# 複数人対話における取りまとめ発話のタイミング予測

仙北谷 知将[1,a)]　嶋田 和孝[2,b)]

**概要**：近年，意思決定を行うツールとしての複数人対話に対する需要が高まっており，特に議論を取りまとめる高い能力が必要とされている．しかし，そのような能力を持つ人間がどこにでもいるわけではない．この問題を解決するため本研究では，人間の代わりに議論を取りまとめるデジタルファシリテータの実現を目指す．一方で，デジタルファシリテータの実現のためには，議論を取りまとめる行動のモデル化が必要である．そこで本研究では，取りまとめ発話を「司会者のような振る舞いで対話に介入する発話」と定義する．その上で，デジタルファシリテータの実現の一部として，取りまとめ発話を行うタイミングを予測する．

# Timing Prediction of Facilitation Utterance in Multi-party Dialogue

Tomonobu Sembokuya[1,a)]　Kazutaka Shimada[2,b)]

***Abstract:*** Supporting consensus-building in multi-party conversations is a very important task in intelligent systems. To conduct smooth, active, and productive discussions, we need a facilitator who controls the discussion appropriately. However, it is impractical to assign a good facilitator to each group in the discussion environment. The goal of our study is to develop a digital facilitator system that supports high-quality discussions. One role of the digital facilitator generates facilitation utterances in the discussions. To realize the system, we need to predict the timing of facilitation utterances. We generate a prediction model with verbal and non-verbal features extracted from discussions.

## 1. はじめに

会議のような複数人での対話は，企業や学校などの組織において意思決定を行うツールとしてよく活用されている．また，近年ではグローバル化やダイバーシティ推進によって，会議の質や意思決定を行うスピードの向上が求められるようになっている．一方で，発言力の違いで平等な意思決定がなされないことや，議論が平行線をたどり必要以上の時間がかかることなどの問題が発生し，質の低下や円滑な進行の妨げの原因となることがある．

このような問題を防ぐ取り組みとして会議ファシリテーションが知られている．会議ファシリテーションは，議論の状況を踏まえ，議論参加者の注意を議論の本質に向けさせる公平な介入を行うことで議論を取りまとめる行為のことである．会議ファシリテーションの多くは発話によるものと考えられる．そのため本研究では，発話による会議ファシリテーションを取りまとめ発話と呼び，これに着目する．取りまとめ発話の具体例には，議論中の適切なタイミングで，発言力の弱い参加者に意見を求めることや，議論の途中で参加者の意見を集約することなどがある．しかし，このようなことが適切に行える高い能力を持つものは，どこにでも存在するわけではない．

そこで，本研究ではデジタルファシリテータと呼ばれるシステムを構築することを目指す（**図 1**）．デジタルファシリテータとは，人間の取りまとめ役に代わり会議ファシリテーションを行うシステムである．デジタルファシリテータを実現するためには，会議ファシリテーションを人間がいつ，どのように行っているかをモデル化する必要がある．

[1] 九州工業大学 情報工学部 知能情報工学科
Department of Artificial Intellijence, Kyushu Institute of Technology 680-4, Kawazu, Iizuka, Fukuoka, 820-8502 Japan
[2] 九州工業大学 大学院情報工学研究院 知能情報工学研究系
Department of Artificial Intelligence, Kyushu Institute of Technology 680-4, Kawazu, Iizuka, Fukuoka, 820-8502 Japan
[a)] t_sembokuya@pluto.ai.kyutech.ac.jp
[b)] shimada@pluto.ai.kyutech.ac.jp

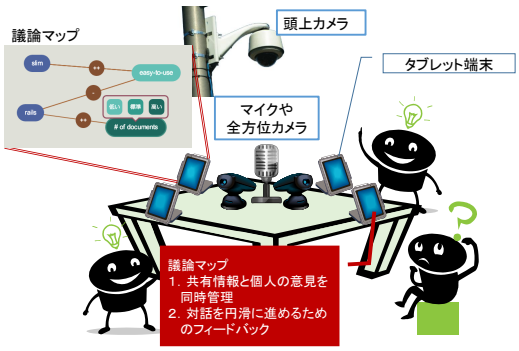


図 1　デジタルファシリテータの例
**Fig. 1**　An example of degital facilitator.

我々は，これまでに議論支援システムのプロトタイプを作成している [1]．この研究では，タブレット端末による議論マップシステムを用いた議論支援システムを提案している．議論マップとは，議論に対する考えをグラフとして表現したものである．このシステムは，議論マップのエディタと，議論に対するフィードバック機能を持つ．この研究では，提案しているシステムを用いて実際に議論実験を行い参加者にアンケートを実施し，システムを評価している．この参加者のアンケートに，「操作に気を取られて議論に集中できなかった」というものがあった．これは，参加者が議論中常にシステムによる議論介入の可能性を意識する状況にあったためであると考えられる．またシステムからフィードバックを得るには，参加者が自発的にアクセスする必要があり，必要以上の議論介入により，逆に議論進行を妨げる可能性がある．これらのことから，参加者ではなくシステムが自発的に議論介入を行うために，議論介入のタイミングは重要である．

本研究では，取りまとめ発話を「司会者のような振る舞いで対話に介入する発話」と定義する．その上で取りまとめ発話を行うタイミングに着目し，人間の会議ファシリテーションの発話行動をモデル化する．具体的には，複数人対話を対象に機械学習を用いて，ある時点で取りまとめ発話を行うかどうかという 2 値分類モデルを構築する．

## 2.　関連研究

複数人対話での取りまとめに関する研究として，Shiota ら [2] の研究がある．Shiota らは，複数人対話を対象として発話の時間情報や対話行為などの素性と決定木により，取りまとめ役を推定するモデルを構築した．さらに，生成された決定木を分析することで，取りまとめ役がもつ発話の特徴を明らかにしている．他の研究として大本ら [3] の研究がある．大本らは，ファシリテーション行動を「発散させる」，「収束させる」，「意見を具体化させる」の 3 種類に分類し行うべきファシリテーション行動の種類を推定するモデルを構築した．素性として非言語情報とパラ言語情報を使用し，素性と推定結果の相関係数を調べることで，ファシリテータの行動選択の要因を分析している．

音声対話を用いた発話タイミングの研究として Lala ら [4] の研究がある．Lala らはユーザーとシステムの 1 対 1 の対話において，相槌を適切に打つためにロジスティック回帰を用いて 500 ミリ秒後までに相槌を打つかどうかを，音声から得られたピッチや強度を用いて予測するモデルを構築している．さらに，複数種類の応答を生成する手法，適切な種類の応答を状況により選択する手法などを組み合わせ，対話システムの改善を試みている．提案した対話システムをアンドロイド Erica [*1] に導入し，実際に人間との対話を評価している．他の研究として Skantze [5] の研究がある．Skantze は人間同士の 1 対 1 の音声対話を対象に，音声情報の系列から直後の 3 秒以内に発話するかどうかをフレーム毎に予測するというモデルを，発話者毎に LSTM によって構築している．構築した 2 つのモデルによって，対話が止まった際に話者交代が生じるかどうかを予測するタスクに適用し従来手法よりも精度が向上することを示している．また，LSTM の出力をフレーム毎の確率値としてみなすことで，発話開始時の発話の長さを予測するタスクにも適用し従来手法よりも精度が向上することを示している．

以上より，会議ファシリテーションのタイミングを予測する研究はいまだ行われておらず，会議ファシリテーションを効率の良いタイミングで行うことが重要であることから，本研究では会議ファシリテーションの一部である取りまとめ発話を行うタイミングを予測するモデルを構築する．

## 3.　予測モデルの構築

本研究での予測モデルの概略図を**図 2** に示す．図 2 の左の表は対話中のある発話列を表しており，橙色の発話が取りまとめ発話である．本研究では，機械学習によって取りまとめ発話を行うかどうかを 10 秒ごとに予測するモデルを構築する．具体的には，複数人対話から $S_a$ 秒毎に素性を獲得し，予測時刻から直前の $5S_a$ 秒分の素性を入力とし，直後の $S_p$ 秒以内に取りまとめ発話を行うかどうか出力するモデルである．本節では，図 2 をもとに，まず入力として使用する素性について述べ，次に学習に用いるデータセットの作成について述べる．

### 3.1　データ作成

本研究では，機械学習によりモデルを構築するための学習データが必要である．そこで，図 2 で示すような学習データを 10 秒ごとに作成する．具体的には，対話中の $S_a$ 秒毎に獲得した直前の $5S_a$ 秒分の素性と，直後の $S_p$ 秒以

[*1] www.erica-android.jp

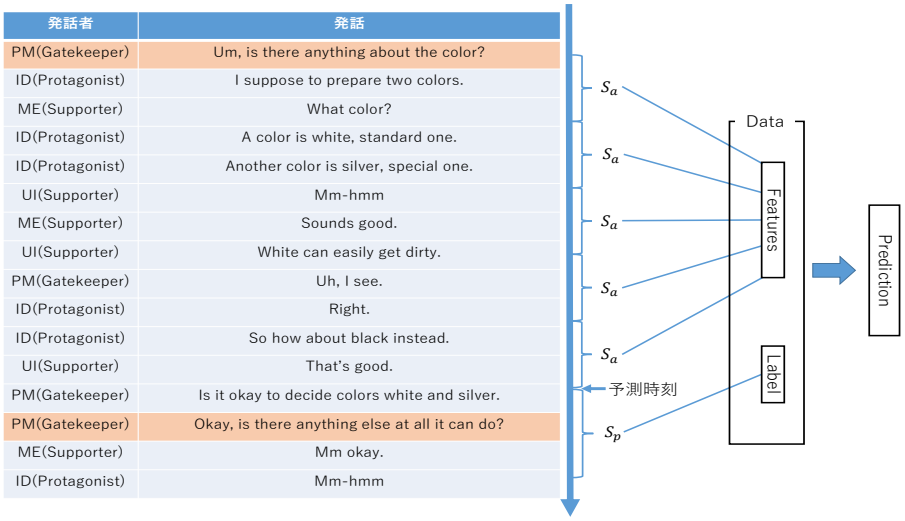


**図 2**　予測モデルの概略図
**Fig. 2**　Outline of prediction model.

内に取りまとめ発話が行われるかどうかのラベルの対である．また，そのためにはどのような発話が取りまとめ発話であるかを決定する必要がある．本節では，データ作成の対象である言語資源について述べ，次に取りまとめ発話について述べる．

### 3.1.1　AMI Meeting Corpus

AMI Meeting Corpus [6] は，英語での複数人対話を収録したコーパスである．このコーパスには，参加者にあらかじめ設定を与えて行うシナリオ会議と，設定を与えないシナリオなし会議の 2 種類の対話が存在し，それぞれ人手と機械による書き起こしデータが公開されている．

シナリオ会議では，4 人の参加者が新たなテレビのリモコンを開発するという設定で対話を行う．4 人の参加者にはそれぞれ，プロジェクトマネージャー（PM），マーケティングエキスパート（ME），ユーザーインターフェースデザイナー（UI），インダストリアルデザイナー（ID）の中のいずれかの役割が与えられる．PM は，プロジェクトを取りまとめる責任者であり，予算や納期などの管理を担当する．ME は，市場の専門家であり，ユーザーの需要や市場の動きを確認し試作機の評価を担当する．UI は，リモコンの技術的な機能とユーザーインターフェースの責任者である．ID は，リモコンの部品を含めた動作設計の責任者である．これらの役割を以降では，シナリオ役割と呼ぶ．図 2 の左列は，このシナリオ役割を表しており，それぞれの発話と対応している．

また AMI Meeting Corpus には，対話行為タグが付与されている．AMI Meeting Corpus では，15 種類の対話行為を定義しており，相槌などの発話部分に付与される「Backchannel」，何らかの情報を示す発話部分に付与される「Inform」などが存在する．

また，シナリオ会議のうちの 59 対話に対して Social role と呼ばれる役割がアノテーションされている [7]．このアノテーションは，まず対話を 1 秒以上の無声区間を基準に，スライスと呼ばれる約 30 秒の区間に区切る．このスライスに対して，4 人の参加者それぞれに 1 つの Social role がアノテーションされている．アノテータは，スライス内の参加者の言動や振る舞いを考慮しアノテーションしている．アノテーションされる Social role は次の 5 つである．

- Protagonist - 現在行われている議題に対し発言権を主張する，もしくは個人的見解を示す参加者
- Supporter - 協調的態度をとり，技術的および関連する支援を提供する参加者
- Neutral - 他の参加者の発言を受動的に受け入れる参加者
- Gatekeeper - 司会者のように振る舞い，他の参加者に対しコミュニケーションを仲介し促進する参加者
- Attacker - 他の参加者の意見に対し反対を表明する，もしくは他の参加者を攻撃する参加者

図 2 の左列の括弧内が，該当スライスにおける Social role を表している．図 2 のように，異なる参加者に同じ Social role が付与される場合もある．

本研究では，AMI Meeting Corpus のうち Social role がアノテーションされているシナリオ会議 59 対話の人手による書き起こしデータを対象とする．また，英語の発話単位として「.」や「?」などの文の終わりを表すパンクチュエーションを基準に区切った書き起こしデータを用いる．

### 3.1.2　取りまとめ発話

本研究では，発話自体，対話行為タグ，Social role の情報を用いて，次の条件を満たす発話を取りまとめ発話とする．取りまとめ発話の判定フローを**図 3** に示す．まず，(1) 司会者のような振る舞いを表す Social role である Gatekeeper の発話であるかを判定する．ここで，シナリ

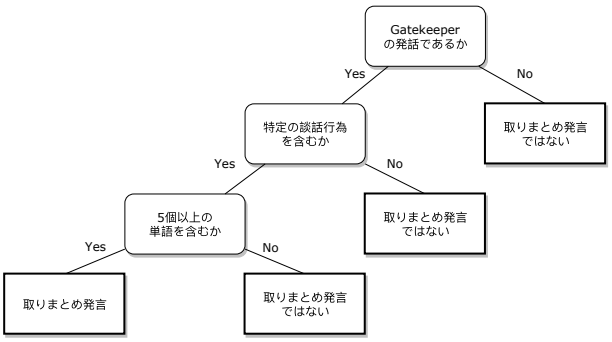


**図 3** 取りまとめ発話の判定フロー
**Fig. 3** Judgement flow of facilitation utterance.

**表 1** AMI Meeting Corpus 上の対話行為
**Table 1** Dialogue acts in AMI Meeting Corpus.

| 対話行為タグ | 内容 |
|---|---|
| Backchannel | 相槌 |
| Stall | 言いよどみ |
| Fragment | 中身のない発話 |
| Inform | 情報の提供 |
| Suggest | チーム，企業での行動に対する意思表明 |
| Offer | 自分の行動に対する意思表明 |
| Assess | 発話や情報に対する評価，感想 |
| Elicit-* | 他の参加者に対する対話行為の要求 |

オ役割の PM も Social role の Gatekeeper も同じ議論を取りまとめる役割であるが，PM の発話ではなく Gatekeeper の発話を抽出する理由は，シナリオ役割の中で取りまとめ発話を行う可能性は PM 以外にもあるためである．以上の理由で，PM 以外が発する取りまとめ発話を見逃さないために Gatekeeper の発話を条件とする．

次に，(2) 特定の対話行為タグが付与されている発話であるかを判定する．図 2 で Gatekeeper が付与された参加者が「Uh, I see.」という発話を行っている．この例に示すように，アノテータはスライス内の参加者の振る舞いを総合的に判断し Social role をアノテーションするため，発話内容と Social role の内容が直接的に関連付かない場合がある．そのため，Social role に加えて対話行為タグによって発話内容に制約を加える必要がある．AMI Meeting Corpus で定義されている対話行為の一例を表 1 に示す．この中で，Inform，Suggest，Offer，Elicit-*の対話行為タグが付与されている発話であるかを判定する．このことにより，Gatekeeper の発話のうち相槌などの議論をまとめる意図を持っていないであろう発話を除く．

最後に，(3) 5 語以上の単語を含む発話であるかを判定する．(1)，(2) では「No.」(対話行為は Inform) のような発話が抽出されるが，これは議論に介入する意図を持つ発話でないと考える．また，5 語以上という設定は (2) の発話の単語数の最頻値が 5 語であったことと，その発話の内容から 5 語以上に設定した．このことにより，より議論を介入する意図を持つ発話に限定することができる．

以上の条件をすべて満たした発話を取りまとめ発話とし，現在時刻から直後の $S_p$ 秒以内に取りまとめ発話が開始されるかというデータを作成し，モデルの構築に使用した．

### 3.2 素性

本研究では，図 2 に示すように素性を獲得し予測に用いる．具体的には，$S_a$ 秒毎に発話などの情報をもとに素性を獲得し，直前の $5S_a$ 秒分の素性を使用する．取りまとめ発話は，議論が停滞しているときや，白熱しすぎたときに行われることが考えられるため，議論の停滞度合いや白熱度合いを表すような素性として次の 5 つを用いた．

I. 単語の分散表現の平均

単語の分散表現 [8] とは，単語を類似する意味の語がベクトル空間の中で近い位置に存在するように固定長の密ベクトルとして表現する方法である．議論が停滞しているときは「uh」や「um」のようなフィラーが増加することが考えられる．以上の理由で，出現する単語の表層的な特徴を捉えるため，$S_a$ 秒の範囲に出現する単語の分散表現の平均を算出し素性として使用する．

II. 時間当たりの単語数

議論の停滞度合いや白熱度合いが極端に高いとき，参加者の発話数が顕著に増減することが考えられる．以上の理由で，議論中の発話数の特徴を捉えるため，$S_a$ に出現する単語の総数を $S_a$ で割ったものを素性として使用する．

III. 無声時間，オーバーラップ時間の割合

II と同様の場合，参加者の発話が時間に対して極端に疎密になることが考えられる．以上の理由で，参加者の発話の疎密さの特徴を捉えるために，$S_a$ 内の無声時間，オーバーラップ時間の総和を $S_a$ で割ったものを素性として使用する．

IV. 一定時間以上の無声時間の回数

II と同様に，III に加えてより時間的な発話の疎密さの特徴を捉えるために，一定時間以上の無声時間を抽出し，その回数を素性として使用する．また，オーバーラップ時間に対してこの素性を用いなかった理由は，無声時間に比べてオーバーラップ時間は比較的小さいためである．

V. 話者交代数

II と同様の場合，参加者が意見を活発に交換する，またはほとんどの参加者が意見を述べない状況が考えられる．以上の理由で，参加者の議論参加の特徴を捉えるために，$S_a$ 内の話者交代数を抽出し素性として使用する．

## 4. 実験

3.1 節で作成したデータセットを用いて，実験を行った．ただし，Social role annotation のスライスの平均時間が 30 秒であることから，$S_p$ =30 秒に設定した．また，対話単位で 10 分割交差検証を行いモデルを評価した．分散表現に

表 2　素性別の予測結果
**Table 2**　Result for every feature.

| 素性 | 適合率 | 再現率 | F 値 |
|---|---|---|---|
| 言語的素性のみ | 0.74 | 0.76 | 0.75 |
| 非言語的素性のみ | 0.73 | 0.55 | 0.63 |
| すべての素性 | 0.74 | 0.76 | 0.75 |

表 3　$S_a$ 別の分類結果
**Table 3**　Result for every $S_a$.

| $S_a$[sec] | 適合率 | 再現率 | F 値 |
|---|---|---|---|
| 5.0 | 0.71 | 0.71 | 0.71 |
| 10.0 | 0.74 | 0.72 | 0.73 |
| 20.0 | 0.75 | 0.75 | 0.75 |
| 30.0 | 0.74 | 0.76 | 0.75 |

は Wikipedia と Web ニュースを言語資源として，fastText [9] で学習したものを使用した．

分類器には，Support Vector Machine（SVM）を使用した．SVM は線形分類器であるが，カーネルトリックを用いることにより非線形分類モデルを構築することができる．SVM は，LIBSVM [10] の実装を使用し，カーネルには RBF カーネルを使用した．本節では，素性別の予測結果について述べ，次に $S_a$ 別の予測結果について述べ，最後に対話行為別の予測結果について述べる．

### 4.1　素性別の予測結果

使用する素性を，言語的素性（3.2 節の I，II）と非言語的素性（3.2 節の III，IV, V）に分類し，素性の有効性について検証した．モデルを，言語的素性のみ，非言語的素性のみ，すべての素性の 3 種類で構築し，予測結果を適合率，再現率，F 値で評価した．ここで，$S_a$ =30 秒に設定した．予測結果を表 2 に示す．表 2 から，非言語的素性のみでの予測結果の評価が言語的素性に比べて低く，言語的素性とすべての素性を用いた予測結果では評価は変わらなかった．まず，非言語的素性のみでの予測結果の評価が言語的素性に比べて低かったことについて考える．これは，素性の中で非言語的素性は比較的有効ではないことを示す．

次に，言語的素性のみとすべての素性での予測結果の評価が変わらなかったことについて考える．このことからも，非言語的素性が比較的有効ではないことがわかった．一方で，言語的素性のみでは予測できなかったが，すべての素性を用いた場合に予測できた時刻が存在した．そのような時刻では，他の予測時刻と比べてオーバーラップ時間の割合が比較的大きくなっていた．このことから，非言語的素性は言語的素性に比べ有効ではないが，非言語的素性の予測への寄与も確認できた．

素性の改良として，音声から得られる他の素性を用いることが考えられる．ピッチや強度などの韻律情報は，音声を対象とした話者の役割に関する研究 [7] [11] でよく用いられるものであるため検討の必要がある．また，言語的素性ではフィラーなどの特定の単語の出現を捉えるために単語の分散表現の平均を用いた．しかし，単語の分散表現の平均では，特定の単語の情報が一部落ちてしまう．このことから，フィラーなどの特定の単語を設定し，言語的素性としてその頻度ベクトルなどを用いることが考えられる．

### 4.2　$S_a$ 別の予測結果

$S_a$ を 30 秒，20 秒，10 秒，5 秒と変化させ，それぞれについてモデルを構築し，予測結果を適合率，再現率，F 値で評価した．予測結果を表 3 に示す．表 3 から，$S_a = 20, 30$ 秒で F 値が最大となっており，$S_a$ が小さくなるにつれて F 値が減少していた．

取りまとめ発話が行われると予測した時刻の直前の状況に $S_a$ による違いが生じているかを分析した．まず，$S_a = 30$ 秒では取りまとめ発話を予測できていたが，$S_a = 5$ 秒では取りまとめ発話を予測できていなかった対話中の時刻について分析した．$S_a$ =30 秒の設定でのみ取りまとめ発話を予測できた時刻の直前では，直前の発話の単語数が極端に少なく，無声時間が比較的長い状況が確認できた．このような状況下において $S_a = 5$ 秒のような設定では，$S_a$ 内に単語が出現しない場合が多く，非言語的素性に比べて有効であった言語的素性の情報が得られず，予測ができなかったと考えられる．次に，$S_a = 5$ 秒では取りまとめ発話を予測できていたが，$S_a = 30$ 秒ではよりまとめ発話を予測できていなかった対話中の時刻について分析した．$S_a$ =5 秒の設定でのみ取りまとめ発話を予測できた時刻の直前では，直前の発話内の単語数が比較的多く，無声時間が比較的短い状況が確認できた．このような状況下において $S_a = 30$ 秒のような設定では，$S_a$ 内に多くの単語が含まれ，分散表現の平均を算出する際に単語の情報が薄まりすぎてしまい，予測ができなかったと考えられる．以上のことから $S_a$ の設定により捉えられる状況が変化するため，$S_a$ の最適化や複数の $S_a$ を組み合わせて用いることが必要と考えられる．

### 4.3　対話行為別の予測結果

$S_a = 30$ 秒に設定し，すべての素性を用いたモデルでの予測結果に対して，3.1.2 節で決定した取りまとめ発話の 2 つ目の条件である，特定の対話行為タグ別に再現率を算出し，比較した．対話行為タグ別の再現率を表 4 に示す．表 4 から Elicit-Assessment（評価の要求）で再現率が最大となっていた．このことから，Elicit-Assessment の直前で比較的特徴が表れており，予測が比較的網羅的に行えていることを確認した．一方で，Elicit-Comment-about-Understanding（理解を示す発話の要求）で再現率が最小となっていた．しかし，テストデータ内全体での Elicit-Comment-about-

表 4　対話行為タグ別の再現率
**Table 4**　Result for every dialogue-act tag.

| 対話行為タグ | 件数 | 再現率 |
|---|---|---|
| Inform | 1305 | 0.75 |
| Suggest | 311 | 0.76 |
| Offer | 122 | 0.78 |
| Elict-Inform | 178 | 0.79 |
| Elicit-Offer-or-Suggestion | 31 | 0.77 |
| Elicit-Assessment | 104 | 0.83 |
| Elicit-Comment-about-Understanding | 2 | 0.5 |

Understanding の出現回数が 2 件と著しく少ないため，他の対話行為と比べて特徴が表れづらいということはできない．また，Elicit-Comment-about-Understanding を除いた対話行為タグの再現率は，いずれも 0.75 以上となっており，予測の大きな偏りは確認されなかった．このことから，少なくとも SVM で分類できる程度で，取りまとめ発話の直前の状況に共通の特徴が表れていたと考えられる．しかし，3.1.2 節で決定した取りまとめ発話の 2 つ目の条件である特定の談話行為タグが，取りまとめ発話の定義と合致しているとは言えない．そのため，人手による評価実験などをすることで 3.1.2 節で定めた条件が，取りまとめ発話の定義と合致しているかを評価する必要がある．

## 5. おわりに

本研究では，複数人対話における取りまとめ発話の予測および予測に用いたモデルの検証を行った．今後は，素性の獲得方法や有効な言語的，非言語的素性を導入し，予測精度の向上を図る．例えば，獲得する素性で $S_a$ を分けることが考えられる．言語的素性では，$S_a$ の値が小さすぎると単語が含まれないことがある．一方で音声のピッチや強度などの素性は，値の変化がフレーム単位である．またこのような素性は，一般的に最大，最小や平均などの回帰関数にかけてから用いられる．そのため，$S_a$ の値が大きすぎると大幅に情報が欠けてしまうことがある．これらのことから，素性の特性ごとに適切な $S_a$ を定めることで予測精度の向上が見込める．

また，デジタルファシリテータとして実世界で運用するためにはリアルタイムな素性の獲得と予測が必要となる．例えば，本研究では言語的素性の獲得の対象として AMI Meeting Corpus の人手による書き起こしデータを使用した．しかし，実際の運用では人手での書き起こしよりも正確性に劣る音声認識の出力などを利用することになる．AMI Meeting Corpus では，機械による書き起こしデータも公開されているため，これを利用し予測モデルを構築した場合の分析などを行う必要がある．また，本研究では，素性の獲得や予測の時間に対する検証は行ったいないため，今後取り組んでいきたい．

**謝辞**　本研究は科研費 17H01840 の助成を受けたものです．

## 参考文献

[1] Kirikihira, R. and Shimada, K.: Discussion Map with an Assistant Function for Decision-Making: A Tool for Supporting Consensus-Building, *International Conference on Collaboration Technologies*, Springer, pp. 3–18 (2018).
[2] Tsukasa, S., Takashi, Y. and Kazutaka, S.: Analysis of Facilitators' Behaviors in Multi-party Conversations for Constructing a Digital Facilitator System, *Collaboration Technologies and Social Computing* (Egi, H., Yuizono, T., Baloian, N., Yoshino, T., Ichimura, S. and Rodrigues, A., eds.), Cham, Springer International Publishing, pp. 145–158 (2018).
[3] 大本義正，戸田泰史，植田一博，西田豊明：議論への参加態度と非言語情報に基づくファシリテーションの分析，情報処理学会論文誌，Vol. 52, No. 12, pp. 3659–3670 (2011).
[4] Lala, D., Milhorat, P., Inoue, K., Ishida, M., Takanashi, K. and Kawahara, T.: Attentive listening system with backchanneling, response generation and flexible turn-taking, *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue*, pp. 127–136 (2017).
[5] Skantze, G.: Towards a General, Continuous Model of Turn-taking in Spoken Dialogue using LSTM Recurrent Neural Networks, *SIGdial Conference* (2017).
[6] Carletta, J.: Unleashing the killer corpus: experiences in creating the multi-everything AMI Meeting Corpus, *Language Resources and Evaluation*, Vol. 41, No. 2, pp. 181–190 (2007).
[7] Sapru, A. and Bourlard, H.: Automatic Recognition of Emergent Social Roles in Small Group Interactions, *Multimedia, IEEE Transactions*, Vol. 17, No. 5, pp. 746 – 760 (online), DOI: 10.1109/TMM.2015.2408437 (2015).
[8] Mikolov, T., Chen, K., Corrado, G. and Dean, J.: Efficient estimation of word representations in vector space, *arXiv preprint arXiv:1301.3781* (2013).
[9] Bojanowski, P., Grave, E., Joulin, A. and Mikolov, T.: Enriching word vectors with subword information, *Transactions of the Association for Computational Linguistics*, Vol. 5, pp. 135–146 (2017).
[10] Chang, C.-C. and Lin, C.-J.: LIBSVM: a library for support vector machines, *ACM transactions on intelligent systems and technology (TIST)*, Vol. 2, No. 3, p. 27 (2011).
[11] Weninger, F., Krajewski, J., Batliner, A. and Schuller, B.: The Voice of Leadership: Models and Performances of Automatic Analysis in Online Speeches, *IEEE Transactions on Affective Computing*, Vol. 3, No. 4, pp. 496–508 (online), DOI: 10.1109/T-AFFC.2012.15 (2012).