

オンライン会議における議論の要点と対話の雰囲気認識技術の開発

後藤啓介^{1,a)} 新美翔太郎¹ 荒川智哉¹ 西田典起² 松本裕治² 廣島雅人¹

概要: オンラインで行われる会議では視聴覚情報の制限や遅延により議論の流れや雰囲気を把握しづらいという課題がある。この課題解決のために、著者らはオンライン会議における議論の要点および対話の雰囲気認識技術の開発に取り組んでいる。本稿では、オンライン会議の対話データセットの構築および対話の雰囲気のアノテーション、ならびに、非言語情報を考慮した議論の要点抽出および発話の感情分析をベースとした対話の雰囲気認識の各技術についての検討内容を報告する。

1. はじめに

情報通信技術の発展により、近年ビデオ通話アプリ等を用いて行われるオンラインの会議の利用機会が増加している。会議において議論の流れや対話の雰囲気を参加者間で共有することは、会議の円滑な進行や建設的な議論を行う上で重要である。しかし、オンライン会議では通信システムを介していることで生じる視聴覚情報の制限や遅延が、会議参加者の内容理解や共感を妨げてしまう。議事録などの会議の記録は議論の流れの理解に役立つが、一般に会議が終わった後に作成されるものであり会議中に確認することはできない。また、雰囲気の情報は明示的に記録されることは少なく、録画された会議映像や書き起こされた会話を確認することで間接的に知ることしかできない。これらのオンライン会議が持つ課題を解決するため、著者らは会議参加者の内容理解と共感を支援するシステムの実現を目的としたオンライン会議の対話解析技術の開発に取り組んでいる。

議論の流れを捉える技術として対面での会議を対象とした自動要約の研究が多数行われている[5, 9, 10, 17]。会議を対象とした自動要約では、会議中の全発話に対する要約ではなく、その会議で扱われた各話題に対する対話内容の要約が求められることも多い[5, 9]。また、会議参加者が互いの話し方や身振り手振りを見て発話の内容や意図を補うことが知られており、言語情報に加えて視覚や聴覚から取得可能な非言語情報の活用が重要な要素として扱われている[10, 17]。対話の雰囲気は良し悪しや盛り上がりの程度といった複数の観点を持つが、特定の観点に着目した雰囲気のアノテーションや推定に関する研究が行われている[6, 7, 12, 14]。また、会議における高次の事象である雰囲気について会議中の発話で言及されることは稀であるため、単語などの表層的な言語情報から直接雰囲気を捉えることは難しい。既存研究においては、発話者の感情や脳波といった観測が容易でない心理的または生理的な情報を用いて雰囲

気を推定する方法が検討されている[6, 12]。さらに、発話者の感情の推定においてはマルチモーダル感情分析の研究が存在する[11, 16]。しかし、マルチモーダル情報からの雰囲気推定はこれまで行われていない。雰囲気は話し合いの場（すなわち発話の系列）に関する情報であり感情に近いものの、各発話の感情の平均といった単純な関係では表現できない[6]。

本稿では、オンライン会議の対話データセットの構築および対話の雰囲気のアノテーション、ならびに、提案手法である非言語情報を考慮した議論の要点抽出および発話の感情分析をベースとした対話の雰囲気認識の各技術についての検討内容を報告する。既存の対話データセットには対面で行われた会議を扱ったものしか存在しないため、オンライン会議の映像および音声の収録から行う必要がある。また、収録したデータに対して、言語化される機会の少ない事象である対話の雰囲気をアノテーションする。

まず、オンライン会議の対話データセットの構築では、会議を行っているアプリケーションの画面の録画録音と共に、各参加者の非言語情報として、各参加者の使用端末からカメラ画像およびマイク音声を収集する。次に、対話の雰囲気のアノテーションでは、感情のアノテーションにおける評価者の感性に基づく評定値の付与を行う方法をベースとして、対話の雰囲気に関する複数の評価尺度を定め、それらの評定値を対象の会議シーンを視聴した第三者が付与する。次に、議論の要点抽出では、オンライン会議の各参加者が使用するカメラ画像およびマイク音声の非言語情報を考慮した発話の重要度推定処理により、話題ごとの要約精度の向上を試みる。最後に、対話の雰囲気認識では、議論の要点抽出と同様に非言語情報を考慮した認識処理を行う。マルチモーダル感情分析を対象としたネットワーク構造をベースとして、発話単位の話者感情の代わりに、発話の系列情報を入力とした雰囲気認識をモデルに学習させる。

以下では、まず2章において関連研究を紹介する。3章

1 京セラ株式会社
KYOCERA Corporation
2 理化学研究所 革新知能統合研究センター

RIKEN Center for Advanced Intelligence Project
a) keisuke.goto.fj@kyocera.jp

においてオンライン会議の対話データセットの構築および雰囲気のアノテーションについて説明し、4章において議論の要点抽出および対話の雰囲気認識についての検討内容を報告する。最後に5章においてまとめと今後の予定を記す。

2. 関連研究

2.1 会議を対象とした自動要約

Kimura ら (2020) [5]では、地方議会における一括質問一括答弁形式の対話とその要約を収録したデータセット（以下 NTCIR15 要約タスク）を構築している。Kadowaki ら (2020) [4]は、NTCIR15 要約タスクに対して重要文抽出と要約テキスト生成の2ブロックからなる手法を提案している。同手法は、それらのブロックにおいてそれぞれ事前学習モデルである BERT [1]とその派生手法である UniLM [2]を再事前学習させたモデルを使用することにより、同タスクにおける最高精度を達成している。4.1 節で述べる要点抽出の手法は、Kadowaki ら (2020) [4]の手法をベースとし、文抽出のブロックにおいて非言語情報および対話文脈の考慮を行う処理を追加することで精度向上を試みる。NTCIR15 要約タスクは、オフラインの会議である点、言語情報のみを扱っている点および一括質問一括答弁という地方議会ならではの会議形式である点において、本研究で構築するデータセットとは異なっている。

中山ら (2021) [9]では、4名の参加者による架空のショッピングモールに関する対話を収集した Kyutech コーパス [15]に対して、話題ごとの人手要約および重要発話のアノテーションを行っている。Zhong ら (2021) [17]では、製品デザインに関する会議や研究機関の週次ミーティングの音声や画像、発話文を収集した既存のコーパスを対象として、話題の分割および事前に用意した質問クエリに対する回答となる要約文の付与を行っている。これらの研究で作成されたデータセットは、NTCIR15 要約タスクと同様にオフラインの会議を収録したものである点において、本研究で構築するデータセットとは異なっている。

Nihei ら (2017) [10]では、会議参加者の音声および頭部の動きを考慮した重要発話の検出手法を提案している。この手法では、畳み込みニューラルネットワークを用いて個々のモーダルデータから特徴ベクトルを抽出している。

4.1 節で述べる要点抽出の手法は、画像および音声のモダリティ情報からの特徴抽出において Nihei ら (2017) [10]と同様に畳み込みニューラルネットワークを使用している。一方、提案手法では、重要発話か否かを認識する Nihei ら (2017) [10]の手法とは異なり、要約元と要約文の類似度に基づいて算出される重要度の値の推定をする。これにより、相対的に重要度の高い発話を優先して抽出することが可能となっている。

2.2 会議を対象とした感情や雰囲気認識

森ら (2005) [8]では、「怒り」や「悲しみ」といった言葉で定義される感情の意味は狭く、自由な会話において発話に含まれることはまれであると指摘している。そこで、パラ言語情報を記述することを目的として、話者自身の感情に加えて、対人関係および態度に関する複数の評価項目を提案すると共に、これらの評価項目についてのアノテーションを行った対話コーパスである UUDB を公開している。近藤ら (2016) [6]では、UUDB に対して雰囲気のアノテーションおよび感情強度からの悪い雰囲気の検出手法を提案している。雰囲気のアノテーションでは、5 発話単位の対話シーンについて悪い雰囲気であるか否かのラベル付けを行っている。悪い雰囲気の検出においては、UUDB に含まれているアノテーション済みの感情強度を説明変数とするポワソン回帰分析を適用した手法を提案している。悪い雰囲気の検出において感情強度が有用であることが示されている。そのため、マルチモーダル情報からの雰囲気認識においては、発話者の感情に関わる特徴量の抽出に優れたネットワーク構造が適していると考えられる。4.2 節で述べる雰囲気認識手法は、そのネットワーク構造をベースとして、マルチモーダル感情分析の手法のものを採用し、発話の系列情報を入力とした雰囲気認識をモデルに学習させる。

白松ら (2017) [12]では、対話の盛り上がりや傾聴の度合いなどの属性からなる対話の雰囲気のモデルおよび脳波、音声、視線および表情の測定値を用いたこれらの属性値の推定手法を提案している。徳久ら (2006) [14]では、雑談における対話の盛り上がりや話題ごとに判定するアノテーション方法を提案すると共に、雑談を対象にした実験において、感情に関する発話が対話の盛り上がりと関連が深い結果が得られたことを報告している。水上ら (2018) [7]では、グループディスカッションを対象として、意見の協調や対立を代表的な相と定めると共に、意見が出尽くしたにもかかわらず結論に至らずにすぶっている状態を停滞相と名付け、その周辺文脈について考察している。これらの研究で扱っている雰囲気は、対面での会議を対象としたものであり、オンラインでの会議の対話の雰囲気認識を目的とする本研究とは異なる。

3. オンライン会議の対話データセット構築

対面で行われた会議を対象としたデータセットは複数存在するが [5, 17]、オンラインで行われた会議の発話を扱ったデータセットはない。そこで、我々はまずオンライン会議の対話データセット構築に取り組む。

まず、オンライン会議の映像および音声の収録を行う。続いて、音声からの発話文の書き起こしを行う。最後に、

要約の作成および雰囲気に関する評定値の付与を行う。

3.1 オンライン会議の収録と発話文の書き起こし

会議は大学や企業など多くの組織において日常的に行われているが、個人情報および秘密情報の保護の観点から収録の対象とすることは難しい。本研究では、会議の参加者として収録に協力してもらう被験者を集めて、企業の採用選考において広く用いられているグループディスカッションを Web 会議アプリによって遠隔で行い、その会議の音声付き映像の録画、ならびに、各参加者が使用しているカメラ画像の録画およびマイク音声の録音を行う。収録データの主な用途としては、会議の音声付き映像は発話文の書き起こしおよびアノテーション、各参加者のカメラ画像およびマイク音声は議論の要約抽出および対話の雰囲気認識における入力情報としてそれぞれ利用する。

グループディスカッションは、複数人の参加者からなるグループが協力して取り組む課題であり、与えられた議題について参加者間で意見を出し合って議論し、決められた時間内でグループとしての結論を出す対話行為である。グループディスカッションの参加者の職業や性格といった属性情報とグループディスカッションの議題は、認識対象である議論の複雑さや雰囲気に影響を及ぼす要素である。しかしながら、どのような属性情報を持った参加者のグループが、どんな議題に取り組めば、どのような議論や雰囲気となるかは分かっておらず、収録したい議論の複雑さや雰囲気がある場合に、これらの条件を適切に設定することは現実的ではない。むしろ、事前準備として、参加者に議論の材料となる資料を渡して読み込んでもらうことや、特定の参加者に雰囲気を整えるあるいは乱す行為を指示しておくことが、有効と考えられる。

3.2 要約の作成

Yamamura ら (2016) [15] や Zhong ら (2021) [17] のように、対面の会議に対しては話題単位で要約が付与されたコーパスが既に公開されている。本研究では、これらの研究と同様に、オンライン会議の対話を対象とした話題単位の要約を作成する。具体的には、音声から書き起こした会話テキストに対して要約テキストを作成する際に、会話テキストに含まれる話題ごとに小見出しと要約テキストを作成する。また、各話題に対して、その要約の根拠となる会話テキスト上の範囲を示す。話題の範囲は発話単位として示されるものとし、要約の作成者が会話テキストから判断する。圧縮率は 10% 以下を想定し、要約の作成者は話題ごとの結論を必要最小限の情報として、圧縮率に応じて、主要なやり取りや話者の情報を含む内容で要約テキストを作成する。

3.3 対話の雰囲気に関する評定値の付与

会議において対話の雰囲気の情報に参加者間で共有す

映像および会話文から推察される総合的な会議の雰囲気について質問します。以下の点について、それぞれ用意された項目から最も近いと思われるものを選択してください。①と②のどちらも該当しない場合は「中間」を選択してください。*

	とても①	①	やや①	中間	やや②	②	とても②
①協調⇔②対立：交わされている意見の関係	<input type="radio"/>						
①進捗⇔②停滞：議論の進み具合	<input type="radio"/>						
①シリアス⇔②談笑：空気の硬さや真面目さ	<input type="radio"/>						
①良い⇔②悪い：会議の目的が達成できそうか	<input type="radio"/>						
①満足⇔②不満：不満のある発話者はいないか	<input type="radio"/>						

図 1 対話の雰囲気についての評価シート

ることは、会議の円滑な進行や建設的な議論を行うために重要な要素である。しかし、一般に議事録などの会議録には雰囲気の情報が明示的に記録されることはなく、録画された会議映像の視聴や、書き起こされた会話文を確認することで間接的にしか知ることはできない。本研究では、このような言語化される機会の少ない事象である対話の雰囲気についてのアノテーションをオンライン会議の対話に対して行う。

対話の雰囲気のアノテーションは、感情のアノテーションにおける評価者の感性に基づく評定値の付与を行う方法をベースとする [3, 13, 16]。具体的には、対話の雰囲気を描写する表現の対（これを形容詞対と呼ぶ）と、評定値として選択可能な段階とを用意し、対象の会議シーンに対して、複数の評価者がその音声付動画および会話文から読み取った値を付与する。図 1 に対話の雰囲気についての評価シートを示す。形容詞対は、対話に含まれる意見間の関係性を評価する「協調-対立」、会議の成否に関わる状況性を評価する「進捗-停滞」、「シリアス-談笑」および「良い-悪い」ならびに会議参加者の心理状態を客観的に評価する「満足-不満」の 5 種類を用意した。選択可能な段階は対となる表現を各 3 段階およびそれらの中間の 1 段階を合わせた 7 段階とした。

評価者は、収録した会議の一部の時間帯（例えば 3 分）についての映像および会話文を評価対象の会議シーンとしてそれらの内容を確認する。そして、評価シートに用意された対話の雰囲気に関する複数の評価尺度のそれぞれについて、確認した会議シーンから読み取れた状態を用意された複数の段階からなる選択肢から 1 つ選ぶ。以上の作業を同一の会議シーンに対して複数人が行い、得られた評定値の平均をその会議シーンの雰囲気に関する属性値として利用する。

4. オンライン会議の対話解析

4.1 議論の要点抽出

本研究では、話題単位の会議の対話要約タスクにおいて優れた精度を達成した Kadowaki (2020) [4] の手法をベース

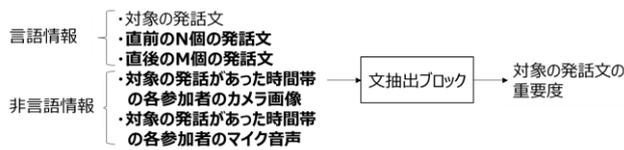


図 2 文抽出ブロックの入出力情報

として、非言語情報の考慮と対話文脈を考慮した発話の重要度推定による要約精度の向上を検討する。

ベースとなる Kadowaki (2020) [4]の手法は、文抽出と要約生成の2ブロックで構成されており、それぞれ BERT およびその派生手法である UniLM の事前学習済みモデルを使用している。文抽出のブロックでは、入力された発話文と要約テキストとの類似度の数値を推定する。訓練サンプルについては、要約元の発話集合と要約テキストの対が与えられるため、この類似度の推定タスクでモデルを訓練する。訓練済みのモデルが推定した類似度は、入力された発話文の重要度として扱い、要約生成のブロックの最大入力長を超えない範囲で、重要度の高い順に発話文を連結して要約生成のブロックへと入力する。要約生成のブロックでは、入力された発話文の系列を要約元とする要約テキストを生成する。訓練時には、発話文と要約テキストの類似度の真の値を用いて、要約生成のブロックの最大入力長を超えない範囲で類似度の高い順に発話文を連結したものを入力データとしてモデルを訓練する。

非言語情報の考慮と対話文脈を考慮した発話の重要度推定は、文抽出ブロックの入力として扱う情報を増やすことで対応する。具体的には、入力する発話ごとに、その発話時間中の会議参加者のカメラ画像およびマイク音声を同時に入力すると共に、言語情報についても対象の発話に加えて対話の時系列上の周辺に位置する発話を入力する (図 2)。非言語情報からの特徴抽出および複数モダリティの特徴ベクトルの統合処理は、Nihei ら(2017) [10]で提案されている方法を採用する。具体的には、前処理としてカメラ画像からは頭部の動き、マイク音声からはインテンションおよびスペクトルグラムに関する波形データを取得し、それらを畳み込みニューラルネットワークを用いて画像的に処理することで、モダリティごとの重要度に関する特徴ベクトルへと変換する。複数モダリティの特徴ベクトルの統合は全結合層により実現する。

以上のような構成の議論の要点抽出の手法をオンライン会議の対話データセットを用いて評価することにより、ベースとなる Kadowaki (2020) [4]の手法と比較した精度改善の確認と、追加した各モダリティ情報の有効性について対面会議との違いの検証を予定している。

4.2 雰囲気認識

本研究では、近藤ら (2016) [6]の感情強度からの対話の

雰囲気推定を参考に、発話者の感情に関わる特徴量の抽出に優れたネットワーク構造を採用する。具体的には、マルチモーダル感情分析タスクにおいて優れた精度を達成した Multilogue-Net [11]をベースとして、複数人の複数発話に対する対話の雰囲気認識を検討する。

ベースとなる Multilogue-Net は、対象の発話の話し手に関する複数のモダリティ情報を入力として、対話の開始から終了までの発話の系列データに対して、それらの発話に対応する話し手の感情を推測する。複数人の話し手を想定しており、複数の再帰型ニューラルネットワークを用いて、話者ごとに隠れ状態を更新する。しかし、発話に対する感情の推定を想定しているため、感情の推定においては発話に対応する話し手1名の隠れ状態しか計算に使用しておらず、聞き手のモダリティ情報は考慮していない。

そこで、対話の雰囲気推定においては、このベース手法に2点の変更を加える。1点目は入力データ系列として発話に加えて発話間の間を扱うようにする。また、同時刻のデータに対して、会議の参加者全員のモダリティ情報を扱うようにし、再帰ニューラルネットワークにおいては、すべての参加者の隠れ状態を同時に更新する。これにより、非発話者の相槌や頷きの有無を考慮できるようになると共に、オンライン会議においてしばしば発生する衝突した発話や不意の沈黙に対しても参加者の隠れ状態を更新することが可能になる。2点目は出力を発話ごとの感情に関する推定値から対話全体の雰囲気に変更し、その算出においては、各参加者に対応する隠れ状態のすべてを使用することである。これにより、特定の発話者ではなく、すべての会議参加者の隠れ状態を用いて、入力された時間帯における対話の雰囲気の推定を行うことが可能となる。

5. おわりに

本稿では、オンライン会議の対話データセットの構築および対話の雰囲気のアノテーション、ならびに、提案手法である非言語情報を考慮した議論の要点抽出および発話の感情分析をベースとした対話の雰囲気認識の各技術についての検討内容を報告した。

データセットの構築では、グループディスカッションを Web 会議アプリによって遠隔で行い、その会議の音声付き映像の録画、ならびに、各参加者が使用しているカメラ映像の録画およびマイク音声の録音を行う。収録データは、会議の音声付き映像は発話文の書き起こしおよびアノテーション、各参加者のカメラ画像およびマイク音声は議論の要点抽出および対話の雰囲気認識における入力情報としてそれぞれ利用する。対話の雰囲気のアノテーションでは、対象の会議シーンの映像および会話文を確認した評価者が、対話の雰囲気に関する複数の評価尺度のそれぞれについて、用意された複数の段階から適当な値を選択することにより

評定値を付与する。

議論の要点抽出では、話題単位の会議の対話要約タスクを想定した手法をベースとして、非言語情報の考慮と対話文脈を考慮した発話の重要度推定の変更を加える。対話の雰囲気認識では、マルチモーダル感情分析タスクを想定した手法をベースとして、非発話者のモダリティ情報を考慮すると共に各参加者に対応する隠れ状態をすべて用いて入力された時間帯における対話の雰囲気の推定を行う。

今後は、これらの検討内容に基づくオンライン会議のデータセットの構築およびアノテーションを行うと共に、議論の要点抽出および対話の雰囲気認識の各手法の実装および性能評価を実施することを予定している。オンラインでは伝わりにくい発話者の意図や感情を考慮した議論の要点や対話の雰囲気をリアルタイムに可視化することで、会議参加者の内容理解と共感を支援するシステムの実現を目指す。

参考文献

- [1] Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, *In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp. 4171–4186 (2019).
- [2] Dong, L., Yang, N., Wang, W., Wei, F., Liu, X., Wang, Y., Gao, J., Zhou, M. and Hon, H.-W.: Unified Language Model Pre-training for Natural Language Understanding and Generation, *In Proceedings of the 33rd International Conference on Neural Information Processing Systems*, pp. 13063–13075 (2019).
- [3] Kajiwar, T., Chu, C., Takemura, N., Nakashima, Y. and Nagahara H.: WRIME: A New Dataset for Emotional Intensity Estimation with Subjective and Objective Annotations, *In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 2095–2104 (2021).
- [4] Kadowaki K.: JRIRD at the NTCIR-15 QA Lab-PoliInfo-2 Task: An Abstractive Dialog Summarization System for Japanese Assembly Minutes, *In Proceedings of the 15th NTCIR Conference on Evaluation of Information Access Technologies* (2020).
- [5] Kimura, Y., Shibuki, H., Ototake, H., Uchida, Y., Takamaru, K., Ishioroshi, M., Mitamura, T., Yoshioka, M., Akiba, T., Ogawa, Y., Sasaki, M., Yokote, K., Mori, T., Araki, K., Sekine, S. and Kando, N.: Overview of the NTCIR-15 QA Lab-PoliInfo-2 Task, *In Proceedings of the 15th NTCIR Conference on Evaluation of Information Access Technologies* (2020).
- [6] 近藤公久, 釜島 萌: 対話者の感情状態による場の空気の推定—対話音声コーパスを用いた検討—, *日本感性工学会論文誌*, Vol. 15, No. 2, pp. 279–285 (2016).
- [7] 水上悦雄, 劉 礫岩, 森本郁代: 話し合いの停滞期境界における参加者の振舞いの分析—話し合いの相移行期の考察(2)—, 第 82 回 言語・音声理解と対話処理研究会, p. 10- (2018).
- [8] 森 大毅, 相澤 宏, 粕谷英樹: 対話音声のパラ言語情報ラベリングの安定性, *日本音響学会誌*, Vol. 61, No. 12, pp. 690–697 (2005).
- [9] 中山友梨, 塩田 幸 and 嶋田和孝: 複数人対話におけるトピック単位の要約データの構築とその要約, *研究報告知能システム (ICS)*, Vol. 2021-ICS-203, No. 6, pp.1-6 (2021).
- [10] Nihei, F., Nakano, Y. I. and Takase Y.: Predicting meeting extracts in group discussions using multimodal convolutional neural networks, *In Proceedings of the 19th ACM International Conference on Multimodal Interaction*, pp. 421–425 (2017).
- [11] Shenoy, A. and Sardana A.: Multilogue-Net: A Context-Aware RNN for Multi-modal Emotion Detection and Sentiment Analysis in Conversation, *In Proceedings of the Second Grand-Challenge and Workshop on Multimodal Language*, pp. 19–28 (2020).
- [12] 白松 俊, 北川 晃, 幸浦弘昂, 熊崎滉大: 議論参加者の行動センシングに基づく場の空気の自動推定に向けた検討, *人工知能学会第二種研究会資料*, Vol. 2017, No. CCI-003, p. 09- (2017).
- [13] Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C. D., Ng, A. and Potts C.: Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank, *In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pp. 1631–1642 (2013).
- [14] 徳久良子, 寺嶋立太: 雑談における発話のやりとりと盛り上がりに関連, *人工知能学会論文誌*, Vol. 21, No. 2, pp. 133–142 (2006).
- [15] Yamamura, T., Shimada, K., and Kawahara, S.: The Kyutech Corpus and Topic Segmentation Using a Combined Method, *In Proceedings of the 12th Workshop on Asian Language Resources*, pp. 95–104 (2016).
- [16] Zadeh, A. B., Liang, P. P., Poria, S., Cambria, E. and Morency L.-P.: Multimodal Language Analysis in the Wild: CMU-MOSEI Dataset and Interpretable Dynamic Fusion Graph, *In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 2236–2246 (2018).
- [17] Zhong, M., Yin, D., Yu, T., Zaidi, A., Mutuma, M., Jha, R. Awadallah, A. H., Celikyilmaz, A. Liu, Y., Qiu X. and Radev, D.: QMSum: A New Benchmark for Query-based Multi-domain Meeting Summarization, *In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (2021).