

機械学習を用いたステークホルダ分析方法の提案と評価

市川 裕也[†] 青山 幹雄[†]南山大学大学院 理工学研究科ソフトウェア工学専攻[†]

1. はじめに

要求獲得においてステークホルダ分析が必要である。本稿では機械学習のクラスタリングと自然言語処理の技術を用いて、大量の文書データから構造化する方法を提案する。さらに、提案方法をステークホルダ分析に適用し、その有効性を示す。

2. 研究課題

本稿では以下の2点を研究課題として設定した。

- (1) 文書データの構造化を行い、構造化データを用いてステークホルダを特定する方法の確立
- (2) 提案方法の妥当性をプロトタイプで評価

3. 関連研究

(1) 機械学習[3]

機械学習に応用した教師データに基づく予測分析は、スパムメール判定などの文書分類に用いられている。しかし、要求獲得において機械学習を用いた研究は行われていない。

(2) ステークホルダ分析方法の提案[2]

データ分析に基づく要求獲得プロセスに基づいて文書を分析し、ステークホルダの発話内容を可視化する研究がある。ステークホルダの役割を特定するために、発話数が多いほどプロジェクトに貢献していることを示す関与度、発話意図の重みからプロジェクトに影響を与える役割を示す影響度を定義している。

4. アプローチ

本稿は要求獲得において、議事録などの文書データの内容からステークホルダの意図を機械学習で抽出を行い、可視化する。さらに、この構造を分析することで、ステークホルダを特定する。

5. ステークホルダ分析プロセスの提案

本稿のステークホルダ分析プロセスを図1に示す。



図1 ステークホルダ分析プロセス

クラスタリングを用いて発話者の意図を抽出する発話意図分類モデルを定義する。形態素解析器と係り受け解析器を用いて発話内容から発話構造を抽出する。抽出した発話意図と発話構造をグラフデータ上で発話意図付き発話構造グラフを生成し、このグラフを生成してステークホルダの特定を行う。

A Stakeholder Analysis Method Based on Machine Learning and Evaluation

[†]Yuya Ichikawa, Mikio Aoyama, Graduate School of Science and Engineering, Nanzan University.

5.1 発話意図分類モデルの定義

発話内容から発話意図を分類する発話意図分類モデルを定義する。研究[1], [2]では、発話意図は語尾の情報から「報告」、「返答」、「受入」、「問い」、「要望」、「示唆」の6種類とした。この定義に基づき形態素解析器から得られた語尾の4単語と4つの品詞を教師データ(表1)として定義する。この教師データを用いることで、発話内容から発話意図を取得する。

表1 発話意図分類モデルの教師データ(部分)

発話意図	語尾単語	品詞1	品詞2	品詞3	品詞4
報告	レクを実施する	名詞	動詞	名詞	動詞
返答	その通り			連体詞	名詞
受入	検討する			名詞	動詞
問い	実施するの	名詞	動詞	名詞	副助詞
要望	宣言して欲しい	名詞	動詞	助詞	形容詞
示唆	中心だと思	名詞	助動詞	助詞	動詞

5.2 発話内容の抽出

係り受け解析器を用いて発話構造を特定し、トリプルを抽出する。トリプルとは、「主語」、「述語」、「目的語」の構造を持つ。

発話内容から形態素解析器を用いて文節を抽出し、抽出した文節から修飾される文節の共通部分を特定し、トリプルを抽出する。

5.3 グラフ生成

5.1節の発話意図と5.2節のトリプルをグラフデータベース[4]で表現する。以下の定義に基づいてグラフデータベース上で表現する。

(1) 話題と発話者の関係の定義(図2)

大規模システム開発では発話者の役割が変わる。話題に対して発話者を特定するために話題ノードと発話者ノードを定義する。話題ノードプロパティは id と topic とする。発話者ノードのプロパティは id, name と設定する。さらにノード間の関係を Participant とする。

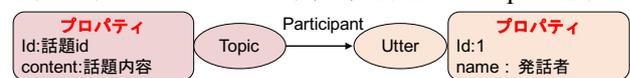


図2 話題と発話者の定義

(2) 発話者と発話内容の定義(図3)

発話者の発話内容を特定するために発話内容ノードを定義する。このノードのプロパティのキーは id, topic_id, label とする。topic_id の付与により、話題毎のステークホルダ分析が可能となる。ノード間の関係は Remark とする。

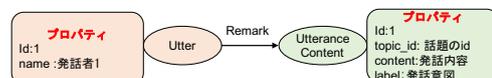


図3 発話者と発話内容の定義

(3) 発話内容と構造化の定義(図4)

発話ノードと発話構造の関係は、5.2節のトリプル抽出から、最初に抽出されたノード間を関係づける。その関係は *Intention* とし、プロパティに 5.1 節で取得した発話意図を値とする *label* を設定する。また、5.2 節のトリプルの関係は *Relation* とし、この関係のプロパティに話題の *id* を値とする *topic_id* を設定することで各話題について分析できる。

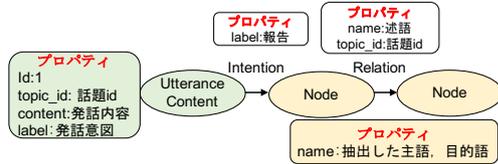


図4 発話内容と発話構造の関係

6. 例題への適用

ステークホルダ分析プロセスの評価を行うために、実際のシステム開発で使用した議事録の約4,000字へ適用した結果を図5に示す。

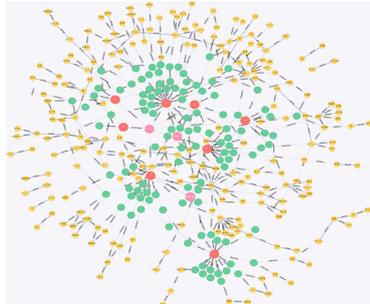


図5 発話意図付き発話構造グラフ

教師データを作成するために、分析対象外の議事録内容に対して約20,000字の発話内容にラベルを付けた。

7. 適用の評価

次の2つで評価する。

- (1) 発話構造を分析し、話題からステークホルダの意図を抽出できるか評価
- (2) 発話意図分類モデルの教師データ数から分類精度を評価

7.1 話題の構造モデルの分析

一つ話題に対して発話内容から抽出されたノードについて接続リンク数から最も最大となった「組織内 LAN」ノードの周辺の関係を図6に示す。



図6 話題ノードの周辺の関係

この結果から、「組織内 LAN」ノードには「削減計画」、「削減率」などのノードが関係していることから、コスト削減が課題であることが明らかとなった。また、議事録からもコスト削減について議論していたことを確認できたため、発話内容から発話構造を抽出したことで、各話題に対するステークホルダの意図を抽出することができた。

7.2 ステークホルダマトリクス

議事録全体の3.5%にあたる5,000字(110文)、7%にあたる10,000字(330文)、14%にあたる20,000字

(660文)を教師データとしてラベル付けを行った。組織内 LAN に関する28文の発話内容の分類結果から発話者4人の影響度と関与度[2]を算出した。

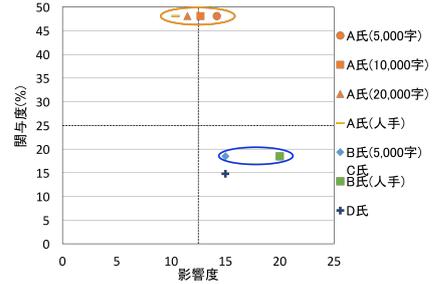


図7 影響度と関与度

この結果と人手による発話意図分類から算出した結果を図7に示す。

C氏、D氏の発話意図分類結果は発話意図分類モデルと人手による分類が一致したため、影響度と関与度の算出結果が一致した。しかし、A氏の分類モデルによる抽出では「報告」と判定されなければならないのに対し、発話意図の重み[2]が高い「示唆」と「要望」が得られたため影響度に差が生じた。

7.3 発話意図分類モデルの評価

A 氏の影響度を図8に示す。教師データを増やすことにより、人手による分類結果に収束することが明らかとなった。

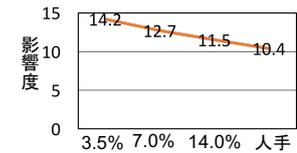


図8 教師データの効果

8. 考察

提案方法を大規模システム開発の文書データを適用し、話題に対する発話者や構造を特定できたことから効率的にステークホルダの意図や役割を推定可能となる。

10. まとめ

本稿は教師データを基にステークホルダの意図の抽出方法と係り受け解析器を用いて議事録を構造化する方法を提案した。提案方法を議事録データへ適用し、話題構造モデルを評価することにより、話題の構造を明らかにした。今後の課題として、発話意図を行う学習モデルの教師データ数の境界値を見つける必要がある。

参考文献

[1] 福本 淳一ほか, 日本語文章の構造化解析, 情報処理学会研究報告, NL-85-11, Sep. 1991, pp. 81-88.
 [2] 藤本 玲子 ほか, 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, 情報処理学会 SES2016 論文集, Sep. 2016, pp. 179-186.
 [3] W. Richert, et al., Building Machine Learning System with Python, Packt Pub., 2013.
 [4] I. Robinson, et al., Graph Databases, O'Reilly, 2013.