

深層学習(RNN/LSTM)を用いた2段階発話意図分析方法の提案と ソフトウェア開発会議への適用評価

久保井 恵里香[†] 野田 菜月[†] 鈴木 陽子[†] 青山 幹雄[†]

ソフトウェア開発会議の発話データ分析を用いてその意図を特定する方法が提案されている。しかし、個別の発話からその意図を正確に分析することは困難である。これに対して機械学習を用いて発話意図を特定する方法が提案されているが、発話毎に分析しているため、その意図が適切に分析できない場合がある。この問題を解決するために、本稿では、機械学習を用いて一連の発話データの文脈を考慮した2段階発話意図分析を提案する。LSTMを用いた再帰型ニューラルネットワーク(RNN)により1段階目で発話データのクラス分類を行い、さらに、2段階目で分類された発話データの文脈から意味を推定する方法である。提案方法を実際のソフトウェア開発者会議での発話データに適用し、文脈を考慮しない分析に比べ意図分析の精度の改善を確認した。

キーワード: ステークホルダ分析, 深層学習, RNN, LSTM, データ駆動, 発話分析

A Two-Step Speech Intention Analysis Method for Software Development Meetings Using RNN/LSTM

ERIKA KUBOI[†] NATSUKI NODA[†] YOKO SUZUKI[†] MIKIO AOYAMA[†]

1. 研究の背景と課題

1.1 研究の背景

近年、システム開発の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いる。しかし、人手による分析では、結果の合理性、公平性などを保証することが困難である。さらに、ステークホルダが多数になった場合にスケールできないという問題やコストの問題もある。このような問題に対して、開発会議などにおいてステークホルダから得られるデータに基づき分析する方法が提案されている[3]。特に、近年の機械学習技術の発展に伴い、教師あり機械学習を用いてステークホルダの発話意図を特定する研究がある[5]。しかし、これらの研究では、一般的に、日本語文法の構造に基づき文末の複数の特定語から発話意図を分類する方法がとられている。このため、発話者の意図が適切に抽出されない問題がある。

1.2 研究課題

本稿では、上記の問題を解決するために、以下の2つの研究課題(RQ: Research Question)を設定する。

- (1) RQ1: ソフトウェア開発者会議などの限定された文脈において、発話データに対して、会話の文脈を考慮した発話意図分析が機械学習を用いて可能か?
- (2) RQ2: 提案した発話意図分析方法が実際の開発者会議で有効か?

2. 関連研究

2.1 データを利用した要求獲得とステークホルダ分析

データを利用した要求獲得方法が提案されている

[3][5][8]。利用するデータにはシステムから生成されるデータと開発者会議などの人の発話データがある。人の発話データは完全性、無曖昧性、一貫性など問題を含み、従来のアルゴリズム処理ではその分析が困難であった[11]。特に、日本語の場合、分かち書きされない問題もある。これに対し、日本語の形態素解析、係り受け解析を精度良く行える自然言語処理ソフトウェアが提供され、分析が容易になった。

一方、会議での発話はある程度の規則性があることが知られている[4]。この特性に基づき、文末の4接続単語と品詞から抽出できる「報告」、「返答」、「受入」、「問い」、「要望」、「示唆」の6種類で発話意図を分類する方法が提案されている[3][5]。

2.2 機械学習の応用

機械学習を応用するには学習方法とアルゴリズムとを問題に応じて選択して組み合わせる必要がある。学習方法としては、教師あり学習、教師なし学習、強化学習などがある。アルゴリズムとしては、決定木、深層ニューラルネットワーク(NN)、SVM (Support Vector Machine)などがある[6]。

NNを用いる場合、データの分布構造に応じてCNN (Convolutional NN)、RNN (Recurrent NN)などがある。本研究で対象とする自然言語の発話や文章は構造を持つことから、RNNが適用されてきた[14][16]。RNNでは、系列データが長くなってNNが深くなるとBPTT(誤差逆伝播)法では勾配の消失や発散の問題が生じる。この問題に対応するために提案されたのがLSTM (Long Short Term Memory)法である[6][15]。

[†] 南山大学 理工学部 ソフトウェア工学科
Dep. of Software Engineering, Nanzan University

2.3 機械学習の要求工学への応用

機械学習を要求工学へ適用する方法の提案が求められている[2]。また、自然言語処理の要求工学への応用も再び活発になっている[1]。例えば、[2]に対して、SVM アルゴリズムを用いた教師あり機械学習により仕様記述文書から機能要求と非機能要求を分類した例が報告されている[7]。著者らは、NN を用いた教師あり学習による発話データからその意図を分類する方法を提案した[5]。また、話題の保存を行いながら要求のクラス分類を行う提案もある[10]。いずれの提案も、分類に留まっておき、推定にまでは至っていない。

2.4 機械学習フレームワーク

機械学習フレームワークは機械学習の様々な機能を提供するフレームワークである。Caffe, Chainer, TensorFlow などがオープンソースソフトウェアとして提供されている。

ほとんどの機械学習フレームワークは、まず計算グラフを構築し、そのグラフでデータを処理する「Define-and-Run」方式を採用している。しかし、この方式では、計算グラフがデータに依存する例が扱いにくい欠点がある。これに対して、Chainer はデータを処理(Run)しながら、計算グラフを構築(Define)する「Define-by-Run」方式を採用している点に特徴がある[13]。データ処理毎に構造を変更できるため、柔軟なネットワーク構築が可能となる。特に自然言語処理など、対象のデータの構造を活かしたネットワークを構築する機械学習に適している。

3. アプローチ

研究課題を解決するために、文末表現に基づく分類と文脈による推定の 2 段階のアプローチを提案する。これは、機械学習の分類と推定とを分け、段階的に学習することにより、機械学習の精度を段階的に高めることを可能とする、機械学習応用の新たなアプローチである。これは、ソフトウェア工学における関心事の分離とそれに基づく段階的な問題解決プロセスの考え方を機械学習へ導入したものである。

(1) 第 1 段階: 文末による分類

会議での発話記録には一般に規則性があり、日本語の文法規則にから文末の複数後により分類できることが知られている[3][4]。本稿ではこの成果に基づき、クラス分類を用いて教師データから抽出した特徴により、発話意図を発話文の文末 4 単語から 6 種類に分類する。本稿では、次の 2 つの理由から分類の対象を文末 4 単語と設定した。本稿では機械学習を適用することから、文中の分析対象範囲や単語数は学習データの拡張によって変更可能である。

- 1) 上述したように、会議などにおける制限された日本語文では文末の文法構造とそれによる意味表現のパターンがあり、文末の 4 単語で 6 つのパターンに分類可能である[3][4]。
- 2) 本稿では意味の分類が目的であり、翻訳などでの全文の正確な理解を必要としない。

(2) 第 2 段階: 文脈による推定

(1) で付与された発話意図ラベルを入力し、文脈教師データを用いた機械学習により、発話意図ラベル間の依存関係から発話意図を再分析する。

4. 2 段階発話意図分析方法

4.1 2 段階分析プロセス

2 段階発話分析のプロセスを図 1 に示す。アプローチで述べた、次の 2 段階プロセスから成る。

- (1) 文末クラス分類を用いた発話毎の意図分類
- (2) 機械学習(RNN/ LSTM)を用いた文脈における発話意図推定

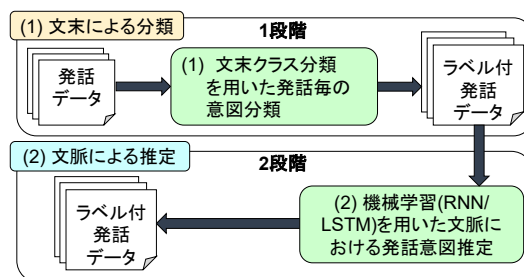


図 1 2 段階発話意図分析プロセス

4.2 文末クラス分類を用いた発話毎の意図分類

発話データの文末の 4 単語から発話意図を分類し、ラベル付けされたデータを生成する。一連のプロセスを図 2 に示す。

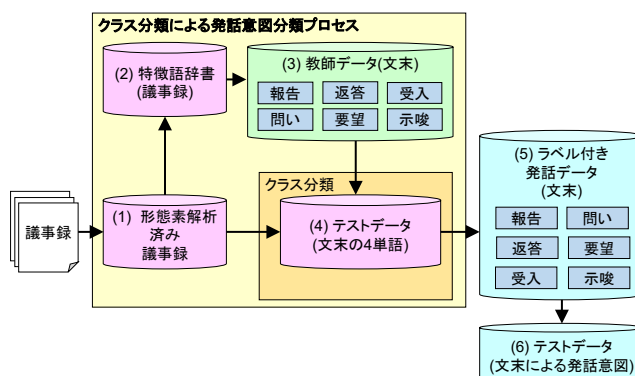


図 2 クラス分類による発話意図分類プロセス

日本語で発話文が記述された発話データを形態素解析器で単語に分割し、文末の 4 単語を取り出す。この発話データを 6 項目(「報告」、「問い合わせ」、「返答」、「要望」、「受入」、「示唆」)に振り分けて文末教師データを作成する。この文末教師データを用いて、テストデータに対してクラス分類を行う。クラス分類の方法としては、テキスト分類で広く用いられているランダムフォレスト[14]を採用する。ランダムフォレストで 6 項目にクラス分類されたテストデータを文末によるラベル付き発話データとする。

2 段階目に入る前に、1 段階目の出力で得られたラベル付き発話データを整形して、2 段階目の文脈による発話意図分析で使用するテストデータを作成する。

4.3 機械学習(RNN/LSTM)を用いた文脈における発話意図推定

2段階目の文脈による発話意図分析では、4.2節のクラス分類による発話意図分類の結果である、発話意図ラベル付き発話データに、教師あり機械学習を用いて、さらに発話意図进行分类する。

(1) 文脈クラスタの作成

まず、機械学習で文脈を学習するために作成する文脈クラスタに、発話者のデータが必要である理由を「返答」を推定する場合を例に挙げて説明する。本研究で扱う、発話文が記述されたテキストデータを分析して、発話意図を「返答」と推定するパターンは、「問い」の後の1文だけではない。例えば、「問い」と分類される発話の後で同じ人が何文にも渡って「問い」に対して答えている場合は、その文すべてが「返答」と推定されるべきだと考える。また、「問い」のすぐ後の発話文が必ずしも「返答」ではない場合もある。例えば、ある人が「問い」と分類される発話をした後に、続けて同じ人が発話をしている場合、その発話は「返答」ではない可能性がある。ある「問い」に対して、同じ人が複数の発話文に渡って答えている場合の「返答」の分類の例を図3に示す。

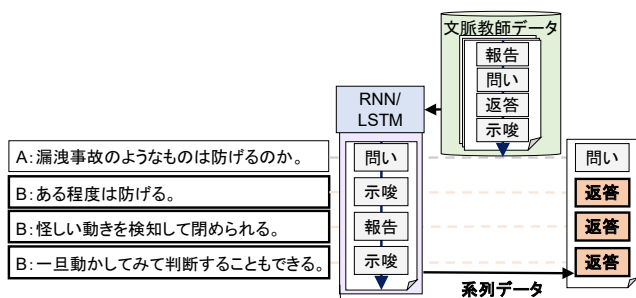


図3 「返答」が続く例

この発話文の例では、発話者 A が「問い」の発話をした後に、続く発話者 B の 3 つの発話文が、前の「問い」に対して答えている場合である。このように、「問い」の後に1人が続けて発話をしている場合は、複数の継続する発話文が「返答」と推定されるべきであると考えられる。しかし、その発話文の文末の構造分析のみでは「報告」や「示唆」と分類されることになる。この結果を用いて2段階目の文脈によるラベル付けデータに対してRNN/LSTMによる機械学習を適用してデータ間の依存関係を分析して、発話者 B の発話3文を「返答」と推定することが可能となる。

その他のパターンとして、「問い」の発話の後に同じ人が再び発話をしているときは、その発話は「返答」ではないと推定できる場合もある。このように、「返答」の発話意図を機械学習で推定するには、「問い」の後の1文だけではなく、「問い」に続けて出現する発話意図の様々なパターンを学習してそれらの発話意図を適切に分析することが必要である。

(2) 文脈教師データの作成

文脈を学習するための教師データの作成するために定義す

るデータの説明を以下に示す。

a) 発話意図ラベル①

1段階目で行う文末によるラベル付けで、発話文の文末の構造を分析して発話意図进行分类し、出力として得られた発話意図ラベル。

b) 正解発話意図ラベル

教師データを作成するときに、人手で発話意図进行分类して付与する発話意図ラベル。

c) 発話意図ラベル②

2段階目で行う文脈によるラベル付けで、RNN/LSTMによって発話の依存関係を分析して付与する発話意図ラベル。

d) 発話データ

発話者データと1段階目の文末によるラベル付けで得られた発話意図ラベル①を保有するデータ。

e) 文脈クラスタ

ある発話からある発話までの文脈をデータにするために、(4)の発話データを発話順に整列したデータである。文脈は、話題をまたいで続くことはないと仮定する。そのため、発話文が記述されたテキストデータから、文脈クラスタを作成するときの前処理として、人手で話題ごとに発話文を分割する必要があると考える。

以下、文脈教師データの作成方法を説明する。発話ごとに発話データを作成し、その発話データまでの文脈クラスタを作成する。文脈教師データは、それぞれの発話までの文脈クラスタに、その発話の正解発話意図ラベルを付与したデータである。例を図4に示す。

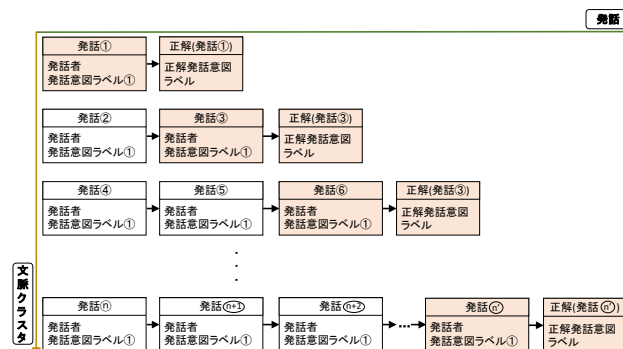


図4 文脈教師データの例

この例では、発話①から発話⑩までの文脈教師データを示す。発話ごとに発話データを作成し、発話者データと1段階目の文末によるラベル付けで得られた発話意図ラベル①を保有する。作成された発話データを、発話①から順に1発話文ごとに整列して文脈クラスタを作成する。これにより、発話ごとにそれぞれの発話が最後に整列された文脈クラスタが作成される。これによって、発話ごとに、その発話以前の発話データをトレースできるデータが得られる。この文脈クラスタに、それぞれの発話に対する正解発話意図ラベルを付与して、文脈教師データが作成される。図5に示すように、この文脈クラスタによって、各

発話文がその発話文以前に出現した発話の文構造から推定される発話意図と、発話者から受ける影響を系列データとして学習することが可能になる。

(3) 2段階目の文脈によるラベル付けの方法

1段階目の文末によるラベル付けから2段階目のプロセスを図5に示す。

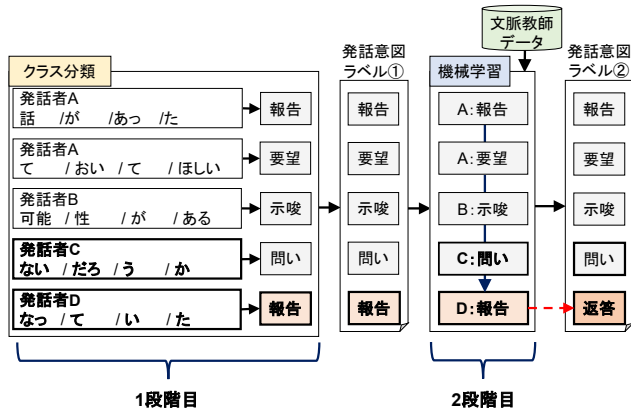


図5 発話意図再分析プロセスの例

4.2節で述べたクラス分類を用いた発話意図分類の結果である、文末を分析して得られた文末によるラベル付き発話データに機械学習を用いて発話意図を分析するプロセスを説明する。入力を4.2節のプロセスから得られた発話意図ラベル①と発話者のデータから作成した文脈クラスタを、文脈を分析して新たに得た文脈クラスタとする。教師データを、発話意図ラベル①と発話者のデータから作成した文脈クラスタ、人手で分類した正解発話意図ラベルから作成した文脈教師データとする。

(4) 文脈によるラベル付けの全体プロセス

2段階目で行う発話の文脈の分析方法を図6に示す。

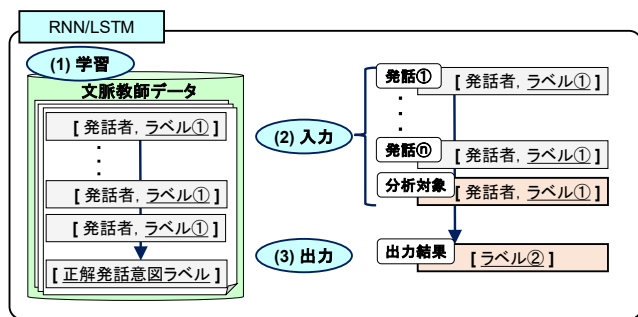


図6 2段階目の文脈による発話意図分析プロセス

2段階目の文脈の分析には、機械学習のフレームワークにChainer [12]を用い、NNにはRNNを改良したLSTMを用いて実装した。Chainerは柔軟なNNの設計ができ、可変長データが扱える。また、LSTMは系列データの解析ができるNNである。文脈を分析するには、分析対象の発話に応じた可変長の文脈クラスタを入力して、文脈クラスタごとに異なる柔軟なNNを構築して学習することが必要である。また、文脈は、前に出現した発話の後続の発話に影響を与える系列データであるので、系列データを学習して分析することができるNNを構築する必

要がある。以上の理由により、本稿では、機械学習のフレームワークとしてChainerを、NNにLSTMを用いることが適切であると判断した。

(1) LSTMによる文脈の学習

先に説明した文脈教師データを系列データとしてLSTMを用いて学習する。これにより、続いて出現する発話者と1段階目の文末によるラベル付けで文の構造を分析して得られた発話意図ラベル①のデータの依存関係から、正解発話意図ラベルの出現パターンを学習することができる。これは、次の2つのデータを教師データとしてLSTMを用いて学習することで、文脈を学習することができる。

- 1) 発話者と文構造を分析して得られた発話意図のデータの系列
- 2) 正解発話意図ラベルの依存関係を示すデータ

この結果、続けて出現する発話間の依存関係を分析することが可能になる。

(2) LSTMへの文脈の入力

学習ができたLSTM学習器に再分析する発話の文脈を入力する。入力する文脈は、文脈教師データと同様に作成した文脈クラスタであり、再分析の対象とする発話までの文脈クラスタとする。

(3) LSTMによる文脈の分析結果の出力

再分析対象の発話に対して、文脈を分析して推定される発話意図ラベル②が出力される。この発話意図ラベル②は、再分析対象の発話以前に出現した発話の、発話者と発話文の構造を分析した結果である発話意図ラベル①が再分析対象の発話に及ぼす影響を分析した結果である。

5. 2段階発話意図分析システムのプロトタイプ

2段階発話意図分析方法を実行する2段階発話意図分析システムを機械学習フレームワークChainer [12]を用いて実装した。システム構成を図7に示す。

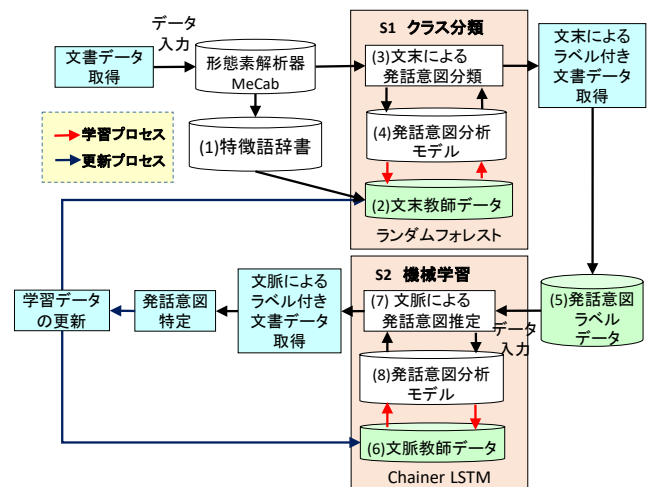


図7 2段階発話意図分析システムの構成

6. プロトタイプの実装と適用

6.1 目的

2 段階発話意図分析システムのプロトタイプを実装し、実データに適用、評価し、提案方法の妥当性を示す。実装環境を以下に示す。

プロトタイプの実装は Python を用いて、1 段階目が 204 LOC, 2 段階目が 944 LOC であった。

表 1 実装環境

実装環境	ハードウェア
OS	Ubuntu16.04 LTS
CPU	Intel®Core™i7-6700CPU@3.40GHz×8
メモリ	16GB

表 2 実装環境

実装環境	ソフトウェア
実装言語	Python 3.5.2
形態素解析	MeCab 0.996
機械学習フレームワーク	Chainer 1.6.0

6.2 適用対象

プロトタイプを適用したデータについて説明する。適用対象としたデータの概要とデータの用途を表 3 に示す。

公共情報システム開発の管理部門における第 3 回から第 10 回までの 8 回にわたる開発会議での議論が記述された発話データである。会議では発話内容がテキストデータ化されている。

第 3 回から第 6 回までの議事録を、1 段階目の文末によるラベル付けで用いる文末教師データの作成に使用し、第 7 回から第 9 回までの議事録と第 10 回の議事録の 60 文を 2 段階目の文脈によるラベル付けで用いる文脈教師データの作成に使用する。第 10 回の議事録の 111 文を、1 段階目の文末によるラベル付けから 2 段階目の文脈によるラベル付けを適用して効果を検証するためのテストデータとした。

表 3 適用データ

議事録回数	発話文数 (文)	用途	報告 (文)	示唆 (文)	要望 (文)	問い (文)	受入 (文)	返答 (文)
第3回	146	文末教師データ (597文)	210	160	39	88	6	94
第4回	140							
第5回	104							
第6回	207							
第7回	201	文脈教師データ (545文)	175	120	20	103	8	119
第8回	91							
第9回	193							
第10回	60	テストデータ (111文)	28	6	5	33	4	35
	111							

文脈教師データとテストデータについて、文脈クラスタの発話データの作成に利用した話題数と発話者数を表 4 に示す。

表 4 文脈教師データとテストデータ

議事録回	話題数 [個]	発話者数 [人]	発話文 [文]
第7回	7	9	201
第8回	6	13	91
第9回	8	12	193
第10回	3	7	60
	6	8	111

6.3 発話データへの適用プロセス

6.3.1 発話データの整形

適用対象は本稿で取り扱う議事録には、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に適さないデータが含まれているので、データを適した形式に整形する。

6.3.2 形態素解析

整形済みデータを、以下のプロセスで MeCab [9]を用いた形態素解析を利用して、発話の文末 4 単語を抽出する。

- (1) 整形済みの発話文データに形態素解析器 MeCab を用いて、形態素に分割する。
- (2) 形態素に分割された発話文データから、発話文の文末から 4 単語を取り出す。
- (3) 発話データでは「(発話者)」のように発話者が括弧内に記述されているので、括弧内から発話者データを抽出する。

6.3.3 文末教師データ作成

1 段階目の発話意図分類として提案した、文末によるラベル付けて用いる文末教師データを作成する。下記にプロセスを示す。

- (1) 会話文それぞれに対して人手で発話意図を特定し、正解発話意図ラベルを付与する。
- (2) 正解発話意図ラベルと文末 4 単語のデータを結び付け、発話意図ラベルごとに文末 4 単語のデータを分類する。

本稿では、表 3 に示したように、第 3 回から第 6 回までの 4 回分の議事録 597 文を文末教師データとして使用した。

6.3.4 クラス分類によるラベル付け

文末教師データから発話文の文末の構造の特徴を抽出して、テストデータの発話文の文末 4 単語に対して発話意図ラベル①を付与する。下記にプロセスを示す。

- (1) 作成した文末教師データから特徴語の辞書を作成する。
- (2) 作成した文末教師データの特徴を学習して、ランダムフォレストの決定木を作成する。
- (3) テストデータの発話文から文末 4 単語を抽出して、作成した辞書を用いて特徴ベクトルに変換する。
- (4) テストデータの特徴ベクトルを決定木によって分析し、テストデータの発話文に文末の構造から分類した発話意図ラベル①を付与する。

本稿では、表 3 に示したように、第 10 回目の会議の会話文から教師データとして利用しなかった 111 文をテストデータとした。

6.3.5 文脈教師データ作成

1 段階目の文末による発話意図分類で得られた発話意図ラベル①と人手で分類した正解発話意図ラベル、発話者データを用いて、文脈教師データを作成した。以下にプロセスを示す。

- (1) 文末教師データ用の発話文に対して、1 段階目の文末によるラベル付けを行い、文末の文構造から分類される発話意図ラベル①を得る。
- (2) 発話者のデータと、発話意図ラベル①、正解発話意図ラベルを用いて、4.3 で提案した文脈教師データを作成した。

本稿では、表 3 に示したように、第 7 回から第 10 回までの 4 回の会議から、文末教師データとほぼ同数の 545 文を文脈教師データとして使用した。

プロトタイプで作成した文脈教師データを図 8 に示す。

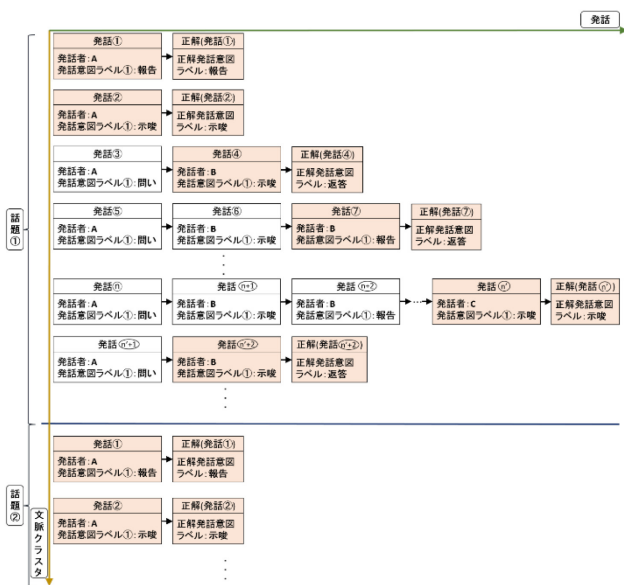


図 8 文脈教師データ

本稿のプロトタイプでは、文脈の分析が最も必要であると考えられる「返答」の分類を、2 段階目の文脈によるラベル付けで実現するため、「返答」の分類基準である「問い」から文脈を作成した。すなわち、文脈クラスターとして整理した発話データは、「問い」から次に出現する「問い」の 1 つ前の発話までである。また、話題をまたいで文脈は続いていないと仮定して、文脈クラスターは話題ごとに作成した。話題の初めに「問い」が出現しない場合は、発話文 1 文ごとに教師データを作成した。

発話者データについて、本研究のプロトタイプでは、教師データの量に対して発話者が多いため、文脈ごとに発話者を区別するようにデータを整形した。

文脈教師データには、文末の 4 単語のテキストデータは取り入れない。文末のテキストデータは、すでに 1 段階目のクラス分類で解析済みであるので、2 段階目で再び解析する必要はないと考える。

1 段階目で、LSTM で分析が難しいとされる日本語をクラス分類で分析することで、日本語の発話文の構造を発話意図ラベル①に変換できる。それにより、LSTM で日本語発話文の解析を正確に行うことができると考える。また、LSTM でテキストの解析を行わないことで、処理時間の増大を抑制することが期待される。

6.3.6 文脈によるラベル付け

文脈教師データから、会話文の発話者と文構造から分類された発話意図ラベル①の依存関係を LSTM によって学習し、テストデータの会話文に対して発話意図ラベル②を付与する。以下にプロセスを示す。

- (1) LSTM を Chainer で構築し、作成した文脈教師データを系列データとして入力して学習する。
- (2) 1 段階目の文末によるラベル付けで得られた、発話意図ラベル①が付与されたテストデータと、発話者のデータを用いて、4.3 で提案した文脈教師データの作成と同様に文脈クラスターを作成する。
- (3) テストデータの文脈クラスターを LSTM によって系列データとして分析し、テストデータの会話文に続けて出現する発話間の依存関係から、分類した発話意図ラベル②を付与する。

文脈によるラベル付けで用いるテストデータは、文末によるラベル付けで用いたテストデータと同じ発話文 111 文である。これはテストデータを同一とすることにより、2 つのテストデータの特性 (Feature) を同一とするためである。

7. 評価

7.1 評価方法の定義

本稿は先行研究[5]との比較方法として、人手でラベル付けしたデータの機械学習と提案方法の機械学習によってラベル付けされたデータの機械学習の正答率を比較する。

(1) 正確率と正答率の定義

「正確率」は機械学習の精度を表す尺度であり、人手で振り分けた場合の正解数と、機械学習の正答数との比率として式(1)で定義する。

「正答率」は機械学習が正しいと判断した回答数に対する正解数の割合と定義する。機械学習の正解数に対する、誤りを含めた機械学習の全回答数の比率として、式(2)で定義する。

「全正答率」は、各発話者の回答数に各発話の正答数を重みづけ、全回答数で割ったもので、式(3)で定義される。

$$\text{正確率} = \frac{\text{機械学習による正答データ数}}{\text{人手による正答データ数}} \quad (1)$$

$$\text{正答率} = \frac{\text{機械学習による正答データ数}}{\text{機械学習による回答数}} \quad (2)$$

$$\text{全正答率} = \frac{\text{各発話回答数} \times \text{各発話正当率}}{\text{全回答数}} \quad (3)$$

(2) 発話意図の重み付けの定義

発話内容の種類により、プロジェクトの計画や予定の変更に及ぼす影響の度合いが異なる。そのため、各発話意図の重みは6種類の合計が100となるよう定義する。

本稿では、先行研究の成果[5]に基づき、表5に示す発話意図の重みを設定した。

表5 発話意図の重み

分類	影響度合い	重み
報告	現段階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。	5
示唆	新たな視点で発話しているため、プロジェクトに対して影響を与える可能性がある。	40
要望	ある事柄に対して実現を求めており、そのためにどうすべきか、どうしてほしいかを述べており、プロジェクトに対して影響を与える可能性がある。	20
問い	事実に対するの質問や確認を行っており、今後のプロジェクトに対して影響を与えるとは考えにくい。しかし、現段階の状況の指摘の意図が含まれている場合があるため、新たな方針を決めるきっかけとなる。	15
受入	他者の意見を参考にすることにより、今後の方針を決定している。	15
返答	事実に対するの返答なため、プロジェクトに与える影響が低い。	5

(3) 出力データの採用について

1段階目の出力結果から「問い」、「要望」、「受入」は文脈に依存しない発話意図分類であるので、1段階目の結果を評価対象とする。

7.2 適用結果の評価

テストデータに2段階発話意図分析方法を適用した結果の正答率を表6に示す。

(1) 1段階目の正確率と正答率

1段階目の結果について、文末による分類で、「要望」、「問い」、「受入」は、正答率と正確率共に高いため、適切に分類されている。しかし、本稿で課題としている「返答」は両方共に低いため、適切に分類されていないことが分かる。この結果から、1段階目の文末から発話意図を分類するだけでは、適切に分類できないことが明らかになった。

表6 1段階目の正確率と正答率

出力/正解	報告	示唆	要望	問い	受入	返答	合計	正答率
報告	17	0	0	0	1	23	41	0.42
示唆	9	4	0	0	0	10	23	0.17
要望	0	0	5	0	0	0	5	1.00
問い	1	0	0	32	0	0	33	0.97
受入	0	0	0	0	3	0	3	1.00
返答	1	2	0	1	0	2	6	0.33
合計	28	6	5	33	4	35	111	-
正確率	0.61	0.67	1.00	0.97	0.75	0.06	-	-

(2) 2段階目の正確率と正答率

2段階目の結果に対し文脈による意図の推定を行った結果を表7に示す。

1段階目から2段階目で変化したカラムを色づけしてい

る。赤色のカラムは2段階目での向上を、青色のカラムは低下を示す。1段階目で適切に分類できなかった「返答」は、正確率と正当率共に向上していることが分かる。

表7 2段階目の正確率と正答率

出力/正解	報告	示唆	要望	問い	受入	返答	合計	正答率
報告	15	0	1	0	0	0	16	0.94
示唆	10	6	0	0	0	1	17	0.35
要望	0	0	4	0	0	0	5	1.00
問い	1	0	0	32	0	0	33	0.97
受入	2	0	0	0	2	1	5	0.40
返答	0	0	0	1	2	33	36	0.92
合計	28	6	5	33	4	35	111	-
正確率	0.54	1.00	0.80	0.97	0.50	0.94	-	-

(3) 1段階目と2段階目を重ね合わせた正確率と正答率

「問い」、「要望」、「受入」は1段階目の結果を、「報告」、「示唆」、「返答」は2段階目の結果を重ね合わせた結果を表8に示す。

表8 1段階目と2段階目の正確率と正答率

出力/正解	報告	示唆	要望	問い	受入	返答	合計	正答率
報告	15	0	0	0	0	0	16	1.00
示唆	10	6	0	0	0	1	17	0.35
要望	0	0	5	0	0	0	5	1.00
問い	1	0	0	32	0	0	33	0.97
受入	2	0	0	0	4	1	5	0.57
返答	0	0	0	1	0	33	36	0.97
合計	28	6	5	33	4	35	111	-
正確率	0.54	1.00	1.00	0.97	0.50	0.94	-	-

(4) 正答率の遷移図

「返答」の1段階目と2段階目の発話意図の変化の遷移を図9に示す。丸の中は1段階目と2段階目、結果のそれぞれ分類された数を示し、左側が正確率、右側が正答率を示す。

(5) 全正答率の改善

全正答率は1段階目の57%から2段階目で86%に改善した。

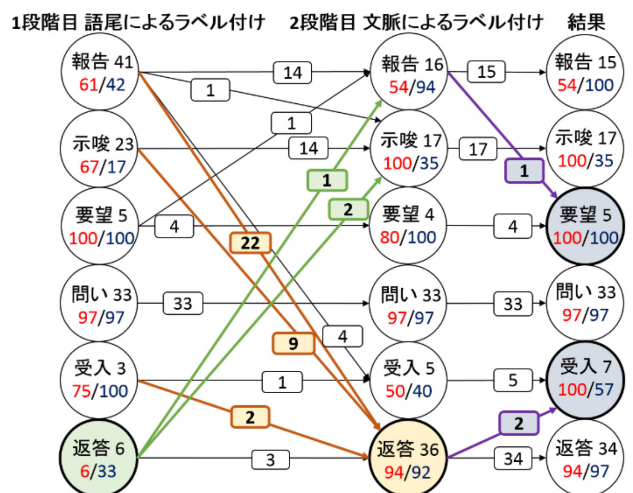


図9 正答率の遷移図

(6) 分類毎の正答率と正確率の推移の

図 10, 図 11 に 6 つの分類毎に, 正答率と正確率の 1 段階目, 2 段階目, 結果をレーダーチャートとして示す。

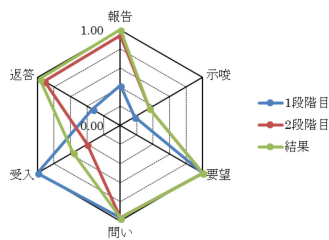


図 10 分類毎の正答率の推移

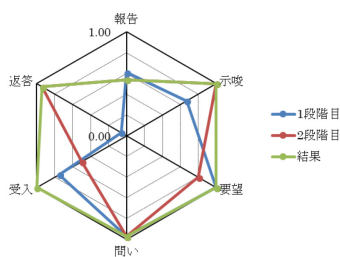


図 11 分類毎の正確率の推移

7.3 2 段階発話意図分析の評価

(1) 2 段階発話意図分析の効果示す例

1 段階目で分析が困難で, 2 段階で文脈を分析した結果, 発話意図を推定した例を表 9~表 11 に示す。

表 9 は「問い」の後に複数の「返答」が続く例である。「問い」と分類される発話の後に, 続けて出現する同じ発話者の発話 7 文を, 正しく「返答」と分類できる。4.3.1 で指摘した「問い」の後の 1 文だけではなく, 複数文にわたって「返答」が続いている場合も正しく推定できている。

表 9 「問い」の後に複数の「返答」が続く例

発話者	分析対象発話データ例	1 段階目	2 段階目	正解
A	報告があるか。	問い	問い	問い
B	打ち合わせを行った。	報告	返答	返答
B	2 点あった。	報告	返答	返答
B	話があった。	報告	返答	返答
B	回答とした。	報告	返答	返答
B	対策についてである。	報告	返答	返答
B	話があった。	報告	返答	返答
B	こととした。	報告	返答	返答

表 10 は「問い」の後の発話が返答でない例である。「問い」と分類された発話の後に出現する同じ発話者の発話を「要望」と正しく推定できている。4.3.1 で指摘した「問い」のすぐ後の発話が「返答」ではない場合も正しく推定できている。

表 10 「問い」の後の発話が「返答」でない例

発話者	分析対象発話データ例	1 段階目	2 段階目	正解
A	ということか。	問い	問い	問い
B	相談しておく。	示唆	示唆	示唆

表 11 では, 「問い」の後に出現する正解データが「受入」の発話を, 2 段階目の文脈によるラベル付けで文脈を分析して「返答」と推測している。また, 「問い」の後以外に出現する「受入」が正解であるような発話であっても, 2 段階目の文脈によるラベル付けで「受入」と正しく推定できている。

表 11 「受入」が正しく推定された例

発話者	分析対象発話データ例	1 段階目	2 段階目	正解
A	記載してほしい。	要望	要望	要望
B	了解した。	受入	受入	受入

(2) 全体としての評価

プロトタイプの実用では, 「返答」の分類の精度を上げることを目標に, 「返答」の分類基準である「問い」を基準に文脈クラスタを作成した。「返答」と同様に発話意図の特定に文脈の分析が必要と考えられる「報告」, 「示唆」に対しては, 文脈クラスタを作成して分析をしていないので, 1 段階目の文末によるラベル付けで得られたラベルを 2 段階目でも出力している場合が多く, 正答率が低い結果になった。しかし, 正解データが「返答」であり, 1 段階目で「報告」や「示唆」に分類されている発話は 2 段階目で正しく「返答」と分類されているので「報告」と「示唆」の正答率は 1 段階目に比べて 2 段階目で向上している。

2 段階目の文脈によるラベル付けの出力結果が「受入」である発話に正しく分類されていない発話が少数見られる。これは, 表 3 に示した文脈教師データの分布の中で, 他の発話意図に比較し「受入」が 8 文と少ないことから, 「受入」の文脈クラスタが十分に生成されずに発話意図ラベルが付与されたと考える。

7.4 ステークホルダマトリクスへの影響評価

先行研究の成果[3][5]に基づき, 表 5 に示した発話意図の重みを用いて発話者をステークホルダとするステークホルダマトリクス生成を第 1 段階, 第 2 段階, 結果それぞれに基づき生成した。ここで, 関与度と影響度は, 先行研究と同じく, それぞれ, 式(4), 式(5)で定義する。

$$\text{関与度} = \frac{\text{ある発話者の発話数}}{\text{話題に対する発話者全員の総発話数}} \times 100 \quad (4)$$

$$\text{影響度} = \frac{\sum (\text{発話意図の重み} \times \text{ある発話意図の発話総数})}{\text{ある発話者の発話総数}} \quad (5)$$

図 12 は 2 段階発話意図分析の 1 段階目と結果により生成されたステークホルダマトリクスを重ねあわせて示し, あわせて, ステークホルダのポジションの 1 段階目から結果への推移を示す。

ここで, 「問い」, 「要望」, 「受入」は文脈に依存しない発話意図分類となっていることから, 1 段階目の出力結果を 2 段階目以降も採用している。

図 12 から、例えば、H 氏は 1 段階目では影響度、関与度共に低いと位置づけられていたが、2 段階発話意図分析を行うことで、影響度が高いと位置づけが変わっている。さらに、E 氏は 1 段階目で影響度、関与度共に高いと位置づけられていたが、最終的に影響度は低いと位置づけられている。

この結果から、1 段階目と 2 段階目は異なったステークホルダの位置づけとなり、2 段階目がステークホルダの位置づけに影響を及ぼしていることが分かる。

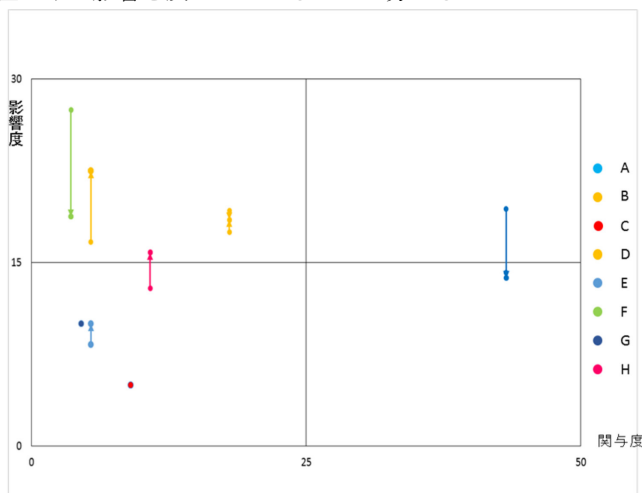


図 12 2 段階発話意図分析による
ステークホルダマトリクスの推移

8. 考察

8.1 研究課題の解決について

本稿で設定した二つの研究課題(RQ1, RQ2)に関して、適用結果に基づき考察する。

(1) Q1: 開発者会議という限定された文脈において発話データから意図分析が機械学習により可能か?

本稿の結果から、可能であると判断できる。提案した発話意図分析では、2 段階目で発話意図を推定するために、文脈クラスごとに異なった柔軟な NN を構築して学習する必要がある。また文脈は、前に出現した発話が後続の発話に影響を与える系列データであるので、系列データを学習して分析する必要がある。そのため、系列データを学習する NN である RNN として LSTM を用いることが適切であると言える。

(2) Q2: 提案した発話意図分析方法の有用性

本稿の結果から、一定水準で有効であると判断できる。提案方法では、1 段階目は文末から、2 段階目は文脈で発話意図を分析する方法をとっている。プロトタイプでは「返答」の分類精度を上げることを目標に、「返答」の分類基準である「問い」を基準に文脈クラスを作成した。1 段階目では「返答」の正答率は低く、適切に分類されていない。しかし 2 段階目で文脈から推定することで正答率が向上した。このプロトタイプの結果から 2 段階発話意図分析は有用で

あると言える。

また、2 段階発話意図分析の 1 段階回目、2 段階目、結果のそれぞれの出力から生成したステークホルダマトリクスにおけるステークホルダのポジションに有意な差が見られることから、2 段階発話意図分析はステークホルダ分析の結果にも影響を及ぼすことが明らかとなった。

8.2 先行研究[5]との比較

現時点でほぼ同一条件で比較できる先行研究である[5]と比較して、本研究の意義を考察する。7 章で述べた評価から、1 段階目の文末によるラベル付けでは、「要望」、「問い」、「受入」の結果を、2 段階目の文脈によるラベル付けでは、「報告」、「示唆」、「返答」の結果に基づき先行研究と比較して考察する。

「要望」、「問い」、「受入」は 1 段階目の文末 4 単語から発話意図分類が適切に行われている。本稿で対象とした「返答」は、1 段階目では適切な発話意図分類はされなかったが、2 段階目では正答率の向上が見られた。よって、適切な発話意図を分類するには、先行研究で適用した分類のための文末 4 単語だけでなく、提案した 2 段階分析プロセスが有効であると言える。

また、本稿で提案した、分類と推定を 2 段階で行う機械学習の要求工学への適用は、著者の知る限り報告されておらず、機械学習の新たな応用方法の提案として意義があると考えている。

9. 今後の課題

今後、以下の課題に取り組み、提案方法の精度の向上を図る。あわせて、機械学習の要求工学への応用に関する研究を推進する。

(1) 「受入」の定義の明確化

1 段階目で「受入」、2 段階目で「返答」と推定された文章があった。文章構造から「受入」とも「返答」とも判断できるので受け取り手の判断になる。そのため定義をより正確にする必要がある。

(2) 「示唆」、「報告」の分類

2 段階目で文脈から発話意図を推定するため、分析対象を「返答」とした結果、「示唆」、「報告」は正解データに対して分析が不十分であった。今後はこの二つも「返答」と同様に、文脈クラスタの基準となる発話意図ラベルを定義、作成を行い、分析する必要がある。

(3) 「返答」の定義の再検討

「要望」に対して返事をする発話は、「返答」と考える場合がある。しかし、本稿では「問い」に対して返す発話を「返答」と定義したため、受け取り手によっては違う分類になったものがある。そのため「返答」の定義を見直す必要がある。

10. まとめ

本研究では、発話データに対して、文単位での文末4単語による分類と文脈を考慮した推定の2段階で発話意図を分析する方法を提案した。発話データに本提案方法を適用し、1段目に文末4単語から発話意図を分類し、2段目に文脈から発話意図を推定した。発話者の意図が抽出されたことを検証するために、正確率と正答率を定義して評価した。この評価に基づいて推定精度の向上を示した。

本研究で提案した、分類と推定を2段階で行う機械学習の要求工学への適用は、著者の知る限り報告されておらず、機械学習の新たな応用方法の提案として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] F. Dalpiaz, et al., 1st Workshop on Natural Language Processing for Requirements Engineering (NLP4RE'18), REFSQ'18, Mar. 2018, <http://fmt.isti.cnr.it/nlp4re/>.
- [2] A. Dekhtyar and V. Fong, RE Data Challenge: Requirements Identification with Word2Vec and TensorFlow, Proc. RE 2017, IEEE, Sep. 2017, pp. 485-489.
- [3] 藤本 玲子, 青山 幹雄, 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, Vol. 59, No. 4, Apr. 2018, pp. 1161-1174.
- [4] 福本 淳一, 安原 宏, 日本語文章の構造化解析, 情報処理学会 SIGNAL, No. NL-85-11, Sep. 1991, pp. 81-88.
- [5] 市川 裕也 ほか, 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. 電子情報通信学会 KBSE2016-49, Mar. 2017. pp. 61-66.
- [6] 神嶌 敏弘(編), 深層学習, 近代科学社, 2015.
- [7] Z. Kurtanović, and W. Maalej, Automatically Classifying Functional and Non-Functional Requirements Using Supervised Machine Learning, Proc. of RE 2017, IEEE, Sep. 2017, pp. 490-495.
- [8] W. Maalej, et al. Toward Data-Driven Requirements Engineering, IEEE Software, Vol. 33, No. 1, Jan./Feb. 2016, pp. 48-54.
- [9] MeCab, <http://taku910.github.io/mecab/>.
- [10] J. Misra, et al., Topic Cohesion Preserving Requirements Clustering, Proc. of RAISE 2016, ACM, May 2016, pp. 22-28.
- [11] 三浦 信幸, 海谷 治彦, 佐伯 元司, 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.
- [12] Preferred Networks, Chainer, <https://chainer.org/>.
- [13] S. Takui, et al., Chainer: A Next-Generation Open Source Framework for Deep Learning, Proc. of Workshop on Machine Learning Systems, NIPS 2015, Dec. 2015, pp. 1-6.
- [14] F. Sebastiani, Machine Learning in Automated Text Categorization, ACM Computing Survey, Vol. 34, No. 1, Mar. 2002, pp. 1-47.
- [15] M. Sundermeyer, et al., LSTM Neural Networks for Language

- Modeling, Proc. of INTERSPEECH 2012, Sep. 2012, pp. 194-197,
https://www.isca-speech.org/archive/interpeech_2012/i12_0194.html.
[16] 坪井 祐太, ほか, 深層学習による自然言語処理, 講談社, 2017.