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

中山 友梨 塩田 宰 嶋田 和孝

† 九州工業大学 〒820-8502 福岡県飯塚市川津 680-4 E-mail: †shimada@ai.kyutech.ac.jp

あらまし 本論文では、4人一組の議論対話コーパスに関する新しいアノテーションデータの構築と要約手法について検証する。我々はこれまでに対話全体のまとめとしての参照要約とその参照要約に関連する発話群のアノテーションデータを公開している。しかしながら、議論のより細かい要素について要約をしたい場合、このデータセットでは十分ではなかった。そこで、トピックの切れ目をセグメントとし、セグメント単位での要約評価データの構築とそのデータを用いた要約手法の提案および比較を行う。要約ではまずセグメント内から重要発話を抽出し、その結果をまとめ上げることで抽出型と生成型の要約結果を得る。実験から、抽出型は要約の正確性の面で、生成型は読みやすさの面で有効であることが確認された。

キーワード 複数人対話コーパス,トピック単位の要約,データ構築

Corpus construction for topic-based summarization of multi-party conversation

Yuri NAKAYAMA[†], Tsukasa SHIOTA[†], and Kazutaka SHIMADA[†]

† Kyushu Institute of Technology 680-4 Kawazu, Iizuka, Fukuoka 820-8502, JAPAN E-mail: †shimada@ai.kyutech.ac.jp

Abstract In this paper, we report corpus construction and topic-based summarization methods for multi-party conversation. We have already constructed reference summaries and the list of important utterances in each discussion. However, fine-grained summaries about topics in a discussion often are desired in many situations. Therefore, we construct topic-based summaries and propose an important utterance extraction method and two summarization processes using the extracted utterances; extractive and abstractive methods. In the experiment, the extractive method was superior in terms of "accuracy as a summary" while the readability of the abstractive method was better

Key words Multi-party conversation, Topic-based summarization, Summarization Corpus

1. はじめに

会議とは複数人が議題について意見を交換し合う場である. 会議は内容を把握したい人全員が参加することが望ましいが, 必ずしも全員が参加できるというわけではない.会議に参加し ていない人への情報共有のために,議事録を作成することが一 般的である.しかし議事録作成を人手で行うと,多大な時間的 労力や,人的労力を要する.よって,議事録を自動で作成する ことができれば,労力の削減を行いつつ,会議内容の共有が可 能であると考えられる.

会議では多くの発言がなされているため、要約に必要な発話 (以降、重要発話)を選定しなければならない、会議全体から 要約をすると、最終決定に関しての情報が圧縮されて要約と してまとめられることが多い、Yamamura ら [1] は参照要約を Kyutech コーパス [2] の対話全体から作成している。しかし、 実際には最終決定に関しての情報のみでなく,最終決定に至るまでの案など様々な情報についても取り逃さず要約としてまとめたい場合も多い.

本研究では Kyutech コーパスを対象とし、図 1 に示すように対話全体よりも細かい粒度で形成されるセグメントを対象とした要約を行う。要約のために必要なデータセットを作成する。作成したデータセットは Web 上で公開する $(^{121})$.

2. 関連研究

自然言語処理の分野においては古くから文書を要約するタスクに関心が持たれている. Mani [3] は自動要約についての定義や,要約に関する自動化手法についてまとめている. 自動要約手法は,抽出型要約と生成型要約の2種類に分類される. 抽出

(注1): http://www.pluto.ai.kyutech.ac.jp/~shimada/resources.html

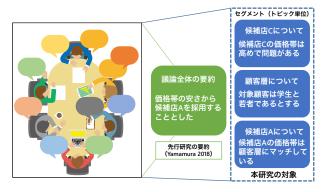


図 1 提案手法概略図.

型要約のメリットは、原文内に出現する単語のみで要約が作成できるため、正しくない情報がほぼ存在しないことである。デメリットとして、原文の情報を抽出しただけでは読みづらい要約が生成されてしまうことがあげられる。生成型要約のメリットは、抽出型要約とは対照的に読みやすい要約が生成できるということがある。近年、ニューラルネットワークに基づく多くの生成型手法が提案されている[4]。デメリットは、原文の単語や言い回しを言い換えてしまうため、抽出型要約よりは原文の内容を保持できない可能性がある。

本研究では複数人対話の要約を行う. Xie ら [5] は会議での抽出型要約のタスクでは、文のトピックに関連した特徴量が有効であることを示し、顕著性に基づいて文をリサンプリングする手法を提案した. Shang ら [6] は会議要約のための、教師無しグラフベースのフレームワークを提案しており、AMI コーパスと ICSI コーパスで評価を行った. Li ら [7] はビデオと音声の両方から抽象化される会議要約手法を提案しており、従来の文に関する特徴量に加えて、話者の注目度をマルチモーダル特徴量として導入していた. 山村ら [8] は時間情報についての素性を用いた重要文抽出と、LCseg [9] を用いたテキストセグメンテーションで複数人対話を要約している. 本研究では関連研究を踏まえ、まず、あるトピックでまとまったセグメントに対して重要発話推定を行う. その結果を利用して抽出型モデルと生成型モデルの適用によって要約を作成する.

3. データセット構築

本節では、複数人対話コーパスを用いたデータセットの構築ついて説明する。まず、本研究で対象とする Kyutech コーパス [2] と本研究での要約の単位であるセグメントについて説明する。次に、3.2 節でセグメントごとの参照要約の作成について、3.3 節でその参照要約とセグメント内の発話の対応付け(重要発話アノテーション)について説明する。最後にその参照要約やアノテーション結果の一致率などを3.4 節で検証する。

3.1 Kyutech コーパスとセグメント

Kyutech コーパスには 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 9 つの対話が含まれている. 具体的な例を表1 に示す. 書き起こされた発話に対して,発話に対する付加的なタグ(表 1 でフィラーを意味する F タグや疑問文を意味する Q タグ)や 28 個のトピックタグ(表 1 の C Cand X や V Vague)が

表 1 Kyutech コーパス書き起こしデータの一部.

#ID	必須	追加	発話と発話タグ
A	CandX		なるほど/
С	CandS		で +
С	CandS		ふうじんと (F この) ボノパスタ (Q)/
A	CandS		うん/
С	CandS		この二つで迷ってて/
A	CandS		うん/
С	CandZ	Vague	で (F えーっと)(?)+
C C	CandZ CandZ	Vague Vague	で (F えーっと)(?)+ あっそう/
			, , , ,
С	CandZ		あっそう/
C C	CandZ CandZ		あっそう/ 結局俺はボノパスタ +
C C	CandZ CandZ CandZ		あっそう/ 結局俺はボノパスタ + これはほんともう三角/

付与されている。各発話の最後に付与されている"+"は発話として次の発話に繋がることを意味し、"/"は1つの発話の終了を意味している。タグなどの詳細については紙面の都合上,[2]を参照して欲しい。

本研究では、対話全体からの要約を試みた先行研究 [1] と異なり、細かいセグメント単位での要約を目的とする。本論文では、必須タグの切り替わりをこのセグメントの切り替わりとする。表 1 において太線がその切り替わりを意味する。 Kyutech コーパスにおける全 9 対話でセグメントは 418 個存在する.

3.2 セグメントごとの参照要約の作成

まず前処理として、全セグメント(418 セグメント)から対話に直接関係のない発話を意味する "Chat"、議事進行の発話を意味する "Meeting"、前後の文脈からも判断できない発話を意味する "Vague"のトピックタグのみが付与された発話からなるセグメントを取り除く、この前処理のあとに得られた 381 セグメントに対して参照要約を作成する.

人手での要約は作成者によって内容に差が出てしまうため、 参照要約の作成はセグメントごとに2人の作業者が行う.本研 究での参照要約作成のルールを以下に説明する.

- 1セグメントに対して1文または2文で参照要約を作成.
- 要約に含まれる内容語は、必ずセグメントの発話中に存在するものでなければならない。セグメント内に出現しない単語や言い回しを用いて要約することは原則禁止.
- ただし、語尾など、セグメント内にある単語だけではどうしても文が完成しない場合は、セグメント内に出現しない単語を使用可.
 - 要約がどうしても作成できない場合は要約無し.

参照要約を作成した例の一部を表 2 に示す. 作業者が作成した 参照要約を A1, A2 とする. 前述のように, 人手の要約は作業 者によって内容にブレがあるため, A1 と A2 両方の内容を網羅 するように統合したデータ M を著者が作成する. ただし, A1 と A2 の内容がほぼ同じ場合は A1 の要約を M とした. 実際に 統合した例を表 3 に示す.

3.3 セグメントごとの重要発話アノテーション

次に 3.2 節で作成したセグメント単位の参照要約とセグメント内の発話の対応付け(重要発話アノテーション)を行う.よ

表 2 セグメントごとの参照要約の作成.

発話	要約
まずラーメンはもう +	
一人暮らし人口が +	一人暮らしや学生が
少ないし学生も +	狙えないためかいぶつはない。
狙えんけん俺もかいぶつはないと	
思ってて/	
(F えーっと) とりあえず +	
うん/	
ふうじんとボノパスタ比較した	
ときの +	ふうじんとボノパスタの
うん/	良いところをまとめる。
(F その) 両方の良いところ (Q)/	
そうね/	
で +	
(L ちょっとまとめる)/	
いやでも、んー +	
中学生が/	
良く分からない/	THE 6/2 ATT. 1
昼も/	要約無し
(?)/	
(?)/	

表 3 要約統合例.

	A1	A2	M
		和食屋さんが	
	不採算化直前	閉店する。閉店	不採算化直前
セグメント 1	だったため、	する理由は売上	だったため、
(C) X Z F I	和食屋さんが	が減ってしまって	和食屋さんが
	閉店する。	不採算化直前	閉店する。
		だったため。	
			三人とも今は
	高橋さんの	三人とも今は	一応ボノパスタ
セグメント 2	推しはボノパスタ	一応 ボノパスタ	推しとる。高橋さん
	とは違う。	推しとる。	の推しはボノパスタ
			とは違う。

り具体的には、3.2節で作成した A1, A2, M に対して、発話 ごとに重要発話か否か(1 or 0)を付与する. 以下にアノテーションルールを示す.

- 参照を作成するために必要だと思う発話には"1",不必要だと思う発話には"0"のラベルを付与.
 - 似た発話がある場合,発話の開始時間が早い発話を優先.
- 3.2 節の処理で要約が作成されなかった場合は、該当セグメント内のすべての発話に"0"と付与.

アノテーションを行った例の一部を表 4 に示す. 例えば,最初の要約「一人暮らしや学生が狙えないためかいぶつはない。」に対してその右にある 4 つの発話が与えられた場合,各作業者は要約と発話を見ながら,発話が要約に必要であるか否かを判断する. この例では,セグメントの中で「一人暮らし人口が +」と「少ないし学生も +」と「狙えんけん俺もかいぶつはないと思ってて/」の発話が要約に必要であると判断できるため,この3 つの発話にラベルの"1"のラベルが付与されている.

事前実験として複数人に少数のセグメントへのアノテーションを依頼したところ、参照要約の作成とは異なり、作業者間での結果の不一致はほとんど発生していないことを確認した (κ

表 4 参照要約に基づく重要発話アノテーション例.

要約	発話	ラベル
	まずラーメンはもう +	0
一人暮らしや学生が	一人暮らし人口が +	1
狙えないためかいぶつはない。	少ないし学生も +	1
	狙えんけん俺もかいぶつはないと	1
	思ってて/	1
	(F えーっと) とりあえず +	0
	うん/	0
	ふうじんとボノパスタ比較した	1
ふうじんとボノパスタの	ときの +	1
良いところをまとめる。	うん/	0
	(F その) 両方の良いところ (Q)/	1
	そうね/	0
	で +	0
	(L ちょっとまとめる)/	1
	いやでも、んー +	0
	中学生が/	0
要約無し	良く分からない/	0
安心無 し	昼も/	0
	(?)/	0
	(?)/	0

表 5 参照要約評価值.

	A1 と A2	A1 と M	A2 と M
ROUGE-1	0.596	0.921	0.662

値は 0.71). そのため A1, A2, M それぞれに 1 人の作業者でデータを作成した.

3.4 データセットの検証

3.2節で作成した参照要約と、3.3節で作成した重要発話データについて検証する。まず、著者が統合して作成した M が、A1 と A2 を適切に統合できているか否かを確認した。具体的には、作成した参照要約が、それぞれどの程度一致しているかのスコアを ROUGE-1 [10] で算出する。

表 5 に A1, A2, M それぞれの参照要約について ROUGE-1 を算出した結果を示す。表 5 の A1 と A2 における ROUGE-1 の値より,A1 と A2 は作業者によって言い回しなどのブレは 多少あるものの,似た要約を作成していることがわかった。また A1 と M, A2 と M についても比較を行った。A1 と M における ROUGE-1 の値が高いことから,A1 と M の参照要約が非常に類似していることがわかる。これは,A1 と A2 を統合する際に,内容がほぼ同じ場合,A1 の要約を M とするというルールによる結果である。A1 と A2 を比較した場合よりも,それぞれを M と比較した場合の方が高くなっているため,M は A1 と A2 を適切に統合できていると考えられる。これらの結果から,以降の実験には A1 と A2 を適切に統合できている M を用いる。

次に先行研究である Yamamura らの要約データ [1] との比較をする. 前述のように先行研究の要約データは、全体をまとめる参照要約を人間が書いた後、その参照要約に関連すると考えられる発話に重要発話のラベルを付与したものである. 一方で、本研究での重要発話はセグメント内において同様の処理を行ったものである. 対話中の全発話に対する重要発話とラベル付け

表 6 重要発話の割合.

A1	A2	M	山村ら[1]
26.90	26.35	27.61	22.98

された発話の割合を表6に示す. 先行研究の重要発話の割合が 最も低い(要約率が高い). 本研究で作成したセグメント内の重 要発話は先行研究と比較して3~4%多いことになる. これは, 先行研究が全体像を追うための要約になっているのに対し,本 研究のデータがトピックを踏まえた網羅性の高い要約になって いることを意味しており, 粒度の異なる要約データになってい ることが分かる.

4. 提案手法

本節では本研究の提案手法である重要発話推定と,それを用いた要約手法についての説明を行う.

4.1 重要発話推定

前節で述べたデータセットのうち、Mを用いて、Support Vector Machine (SVM) [11] で重要発話推定を行う。素性は、徳永ら [12] の手法を参考にした。発話を SVM に入力するためのベクトル化には以下に示す計 12 種類の素性を利用した。なお、素性を作成する際、形態素解析が必要なものに関しては $MeCab^{(\pm 2)}$ を用いた。

- 発話の長さ:発話中の文字数に基づく素性. ただし, 括 弧や記号は数えない.
- 頻出単語の有無:対話内に含まれる名詞,動詞,形容詞の上位50単語のうち,その中の単語を発話中に含む/含まないの二値(1 or 0).
- 特定タグの有無: "Chat" か "Vague" のタグを含む/含まないの二値(1 or 0).
 - 指示詞の有無: 含む/含まないの二値(1 or 0).
 - 接続詞の有無:含む/含まないの二値(1 or 0).
 - フィラーの有無: 含む/含まないの二値 (1 or 0).
 - 名詞の有無:含む/含まないの二値(1 or 0).
 - 動詞の有無:含む/含まないの二値(1 or 0).
 - 形容詞の有無:含む/含まないの二値(1 or 0).
- 直前の発話:直前の発話に(Q)タグを含む/含まないの 二値(1 or 0).
- 話者の連続:直前の発話と今の発話の話者が同じ/異なるの二値(1 or 0). ただし,対話の1発話目は,必ず「話者は連続していない」と処理する.
- 発話の出現位置:対象の発話がセグメント内で何番目に発言されたかものかを表す. 例えば1番目に発言されていれば"1",2番目に発言されていれば"2"とする.

4.2 要約手法

本節では重要発話推定の結果を用いた 2 種類の要約生成手法を説明する。まず 4.2.1 節で重要発話をそのまま連結させて要約を生成する抽出型手法について説明し、4.2.2 節で mBARTを用いて要約を生成する生成型手法について説明する.

(注2): http://taku910.github.io/mecab/

表 7 抽出型要約例.

発話	推定	要約結果(抽出)
まず +	0	
(D その)+	1	
和食屋さんが閉店するやん/	1	
うん/	0	
和屋さんが/	0	
そうそう (L 閉店して)+	1	
で、閉店する理由は +	0	和食屋さんが閉店するやん
売り上げが減ってしまっ +	1	そうそう売り上げが減ってしまっ
7/	0 0	とうとう処り上のが減らとしよう
減ってからの/		
その (D ふさ)/	0	
(D ふさ)/	0	
不採算化直前だったため/	0	
(L 不採算化)/	0	
(L うん)/	0	

4.2.1 抽出型手法

この手法では、4.1節で推定された重要発話を時系列順に以下のルールを適用して連結する.

- 重要("1")と推定された発話群を重要発話群とする.
- 括弧で囲まれた文字(タグ)や発話末尾の記号を除去.
- 重要発話群を時系列順にすべて連結.

この手法の要約例を表 7 に示す。前述のルールの通り,推定 ラベルが 1 となっているものが取り出され,タグなど("(D その)" や末尾の "+")削除され,結合されたものが要約結果と なる。

4.2.2 生成型手法

生成型手法には mBART を用いる. mBART は、多言語の大規模コーパスに Lewis ら [13] の BART を適用して学習したニューラル翻訳モデルである. ここで要約を単一言語内の長さを変える翻訳と置き換える. 生成型手法では、抽出型と同様に、推定された重要発話からタグなどを除去し、それを時系列に連結させたものを mBART の入力とする. すなわち、抽出型手法の出力を mBART で長さ制約のもと整形することが生成型手法となる. この手法で生成された要約例を表 8 に示す.

表 8 に生成型手法の例を示す. 重要発話("1")と推定された発話そのものではなく, mBART が文章として再構成したものが要約として出力されている.

5. 実 験

本節では前節で述べた提案手法の実験について説明する. 1 つめの実験は、SVM による重要発話の推定である. この結果は、要約処理のための入力を得るための前処理に相当する. 2 つめはその重要発話推定の結果を利用した 2 つの要約手法の結果の検証である.

5.1 重要発話推定

構築した 3つのデータ (A1, A2, M) のうち,データ M に関して重要発話推定を 9 対話交差検証(対話単位で交差検証)で行った. ハイパーパラメータは訓練データにおいて Grid Search で最適化を行った(表 9). また,重要発話と非重要発話のデータ数がアンバランスなデータセットであるため,class weight で

表 8 生成型要約例.

発話	推定	要約結果 (生成)
で +	0	
これは (F その) 和屋も共通してて	1	
メニューが少ないってところがあるのね/	1	
なるほどね/	0	和屋が閉店した理由に
で、基本的に +	0	「女性受けが悪い」
和屋が +	0	との声が上がっている。
(F その一何)+	1	閉店した理由として、
(F その) 閉店した理由ってのが	1	メニューが少ないことが
やっぱり女性受けが悪い +	1	あげられている
ふんふんふん/	0	
ってとこもあんじゃん (Q)/	1	
どっちかっていうと (D これ) この店	1	
ほんと +	1	
(F その一何) おかわりも自由だったり	1	
して、男性向けの店じゃん (Q)/	1	
ふんふんふん/	0	

表 9 交差検証ごとのハイパーパラメータ.

テスト	パラメータ		
データ	C	gamma	kernel
20150313_C1	10	0.0001	rbf
20150320_C1	1	0.0001	rbf
20150320_C4	10	0.0001	rbf
20150323_C3	1	0.0001	rbf
20150326_C1	1	0.0001	rbf
20150326_C2	1	0.0001	rbf
20150326_C4	100	0.0001	rbf
20150327_C2	1	0.0001	rbf
20150327_C3	1	0.0001	rbf

表 10 重要発話推定の評価値.

正解率	適合率	再現率	F1
0.708	0.507	0.749	0.596

ラベルの重みを "1":"0" = 3:1 とした.

重要発話推定の正解率、適合率、再現率、F1 値を表 10 に示す。表 10 の再現率の値から、要約作成に必要な重要発話の多くを識別できており、その網羅性が高いことが分かる。適合率は 0.507 と必ずしも高くないが、重要発話推定の結果が要約生成の入力であることを考えると、再現率が高いことが望ましく、好ましい結果であるといえる。

実際の推定結果ついて分析をすると、文字数の多い発話や頻出単語を含む発話は正しく推定できている傾向があった.一方で、発話の長さ自体は長い(文字数が多い)が、発話の内容がほぼ相槌のみである場合に、重要発話であると誤って推定されている傾向があった.より高精度な重要発話推定のため、素性の改良が必要だろう.また、SummaRuNNer [14] など、ニューラルネットワークベースの手法の適用についても考察が必要である.

5.2 要約作成

本節では 4.2 節で生成した要約の評価を行う. 重要発話推定の有効性を検証する目的で、セグメント内の全発話を mBART によって要約する手法をベースラインとした. ただし、ベースラインにおいてもタグや記号を除去による前処理は行う.

ベースラインおよび生成型手法のための mBART のモデルは、livedoor $^{(\+ i 3)}$ の 3 行要約データ約 10000 件で学習されたもの $^{(\+ i 4)}$ を利用した。出力長のトークン数の制限は 100 とした。

評価は無作為にサンプリングした 100 個のセグメントに対して、3名の被験者が行った。セグメントごとに、3つの要約(ベースライン、抽出型、生成型) それぞれに対して評価をする。評価尺度は2つあり、1つは要約された情報の正しさ、もう1つは可読性である。

まず情報の正しさとは、元の発話の内容をどれだけ過不足なく保持できているかどうかを表す尺度である。生成した要約とベースラインについての正しさを $0\sim3$ の4段階の数値で評価する

3 : 原文の内容を過不足なく正しく保持している

2 : 原文の内容を概ね正しく保持している

1:原文の内容をある程度保持している

0 : 原文の内容を十分に保持していない

3つの要約に対して、正しさのスコアが等しい要約が複数存在 しても良いこととした。可読性とは、要約が日本語として読み やすいかどうかを表す尺度である。生成した要約とベースライ ンについての可読性の高い順に順位付けし、1位を3点、2位 を2点、3位を1点とした。

3名の被験者の数値の平均を表 11 に示す.表 12 に要約元のセグメント内の発話と 3 つの手法の出力,ある被験者が実際に付けた点数を示す(注5).

表 11 より、抽出型は正しく元の情報を含んでいることが分かる。これは要約が原文に含まれている情報のみで構成されていることと、5.1 節で行った重要発話推定の再現率が高いことに起因している。生成型とベースラインは生成型の手法であるため、正しさは低くなり、可読性は高い結果となった。生成型とベースラインの違いは入力に重要発話推定を用いるかどうかである。したがって、生成型はベースラインと比較して、重要な情報を維持した状態で可読性が向上することを予想していたが、可読性の向上はわずかであった。これは、重要発話推定で抽出された発話群は抽出型の結果から分かるように、本来つながりのない複数の発話が連続しており、不自然な入力が不自然な要約の生成に繋がったものと考えられる。

全体として、抽出型は情報の正しさの面で、生成型は可読性の高さの面で優れているという予測通りの結果が得られた。可読性が高く、情報の正しさを保証するためには、Pointer-Generator [15] に代用されるような原文の情報をある程度保持する枠組みの導入などが必要となる。

6. おわりに

本研究では重要発話推定を用いた複数人対話の要約手法を提案した。まず、Kyutech コーパスに対してトピックごとの参照要約の作成と参照要約に基づく重要発話のアノテーションを

⁽注3): http://www.livedoor.com/

⁽注5):実際に被験者に提示するときには手法名の部分は削除している.

表 11 正しさと可読性の評価値.

手法名	重要発話推定	要約作成法	正しさ	可読性
ベースライン	無	mBART	1.54	2.28
抽出型	有	連結	2.44	1.40
生成型	有	mBART	1.30	2.32

表 12 要約評価例.

発話内容					
で +					
これは (F その)) 和屋も共通しててメニューが少ないってと	ころがあ	るのね/		
なるほどね/					
で、基本的に	+				
和屋が +					
(F その一何)-	+				
(F その) 閉店した理由ってのがやっぱり女性受けが悪い +					
ふんふんふん/					
ってとこもあん	しじゃん (Q)/				
どっちかってい	どっちかっていうと (D これ) この店ほんと +				
(F その一何) おかわりも自由だったりして、男性向けの店じゃん (Q)/					
ふんふんふん/	ふんふんふん/				
エナク	亜 奶山穴	×	=T =± A+		

手法名	要約内容	正しさ	可読性
ベースライン	和屋が閉店した理由について、筆者が考察 している。女性受けが悪いことや、男性向 けの店であることがあげられている。閉店 した理由について、筆者は「おかわりも自 由だったりして」と述べている。	2	2
抽出型	これは和屋も共通しててメニューが少ないってところがあるのね閉店した理由ってのがやっぱり女性受けが悪いってとこもあんじゃんどっちかっていうとこの店ほんとおかわりも自由だったりして、男性向けの店じゃん	3	1
生成型	和屋が閉店した理由に「女性受けが悪い」 との声が上がっている。閉店した理由とし て、メニューが少ないことがあげられてい る。	2	3

行った. 参照要約については,2名の作業者によるもの(A1およびA2)とその2つを統合したもの(M)を作成した. それぞれに対して重要発話のアノテーション済みデータを構築した.

その後、作成したデータに対して、SVM による重要発話推定のモデルを構築した。得られた重要発話を入力として2種類の要約手法を適用し、情報の正確さ(要約前の情報を適切に含んでいる度合い)と可読性について被験者による評価を行った。抽出型は正しさが保持されやすい一方、生成型は可読性が高くなる傾向にあることがわかった。

今後の課題として、生成型手法の情報の正確さを上げるための工夫が必要となる。また、今回はトピックタグに基づくトピック単位の要約を対象としたが、話者ごとの重要発話や議論の対象ごと(Kyutech コーパスでは候補店ごとなど)の要約など、要約するべき対象は様々あり、これらへの対応も今後の課題である。

謝辞 本研究は科研費 20K12110 の助成を受けたものです.

文 献

- Takashi Yamamura and Kazutaka Shimada. Annotation and Analysis of Extractive Summaries for the Kyutech Corpus. In Proceedings of the 11th International Conference on Language Resources and Evaluation, pp. 3216–3220, 2018.
- [2] Takashi Yamamura, Kazutaka Shimada, and Shintaro Kawahara. The Kyutech Corpus and Topic Segmentation Using a Combined Method. In *Proceedings of the 12th Workshop on Asian Language Resources*, pp. 95–104, 2016.
- [3] Inderjeet Mani. Automatic Summarization. 2001.
- [4] Alexander M. Rush, Sumit Chopra, and Jason Weston. A neural attention model for abstractive sentence summarization. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pp. 379–389, 2015.
- [5] Shasha Xie, Yang Liu, and Hui Lin. Evaluating the Effectiveness of Features and Sampling in Extractive Meeting Summarization. In Proceedings of IEEE Spoken Language Technology Workshop, pp. 157–160, 2008.
- [6] Guokan Shang, Wensi Ding, Zekun Zhang, Antoine Tixier, Polykarpos Meladianos, Michalis Vazirgiannis, , Jean-Pierre Lorre. Unsupervised Abstractive Meeting Summarization with Multi-Sentence Compression and Budgeted Submodular Maximization. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, pp. 664-674, 2018.
- [7] Manling Li, Lingyu Zhang, Heng Ji, Richard J. Radke. Keep Meeting Summaries on Topic: Abstractive Multi-Modal Meeting Summarization. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pp. 2190–2196, 2019.
- [8] 山村崇, 徳永陽, 嶋田和孝. 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. 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), pp. 43–48, 2015.
- [9] Michel Galley, Kathleen McKeown, Eric Fosler-Lussier, and Hongyan Jing. Discourse Segmentation of Multi-Party Conversation. In Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics, pp. 562–569, 2003.
- [10] Chin-Yew Lin. ROUGE: A Package for Automatic Evaluation of Summaries. In *Text Summarization Branches Out*, pp. 74–81, 2004.
- [11] Vladimir Vapnik. The Nature of Statistical Learning Theory. Springer-Verlag, 1995.
- [12] Yo Tokunaga and Kazutaka Shimada. Multi-party conversation summarization based on sentence selection using verbal and nonverbalinformation. In 2nd International Conference on Smart Computing and Artificial Intelligence, pp. 464–469, 2014.
- [13] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. BART: Denoising Sequenceto-Sequence Pre-training for Natural Language Generation, Translation, and C omprehension. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 7871–7880, 2020.
- [14] Ramesh Nallapati, Feifei Zhai, and Bowen Zhou. Sum-maRuNNer: A recurrent neural network based sequence model for extractive summarization of documents. In Proceedings of the 31th AAAI Conference on Artificial Intelligence, pp. 3075–3081, 2017.
- [15] Abigail See, Peter J. Liu, and Christopher D. Manning. Get to the point: Summarization with pointer-generator networks. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, pp. 1073–1083, 2017