

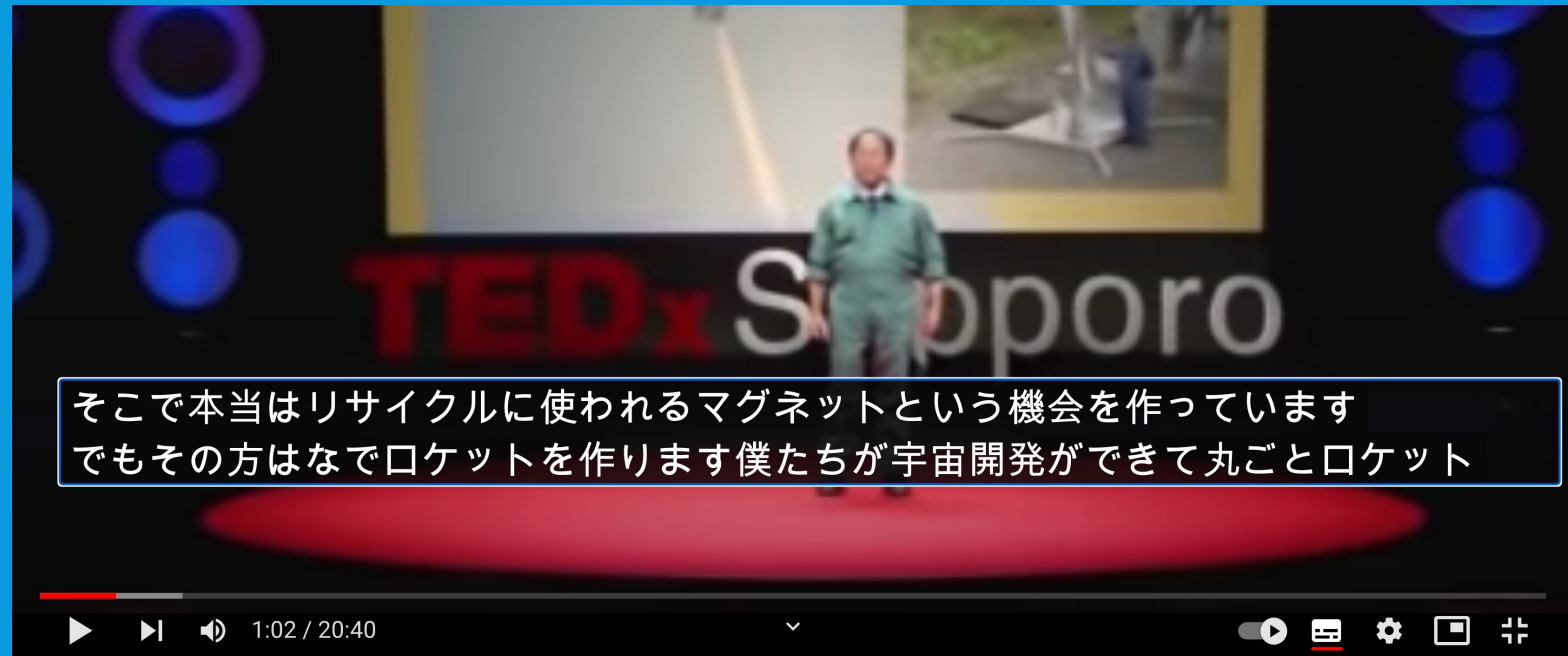
# 字幕生成システムの改良

令和4年度 情報工学ゼミⅡ  
豊田工業高等専門学校 情報工学科 4学年

村田研究室  
加藤愛斗, 田中涼雅, 仲田泰都, 村井源太, ミラ

# 開発の背景

- ・ 一般の字幕生成システムでは認識した音声をそのまま表示するため、句読点がなく字幕が見づらい
- ・ 字幕では長い文を読み取りづらい



# 解決方法

- ・改行、句読点がなく文の区切りが分かりづらい  
→ 句読点・改行挿入モデルを作成
- ・長文字幕の要点が読み取れない  
→ 要約モデル（文短縮モデル）を作成

# 開発の目的と対象ユーザ

## 目的

字幕をより読み取りやすくする

## 対象ユーザ

字幕利用者 (聴覚障がい者など)

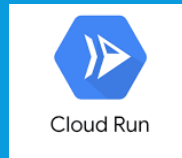
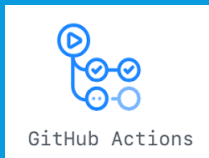
# 実現方法（技術）

## 字幕表示のフロント



Speech  
Recognition  
API

## モデルを配信するバックエンド



## 句読点・改行挿入モデル



## 要約モデル(文章短縮モデル)



# 句読点・改行モデル（目標）

入力：

そのことをふまえて二十一世紀に何をすべきかということにつきましてお話しをさせていただきたいと思いますまずはじめの冷戦後の世界

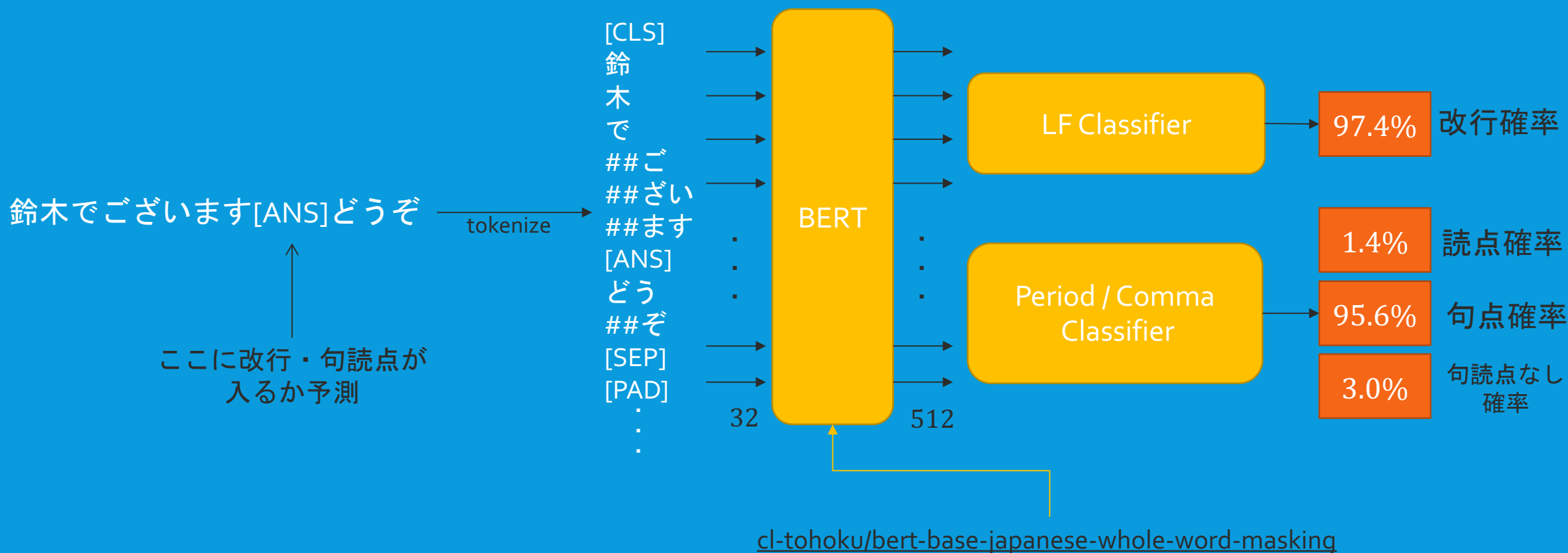
出力：

そのことをふまえて

二十一世紀に何をすべきかということにつきまして、  
お話しをさせていただきたいと思います。

まずはじめの冷戦後の世界

# 句読点・改行モデル（ネットワーク）



# 改行モデル（結果）

予測

改行あり(Positive)

改行なし(Negative)

改行あり  
(Positive)

正解(TP)

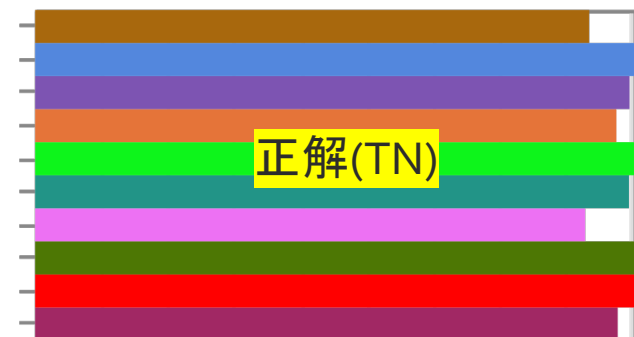
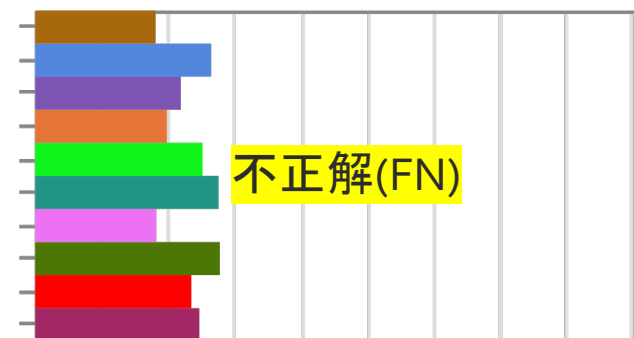
不正解(FN)

改行なし  
(Negative)

不正解(FP)

正解(TN)

正解

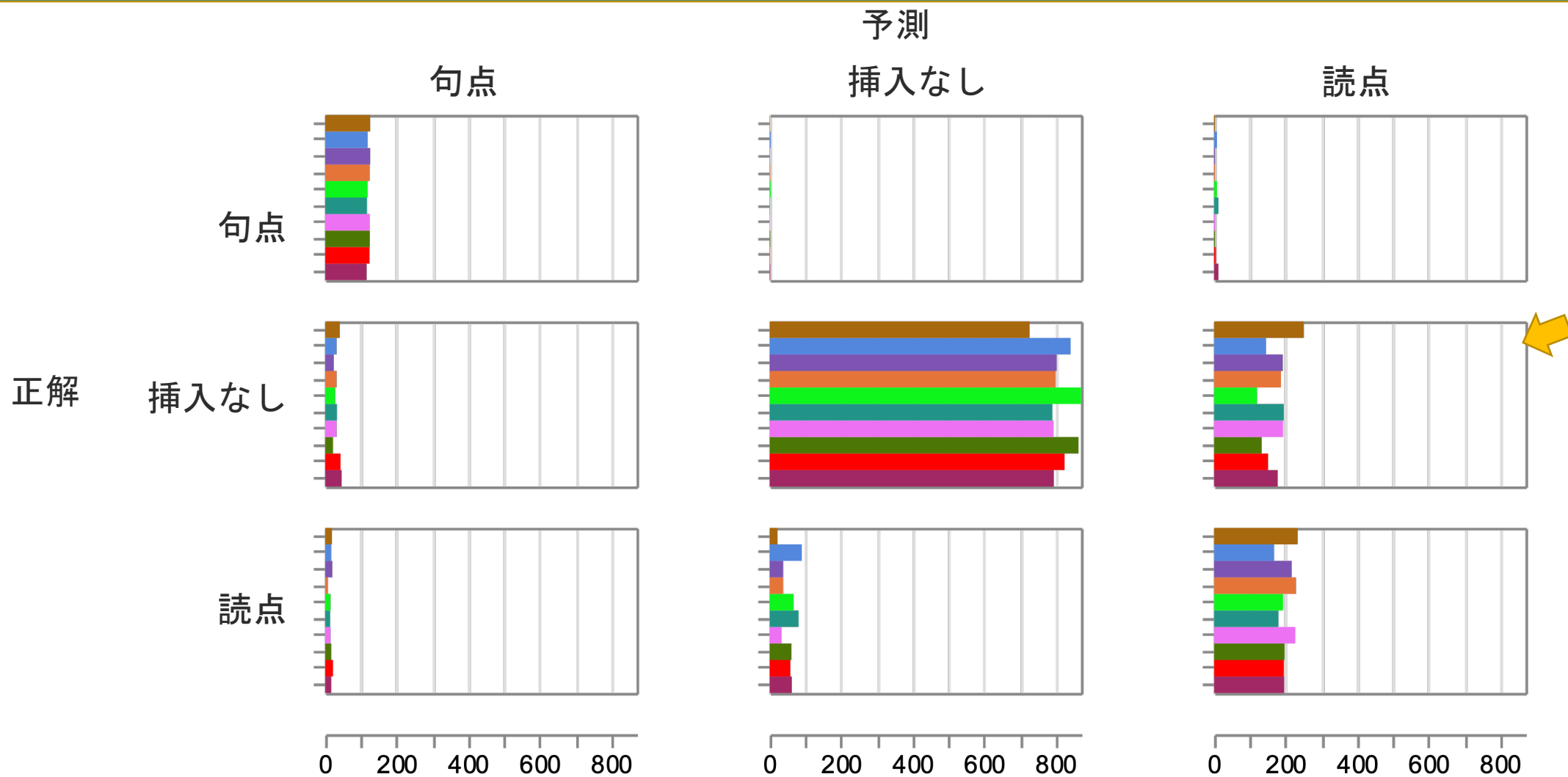


0 100 200 300 400 500 600 700 800 900

0 100 200 300 400 500 600 700 800 900



# 句読点モデル（結果）

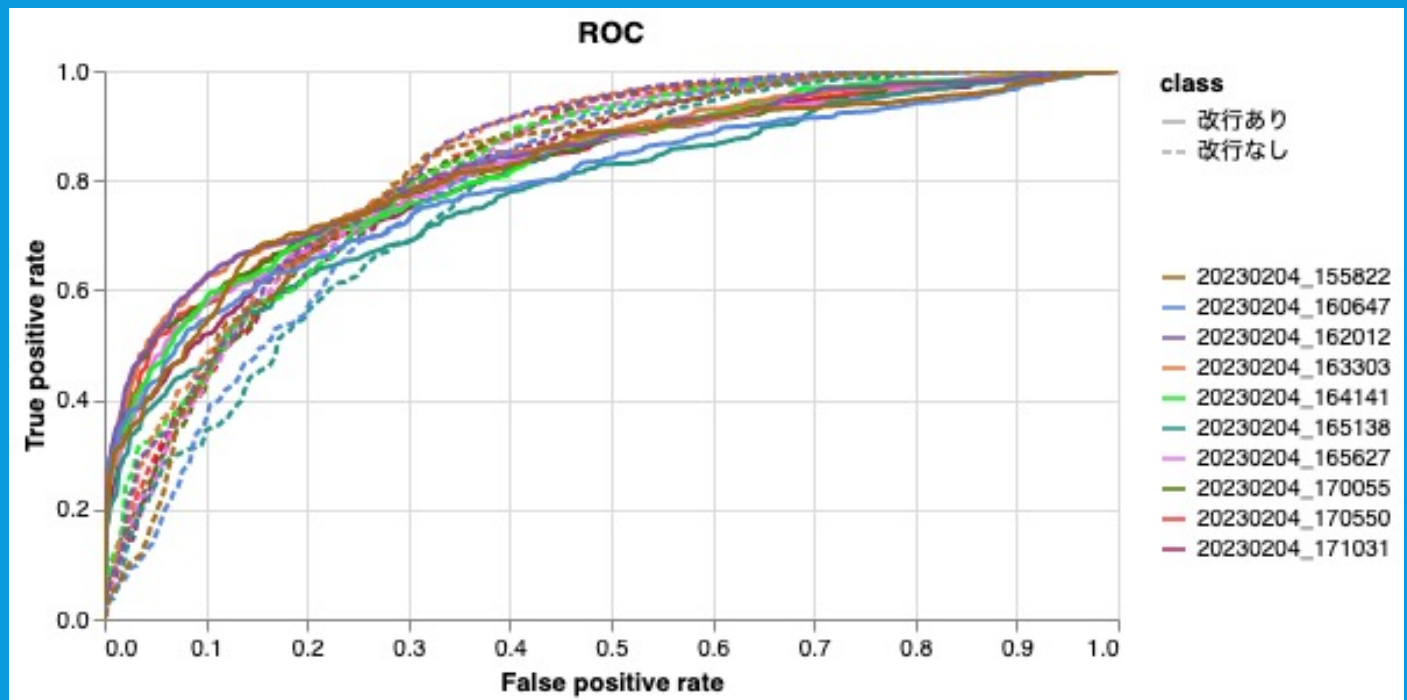
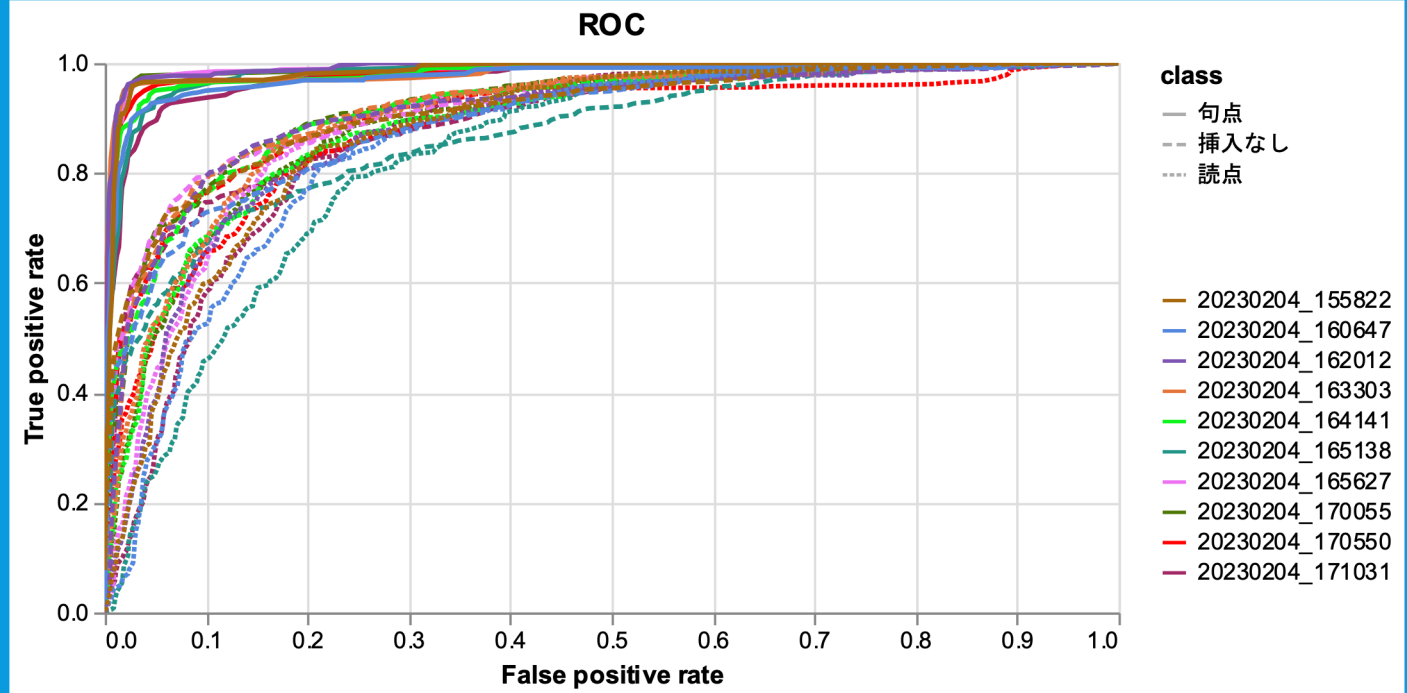


# 句読点・改行モデル（スコア）

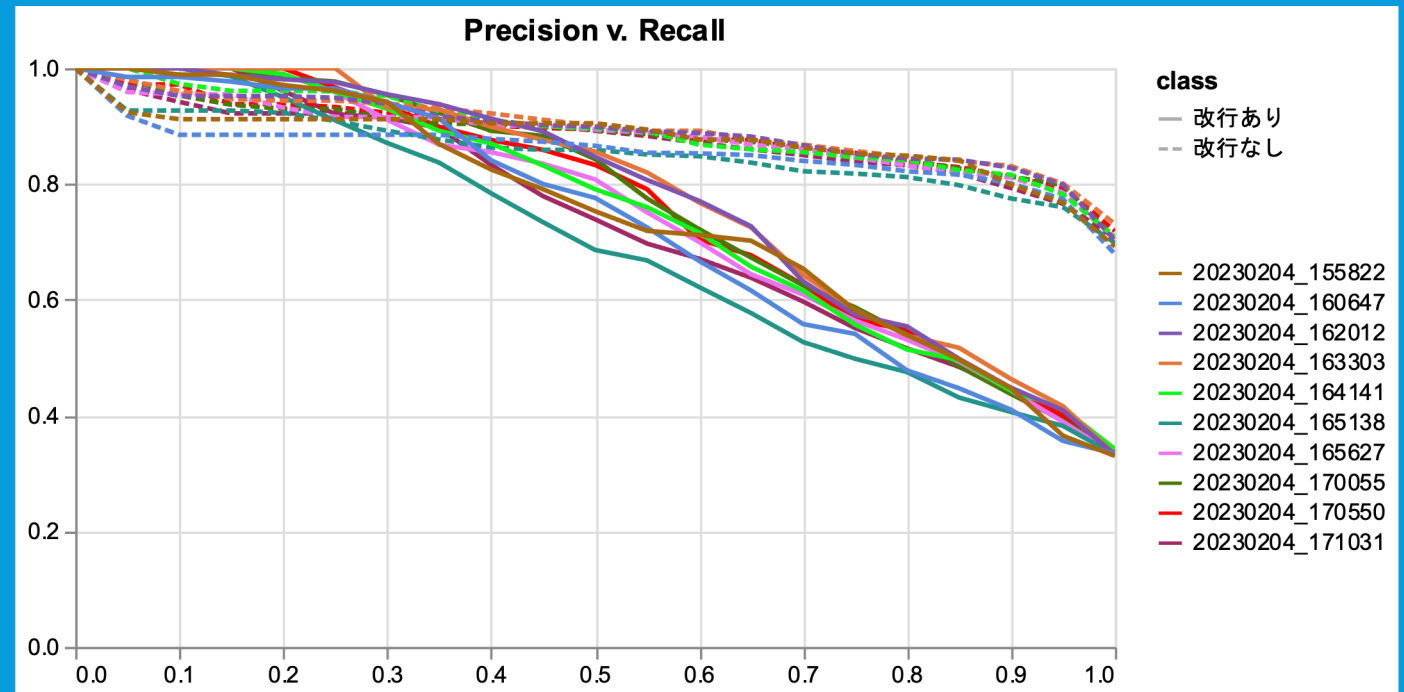
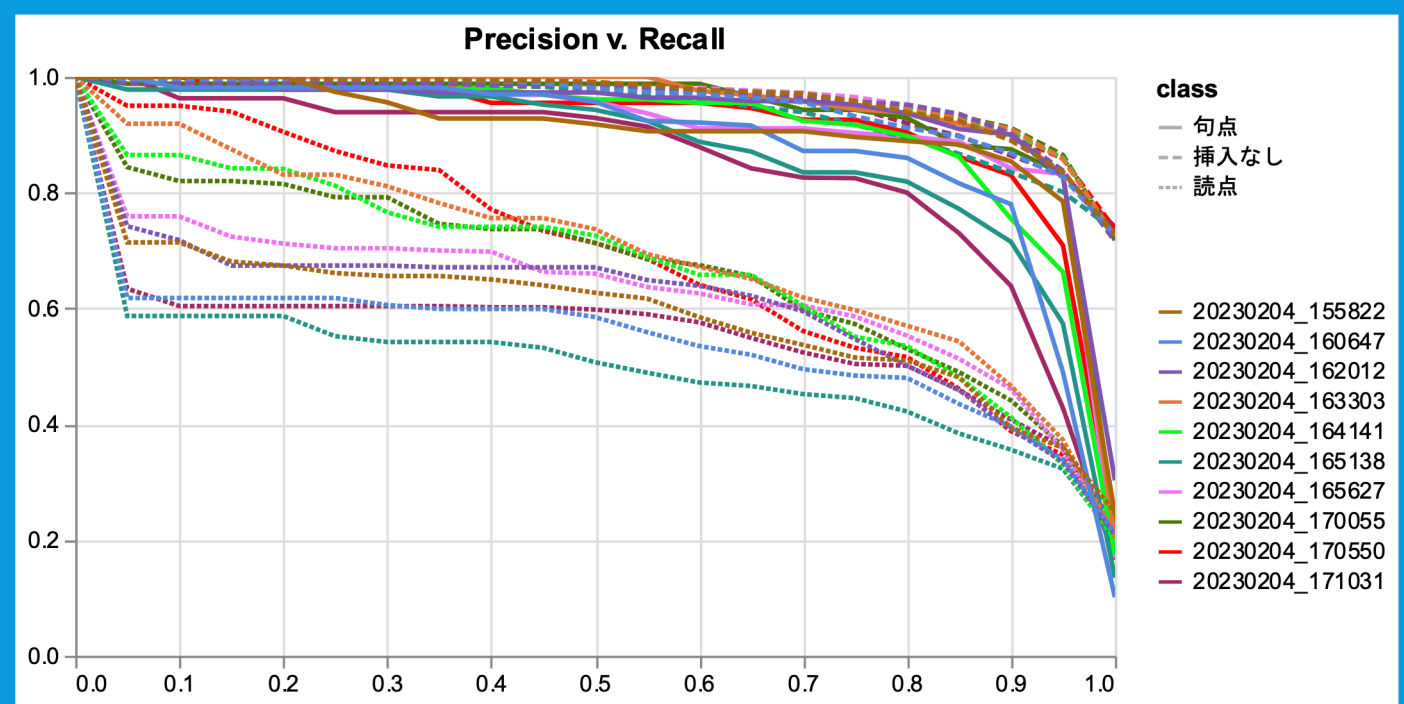
	Accuracy	Precision	Recall	F1 Score	Matthews Corr Coef
改行モデル	0.806	0.780	0.576	0.663	0.544
句読点モデル	0.859	0.751	0.859	0.789	0.649

※ MCCが、+1の場合は完全予測を表し、0の場合は平均ランダム予測を表し、-1の場合は逆予測を表す。

# ROC Curve



# PR Curve



# 文要約モデル（目的）

入力と出力:

21~25文字以内の文 → 20文字以下の文

理由:

- 字幕表示において1行は20文字としたい
- 20文字を少し超える文を、改行挿入し表示するとアンバランスで読みにくい

その新しい動きとは一体何なんだろうかと考えますと

改行

その新しい動きとは一体何なんだろうかと考えますと

要約

その新しい動きは一体何なのかを考えると

# 文章要約（構成）

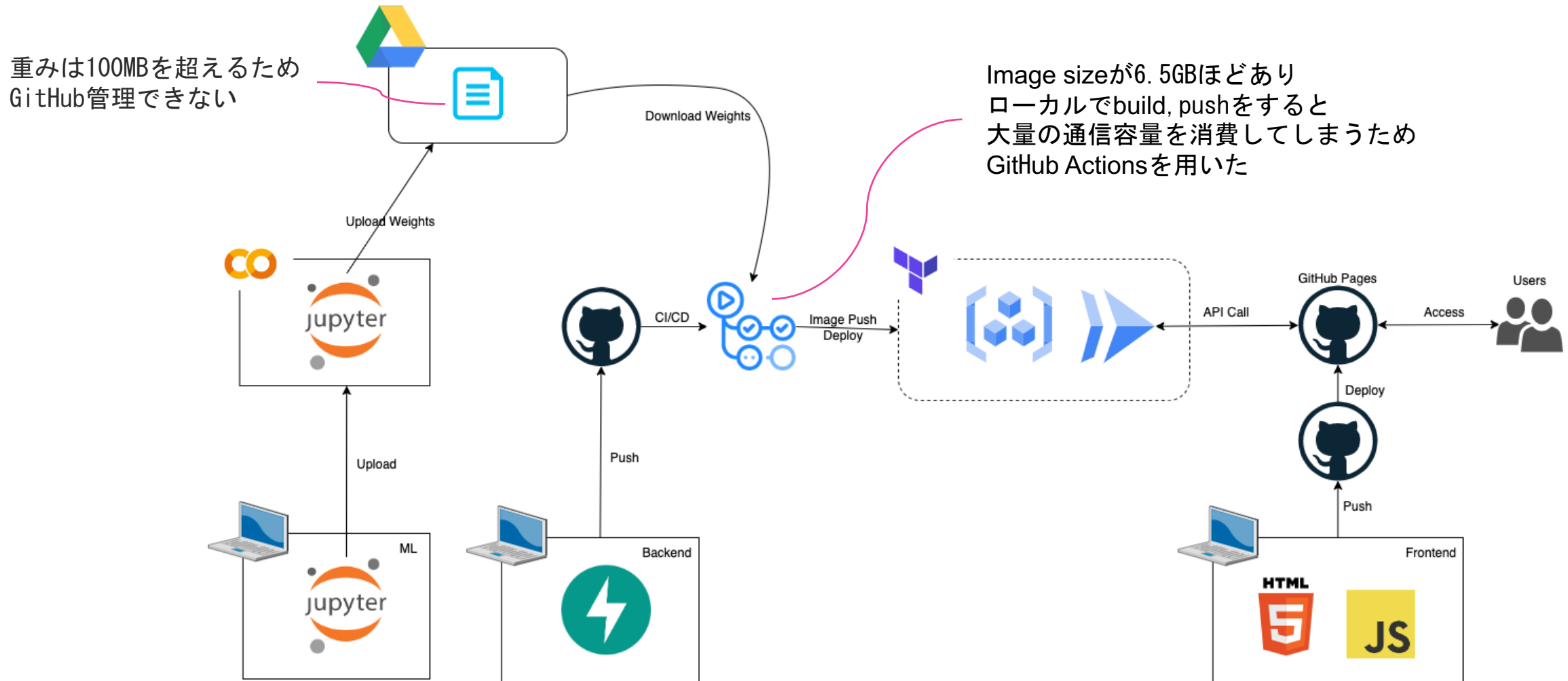
original	ans_len	ans	ans_len
1 ただ今ご紹介いただきました細川でございます	21	ただ今ご紹介いただいた細川です	15
2 次にそのことをふまえて二十一世紀に何をすべきか	23	次にそれをふまえ二十一世紀に何をすべきか	20
3 世界的に大規模な戦争が起こる必然性というものは	23	世界的な戦争が起こる必然性というものは	19
4 むしろ頻発してきたということが言えると思います	23	むしろ頻発してきたことが言えると思う	18
5 国連の安保理における拒否権の行使が断然減りました	24	安保理における拒否権の行使が断然減った	19
6 千九百九十年以降は拒否権の行使が殆どなくなった	23	<CANNOT>	8
7 冷戦時代には誰も国連に何も期待しなかったと	21	冷戦時代の国連に誰も何も期待しなかったと	20
8 国連の能力以上にいろんな問題が国連に持ち込まれる	24	国連の能力以上に大量の問題が持ち込まれる	20
9 現在の常任理事国は当時の戦勝大国でございます	22	今の常任理事国は当時の戦勝国である	17
10 この国連という組織に大きな貢献が出来るようになった	25	<CANNOT>	8
11 もうひとつはジーセブンといったようなところで	22	もうひとつはジーセブンというところで	18
12 この国連の分担金をちゃんと払わない国がある	21	国連の分担金をちゃんと払わない国がある	19
13 ちゃんと払わない国があるというような事もありまして	25	ちゃんと払わない国がある事もあり	16
14 国連の改革の問題は色々複雑な問題がございます	23	国連の改革の問題は複雑な問題があり	17
15 これだけでも二三時間かかるような話でございますので	25	これだけで二三時間かかる話ですので	17
16 その意味がヨーロッパの場合とアジア太平洋の場合で	24	その意味がヨーロッパとアジア太平洋で	18
17 ナトーとワルシャワ条約機構との間の二極対立と	22	ナトーとワルシャワ条約機構間の二極対立と	20
18 グローバルに言えばそういう事だと思いますが	21	世界的に言えばそういう事だと思いますが	18
19 五十年百年の先は正直なところ到底見通しがたちません	25	五十年百年の先の未来の見通しがたちません	20
20 適切に処理する国際的な機関もまだ十分できていない	24	適切に処理する国際機関も十分できてない	19
21 まだ十分できていないということもあると思います	23	十分できてないということもあると思います	20
22 非常に深刻な問題ということが出てきております	22	とても重い問題ということが出てきている	19
23 引き続き行う必要があるということでございます	22	引き続き行う必要があるということである	19
24 恒定的に二つは分けて整理したいと思っております	22	恒定的に二つは分けて整理したいと思っております	17

- ・ 入力と出力のペアを作成(700文程度)
- ・ Text to TextのタスクであるためT5を  
ファインチューニング
- ・ Pre trained modelは、日本語で学習された  
t5-Japanese-baseを使用

# 文章要約（成果）

データ数が少なかったことにより求めていた結果がでなかった  
そのため今回は文章要約モデルはバックエンドに組み込んでいない

# システム構成（機器のつながり）





デモ動画

[ここに結果表示（音声認識）]  
[前回の認識結果を表示]

文字色 フチ色

認識 :   背景色 :

認識 :  サイズ  22.1px  ウェイト  900  縁取り  6pt  
テキストの位置 : 左右 :  中央寄せ  上下 :  上   
テキストの改行 :  画面端で改行

音声認識言語 :  日本語  フォント名 :

↑言語設定は、即反映されないの、設定後リロードしてみてね。（なんで変わらんのかなー(´・ω・`))

何ミリ秒後に文を消す? :  ミリ秒（半角数字で入れてね、1000ミリ秒 = 1秒）  
何ミリ秒黙ったら文を区切る? :  750 ミリ秒（半角数字で入れてね、1000ミリ秒 = 1秒）  
（500：すぐ切る | 750：普通 | 1000：つなぐ 調整して、何も入れないと、これまでと同じ動作になる。）

# ほかの作品・技術との相違点（知的財産権）

## 音声字幕ちゃん

([http://www.sayonari.com/trans\\_asr/asr.html](http://www.sayonari.com/trans_asr/asr.html))

→句読点や改行がない

→このシステムに改良を加える

目的で開発を行った

村田研究室です 今回 ゼミでは改行 句読点 モデルの追加を行いました  
[ここに結果表示(翻訳1)]

Google Script API-KEY:

↑ここにKEYを入れたら翻訳結果も出る。キーの作成方法はこちら: [http://www.sayonari.com/trans\\_asr/index\\_asr.html](http://www.sayonari.com/trans_asr/index_asr.html)

文字色 フチ色  
認識:   背景色:

翻訳1:

翻訳2:

サイズ ウェイト 線取り  
認識:  25px  900  6pt  
翻訳1:  25px  900  6pt  
翻訳2:  25px  900  6pt

テキストの位置: 左右:  中央寄せ 上下:  上  
テキストの改行:  画面端で改行

音声認識言語:  日本語 フォント名:   
翻訳言語(1):  英語 フォント名:   
翻訳言語(2):  なし フォント名:

↑言語設定は、即反映されないので、設定後リロードしてみてください。(なんで変わらんのかなー(´・ω・`))

何ミリ秒後に文を消す?  ミリ秒(半角数字で入れてね。1000ミリ秒 = 1秒)  
何ミリ秒黙ったら文を区切る?  750 ミリ秒(半角数字で入れてね。1000ミリ秒 = 1秒)  
(500: すぐ切る | 750: 普通 | 1000: つなが 調整して。何も入れないと、これまでと同じ動作になる。)

棒読みちゃん連携を使う: ☐ ← 使う場合チェック入れる。詳細 → [twitter](#)

センシティブワード対応(\*\*\*に置き換え)をやめる: ☐

# SDGsとの関連

- 4 遠隔やオンデマンドの授業の効果を向上
- 10 視覚に障害があるかた用の字幕システム

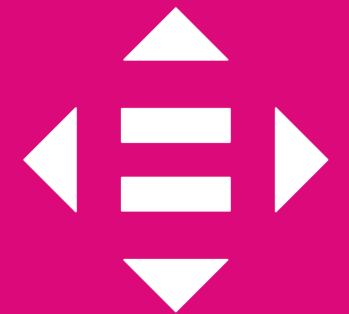
## 参考文献

<https://www.mofa.go.jp/mofaj/gaiko/oda/sdgs/index.html>

**4** 質の高い教育を  
みんなに



**10** 人や国の不平等  
をなくそう



THANK YOU FOR LISTENING