

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

2013SE090 小林 勇也 2013SE192 白崎 悠太

指導教員 青山 幹雄

1. 研究背景と目的

近年、ソフトウェア要求獲得における要求仕様の高品質化が重要になっているため、要求獲得の源流であるステークホルダ分析の重要性が高まっている[3, 7]。しかし、ステークホルダ分析で必要となる情報量は膨大であり、人手による十分にデータを活用したステークホルダ分析は困難である。本稿では、要求獲得者の能力に依存しない要求獲得の支援を行うことによって要求仕様の高品質化を実現することを目的とする。

2. 研究課題

本稿では、データに基づく高品質な要求獲得を支援するステークホルダ分析方法の確立を課題とする。要求獲得で扱うデータのステークホルダ間の関係に着目し、セマンティックグラフモデルで表現して、ステークホルダの意味分析を行う方法を提案する。セマンティックグラフモデルのデータベースとして、グラフデータベースを採用する。

- (1) テキストデータに対して日本語テキスト解析を行い、グラフデータベースに格納する方法の提案
- (2) グラフ化したテキストデータから、ステークホルダ、目的語、目的語に対する述語に着目する、発話者ステークホルダ分析方法の提案

3. 関連研究

3.1. グラフデータベース

グラフデータベースとは、グラフ構造を持ったデータベースである。グラフデータベースにおける最小単位はノードであり、人やデータなど、調査する対象を表す。また、ノード間のエッジをリレーションシップといい、ノード間の関連を表し、始点から終点への向きを持つ。さらに、ノード間の関係をタイプとして定義できる[4, 5, 6]。

3.2. 日本語テキスト解析技術

3.2.1. CaboCha

CaboCha とは、高性能な日本語係り受け解析器である。形態素解析だけでは分からない、「何がどうしたのか」という、文節間の関係を明確にできる[1]。

4. アプローチ

本稿では、セマンティックグラフモデルを用いてステークホルダ間の関係を分析し、そのステークホルダの構造を可視化することによって、要求獲得を可能とする。

5. 提案方法

本稿では、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を提案する。提案プロセスを図1に示す。

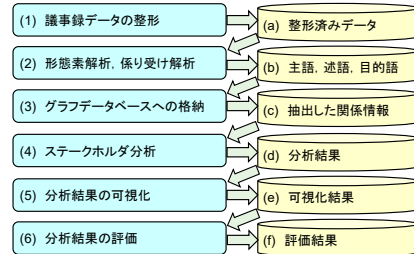


図1 提案プロセス

提案プロセスの詳細は以下の通りである。

(1) 議事録データの整形

議事録データを、形態素解析と係り受け解析を行える形式にテキストデータとして整形する。

(2) 形態素解析, 係り受け解析

McCab, CaboCha を用いて文毎に形態素解析, 係り受け解析を行い、主語, 述語, 目的語を抽出する。

(3) グラフデータベースへの格納

本稿ではグラフデータベースとして、Neo4j を用いる。データモデルを作成し、(2)で抽出したデータをグラフデータベース Neo4j に格納する。

(4) ステークホルダ分析

Cypherクエリを用い、ステークホルダのグラフ分析を行う。

(5) 分析結果の可視化

結果に基づき、Neo4j のデータを更新し、可視化する。

(6) 分析結果の評価

可視化した分析結果がどのように要求獲得支援が可能か評価する。

6. ステークホルダ分析システムのプロトタイプ

議事録データは膨大で、手作業でのテキスト解析とグラフデータベースへの格納処理は困難である。よって、図2に示すプロトタイプを実装し、一連の処理を自動化した。

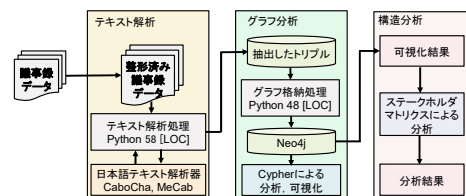


図2 システム構成図

7. 議事録データへの適用

7.1. 目的と適用対象

本稿での提案プロセスを実データに適用、評価し、提案プロセスの妥当性を示す。対象は、公共情報システム開発の約27,000字で構成される実際の議事録へ適用した。

7.2. 議事録データの適用プロセス

7.2.1. 議事録データの整形

本稿で取り扱う議事録データには、表形式のデータ、タイトル、日付など、そのままの形式では形態素解析と係り受け解析に適さないデータが含まれる。よって、データを処理に適した形式に整形した。

7.2.2. 形態素解析, 係り受け解析

McCab と CaboCha を用いて、整形済みデータを形態素解析, 係り受け解析することで、主語, 目的語, 述語を抽出する。下記にプロセスを示す。

- (1) 係り先のない文節から述語を抽出する。
- (2) (発話者)のように、発話者が括弧で囲まれた形で書かれているので、括弧内を主語として抽出する。
- (3) 主節の係り先の名詞を目的語として抽出する。目的語が発話文中に存在する場合は、前文の目的語を目的語として採用する。

7.2.3. グラフデータベースへの格納

Neo4j のクエリ言語である Cypher を用いて、データモデルに基づき、データを格納する。

- (1) 主語ノードの格納
発話者を主語ノードとして格納する。格納したそれぞれの主語ノードをステークホルダとする。ここでは、発話者に同じ名前の者は存在しないと仮定する。
- (2) 目的語ノードの格納
目的語ノードも同様に、重複を許さず格納する。
- (3) 議題ノードの格納
議題ノードもまた同様に、重複を許さず格納する。
- (4) リレーションシップの追加
存在するノード間のリレーションシップを追加する。リレーションシップを追加した結果を図3に示す。

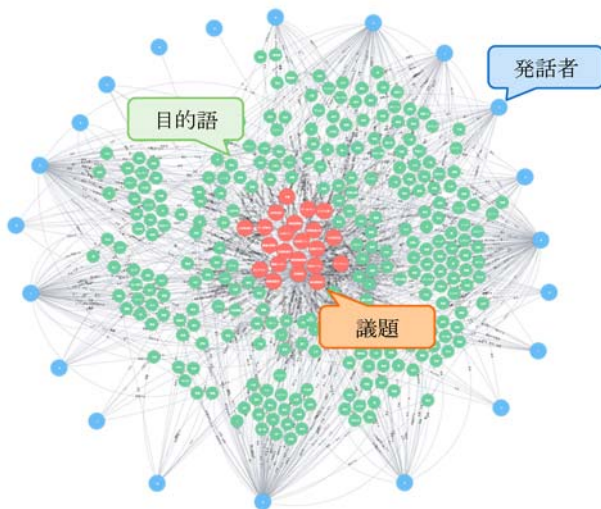


図3 リレーションシップ追加後のグラフ

7.2.4. ステークホルダ分析

格納したデータに対し、Cypher を用いてステークホルダ分析を行う。議事録は図3に示した通り、多数の議題ノードがあり、議題ごとに関わるステークホルダは異なる。本稿では、重要性が高いと推定される特定の議題に着目して、構造を明らかにする。下記にプロセスを示す。

7.2.4.1. グラフ出力

「議題 No.8 データベース登録支援」に関わる主語ノードと目的語ノードを含むグラフを出力する。

7.2.4.2. グラフ分析

出力したグラフ全体を見る大域的な分析と具体的な目的語とステークホルダの関係を見る局所的な分析の2段階分析を行う方法を提案する。

- (1) グラフの大域的構造分析

図4に議題No.8データベース登録支援のグラフを示す。図4より、主語ノードを均等に配置しているにも関わらず、目的語ノードが右に偏り、その影響で、議題ノードも中心よりも右寄りであり、グラフの重心がずれていると言える。

以上より、J が特に多く関係を持ち、発言数が多いことが分かる。

また、Jと関係を持つノードとB, D, E, F, Gも関係を持っていることが分かる。

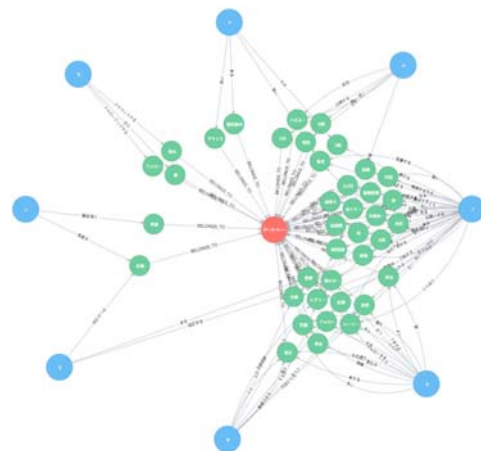


図4 議題 No.8 データベース登録支援

さらに、議題 No.8 に関する発言数分布を図5に示す。グラフ分析と同様に、Jの発言数が最も多かった。

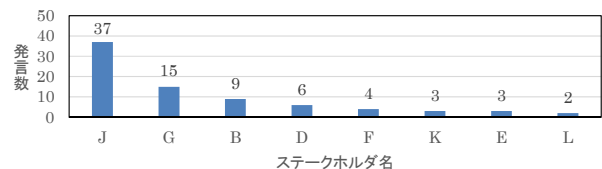


図5 ステークホルダと目的語の発言数の比較

(2) グラフの局所的構造分析

ステークホルダ間の関係を分析するため、複数のステークホルダと関係を持つ目的語を含むサブグラフを図 6 に示す。

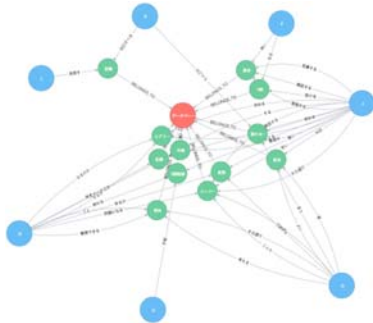


図 6 抽出したサブグラフ

抽出したサブグラフ上の単項、二項関係、三項関係に着目し、ステークホルダの役割を分析する。

a) 単項分析

i) K の役割

K は、図 4 に含まれているが、図 6 に含まれていない。したがって、他のステークホルダと共通の目的語を持たず、他のステークホルダ全体に対して一方的に報告を行なっていることが分かる。

ii) J の役割

図 4 と図 6 を比較すると、J の関係数が著しく減少している。また、J の目的語への関係を見ると、報告と質問への返答が多い。よって、J は議題に対して情報を与える立場であることが推定される。

b) 二項関係分析

i) E と L の関係

E と L の関係に注目する。E の改定すべき、という要望に対して、L が見直すかと返答しており、E が L を監督する立場であることが分かる。

ii) D と J, F と J の関係

D と F に関して、どちらの発話も J が情報を与えた目的語に対する意見であることから、議題自体には深く関係していないが、特定の話題について意見する立場であることが推定される。

c) B, G, J の三項関係分析

B, G, J に共通の目的語が存在する。したがって、その部分をサブグラフとして抽出し、三項関係を分析する。B, G, J の三項関係を図 7 に示す。

変更に対して B があるか、と質問をすると、J が無いと返答し、G が了承すると返答している。また、シーリング額も B が問題提起をし、G が J に対して質問し、J が返答している。

よって、B と G は J よりも上位の監督する立場のステークホルダである。また B と G は、議題に対する影響力が高いステークホルダであることが分かる。



図 7 B, G, J の三項関係

7.2.5. 分析結果の可視化

2 段階分析の結果を踏まえ、報告を行う関係には REPORTS_TO, 管理する関係には MANAGES, 意見や助言を行う関係には OPINION_TO の関係を追加した。可視化した発話者ステークホルダ間の関係を図 8 に示す。



図 8 ステークホルダ間の関係

7.2.6. 分析結果の評価

グラフ分析の結果を受けて、主たる発言者である J を中心としたステークホルダの関係を分析する。E, L, K の 3 つと J との関係は、会議においての一方的な報告の関係のみであるので、本稿での分析の対象とはしない。

MANAGES に属する発言は、問題提起など他のステークホルダの行動を管理する発言とし、OPINION_TO に属する発言は、話題について評価、意見を述べる発言とする。発話意図の重みを定義し、表 1 に示す。

表 1 発話意図の重み

MANAGES	OPINION_TO
0.8	0.2

関与度、影響度をそれぞれ式(1)、式(2)で定義する。

$$\text{関与度} = \frac{\text{各ステークホルダの発言数}}{\text{議題に関するステークホルダ全体の合計発言数}} \quad (1)$$

$$\text{影響度} = \frac{\sum(\text{発話意図の重み} \times \text{各発話意図の発言数})}{\text{各ステークホルダの発言数}} \quad (2)$$

ステークホルダに対する発話意図の発言数、関与度、影響度の集計結果を表 2 に示す。

表 2 発言集計結果

	MANAGES	OPINION_TO	関与度	影響度
B	6	1	0.13	0.44
D	0	3	0.08	0.2
F	0	2	0.06	0.2
G	4	5	0.21	0.29
J	0	2	0.52	0.01

ステークホルダの影響度と関与度を図 9 に示すステークホルダマトリクス上に表現し、3 つのクラスにクラスタリングを行なった。

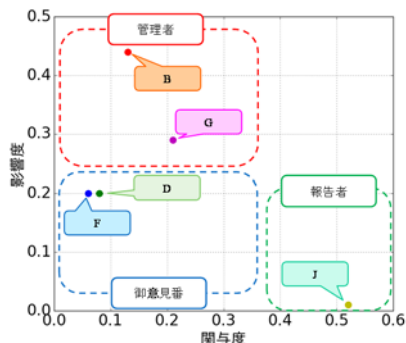


図 9 ステークホルダマトリクス

B は、関与度が他のステークホルダと比較して高いとはいえないが、影響度は著しく高い。よって、B は J に対しての意思決定を行なっていると推定される。

また、G も J を管理する役割であるが、B と比べ関与度が高く、影響度が低い。よって、話題について、包括的な意見や助言を与える役割であることが推定される。

D と F は、関与度と影響度ともに低く、ある特定の話題に関して意見をする役割であることが推定される。

下記のように、役割によってクラスを定義した。

- a) 管理者ステークホルダ
他のステークホルダを管理するステークホルダ
- b) 御意見番ステークホルダ
意見、助言を行うステークホルダ
- c) 報告者ステークホルダ
議題についての報告を行うステークホルダ

8. 議事録データへの適用評価、考察

実データに本提案方法を適用することで、グラフ分析が可能になった。共通の目的語を持つステークホルダの関係を分析することで、ステークホルダの役割を決定し、構造化が可能になることを示した。

また、グラフ分析によってステークホルダの役割を限定した上で、統計的に影響度、関与度を計算し、ステークホルダマトリクスを作成して、より詳細なステークホルダの役割を推定できることを示した。

本稿では、会議 5 回分の議事録データに対して本提案方法を適用し、妥当性を検証した。今後、会議を重ねる度にデータ量が増加しても、データモデルを定義し直すことなく、追加データの格納と分析が行える拡張性があり、分析プロセスを繰り返し実施することも可能である。

しかし、テキストデータに対して日本語テキスト解析を行い、グラフデータベースに格納する方法の提案については、形態素解析の精度の問題があり、手作業での修正が必要であるという問題が生じた。

9. 考察

従来の統計的分析方法だけでは、分析が困難なステークホルダ間の構造分析が可能だという点で有効性がある。

また、構造化したステークホルダの関係をセマンティックグラフモデルで可視化することで、要求獲得者がステークホルダの関係を視覚的に把握することで、効率的な要求獲得が可能になる。

さらに、Sharp らが提案したステークホルダ識別方法[7]で抽出されるベースラインステークホルダ、サテライトステークホルダ間の情報のやりとりをセマンティックグラフモデルで表現することができ、それらにグラフ分析が適用可能である。

10. 今後の課題

(1) セマンティックグラフモデルの応用

グラフの意味構造を利用した高度な分析を実現し、より詳細な、発展した分析を検討する。また、データの適用範囲を拡大し、本稿の妥当性を確認する必要がある。

(2) 日本語テキスト解析の改善

同一の意味を表す代名詞がノードとして存在してしまい、手直しの手間が生じた。よって、代名詞が指す名詞を一元管理し、同一名詞の割り当て処理を可能にする。

11. まとめ

本稿では、テキストデータに対して日本語テキスト解析を行い、グラフデータベースに格納する方法と、グラフデータベースに格納したテキストデータから、ステークホルダ、目的語、目的語に対する述語に着目し、発話者ステークホルダの分析を行う方法の提案をした。

実際のプロジェクト議事録に本提案方法を適用し、提案方法の有効性、妥当性を示した。

本提案を行うことで、従来のデータモデルでは表現することができないステークホルダの関係に着目した構造分析が可能である。

12. 参考文献

- [1] CaboCha, Yet Another Japanese Dependency Structure Analyzer, <http://taku910.github.io/cabocha/>.
- [2] 藤本 玲子, 他, セマンティックグラフモデルによるデータ駆動要求工学の提案とステークホルダ分析への適用評価, SES 2016 論文集, Sep. 2016, pp. 179-186
- [3] 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.
- [4] M. Lal, Neo4j Graph Data Modeling, Packt Pub., 2015.
- [5] Neo Technology, neo4j, 2016, <http://neo4j.com/>.
- [6] I. Robinson, et. al., Graph Databases, O'Reilly, 2015.
- [7] H. Sharp, et. al., Stakeholder Identification in the Requirements Engineering Process, Proc of DEXA, 1999, pp. 387-391.