

セマンティックグラフモデルによる 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*} 青山 幹雄^{†2}

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 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関係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、有効性を示す。

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

REIKO FUJIMOTO^{†1*} MIKIO AOYAMA^{†2}

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人手に頼っていることから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、ビッグ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[3, 4]。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[5, 7, 11]。また、ソーシャルネットワークの適用[10]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言えない。

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)[19]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の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(RQ)とする。

- (1) RQ1: 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。
- (2) RQ2: (1)を用いた要求獲得プロセスを明らかにする。
- (3) RQ3: 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, 9]。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[9]。しかし、その分析は組織上での役割やインタビューなどの情報を用いた人手に頼っており、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[20]を応用したステークホルダ分析方法が提案されている[10]。しかし、ステークホルダの識別は影響度に基づくことされ、その合理的な識別方法は提案されていない。

^{†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 Program of Software Engineering, Nanzan University

* 現在、中部電力株式会社勤務。

^{†2} 南山大学 理工学部 ソフトウェア工学科
Dep. of Software Engineering, Nanzan University

3.2 BI と BA の要求工学への応用

BI (Business Intelligence)[8]や BA (Business Analytics)[4]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[17]. 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. 例えば, 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[14]. 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[7]. 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[5]. 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等のデータ分析

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[12]. 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.

4. 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, 人手に頼らざるおえない場合がある.
- (2) データから得られる要求に関する情報は完全とは言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[17]のアプローチをとる.

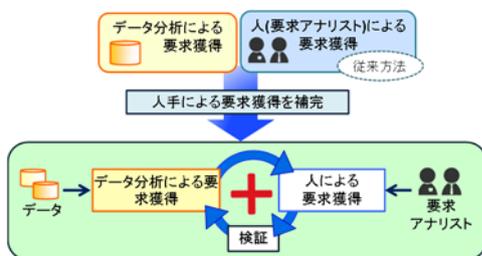


図 1 D2RE のアプローチ
Figure 1 Approach of D2RE.

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.

5. D2RE システム

5.1 D2RE システムのフレームワーク

D2RE システムのフレームワークを図 2 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.

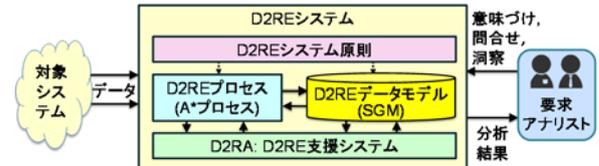


図 2 D2RE システムのフレームワーク
Figure 2 Framework of D2RE System

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 の A*プロセス

D2RE の A*(A Star)プロセスモデルを図 3 に示す. A*プロセスは, Aim (目的), Acquisition (データ収集), Analysis (分析), Assessment (評価), Action (アクション)の 5 つのアクティビティ群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.

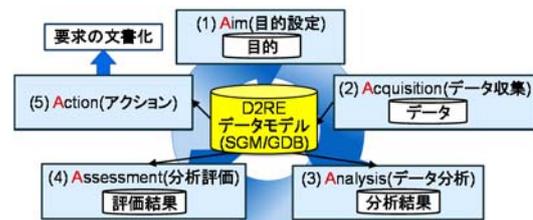


図 3 A*プロセス
Figure 3 A* Process.

- (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.
- (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する.

(3) Analysis (分析)

収集したデータを分析する。分析方法は一つとは限らず、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、収集したデータに応じて、目標を満たすまでアナリストが分析を行う。

(4) Assessment (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、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。

(5) Action (アクション)

評価結果から、次に取りうるべきアクションを決定し、行動する。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、結果を文書化する。

5.3 D2RE データモデル

5.3.1 D2RE メタモデル

D2RE が対象とするデータのメタモデルを図4に示す。一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。一つの目的に対して分析結果は一つであり、一つの分析結果に対して目的は一つである。

一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、一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える。

一つの分析結果に対してその評価は一つであるが、複数の分析結果から一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。

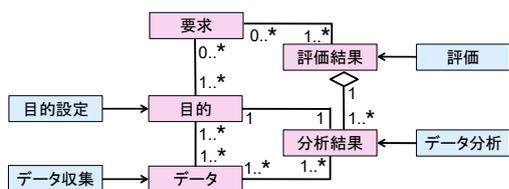


図4 D2RE のメタモデル
Figure 4 Metamodel of D2RE.

5.3.2 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、図5に示すSGM(Semantic Graph Model)[19]を導入する。SGMの実装としてneo4j[15]などのGDB(Graph Data Base)[18]がある。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

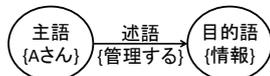


図5 トリプルグラフ
Figure 5 Triple Graph.

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) 拡張可能性

多様なデータを実行時に追加できる必要があるため、データ構造が拡張できる必要がある。SGMとその実装であるGDBは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る必要がある。

5.4 D2RA (D2RE Analyzer)

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。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

ビジネス活動から直接得られるデータを直接データと呼ぶ。一般に抽象度が低い。例えば、工場の機械から送られる生産状況のイベントやデータである。

(2) 間接データ

ビジネス活動から得られる直接データに対して何らかの形で手を加えたデータやステークホルダが対象システムに関して記述や発言したデータである。一般に抽象度が高い。例えば、顧客からの要望レポートやメールなどのメッセージ、会議の議事録などのデータである。

本稿では抽象度の高い間接データを分析対象とするが、D2REはいずれのデータにも対応できる必要がある。

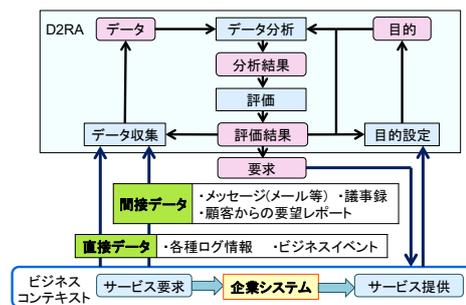


図6 D2RAの参照アーキテクチャ
Figure 6 Reference Architecture of D2RA.

6. データ駆動ステークホルダ分析

6.1 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文字のデータである。これを、以後、議事録データと呼ぶ。

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*に基づくステークホルダ分析プロセス

ステークホルダ分析をトップゴールとして、それに対して3つのサブゴールを設定し、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: 発話構造の可視化
- (2) イテレーション2: 発話者ステークホルダ分析
- (3) イテレーション3: 話題ステークホルダ分析

6.1.2 ステークホルダ分析を支援するD2RAプロトタイプ

図5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、グラフデータベース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[2]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[15]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でPythonからneo4jに接続し、解析結果をグラフデータベースへ格納する。このために、neo4jのクエリ言語であるCypherを用いている。



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
Figure 7 A Prototype for Generating Graph Data Base from Text Data

6.2 イテレーション1: 発話構造のモデル化と可視化

6.2.1 A*に基づく分析プロセス

分析対象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。日本語テキストデータから発話構造をSGMに変換し、可視化するプロセスを図8に示す。

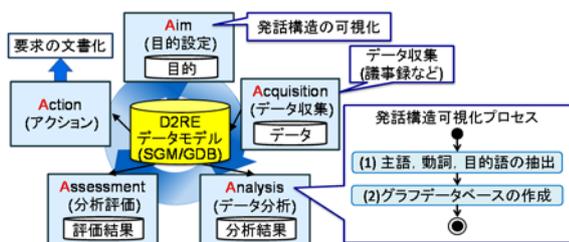


図8 発話記録の可視化プロセス
Figure 8 Visualization Process of Speech Record.

「発話構造の可視化」をデータ分析の目的とし、「主語」、「動詞」、「目的語」のトリプルで内容を表現する。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可視化する。

(1) 主語、述語、目的語の抽出

日本語係り受け解析器CaboChaを用いて、「主語」、「述語」、「目的語」のトリプルを抽出する。

(2) グラフデータベースの作成

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。例えば、「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5に示すグラフを生成できる。

6.2.2 発話構造のセマンティックグラフモデルによる可視化

議事録から抽出したトリプルを用いて生成した発話グラフを図9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。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わかった。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わかった。この分析結果から、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。

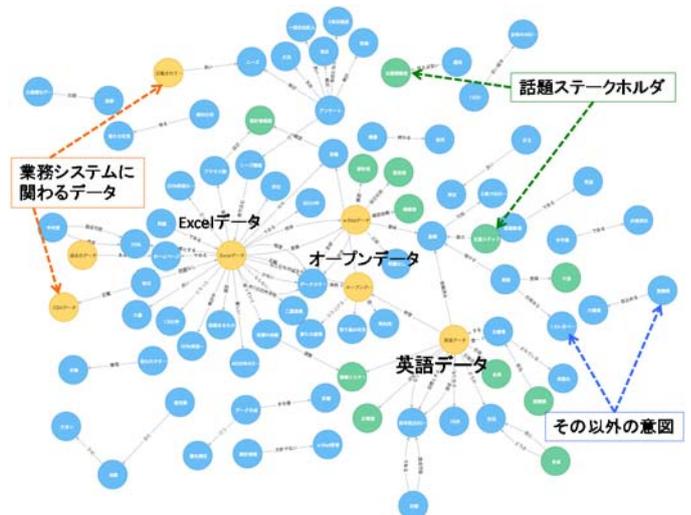


図9 発話グラフ
Figure 9 Speech Graph

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

「発話者ステークホルダの特定」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、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[6]。

このような前提に基づき、対象とする議事録を含む約200件の発話記録を文法とその用例[13]も参考にして分析した発話意図の分類を表1に示す。「報告」、「示唆」、「要望」、「問い」、「受入」、「返答」の6つに分類した。これに基づき、発話ごとに分類をメタデータとして付加し、以後の構造分析の手掛かりとした。

表1 発話意図の分類例

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。

(2) 各発話内容への重み付け

発話の種類によって、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に示す重みを付けた。各発話意図の重みは、6種の合計が100となるように配分した。例えば、「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。

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

(3) 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。

発話者が多いほどプロジェクトに大きく関与していると仮定し、関与度を式(1)で定義する。影響度は発話意図の重みと発話数を用いて式(2)で定義する。

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

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

(4) 発話者ステークホルダマトリクス生成

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。

6.3.2 発話者ステークホルダ分析への適用

図10に発話者毎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。

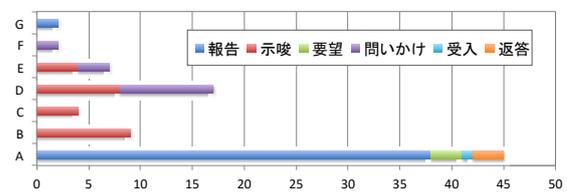


図10 発話者ごとの発話意図

Figure 10 Speech Intention per Speaker.

関与度と影響度の関係をステークホルダマトリクスに準じて図11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が、関与度が高いことからプロジェクトに大きく関与していることがわかる。

よって、Aはプロジェクトを進める上で、重要な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人の発話者の重要度は高くないため、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

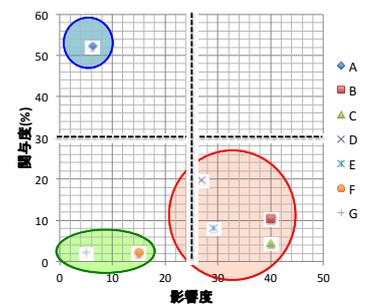


図11 関与度と影響度

Figure 11 Involvement Rate and Influence Rate

及ぼす集団であると判断できた。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言い難い。

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

「話題ステークホルダ分析」をゴールとして設定し、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

トークホルダも含まれる可能性があるので、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 12 に示す。発話者ステークホルダの分析結果を用いて、発話記録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。

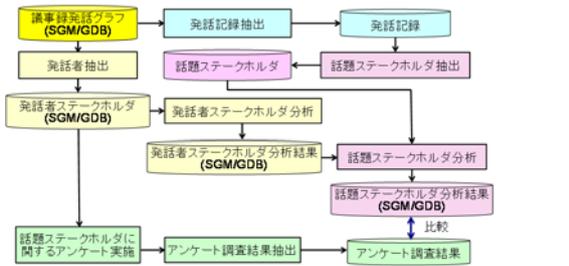


図 12 ステークホルダ分析のプロセス
Figure 12 Process of Stakeholders Analysis.

(1) グラフデータベースへの発話者データの付加

6.1 節で作成したグラフデータベースで発話者を表現するように発話者情報をグラフデータへ付加する。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。

図 13 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(Triadic Closure)[18]として付加される。

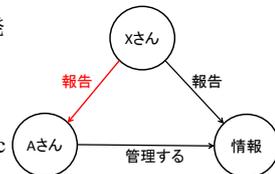


図 13 3 者閉包

(2) SGM の構造分析によるステークホルダ分析

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[20]を式(3)に示す。

$$\text{測地線距離} = d(i, j) \quad (3)$$

ステップ(1)で情報を付与した新たなグラフデータベースから、各ステークホルダと各発話者ステークホルダの測地線距離 $d(i, j)$ を評価する。

図 14 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。

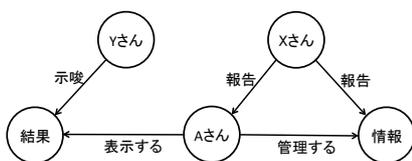


図 14 グラフの測地線距離
Figure 14 Geodesic Distance

(3) 重要度と期待度の評価

ステップ(2)で求めた測地線距離と、6.3 節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 $i(n, m)$ 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 $i(n)$ を話題ステークホルダ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。

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

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

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。

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

$$e(n) = \frac{\text{話題ステークホルダ } n \text{ の期待度要素の総和}}{\text{発話者ステークホルダ数}} = \frac{\sum_{j=1}^g e(n, j)}{g} = \frac{e(n, 1) + e(n, 2) + e(n, 3) + \dots + e(n, g)}{g} \quad (7)$$

(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。

6.4.2 話題ステークホルダ分析への適用

図 9 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 15 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 16 に示す。

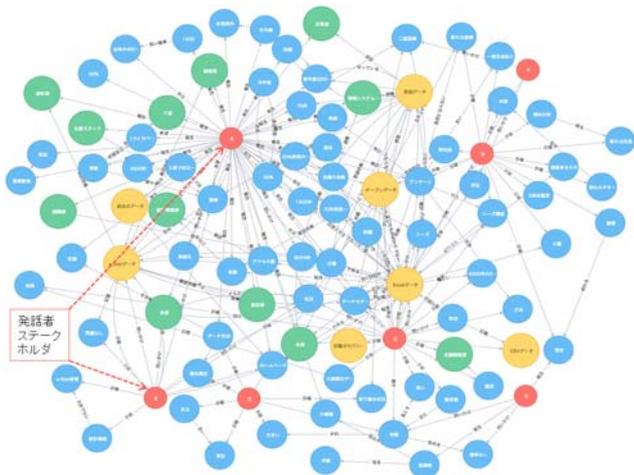


図 15 発話者付き発話グラフ
Figure 15 Speech Graph with Speaker

図 16 の右上(第 1 象限)に位置する「各局」は、重要度、期待度共に高く、現在も今後もプロジェクトに大きく関わることが推定できることから、プロジェクトを進める上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、重要度が高いため现阶段で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

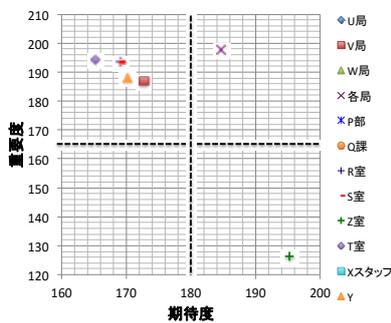


図 16 重要度と期待度
Figure 16 Importance Rate and Expectation Rate.

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ーであるが、特に先述した 3 つの局が注目されていることが指定できる。

7. 評価

実際に発話者ステークホルダーを対象に話題ステークホルダーの重要度に関するアンケート調査を行っ

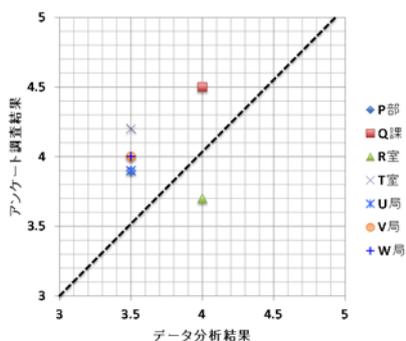


図 17 データ分析結果とアンケート調査結果の比較
Figure 17 Comparison of Survey Result with Analysis Result

た。アンケート調査は、発話者ステークホルダーとして特定した 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、影響度による重みを付け(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しての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。

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

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 17 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データ分析の意義であると共に、アンケート調査を併用する意義でもある。

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、ステークホルダー分析を行った結果、発話者ステークホルダーと話題ステークホルダーの識別を行い、ステークホルダーマトリクスを生成することで、ステークホルダー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言える。

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) RQ1: 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。さらに、ステークホルダ分析の例で示したように、3者閉包などのグラフ構造に基づく意味推論が可能となることから、従来のデータをそのまま解析する方法[11, 14]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) グラフデータベースを用いたデータ解析の活用: 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解析が可能となることから、要求獲得に求められる高度なデータ解析が可能となる。

(2) RQ2: 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、予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[16]などの繰り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、繰り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言える。

(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[17]。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、有効性を示していると言える。

しかし、日本語の発話分析だけでは、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

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。本提案方法は、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。

9. 今後の課題

今後、下記の課題を研究する必要がある。

- (1) 複雑な自然言語表現からトリプルを抽出する方法
- (2) 異なる名称などを意識別するデータ品質向上方法
- (3) 機械学習を適用したデータ解析と大量データ処理方法

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れを

利用した仮説検証を繰返す 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議事録データへ適用し、ステークホルダの特定とその構造を明らかにした。さらに、公共サービス開発の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と提案方法により特定したステークホルダとを比較し、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、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頂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 ほか, 動的利害相互作用に基づくステークホルダ分析方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, Aug. 2013, pp. 102-108.
- [2] CaboCha, Yet Another Japanese Dependency Structure Analyzer, <http://taku910.github.io/cabocho/>.
- [3] V. Dhar, Data Science and Prediction, CACM, Vol. 56, No. 12, Dec. 2013, pp. 64-73.
- [4] J. R. Evans, Business Analytics, 2nd ed., Pearson, 2015.
- [5] 藤本 玲子, ほか, 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, Nov. 2015, pp. 109-114.
- [6] 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, Sep. 1991, pp. 81-88.
- [7] A. Ghose, et al., Data-Driven Requirements Modeling: Some Initial Results with i*, Proc. of APCCM 2014, Jan. 2014, pp. 55-64.
- [8] W. Grossmann, et al., Fundamentals of Business Intelligence, Springer, 2015.
- [9] 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, Apr. 2013, pp. 152-160.
- [10] S. L. Lim, et al., StakeRare: Using Social Networks and Collaborative Filtering for Large-Scale Requirements Elicitation, IEEE Trans. on Software Engineering, Vol. 38, No. 3, May/June. 2012, pp. 707-735.
- [11] W. Maalej, et al. Toward Data-Driven Requirements Engineering, IEEE Software, Vol. 33, No. 1, Jan./Feb. 2016, pp. 48-54.
- [12] 三浦 信幸 ほか, 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, Jan. 1994, pp. 9-16.
- [13] 永山 勇, 国文法の基礎, 洛陽社, 2009.
- [14] S. Nalchigar et al., From Business Intelligence Insights to Actions, Proc. of PoEM 2013, LNBIP Vol. 165, Springer, Nov. 2013, pp. 114-128.
- [15] neo4j, Neo Technology, 2016, <http://neo4j.com/>.
- [16] C. Potts, et al., Inquiry-Based Requirements Analysis, IEEE Software, Vol. 11, No. 2, Mar. 1994, pp. 21-32.
- [17] F. Provost, and T. Fawcett, Data Science for Business, O'Reilly, 2013.
- [18] I. Robinson, et al., Graph Databases, 2nd ed., O'Reilly, 2015.
- [19] K. Shaban, A Semantic Graph Model for Text Representation and Matching in Document Mining, PhD Thesis, Univ. of Waterloo, 2006.
- [20] S. Wasserman, et al., Social Network Analysis, Cambridge Univ. Press, 1994.