

発話間関係の構造化による会議録からの議論マップ自動生成システム

林 佑磨[†] 山名 早人^{††,†††}

[†] 早稲田大学基幹理工学研究科 〒169-8555 東京都新宿大久保 3-4-1

^{††} 早稲田大学理工学術院 〒169-8555 東京都新宿大久保 3-4-1

^{†††} 国立情報学研究所 〒101-8430 東京都千代田区一ツ橋 2-1-2

E-mail: {yuma.hayashi, yamana}@yama.info.waseda.ac.jp

あらまし 組織では、活動の現状を共有し、今後の方針や重要事項を決定するために会議を行う。通常、会議の時間は限られており、しばしば議題が持ち越される。そのため、参加者が過去の会議内容を事前に把握することは、会議を円滑に進める上で重要である。そのような背景から、会議録の自動要約や閲覧システムに関する研究が行われている。しかし、自動要約では発話間の関係が構造化されておらず、閲覧システムでは会議録全体が対象であるため、いずれも会議の流れや決定事項などの要旨が掴みづらいという問題がある。そこで本研究では、1) 会議録からの重要発話抽出を行い、2) 発話間における質疑応答の対応付けを行った後に、3) 発話間の構造を保持した議論マップとして出力することで、会議要旨の素早い把握を可能にする統合システムを提案する。発話間の構造化により、重要発話抽出では既存手法の MMR-centroid と比べ ROUGE-1 が最大 0.111 向上した。また、4 名の被験者実験により、構造化をせずに出力した場合と比較して、構造化を行った議論マップによる出力の有用性が示された。

キーワード 会議録, 質疑応答対応付け, 議論マップ, AMI

1. はじめに

企業や学術機関などの組織では、現状の共有を行い、今後の方針や重要事項を決定するために会議を行う。多くの場合、会議の時間は限られており、しばしば複数回に渡り議題が持ち越される。参加者が会議出席前に過去の会議内容を把握しておくことは、会議を円滑に進める上で重要である。しかし、会議中の発話を書き起こした会議録は膨大な量であり、全てを読むには多くの時間を要する。

そのような背景から、ICSI Meeting Corpus^(注1) [1] や AMI Meeting Corpus^(注2) [2] といった、会議の円滑な進行や会議録の効率的な閲覧に関する研究の支援を目的とした、人手により様々なアノテーション情報が付与された会議録データが提供されている。人手で付与されたアノテーション情報を機械的に付与する、という研究も同時に行われているため、多くの会議録を対象とした研究では、会議録データに付属のアノテーション情報を効果的に用いている。

これらのデータセットを利用した会議録の自動要約に関する研究は、1) 重要な発話を取り出す抽出型の要約手法 [3] [4] [5] [6] [7] [8] [9] [10] や、2) 発話中の冗長部分削除や複数の発話をまとめる生成型の要約手法 [11] [12] [13] など、今までに多く行われている。これらの研究では、読むべき対象となる発話の量を大幅に削減することが可能であるが、発話間の関係が構造化されておらず、質疑応答などの対応する発話の紐付けは読み手が行わなければならないという問題がある。

一方で、構文解析や意味解析を応用することで、文内および文間の構造化を行い、文章からのマインドマップ自動生成を行

うような研究 [14] [15] [16] [17] が存在する。しかし、会議中の発話は自然発生的なものであるため、言い間違えなどによる不自然な単語の連続や構文的に正しくない言い回しなどが多く含まれている。このため、会議録中の発話に対しては構文解析がうまく機能せず、これらの手法の応用は困難である。また、会議内の発話を対象とした、構文解析を用いない会議内容の効率的な閲覧システムの提案 [18] [19] [20] [21] も行われているが、これらの多くは閲覧すべき発話の選定および構造化が行われておらず、会議の流れや決定事項などの要旨が掴みづらいという問題がある。

そこで本研究では、AMI Meeting Corpus [2] を対象に、1) 発話間の関係が構造化されていない、2) 確認すべき発話が限定されていない、という従来の問題点を解決し、会議の流れや決定事項などの要旨を効率良く把握することを可能にする統合システムを提案する。具体的には、以下の 3 手順で実現する。

1. 会議録からの重要発話抽出
確認対象となる議論内容を限定し、2) の問題を解決する
2. 発話間関係の構造化
重要発話内にある質問発話を特定し、元の会議録からその回答を紐付け構造化を行うことで、1) の問題を解決する
3. 議論マップによる出力
回答が補充された重要発話を、発話間の構造を保持した議論マップとして出力することで、1) と 2) の問題を解決する

本稿の構成は以下である。まず 2 節で関連研究について説明し、3 節で提案システムの詳細を説明する。4 節では重要発話抽出に関する評価、および被験者実験による議論マップの有用性評価とその結果を示す。5 節でまとめと今後の課題を述べる。

(注1) : ICSI Meeting Corpus: <http://www1.icsi.berkeley.edu/Speech/mr/>

(注2) : AMI Meeting Corpus: <http://groups.inf.ed.ac.uk/ami/corpus/>

2. 関連研究

2.1 会議録の自動要約に関する研究

会議録の自動要約は、会議録の閲覧にかかる時間を短縮し、会議の流れや決定事項などの重要項目を効率よく把握できるようにすることが目的である。

2.1.1 抽出型要約に関する研究

既存の自動要約に関する研究の大部分が抽出型要約によるものである。抽出型の要約では、会議録中から重要であると思われる発話を抜き出し、それらを並べたものを要約とする。従って、抽出型要約の目標は、議論内容における重要箇所の網羅的な抽出、および同一内容の文を含めないよう冗長性を最小化する、ということになる。

この観点に基づく問題の定式化を行った研究として、Gillickら [4] は、2009年に教師なし学習の枠組みである ILP (Integer Linear Programming) を利用して、AMI meeting corpus [2] に含まれる会議録からの重要発話抽出を行った。Gillickらは、会議録中の発話に含まれる会議内容を特徴づけるような形態素あるいは形態素列を会議の concept として定義し、なるべく広く concept を含むよう ILP による定式化を行った。

Gillickらは、会議録中の発話構造や発話間の対応を考慮していない点や、発話の抽出による要約が人手で作成された要約とは本質的に大きく異なる点を課題として挙げている。

また、Xieら [8] は、2010年に教師あり学習の枠組みである SVM (Support Vector Machine) および SVR (Support Vector Regression) を利用して、ICSI meeting corpus [1] に含まれる会議録からの重要発話抽出を行った。Xieらの研究における主な貢献は、1) 重要発話抽出のモデル学習時にサンプリングを適用した点と、2) 二値分類問題ではなく回帰問題として重要発話抽出問題を解いた点である。

前者に関して、モデル学習時のサンプリングにより、重要発話が会議録全体に対して相対的に少ない中での効率的な学習方法を提案した。また後者に関して、重要発話抽出を回帰問題に帰着することで、会議録中の重要発話を選定した人による、選定基準の違いから生じるラベルノイズに頑健な学習を実現し、精度が向上するケースがあることを示した。

Xieらは、サンプリングや回帰による解き方に有効な素性の特定や、会議内容の閲覧ブラウザなどへ応用した先での評価を課題として挙げている。

2.1.2 生成型要約に関する研究

近年になり、元の会議録に含まれない文を創ることで、人手による要約に近づくことを目的とした生成型要約が行われるようになってきた。

Mehdadら [12] は、2013年に単語グラフを利用した複数発話の集約による文生成を行い、入力としての会議録を受け取ってから、出力としての要約を生成するまでの全過程を扱うような end-to-end の生成型要約フレームワークを提案した。また、Murrayら [13] は、2015年にマルコフ決定過程 (MDP: Markov Decision Process) を用いた、生成型要約手法を提案し、会議録データへの適応を行った。

しかし、いずれの生成型要約の手法においても、大きな問題として、構文的に不自然な文や意味を成さないような文を多く生成してしまう、という点が挙げられている。

本稿で提案するシステムでは、発話間における質疑応答の対応付けをすることで構造化を行い、その構造を保持した上で重要発話の抽出を行う。

2.2 テキストの構造化による可視化に関する研究

テキスト情報のみからでは、そこに含まれる文の構造が分からないという問題がある。そこで、テキストを構造化し可視化を行うことで、それらの情報を補完するという研究が存在する。

Adbeenら [14] は、2009年に純粋なテキストに対し構文解析や意味解析を適応することで、マインドマップを自動生成する最初のシステムを提案した。また2013年には、Purwariantiら [16] が、インドネシア語に特化したテキストの構造解析を用いてマインドマップの自動生成を行った。

Elhoseinyら [17] は、2015年にそれまでに提案されていたテキストからのマインドマップ生成手法を拡張し、階層的な表示が可能なマインドマップの自動生成を行うシステムを提案した。

しかし、これらの研究はいずれも、構文的な解析が可能な書き言葉のテキストを対象としている。そのため、これらの手法を構文的な解析がうまく機能しないような、会議録中の発話を対象として適用することは不可能である。

一方で、構文的な解析を用いない、会議録を対象とする効率的な閲覧システムに関する研究が存在する。松村ら [20] は、2003年に議事録からトピックのセグメント分割を行い、セグメント単位での構造化を行うことで構造化されたマップの作成を行った。構造化においては、セグメント内の発話に含まれる単語を利用してセグメント間の類似度をとり、ある閾値以上であればエッジを張るという方法により行っている。

趙ら [19] は、2006年に議論の中で展開されるテーマは名詞の集合で表現できるという過程の元で、議事録内に含まれる名詞をノードとし、関係のあるノード間にエッジを貼った議論マップの自動生成を行った。また、森ら [21] は、2007年にリフレクションのための、発話間の関係に着目した議論構造モデルの提案と、そのモデルに基づく議事録の自動構造化手法の提案を行った。

これらの研究ではそれぞれ、1) 一つのトピックセグメントに含まれる文量が多すぎる点や、2) 議論の流れに沿って正しく可視化しないと余計に見づらくなる点、3) ノードが名詞であるため分かりづらい点などが問題に挙げられている。これらの問題は、議事録に含まれる発話全体が対象であるために生じている。

本稿で提案するシステムでは、重要発話の抽出を行い、確認すべき対象となる発話を制限することでこの問題を解決する。

3. 発話間関係の構造化による会議録からの議論マップ自動生成システム

3.1 システムの概要

本システムは、会議録中から重要発話を抽出し、発話間の質疑応答関係を構造化した上で、構造を保持した議論マップの自動生成を行う統合的なシステムである。ここで、議論マップと

は、1) 会議全体の流れ、2) 会議中での重要な発話、3) 会議中の重要な質問に対する回答の紐付け、をそれぞれ構造化し表現したものである。図 1 にシステム全体の処理フローを示す。

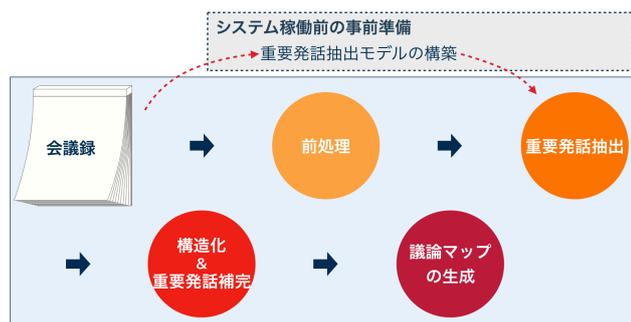


図 1: 議論マップ自動生成システムの処理フロー

本システムは、まず入力としての会議録データを受け取った後、1) 前処理、2) 重要発話抽出、3) 質疑応答の対応付けと重要発話の補完、4) 議論マップとして可視化、という大きく 4 つの処理を行い、出力としての議論マップを生成する。システムが内部で行う 4 つの主な処理に関する大まかな内容は以下である。

システム内部で行う主な処理

処理 1. 前処理

Stopwords 除去や言い直しの除去を行い、発話内の重要フレーズの特定を行う。

処理 2. 重要発話抽出

事前に構築した、重要発話識別のための教師ありオンライン学習 AROW のモデルを利用して、議論マップに残すべき重要発話の抽出を行う。

処理 3. 質疑応答の対応付けと重要発話の補完

取り出された重要発話中から質問であるような発話を選定し、その回答にあたるような発話の抽出と紐付けを行う。また、取り出された回答発話を重要発話として補充する。

処理 4. 議論マップによる可視化

抽出された重要発話を、会議中の主なトピックの流れや発話者、質疑応答の情報が素早く分かるような議論マップとして可視化を行う。

なお、本システムへの入力は、3.2 にて詳細を説明する、AMI Meeting Corpus [2] に含まれる人手により書き起こしが行われた会議録のデータである。このデータには、各発話に対する意味役割である Dialogue Act や発話開始/終了時間や話者情報などがアノテーションとして付与されている。

また、本システムは稼働前の事前準備として、訓練用の会議録データから、重要発話抽出のためのモデル構築を行っておく。事前のモデル構築に関する詳細は、3.6.2 にて述べる。

3.2 対象とする会議録データセットについて

本研究では、会議録のデータセットとして AMI Meeting

Corpus [2] を用いる。AMI Meeting Corpus には、1) 議題が与えられているシナリオ有りの会議と、2) 議題が与えられていないシナリオ無しの会議に関する会議録があり、それぞれ人手による書き起こしと機械による書き起こしのデータが存在する。本研究では、シナリオ有り会議録の人手による書き起こしデータを対象とする。なお、シナリオ有りの会議では、4 人のそれぞれ異なる役をもつメンバが、架空の会社で新たなリモートコントロールの製作に関する会議を 4 回にわたり行う。ただし、あらかじめ定まった発話内容などは存在せず、会議中の発話は全て自然発生的なものである。

なお、シナリオ有りの会議において、4 名の会議参加者に割り当てられる役は以下である^(注3)。

表 1: AMI Meeting Corpus のシナリオ有り会議における参加者の役

役名	役の説明
PM	プロジェクトマネージャー (Project Manager)。会議全体の司会進行を中心的に務める。
ME	マーケティング専門家 (Marketing Expert)。ユーザからの要求や市場傾向の調査、プロトタイプの評価を担当する。
UI	ユーザインタフェースデザイナー (User Interface Designer)。開発するリモートコントロールの外見的デザインや、提供する機能に関する責任を担当する。
ID	工業デザイナー (Industrial Designer)。開発するリモートコントロールに搭載するボタンなどの要素決定とその機能に関する設計を担当する。また、UI と共に外見的デザインも手掛ける。

また、AMI Meeting Corpus には、人手による豊富なアノテーション情報が存在する。本研究では最終的に、以下に示す情報が付与されている、計 134 の会議録データを対象とした。

- 人手により記載された会議の要約文
各会議に対し、元の会議録の約 6% (平均で約 290 単語) に相当する分量でまとめられた、人手による会議内容の要約である。この人手要約は、重要文抽出の評価における正解の参照要約として利用する。
- 人手により選択された会議録中の重要発話
各会議録中から、重要だと思われる発話に人手でラベルをつけたもの。この人手により取り出された重要発話の情報は、重要発話抽出のためのモデル構築時に利用する。
- Dialogue Act (DA)
会議録中の各発話に人手で付与された意味役割^(注4)。DA には、何らかの提案を行っているような発話を表す “SUGGEST” や、何らかの情報を与えるような発話を表す “INFORM” など、全 15 種類が存在する。
- 会議中のトピックセグメント (主な話の流れの境界)

(注3): 各役の詳細は、<http://groups.inf.ed.ac.uk/ami/corpus/scenariomeetings.shtml> に記載されている。

(注4): Dialogue Act の詳細および付与の基準は、AMI Meeting Corpus のアノテーションガイド (http://groups.inf.ed.ac.uk/ami/corpus/Guidelines/dialogue_acts_manual_1.0.pdf) に記載されている。

AMI Meeting Corpus にはテスト用の会議録が用意されている。テスト用会議録の会議シリーズ（1つの会議シリーズは4回の会議からなる）は“ES2004”，“ES2014”，“IS1009”，“TS3003”，“TS3007”であり，合計20回分の会議録から構成される。

重要発話抽出のためのモデル構築時は，上記のテスト用会議録データを除く計114個の会議録データを，訓練用データとして利用する。

3.3 システムで用いる要素技術の説明

提案システムでは，高精度かつ高速なオンライン型の線形識別器である AROW (Adaptive Regularization of Weight Vectors) [22] を用いて重要発話抽出を行う。AROW は CW (Confidence-Weighted) [23] と呼ばれる手法の拡張であるため，以下ではまず CW について説明し，次に CW の欠点を解決した AROW の説明を行う。なお，以下の説明は，我々が以前執筆した [24] で行った説明に基づいている。

3.4 CW (Confidence-Weighted)

CW は AROW の元となるオンライン学習の二値分類器である。オンライン学習のため，素性ベクトル \mathbf{x}_t が与えられる度に予測ラベル \hat{c}_t を求め，正解ラベル c_t と比較することでモデルの重みベクトル \mathbf{w} を更新する。CW では重みベクトル \mathbf{w} が平均 $\boldsymbol{\mu} \in \mathbb{R}^m$ ，分散 $\Sigma \in \mathbb{R}^{m \times m}$ の正規分布 $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$ に従うと仮定されている。重みベクトルの中で分散値が大きいパラメータに関しては，まだ自信 (confidence) があまりない状態と考え，大きくパラメータを更新する。逆に，分散値が小さなパラメータに関しては，頻出な特徴のためにもう既に十分な情報が得られていると考え，小さくパラメータを更新する。実際にラベルを推定する際には，重みベクトルの期待値 $\mathbb{E}[\mathbf{w}] = \boldsymbol{\mu}$ を用いて行う。

t 回目の学習において，学習用の素性ベクトル \mathbf{x}_t および正解ラベル c_t が与えられたとする。この際，CW は以下の式 (1) に示す制約付き最適化問題を解くことで，重みベクトルに関する分布を更新し，重みベクトルの新たな平均 $\boldsymbol{\mu}_t$ と分散 Σ_t を得る。

$$\begin{aligned} (\boldsymbol{\mu}_t, \Sigma_t) = \arg \min_{\boldsymbol{\mu}, \Sigma} \text{KL}(\mathcal{N}(\boldsymbol{\mu}, \Sigma) \parallel \mathcal{N}(\boldsymbol{\mu}_{t-1}, \Sigma_{t-1})) \\ \text{s.t. } \mathbb{P}_{\mathbf{w} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)}[c_t(\mathbf{w} \cdot \mathbf{x}_t) \geq 0] \geq \eta \end{aligned} \quad (1)$$

ここで， $\mathcal{N}(\boldsymbol{\mu}_{t-1}, \Sigma_{t-1})$ は t 回目の学習による更新を行う前の重みベクトルに関する分布であり， $\text{KL}(\mathcal{N}(\boldsymbol{\mu}, \Sigma) \parallel \mathcal{N}(\boldsymbol{\mu}_{t-1}, \Sigma_{t-1}))$ は $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$ と $\mathcal{N}(\boldsymbol{\mu}_{t-1}, \Sigma_{t-1})$ 間のカルバック-ライブラーダイバージェンスである。また， $\eta \in (0.5, 1]$ は更新の度合いを制御するハイパーパラメータである。

式 (1) の制約条件において， $\mathbb{P}_{\mathbf{w} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)}[c_t(\mathbf{w} \cdot \mathbf{x}_t) \geq 0]$ は与えられた素性ベクトル \mathbf{x}_t に対して正しくラベル予測される確率を表している。したがって，CW は $\eta \in (0.5, 1]$ 以上の確率で正しく分類されるという条件を満たした上で，更新前の重みベクトルの正規分布に最も近い正規分布を求めることで学習を行う。この制約条件は，与えられた素性ベクトル \mathbf{x}_t を常に正

しく分類出来るようにモデルを更新することを意味するため，ノイズデータに極めて弱く過学習を起こしやすいという欠点が存在する。

3.5 AROW (Adaptive Regularization of Weight Vectors)

AROW は，上記の CW が持つ欠点を，制約条件を目的関数の一部に正則化項として持たせることにより解決した手法である。具体的には，以下の式 (2) に示す最適化問題を解くことで，重みベクトルに関する分布を更新する。

$$\begin{aligned} (\boldsymbol{\mu}_t, \Sigma_t) = \arg \min_{\boldsymbol{\mu}, \Sigma} \text{KL}(\mathcal{N}(\boldsymbol{\mu}, \Sigma) \parallel \mathcal{N}(\boldsymbol{\mu}_{t-1}, \Sigma_{t-1})) \\ + \frac{1}{2r} \ell_{h^2}(c_t, \boldsymbol{\mu} \cdot \mathbf{x}_t) + \frac{1}{2r} \mathbf{x}_t^T \Sigma \mathbf{x}_t \end{aligned} \quad (2)$$

ここで， $r > 0$ はモデルの更新を調節するハイパーパラメータである。式 (2) は3つの項から構成され，それぞれの項は以下のような意味を持つ。

- (1) $\text{KL}(\mathcal{N}(\boldsymbol{\mu}, \Sigma) \parallel \mathcal{N}(\boldsymbol{\mu}_{t-1}, \Sigma_{t-1}))$
この項を小さくすることは，パラメータの更新を小さく抑え，更新前の重みベクトルの正規分布に最も近い正規分布を求めることを意味する。
- (2) $\frac{1}{2r} \ell_{h^2}(c_t, \boldsymbol{\mu} \cdot \mathbf{x}_t)$
 $\ell_{h^2}(c_t, \boldsymbol{\mu} \cdot \mathbf{x}_t) = (\max\{0, 1 - c_t(\boldsymbol{\mu} \cdot \mathbf{x}_t)\})^2$ は二乗ヒンジ損失である。この項を小さくすることは，現在与えられているデータに対する予測間違いをなるべく少なくするような重みベクトル \mathbf{w} の平均 $\boldsymbol{\mu}$ を求めることを意味する。ただし，この項には二乗ヒンジ損失関数以外の損失関数を適用することも可能である。
- (3) $\frac{1}{2r} \mathbf{x}_t^T \Sigma \mathbf{x}_t$
この項を小さくすることは，重みベクトル \mathbf{w} の各素性に関する分散 (自信のなさ) を，学習を進めるにつれて小さくしていくことを意味する。

以上から，AROW は 1) \mathbf{w} の分布を今までの正規分布になるべく近く，2) 現在の学習データを正しく分類し，3) \mathbf{w} の各素性に関する自信を少しずつ上げていくことで，CW の欠点であったノイズのあるデータに頑健なオンライン学習を実現している。

3.6 システムの詳細

本項では，提案システム内部で行う，1) 前処理，2) 重要発話抽出，3) 質疑応答の対応付けと重要発話の補完，4) 議論マップとしての可視化，という主な4つの処理に関する詳細をそれぞれ説明する。

3.6.1 前処理

前処理では，主に1) 各発話からの Stopwords 除去や言い直しによる冗長性の除去と，2) 発話中の重要フレーズの特定を行う。なお，英語テキストの形態素解析および品詞推定に Stanford 大学が提供している CoreNLP [25] を利用する。

本研究では，Stopwords として，“the” や “a” などの冠詞に代表される英語テキストに多々含まれる単語に加え，“uh” や

“ah”などのフィラーをはじめとする発話にしばしば含まれるが特に重要な意味を持たないような単語も含む。また、会議中の発話には、しばしば同じことを繰り返している場面が見受けられる。そこで、言い直しによる冗長性の除去として、発話中において連続する同じ単語を取り除く。

本研究において、重要フレーズとは、発話内に含まれ会議内容を特徴づけるような形態素あるいは形態素列である。重要フレーズになりうる形態素列は [5] に基づいて、以下の品詞列に適合するものとして定義する。

重要フレーズの品詞列

$JJ^*(NN|NNS|FW|CD)^+((DT|IN)^+JJ^*(NN|NNS|FW|CD)^+)^*$

ただし、上記の品詞記号は Penn TreeBank の品詞記号に従っている。それぞれの意味を表 2 に示す。

表 2: Penn TreeBank の品詞記号と意味

品詞記号	品詞記号の意味
JJ	形容詞
NN	名詞 (単数形)
NNS	名詞 (複数形)
FW	外来語
CD	数字
DT	限定詞
IN	前置詞, あるいは従属接続詞

例えば、上記の定義により “we are asked to make new remote control for television.” という発話から取り出される重要フレーズは、“new remote control” と “television” となる。これらの取り出した重要フレーズは、次項にて説明する、重要発話抽出のためのモデル構築の際の素性として利用する。

3.6.2 重要発話抽出

本フェーズでは、会議録からその会議における重要な発話の抽出を行う。提案システムでは、3.5 にて説明した、オンライン型の線形識別器である AROW (Adaptive Regularization of Weight Vectors) [22] を用いて重要発話抽出を行う。

AROW は教師あり学習の二値分類器であるため、モデルの構築手順は以下となる。

重要発話抽出のための AROW モデル構築手順

1. AMI Meeting Corpus からの、学習に用いる素性の選択
2. 選定した素性を用いて、各発話の素性ベクトルを作成
3. AMI Meeting Corpus に含まれている、会議録中の各発話が人手により選択されたかどうかの情報を正解ラベルとして、AROW を学習

なお、AROW の持つハイパーパラメータ r に関しては、訓練データセットにおける 10-fold cross validation を行い、グリッドサーチをかけることで適切に設定を行う。

表 3: AROW モデルの構築に用いる素性一覧

素性	素性の説明
話者情報	発話を行った話者の役に関する情報。PM, ME, UI, ID の 4 種類。
発話役割	3.2 にて記載した、Dialogue Act。
発話位置情報	会議全体を 5 分割したうち、対象の発話がどの区分に属するかという、発話の時間位置情報。
発話時間長	発話を始めてから終わるまでにかかった発話時間の長さ。
単語数	発話に含まれるユニークな単語の数。
総単語数	発話に含まれる全ての単語の数。
品詞数	発話に含まれるユニークな品詞の数。
単語 N-gram	単語の N-gram 情報。 ($1 \leq N \leq 3$)
品詞 N-gram	品詞の N-gram 情報。 ($1 \leq N \leq 3$)
重要フレーズ	前処理にて特定した、発話に含まれる重要フレーズの情報。
重要フレーズ数	発話に含まれる重要フレーズの数。

本研究において、使用した素性の一覧を表 3 に示す。

また、抽出された発話 u の重要度は、AROW が識別時に利用するマージンの値 $\mu \cdot x_u$ を用いて表現する。すなわち、マージンが大きい発話ほど重要であると定義する。

3.6.3 質疑応答の対応付けと重要発話の補完

本フェーズでは、まず重要発話抽出のフェーズにて抽出された重要発話に含まれる質問発話の選定を行い、次にその質問の発話に対する回答となるような発話の特定と紐付けを行う。最後に、特定した回答に相当する発話を重要発話として補充する。

重要発話に含まれる質問は、1) Dialogue Act (DA) の情報と、2) 文末が “?” かどうかにより、以下の擬似コード 3.1 に記載するルールに基づき推定する。

擬似コード 3.1 質問発話推定

入力: 対象発話 u , 質問 DA の集合 DA_q

出力: 質問発話かどうかの True, False

```

1: procedure ISQUESTION( $u$ )
2:   if  $DA(u) \in DA_q$  :                                ▷  $DA(u)$  は発話  $u$  の DA
3:     return True
4:   else if  $u$  の末尾が “?” :
5:     return True
6:   else
7:     return False

```

ただし、質問 DA は “Elicit-Comment-Understanding”, “Elicit-Assessment”, “Elicit-Offer-Or-Suggestion”, “Elicit-Inform” の 4 種類として定義する。

また、質問に対する回答の発話は、1) Dialogue Act (DA) の情報と、2) 発話が同じトピックセグメント内にあるかどうか、および 3) 話者情報を用いて、擬似コード 3.2 に記載するルールに基づき推定する。なお、取り出した回答発話が質問の場合は、その回答も再帰的に取り出す。

擬似コード 3.2 回答発話抽出

入力: 質問発話 q , 回答 DA の集合 DA_a , 会議録中の全発話 U

出力: 質問 q に対する回答発話の集合 U_a

```
1: function GETANSWER( $q, U, depth = 0, q_{orig} = \text{Null}$ )
2:   if  $depth = 0$  : ▷  $depth$  は 再帰の深さ
3:      $q_{orig} \leftarrow q$  ▷ ルートとなる質問発話を設定
4:      $U_a \leftarrow \emptyset$ 
5:     for  $u \in U$  : ▷ 会議録中の全発話に対する処理
6:       if  $topic(u) \neq topic(q)$  :
7:         break
8:       if isAnswerCandidate( $u, q, q_{orig}$ ) :
9:          $U_a \leftarrow U_a \cup \{u\}$ 
10:      if isQuestion( $u$ ) : ▷ 回答発話が質問の場合
11:         $U'_a \leftarrow \text{getAnswer}(u, U, depth + 1, q_{orig})$ 
12:         $U_a \leftarrow U_a \cup U'_a$ 
13:      break
14:    else
15:      continue
16:  return  $U_a$ 
```

擬似コード 3.3 回答発話候補の推定

入力: 対象発話 u , ルートとなる質問発話 q_{orig} , 直前の質問発話 q , 回答 DA の集合 DA_a

出力: 発話 u が質問 q に対する回答の候補であるかどうかの True, False

```
1: function ISANSWERCANDIDATE( $u, q, q_{orig}$ )
2:   if  $DA(u) \notin DA_a$  or  $topic(u) \neq topic(q)$  : ▷  $topic(u)$  は発話  $u$  のトピック
3:     return False
4:   if  $speaker(u) = speaker(q_{orig})$  : ▷  $speaker(u)$  は発話  $u$  の話者
5:     return False
6:   return True
```

上記ルールに記載したとおり、質問に対する回答発話は、1) 質問発話を発した話者以外の話者による、2) 同一トピックセグメント内の、3) 回答になりうる DA に適合し、かつ質問発話に最も近い発話として抽出される。なお、これらの抽出された回答発話は、どの質問に対する回答かという対応を保持したまま、元の重要発話に補充する形で追加する。ただし、補充の際は、元の重要発話に含まれていない場合のみ追加を行う。

3.6.4 議論マップによる可視化

本フェーズでは、重要発話抽出および質疑応答の対応付けを行った後の、構造化された会議録情報を元に、議論マップを生成する。議論マップとは、1) 会議全体の流れ、2) 会議中での重要な発話、3) 会議中の重要な質問に対する回答の紐付け、をそれぞれ構造化した上で素早く把握できるように可視化したものである。

具体的な議論マップの生成は、以下の手順に従って行う。

議論マップ生成の手順

1. 各発話を発した話者情報を、発話の先頭に付与する。
2. 各発話に含まれる重要フレーズを太字で強調する。
3. 抽出された重要発話をトピックセグメントで分割。
4. トピックセグメント内で、発話を時系列順で並べる。
5. 発話が質問の場合、その質問に紐づく回答の発話を、質問にぶら下げて表示する。

以上の手順により、生成される議論マップの例を図 2 に示す。なお、この例では AMI Meeting Corpus のテスト用データの一つである“ES2004c”から議論マップの自動生成を行った。



図 2: 提案システムが“ES2004c”から自動生成した議論マップ

4. 実験と評価

本システムの目的は、膨大な会議録から効率的に会議の流れや会議内で話された重要事項を得ることが出来るような議論マップの自動生成である。そこで、システムの評価に関しては以下の 2 点に関する評価実験を行う。

システム評価実験の内容

1. 会議録からの重要発話抽出に関する評価
会議録中から重要発話をきちんと取り出せているかという点に関する評価。
2. 構造化された議論マップによる出力の有用性評価
抽出した構造化済み重要発話の出力を、i) テキストとした場合、ii) 議論マップとした場合のどちらがより効率的に会議内容を把握可能か、という点に関する評価。

以下では、まず 4.1 で、重要発話抽出に関する評価を行い、続く 4.2 で、被験者実験による議論マップによる出力の有用性に関する評価を行う。

4.1 会議録からの重要発話抽出に関する定量的評価

本項では、会議録からの重要発話抽出に関する評価を行う。評価指標には、文書要約のタスクにおいて最も広く用いられ

表 4: 比較対象の重要発話抽出手法

既存/提案	手法名	
既存手法	MMR-centroid	Carbonell ら [27] による MMR-centroid System
	MMR-cosine	Gillick ら [4] による MMR-cosine System
	ILP-based	Gillick ら [4] による ILP-based System
	TextRank	Mihalcea ら [28] による TextRank System
	ClusterRank	Garg ら [6] による ClusterRank System
提案手法	AROW	質疑応答の対応付けを行わない提案手法
	AROW-QA	質疑応答の対応付けを行った提案手法

ている ROUGE [26] を用いた。ROUGE は、正解である参照要約と、システムが生成した要約との間で、どれだけ形態素の N-gram が一致するかの評価を行う。なお、会議中の自然発生的な発話と、人手で記載された要約とは、bi-gram より uni-gram で重なりを評価する方が適切であるという Gillick ら [4] の見解に基づき、本評価実験においては uni-gram の一致率を評価する ROUGE-1 における F-measure を評価指標として用いた。

実験のテストデータには、3.2 で説明を行った、AMI Meeting Corpus が提供する 20 回分のテスト用会議録データを利用した。また、参照要約にはテストデータに付属の人手により作成された会議の要約文を用いた。比較対象とする重要発話抽出手法を表 4 に示す。

重要発話の抽出に関する既存手法は、抽出した発話に含まれる単語数の合計が、予め与えられた単語数以下となるように重要発話の抽出を行う。そこで本実験においては既存手法との比較のため、多くの研究で採用されている、要約文に含まれる単語数が元の会議録に含まれる総単語数の 6% 以下となるような制約を設けた。この 6% (平均で約 290 単語) とは、人手で作成された要約文のに含まれる単語数の平均値と概ね一致する割合である。

なお、提案手法で用いている AROW は教師あり学習の二値分類器であるため、本来は各発話が要約に含まれるか否かを識別しており、設けられた単語数制限の元で重要発話と取り出す手法ではない。そこで、本実験における単語数の制約を満たすために、AROW のマージンにより定義した発話の重要度が高いものを優先的に要約に含めた。

ROUGE-1 による評価実験の結果を表 5 に示す。なお、表 5 に記載した既存手法の値は、過去に報告された同じデータセットおよび条件による実験結果の値を引用している [4] [6] [12]。

以上の結果から、本手法による重要発話抽出がより人手による要約に近い発話を取り出していることが分かる。また、質疑応答を行った AROW-QA のスコアが、行わなかった場合の AROW と比べ高いスコアを出している。このことから、重要な質問に対する回答を要約に含めることで、有益な情報を追加

表 5: ROUGE-1 による重要発話抽出の評価

手法名	ROUGE-1
MMR-centroid	0.182
MMR-cosine	0.21
ILP-based	0.24
TextRank	0.250
ClusterRank	0.275
AROW	0.291
AROW-QA	0.293

表 6: 被験者実験による議論マップの有用性評価

被験者	正答した質問数	
	構造化あり	構造化なし
被験者 A	2	0
被験者 B	3	2
被験者 C	2	1
被験者 D	1	0

出来ていることが分かる。

4.2 被験者実験による議論マップの有用性評価

本項では、重要発話抽出と構造化を行った結果の出力である議論マップの有用性に関して、被験者実験による評価を行う。実験では、以下の 2 通りの出力を用いて、被験者にこちらが予め用意した会議内容に関する質問に答えてもらった。

比較対象の出力

1. 構造化あり：1) 重要フレーズ強調、2) トピック分割、3) 質疑応答の紐付け等の構造化を行った議論マップ
2. 構造化なし：構造化を行わず、抽出した重要発話をプレーンテキストとして並べたもの

ここで、予め用意した質問は全て 4 択であり、重要発話全て読めば回答が可能な内容となっている。質問は全 3 問であり、回答時間は 3 分以内である。なお、AMI Meeting Corpus は英語テキストによる会議録であるため、被験者はなるべく英語力に偏りのない、20 代の理工系学生 4 名に依頼した。また、各被験者には、構造化ありの出力で会議 1 を、構造化なしの出力で会議 2 を評価という具合に、それぞれの出力で異なる 1 つずつの会議録に対して評価を行ってもらった。

本実験で用いる会議録データは、テストデータに含まれる 20 個の会議録から、無作為に取り出した “ES2004c” と “TS3003b” の 2 つを対象とし、被験者毎に出力との組み合わせを入れ替えた。実験の結果を表 6 に示す。

得られた表 6 の結果に対し、有意水準 5% の t 検定を行った結果、両出力の間には正答した質問数の差がないという帰無仮説が棄却され、構造化を行った議論マップによる出力が有意に有効であることが示された。

5. まとめ

本稿では、会議の流れや決定事項などの要旨を素早く把握す

ることを可能とする、発話間関係の構造化による会議録からの議論マップ自動生成システムを提案した。提案システムは、1) 会議録からの重要発話抽出を行い、2) 発話間における質疑応答の対応付けを行った後に、3) 発話間の構造を保持した議論マップとして出力を行う。これらの処理により、1) 発話間の関係が構造化されていない、2) 確認するべき発話が限定されていない、という従来の問題点を解決した。

重要発話抽出に関する評価では、発話間における質疑応答の対応付けを行うことで、既存手法と比べ ROUGE-1 のスコアを最大 0.111 上げることに成功した。また、4 名の被験者実験を行った結果、重要発話を構造化せずに表示した時と比べ、構造化を行った議論マップとして出力を行った時の方が素早く会議内容を把握できるということが検証され、システムの有用性が示された。

今後の課題としては、システムの完全自動化に向けて、トピックセグメントの自動推定などを行い、現在用いている AMI Meeting Corpus が提供するアノテーション情報の自動付与に取り組む必要がある。また、質を向上させるため、重要発話抽出と、現在はルールベースで行っている質疑応答の対応付けの精度向上を行い、より多くの被験者による実験を行う必要がある。

文 献

- [1] Adam Janin, Don Baron, Jane Edwards, Dan Ellis, David Gelbart, Nelson Morgan, Barbara Peskin, Thilo Pfau, Elizabeth Shriberg, Andreas Stolcke, and others: “The ICSI meeting corpus”, in *Proc. of ICASSP*, pp. 364–367, 2003.
- [2] Jean Carletta, Simone Ashby, and Sebastien Bourban et al.: “The AMI meeting corpus: A pre-announcement”, in *Proc. of MLMI*, pp. 28–39, 2005.
- [3] Shasha Xie, Benoit Favre, Dilek Hakkani-Tr, and Yang Liu: “Leveraging sentence weights in a concept-based optimization framework for extractive meeting summarization”, in *Proc. of INTERSPEECH*, pp. 1503–1506, 2009.
- [4] Dan Gillick, Korbinian Reidhammer, Benoit Favre, and Dilek Hakkani-Tür: “A global optimization framework for meeting summarization”, in *Proc. of ICASSP*, pp. 4769–4772, 2009.
- [5] Korbinian Riedhammer, Benoit Favre, and Dilek Hakkani-Tr: “Long story short Global unsupervised models for keyphrase based meeting summarization”, *Speech Communication*, Vol. 52, No. 10, pp. 801–815, 2010.
- [6] Nikhil Garg, Benoit Favre, Korbinian Reidhammer, and Dilek Hakkani-Tür: “ClusterRank: a graph based method for meeting summarization”, in *Proc. of Annual Conf. of the Inter-Speech*, pp. 1499–1502, 2009.
- [7] Fei Liu and Yang Liu: “Using spoken utterance compression for meeting summarization: A pilot study”, in *Proc. of IEEE SLT*, pp. 37–42, 2010.
- [8] Shasha Xie and Yang Liu: “Improving supervised learning for meeting summarization using sampling and regression”, *Computer Speech & Language*, Vol. 24, No. 3, pp. 495–514, 2010.
- [9] Andreas Kathol and Gokhan Tur: “Extracting question/answer pairs in multi-party meetings”, in *Proc. of ICASSP*, pp. 5053–5056, 2008.
- [10] Michel Galley: “A skip-chain conditional random field for ranking meeting utterances by importance”, in *Proc. of EMNLP*, pp. 364–372, 2006.
- [11] Fei Liu and Yang Liu: “Towards abstractive speech summarization: Exploring unsupervised and supervised approaches for spoken utterance compression”, *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 21, No. 7, pp. 1469–1480, 2013.
- [12] Yashar Mehdad, Giuseppe Carenini, Frank W. Tompa, and Raymond T. NG: “Abstractive meeting summarization with entailment and fusion”, in *Proc. of ENLIG*, pp. 136–146, 2013.
- [13] Gabriel Murray: “Abstractive Meeting Summarization as a Markov Decision Process”, *Advances in Artificial Intelligence*, Springer International Publishing, pp. 212–219, 2015.
- [14] M. Abdeen and M. C.E. Yagoub: “Direct Automatic Generation of Mind Maps from text with M2 Gen”, in *Proc. of TIC-STH*, 2009 IEEE Toronto International Conference. IEEE, 2009.
- [15] Robert Kudeli, Mladen Koneck, and Mirko Malekovi: “Mind map generator software model with text mining algorithm”, in *Proc. of ICITI*, 2011.
- [16] Au Purwarianti, Atria Saelan, Irfan Afif, Filman Freudian, and Alfian Farizki Wicaksono: “Natural Language Understanding Tools with Low Language Resource in Building Automatic Indonesian Mind Map Generator”, *International Journal on Electrical Engineering and Informatics (IJEI)*, Vol.5, No.3, 2013.
- [17] Mohamed Elhoseiny, Ahmed Elgammal: “Text to multi-level MindMaps”, *Multimedia Tools and Applications*, Springer, pp. 1–28, 2015.
- [18] 松村真宏, 加藤優, 大澤幸生, 石塚満: “議論構造の可視化による論点の発見と理解”, *日本ファジィ学会誌*, Vol. 15, No. 5, pp. 554–564, 2003.
- [19] 趙雲超, 松村真宏, 谷内田正彦: “音声認識された議事録からの議論マップ自動生成”, *人工知能学会全国大会論文集*, Vol. 6, pp. 221–221, 2006.
- [20] 友部博教, 土田貴裕, 大平茂輝, 長尾確: “ディスカッションメディア: 会議コンテンツの構造化と効率的な閲覧システム”, *人工知能学会第 21 回全国大会論文集*, No. 2F3-5, 2007.
- [21] 森幹彦, 八村太輔, 喜多一: “リフレクションのための逐語議事録を用いた議論の構造化法”, *人工知能学会第 21 回全国大会論文集*, No. 2D4-1, 2007.
- [22] Koby Crammer, Alex Kulesza, and Mark Dredze: “Adaptive regularization of weight vectors”, in *Proc. of Advances in Neural Information Processing Systems*, NIPS’09, pp.414–422, 2009.
- [23] Mark Dredze, Koby Crammer, and Fernando Pereira: “Confidence-weighted linear classification”, in *Proc. of the 25th international conference on Machine learning*, ICML’08, pp.264–271, 2008.
- [24] 林佑磨, 諏訪晴士, 山名早人: “非テキスト情報のみを用いた AROW による効率的な CTR 予測モデルの構築”, 第 7 回データ工学と情報マネジメントに関するフォーラム (DEIM2015), 2015.
- [25] Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky: “The Stanford CoreNLP Natural Language Processing Toolkit”, in *Proc. of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pp. 55–60, 2014.
- [26] Chin-Yew Lin: “Rouge: A package for automatic evaluation of summaries”, in *Proc. of Workshop on Text Summarization Branches Out (WAS)*, pp25–26, 2004.
- [27] Jaime Carbonell and Jade Goldstein: “The use of MMR, diversity-based reranking for reordering documents and producing summaries”, in *Proc. of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 335–336, 1998.
- [28] Rada Mihalcea and Paul Tarau: “TextRank: Bringing order into texts”, in *Proc. of EMNLP*, pp. 404–411, 2004.