基づく仮説検証を繰り返す反復型プロセス A* を提案し, あわせて, 支援ツール D2RA のアーキテクチャを提示した.

さらに, 実際の議事録データへ適用し, ステークホルダ分析を行った結果, 発話者ステークホルダと話題ステークホルダの識別を行い, ステークホルダマトリクスを生成することで, ステークホルダの役割の特定と構造化が可能となった.

これらの結果は, 本稿の提案方法が要求工学の新たなアプローチとその効果を示した点で意義があるといえる.

8.1 研究課題に対する提案方法の妥当性に関する考察

本稿の研究課題について, 提案方法の妥当性を考察する.

(1) 研究課題 1: SGM の有効性について

提案方法では, SGM を導入したことにより, 次の 2 点で要求獲得に有効と考えられる.

1) データの意味表現: データとその間の関係が意味づけられることにより, 要求獲得に求められるデータ表現が可能となった, 従来のデータをそのまま解析する方法 [21], [24] に比べ, 獲得できる情報の量と質が高まること

が期待できる。

2) GDB を用いたグラフ理論に基づくデータ解析の活用：データ駆動ステークホルダ分析で示したように、3 者閉包などのグラフ構造に基づく意味推論と測地線距離に基づくステークホルダ間の関係の定量的評価が可能となることから、従来の役割などに基づく定性的なステークホルダ分析と比較し、合理的な分析を実現できる可能性がある。

このようなグラフ理論に基づく分析は、ソースデータを SGM へ変換後は Neo4j が提供するクエリ言語 Cypher を利用して、ソースデータの種類によらず、かつ、意味的構造を含むデータとして一様に解析が可能となることから、多様なデータを入力として要求獲得に求められる高度なデータ分析が可能となる。あわせて、従来は人手で作成していたステークホルダマトリクスが自動生成可能となったことにより、ステークホルダ分析において D2RA による一部のアクティビティの自動化とその有効性を見通しを得た。

(2) 研究課題 2：要求獲得プロセスについて

本稿の対象とする要求獲得では、不確実性が高く、あらかじめ結果が予測できない可能性が高い。このような課題に対して、これまで Inquire Cycle [26] などの繰返し型要求獲得プロセスが提案されている。本稿では、このようなプロセスを仮説検証プロセスとしてモデル化し、さらに、データ解析と統合することにより、有効性を高めることを可能としている。これは、繰返し型要求獲得プロセスの新たなモデルを提示している点で意義があるといえる。

(3) 研究課題 3：実データへの適用評価について

従来のステークホルダ分析ではインタビューなどによる人からのデータを要求アナリストが人手で分析するため人間的要因が 2 重の障壁となっている [28]。さらに、ステークホルダの表現として利用されているステークホルダマトリクス上のステークホルダの位置づけも定性的になりがちである。提案方法は、データ分析に基づき、ステークホルダのプロジェクトへの関与を定量的に分析可能とする点で有効性が高いといえる。

しかし、日本語の発話データ分析だけでは、主語と述語の対からステークホルダを抽出できない場合や、同一ステークホルダであっても、異なる名称で表現されていることがあるという、自然言語処理上の課題も明らかとなった。今後、検討すべき課題である。

8.2 従来の要求工学と提案方法の比較

従来の要求獲得方法では、主としてステークホルダなどの人から獲得したデータを要求アナリストが人手により分析する方法が主体である [33]。また、ステークホルダ分析を支援するツールの提案もあるが、構造の表示にとどまっている [3]、[34]。提案方法はデータ分析を活用することにより、事実に基づき、かつ、人手に頼る分析作業を回避し、より合理的で妥当な要求の獲得が期待でき、獲得した要求

の品質向上も期待できる。

8.3 提案方法と分析対象データの適用拡張性

本稿では、D2RE をステークホルダ分析へと詳細化し、かつ、分析対象データとして会議における発話データを利用した。しかし、D2RE をより広範な要求獲得へ適用するためには、提案方法と利用データの両面で適用の拡張が必要である。

(1) 提案方法の適用拡張性

提案方法である A* プロセスは、データに基づくモデル化と分析の一般的プロセスとなっていることから、その拡張性は高いと考えている。たとえば、関連研究 [31] では Facebook などの SNS のデータ分析を要求獲得、優先順位付け、交渉などに利用する可能性を議論している。しかし、データに基づくモデリングと分析の方法は対象データとその構造によることから、本稿のデータ駆動ステークホルダ分析で示したようにカスタマイズが必要である。今後、このカスタマイズ方法とその適用可能性を研究する必要がある。

(2) 利用データの適用拡張性

利用データの適用拡張性に関しては、5.5 節で述べた分析対象データの 2 つの分類基準に沿って議論する。

1) 直接データへの適用可能性

データ駆動に基づく要求獲得における直接データの利用については、関連研究 [7]、[21] でアプリケーションに対するユーザの利用履歴などの利用が提案されている。本提案方法においても、このようなデータの利用を想定している。特に、関連研究との違いは、利用履歴をトリプルで定義し、SGM でその間の関係を定義し、GDB による分析を行うことにより、従来の統計的分析と比較し、意味的な分析を通じたユーザ要求への洞察を得ることが期待できる。これについては、今後の課題である。

2) 間接データにおける他のデータの利用可能性

本稿で用いたデータに加え、著者らは GitHub 上で OSS (オープンソースソフトウェア) に対する開発者やユーザなどのコメントへ適用し、SGM に基づく分析方法の有効性を確認している [11]。また、関連研究 [4]、[13] においてもユーザコメントデータなどの利用事例がある。このような間接データは直接データと比較し、ユーザの利用に関するコンテキスト情報を含むことがあり要求獲得に利用可能であることも指摘されている。しかし、このようなコメントが必ずしもユーザなどの意図を表さないことも指摘されている [5]。したがって、データの種類とその利用方法について、留意する必要がある。

9. 今後の課題

今後、下記の課題を研究する必要がある。

(1) データ駆動ステークホルダ分析における発話データ以

外のデータを含むステークホルダ分析への拡張性の評価

(2) D2RE のステークホルダ分析以外の要求獲得への拡張性の評価

要求分析における要求の優先順位付けや要求交渉など、多様なステークホルダの利害が関与する分析 [18], [31] への適用方法とその効果の評価。

(3) 機械学習を適用することによるデータ解析の精度向上と大量データ処理方法の研究

10. まとめ

ビジネスデータ分析に基づく要求獲得方法 D2RE の概念を実現するために、セマンティックグラフモデルとその分析を利用して仮説検証を繰り返す A* プロセスを提案した。この具体化として、データに基づくデータ駆動ステークホルダ分析の詳細なプロセスとデータモデルを定義し、支援環境 D2RA のプロトタイプを実装した。

提案したステークホルダ分析方法を実際の公共サービス開発の会議における発話データへ適用し、ステークホルダの特定とその構造を明らかにし、ステークホルダの自動生成を可能とした。さらに、公共サービス開発の会議に参加している管理者にアンケート調査を行い、人手によるステークホルダの構造と提案方法により特定したステークホルダの構造とを比較し、提案方法の妥当性、有効性を示した。

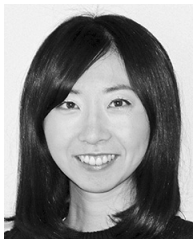
本研究は、ビジネスデータ分析を要求工学へ応用する具体的方法を提案し、グラフ理論に基づく要求工学の新たなアプローチとその実現技術を提案した点で、意義があると考えている。

謝辞 議事録データをご提供いただいた伊藤忠テクノソリューションズ株式会社の野村典文氏と関係各位に感謝する。

参考文献

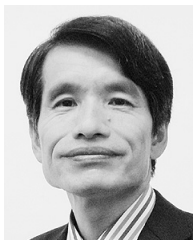
- [1] 青山幹雄, 木下康介, 山下和希: 動的利害相互作用に基づくステークホルダ分析手法の提案と節電問題への適用評価, *コンピュータソフトウェア*, Vol.30, No.3, pp.102–108 (2013).
- [2] Behal, A. et al.: Business Insights Workbench – An Interactive Insights Discovery Solution, *Proc. HCII 2007, Part II*, LNCS, Vol.4558, pp.834–843, Springer (July 2007).
- [3] Bourne, L.: *Stakeholder Relationship Management, Revised ed.*, Gower Pub. (2012).
- [4] Burnay, C.: Are Stakeholders the Only Source of Information for Requirements Engineers? Toward a Taxonomy of Elicitation Information Sources, *ACM Trans. Management Information Systems*, Vol.7, No.3, Article 8, p.29 (Oct. 2016).
- [5] Burnay, C., Jureta, I. and Faulkner, S.: What Stakeholders Will or Will not Say: A Theoretical and Empirical Study of Topic Importance in Requirements Engineering Elicitation Interviews, *J. Information Systems*, Vol.46, pp.61–81 (Dec. 2014).
- [6] CaboCha, Yet Another Japanese Dependency Structure Analyzer, available from (<http://taku910.github.io/cabocho/>).
- [7] Chen, N. et al.: AR-Miner: Mining Informative Reviews for Developers from Mobile App Marketplace, *Proc. ICSE 2014*, pp.767–778, ACM (May–June 2014).
- [8] Dhar, V.: Data Science and Prediction, *Comm. ACM*, Vol.56, No.12, pp.64–73 (2013).
- [9] Dumas, M., La Rosa, M., Mendling, J. and Reijers, H.A. (Eds.): *Fundamentals of Business Process Management*, Springer (2013).
- [10] Evans, J.R.: *Business Analytics, 2nd ed.*: Pearson (2015).
- [11] 藤本玲子, 原 起知, 青山幹雄: データ駆動要求工学 D2RE の提案, ソフトウェア科学会 FOSE2015 論文集, pp.109–114, 近代科学社 (Nov. 2015).
- [12] 福本淳一, 安原 宏: 日本語文章の構造化解析, 情報処理学会 SIGNL, No.NL-85-11, pp.81–88 (Sep. 1991).
- [13] Carreño, L.V.G. and Winbladh, K.: Analysis of User Comments: An Approach for Software Requirements Evolution, *Proc. ICSE 2013*, pp.582–591, IEEE (May 2013).
- [14] Ghose, A. et al.: Data-Driven Requirements Modeling: Some Initial Results with i*, *Proc. APCCM 2014*, pp.55–64 (Jan. 2014).
- [15] Grossmann, W. and Rinderle-Ma, S.: *Fundamentals of Business Intelligence*, Springer (2015).
- [16] Ide, M., Amagai, Y., Aoyama, M. and Kikushima, Y.: A Lean Design Methodology for Business Models and Its Application to IoT Business Model Development, *Proc. Agile 2015*, IEEE, pp.107–111 (Aug. 2015).
- [17] 位野木万里: 要求獲得におけるステークホルダ識別手法の実適用評価, 情報処理学会デジタルプラクティス, Vol.4, No.4, pp.152–160 (2013).
- [18] JISA REBOK 企画 WG (編): 要求工学知識体系 (REBOK), 第1版, 近代科学社 (2011).
- [19] Leffingwell, D.: *Agile Software Requirements*, Addison-Wesley (2011).
- [20] Lim, S.L. and Finkelstein, A.: StakeRare: Using Social Networks and Collaborative Filtering for Large-Scale Requirements Elicitation, *IEEE Trans. Softw. Eng.*, Vol.38, No.3, pp.707–735 (2012).
- [21] Maalej, W. et al.: Toward Data-Driven Requirements Engineering, *IEEE Software*, Vol.33, No.1, pp.48–54 (2016).
- [22] 三浦信幸, 海谷治彦, 佐伯元司: 仕様作成会議の発話履歴を用いて仕様書を作成する方法, 電子情報通信学会 SIGKBSE, No.93-41, pp.9–16 (Jan. 1994).
- [23] 永山 勇: 国文法の基礎, 洛陽社 (2009).
- [24] Nalchigar, S. and Yu, E.: From Business Intelligence Insights to Actions, *Proc. PoEM 2013*, LNBIP, Vol.165, pp.114–128, Springer (Nov. 2013).
- [25] Neo4j, Neo Technology, available from (<http://neo4j.com/>).
- [26] Potts, C., Takahashi, K. and Antón, A.I.: Inquiry-Based Requirements Analysis, *IEEE Software*, Vol.11, No.2, pp.21–32 (1994).
- [27] Powell, S.G. and Batt, R.J.: *Modeling for Insight*, Wiley (2008).
- [28] Provost, F. and Fawcett, T.: *Data Science for Business*, O’Reilly (2013).
- [29] Ries, E.: *The Lean Startup*, Crown Business (2011).
- [30] Robinson, I., Webber, J. and Eifrem, E.: *Graph*

- Databases, 2nd ed.*, O'Reilly (2015).
- [31] Seyff, N. et al.: Using Popular Social Network Sites to Support Requirements Elicitation, Prioritization and Negotiation, *J. Internet Services and Applications*, Vol.6, No.7, p.16 (2015).
 - [32] Shaban, K.: A Semantic Graph Model for Text Representation and Matching in Document Mining, Ph.D. Thesis, University of Waterloo (2006).
 - [33] Sharp, H., Finkelstein, A. and Galal, G.: Stakeholder Identification in the Requirements Engineering Process, *Proc. DEXA*, pp.387–391 (1999).
 - [34] 鶴飼孝典, 林 晋平, 佐伯元司: 要求獲得におけるステークホルダの偏りと不足を検出する可視化ツール, 情報処理学会論文誌, Vol.53, No.4, pp.1448–1460 (2012).
 - [35] Wasserman, S. and Faust, K.: *Social Network Analysis*, Cambridge University Press (1994).
 - [36] Wellsandt, S., Hribernik, K.A. and Thoben, K.-D.: Qualitative Comparison of Requirements Elicitation Techniques that are Used to Collect Feedback Information about Product Use, *Proc. 24th CIRP Design Conference*, pp.212–217, Elsevier (April 2014).
 - [37] W3C: RDF, available from (<https://www.w3.org/RDF/>).
 - [38] 安田 雪: 実践ネットワーク分析, 新曜社 (2001).



藤本 玲子

2016年3月南山大学大学院理工学研究科ソフトウェア工学専攻修士課程修了。修士(ソフトウェア工学)。現在、中部電力株式会社にて設計業務に従事。在学中、要求工学の研究に従事。



青山 幹雄 (正会員)

1980年岡山大学大学院工学研究科修士課程修了。同年富士通株式会社入社。大規模ソフトウェア開発とプロジェクト管理、ソフトウェア工学の実践に従事。1986~1988年米国イリノイ大学客員研究員。1995年4月~2001年3月新潟工科大学情報電子工学科教授。2001年より南山大学数理情報学部情報通信学科教授, 2009年より情報理工学部ソフトウェア工学科教授。博士(工学)。クラウドコンピューティング, 自動車組込みソフトウェア, 機会学習ソフトウェア等を対象として, 要求工学, ソフトウェアアーキテクチャ技術の研究と教育, 人材育成に取り組む。著書『要求工学知識体系』(2011年刊:共著)ほか多数。IEEE Software, IEEE Transactions on Services Computing等の編集委員, 本会理事を歴任。1993年情報処理学会研究賞受賞。ソフトウェア科学会, 自動車技術会, IEEE, ACM, SAE各会員。