

# GANによる質問答弁の対話関係評価を導入した 議会会議録の要約文章生成

## Summarization of parliamentary proceedings introducing evaluation of question-answer dialogue relationships by GAN

大野木裕睦<sup>1\*</sup> 森辰則<sup>1</sup>  
Hiromu Onogi<sup>1</sup> Tatsunori Mori<sup>1</sup>

<sup>1</sup> 横浜国立大学大学院環境情報学府

<sup>1</sup> Graduate School of Environment and Information Sciences, Yokohama National University

**Abstract:** There is a great demand for automatic summarization of parliamentary proceedings. It is necessary considering the correspondence between questions and answers when summarizing, but an effective method has not yet been established. We tested a new method to train a T5-based generative summary model to consider the correspondence between questions and answers by using an adversarial generative network (GAN). We also introduced a new framework for the loss function of the generator and tried a novel method that uses the generated question summaries as input when generating answer summaries. The summaries generated by the proposed method obtained higher accuracy than the ones of many existing models. Furthermore, we believe that the "readability" of the generated summaries was improved as a result of human evaluation.

### 1 はじめに

政治課題に関する議論は、議会において行われており、会議録には議論の質問や答弁が記録されている。しかし会議録は、議会での発言を書き起こされただけでまとめられておらず、読みづらいという問題がある。そのため「議会だより」等の文書に、会議録の要約が掲載されるが、50字程度と短く、人手で作成するには時間とコストが発生する。そこで、精度の高い議会会議録の自動要約システムが求められている。

議会会議録の自動要約には、次のような課題がある。

(i) 先行研究は従来、抽出型要約手法を採用してきたが、要約ペアを作る上で文抽出型的手法には限界がある上、昨今は生成型要約手法の性能向上が著しいため、これを試すべきである点と、(ii) 質問答弁の対話関係を考慮しながら要約する必要があるという点である。

(i) の点に対して、我々は T5 を用いて生成型要約モデルを構成・訓練する。(ii) の点に対して、我々は質問と答弁の要約ペアを生成するモデルと、その対話関係の妥当性を検証するモデルを同時に訓練していくことが有効ではないかという仮説を立てる。このような枠組みを、敵対的生成ネットワーク (GAN) を利用して

実現する。生成器  $G$  は会議録の質問発言と答弁発言を入力して質問と答弁の要約ペアを生成する。一方で識別器  $D$  は、質問要約と答弁要約のペアを受け取り、それが人手で作られたものか、あるいは、生成器  $G$  により生成されたものか、を判別する。これら両モデルを交互に学習させることで、生成器  $G$  は識別器  $D$  を騙せるほど精度の高い要約生成をするよう洗練されていき、結果的に自然な対話構造のある質問要約・答弁要約のペアを出力することを期待する。また、生成器  $G$  の損失関数について、教師強制を行いつつ、識別器  $D$  の損失を加味するような新たな枠組みを導入する。

さらなる性能向上を目指して、次のような新規手法を試す。(1) 対話構造を加味しながら要約生成を行うために、生成された質問要約を、答弁要約生成の入力として利用することで、対話の流れを考慮させる。この枠組みをカスケード型生成器と呼ぶ。(2) 質問要約生成・答弁要約生成それぞれについて、入力に質問・答弁両方の原文を与える工夫を行う。(3) 原文と要約を入力し、与えられた要約が人手で書かれた要約か、生成器  $G$  により生成された要約かを判別する識別器  $D_{QPS}$ 、 $D_{APS}$  を導入する。これにより、原文と要約の対応を考慮できるようになることが狙いである。

東京都議会会議録に対して、提案手法を利用し実験を行う。会議録を訓練用データと評価用データに分割し、モデルの訓練と評価を実施する。評価にあたって

\*連絡先： 横浜国立大学大学院環境情報学府  
〒240-8501 神奈川県横浜市保土ヶ谷区常盤台 79-1  
E-mail: ridivarg@outlook.com

は、自動評価および人手評価を行う。

## 2 関連研究

議会議録を自動要約する試みは既に行われている。例えば NTCIR QALab-PoliInfo という評価型ワークショップにおいて、地方議会議録の要約を行うタスクが設定されている。これに関連して、多くの先行研究がなされている [4][5][2]。

### 2.1 要約手法の種類

自動要約手法は、抽出型要約と生成型要約に大別される。抽出型要約は要約対象の文章から、代表文を抽出する手法であるのに対し、生成型要約は文の表現を抽象化し、1 から要約を生成する手法である。

議会議録の要約においては、長らく抽出型要約手法が主流であった [4][5]。しかし、抽出型要約は、質問と答弁の対話関係を保ちつつ、質問答弁それぞれの代表文を抽出することが難しい。

一方で生成型要約は、文法が正確で、文脈に整合性のある要約を生成することが難しいとされていた。しかし近年、生成型要約が目覚ましい発展を遂げている。先述した課題をほとんど克服しつつあり、人間の性能にどんどん迫っている。代表的な生成型要約モデルのほとんどは、とりわけ Transformer ベースの事前学習モデルをファインチューニングすることで構成される。例えば、GPT や BART、T5[6] などが挙げられる。

### 2.2 対話構造を把握することの重要性

木村ら [7] は、議会議録の要約の実現において、精度向上のために質問と答弁の対応関係を把握することが重要と指摘している。しかし、先行研究の多く ([5] 等) は質問と答弁の対話構造を十分に捉えられておらず、いまだ質問答弁の対話構造を把握するための確立された手法は無い。

### 2.3 GAN を用いた文書要約

本論文では、敵対的生成ネットワーク (GAN)[1] を用いた議会議録の要約モデルを提案する。自動要約において GAN を利用する手法は、Liu ら [3] により提案されているが、この手法は単テキストの要約のために GAN を導入することで、品質向上を図るものである。一方で我々は、2つのテキスト間の対応関係を勘案するため、GAN を導入する点に違いがある。

## 3 GAN による対話関係評価を導入した生成型要約モデル

### 3.1 設計方針

2章で、(i) 生成型要約モデルを試す必要があると考える点 (2.1 節)、(ii) 文脈や論理展開を考慮しながら要約する必要があるという点 (2.2 節) を指摘した。これを踏まえ、以下のような方針を立てた。

#### 3.1.1 生成型要約モデルの利用

i. に関して、我々は、T5 ベースの日本語事前学習モデルをファインチューニングし、生成型要約モデルを構成する。T5 モデルを選択する理由は、通常の Transformer とは異なり Relative-Positional-Encoding を利用しているため、実質的に入力長に制限が存在しないからである。長大な議会議録を扱う際に、入力長に制限があると対応できないケースがあり得る。

#### 3.1.2 GAN の枠組みを用いたモデルの構築

ii. に関して、我々は質問と答弁の要約ペアを生成するモデルと、その対話関係の妥当性を検証するモデルを同時に訓練していくことが有効ではないかと仮説を立てる。これによって、対話関係の「自然さ」を担保する狙いである。さらにこの際、両モデルを敵対的生成ネットワーク (GAN) の枠組みにおいて共同で訓練することを考える。

### 3.2 基本設計

敵対的生成ネットワーク (GAN) を利用し生成型の要約生成を行う手法の基本設計について述べる。基本設計には、生成器  $G_Q$ 、 $G_A$  の訓練において、人手要約で教師強制しながらも、識別器  $D$  の損失を利用して学習を進める手法 (3.2.5 節) のような新しい枠組みを含んでいる。また本研究においては、基本設計に加えて、さらなる新規手法も試す。これらについて、3.3 節、3.4 節、3.5 節にて詳述する。

#### 3.2.1 GAN による訓練

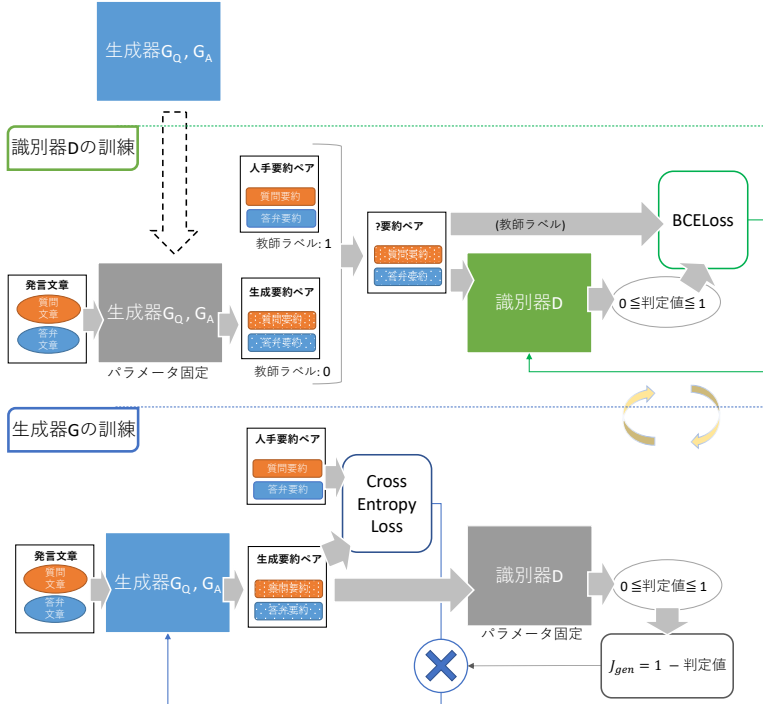
GAN による訓練の模式図を図 1 に示す。

まず、議会議録コーパスから質問原文・答弁原文を用意し、さらにそれらについての人手による要約 (人手要約) も用意する。

生成器は質問要約用生成器  $G_Q$ 、答弁要約用生成器  $G_A$  に分かれている (3.2.3 節に詳述)。それぞれ議会議録

## GANによる訓練

通常の教師強制でファインチューニング済みの生成器



## 実際の運用

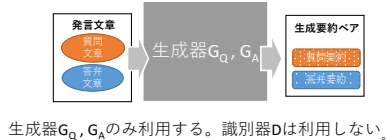


図 1: GAN による訓練と、生成器の実運用についての模式図

の質問原文・答弁原文を入力として受け付け、それをもとに生成質問要約・生成答弁要約を出力するモデルである。GAN における訓練については、生成要約と人手要約のクロスエントロピー損失を算出し、この損失に識別器の出力を掛け合わせ、フィードバックすることで進める (3.2.5 節)。ただし、GAN 訓練前に、人手要約を用いた教師あり学習でファインチューニングを行っておき、これをもとに GAN で学習する (3.2.6 節)。

一方で識別器  $D$  は、質問要約と答弁要約のペアを受け取り、それが人手で作られたものか、あるいは、生成器  $G_Q, G_A$  が自動生成したものか、を判定する (3.2.4 節)。

これらを交互に学習させることで、生成器は識別器を騙せるほど精度の高い要約生成をするようになるよう洗練されていき、結果的に自然な対話構造のある質問要約・答弁要約のペアを出力することを期待する。つ

まり、手本となる人手の要約により近い要約を生成することを期待している。

### 3.2.2 GAN によって得られるものはなにか

識別器  $D$  は、ひとまず人間生成の要約か、機械生成の要約かという判断だけで、内容の良さは評価していない。一般に機械翻訳や生成型の自動要約では、流暢さ、重要内容の保持、その他で生成されたテキストの良さを計測するが、上記の識別器の設計だと、原文を与えない方式なので、「流暢さ」がおもに判断されるのではないかと推測する。

### 3.2.3 生成器 $G_Q, G_A$ の詳細

生成器  $G$  は、質問要約用生成器  $G_Q$ 、答弁要約用生成器  $G_A$  に分かれている。生成器  $G_Q, G_A$  は、会議録の原文を入力として受け付け、それをもとに生成質問要約・生成答弁要約を出力する。

質問要約用生成器  $G_Q$ 、答弁要約用生成器  $G_A$  ともに、Huggingface Model Hub 上に公開されている T5 事前学習モデル<sup>1</sup>を利用して構成する。

運用上、生成器  $G$  の入出力テキストは、すべて日本語 T5 用トークナイザー (SentencePiece) によるトークン列である。

GAN における生成器  $G$  の訓練は、識別器  $D$  による判定値を損失としてフィードバックすることで行う。ここで我々は、生成器  $G$  を人手要約によって教師強制しながら、識別器  $D$  から得られた損失をフィードバックして学習を進める新手法を導入する。3.2.5 節にて詳述する。

### 3.2.4 識別器 $D$ の詳細

識別器  $D$  は「生成質問要約と生成答弁要約のペア」(生成要約ペア) あるいは、「人手質問要約と人手答弁要約のペア」(人手要約ペア) を受け付ける。出力は、与えられた要約ペアが人手により書かれたものか、生成器  $G_Q, G_A$  により生成されたか、の判定値である。

$$D(\text{質問要約, 答弁要約}) = \text{判定値} \quad (0.0 \leq \text{判定値} \leq 1.0)$$

判定値については、0 に近いほど生成要約ペア、1 に近いほど人手要約ペアらしいとする。

識別器  $D$  は、東北大学が Huggingface Model Hub 上で公開している日本語 BERT 事前学習モデル<sup>2</sup>を利

<sup>1</sup><https://huggingface.co/sonoisa/t5-base-japanese>

<sup>2</sup><https://huggingface.co/cl-tohoku/bert-base-japanese-whole-word-masking>

用し、これを転移学習 (ファインチューニング) することで構成する。

運用上、識別器  $D$  の入力テキストは、日本語 BERT 用トークナイザー (MeCab) によるトークン列である。

日本語 BERT 事前学習モデルを、生成要約ペアと人手要約ペアを与えることでファインチューニングし、識別器を訓練する。ファインチューニングは、人手要約の教師ラベルを 1、生成要約ペアの教師ラベルを 0 とする教師あり学習で行う。特別トークン [SEP] で質問要約と答弁要約を分割して入力する。

### 3.2.5 GAN の訓練における生成器 $G_Q$ 、 $G_A$ の損失関数

従来の GAN に基づく設計では、生成器  $G_Q$ 、 $G_A$  の訓練を次のような仕組みで行う。

$$\text{判定値} = D(G_Q(\text{質問原文}), G_A(\text{答弁原文}))$$

$$BCELoss(1, \text{判定値}) \quad =: \text{生成器 } G_Q, G_A \text{ の損失}$$

しかし、本研究で設計する生成器  $G_Q$ 、 $G_A$  と識別器  $D$  の入出力の形式は大きく異なっている。したがって、生成器  $G_Q$ 、 $G_A$  と識別器  $D$  を連続したニューラルネットワークとして扱っても、損失を逆伝播させる際に障害が生じる。ゆえに、生成器  $G_Q$ 、 $G_A$  と識別器  $D$  を独立したモデルとして扱いつつ、生成器  $G_Q$ 、 $G_A$  に識別器  $D$  の損失をフィードバックさせる方法が必要である。

我々はこの課題に対して、次のような新規手法を適用することで対処する。生成器  $G_Q$ 、 $G_A$  の訓練において、人手要約で教師強制しながらも、識別器  $D$  の判定値を利用して学習を進める。

$$J_{gen} = 1 - D(G_Q(\text{質問原文}), G_A(\text{答弁原文}))$$

$$CrossEntropyLoss(G_Q(\text{質問原文}), \text{人手質問要約})$$

$$\times J_{gen} \quad =: \text{生成器 } G_Q \text{ の損失}$$

$$CrossEntropyLoss(G_A(\text{答弁原文}), \text{人手答弁要約})$$

$$\times J_{gen} \quad =: \text{生成器 } G_A \text{ の損失}$$

通常、T5 モデルの訓練時にはクロスエントロピー損失を利用するが、本手法ではクロスエントロピー損失に  $J_{gen}$  を掛け、これを生成器の損失とする。ここで  $J_{gen}$  は、人手要約ペアのラベル (=1) と生成要約ペアの判定値の差分、すなわち「本物の要約ペア」に対する「偽物の要約ペア」の損失と見なすことができる。

この手法では、もとのクロスエントロピー損失に対してスカラー値である  $J_{gen}$  を乗算しているだけなので、損失の勾配の大きさは制御されるが、勾配の向きは制御されない。しかし、損失が多層 NN を逆伝播するなかで、各ニューロンにおける勾配ベクトルは複雑に変

化していくはずであるため、有効な手法であると期待する。

画像生成 GAN 等においては、ランダムノイズを生成器に入力して訓練を行うため、このような訓練手法を導入できない。一方で、文章要約生成は、明確な教師データが存在しているため、この訓練手法を導入できる。

生成器の損失関数を以上のように定義する手法は、我々の知る限り無い。

### 3.2.6 GAN 訓練前に生成器 $G_Q$ 、 $G_A$ をファインチューニングする手法

生成器  $G_Q$ 、 $G_A$  の学習が不十分な場合、識別器  $D$  が質疑応答の対応関係を見逃して生成文章の自然さのみで判断する可能性があることが懸念される。この対処として、先に生成器  $G_Q$ 、 $G_A$  のみ、人手要約を用いた教師あり学習でファインチューニングを行う。これをもとに、GAN で学習する。

## 3.3 カスケード型生成器

上述の生成器の設計では、質問要約と答弁要約を独立に生成しているため、質問と答弁の対話関係を加味しながら要約しているとは言い難い。これを加味するには、独立せずに 2 テキストを出力できる、すなわち質問要約と答弁要約のペアを出力できるようなモデルが必要である。しかし、この条件を満たすような Transformer ベースの事前学習モデルは、今のところ存在しない。

そこで、質問要約生成器  $G_Q$  が生成した質問要約を、答弁要約生成器  $G_A$  に入力することにより、「質問要約を加味して答弁要約を生成できる」ため、対話の流れに則って、質問答弁要約ペアを生成できると考えた。以下、この仕組みを導入した生成器を「カスケード型生成器」と呼ぶ。また、略称 **cascade** でも示す。

$$G_Q(\text{質問原文}) = \text{生成質問要約}$$

$$G_A(\text{答弁原文}, \text{生成質問要約}) = \text{生成答弁要約}$$

カスケード生成器の枠組みについて、図 2 に示す。

より詳細には、T5 に特別トークン <q\_sum> を導入して、表 1、表 2 のように入出力を行う。

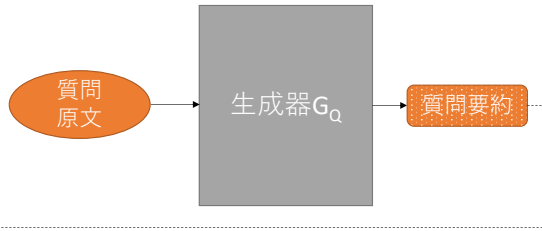
表 1: 質問要約を生成 (Step1)

入力	出力 (目標)
質問原文	生成質問要約

表 2: 生成質問要約を利用し、答弁要約を生成 (Step2)

入力	出力 (目標)
答弁原文 <q_sum> 生成質問要約	生成答弁要約

Step1 質問要約生成



Step2 生成質問要約を再帰的に利用し、答弁要約生成

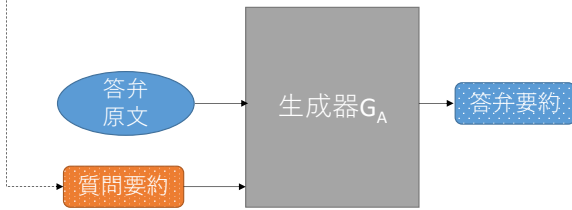


図 2: カスケード型生成器

3.4 各生成器に質問原文と答弁原文の両者を入力する手法

質問要約生成・答弁要約生成それぞれについて、入力に質問・答弁両方の原文を与えるほうが良い可能性を考え、検証する。

$$G_Q(\text{質問原文}, \text{答弁原文}) = \text{生成質問要約}$$

$$G_A(\text{質問原文}, \text{答弁原文}) = \text{生成答弁要約}$$

より詳細には、T5 に特別トークン<question>、<answer>を導入して、表 3、表 4 のように入出力を行う。以下、この手法について、略称 **SET** で示す。

表 3:  $G_Q$  に、質問・答弁原文両方を入力する場合

入力	出力
<question>質問原文<answer>答弁原文	質問要約

表 4:  $G_A$  に、質問・答弁原文両方を入力する場合

入力	出力
<question>質問原文<answer>答弁原文	答弁要約

3.5 原文と要約の対応を判定する識別器  $D_{QPS}$ 、 $D_{APS}$  の導入

上述の識別器 (3.2.4 節) は質問要約と答弁要約のペアを受け付けるが、原文を考慮しない。したがって、原文と要約の対応を加味できない。追加の識別器として、質問と答弁それぞれについて、原文と要約を入力し、与えられた要約が人手要約か生成要約か判定するようなモデルを構成する。

$$D_{QPS}(\text{質問要約}, \text{質問原文}) = \text{判定値} (0.0 \leq \text{判定値} \leq 1.0)$$

$$D_{APS}(\text{答弁要約}, \text{答弁原文}) = \text{判定値} (0.0 \leq \text{判定値} \leq 1.0)$$

判定値については、0 に近いほど生成要約、1 に近いほど人手要約らしいとする。

以下、質問原文と質問要約の対応を判定する識別器を  $D_{QPS}$ 、質問原文と質問要約の対応を判定する識別器を  $D_{APS}$  と呼称する。 $D_{QPS}$ 、 $D_{APS}$  と、先述した識別器  $D$  を併用し、検証を行っていく。

なお、これらを併用する際、生成器  $G_Q$ 、 $G_A$  の訓練は次のようになる。

$$J_1 = 1 - D(G_Q(\text{質問原文}), G_A(\text{答弁原文}))$$

$$J_2 = 1 - D_{QPS}(G_Q(\text{質問原文}), \text{質問原文})$$

$$J_3 = 1 - D_{APS}(G_A(\text{答弁原文}), \text{答弁原文})$$

$$CrossEntropyLoss(G_Q(\text{質問原文}), \text{人手質問要約})$$

$$\times J_1 J_2 J_3 \quad =: \text{生成器 } G_Q \text{ の損失}$$

$$CrossEntropyLoss(G_A(\text{答弁原文}), \text{人手答弁要約})$$

$$\times J_1 J_2 J_3 \quad =: \text{生成器 } G_A \text{ の損失}$$

以下、この手法について、略称 **AQPS** で示す。

4 評価実験

実際の地方議会会議録 (平成 23 年 6 月から平成 27 年 3 月までの東京都議会会議録) に対して、提案モデルがどれほど有効なのか評価する。NTCIR-15 QALab-PoliInfo2<sup>3</sup> Dialog summarization task の設定するデータセット、及び評価手法を利用して実験を行う。

4.1 データセット

東京都議会における一般質問および代表質問の概要は都議会だよりに掲載されている。都議会だよりは、議会で記載された内容が議会事務局の職員により作られていることから、人手により作成された「正解の要約」とみなすことができる。また、都議会だよりでは、質問項目ごとに質問者と答弁者が示されており、質問とそれに対応する答弁が簡潔にまとめられている。そこで、都議会だよりに記載された質疑の要約 (50 字程度) を正解として用いる。

データは全 692 個の出題項目からなり、それぞれには質問原文と答弁原文のペア、人手による質問要約と答弁要約のペアが含まれている。438 個を訓練用データとして扱い、残り 254 個を評価用データとして扱う。

<sup>3</sup><https://poliinfo2.github.io/>

## 4.2 マイナーチェンジモデルの訓練

表5のように追加の新規手法を適用したマイナーチェンジモデルも訓練する。

表 5: マイナーチェンジモデル

cascade	×	×	×	×	○	○	○	○
SET	×	×	○	○	×	×	○	○
AQPS	×	○	×	○	×	○	×	○

## 4.3 学習のハイパーパラメータ

生成器  $G_Q, G_A$  訓練時のバッチサイズは1、エポック数は3である。

識別器  $D, D_{QPS}, D_{APS}$  訓練時のバッチサイズは6、エポック数は4である。

GANの訓練においては、生成器の学習と識別器の学習を、交互に25回繰り返す。

## 4.4 評価

評価用データを対象に要約を生成し、これと人手要約の差を ROUGE スコアによって自動評価し、比較する。また11個の評価用データを無作為に取り出し、これに対応する生成要約を10人に人手評価してもらった。(6人: 情報工学を専攻する学生、1人: 気象関係の研究者、1人: デザイナー、1人: 動画配信者、1人: 主婦) 評価観点は表6のとおりである。

表 6: 人手評価の評価観点

Q-content	質問部分について、人手要約との内容の合致
A-content	答弁部分について、人手要約との内容の合致
Well-formed	文法的な正しさ
Dialog goodness	質問と答弁の対話関係が成立しているか
Overall goodness	質問・答弁要約ペアとしての全体的な良さ

人手評価の対象は、次の5つのモデルに限定した。

- モデル1(ベースライン): 入力を原文、出力を人手要約とするよう、通常の教師強制でファインチューニングした T5 モデル (質問要約生成器、答弁要約生成器)

表 7: 人手評価の結果

	Q-content	A-content	Well-formed	Dialog goodness	Overall goodness	(Q-Rouge)	(A-Rouge)
モデル1 (ベースライン)	1.308	1.503	2.562	1.921	1.528	0.275	0.247
モデル2 (_cascade.AQPS)	<b>1.892</b>	1.595	<b>2.708</b>	2.025	1.707	0.300	<b>0.301</b>
モデル3 (_SET.cascade)	1.508	1.558	2.437	1.756	1.453	0.296	0.291
モデル4 (_SET.AQPS)	1.608	1.595	2.321	1.974	1.657	<b>0.304</b>	0.294
モデル5 (_SET.cascade.AQPS)	1.808	<b>1.641</b>	2.491	<b>2.053</b>	<b>1.798</b>	0.299	0.290

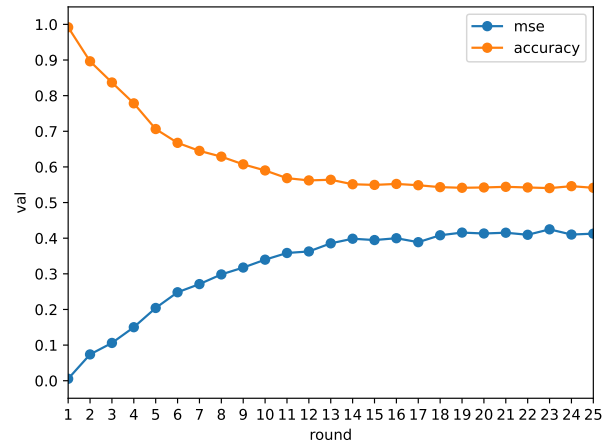


図 3: 識別器  $D$  の性能推移

- モデル2~5: マイナーチェンジモデルのうち、4つ(表8)

表 8: 人手評価対象のマイナーチェンジモデル

	モデル2	モデル3	モデル4	モデル5
cascade	○	○	×	○
SET	×	○	○	○
AQPS	○	×	○	○

評価結果を、表7に示す。

## 5 考察

### 5.1 GANの訓練における、生成器 $G_Q, G_A$ への新たな枠組みの損失関数の導入

識別器  $D$  が GAN の学習を繰り返す中で、どのように性能が推移するかを図5に示す。段々と正解率 (accuracy) が0.5に向かっており、識別器が生成要約と人手要約を区別できなくなっていることがわかる。また、平均二乗誤差 (mse) も、段々と0から大きくなっている。したがって、生成器  $G_Q, G_A$  によって識別器  $D$  が騙されるよう、学習が進んでいることがわかる。



ゆえに、3.2.5節で述べたような、生成器  $G_Q$ 、 $G_A$  の訓練において、人手要約で教師強制しながらも識別器  $D$  の損失を利用して学習を進める新規手法を導入しても、生成器と識別器とで敵対的に学習が進むと言える。

## 5.2 先行研究との比較

評価型ワークショップ NTCIR-15 QALab-PoliInfo2 Dialog summarization task に提出されたモデルと、我々が訓練したモデルの ROUGE スコアを表 9 に示す。なお表中の、JRIRD、nukl、TO、Forst は、本 task へ他の参加者が提出したモデルを表す。

表 9: 各モデルの ROUGE スコア

モデル	ROUGE
JRIRD	0.321
モデル 2(.cascade_AQPS)	<b>0.301</b>
基本設計	<b>0.296</b>
モデル 5(.SET_cascade_AQPS)	<b>0.295</b>
モデル 1(ベースライン)	<b>0.261</b>
nukl	0.258
TO	0.244
Forst	0.241
(本 task 参加者の平均)	0.185

我々の制作したモデルは、他のモデルの平均を大きく超えた。基本設計であっても、高い水準を獲得しており、GAN の導入が、要約生成の品質改善に寄与していることがわかる。

表 10: モデルの出力例

	質問	答弁
原文	我が国の経済は、自由民主党政権による景気対策や金融対策により、少しずつ明るい兆しが見え始めています。こうした景況の改善が、東京の産業を支えている数多くの中小零細企業にしっかりと行き渡る施策の展開こそが都政の重要な役割であると考えます。また、東京の経済の中長期的な発展を確実なものとするためには、中小企業の成長を行政として戦略的にサポートすることも不可欠です。現政権のアベノミクスの打ち出す成長戦略と連携して、東京の産業振興を着実に推し進めて、中小零細企業の活力を高め、雇用就業の場も生み出していくという総合的な政策の展開こそが必要となっています。都は、中小零細企業への支援を初め、東京の産業振興に向けてどう取り組んでいく考えか、知事の所見を伺います。	次に、雇用就業対策についてであります。東京の産業の成長と都民生活の安定を図るため、少子高齢化の影響を踏まえ、労働力を確保し、一人一人の能力や個性を生かした働き方を実現することが重要であります。現在、非正規雇用で働かざるを得ない若者や、出産等で仕事を離れた女性、これまでの経験等を生かして働くことを目指す高齢者に対して、就業の推進に向け、きめ細かく支援することが課題となっております。このため都は、正社員を希望する若者への重点的な就業支援の実施や、女性の再就職を後押しする新たな取り組みの展開、高齢者が希望や能力に応じて働くための支援の強化など、雇用就業施策の一層の充実を図ってまいります。今後とも、産業の成長を支え、多様な人材が活躍できる雇用就業の仕組みづくりを着実に進めてまいります。
人手要約	若者や女性、高齢者が能力を十分発揮し働ける社会の仕組みづくりを。	きめ細かい支援が課題。能力や個性を生かした働き方の実現へ施策を充実。
モデル 1(ベースライン)	活力を高め雇用就業の場も生み出していく政策の展開が必要。所見は。	雇用就業の仕組みづくりを着実に進め、多様な人材が活躍できる雇用就業の仕組みづくり。
モデル 3(.SET_cascade)	中小零細企業への支援をはじめ東京の産業振興にどう取り組むか。	若者の重点的な就業支援実施や高齢者の再就職支援等、雇用就業施策の一層の充実を図る。
モデル 5(.SET_cascading_AQPS)	[1] 東京の産業振興に向けどう取り組むのか。 [2] 中小企業の活力を高め雇用就業の場も生み出していく総合的な政策展開が必要。所見は。	[2] 若者の重点的な就業支援の実施や高齢者の希望や能力に応じて働くための支援の強化等雇用就業施策の一層の充実を図る。

## 5.3 質問文章と答弁文章をまとめて入力する手法 (SET)

モデル 2(.cascade\_AQPS) とモデル 5(.SET\_cascade\_AQPS) を比較し、SET の有効性を検討する (表 7)。Q-Content 指標、すなわち質問要約内容の精度は減少してしまっ。このことから、質問要約生成時に、答弁原文を加味することはむしろ悪影響を及ぼすと推測される。一方で、答弁要約生成時には、質問原文も答弁原文も与えた方が、答弁要約内容の精度が向上した。以上から、時系列の流れに沿って原文を入力することが最善であると考察する。

また、現在のモデルは、質問原文 1 つと答弁原文 1 つのペアに対して学習を行うが、実際の質疑応答では、質問原文 1 つに対して答弁原文が複数あることがある。この場合、質問要約生成時に答弁原文を加味すると、質問原文に対して質問要約が一意に定まらない。このことから、質問要約生成時に答弁原文を加味すべきではない。

## 5.4 カスケード型生成器の導入

モデル 4(.SET\_AQPS) とモデル 5(.SET\_cascade\_AQPS) を比較し、カスケード型生成器の有効性について検証する (表 7)。カスケード型生成器を導入することで、すべての評価指標が向上した。A-content 指標のみならず Q-Content 指標も向上していることから、答弁要約の品質が向上すると、GAN の学習の中で、質問要約の品質も向上するよう、間接的に学習圧がかかっている可能性がある。

## 5.5 原文と要約の対応を判定する識別器 $D_{QPS}$ 、 $D_{APS}$ の導入 (AQPS)

モデル3(\_SET\_cascade)とモデル5(\_SET\_cascade-AQPS)を比較し、AQPSの有効性について検証する(表7)。AQPSを導入することで、すべての評価観点において良い結果が得られた。AQPSの導入により、生成要約内容の品質が向上し、これに伴ってDialog-goodness, Overall-goodness 指標も向上したと思われる。

## 5.6 GANによって何が得られたのか

GANの導入によって、質問答弁要約ペアの「流暢さ」が向上すると予想した(3.2.2節)。これを踏まえ、GANを導入することで何が得られたのか、定性的な判断も含めて検討する(表10に例を示す)。

まず、流暢さの1つの側面として、文法の正しさが考えられるが、well-formed 指標は、GAN適用・不適用の間で、ほとんど差が無い。もともと、事前学習済T5モデルが、洗練された文法構造を出力できる言語モデルだからであろう。

GANの適用により、特に政治課題の議論において頻繁に見かけるような表現や語句を出力するようになり、政治トピックの要約として自然になる印象がある。また、GANを適用すると、質問側と答弁側で対応関係のある単語や表現を利用することが増える。

GANを適用していないモデル1(ベースライン)の出力については、質問要約と答弁要約のどちらを先に生成したのか解らず、読みづらい印象がある。GANを利用すると、この点が解消し、質問要約・答弁要約の順に読むと理解しやすくなる印象がある。これは、単語や表現の対応に起因する部分が大いかもしれない。

以上より、「読みやすさ」とも言うべき要素が向上したと考える。「読みやすさ」は流暢さの1つの指標であると思われるので、我々の予想は正しかったと言える。

## 6 おわりに

事前学習モデルT5を生成型要約モデルとして訓練することで、議会議録の要約文章生成を行った。ここで、質問答弁の対話関係を加味するために、GANを利用する新規手法を提案した。また、GANの訓練において、生成器Gの損失関数について、教師強制を行いつつ、識別器Dの損失を加味するような新たな枠組みの損失関数を導入した。この枠組みは、両モデルの入出力の形式が大きく異なり、独立に運用するしかない場合でも敵対的に訓練できる点に強みがある。様々な応用に可能性を見込める手法と考える。

さらに3つの追加の新規手法、すなわちカスケード型生成器、各生成器に質問原文と答弁原文の両者を入力する手法、原文と要約の対応を判定する識別器  $D_{QPS}$ 、 $D_{APS}$  の導入は、いずれも性能向上に寄与することがわかった。総合的には自動評価・人手評価ともに良い結果を得られるモデルを構成できた。特に、カスケード型生成器は、対話関係を加味しつつ要約を行うための容易な工夫である。議会議録の要約生成に限らず、対話関係の見られるテキストに対する要約生成において、応用の効く手法であると考えられる。

今後の課題としては、以下が挙げられる。まず、5つのモデルを対象に人手評価を行ったが、GANや各種手法の寄与についてより詳細に調査するため、追加の実験を実施すべきである。さらに、訓練したモデルは、年度や個数などの数字を扱うことを苦手としている。最後に、学習や評価で扱う議会議録データセットは、人手要約と原文の内容が対応していない場合があるなど、看過できない問題点があることが挙げられる。他の地方議会議録についても利用しながら、この課題に対する解決策を探るべきと考える。

## 参考文献

- [1] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems*, Vol. 27. Curran Associates, Inc., 2014.
- [2] Kazuma Kadowaki. Jrird at the ntcir-15 qa lab-poliinfo-2 task: An abstractive dialog summarization system for japanese assembly minutes. *Proceedings of the 15th NTCIR Conference on Evaluation of Information Access Technologies*, Dec 2020.
- [3] Linqing Liu, Yao Lu, Min Yang, Qiang Qu, Jia Zhu, and Hongyan Li. Generative adversarial network for abstractive text summarization. *CoRR*, Vol. abs/1711.09357, , 2017.
- [4] Yasuhiro Ogawa, Yuta Ikari, Takahiro Komamizu, and Katsuhiko Toyama. Nukl at the ntcir-15 qa lab-poliinfo-2 task. *Proceedings of the 15th NTCIR Conference on Evaluation of Information Access Technologies*, Dec 2020.
- [5] Hiromu Onogi, Kiichi Kondo, Younghun Lim, Xinnan Shen, Madoka Ishioroshi, Hideyuki Shibuki, Tatsunori Mori, and Noriko Kando. Forst: A challenge to the ntcir-15 qa lab-poliinfo-2 task. *Proceedings of the 15th NTCIR Conference on Evaluation of Information Access Technologies*, Dec 2020.
- [6] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. Exploring the limits of transfer learning with a unified text-to-text transformer. *CoRR*, Vol. abs/1910.10683, , 2019.
- [7] 木村泰知, 関根聡, 乾健太郎. 地方議会議録の要約に向けて. 言語処理学会 第24回年次大会 発表論文集(NLP2018), Vol. P5-3, pp. 596-599, Mar 2018.