

# 文符号化器のマルチタスク学習 によるテキスト分類モデルの頑健化

大橋 空\* 高山 隼矢\* 梶原 智之\*\* Chenhui Chu\*\* 荒瀬 由紀\*

\* 大阪大学大学院情報科学研究科

\*\* 大阪大学データビリティフロンティア機構

# 目次

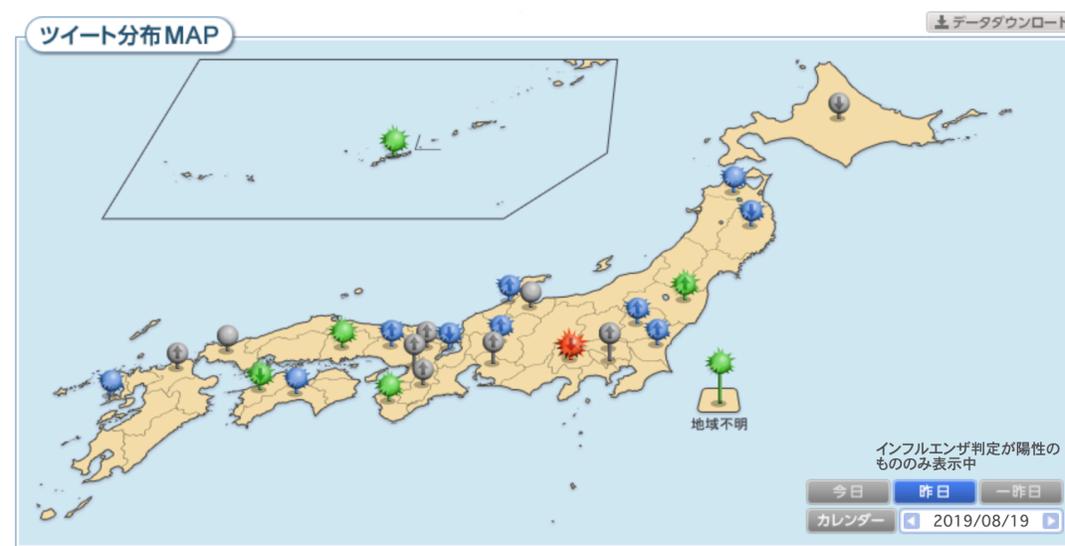
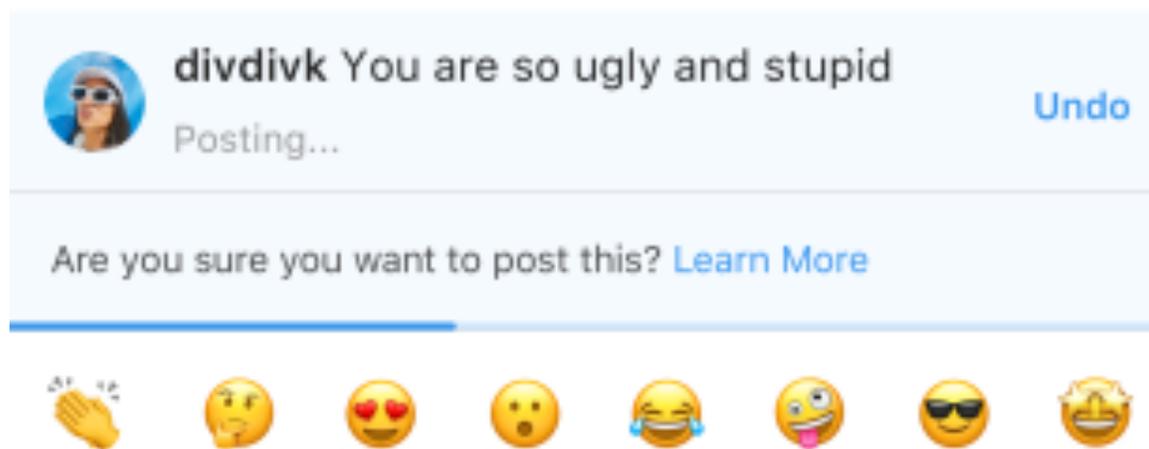
1. 背景
2. テキスト分類モデルの問題点
3. 提案手法
4. 実験
5. 考察
6. まとめ

# 目次

1. 背景
2. テキスト分類モデルの問題点
3. 提案手法
4. 実験
5. 考察
6. まとめ

# 背景

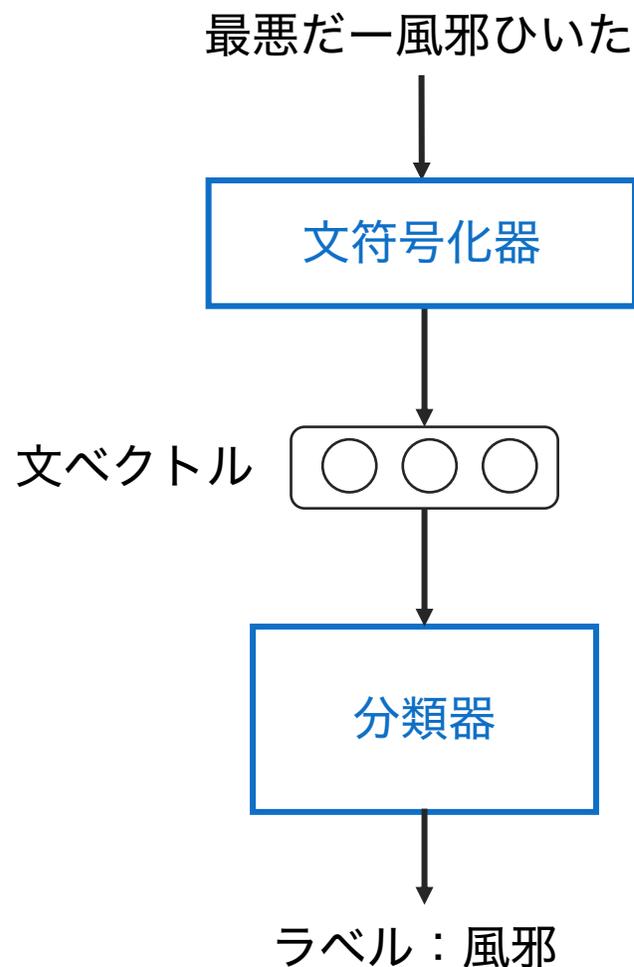
- テキスト分類は様々なシステムに有用
  - 不適切なコメントの判別 (例: Instagram)
  - 疾病流行検出 (例: インフルくん)



- テキスト分類モデルの需要が高い

<https://jp.techcrunch.com/2019/07/09/2019-07-08-instagram-anti-bullying-features/>  
[http://mednlp.jp/influ\\_map/](http://mednlp.jp/influ_map/)

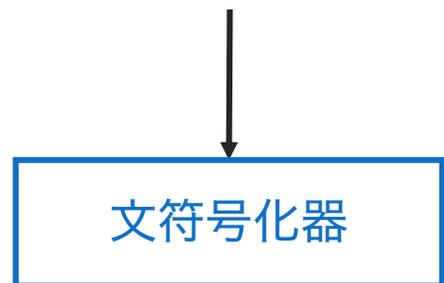
# 一般的なテキスト分類モデル



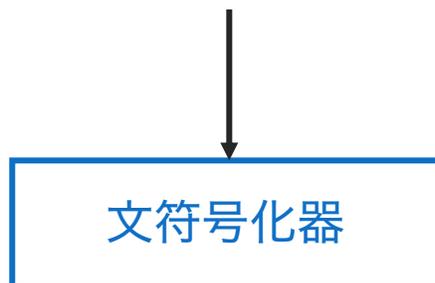
1. 文符号化器を用いて、文をベクトルに変換
2. 分類器を用いて文ベクトルをもとに、各ラベルが付与される確率を計算
3. 確率が最大（確率が閾値以上）のラベルを出力

# 一般的なテキスト分類モデル

最悪だー風邪ひいた



風邪が治らん…



- 二つの文ベクトルが類似している  
➡ 同一のラベルを付与しやすい
- 二つの文ベクトルが類似していない  
➡ 異種のラベルを付与しやすい
- 同一ラベル文の文ベクトルのみ類似するのが理想



出力：風邪 ○

正解：風邪



出力：風邪 ○

正解：風邪

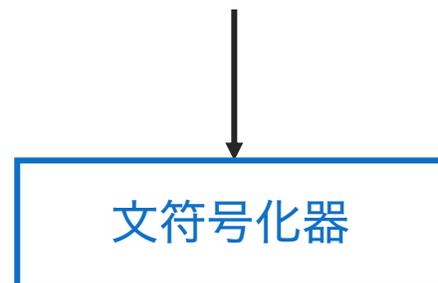
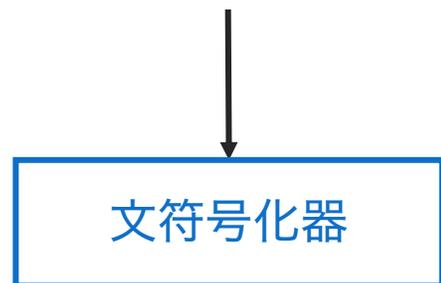
# 目次

1. 背景
2. テキスト分類モデルの問題点
3. 提案手法
4. 実験
5. 考察
6. まとめ

# 問題点：ラベルとよく共起する単語への依存

最悪だー風邪ひいた

風邪も立派な病気の一つ



■ 特定の単語とよく共起するラベルを出力すると平均的に損失が低下

■ ラベルとよく共起する単語へ強く依存

➡ 誤分類が発生

原因の一つ：

異種ラベルの文同士で

文ベクトルの類似度が高い



出力：風邪



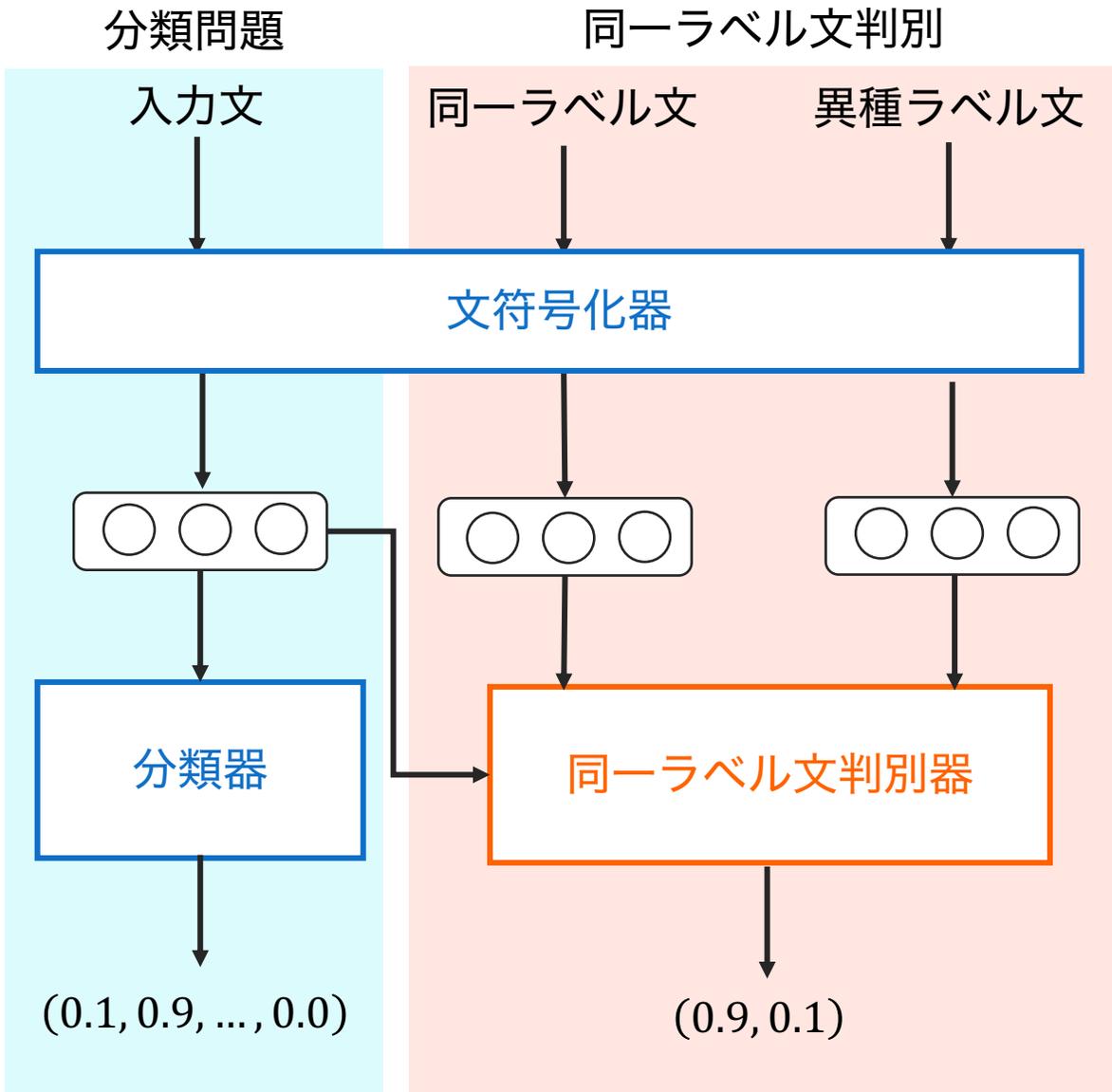
正解：風邪

出力：風邪



正解：なし

# 解決策：分類問題と同一ラベル文判別のマルチタスク学習



同一ラベル文判別：

**同一ラベル文の文ベクトルのみが類似するように訓練**

- 同一ラベル文は類似度上昇
- 異なるラベルの文は類似度減少

# 目次

1. 背景
2. テキスト分類モデルの問題点
3. 提案手法
4. 実験
5. 考察
6. まとめ

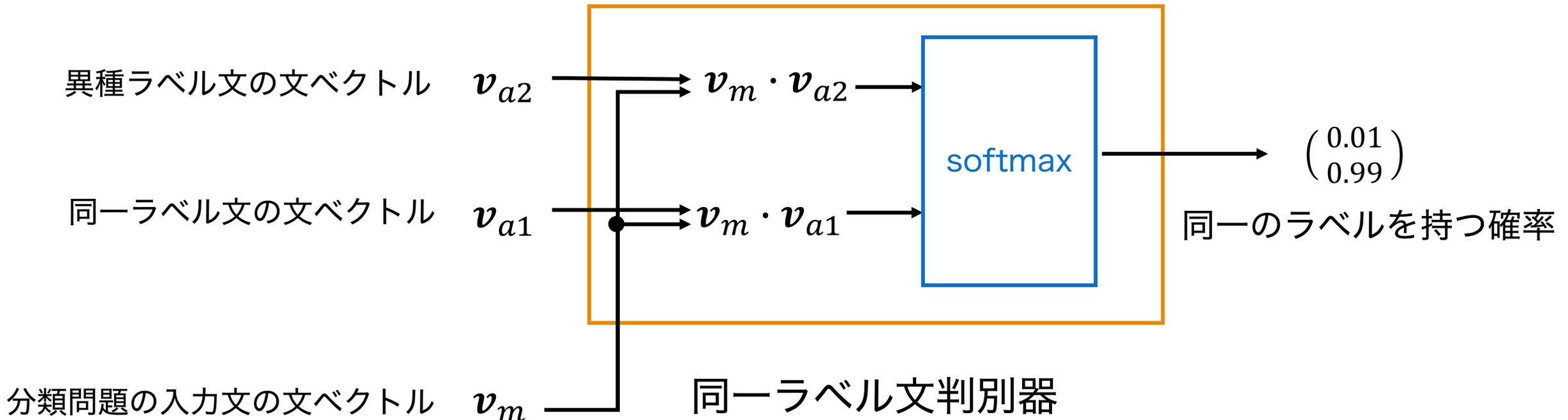
# 同一ラベル文判別

- 複数の入力文のうち、  
**どれが分類問題の入力文と同一のラベルを持つ**かを判別
- 各文に付与されているラベルが完全に一致している時  
同一ラベルとみなす
- 例：「風邪ひいた」 （ラベル：風邪）

入力文	ラベル1	ラベル2	同一ラベル
風邪が流行ってる	風邪	(なし)	0
風邪のせいで遊べない	風邪	風邪	1
風邪と咳がやばい	風邪	風邪、咳	0

# 同一ラベル文判別器

- 入力文が分類問題の入力文と**同一のラベルを持つ確率**を計算
- 文ベクトル間の**内積**を用いて確率を計算



# 目次

1. 背景
2. テキスト分類モデルの問題点
3. 提案手法
4. 実験
5. 考察
6. まとめ

# 実験：データセット

名前	内容	ラベル数	種類	入力単位
MR	映画レビューの感情分析	2	単一	文
CR	商品レビューの感情分析	2	単一	文
SST-2	映画レビューの感情分析	2	単一	文
SST-5	映画レビューの感情分析	5	単一	文
TREC	質問の種類判別	6	単一	文
SUBJ	主観 / 客観分類	2	単一	文
MedWeb	疾病情報の抽出	8	複数	文
arXiv	論文が属する分野の分類	40	複数	文書

# 実験：実験設定

## ■ 評価指標

- 単一ラベル分類：Accuracy
- 複数ラベル分類：Exact Match  
予測ラベルと正解ラベルが完全に一致しているものの割合

## ■ 使用する文符号化器

- 文分類
  - 双方向GRU<sub>[1]</sub> (bi-GRU)
  - BERT<sub>[2]</sub>
- arXiv データセット
  - Hierarchical Attention Network<sub>[3]</sub> (HAN)

[1] Chung et al. Empirical evaluation of gated recurrent neural networks on sequence modeling. NeurIPS. 2014

[2] Devlin et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. NAACL. 2019

[3] Yang et al. Hierarchical Attention Networks for Document Classification. NAACL. 2016

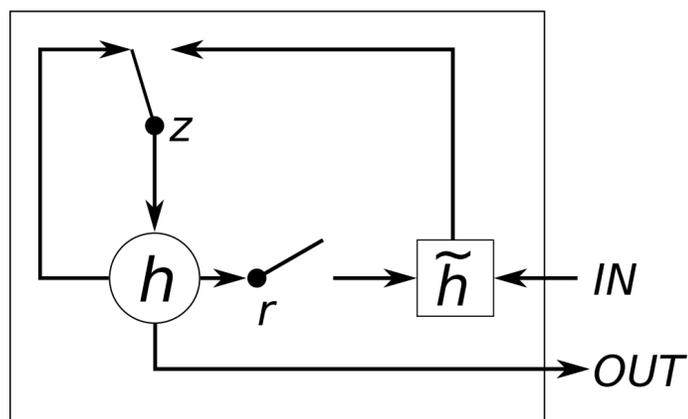
# 実験：文符号化器の概要

## ■ GRU

- ゲート機構を持つ再帰的ニューラルネット
- 更新ゲート、リセットゲートを用いて情報の取捨選択を行う

## ■ 双方向GRU

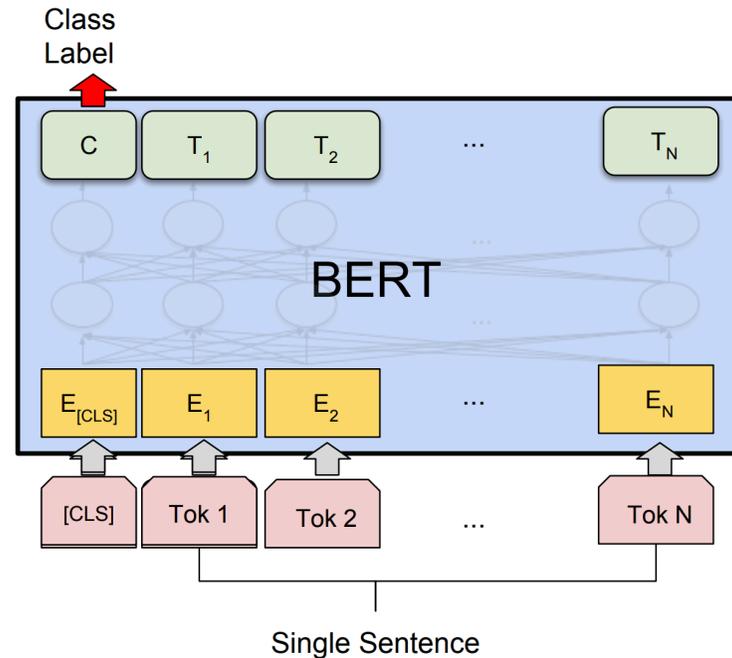
- 順方向、逆方向にGRUを適用し、その出力結果を連結



# 実験：文符号化器の概要

## ■ BERT

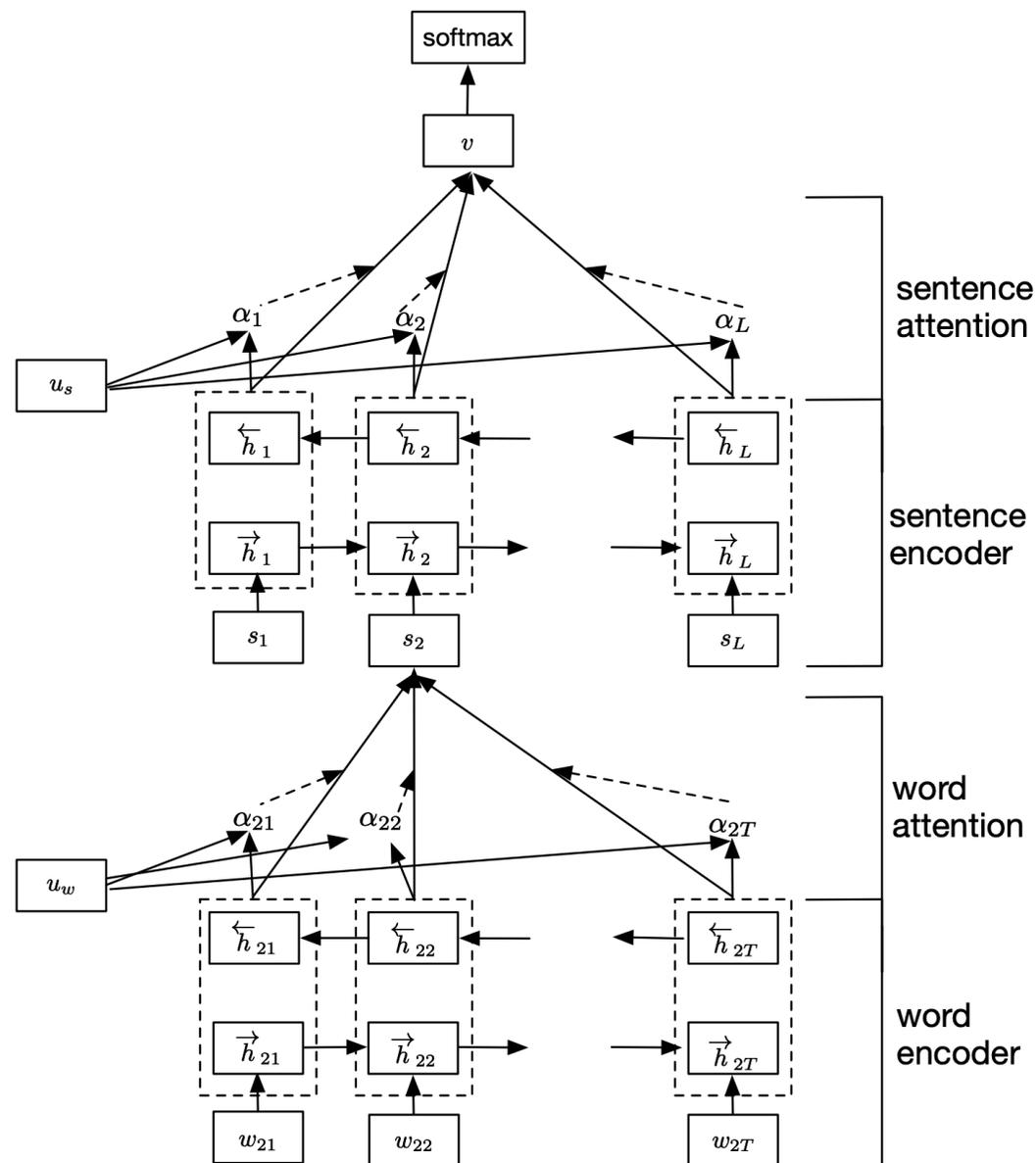
- Transformer の encoder 部分を12層に拡張したモデル
- 複数ヘッドを持つ自己注意機構
- [CLS]に対応するトークンの出力を文ベクトルとする



# 実験：文符号化器の概要

## ■ HAN

- 文書分類に特化したモデル
- 文書中の各文を一つのベクトルに変換
- 出力された各ベクトルをさらに一つのベクトルに変換



## 結果：単一ラベル分類

	MR	CR	SST-2	SST-5	TREC	SUBJ
bi-GRU	77.1	<b>78.8</b>	85.8	42.6	86.8	92.2
bi-GRU + prop	<b>77.9</b>	78.0	85.8	<b>42.7</b>	<b>87.4</b>	<b>92.4</b>
BERT	86.3	89.3	<b>92.8</b>	51.7	96.6	<b>96.6</b>
BERT + prop	<b>86.7</b>	<b>89.7</b>	92.5	<b>52.0</b>	<b>97.2</b>	96.5

- MR、SST-5、TRECでは**両方の文符号化器で改善**
- SST-2はどちらの文符号化器でも改善せず
  - ➡ SST-2は他と比べ、ラベルとよく共起する単語が少ないことが原因

## 結果：複数ラベル分類（文分類）

	MedWeb(ja)	MedWeb(en)	MedWeb(zh)
bi-GRU	83.0	76.3	81.3
bi-GRU + prop	<b>83.6</b>	<b>78.3</b>	<b>81.5</b>
BERT	85.5	83.5	85.8
BERT + prop	<b>86.0</b>	83.5	<b>86.0</b>

**MedWeb (en) での BERT 以外全ての場合で性能を改善**

## 結果：複数ラベル分類（文書分類）

	arXiv
HAN	41.7
HAN + prop	<b>42.8</b>

入力が文書である場合でも性能改善に成功

➡ **文書分類においても一定の効果**がある

# 考察

- 提案手法は単語とラベルの依存解消が目標
- 以下の二つを検証
  1. 性能の上昇幅と、単語とラベルに関する相互情報量の期待値の関係
    - **ラベルと単語の依存の度合いが強くなるにつれ、提案手法の効果が大きくなる**ことが期待
  2. 文ベクトル生成におけるBERTの注意箇所の違い
    - **異種ラベル間では異なる箇所へ注意する**ことが期待

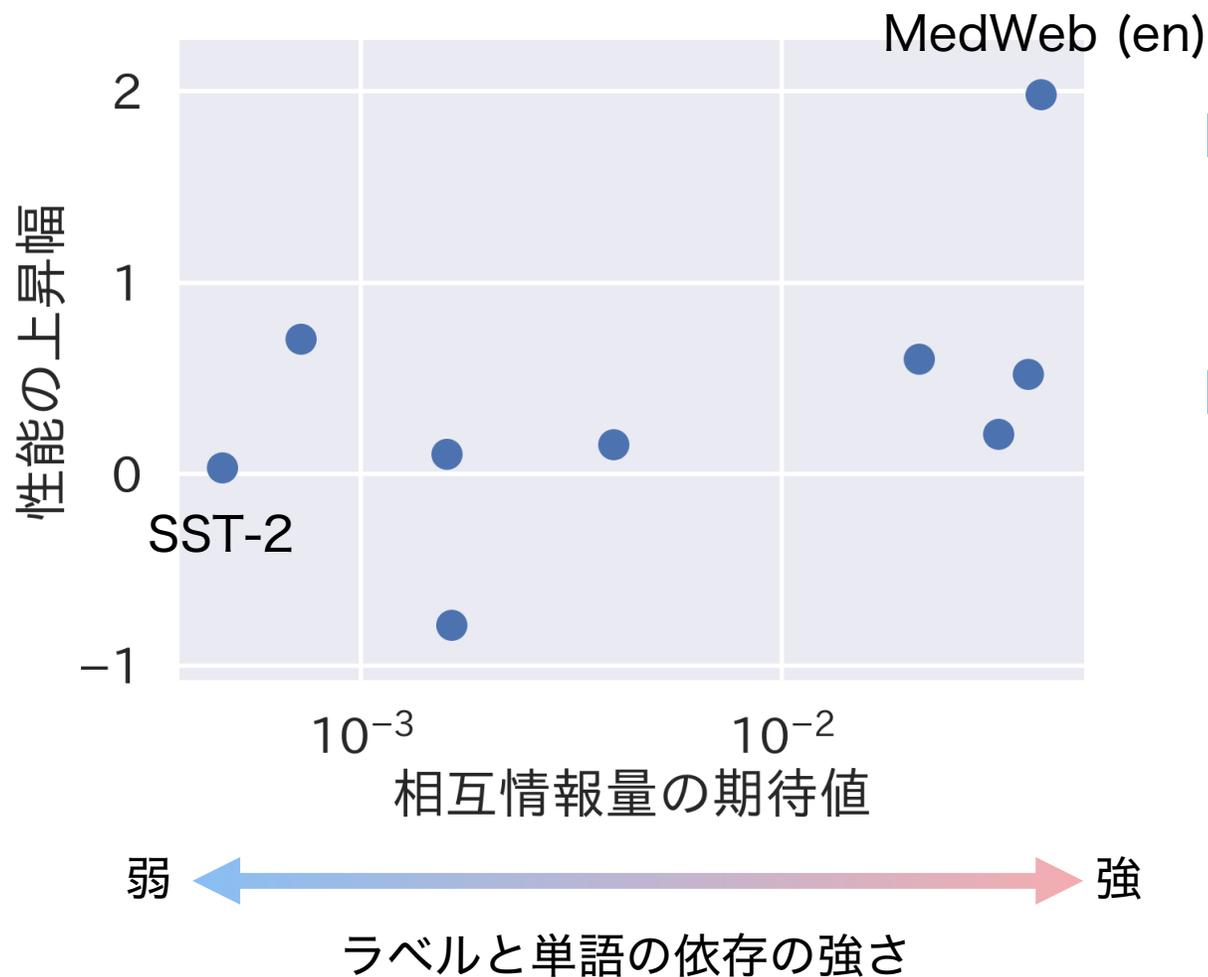
# 目次

1. 背景
2. テキスト分類モデルの問題点
3. 提案手法
4. 実験
5. 考察
6. まとめ

# 考察

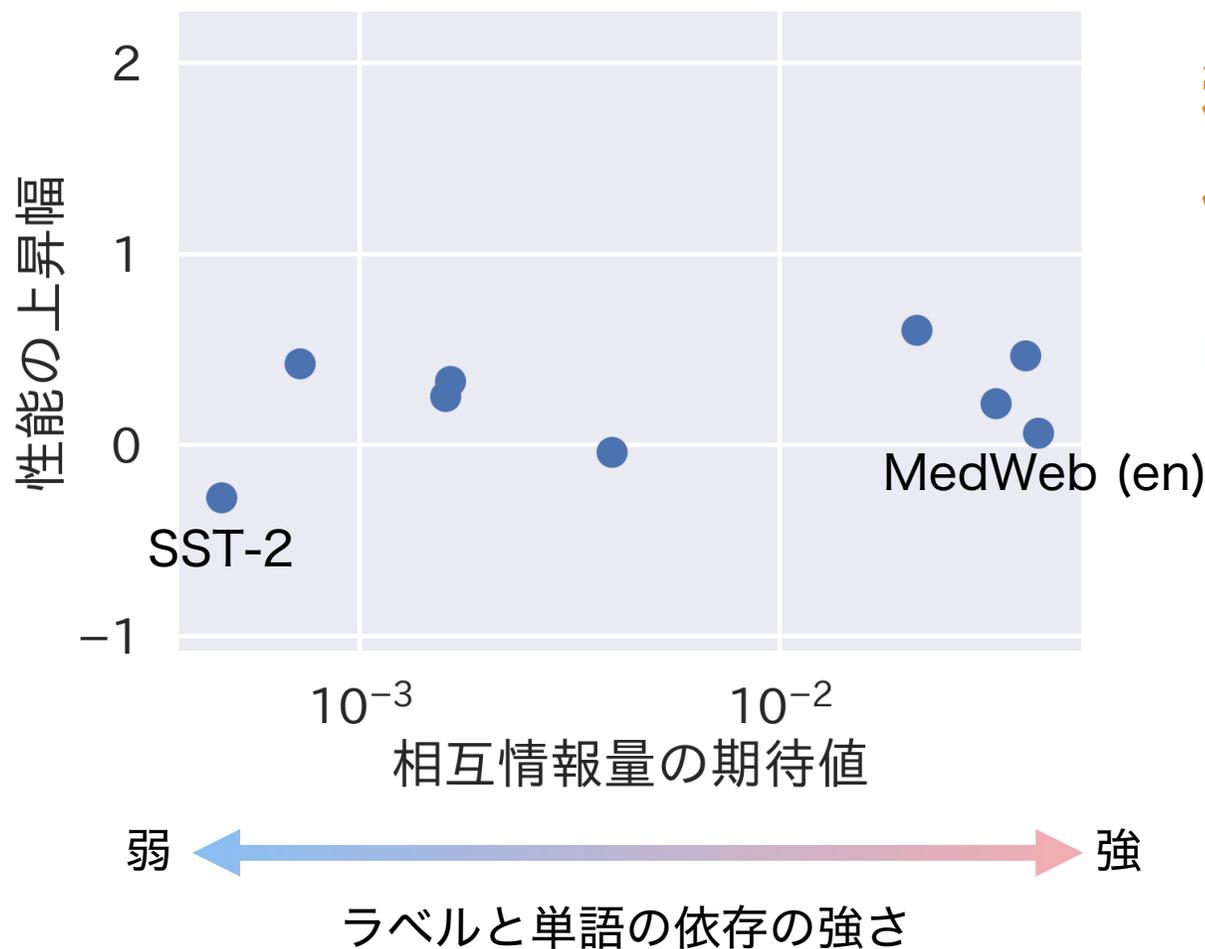
- 提案手法は単語とラベルの依存解消が目標
- 以下の二つを検証
  1. 性能の上昇幅と、単語とラベルに関する相互情報量の期待値の関係
    - **ラベルと単語の依存の度合いが強くなるにつれ、提案手法の効果が大きくなる**ことが期待
  2. 文ベクトル生成におけるBERTの注意箇所の違い
    - **異種ラベル間では異なる箇所へ注目する**ことが期待

# 考察：ラベルとよく共起する単語への効果（GRU）



- 期待値が増加するにつれ、性能の上昇幅が増加
- ラベルとよく共起する単語への依存が防げている

# 考察：ラベルとよく共起する単語への効果（BERT）



程度は低いですが、期待値の上昇に伴い上昇幅が増加

■ BERTでも、ラベルとよく共起する単語への依存が防げている

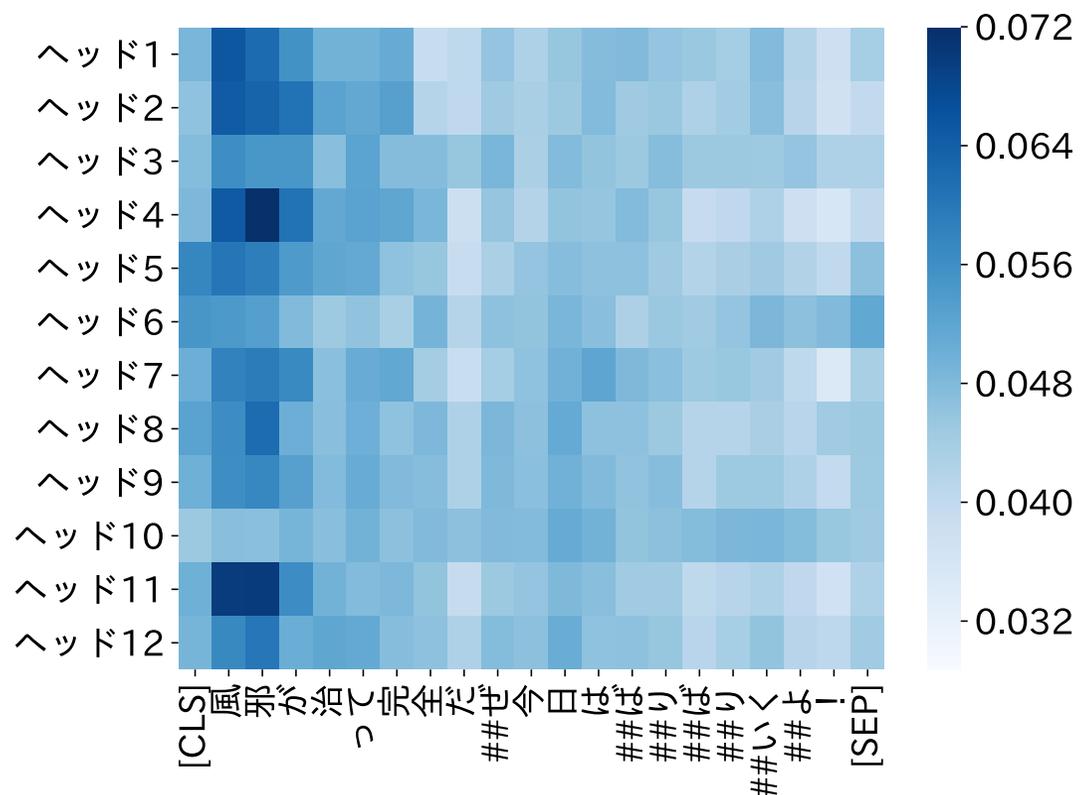
# 考察

- 提案手法は単語とラベルの依存解消が目標
- 以下の二つを検証
  1. 性能の上昇幅と、単語とラベルに関する相互情報量の期待値の関係
    - **ラベルと単語の依存の度合いが強くなるにつれ、提案手法の効果が大きくなる**ことが期待
  2. 文ベクトル生成におけるBERTの注意箇所の違い
    - **異種ラベル間では異なる箇所へ注目する**ことが期待

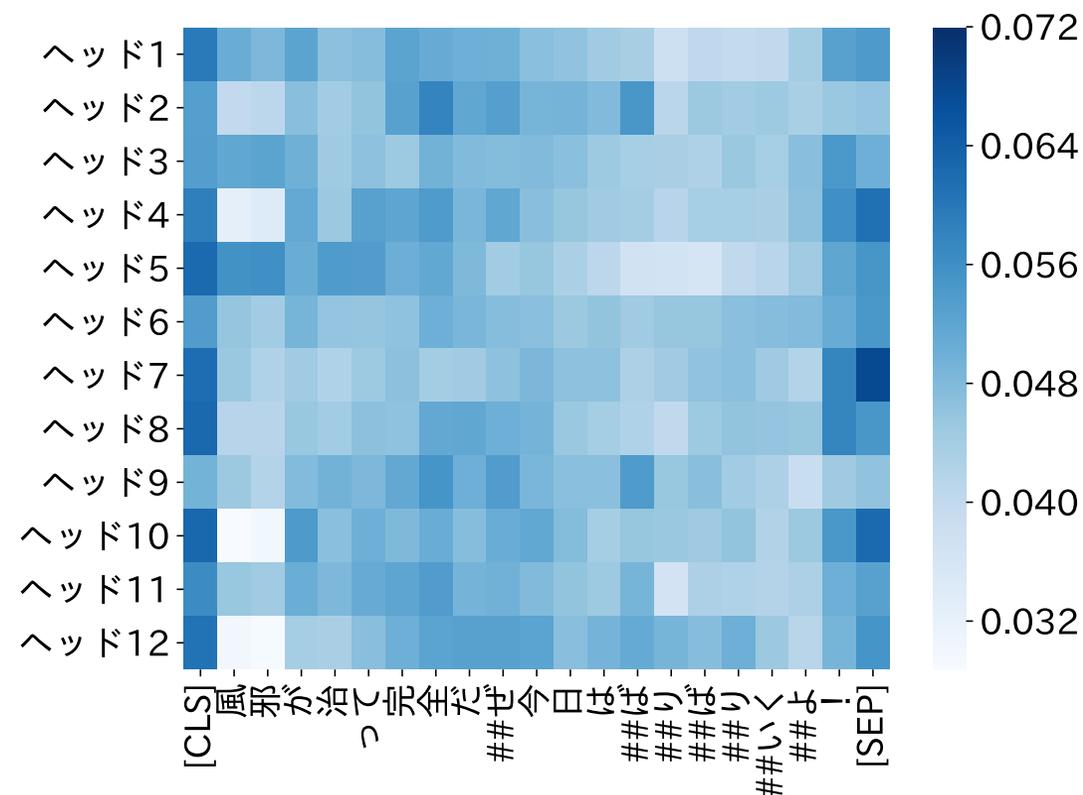
# 考察：注目箇所の違い（BERT）

■ 入力文：風邪が治って完全だぜ今日はバリバリいくよ

■ 提案手法により適切な箇所へ注意するように



BERT（出カラベル：風邪）

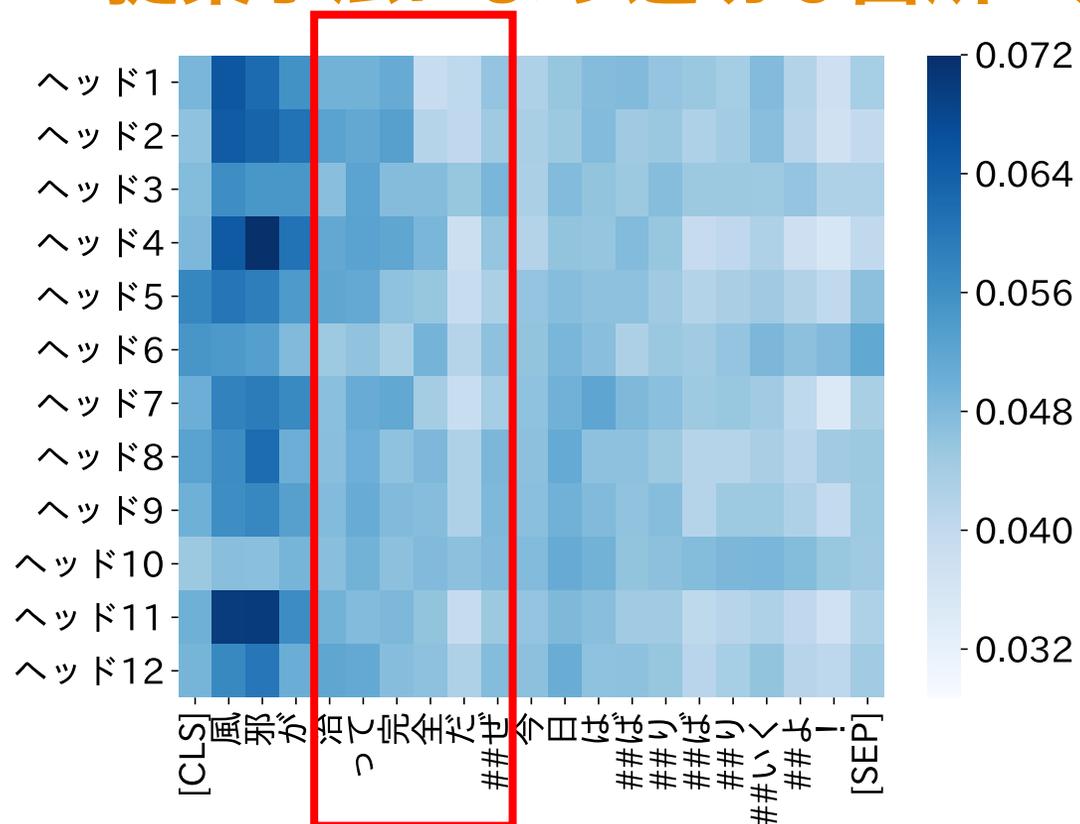


BERT + prop（出カラベル：なし）

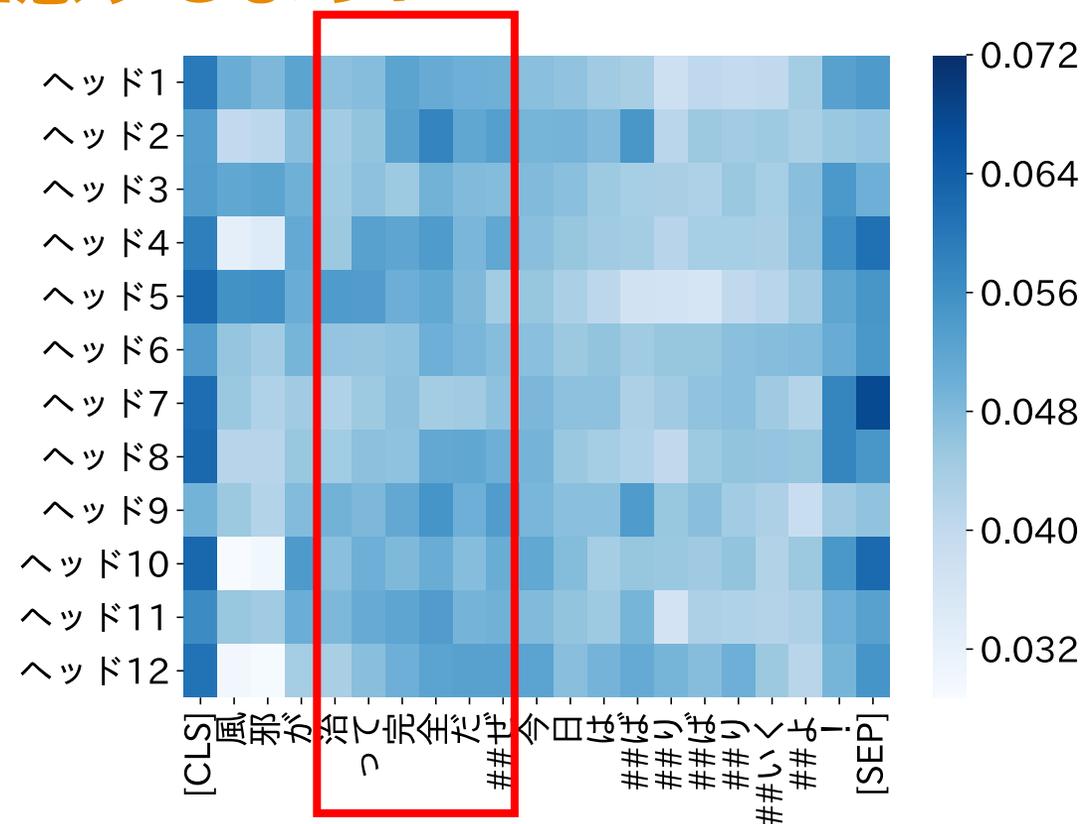
# 考察：注目箇所の違い（BERT）

■ 入力文：風邪が治って完全だぜ今日はバリバリいくよ

■ 提案手法により適切な箇所へ注意するように



BERT（出カラベル：風邪）



BERT + prop（出カラベル：なし）

# 目次

1. 背景
2. テキスト分類モデルの問題点
3. 提案手法
4. 実験
5. 考察
6. まとめ

## まとめ

- 一般的なテキスト分類モデルはラベルとよく共起する単語へ依存しやすい
- 同一ラベル文判別とのマルチタスク学習を提案
- 単一、複数ラベル分類両方のタスクで性能の改善を確認

