

時間情報とテキストセグメンテーションに基づく複数人対話要約手法

山村 崇[†] 徳永 陽[†] 嶋田 和孝[†]

[†]九州工業大学大学院 情報工学府 先端情報工学専攻

〒820-8502 福岡県飯塚市川津 680-4

E-mail: †{t.yamamura,y.tokunaga,shimada}@pluto.ai.kyutech.ac.jp

あらまし 本論文では、複数人対話を対象とした重要文の抜粋による要約手法について提案する。複数人による対話では、話者が自由に発言し、トピックの切り替わりが多く存在するなどの特徴が存在する。このような対話文の要約を行なう際には、これらの要素を考慮することが重要である。本研究ではこれらの特徴を捉えるため、発話速度などの時間情報や対話文をトピック毎に分割するテキストセグメンテーションを導入し、その有効性を検証する。

キーワード 複数人対話, 対話要約, 時間情報, テキストセグメンテーション

Multi-party conversation summarization using time information and text segmentation

Takashi YAMAMURA[†], Yo TOKUNAGA[†], and Kazutaka SHIMADA[†]

[†] Department of Artificial Intelligence, Kyushu Institute of Technology

Kawazu 680-4, Iizuka-shi, Fukuoka, 820-8502 Japan

E-mail: †{t.yamamura,y.tokunaga,shimada}@pluto.ai.kyutech.ac.jp

Abstract In this paper, we propose a method to summarize a multi-party conversation. The method consists of two processes; important utterance extraction and summary generation using the extracted utterances. For the extraction process, we apply time features as new nonverbal information. For the generation process, we utilize a text segmentation method, LCSEg, to improve the coverage of the summary. The experimental result shows the effectiveness of the time features in the extraction process.

Key words Multi-party conversation, conversation summarization, time information, text segmentation

1. ま え が き

テキスト要約は自然言語処理における重要な研究課題の一つである [1]。これまでの研究は新聞記事や論文のような書き言葉を対象としたものが多かった。一方で、会議や雑談のような複数人による対話に関する要約の重要性も増している。このような話し言葉の要約では、書き言葉とは異なり、発話の自由度が高いことなどから、従来の書き言葉の手法をそのまま適用することは困難である。また、対話に含まれる非言語情報を考慮することも必要である。

徳永ら [2] は、複数人自由対話を構成する各発話の重要性や発話同士の関係性などの言語情報や、笑いなどの非言語情報を考慮し、発話の抜粋によるテキスト要約手法を提案している。本論文の手法もこれに倣い、重要文抽出部と要約生成部の2段階による抜粋型の要約手法をとる。重要文抽出部では、要約に必要な文を高い精度で抽出することが重要である。また、要約

生成部では、重要文抽出の際に算出された重要度に基づき、元の対話の文意を崩さないように、読みやすい文を生成することが重要である。

本論文では、先行研究が触れていない他の非言語情報を考慮することで、重要文抽出部における精度の改善を目指す。例えば、話者の発話タイミングや発話速度、発話オーバーラップが話者に依存するというような情報も、自由対話特有の非言語情報である。また、先行研究の要約生成手法の問題点を挙げ、その解決方法を提案し、有効性を検証する。

2. 先行研究

まず、基本となる徳永らの手法について説明する。徳永らは、対話文を構成する各発話の重要性や発話同士の関係性を考慮し、発話の抜粋による要約を行なっている。図1に、徳永らの要約手法の全体図を示す。図1のように、まず対話文から、スコアによる分類器と機械学習による分類器を用いて、要約に必要な

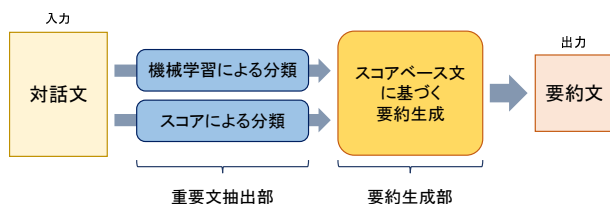


図 1 先行研究の全体像

重要文を抽出する。その後、スコアによる分類器によって抽出された重要文を基盤として、重要文と関連が強い発話を組み合わせることで、要約文を生成している。

2.1 重要文抽出

スコアによる分類器では、スコアベース文と呼ばれる対話の要点となる重要文を高い適合率で抽出することを目的としている。スコアベース文を抽出するために、スコアによる分類器では、キーワード抽出に基づく重要文抽出の手法の1つである展望台システム [3] を参考にして、スコアリングを行なう。各発話をスコアリングし、そのスコアを基に上位 10% の発話をスコアベース文として抽出する。

機械学習による分類器では、発話単体や発話間の関係性に関する特徴などの言語情報と、笑いと盛り上がりなどの非言語情報の素性の計 23 個の素性を用いて重要文を抽出している。これらの素性は、大別すると言語情報に関する 4 つの素性群と、非言語情報に関する 1 つの素性群に分類される。

素性群 A: 発話単体の特徴に関する素性

(発話の長さ、高頻度単語の有無など)

素性群 B: 発話間の関係性に関する素性

(直前の発話が疑問文か、発話者の連続性など)

素性群 C: 照応性に関する素性

(指示表現、接続表現の有無など)

素性群 D: キーワード評価値に関する素性

(展望台システムにおける 3 つの評価値)

素性群 E: 非言語情報に関する素性

(笑いの有無、盛り上がり度など)

最後に、以上の 23 個の素性を機械学習に適用し重要文の抽出を行なう。機械学習のアルゴリズムには SVM [4] を用いて、機械学習による分類器から得られた出力結果を、各発話の重要文の信頼値として利用している。

2.2 要約生成

先行研究では、スコアベース文を主軸に対話の流れを意識した要約を生成するために、以下の 3 つの手順を踏んで要約を生成している。

まず、スコアベース文を要約文として抽出する。しかし、スコアベース文は対話文全体の 10% の抽出であるため、これだけでは要約として不十分である。そこで、スコアベース文の前後文も要約として抽出する。話者が頻繁に変わり、対話の流れが連続的である自由対話において、前後文の有無が要約の読みやすさや内容の理解に大きく影響する。ただし、スコアベース文の前後文には、明らかに要約として不適切な相槌などの文が

含まれる場合がある。そこで、スコアベース文の各前後文を抽出する際に、発話ごとの機械学習による分類器の出力を信頼値として用いて選定を行なう。これは、相槌などの要約に不適切な発話は、機械学習による分類器の出力が低くなりやすいためである。最後に、対話の流れをより自然にするために、スコアベース文間に機械学習の分類器による信頼値が最も高い 1 文を要約として追加する。

2.3 問題点

先行研究の重要文抽出と要約生成の手法には、それぞれに問題点がある。本節では、先行研究の重要文抽出と要約生成の問題点について説明する。

2.3.1 重要文抽出の問題点

2.1 節の機械学習による分類器の素性では、非言語情報である盛り上がりと笑いを自由対話特有の特徴の 1 つと捉え、これらを非言語情報の素性として用いている。しかし、これらの非言語情報は、素性として有効に機能しないという問題が実験結果より確認されている。これは、実験データに対して笑いを含む発話が少なかったことなどの、データに対してばらつきがある素性であったことが問題点として挙げられる。

2.3.2 要約生成の問題点

要約生成の手法には、大きく分けて 2 つの問題点がある。

1 つ目の問題点として、先行研究の手法では要約率が不変であるという点が挙げられる。先行研究の手法では要約率を考慮しておらず、基本的にスコアベース文 (元のテキストの発話の 10%) とスコアベース文の各前後 1 文、そしてスコアベース間で 1 文を要約文として抽出しているため、要約率は約 33% で固定されている。元の対話の内容や長さによって、適切な要約率、要約の文数は異なるため、要約率は可変であることが望ましいと考えられる。

2 つ目の問題点として、先行研究の手法では対話中のトピック情報を考慮していないため、対話中の全てのトピックを網羅できていない可能性がある点が挙げられる。これは、スコアベースを起点にした要約しか生成できないことが原因であり、スコアベース文とその周辺が主に要約文として抽出される傾向にあるためである。逆にいえば、あるトピックの周辺にスコアベース文が存在しない場合は、そのトピックが要約文として抽出され難いため、結果的にそのトピックが損なわれる可能性があると考えられる。

3. 提案手法

本節では、2.3 節で説明した先行研究の問題点を解決するための手法を提案する。図 2 は、提案手法の要約手法の全体図である。図 2 のように、本研究では、重要文抽出部の改善として、機械学習の分類器に時間情報の素性を追加する。また、要約生成部の改善として、テキストセグメンテーションを用いた要約手法を提案する。

3.1 時間情報を追加した重要文抽出

2.3.1 節で説明した問題点を解決するために、先行研究が触れていないその他の非言語情報の素性として、発話時間情報を追加する。発話時間情報は自由対話特有の情報であり、かつ実

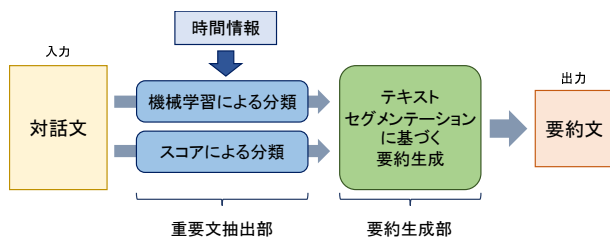


図 2 提案手法の全体像

験データに対してばらつきがない非言語情報であることから、先行研究の問題を解決できると考えられる。藤原ら [5] は、対話意図を考慮した対話リズムの分析を行っており、対話の応答において重要度の高い発話は、発話速度がゆっくりになり、発話タイミングは遅く、しばしば強調される傾向があると述べている。このようなことから、発話時間情報は有効に機能すると考えられる。

本論文では、発話速度と発話タイミング、発話オーバーラップの 3 つの素性を、発話時間に関する新たな非言語情報の素性群 F として追加する。この 3 つの素性について、以下で詳しく説明する。

素性群 F ： 発話時間情報に関する素性

- F_{1-1} ：発話速度
- F_{1-2} ：発話速度 (平仮名)

発話速度は発話の速さを表す。発話速度が遅い発話はゆっくり強調された発話が多いため、重要な発話が多いと考えられる。発話文の句読点を除く文字数から、その発話の発話時間の除算で得られる値を、発話速度の特微量として用いる。また、発話文は漢字で表記されているため、漢字のまま発話文の文字数を計算した素性を F_{1-1} 、漢字を平仮名に変換し文字数を計算したものを F_{1-2} とする。

- F_2 ：発話タイミング

発話タイミングは直前の発話との時間差を表す。発話タイミングが遅い発話は思考時間が長いと考えられることから、このような発話では重要な発話が多いと考えられる。ある発話の開始時間から、直前の発話の終了時間の差で得られる値を、発話タイミングの特微量として用いる。

- F_3 ：発話オーバーラップ

発話オーバーラップは、誰かが発話している最中に応答するような発話を表す。発話オーバーラップが発生する状況では、活発に話されていることが多いため、重要な発話が多いと考えられる。ある発話の開始時間と、それまでの発話の終了時間を比較し、発話オーバーラップが存在するかどうかの 2 値を特微量として用いる。

3.2 テキストセグメンテーションを用いた要約の生成

2.3.2 節で説明したように、先行研究の要約手法ではスコアベース文を主軸にした要約生成手法が原因で、要約率が不変であることや、対話中の全てのトピックに触れられていないという問題点が存在していた。そこで本研究では、スコアベース文を用いる代わりとして、テキストセグメンテーションを用いた要約生成手法を提案する。

テキストセグメンテーションでは、非構造的な文書を意味的なまとまりを表すトピック毎 (セグメント単位) に分割する。スコアベース文に依存せずに、各セグメントからセグメントの長さに応じて要約文を抽出することで、全体のトピックを網羅し、要約率を任意の値に変更できると考えられる。テキストセグメンテーションのプログラムには、Michel ら [6] の複数人会議対話における語彙的結束性に基づくテキストセグメンテーション手法 (以下、LCseg) を参考に実装した。LCseg では、文章の意味的に関連の深い部分には、同一の語が繰り返し出現するという語彙的連鎖 [7] を利用して、テキストセグメンテーションを行なう。

要約生成は、LCseg や 2.1 節の展望台システムのスコア、3.1 節の機械学習による分類結果の信頼値を用いて、以下のような手順で行なわれる。

ステップ 0： LCseg の適用

対話文書に対して LCseg を適用し、セグメント毎に分割する。

ステップ 1： 要約文数の決定

まず各セグメント毎に、要約文として抽出する文数の決定を行なう。本手法では、要約文数を任意の値で扱うために、各セグメントに対しセグメントの長さに応じて要約文数を決定する。具体的には、式 (1) で表されるように、あるセグメント i の長さを l_i とし、パラメータ β に任意の値を指定することで、そのセグメントの要約文数 S_i を決定する。ただし、 S_i は四捨五入して整数値にし、対話の流れや文意の流れを保つために $S_i = 1$ の場合 (セグメントから 1 文しか要約文に選ばれないとき) は、 $S_i = 2$ に補正を行なう。最終的な要約文数 S は、式 (2) で示されるように、各セグメントの要約文数の総和となる。

$$S_i = \beta \cdot l_i \quad (1)$$

$$S = \sum_i S_i \quad (2)$$

ステップ 2： セグメントの重要度の計算

明らかに重要文が少ないセグメントに対しては、次ステップである要約文の抽出を行なう必要がないと考えられる。このステップでは、各セグメントに対して重要度を算出し、明らかに重要文が少ないセグメントを識別することが目的となっている。そのため、展望台システムのスコアと機械学習による分類結果の信頼値を用いて、重要度の低いセグメントを識別し、残ったセグメントに対してのみ次ステップの処理を行なう。

- セグメントの長さが 5 文以内かつセグメント内の信頼値の最大が 0.55 未満、かつ展望台システムのスコアの平均が 0.4 未満のセグメントに対しては、要約文の抽出を行なわない

ステップ 3： 要約文の抽出

ステップ 2 で残ったセグメントに対して、要約文数の数だけそのセグメントから要約文の抽出を行なう。要約文の抽出には、展望台システムのスコアと、機械学習による分類結果の信頼値を用いる。より高いスコアや信頼値を持つ発話を選定することで、相槌などの要約に不適切な発話を除外できると考えられる。

要約文の抽出は、トピックの切り替わりを考慮し、対話の流れや文意を保つように、以下の手順に沿って行なう。図 3 に、このステップの概要図を示す。

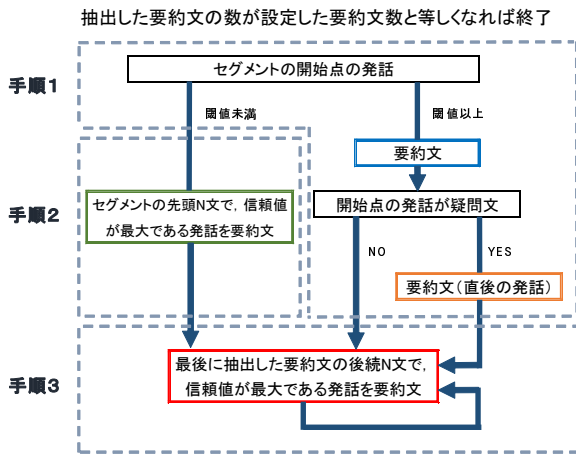


図3 要約文抽出の概要図

発話	信頼度	
S₁ (not疑問文)	0.476	←手順1
S₂	0.979	←手順3
S ₃	0.494	
S ₄	0.409	
S ₅	0.292	
S ₆	0.067	
S₇	0.831	←手順3
S ₈	0.161	
S ₉	0.042	
S ₁₀	0.340	

図4 要約文の抽出例 (要約文数3, 開始点閾値0.4, N=5)

ステップ3-1: まず、セグメントの開始点の発話は、トピックの切り替わりを表していることが多いため、低めの閾値を用いてなるべく抽出する。セグメント開始点の発話が閾値以上であれば要約文として抽出し、閾値未満であれば、ステップ3-2に移る。また、発話の照応関係を考慮するために、要約文を抽出した際に、開始点の発話が疑問文であれば、直後の発話も要約文として抽出する。この時点で、抽出した要約文の数が設定した要約文数に到達した場合は、このセグメントからの要約文の抽出は終了する。要約文数に満たなければ、ステップ3-3に移る。

ステップ3-2: セグメントの先頭 N 文の発話から、機械学習の信頼値が最大である発話を要約文として抽出する。抽出後、ステップ3-3に移る。

ステップ3-3: 最後に要約文として抽出した発話に後続する N 文の発話から、機械学習の信頼値が最大である発話を要約文として抽出する。そして、抽出した要約文の数が設定した要約文数に到達するまで、このステップ3-3を繰り返し行い要約文の抽出を行なう。

ステップ3-2, 3-3におけるパラメータ N は、ある要約文から次の要約文までの発話の範囲を意味している。

図4に、あるセグメントに対する要約文抽出の例を示す。図4においてまずステップ3-1が適用され、セグメント開始点の発話 S_1 の信頼値は開始点閾値以上であるため S_1 は要約文として決定される。 S_1 は疑問文ではないため、次はステップ3-3

のルールが適用される。最後に抽出した要約文は S_1 であるため、 S_1 から後続5文の発話 $S_2 \sim S_6$ の中で最も高い信頼値をもつ発話 S_2 を要約文として決定する。この時点では、抽出した要約文の数が設定した要約文数に満たないため、再びステップ3-3のルールを適用して、発話 $S_3 \sim S_7$ の中で最も高い信頼値をもつ発話 S_7 を要約文として決定する。ここで、設定した要約文数と等しくなったため、このセグメントからの要約文の抽出を終了する。

ステップ4: 要約文の補填

重要度の低いセグメントが対話中に存在した場合はそのセグメントから要約文を抽出しないため、実際に抽出した要約文数が式(2)で求めた要約文数よりも少ない場合がある。そこで、式(2)で設定した要約文数と同数になるように要約文の補填を行なう。提案手法では後続の5文に対して要約文数の抽出を行なうため、図4の $S_8 \sim S_{10}$ のように、セグメントの後半部分は要約文の判定が行なわれない場合がある。このような未判定の発話を考慮するために、要約文の補填はこれらの発話文に対して行なう。未判定の発話文の中から、最大の信頼値をもつ発話を要約文として補填し、要約文数と同数になるように繰り返し補填処理を行なう。

4. 実験

実験対象データは、4名の話者が特定の話題について自由に発話している8つの対話(全1295発話)とした。また正解データには、対話データの各発話に、正例文、負例文の評価をタグ付けしたものを使用した。

はじめに、3.1節で説明した新たな素性群 F を追加した分類器を作成し、抽出結果を正解データと比較することで重要文抽出実験を行なった。次に、重要文抽出の精度実験で用いた8対話のうち3対話について、3.2節で説明した提案手法に基づいて要約を作成し、先行研究の要約と被験者テストを行い、評価実験を行なった。

4.1 重要文抽出実験

重要文抽出実験では、新たな時間情報の素性を追加した機械学習による分類器を作成した。正解データを用いて、抽出した重要文と正例文の適合率、再現率から F 値を算出することで、時間情報の有効性を検証した。実装には、データマイニングツール Weka^(注1) の学習器 SVM [4] を用いて機械学習による分類器の抽出精度実験を行なった。8対話交差検定を用いて、得られた8回の結果を平均して F 値を算出した。

新たな時間情報の素性群 F の有効性を検証するために、重要文抽出の実験として2つの実験を行なった。まず1つ目に、先行研究で用いられた素性群に、時間情報の素性群 F を加えて、重要文抽出の精度を確認する実験を行なった。次に、先行研究の素性群から非言語情報の素性群 E を除いたものに、時間情報の素性群 F を加えて、重要文抽出の精度を確認した。

4.1.1 先行研究 (Baseline1) との比較実験

先行研究の素性群 ABCDE を Baseline1 として、Baseline1

(注1): <http://www.cs.waikato.ac.nz/~ml/weka/>

表 1 重要文抽出実験結果 1

組み合わせパターン	F 値
(1) Baseline1 (素性群 ABCDE)	0.785
(2) (1) + F_{1-1}	0.792**
(3) (1) + F_{1-2}	0.793
(4) (1) + F_2	0.780
(5) (1) + F_3	0.784

と新たな時間情報の素性群 F と組み合わせると素性の有効性を検証した。

結果を表 1 に示す^(注2)。表 1 において、Baseline1 と比較すると、今回新たに時間情報の素性を追加した場合は (2), (3) の発話速度を追加したときだけでなく、(1) の先行研究の F 値より上回っていることが分かる。

そこで、(2), (3) の違いを検証するために、各対話における抽出精度の評価を行なった。(1) に対して、(2), (3) が本当に有意差が存在するか確認するために、それぞれの平均値間に有意な差があるか対応のある両側検定の t 検定を行なった。t 検定を行なった結果、(1) と (2) のみ有意水準 5 % で有意差が見られた ($t=2.36$, $df=7$, $p<0.05$)^(注3)。これより、先行研究に発話速度 F_{1-1} を追加することで、精度向上に貢献していることが確認された。

4.1.2 非言語情報なし素性群 (Baseline2) との比較実験

2.3.1 節で説明したように、先行研究の笑いと盛り上がりに関する非言語情報の素性群 E が有効に機能しないことが、先行研究の実験結果より確認されている。そこで次に、先行研究の素性群 ABCDE (Baseline1) から素性群 E を除いた素性群 ABCD を新たな Baseline2 として、Baseline2 と時間情報の素性 F 群と組み合わせることで再度、素性の有効性を検証した。

結果を表 2 に示す。表 1 と表 2 を比較すると、(6)~(10) 全てにおいて、素性群 E を用いている (1)~(5) よりも精度が向上していることが確認できる。また、(7)(8) 発話速度 F_{1-1} , F_{1-2} を追加したとき、(10) 発話オーバーラップ F_3 を追加したときが、(6) Baseline2 よりも上回る結果となった。

そこで 4.1.1 節と同様に、これらの違いを検証するために、各対話における抽出精度の評価を行なった。(6) に対して (7), (8), (10) が有意差が存在するか t 検定を行なった結果、(6) と (8) においてのみ、有意水準 5 % で有意傾向がみられた ($t=2.36$, $df=7$, $0.05<p<0.10$)。

4.1.1 節、4.1.2 節の 2 つの実験より、発話速度を追加することで、共に一定の有効性が確認された。このことより、発話速度は非言語情報の素性において、重要文抽出の精度改善に貢献しているといえる。一方で、発話タイミング F_2 と発話オーバーラップ F_3 が有効に機能しなかった原因としては、発話タイミングや発話オーバーラップは話の重要度だけでは一意に定まら

表 2 重要文抽出実験結果 2

組み合わせパターン	F 値
(6) Baseline2 (素性群 ABCD)	0.789
(7) (6) + F_{1-1}	0.794
(8) (6) + F_{1-2}	0.793*
(9) (6) + F_2	0.788
(10) (6) + F_3	0.794

表 3 要約生成結果

手法	適合率	再現率	F 値
先行研究	0.732	0.509	0.600
提案手法	0.779	0.543	0.639

ず、発話者の外向性や発話者同士の社会的関係など様々な要因からも生じるためであると考えられる。これは、話者あたりの発話オーバーラップ数の割合が、ある話者では 60 % であったのに対し、一方では 14 % と話者によって異なるという結果からも確認できた。

4.2 生成した要約の評価実験

評価対象は、先行研究の要約手法 2.2 節と、テキストセグメンテーションを用いた提案手法 3.2 節の 2 つの要約を対象とした。式 (1) で説明した要約文数のパラメータ β は、先行研究と要約率が同一になるように設定した。また、3.2 節で説明した要約文抽出のパラメータ N は、実際に要約を生成する対話のテーマとは異なるデータを用いてパラメータのチューニングを行った結果、本手法では $N=5$ と設定した。

評価実験は、要約として抽出された個々の発話に対する抽出精度による検証と、生成された要約文全体に関する検証の 2 つを行なった。

4.2.1 要約文抽出精度の検証

まず、4.1 節と同様に、正解データと生成された要約文を比較することで、F 値を算出する。この評価尺度では、それぞれの要約手法が生成した要約に、どの程度正しく正例文が含まれているかが評価される。要約文の抽出精度の結果を表 3 に示す。

表 3 より、要約文を生成した際の重要度評価値、適合率、再現率、F 値はどれも提案手法が先行研究を上回る結果となっている。提案手法では、各セグメントから連続する N 文の発話から機械学習の信頼値が最も高い発話を繰り返し要約文として抽出しているため、先行研究より多くの重要文を抽出できていると考えられる。先行研究では、対話の要点となるようなスコアベース文を主軸にして、要約としての読みやすさや対話の流れを考慮した要約手法を取っているため、重要度があまり高くなくてもスコアベース文の前後に存在する文を抽出している。そのため、要約文の抽出精度は提案手法より低い値になっていると考えられる。

4.2.2 要約文全体の評価検証

次に、要約文全体の評価として、対話の構造や流れ、原文との文意の違いなどについて評価を行なう。4.2.1 節の評価は、出力された要約文中の各発話に対して、それが正例文か負例文であるかの評価を示している。つまり、より多くの正例文を要約として抽出していたとしても、発話の繋がりのない要約に

(注2)：表中の ** と * は、ベースラインに対してそれぞれ有意差がある、有意傾向があることを示している。

(注3)：(2) よりも F 値が高い (3) に有意差がなかった理由は、(2) が全ての対話で F 値が (1) よりも向上しているのに対して、(3) は対話によっては大きく精度が低下しているものが存在したためである。

なっている可能性がある。要約において重要なことは、原文の文意を保ち、読みやすい文であるため、これらについて評価を行なう。

本論文では、生成された要約文全体の評価基準として、「読みやすさ」と「要約の正しさ」を定義した。先行研究と提案手法に対して、9名の被験者にこの2つの基準をそれぞれ5段階^(注4)で評価してもらい、その平均値に基づいて各手法の有効性を検証した。評価の際に被験者に伝えた「読みやすさ」と「要約の正しさ」の定義を以下に示す。

読みやすさ：生成された要約文のスムーズさ、対話の流れの自然さ

要約の正しさ：元の対話の文意を保っているか、文脈は正しいか

この2つの項目について、被験者には次の手順で評価してもらった。

手順1：被験者には各手法によって生成された要約文を同時に与え、要約だと伝えずに機械的に生成された文だと説明し、「読みやすさ」の評価をしてもらう。

手順2：その後、読んだ文が自由対話の要約文だと伝え、対話の原文を提示し、その内容と比較して、生成された要約文の「要約の正しさ」の評価をしてもらう。

手順1で、被験者に要約文であることを伝えずに「読みやすさ」を評価してもらうことで、被験者が単純に対話文の「読みやすさ」について正しく評価できると考えられる。手順2については、2手法の説明はせずにそれぞれ異なった方法で生成された要約だと伝え、要約文と原文は何度も読み直しても良いとすることで、「要約の正しさ」をより正確に評価できると考えられる。

結果を表4に示す。表4より、「読みやすさ」と「要約の正しさ」の2つの項目に対して、提案手法が先行研究より低い評価を取っていることが分かる。また、「読みやすさ」の評価に対しては、「要約の正しさ」の評価よりも先行研究と提案手法に大きな差があった。提案手法が、表3の評価において先行研究よりも優れていた結果であったのに対し、「読みやすさ」や「要約の正しさ」が低い評価になったのは、提案手法は要約文の発話に繋がりが弱かったことが原因として挙げられる。提案手法では、各セグメント毎に発話の繋がりを保つように要約文を抽出していたが、セグメント間では発話の繋がりを考慮していなかったため、このような低い評価になったと考えられる。また先行研究では、対話全体のトピックを網羅していない場合があるものの、スコアベース文を主体に要約文を組み立て、前後文の選定を行なうことで、対話の流れが自然で、文意も比較的保たれていたと考えられる。つまり、提案手法では各トピック毎に要約文を抽出しているため、要約文の内容という点では良いが、トピック間の繋がりががないため要約全体としては読みにくい要約になっている。それに対し先行研究では、対話全体のトピックには触れていないため、要約文の内容としては不十分かもしれないが、その分より正確で分かりやすい要約を生成していると

表4 被験者テストによる評価結果

手法	読みやすさ	要約の正しさ
先行研究	3.07	3.15
提案手法	2.15	3.00

考えられる。

5. まとめ

本論文では、複数人による自由対話を対象とした要約手法について、重要文抽出部に時間情報を追加し、テキストセグメンテーションを用いた要約手法を提案し、その有効性を検証した。

重要文の抽出精度の実験では、発話速度や発話タイミング、発話オーバーラップを新たな時間情報の素性として追加し、重要文抽出の精度向上を検証した。その結果、発話速度の素性を追加することで、に一定の有効性があることが確認された。

また、新たに提案したテキストセグメンテーションを用いた要約手法は、要約文として重要な発話を先行研究より多く抽出していた。しかし、被験者テストによって、要約文全体の評価である「読みやすさ」や「要約の正しさ」といった観点では、先行研究を下回る結果となった。このことから、今回の提案手法は先行研究と比べ、対話のトピックを網羅し、要約率を任意の値に変更できるようになったものの、要約文全体としては不十分であることが確認された。

今後の課題は、重要文抽出の精度を向上するために今回触れていない非言語情報を取り入れることや、トピック間の関係を考慮することで、より読みやすく、原文の文意を保った要約を生成できると考えられる。また、今回の手法は重要文の抜粋による要約であったが、より良い要約を生成するためには文生成に基づく要約[8]も不可欠であり、この点も今後の課題の一つである。

文 献

- [1] 奥村学, 難波英嗣 (2005), “知の科学 テキスト自動要約”, オーム社.
- [2] Y. Tokunaga and K. Shimada, “Multi-party conversation summarization based on sentence selection using verbal and nonverbal information”, Proceedings of ICSCAI 2014, 2014.
- [3] W. Sunayama and M. Yachida, “A Panoramic View System for Extracting Key Sentences with Discovering Keywords Featuring a Document”, Systems and Computers in Japan, Vol.34, No.11, pp.81-90, 2003.
- [4] V.Vapnik, “The Nature of Statistical Learning Theory”, Springer-Verlag, 1995.
- [5] 藤原敬記, 伊藤敏彦, 荒木健治, “タスク指向対話における相互の対話意図を考慮した対話リズムの分析”, 人工知能学会言語・音声理解と対話処理研究会, SIG-SLUD-A701, pp.45-50, 2007.
- [6] M. Galley, K. McKeown, E. Fosler-Lussier, and H. Jing, “Discourse segmentation of multi-party conversation”, Proceedings of ACL 2003, 2003.
- [7] 望月源, 奥村学, 岩山真, “語彙的結束性に基づく語彙的連鎖の計算”, JAIST Technical Memorandum, IS-TM-2000-002, 2000.
- [8] T. Oya, Y. Mehdad, G. Carenini, and R. Ng, “A Template-based Abstractive Meeting Summarization: Leveraging Summary and Source Text Relationships”, Proceedings of INLG 2014, 2014.

(注4) : 1: 悪い ~ 5: 良い