

# 非言語情報と重要文の選定に基づく 複数人自由対話要約手法の改善

徳永 陽

嶋田 和孝

九州工業大学 情報工学部 知能情報工学科

{y\_tokunaga,shimada}@pluto.ai.kyutech.ac.jp

## 1 はじめに

自動要約に関する研究の多くは、新聞記事や論文のような、書き手が1人で、かつ文体が書き言葉であるものが多かった [1]。自動要約に関する研究が進むにつれて、テキストだけでなく、複数人の発話で構成される対話文を対象とした研究も行われるようになってきた。岩崎ら [2] は、コールセンターにおける対話データの要約として、対話データから営業日報を自動生成するための重要文抽出手法を提案している。

しかし、雑談のような複数人による自由対話では、発話の自由度が高く、ルールや規則性が少ないため、コールセンターや会議のような話の内容や構成がある程度形式化されている対話の要約より難しいと考えられる。このような自由対話を要約するには、自由対話特有の特徴を捉えた重要文の抽出が必要になってくる。例えば、日本語はよく先行詞を省略して話すことが多いため、照応関係を正確に把握することが重要である。また、会議などではあまり見られない笑いなどが起こることも自由対話特有の特徴だといえる。笑いが起こったり、場が盛り上がっている状況は、話者が話している内容に興味を持っていると考えられ、比較的重要な発話が多くなると考えられる。

嶋田ら [3] は、複数人の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を考慮し、発話の抜粋による自動要約手法を提案している。図1は、嶋田らの要約手法の全体像であり、要約は二つの段階を経て生成される。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り、要約生成部では、それらの文意を崩さず、読みやすい文を生成することが必要である。

本論文では、先行研究が触れていない非言語情報を重要文抽出の素性に追加することで、抽出精度の改善を目指す。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。



図1: 要約手法の全体像。

## 2 先行研究

まず、基本となる嶋田らの手法について説明する。スコアによる分類器と機械学習による分類器を用いて抽出した重要文と、その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、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、展望台システム [4] などを用いて各発話をスコアリングし、このスコアを基に上位10%以内の発話(以下、スコアベース文)の抽出を行っている。

また、機械学習による分類器は、発話単体や発話間に関する特徴など計19の素性を用いて、重要文を抽出している。これらの素性は、大きく4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:** 発話間の特徴に関する素性  
(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)

素性の詳細については、文献 [3] を参照してほしい。

### 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、以下の方法を提案している。

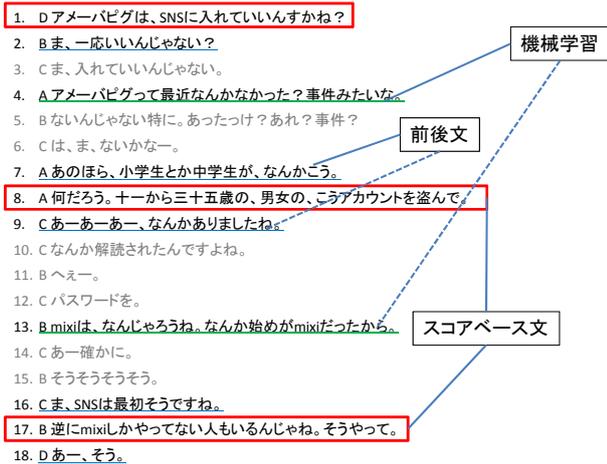


図 2: 先行研究の要約.

まず、スコアベース文を要約文として抽出する。しかし、スコアに基づく処理で抽出されるのが対話文全体の 10% の抽出であるため、要約としては不十分である。そこで、スコアベース文の前後文も要約として抽出する。前後文の有無が、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、要約の読みやすさや内容の理解に大きく影響するとしている。また、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分類器の信頼度が最も高い 1 文を要約として追加する。

図 2 がその抽出例である。枠に囲まれている発話がスコアベース文、その前後の下線付き発話が前後文、それらのある下線付き発話が機械学習分類器の信頼度によって抽出された発話である。

### 3 問題点と解決方法

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。一つ目は、機械学習による分類器の素性群 A~D において、会議のような形式的な対話では見られないが、発話の自由度の高い自由対話で見られる非言語情報や場の状況について一切触れられていない点である。二つ目は、要約を生成する際、対話の流れを考慮するために無条件で追加したスコアベース文の前後文に、明らかに要約に不適切な相槌などが含まれている点である。

本論文は、先行研究に挙げられた二つの問題に対して、それぞれの解決方法について提案する。

#### 3.1 非言語情報を追加した機械学習分類器

先行研究で触れられていない自由対話特有の特徴として、非言語情報や場の状況などが考えられる。これらを機械学習の分類器の素性として追加することで重

要文抽出の精度向上を試みる。

徳久ら [5] は、発話行為や修辞構造の分類を行い、これらを分析した結果、主観的な発話やユーモラスな発話、相手の発話を補完するような発話は、盛り上がり時に頻出すると述べている。また、須見ら [6] は、笑いは盛り上がり強い影響を与えると考え、笑いを検出する手法について述べている。ここで、笑いが起こっているときや盛り上がっているときは、話者は意味のある発話をしていると仮定し、盛り上がりと笑いについての非言語情報を新たに素性群 E として追加する。

追加する素性は、盛り上がりに関する素性 1 個と笑いに関する素性 10 個の計 11 個で、これらは大きく 5 つに分類される。以下で詳しく説明する。

$E_1$  : 盛り上がり度

$E_2$  : 笑いの有無

$E_3$  : 笑いと言話のタイミング

$E_4$  : 笑いの要因 (internal, external)

$E_5$  : external な笑いからの距離

盛り上がり度 ( $E_1$ ) は、発話毎に二人のアノテータが 5 段階で評価した平均値を利用する [7]。笑いについては、発話中の笑いの有無 ( $E_2$ ) だけでなく、笑いと言話のタイミングで分類し、詳細化することで発話中の笑い情報が有効か確認できると考える。そこで、発話と笑いが起こったタイミングについて以下の 4 パターンに分類した情報を  $E_3$  とする。

- 発話中に笑いが発生し、終了する。
- 発話前に笑いが発生し、発話中に笑いが終了する。
- 発話中に笑いが発生し、発話後に笑いが終了する。
- 笑い発生中に発話が始まり、終わる。

また、嶋田ら [7] は、笑いを外部からの発話 (行動) に対して発生する笑い (external) と自発的な笑い (internal) に分類し、external な笑いの周辺には、笑いを引き起こした原因となる発話が存在するとしている。起こった笑いを external と internal に分類した情報 ( $E_4$ ) を用いることで、笑いと言話の関係をより詳しく分類できる。さらに、起こった笑いが external の場合、周辺には笑いを引き起こした発話があるため、その直前 3 発話以内の発話に対し、external な笑いを含む発話からの距離 ( $E_5$ ) を素性として割り振る。

#### 3.2 信頼度を用いた要約の生成

先行研究の手法では、スコアによる分類器によって抽出された発話の前後の文に、要約として不適切な文 (例えば図 2 中の 9 番目の発話「あーあーあー…」のような相槌) が多く含まれてしまうという問題があった。

	閾値0.70	信頼度	
$S_1$ : 文Aの信頼度		0.83	
$S_2$ : 前文の信頼度	0.75		事例(a)
$S_3$ : <b>スコアベース文</b>		0.91	
$S_4$ : 後文の信頼度		0.08	
$S_5$ : 文Bの信頼度	0.67		事例(b)
$S_6$ : 文Cの信頼度		0.33	
$S_7$ : <b>要約決定済み文</b>		0.87	

図 3: 信頼度に基づく要約文選択の例.

本論文では、この問題を解決するため、2.2 節で説明した手法に対し、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度として用いて選定する手法を提案する。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の信頼度が低くなりやすいためである。選定の具体的な流れは以下のようになる。

**手順 1:** スコアベース文の前後の文について、その文の信頼度が閾値以上であれば、その文を選択する。

**手順 2:** 前後の文の信頼度が閾値以下であれば、スコアベース文とすでに要約文として抽出された文の間で信頼度が(その値が閾値以上かどうかにかかわらず)最大の文を選択する。

図 3 に例を示す。図 3 の (a) は手順 1 の例であり、(b) が手順 2 の例である。この例では、スコアベース文の前の文 ( $S_2$ ) は手順 1 の条件を満たす。従って、それ以前により高い信頼度を持つ文 ( $S_1$ ) が存在しても、スコアベース文に近い  $S_2$  が優先される。一方で、スコアベース文の後ろの文 ( $S_4$ ) の信頼度は閾値よりも低いため、手順 2 が適用される。この場合は、その値が閾値以上かどうかにかかわらず、最大の信頼度を持つ文(ここでは  $S_5$ ) が選択されることになる。

ここで、閾値以下の文であっても必ず文を選択するのは、スコアベース文間の発話を残すことが要約文としての読みやすさや発話の流れ理解するのに貢献する傾向が強いためである。このように信頼度を有効利用することで、先行研究と比較してより適切な要約文が選定可能になる。

## 4 実験

はじめに、3.1 節で説明した新たな素性群 E を追加した素性群 A~E でなる機械学習による分類器を作成し、抽出結果を正解データと比較することで重要文の抽出精度実験を行った。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話している 8 つの対話

表 1: 重要文抽出実験結果.

組み合わせパターン	F 値
(1) 素性群 ABCD(ベースライン)	0.790
(2) ベースライン+ $E_1$	0.792
(3) ベースライン+ $E_2$	0.788
(4) ベースライン+ $E_3$	0.787
(5) ベースライン+ $E_1, E_2$	0.790
(6) ベースライン+ $E_4$	0.789
(7) ベースライン+ $E_5$	0.786
(8) ベースライン+素性群 E	0.788

(全 1295 発話)とし、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。

次に、重要文の抽出精度実験で用いた 8 対話のうちの 3 対話に対して、先行研究、提案手法、機械学習分類器の抽出結果のみを用いた手法の 3 つを用いて生成した要約について評価実験を行った。

### 4.1 重要文抽出実験

先行研究で用いられた素性群に、盛り上がりや笑い情報の素性群 E を加えて、重要文抽出の精度を確認する実験を行った。実装には、データマイニングツール Weka<sup>1</sup> の学習器 SVM[8] を用いて機械学習による分類器の抽出精度実験を行った。

結果を表 1 に示す。表 1 から今回追加した素性では、もとの素性群 A~D を用いた時の F 値と僅かな値の差しか見られなかった。これは、今回の実験データでは、笑い含む発話が全体の 25 %弱しかないため、素性として十分に機能しなかったものと考えられる。

### 4.2 生成した要約の評価実験

評価対象は、先行研究の手法(2.2 節)と信頼度を用いた提案手法(3.2 節)、さらに機械学習による分類器によって抽出された重要文の信頼度上位  $n$  発話のみを要約文とした 3 つの要約を対象とした。 $n$  の値は先行研究や提案手法で得られる要約文数と同一にした。

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 つを行う。

それぞれの対話文には、各発話毎に要約として適切な発話は正例文、不適切な発話は負例文と評価されたタグ付けが行われている。これを用いて生成された要約文と正例文の適合率、再現率、F 値を算出する。すなわち、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。

<sup>1</sup><http://www.cs.waikato.ac.nz/ml/weka/>

表 2: 要約生成結果.

手法	適合率	再現率	F 値
機械学習	<b>0.859</b>	<b>0.605</b>	<b>0.709</b>
提案手法	0.732	0.509	0.599
先行研究	0.628	0.440	0.567

要約文の抽出精度の結果を表 2 に示す. 要約文を生成した際の適合率, 再現率, F 値はともに機械学習による手法の要約文がとて高い値を取っている. 機械学習による分類器の重要文抽出の実験結果 (表 1) が精度 8 割弱と高い値を取っており, さらにその上位  $n$  文を抽出しているため, 今回のような結果になったと考えられる. しかし, 提案手法は, F 値は機械学習による要約手法に劣るものの, 先行研究の手法よりも勝っていることから, 要約に不必要な発話を完全ではないが取り除くことができていると考えられる.

ここまで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否かのための判断であり, 対話の構造や流れ, 原文との文意の違いについての評価は含まれていない. すなわち, 要約全体としての評価ではない. 一方で, 要約において重要なのは, 原文の文意を保ち, 読みやすい文になっていることである.

そこで, 要約文全体の評価が必要となる. 本論文では, 評価方法として「読みやすさ」と「要約の正しさ」を定義し, 7 名の被験者に 3 つの手法の出力を 5 段階<sup>2</sup>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. 評価時に被験者に伝えた「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

この二つの項目について, 次の手順で評価してもらった. まず, 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」の評価をしてもらった. 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」の評価をしてもらった.

結果を表 3 に示す. 「読みやすさ」と「要約の正しさ」の二つの項目に対して, 提案手法が他の手法より高い値を取っている. 表 2 の結果では, 機械学習の手法と大きな差があるのに対し, 表 3 の尺度では提案手法の方が高くなったのは, 機械学習による手法は, 対

<sup>2</sup>1: 悪い ~ 5: 良い.

表 3: 人手による評価結果

手法	読みやすさ	要約の正しさ
提案手法	<b>3.52</b>	<b>3.47</b>
機械学習	3.09	3.04
先行研究	2.66	2.57

話の流れが途切れていたり, 原文の文意があまり保たれていなかったためだと考えられる. これは, 機械学習のみの手法では要約文生成の際に文間の関係を反映することができないためである. それに比べ, 提案手法は, 精度の高いスコアベース文を主体に要約文を組み立てていくため, 対話の流れが自然で, 文意も比較的保たれていたため評価が高くなった.

## 5 おわりに

本研究では, 複数人による自由対話を対象とした要約手法について, (1) 非言語情報を機械学習の素性に追加することおよび (2) 信頼度と閾値に基づく要約文選択の 2 点について拡張した. 非言語情報は有効に機能しなかったが, 要約文選択手法についてはその有効性が確認された.

## 参考文献

- [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] 岩崎礼次郎, 荒木健治, “コールセンターにおける対話データを用いた営業日報の自動生成”, 人工知能学会ことば工学研究会資料, pp.87-94, 2006.
- [3] Kazutaka Shimada, Shinpei Toyodome, and Tsutomu Endo, “Conversation summarization using machine learning and scoring method”, Proceedings of the 13th Conference of the Pacific Association for Computational Linguistics (PAACLING2013), 2013.
- [4] 砂山渡, 谷内田正彦, “文章要約のための特徴キーワードの発見による重要文抽出法-展望台システム-”, 情報処理学会研究報告, 自然言語処理研究会報告, pp.103-110, 2000.
- [5] 徳久良子, 寺島立太, “雑談における発話のやりとりと盛り上がりの関連”, 人工知能学会論文誌, vol.21, No.2, pp.132-142, 2006.
- [6] 須見康平, 河原達也, 緒方淳, 後藤真孝, “ポッドキャストを対象とした音リアクションイベント検出”, 情報処理学会研究報告, vol.2009-SLP-77, No.24, pp.132-142, 2009.
- [7] 嶋田和孝, 楠本章裕, 横山貴彦, 遠藤勉, “複数人談話における笑いの情報を考慮した盛り上がり判定”, 電子情報通信学会, 言語理解とコミュニケーション研究会 (NLC), NLC2012-7, pp.25-30, 2012.
- [8] V.Vapnik, “The Nature of Statistical Learning Theory”, Springer-Verlag, 1995.