

ステークホルダ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評価

M2017SE005 小林 勇也
指導教員 青山 幹雄

1 はじめに

近年、ステークホルダ分析の重要性が高まっている[7]. ステークホルダの発話データから主語、述語、目的語をトリプルとして抽出し、ステークホルダマトリクスを用いたステークホルダ分析方法が提案されている[4]. しかし、要求獲得方法の提案までには至っていない.

本稿では、プロパティグラフモデルで表現した発話構造に対してグラフ構造に基づいてステークホルダ分析と目的語類似性分析を行うデータ駆動要求獲得方法を提案する. 提案方法を実際の発話データに適用し、有効性を評価する.

2 研究課題

2.1 データ駆動要求獲得の範囲

研究のコンテキストを明らかにするため、本稿でのデータ駆動要求獲得の範囲を下記のように設定した.

(1) データ駆動要求獲得のデータソース

本稿では、データ駆動要求獲得のソースとして発話のテキストデータを想定している. さらに、下記に示す人々をステークホルダとして想定している.

- a) 発話機会の場に参加し、1回以上の発言をした発話者
- b) 発話の中で登場した人物

(2) データ駆動要求獲得の目的

これまで、発話データからステークホルダの発話構造を可視化し、各ステークホルダの役割を特定する研究がなされてきた[4]. しかし、どのような要求が存在するか明らかになっていない. 本稿では、ステークホルダの発話構造から要求を明らかにすることを目的とし、データ駆動要求獲得を行う.

2.2 研究課題

本稿では、以下の3点を研究課題とする.

RQ1: 発話データから生成したSIG上の目的語に着目して要求を特定することは可能か

RQ2: プロパティグラフモデルを用いて発話データから要求を獲得することは有効か

RQ3: 実際の発話データに対して、提案方法は有効か

3 関連研究

3.1 ステークホルダ分析

ステークホルダ分析方法は要求獲得のために初めて提案され、多くの研究がある[7, 9]. しかし、要求分析者の能力に依存しており、合理的な結論を得ることは困難である. また、ステークホルダと収集されたデータの増加時のスケーラビリティがないことも課題として挙げられる. さらに、グラフの構造に基づいた提案まで至っていない.

3.2 データ駆動要求獲得

データ駆動要求獲得としてグラフ上でのステークホルダ分析方法が提案されている[3]. しかし、グラフ構造を活用した分析方法は確立されていない.

3.3 プロパティグラフモデル(PGM)

PGMは、グラフモデルの1つのクラスである. グラフのノードとエッジに対して属性集合としてプロパティを付与したものである[6, 8].

3.4 グラフデータベース

グラフデータベースは、グラフモデルを操作するためのNoSQLデータベースの1つのクラスであり、プロパティグラフを保存、操作できるデータベースが提供されている[8]. OSSとしてNeo4j[6]が知られている. スキーマレスであるため、実行時にスキーマ定義が可能である.

3.5 NLP

要求の源として、テキストデータをNLP(Natural Language Processing)システムで分析可能である. 日本語データは分かち書きされておらず、単語と単語間の関係を特定することが必要である. そのため、CaboCha[2]やMcCab[5]等の多くの日本語NLPツールが開発されている.

4 アプローチ

データ駆動要求獲得のソリューションフレームワークを図1に示す.

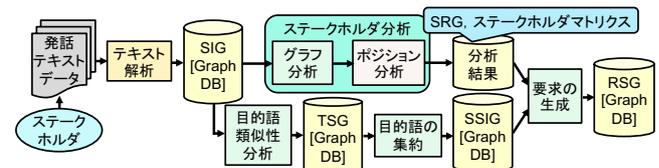


図1 データ駆動要求獲得のソリューションフレームワーク

NLP技術を用いて、発話テキストデータから、ステークホルダの意図とその関係を表現可能なPGMの意味的拡張であるSIG(Stakeholder Intention Graph)[4]を生成する. SIG上の発話意図に着目し、ステークホルダ分析と目的語間の類似性分析を行うことで、データ駆動要求方法を実現する.

(1) ステークホルダ分析

グラフデータベース上でステークホルダ間の関係を分析し、その構造を可視化する. ステークホルダ間の役割構造を表現可能なPGMの意味的拡張であるSRG(Stakeholder Role Graph)[4]とステークホルダマトリクス[1]は、グラフ構造に基づくステークホルダ分析により自動的に生成される.

(2) 目的語間の類似性分析

グラフデータベース上のステークホルダが発話した目的語間の類似関係を分析し、その構造を可視化する. 可視化した目的語間の類似構造に基づき、SIG上の目的語を集約しその構造を可視化する. TSG(Topic Structure Graph)は各目的語間の類似関係の構造を表現可能なプロパティグラフモデルの意味的拡張である. SSSIG(Small Scale SIG)は縮約したSIGの構造を表現可能なPGMの意味的拡張である.

ここで、目的語間の類似性とは、グラフ構造における、ある目的語間にis-a関係とhas-a関係が存在するとき、その目的語間は意味的に類似し、類似性があると定義する.

is-a関係とhas-a関係を式(1)、(2)で定義する.

$P \text{ is a } X \Rightarrow P \rightarrow X$	(1)
$Y \text{ has a } Q \Rightarrow Q \rightarrow Y$	(2)

(1)より, is-a 関係は継承関係であり, 目的語 P は目的語 X に集約される. (2)より, has-a 関係は包含関係であり, 目的語 Q は目的語 Y に集約される.

(3) 要求の生成

(1)と(2)から得られた分析結果を用いて要求の生成を行い, その構造を可視化する. RSG(Requirements Structure Graph)は, 要求とステークホルダとの関係を表現可能な PGM の意味的拡張である.

5 データ駆動要求獲得方法

5.1 ステークホルダグラフモデルの定義

著者らの先行研究では, SIG と SRG が定義され, ステークホルダの役割特定を行っている[9]. 本稿では, 新たに TSG, SSSIG, RSG という 3 つのステークホルダグラフモデル(SGM)を定義する. また, SIG と SRG についてもモデルの見直しを行った. 5 つの SGM 間の関係を図 2 に示す.

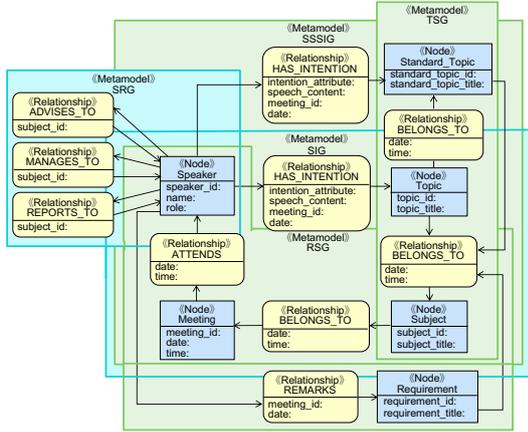


図 2 5 つのステークホルダグラフモデル間の関係

テキストデータの前処理後に SIG が生成される. Cypher クエリを用いて SIG を分析することで, SRG が生成される. SIG から定量的な分析を経てステークホルダマトリクスが生成される. ステークホルダはマトリクス上で SIG のグラフ分析を用いてクラスタリングされる. また, Cypher クエリを用いて SIG 上の目的語間の類似性に着目して分析することで, TSG が生成される. Cypher クエリを用いて TSG に基づき, SIG を縮約することで SSSIG が生成される. ステークホルダ分析と目的語類似性分析の分析結果に対して Cypher クエリを用いて分析することで, RSG が生成される.

5.1.1. SIG と SRG の拡張

SIG はステークホルダの意図とその関係を表示し, SRG はステークホルダの役割とステークホルダ間の関係を示すステークホルダの役割構造を表示する. 共にメタモデルとしてノード, エッジ, プロパティが定義されている[4]. 本稿では, 全てのノードとエッジの定義を見直し拡張する.

5.1.2. 目的語構造グラフ(TSG)

TSG は, 各目的語間の類似関係の構造を表示し, ノード, エッジ, プロパティは, それぞれ表 1, 2, 3 で定義する.

表 1 TSG のノードの定義

ノード	定義	プロパティ
Topic	会議における発話者	speaker_id, name
Standard_Topic	会議における, ある話題と継承(包含)関係にある, 粒度の粗い話題	standard_topic_id, standard_topic_title
Subject	会議における議題	subject_id, subject_title

表 2 TSG のエッジの定義

エッジ	定義	プロパティ
BELONGS_TO	話題(粒度の粗い話題)が粒度の粗い話題(議題)に属する関係	date, time

表 3 TSG のプロパティの定義

プロパティ	定義
topic_id	話題の ID
subject_id	議題の ID
standard_topic_id	粒度の粗い話題の ID
topic_title	話題のタイトル
subject_title	議題のタイトル
standard_topic_title	粒度の粗い話題のタイトル
date	日付
time	時刻

5.1.3. 縮約した SIG(SSSIG)

SSSIG は TSG に基づき, SIG のサブグラフを縮約した構造を表示し, ノード, エッジ, プロパティは, それぞれ表 4, 5, 6 で定義する.

表 4 SSSIG のノードの定義

ノード	定義	プロパティ
Speaker	会議における発話者	speaker_id, name
Standard_Topic	会議における, ある話題と継承(包含)関係にある, 粒度の粗い話題	standard_topic_id, standard_topic_title
Subject	会議における議題	subject_id, subject_title
Meeting	組織における会議	meeting_id, date, time

表 5 SSSIG のエッジの定義

エッジ	定義	プロパティ
HAS_INTENTION	発話者が持つ粒度の粗い話題への意図	intention_attribute, speech_content, meeting_id, date
BELONGS_TO	話題(議題)が議題(会議)に属する関係	date, time
ATTENDS	発話者の会議参加日時	date, time

表 6 SSSIG のプロパティの定義

プロパティ	定義
speaker_id	発話者の ID
subject_id	議題の ID
standard_topic_id	粒度の粗い話題の ID
meeting_id	会議の ID
name	発話者の名前
subject_title	議題のタイトル
standard_topic_title	粒度の粗い話題のタイトル
intention_attribute	意図属性
speech_content	発話内容
date	日付
time	時刻

5.1.4. 要求優先度グラフ(RSG)

RSG は, 各要求とステークホルダ間の関係を表示し, ノード, エッジ, プロパティを表 7, 8, 9 で定義する.

表 7 RSG のノードの定義

ノード	定義	プロパティ
Speaker	会議における発話者	speaker_id, name, role
Requirement	会議における要求	requirement_id, requirement_title
Subject	会議における議題	subject_id, subject_title
Meeting	組織における会議	meeting_id, date, time

表 8 RSG のエッジの定義

エッジ	定義	プロパティ
REMARKS	発話者が要求に対して発言する関係	meeting_id, date
BELONGS_TO	話題(議題)が議題(会議)に属する関係	date, time
ATTENDS	発話者の会議参加日時	date, time

表 9 RSG のプロパティの定義

プロパティ	定義
speaker_id	発話者の ID
requirement_id	要求の ID
subject_id	議題の ID
meeting_id	会議の ID
name	発話者の名前
requirement_title	要求のタイトル
subject_title	議題のタイトル
role	発話者の役割
date	日付
time	時刻

5.2 データ駆動要求獲得プロセス

データ駆動要求獲得方法のプロセスとプロセスで得られる成果物を図 3 に示す。

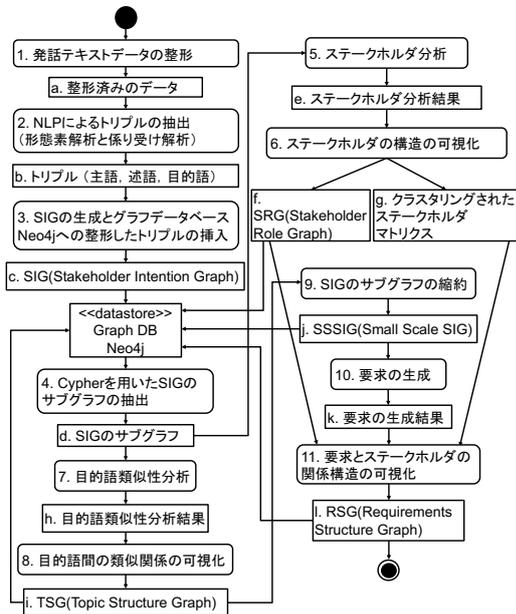


図 3 データ駆動要求獲得プロセス

分析プロセスの詳細を以下に示す。

(1) SIG の生成とグラフデータベース Neo4j への挿入

発話テキストから抽出したトリプルより SIG は生成される。ステークホルダの発話意図を明らかにするため抽出した述語に基づき、発話意図プロパティを述語に付与する。

(2) Cypher クエリを用いたサブグラフの抽出

1 つの議題に関するサブグラフを分析するため、Cypher クエリを用いて SIG のサブグラフを抽出する。

(3) ステークホルダ分析

ステークホルダ分析はグラフ構造に基づき、Neo4j における SIG, SRG に対して Cypher クエリを用いて実行され、ステークホルダマトリクスが生成される。

(4) 目的語間の類似性分析

目的語間の類似関係を分析して目的語構造を明らかにし、目的語構造に基づいて発話構造を縮約する。グラフ構造に基づき、Neo4j 上の SIG, TSG, SSSIG に対して行う。

(5) 目的語間の類似関係と縮約発話構造の可視化

類似性分析結果に基づいて Neo4j のデータを更新し、目的語間の類似関係を TSG, TSG に基づいて SIG を縮約した発話構造を SSSIG として可視化する。縮約することにより目的語の粒度を統一することが可能になる。

(6) 要求の生成

明らかになった類似関係を持つ目的語集合周辺の局所的グラフ構造に着目する。特定の発話意図を持つエッジとそのエッジがリンクする目的語に基づき、要求を生成する。

(7) 要求とステークホルダの関係構造の可視化

(3)と(6)の結果に基づき、RSG を生成する。

5.3 SGM に基づく要求獲得方法

5.3.1. 目的語間の類似性分析

(1) 分析内容と対象

発話テキストから生成される SIG のサブグラフ上の目的語に対して実行し、目的語間の類似関係を明らかにして目的語間の類似関係に基づいて SIG を縮約する。

(2) 分析方法

各目的語の内容に基づいて、目的語間の類似関係を分析し、目的語ノード間をエッジで直接連結することで TSG を生成する。生成した TSG に基づき、SIG 上の目的語を集約して縮約することで、SSSIG を生成する。

5.3.2. 要求の生成

(1) 分析内容と対象

ステークホルダ分析結果と目的語間の類似性分析によって生成した SSSIG に基づいて要求を明らかにする。

(2) 分析方法

SSSIG 上の特定の発話意図を持つエッジとそのエッジがリンクする目的語に基づいて要求を生成し、ステークホルダ分析結果と生成した要求に基づき、RSG を生成する。

6 プロトタイプ

テキスト解析、発話構造可視化、ステークホルダ分析、目的語類似性分析、要求生成というサブシステムで構成される要求獲得システムのプロトタイプを図 4 に示す。

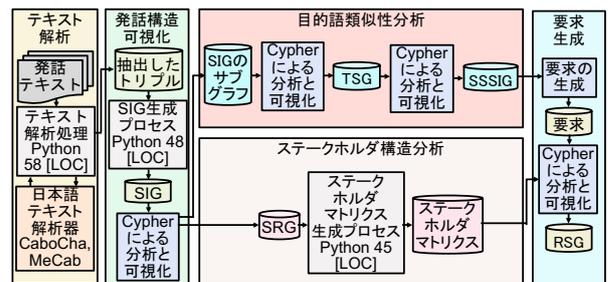


図 4 プロトタイプシステムの構成図

7 実開発への適用方法

本稿での提案プロセスを実際の会議の議事録に適用、評価し、提案プロセスの妥当性を示す。

7.1 SIGの生成と分析対象のサブグラフ出力

図2のSIGメタモデルに基づき、Cypherクエリを用いて係り受け解析から特定されたトリプルをNeo4jに挿入してSIGを生成し、議論が活発な「議題No.8 データベース登録支援」の周りのサブグラフを抽出する[4]。

7.2 目的語間の類似性分析

(1) TSGの生成

抽出したSIGのサブグラフ上の目的語に基づき、目的語間の類似関係を分析し、TSGを生成する。類似性分析の結果、目的語を7つの目的語集合に分類した。

(2) SSSIGの生成

生成したTSGに基づき、分析対象の抽出したSIGのサブグラフ上の目的語を集約して縮約することでSSSIGを生成する。生成したSSSIGを図5に示す。

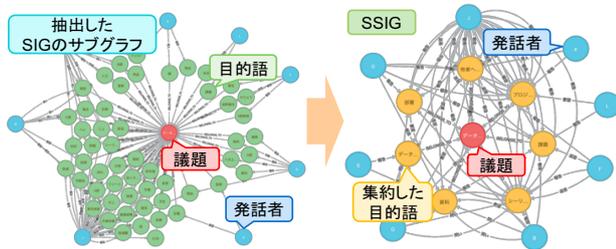


図5 抽出したSIGのサブグラフとSSSIG

7.3 発話意図分析に基づく要求の生成

(1) 要求の生成

目的語間の類似性分析によって得られたSSSIGに着目する。図5より、SSSIGにおいて、今後に対する新たな意見は含まれないと判断した「報告」と「返答」以外の発話意図を持つエッジとそのエッジがリンクするStandard_TopicノードをSSSIGのサブグラフとして抽出し、抽出したサブグラフに基づいて要求を生成する。

(2) 要求とステークホルダの関係構造の可視化

ステークホルダ分析結果[4]と生成した要求に基づき、RSGを生成する。図6にRSGを示す。

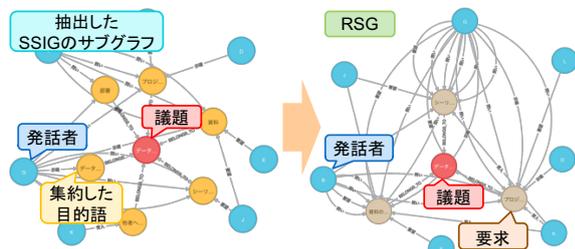


図6 抽出したSSSIGのサブグラフとRSG

8 評価

(1) RQ1: SIG上の目的語に着目した要求の特定の評価

目的語間の類似性分析により縮約したグラフ構造であるSIGにおける、集約した目的語と特定の発話意図に着目した要求特定が可能である。これにより、全てのステークホルダの発話内容と意図に基づく要求の特定が可能である。

(2) RQ2: PGMを用いた発話データからの要求獲得の評価

PGMを用いてSIG, SRG, TSG, SSSIG, RSGという5つのSGMで表現することで発話構造のモデル化が可能である。これにより、グラフモデルに基づくステークホルダ分析と目的語の類似性分析の結果から要求獲得可能である。(3)RQ3: 実データに対する提案方法の有効性評価

本提案方法を実データに対して適用することで、実際に要求を生成することが可能となった。これにより、本提案方法が有効であることが明らかになった。

9 考察

従来のステークホルダ分析方法では、ステークホルダマトリクスを用いたステークホルダ分析方法が提案され、ステークホルダの役割を特定している。しかし、要求獲得方法は提案されておらず、どのような要求があるのか明らかでない。一方で本提案方法は、グラフモデルで表現した発話構造に対してグラフ構造に基づく要求獲得が可能である。

10 今後の課題

10.1 提案方法の適用範囲の拡大

本提案方法を他の発話データや議事録データに対して適用することが必要である。

10.2 生成した要求の正しさ

グラフ構造を用いて生成した要求の正しさに関しては議論していないため、評価できない。したがって、生成した要求の正しさの評価が必要である。

10.3 要求の優先順位付け結果の表現と分析方法の拡張

RSG上の各要求の優先度は明らかになっていないため、要求の優先順位付け結果の可視化方法の検討が必要である。

11 まとめ

本稿では、発話データからNLP技術によってトリプルを抽出し、5つのSGMを生成するデータ駆動要求獲得方法を提案した。本稿により、グラフモデルに基づいてステークホルダ分析と目的語構造分析を行うことで、データに基づく要求獲得支援が可能になることが期待できる。

謝辞 議事録データをご提供頂いた伊藤藤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] M. Aoyama, et al., A Model and Architecture of REBOK (Requirements Engineering Body Of Knowledge) and Its Evaluation, Proc. of APSEC 2010, IEEE, Nov.-Dec. 2010, pp. 50-59.
- [2] CaboCha, Yet Another Japanese Dependency Structure Analyzer, <http://taku910.github.io/cabocha/>.
- [3] 藤本 玲子, 他, 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, 情報処理学会論文誌, Apr. 2018, pp. 1161-1174.
- [4] 小林 勇也, 他, セマンティックグラフモデルを用いたステークホルダ分析方法の提案と評価, SES 2017 論文集, Sep. 2017, pp. 98-105.
- [5] MeCab, Yet Another Part-of-Speech and Morphological Analyzer, <http://taku910.github.io/mecab/>.
- [6] Neo Technology, neo4j, 2016, <http://neo4j.com/>.
- [7] C. Pacheco, et al., A Systematic Literature Review of Stakeholder Identification Methods in Requirements Elicitation, J. of Systems and Software, Vol. 85, No. 9, Sep. 2012, pp. 2171-2181.
- [8] I. Robinson, et al., Graph Databases, 2nd ed., O'Reilly, 2015.
- [9] H. Sharp, et al., Stakeholder Identification in the Requirements Engineering Process, Proc. of DEXA, 1999, pp. 387-391.