

# データ駆動要求獲得方法の提案と評価

M2014SE001 藤本 玲子

指導教員 青山 幹雄

## 1. はじめに

現行の要求獲得は人手に頼っていることから、獲得した要求の合理性は明らかとはいえない。そこで、ビジネスデータ分析に基づく要求獲得方法 D2RE (Data-Driven Requirements Engineering) の概念を提案する。また、D2RE のプロセスとして、目的を明確にし、その目的に向かって、データを収集し分析を行うというアクティビティを繰り返す A\*プロセスを提案する。さらに、データに基づくステークホルダ分析の詳細なプロセスを定義した。このプロセスを、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。分析結果とアンケート調査を比較し、提案方法の妥当性を示す。

## 2. 研究の背景

要求工学の最上流工程である要求獲得の重要性が認識されている。要求定義は、成果物として構築する情報システムの仕様書となるため、その成否がシステム開発の成否を決めることになる。しかし、現行の要求獲得では、インタビューなどを利用した人の意見や主観に頼っているため、合理的な要求獲得の欠如という問題がある。

## 3. 研究の課題

前述の背景を踏まえ、以下の3点を研究課題とする。

- (1) ビジネスアナリティクス技術を応用した要求獲得方法を明らかにする。
- (2) (1)に基づく分析プロセスを明らかにする。
- (3) 提案方法の妥当性を実データで明らかにする。

## 4. 関連研究

### 4.1. 要求獲得とステークホルダ分析

要求工学において、ステークホルダ分析が研究され、広く実践されているが[1]、現行のステークホルダ分析は人手に頼っているため、合理性は明らかとはいえない。

### 4.2. データ分析の要求工学への適用

企業を感知応答モデルで捉え、ビジネスインテリジェンス (BI)を適用し、ビジネス要求を選択し、対応する方法の提案がある[5]。しかし、人手に頼る部分が大きく、複雑な組織では対処しきれない。

## 5. アプローチ

データ駆動要求工学 D2RE の枠組みと、D2RE に基づ

く分析プロセスを提案する。データから得る情報は完全とは言えず、また要求の意味づけには人の意見などが必要である。そこで D2RE は、従来の人手による要求獲得を、データ分析によって補完するアプローチとする。

## 6. D2RE の A\*プロセス

図1にD2REのA\*(A Star)プロセスモデルを示す。Aim, Acquisition, Analysis, Assessment, Actionの5つのアクティビティ群を繰り返す構造をとり、それぞれの頭文字をとってA\*と呼ぶこととする。

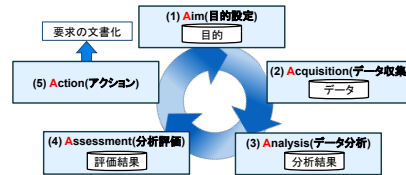


図1 A\*プロセス

## 7. データ駆動ステークホルダ分析

本稿ではテキストデータに焦点を当て、さらに、データ分析の目的をステークホルダの特定、分析を目的に A\*プロセスの詳細化を行った。ステークホルダ分析をトップゴールとし、それに対して3個のサブゴールを設定し、A\*プロセスを3回繰り返した結果を、それぞれ7.1から9.2に示す。

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、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。

### 7.1. 発話記録の可視化

#### 7.1.1. A\*に基づく分析プロセス

発話記録の可視化プロセスを図2に示す。

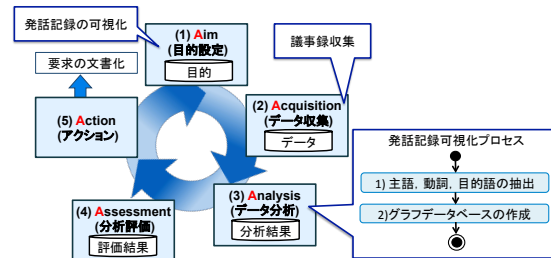


図2 発話記録の可視化プロセス

「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し、その構造をグラフで可視化する。

(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:

(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作成する。例えば、「A さんは情報を管理する」という文

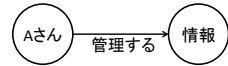


図 3 トリプルグラフ

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 3 に示すグラフを生成できる。

### 7.1.2. 発話記録可視化のプロトタイプ

データの収集から分析を支援し、グラフデータベース生成するプロトタイプの構成を図 4 に示す。

日本語係り受け解析器である CaboCha[2]モジュールの拡張を Ruby で実装した。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[4]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 に接続する。neo4j のクエリ言語として Cypher を用いている。

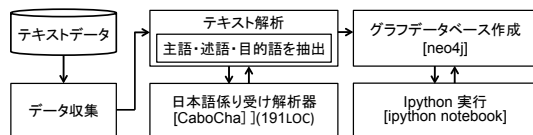


図 4 プロトタイプ構成図

### 7.1.3. 可視化への適用

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 5 に示す。黄色ノードがデータ、

緑が話題ステークホルダ、

青色がその他である。本グラフ

から、英語と Excel データをオープン

データとして記載すること、特に

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。

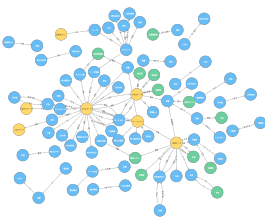


図 5 発話グラフ

## 7.2. データ駆動発話者ステークホルダ分析

### 7.2.1. 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、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。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:

発話を句点で分け、語尾に着目することで、発話の目的を特定して発話意図を分類する。対象とする議事録を含む約 200 件の発話記録の分析を行い、意図を「報告」、「示唆」、「要望」、「問い」、「受入」、「返答」の 6 つに分類した (表 1)。発話ごとに分類をメタデータとして付加した。

表 1 発話意図分類例

分類	例
報告	～である。～<名詞>。～した(過去形)。
示唆	～良い。～と考える。～と思われる。

(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、「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与える可能性があるかと判断した。

表 2 発話意図の重み

発話意図	報告	返答	受入	問い	要望	示唆
重み	5	5	15	15	20	40

(3) 影響度と関与度の算出:

各発話者に対して影響度と関与度を算出する。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作成:

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作成する。

### 7.2.2. 発話者ステークホルダ分析への適用

図 6 は発話者毎に発話内容の頻度分布である。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。

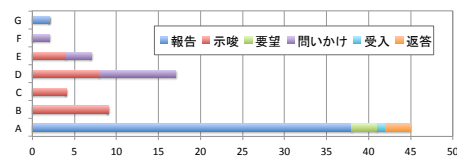


図 6 発話者ごとの発話内容

関与度と影響度の関係を図 7 に示す。左上枠内に位置する A の影響度は高くはないが、プロジェクトに大きく関与している。右下枠内に位置する発話者の重要度は高くないが、影響度が高く、意思決定の役割を果たしていることが分かる。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、影響度ともに低いことが分かる。

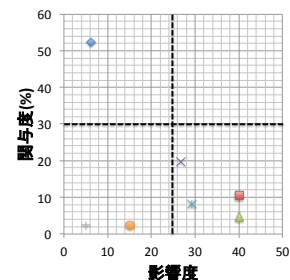


図 7 関与度と影響度

### 7.3. データ駆動話題ステークホルダ分析

#### 7.3.1. 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間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を指すこととし、発話者ステークホルダと区別する。図8に発話者ステークホルダ分析と話題ステークホルダ分析との関係を示す。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の分析を行う。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を評価する。

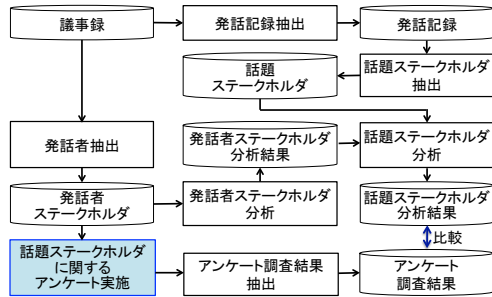


図8 ステークホルダ分析の関係

(1) グラフデータベースへの情報付与:

7.1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付与する。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与する。例えば、「Aさんは情報を管理する」とXさんが発言したとする。この文の発話意図を報告と判断し、図9のように表す。

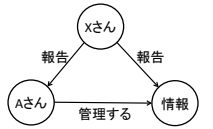


図9 トリプルグラフ 2

(2) 測地線距離の算出:

ノードiからノードjまでの最短のパスである測地線距離を式(3)に示す。

$$\text{測地線距離} = d(i, j) \quad (3)$$

(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を調べる。図10で示す例では、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。

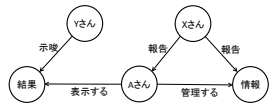


図10 トリプルグラフ 3

(3) 重要度と期待度の算出:

(2)で求める測地線距離と、7.2で求める発話者ステークホルダの重要度と影響度を用いて、話題ステークホルダの関与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、現状報告の発言が増えると考えられる。この発言の度合いである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  $n$  の重要度要素  $i(n, m)$  を式(4)で定義した。発話から得られた話題ステークホルダ  $n$  の重要度  $i(n)$  として、話題ステークホルダの重要度要素の平均として式(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。

$$i(n, m) = \frac{\text{測地線距離の総和}}{m \text{ の測地線距離}} \times m \text{ の関与度} = \frac{\sum_{j=1}^g d(n, j)}{d(n, m)} \times m \text{ の関与度} \quad (4)$$

$$i(n) = \frac{i(n, 1) + i(n, 2) + i(n, 3) + \dots + i(n, g)}{g} = \frac{\sum_{j=1}^g i(n, j)}{g} \quad (5)$$

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、重要度要素と同様に期待度要素  $e(n, m)$  を式(6)で定義し、さらに、重要度と同様に期待度  $e(n)$  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。

$$e(n, m) = \frac{\text{測地線距離の総和}}{m \text{ の測地線距離}} \times m \text{ の影響度} = \frac{\sum_{j=1}^g d(n, j)}{d(n, m)} \times m \text{ の影響度} \quad (6)$$

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

(4) 発話者ステークホルダマトリクス作成:

算出し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。

#### 7.3.2. 話題ステークホルダ分析への適用

図5に発話者を赤色ノードとして付与した結果を図11に示す。Cypherクエリを用いて、発話者ステークホルダと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12に示す。

右上の枠内に位置する「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予測できる。左上の枠内に位置する集団は现阶段でプロジェクトに大きく関わっているが、今後その

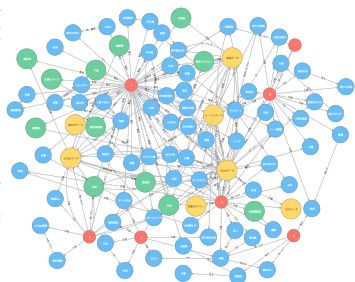


図11 発話者付き発話グラフ

役割は小さくなる可能性がある。右下に位置するZ室は、重要度は低く、プロジェクトに対して大きな役割を果たしているとは考え難いが、今後大きく関わると考えられる。また、各局にはU局やV局、W局が含まれている。本プロジェクトには各局が重要なステークホルダであるが、特に先述した3つの局が目まされていることが予測できる。

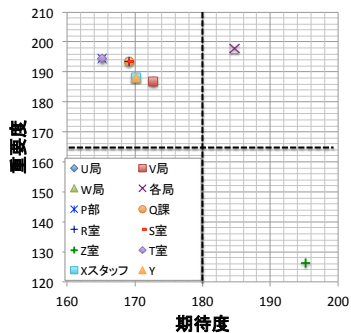


図 12 重要度と期待度

## 8. 評価

実際の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った。アンケートは、発話者ステークホルダとして特定した7人中6人の協力を得た。

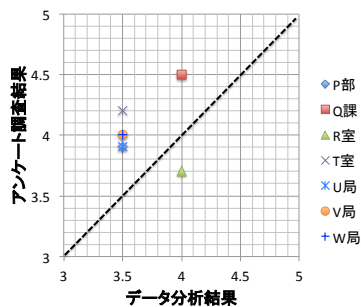


図 13 分析結果とアンケートの比較

アンケート調査結果とデータ分析結果の比較を、図13に散布図として示す。アンケート調査結果は、発話者ごとの影響度を考慮し、影響度による重みを付け、加重平均を求めた値である。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、その平均を求めた値である。

$y < x$  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つのステークホルダは、アンケート調査結果で、より重要であると判断された。

## 9. 考察

議事録データを分析し、トリプルとして表現することにより会議内容の可視化を行った。さらに、データに基づき発話者ステークホルダと話題ステークホルダの分析を行い、ステークホルダマトリクスを生成することで、役割の特定が可能となった。

### 9.1. D2REの妥当性

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のフレームワークとその具体的プロセスを提案した。これにより、本稿は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。

### 9.2. ステークホルダ分析の妥当性

従来のステークホルダ分析は人手で行っていたが、データとその分析に基づき、ステークホルダのプロジェクトへの関与を合理的に分析可能となった。しかし、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がある。また、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課題があることも明らかとなった。

### 9.3. 従来の要求工学と提案方法の比較

人手で要求獲得を行う従来の方法と比べ、人手に加えてデータ分析を行うことで、主観的な発言などに頼る要求獲得を回避し、より妥当な要求を獲得し、獲得する要求の品質向上を図ることができると考えられる。

## 10. 今後の課題

- (1) 自然言語の表現が複雑であった場合、うまくトリプルを抽出できない場合があったため、情報の抽出方法を再検討する必要である。
- (2) ステークホルダなどの名称が異なる場合も一意に識別できるよう、メタデータを付与し、一元管理するなど、データ品質の向上を検討する必要がある。
- (3) 機械学習などを適用したデータ解析を応用し、より大量のデータを効率的に分析できる方法を検討する。

## 11. まとめ

ビジネスデータ分析に基づく要求獲得方法D2REの概念と、そのプロセスとして、5つのアクティビティを繰り返すA\*プロセスを提案した。さらに、データに基づくステークホルダ分析の詳細なプロセスを定義した。このプロセスを、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ステークホルダの構造化を明らかにした。

実際のビジネスデータである議事録へ提案方法を適用し、ステークホルダの特定や発話者の意図の構造を分析することで、ステークホルダマトリクスを作成し、ステークホルダ分析を行い、提案方法の有効性、妥当性を評価した。

**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, pp. 102-108.
- [2] CaboCha, Yet Another Japanese Dependency Structure Analyzer, <http://taku910.github.io/cabocha/>.
- [3] A. Ghose, et al., Data-Driven Requirements Modeling: Some Initial Results with i\*, Proc. of APCCM 2014, Jan. 2014, pp. 55-64.
- [4] neo4j, Neo Technology, INC., 2016, <http://neo4j.com/>.
- [5] S. Nalchigar et al., From Business Intelligence Insights to Actions, Proc. of PoEM 2013, Nov. 2013, pp. 114-128.