

## 音声言語基盤モデルを用いた講義動画からの重要箇所検出

○米田陸太(阿南高専)

## 概要

- ✓ 課題: 講義動画などの教育コンテンツが急増。学習に必要な視聴時間が膨大に。
- ✓ 手法: 音声言語基盤モデルと分類器を用いて重要箇所を抽出し、効率的な視聴を支援。
- ✓ 成果: 新しい基盤モデルやCNNを組み込んだ分類器により、検出精度の向上を確認。

## 研究背景・目的

- コロナ禍により、大学や高専において授業のオンデマンド化が進み、講義動画が急増。
- 講義動画が増加したことで学習の利便性は向上したが、必要な視聴時間が膨大に。

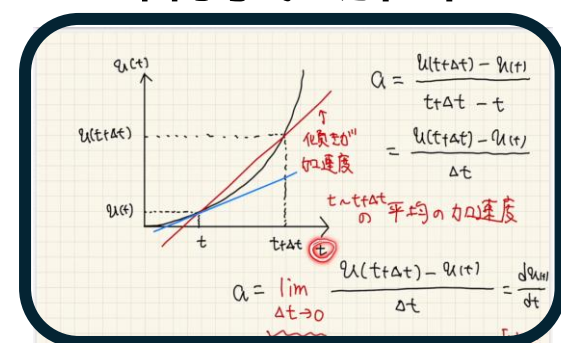
➔ 効率的な動画視聴を支援する技術が必要とされている。

◆ そこで本研究では、

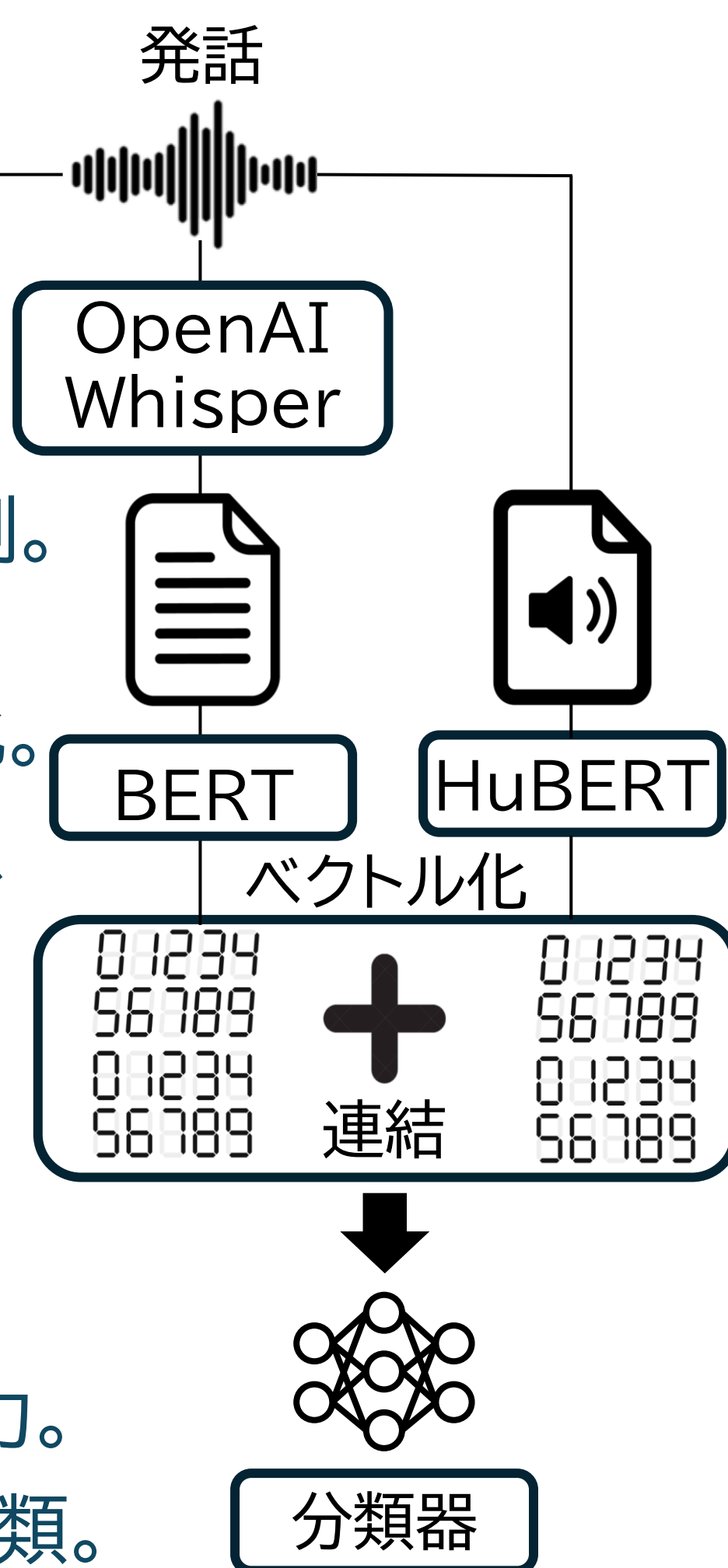
- ✓ 講義動画から重要箇所を抽出する手法を音声言語基盤モデルを用いて構築。
- ✓ ユーザの学習効率向上を目指す。

## 提案手法

講義動画



1. 動画から音声を抽出。発話(セグメント)に分割。
2. 各発話をWhisperで書き起こしてテキスト化。
3. 音声言語基盤モデルで、テキストと音声をそれぞれベクトル化。
  - テキスト: BERT
  - 音声: HuBERT
- ➔ 得られたベクトルを連結して分類器に入力。
4. 重要or非重要を二値分類。
  - ✓ ニューラルネットに基づく以下の分類器を比較:
    - Dense(全結合層)
    - Dense+CNN(畳み込み層)



## 評価実験

## ■ データセット

- ✓ 講義動画に人手で重要/非重要ラベル付与。
- 男性話者A「物理2」「理科総合」(国立高専の授業)
- 男性話者B「Python入門」(YouTube動画)

## ■ 音声言語基盤モデル(BERT+HuBERT)

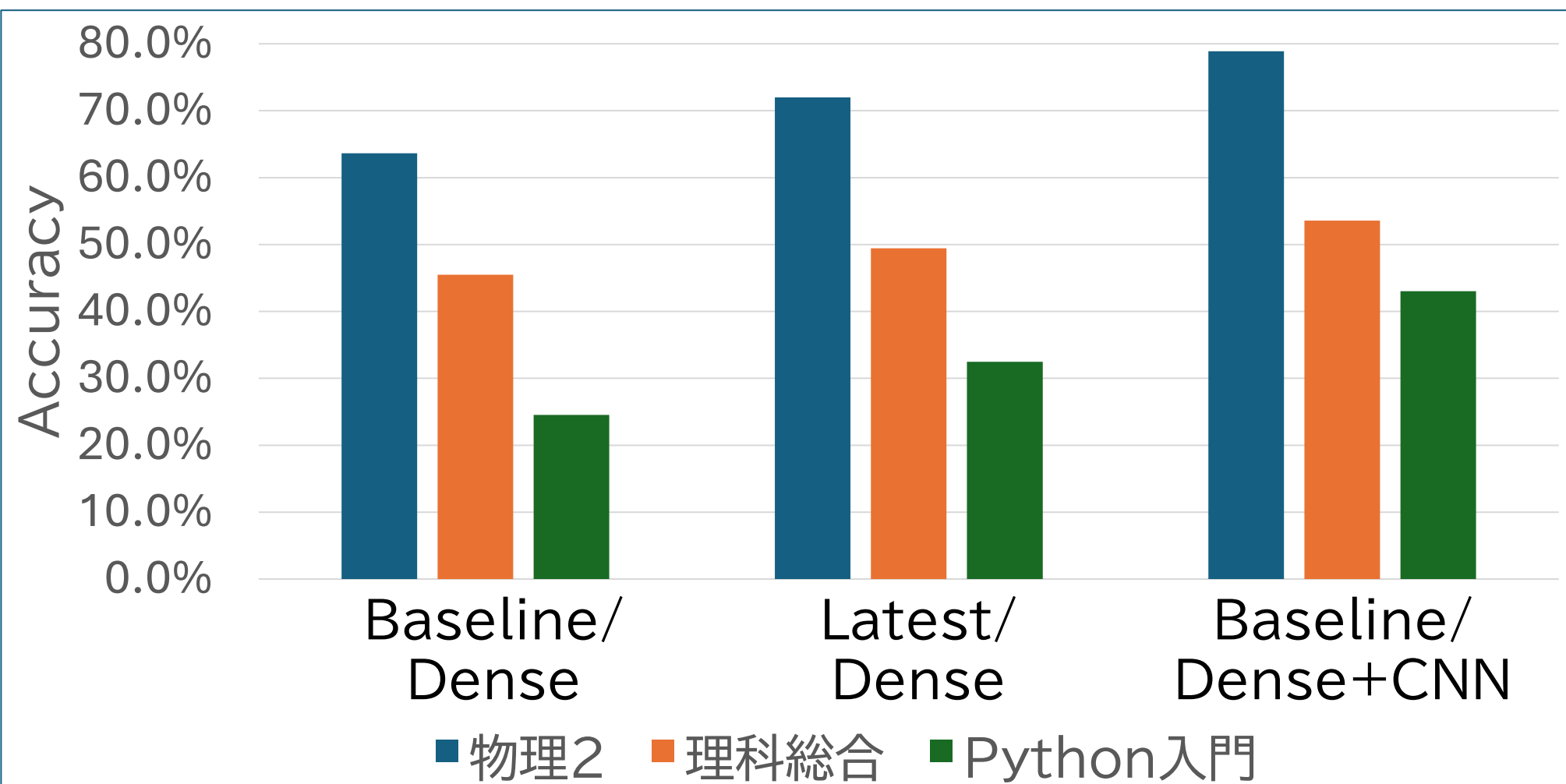
## ● Baseline(従来モデル):

- rinna/japanese-hubert-base
- tohoku-nlp/bert-base-japanese

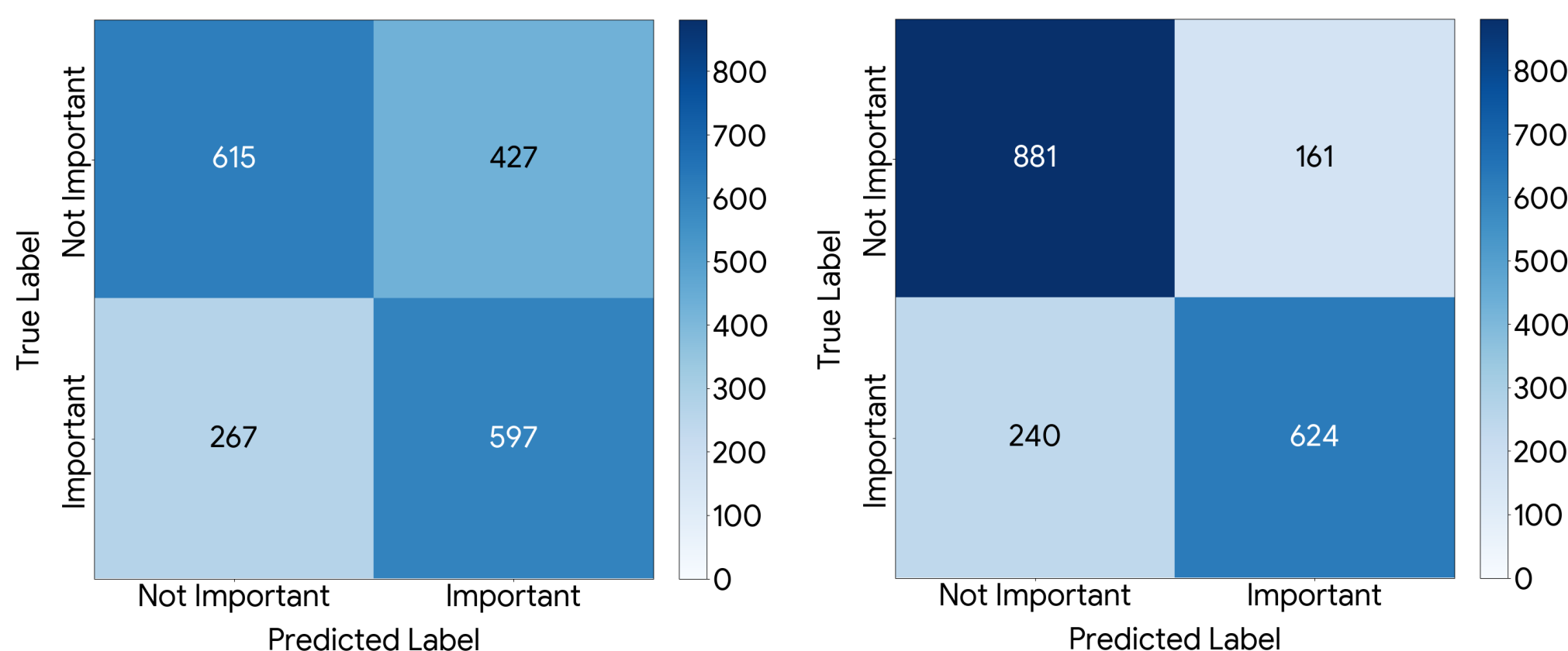
## ● Latest(改良版モデル):

- imprt/kushinada-hubert-base
- tohoku-nlp/bert-base-Japanese-v3

## 実験結果



➔ 基盤モデルや分類器の改善が有効。



➔ CNNの追加で偽陽性が大幅に減少。

- ✓ Denseは、専門用語・数式を過度に重要視。
- ✓ Dense+CNNは、短すぎる重要発言や、数式だけの重要な部分を見逃す傾向。

## まとめ

- ✓ 講義動画から重要箇所を自動抽出する手法を構築。
- ✓ 基盤モデル(ベクトル化手法)よりも分類器の改善による効果大きい。
- ✓ 特定の話者と分野に依存しない、汎用的なモデルの構築が課題。