





# Section5の概要

여보세요

CIAO

HOLA

你好

HALLO

HELLO

BONJOUR

สวัสดี

こんにちは


# 講座の内容

Section 1. 講座とBERTの概要

Section 2. シンプルなBERTの実装

Section 3. BERTの仕組み

Section 4. ファインチューニングの活用

 **Section 5. BERTの応用**

# 今回の内容

1. Section5の概要
2. BERTの活用例
3. BERTによる日本語ニュースの分類
4. BERTSUMの紹介
5. 最後に

# 教材の紹介

- 01\_news\_classification.ipynb
- ldcc-20140209.tar.gz (livedoor ニュースコーパス)



# BERTの活用例

여보세요

CIAO

HOLA

你好

BONJOUR

こんにちは

สวัสดี

HAHO

HELLO

# BERTの活用例1

- 検索エンジン

- Googleは2019年10月25日、BERTを検索エンジンに導入したと発表

- <https://blog.google/products/search/search-language-understanding-bert>

- 翻訳

- Language-agnostic BERT sentence embedding model(LaBSE)

- 109もの言語を事前学習、学習データにない言語でも性能を発揮する

- <https://ai.googleblog.com/2020/08/language-agnostic-bert-sentence.html>

# BERTの活用例2

- 文章分類

- 日立ソリューションズは「活文 知的情報マイニング」にBERTを導入

- <https://www.hitachi-solutions.co.jp/company/press/news/2019/1125.html>

- 文章要約

- BERTSUMなど



# BERTの活用例3

- **特許の分類**

- Googleは、様々な国の1億件以上の特許公報に対してBERTモデルを訓練する方法を提案

- <https://cloud.google.com/blog/products/ai-machine-learning/how-ai-improves-patent-analysis>

- **金融**

- <https://www.nttdata.com/jp/ja/data-insight/2020/101202/>

- **医療言語処理**

- etc...

# BERTによる日本語ニュースの分類

여보세요

CIAO

HOLA

你好

สวัสดี

こんにちは

BONJOUR

# 事前学習済みモデル

- **Pretrained Japanese BERT models**
  - 東北大学、乾研究室が作成
  - <https://github.com/cl-tohoku/bert-japanese>



# ファインチューニング用データ

- **livedoor ニュースコーパス**

→ <https://www.rondhuit.com/download.html#ldcc>

# BERTによる日本語ニュースの分類

- 01\_news\_classification.ipynb
- ldcc-20140209.tar.gz (livedoor ニュースコーパス)

# BERTSUMの紹介

여보세요

CIAO

HOLA

你好

สวัสดี

こんにちは

BONJOUR



# BERTSUMとは？

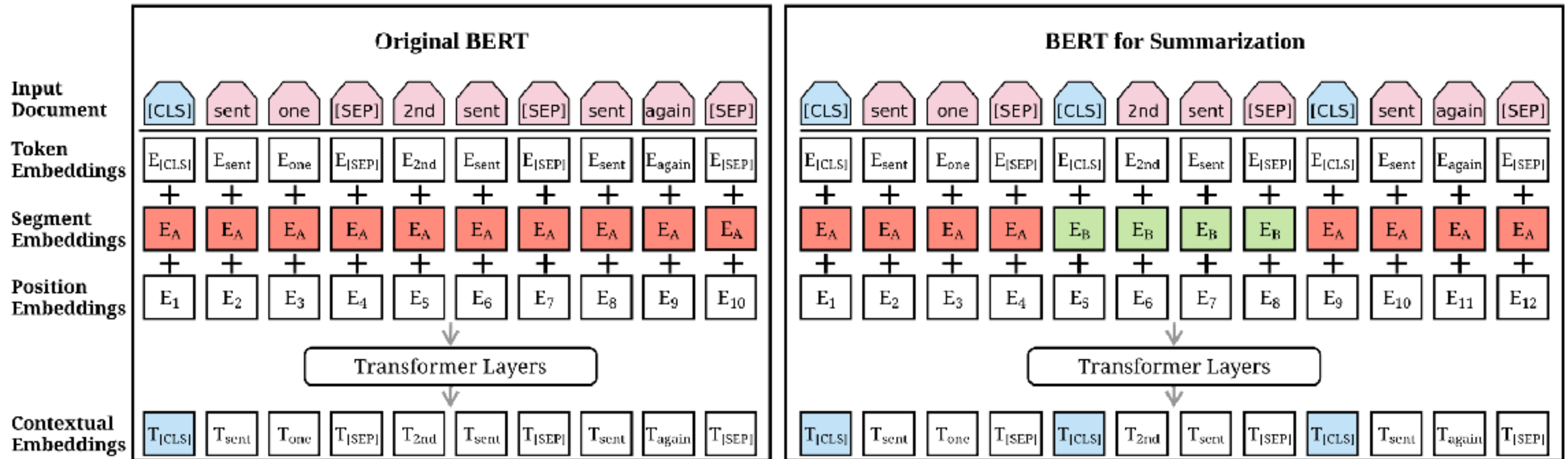
- **BERTSUM**

- BERTを拡張した文章要約のためのモデル

- <https://arxiv.org/abs/1903.10318>

- <https://arxiv.org/abs/1908.08345>

# BERTSUMのモデル



- Token Embedding: オリジナルでは先頭のみ挿入する[CLS]を、文の区切りにも挿入
- Segment Embedding: 奇数と偶数の文で異なるラベル（A、B）を割り当てる
- Encoder-Decoderの形式。Encoderは事前学習を行い、Decoderはゼロから訓練。

# BERTSUMの実装

- 英語

→ <https://github.com/nlpyang/BertSum>

- 日本語

→ <https://github.com/neilctwu/YouyakuMan>





# 最後に...

受講いただき

ありがとうございました！