一般社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
IEICE Technical Report
NLC2022-13,SP2022-33(2022-11)

前後文脈を用いた対話文の言い淀み検出

中島　寛人†　　嶋田　和孝††

† 九州工業大学 大学院情報工学府 情報創成工学専攻 〒820-8502 福岡県飯塚市川津 680-4
†† 九州工業大学 大学院情報工学研究院 知能情報工学研究系 〒820-8502 福岡県飯塚市川津 680-4
E-mail: †nakashima.hiroto810@mail.kyutech.jp, ††shimada@ai.kyutech.ac.jp

あらまし　ASR に代表される音声認識技術の進歩により正確な書き起こし文が得られるようになった．しかしながら，言い淀みやフィラーといった話し言葉に特有の表現は，依然として書き起こし文の可読性や後段の処理への応用性を下げてしまう一因となっている．中でも言い淀みは文意に影響せず，定型も持たない表現であり，これを人手で取り除くことはコストが掛かる．したがって自動的に検出および整形処理を行うことが求められている．近年の機械学習モデルを使用した言い淀み検出では文脈情報を利用した検出が盛んである．しかし，複数人の話者が存在する対話文では，ある発話者の発話に対して他発話者の割り込みによる中断が発生し，得られる文脈情報も少なくなる傾向がある．そこで本研究では，モデルへの入力文を結合や生成によって前後に拡張し，文脈情報を補完する手法を提案する．実験では，文脈補完の対象や補完の窓幅，手法によっていくつかの入力データの拡張を行う．これらのデータで学習されたモデルによって言い淀み検出に取り組むことで，提案手法の有効性を確認すると共に，より良い文脈補完の手法について検討する．

キーワード　言い淀み検出，非流暢性，対話文生成，文脈補完

Dialogue disfluency detection using context

Hiroto NAKASHIMA† and Kazutaka SHIMADA††

† Department of Creative Informatics, Kyushu Institute of Technology
680-4 Kawazu, Iizuka, Fukuoka 820-8502, JAPAN
†† Department of Artificial Intelligence, Kyushu Institute of Technology
680-4 Kawazu, Iizuka, Fukuoka 820-8502, JAPAN
E-mail: †nakashima.hiroto810@mail.kyutech.jp, ††shimada@ai.kyutech.ac.jp

Abstract　Recent automatic speech recognition (ASR) techniques have been improved by a large amount of training data and machine learning, such as deep learning technology. Problems in the outputs from ASR are not only recognition errors but also outputs caused by disfluency from speakers. It is difficult to remove them automatically, and removing them by hand is costly. In this paper, we propose a disfluency detection model with BERT. The model utilizes context information of target utterances. We introduce two types of context information. The first one is real utterances that appear around the target utterance. We compare several sequence lengths of the previous and following utterances. The second one is a generated utterance by GPT-2. Our model adds the utterance generated from the target utterance as the following context. In the experiment, the long sequence improves the disfluency detection accuracy, and real context outperforms generated context.

Key words　disfluency detection, disfluency, dialogue generation, context complement

1. はじめに

近年，音声認識技術の進歩により，話し言葉の正確な書き起こし文を得ることができるようになっている．しかし，話し言葉には言い淀みやフィラーなど様々な話し言葉に特有の表現が含まれており，認識誤りの無い完全な書き起こし文であっても，可読性や後段の処理への応用性といった観点でのノイズになり得る．例えば，以下のような例文を考える．

(1) このように，**えー**高度の**たか**，高い場所においても

(2) でもこんな急に**高い**，高くなってたら違和感凄くないか

例文 (1) における「えー」は発話の間を埋める表現のフィラーである．また，「たか，高い」の「たか」の部分のように，単語

This article is a technical report without peer review, and its polished and/or extended version may be published elsewhere.
Copyright ©2022 by IEICE

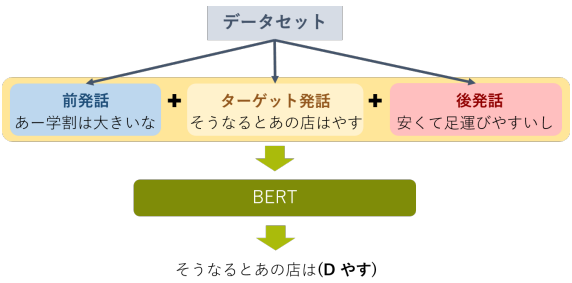

図 1　提案手法の一例．実際の前後発話を結合してターゲットとなる発話の前後文脈を補完している．

を途中まで言いかけて中断する表現が言い淀みである．これらの表現は文意にも関係しないため，話し言葉を処理する際には事前の整形が必要不可欠である．フィラーは「えー」「あの」のような定型表現を持っているため，比較的容易に対処できる．一方で言い淀みは，例文 (1) では言い淀みではなかった「高い」が例文 (2) では言い淀みとなっているように，言い淀んだ単語や状況によって多様に変化するため，人手で整形する作業は労力を要する．したがって，言い淀みの自動的な整形や検出は，話し言葉を対象とする様々な言語処理タスクの前段処理として重要なタスクとなっている．

話し言葉の中でも，話者が複数人存在する対話文では，割り込みや相槌による発話の中断が発生する．このような中断の影響で，対話文においては個々の発話の文長が短くなってしまう傾向がある．言い淀み検出では深層学習モデルによる文脈情報を考慮した検出の有効性が確認されているが，発話が短い状態では文脈を十分に得られない．したがって，途切れた文脈を補完する必要がある．

そこで，本論文では深層学習モデルを使用した対話文の言い淀みの検出を行うにあたって，モデルに入力する発話の文脈を前後方向に補完する入力データの拡張手法を提案する．提案手法の一例を図 1 に示す．我々は先行研究 [1] で，文脈情報の補完において前後に同一話者の発話を結合することで言い淀み検出の精度が向上することを示している．

本論文では大きく分けて二つの入力データの拡張手法を導入する．一つ目は，先行研究でも行った実際の発話を利用する入力データの拡張である．元の対話文で実際に対象の発話の前後で行われた発話のみを利用して，補完する発話の選択基準や補完する窓幅を変更しながら入力データの拡張を行う．二つ目は生成モデルを併用したデータ拡張の手法である．この手法では，対象となる発話を入力として生成モデルによって得られた後文脈を結合する形で入力データの拡張を行う．その後，作成した各入力データによって分類モデルを学習させて言い淀み検出実験を行うことで，対話文言い淀み検出における提案手法の有効性の検証や，より良い文脈補完手法の検討を行う．

2. 関 連 研 究

言い淀みの検出において，著名な手法として RIM (Repair Interval Model) [2] が挙げられる．これは，言い直しのような自己修復 (self-repairing) が発生する場合は被修復部，言い淀み，修復部の系列で表れると仮定し，文中からこれを探索することで検出を行う手法である．しかし，船越ら [3] は，RIM による自己修復検出の問題点として被修復部と言い淀みの間に流暢文が入るパターンのように，RIM によっては検出できない構造の存在を指摘している．また，尾嶋ら [4] は自己修復や言い淀みの周囲に，繰り返しや定型表現のような特定の文脈的特徴が存在することを示している．したがって，言い淀み検出を行う際にはこのような文脈的特徴を考慮する必要がある．

文脈情報を考慮した検出においては様々なアプローチでの取り組みがなされている．従来の日本語話し言葉的特徴の検出に関する研究では，SVM ベースのテキストチャンカである YamCha [5] を使用し，素性として文脈情報を使用して各形態素に対してチャンク付けを行う手法がある [4]．近年は深層学習モデルを用いた研究も盛んであり，Johann ら [6] は，事前学習モデルによる言い淀みの検出は英語の話し言葉の書き起こし文に対して有効であることを示している．BERT [7] のようなモデルでは文脈情報を考慮することで従来よりも高精度な言い淀みの検出が可能になる．しかし，対話文においては割り込みなどの影響で不安定な文長への対処を行わなければ，文脈情報も不完全なものとなってしまう．したがって，文脈を補完できるような発話周辺の情報が必要である．我々の先行研究 [1] では，言い淀み検出のターゲットとなる発話に対して，前後に存在する同一話者による発話を結合することでデータを拡張し，これをモデルへの入力とすることで，対話文における言い淀み検出精度の向上が示されている．しかし，同一話者発話が他の周辺発話より文脈補完に適しているかや，補完する範囲の長さについては十分に議論されていない．そこで，本論文では対話文における文脈補完を行う際に補完の対象とする発話の選択や，結合する範囲の増加などについてさまざまなデータを作成し，これらの検証を行う．

また，本論文では欠落のある文への補完について取り扱うが，このような文脈欠如に対する補完の試みは，主にリアルタイム処理を必要とする分野でも行われている．Skidmore ら [8] は，言語学習者の発話をリアルタイム処理するにあたって，笑いや休止の情報を組み合わせることが言語学習者の不安定な言語情報や，リアルタイム処理による不完全な入力を補うことができることを示した．また，Rohanian ら [9] はリアルタイム処理の際に断片的になってしまう入力文に対して，現在までに得られた文の続きを生成モデルを利用して仮文脈として予測することで文脈の補完を行い，言い淀みやフィラーの検出における精度の向上を示した．本論文での対象は音声認識により書き起こされたテキストであり，次発話が未知であったり，欠落している状態ではない．しかし，対話においては発話そのものが割り込みなどで途切れてしまう場合もありうる．また，実際の後続発話よりも言語モデルによって生成される文脈的にはきれいな発話の方が，言い淀み検出には有効である可能性もある．そこで，本論文では生成モデルによる文脈補完の有効性について検証する．

表 1　CSJ に付与されているタグの一例．本論文で使用する言い淀み (D タグ) は言い淀んでいる部分のみを範囲としている．

タグ	例文
言い淀み	…原型と変形の **(D しょうちゅし)** 聴取時間の差…

表 2　前処理を行った発話群の一例 (ただし，話者交替を明示するために話者情報を付与している)．相槌による発話の中断がみられる．

話者	発話
A	(F あのー)[SEP] 個別の部品がたくさん使われてたんで
B	(F うーん)(F うん)(F うん)(F うん)
A	(D わ)(D す) そういう意味では再利用が (D し) し易かったですね
B	(F ふーん)
A	(F あの) ＩＣばかりの [SEP] プリント基板だと
A	何か (F ん) もうどうしようもないんですけど
B	(F うーん)[SEP] なるほどね [SEP] そうですか
B	最近何かこう作ったものは

3. データセット

本論文では書き起こし文のデータセットとして，話し言葉コーパスにおける書き起こし文にあたる転記テキストを使用する．『日本語話し言葉コーパス』(以降 CSJ) [10] は，全国共通語のモノローグ音声を主な対象とした，全 3302 講演の大規模な話し言葉コーパスである．各独話・対話データは転記テキストの形で使用できる．転記テキスト内には言い淀みを含む様々な言語現象に対して，表 1 のように対応する記号ととも一定範囲を囲むことでタグ付けが行われている．そこで，本論文ではこのタグ情報を正解ラベルの作成に使用する．本論文では，CSJ に含まれる音声タイプの中からインタビューや自由対話からなる全 58 対話の転記テキストを対象として，対話による話し言葉データとして扱う．

また，CSJ の転記テキストは，原則「200 ミリ秒以上のポーズで挟まれた範囲」によって発話が区切られている．しかし本論文では，割り込みによる中断で発生する文脈の不足を補うことを主な目的としているため，話者転換を伴わないような単位では発話を区切らない．具体的には，前処理として割り込みの発生である話者転換か，発話の完遂である終止形文末が発生するまで，同一話者の発話を結合したものを使用する．ただし，一定のポーズがあるということ自体には，特に言い淀みという現象に関しては関連がある可能性がある．そこでポーズがあった情報を付与することとする．具体的には，BERT で利用される “[SEP]” という特殊トークンをそのマーカとして利用する．このような結合処理の上で，今回使用する対話に含まれる発話は表 2 のようなものになっている．

4. 提 案 手 法

本節では，3 節のデータセットを対象に入力文脈を補完する手法と，利用する分類モデルについて説明する．

4.1 文脈の補完

本節では，文脈を補完する方法について説明する．本論文では，対話文の各発話 (以降，ターゲット発話) において途切れた前後の文脈を補完するにあたって，前後双方の発話について実際の発話を利用する方法と，前発話は実際の発話を，後発話部分は言語モデルによって生成して利用する方法の 2 つを用いる．それぞれについて順に説明する．

表 3　使用する入力データと内訳．

データ	内容
Same4	同一話者発話による 4-gram
Same5	同一話者発話による 5-gram
Same7	同一話者発話による 7-gram
Other3	非同一話者発話による 3-gram
Near3	近傍発話による 3-gram

4.1.1 実際の発話による補完

対話において相槌のような割り込みが行われた場合，割り込み前後の同発話者による発話 (以降，同一話者発話) では，内容が維持されることがある．私たちの先行研究では，このような同一話者発話が最も文脈を維持できているという仮定のもと同一話者発話による補完を行っており，本論文でもこれに従って拡張を行っている．

しかし一般に，話題の遷移は流動的に行われるため，隣り合っている発話 (以降，近傍発話) ほど文脈情報を維持していると考えられる．また，割り込み発話がターゲット発話との関連性を維持していた場合最も文脈を保持できているのは割り込んだ他者の発話 (以降，非同一話者発話) である．そこで，ターゲット発話に対して周囲で様々な話者によって行われた近傍発話や，ターゲット発話と異なる話者による非同一話者発話を利用して文脈情報を補完したデータも使用する．

本論文で使用する入力データ群は表 3 に示す通りである．各データ名は結合された発話の種類とその *n*-gram を表しており，例えば Same5 であれば同一話者発話の 5-gram である．つまり，対象となる発話に対して前 2 発話と後 2 発話をそれぞれターゲット発話の前後に結合している．また，我々の先行研究において，前発話のみの補完より後発話のみの補完の方が高精度となる傾向があった．そこで本論文では不均衡な *n*-gram となっている Same4 では後発話側の補完を優先し，前 1 発話と後 2 発話をそれぞれターゲット発話の前後に結合している．

4.1.2 生成モデルによる後方向文脈の補完

Rohanian ら [9] は，逐次処理の影響で途切れている発話の後文脈を言語モデルによって予測することで言い淀み検出の精度を向上させた．割り込みが発生して発話の中断や話題の転換により文脈が維持されなかった場合，その発話も途切れていると考えられる．そこで，本論文では Rohanian らの手法を本タスクに応用し，言語モデルの次単語予測によって途切れた発話の後文脈を生成する．生成は前文脈を考慮して行われるため，話題転換などで必ずしも文脈を維持していない実際の発話より，文脈的に生じうる安定した文脈の補完が行える可能性がある．

そこで，本節では言語モデルを用いた後文脈の補完を行いデータを拡張する．図 2 に示すように，前文脈の補完には 4.1.1 節の Same5 などと同様に同一話者発話を結合するが，後文脈の補完では言語モデルを用いて予測し，これを結合する．言語モ

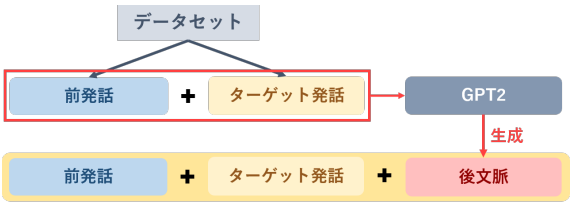

図 2　GPT2 を用いた補完の例．前発話とターゲット発話からなる前文脈を入力として，続きの文を後文脈として生成して結合する．

デルには GPT2 [11] を用いる．GPT2 は事前学習知識を利用して，ファインチューニングすることで様々なタスクへと応用することができる汎用言語モデルである．今回は入力文を文の途中で途切れているとしたとき続く文を予測する言語モデルとして使用する．本論文では rinna 社が公開している日本語事前学習済み GPT2 モデル[1]を用いた．ファインチューニングには同一話者発話による 3-gram の対話文を使用し，対話文の後文脈予測が行えるモデルとした．GPT2 の入力には，ターゲット発話の前に直前の同一話者発話を結合したものを用い，生成時のトークン数は (入力部のトークン長)+10 とした．

4.2　トークナイズとラベリング

本節では，実際にモデルの入力とする際のデータへの前処理について説明する．

トークナイズとラベリングの例を図 3 に示す．後述するように本論文では分類器として BERT を使用するため，入力する発話を単語のような小さな単位で分割されたトークン列に変換する．この際，トークン列長が 5 未満および 512 以上となる発話はデータセットに含めない．また，正解データのラベリングには，従来の研究でも使用されてきたチャンキング形式を使用した．チャンクの始点に B-タグ，チャンクに属している部分に I-タグ，チャンクに属さない部分に O-タグを付与する IOB2 (Inside-Outside-Beginning) 方式に従って，言い淀み部分へのラベリングを行う．本論文で使用したラベルを表 4 に示す．各トークンについて IOB2 方式に対応するようにラベリングを行い正解ラベルデータとする．ただし，提案手法によって補完された範囲のトークンに対しても同様にラベリングを行うと，文脈補完を行うモデルが学習時に有利になってしまう．そこで，図 4 のように補完された範囲のトークンに対しては，本来言い淀みチャンクである部分も含めて，全てチャンクの外側を示す O-タグでマスクしてラベリングする．このように作成されたトークン列と対応する正解ラベルデータの組を入力データセットとして使用する．

4.3　分類モデル

本節では，BERT [7] を使用した話し言葉的特徴の検出を行う系列ラベリングモデルについて説明する．BERT は事前学習済み言語モデルであり，ファインチューニングを行うことで多くの言語処理タスクへ応用することができる．本論文では東北大学が公開している事前学習済みモデル[2]を使用する．

(注1)：https://huggingface.co/rinna/japanese-gpt2-medium
(注2)：https://huggingface.co/cl-tohoku/bert-base-japanese

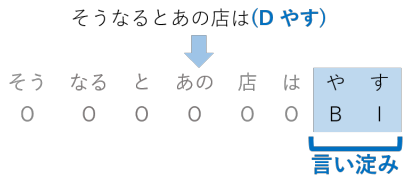

図 3　入力データ前処理の例．トークナイズによって「やす」という文字列からなる言い淀み部は 2 トークンに分割されたため，それぞれに B と I のラベルが付与されている．

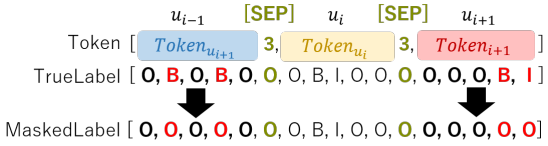

図 4　補完発話へのマスクラベリング．ターゲット発話である u_i のラベルのみ真のラベル情報を用い，前後の u_{i-1}，u_{i+1} のラベルはマスクした MaskedLabel を正解ラベルとして使用する．

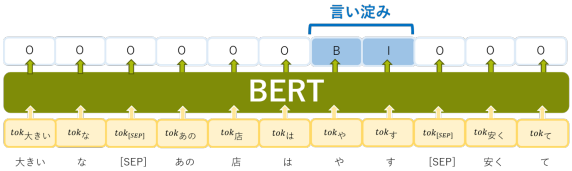

図 5　分類モデルの入出力．tok_{word} はトークナイズによって得られた，文字列に対応するトークンを表している．出力されている各ラベルの詳細は表 4 の通り．

表 4　使用するラベルの一覧．

ラベル	説明
B	言い淀みチャンクの始点
I	言い淀みチャンクの内側 (始点以外)
O	チャンクに属さない部分

本論文では，BERT による系列ラベリングを行う．この際使用する BERT の構造を図 5 に示す．文章を分割したトークン列と，各トークンに対応する正解ラベル列の組を使用した学習により，BERT へ入力したトークンの出力に対して softmax 関数を適用することで推定ラベル列を得ることができるようなモデルにファインチューニングする．トークン化の際には，通常文頭に “[CLS]” という特殊なトークンが付与され，クラス分類などのタスクに使用される．しかし，本論文では分類ラベルは文章毎ではなくトークン毎に付与するためこれを除いている．また，本論文で行われる発話同士の結合処理では，文の区切りを表す “[SEP]” トークンを区切り文字として使用する．

実際のモデルへの入出力を図 5 に示す．例における出力では「やす」を言い淀みのチャンクとして検出している．

5.　実　　験

本節では，4.1.1 節および 4.1.2 節で示した手法によって拡張された入力データを使用して，4.3 節の分類モデルで学習と実験を行い，補完した文脈情報の有効性について検証する．

5.1 実験設定

実験では，作成したデータ毎にモデルを作成し，補完対象の選択や文脈補完の程度・補完手法による精度の違いを検証する．ベースラインには，文脈補完を行わないデータによって学習されたモデルを $BERT_{base}$ として，同一話者発話による 3-gram 文脈補完で構成されたデータによって学習されたモデルにあたるモデルを $BERT_{Same3}$(先行研究) として使用する．これらを含め，今回行う実験と各実験に使用するモデルを表 5 に示す．モデルの添え字は基本的には学習に使用した表 3 のデータ名に対応しているが，$BERT_{GPT}$ は 4.1.2 節の後方向の補完に生成モデルを用いて拡張されたデータを学習に使用したモデルとなっている．

モデルの訓練の際に，同じ対話のデータが訓練・テストの双方に存在すると，登場する品詞の種類などによって表層が変化する言い淀みの判定には有利となってしまう．そこで，本論文では各データセットを対話単位で区切り，対話単位での k 分割交差検証を行う．分割数 k は 9 とし，訓練：テストが 8：1 となるように対話単位で分割して 9 分割交差検証を行う．なお，各モデルについて，学習のエポック数は 10，バッチサイズは 32 とした．この際，言い淀みの含まれない文字列は学習の妨げになるため，訓練データでは言い淀みを含まない文章を除外する．

5.2 実験結果

まず，補完対象とする発話の選択による影響について，同一話者発話，非同一話者発話，近傍発話それぞれの 3-gram 補完モデル $BERT_{Same3}$，$BERT_{Other3}$，$BERT_{Near3}$ について比較を行う．表 6 に実験結果を示す．ベースラインの $BERT_{base}$ と比較すると各モデルで精度の向上が見られるが，$BERT_{Same3}$ がより良い F 値を獲得している．したがって，対話においては単純に近傍に存在する発話やよりも周囲の文の中では同一話者による発話がより元の文脈を保持できていると考えられる．

次に，文脈補完の窓幅による影響について，同一話者発話を対象として n-gram 文脈補完モデルから $BERT_{Same3}$，$BERT_{Same4}$，$BERT_{Same5}$，$BERT_{Same7}$ の比較を行う．表 7 に実験結果を示す．文脈補完の程度が増加するに従って，モデルの F 値も向上しており，利用できる文脈情報が増えるほど言い淀みの検出には良い影響を与えると考えられる．

最後に，文脈補完の手法による影響について，前後を実際の同一話者発話によって補完したデータによるモデル $BERT_{Same3}$ と，後発話相当を GPT2 によって補完したデータによるモデル $BERT_{GPT}$ の比較を行う．表 8 に実験結果を示す．文脈補完を行わない $BERT_{base}$ よりは良い F 値を得ているため，文脈の補完はある程度行われていると考えられる．一方で $BERT_{Same3}$ を F 値で上回ることはできなかった．同じく 3-gram 補完による結果である表 6 の $BERT_{Other3}$ や $BERT_{Near3}$ と比較しても低い F 値となっており，予測による文脈よりも実際の文脈の方が言い淀みの検出には適していることが示された．

表 5　実験内容と各実験で用いたモデル．

実験	モデル名
ベースライン	$BERT_{base}$
	$BERT_{Same3}$
補完対象の選択	$BERT_{Other3}$
	$BERT_{Near3}$
補完の窓幅	$BERT_{Same4}$
	$BERT_{Same5}$
	$BERT_{Same7}$
補完の手法	$BERT_{GPT}$

表 6　補完対象とする発話の選択による比較結果．

モデル	Pre.	Rec.	F1
$BERT_{base}$	0.223	0.564	0.321
$BERT_{Same3}$	**0.462**	0.537	**0.496**
$BERT_{Other3}$	0.421	0.556	0.479
$BERT_{Near3}$	0.399	**0.608**	0.482

表 7　文脈補完の窓幅による比較結果．

モデル	Pre.	Rec.	F1
$BERT_{base}$	0.223	0.564	0.321
$BERT_{Same3}$	0.462	0.537	0.496
$BERT_{Same4}$	0.529	**0.675**	0.594
$BERT_{Same5}$	**0.569**	0.638	0.602
$BERT_{Same7}$	0.566	0.647	**0.604**

表 8　文脈補完の手法による比較結果．

モデル	Pre.	Rec.	F1
$BERT_{base}$	0.223	**0.564**	0.321
$BERT_{Same3}$	**0.462**	0.537	**0.496**
$BERT_{GPT}$	0.343	0.548	0.422

6. おわりに

本論文では，対話文での言い淀み検出を対象として，割込みなどで中断が発生して不足する文脈情報を補う目的で実際の発話や生成モデルを用いた文脈の補完を行った．補完の対象とする発話や補完の窓幅，補完手法について様々な観点から実験を行い，最適な補完手法について検証を行った．先行研究で利用されていた同一話者発話の文脈補完における有効性を確認すると共に，より窓幅の大きな n-gram データを使用することで精度の向上を確認した．また，生成モデルを使用した文脈補完についても検証を行ったが，実際の発話を用いたものより良い結果は得られなかった．

今後は，本論文で最も良い結果を得られた実際の発話を使用する同一話者発話の n-gram 補完モデルについては，n-gram の上限検証や，補完の対象とする発話の選択において同一話者発話と近傍発話の双方を組み合わせるなど，より適切に文脈を補完できるような手法について検討を行う．

また，今回使用した CSJ の対話文では，対話の種類はインタビューや自由対話となっており，すべての対話において話者が 2 人だけであった．対話中に発生する中断の回数は，話者が多いほどより頻繁に発生すると考えられ，また会議やディベートのような意見の衝突が発生する場でもより顕著になると考えられる．そこで，3 人以上の話者による意思決定タスク対話やディベートが収録されている Kyutech コーパス [12] や Kyutech ディ

ベートコーパス [13] のように，より多くの中断が発生するコーパスを使用しての実験も行いたいと考えている．我々の研究室では，先述のコーパスのような複数人議論を対象としたファシリテーションの研究が存在する [14]．議論のファシリテーションではリアルタイムな処理が求められる．したがって，本論文で良い結果を得られた実際の発話による文脈を用いる方法では，未来の発話である後文脈は使用することができず，精度が著しく低下する可能性がある．そこで，このような状況下での言い淀み処理における，生成モデルを利用した手法の有効性についても更なる検証が必要であると考える．

本論文では，生成モデルである GPT2 のファインチューニングは CSJ の対話文で行った．より文脈を補完するために，話し言葉の中でも成文が得られやすい独話文や，書き言葉を用いたファインチューニングを行うことを予定している．また，本論文では生成時のトークン長を 10 としているが，これについてもより最適な値を探索する必要があると考える．T5 [15] や GPT3 [16] のような異なる生成モデルの利用も検討し，生成モデルを使用した文脈の補完についても引き続き取り組んでいく予定である．

謝辞　本研究は科研費 20K12110 の助成を受けたものです．

文　　献

[1] 中島寛人, 嶋田和孝. 話し言葉における双方向文脈補完を用いた言い淀み検出. 電子情報通信学会九州支部 第 30 回学生会講演会, D-44, 2022.

[2] Christine Nakatani and Julia Hirschberg. A speech-first model for repair detection and correction. In *Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics*, pp. 46–53, 1993.

[3] 船越孝太郎, 徳永健伸, 田中穂積. 音声対話システムにおける日本語自己修復の処理. 自然言語処理, Vol. 10, No. 4, pp. 33–53, 2003.

[4] 尾嶋憲治, 河原達也, 秋田祐哉, 内元清貴. 話し言葉の整形作業における削除箇所の自動同定. 情報処理学会研究報告自然言語処理 (NL), Vol. 185, pp. 85–91, 2008.

[5] Taku Kudo and Yuji Matsumoto. Chunking with support vector machines. In *Proceedings of Second Meeting of the North American Chapter of the Association for Computational Linguistics*, pp. 192–199, 2001.

[6] Johann C. Rocholl, Vicky Zayats, Daniel D. Walker, Noah B. Murad, Aaron Schneider, and Daniel J. Liebling. Disfluency Detection with Unlabeled Data and Small BERT Models. In *Proceedings of Interspeech 2021*, pp. 766–770, 2021.

[7] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 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.

[8] Lucy Skidmore and Roger K Moore. Incremental disfluency detection for spoken learner english. In *Proceedings of the 17th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2022)*, pp. 272–278, 2022.

[9] Morteza Rohanian and Julian Hough. Best of both worlds: Making high accuracy non-incremental transformer-based disfluency detection incremental. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pp. 3693–3703, 2021.

[10] 前川喜久雄. 『日本語話し言葉コーパス』 の概要. 日本語科学, Vol. 15, pp. 111–133, 2004.

[11] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. Language models are unsupervised multitask learners. *OpenAI blog*, Vol. 1, No. 8, p. 9, 2019.

[12] 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 (ALR12)*, pp. 95–104, 2016.

[13] Tsukasa Shiota and Kazutaka Shimada. Annotation and multi-modal methods for quality assessment of multi-party discussion. In *Proceedings of PACLIC36*, pp. 178–185, 2022.

[14] 西山空良, 嶋田和孝. 議論の分析とファシリテーションのための可視化ツールの構築. 電子情報通信学会 HCG シンポジウム 2021, I-2-2, 2021.

[15] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, Peter J Liu, et al. Exploring the limits of transfer learning with a unified text-to-text transformer. *J. Mach. Learn. Res.*, Vol. 21, No. 140, pp. 1–67, 2020.

[16] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, Vol. 33, pp. 1877–1901, 2020.