

# ラベル間の意味の違いを考慮した Few-shot テキスト分類

大橋空<sup>1</sup>、高山隼矢<sup>1</sup>、梶原智之<sup>2</sup>、荒瀬由紀<sup>1</sup>

<sup>1</sup>大阪大学大学院情報科学研究科

<sup>2</sup>愛媛大学大学院理工学研究科

# 背景：few-shot テキスト分類

- BERTなどの大規模言語モデルは分類問題に極めて有効
  - 高い分類性能
  - あらゆる分類タスクに有効
- 事例が数個しかない場合、容易に過学習<sup>[1]</sup>
  - ➡ Few-shot テキスト分類

# Few-shot 分類：問題設定

## ■入力:

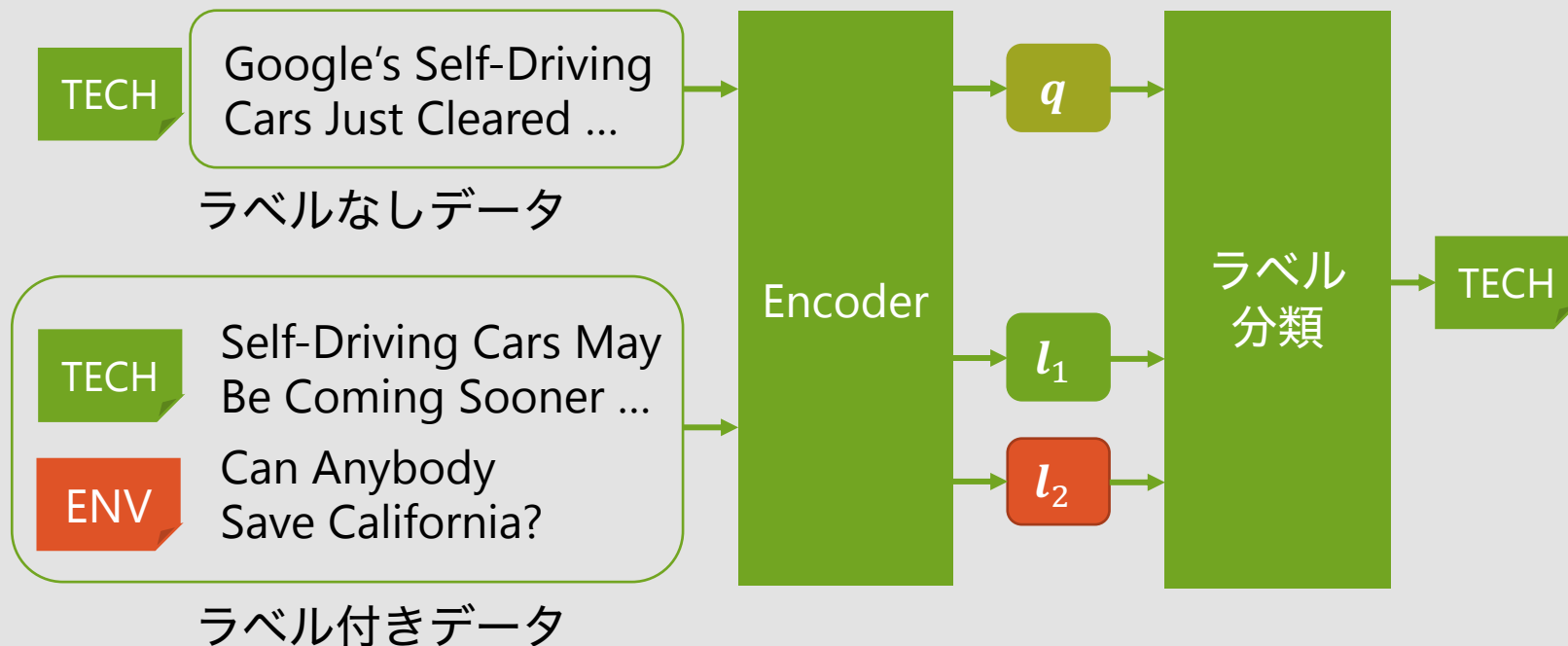
- ラベル付きデータ：テキストとラベルの組の集合（訓練データ）
- ラベルなしデータ：テキストの集合（テストデータ）

## ■出力： ラベルなしデータの予測ラベル



# 関連研究：ProtoNet

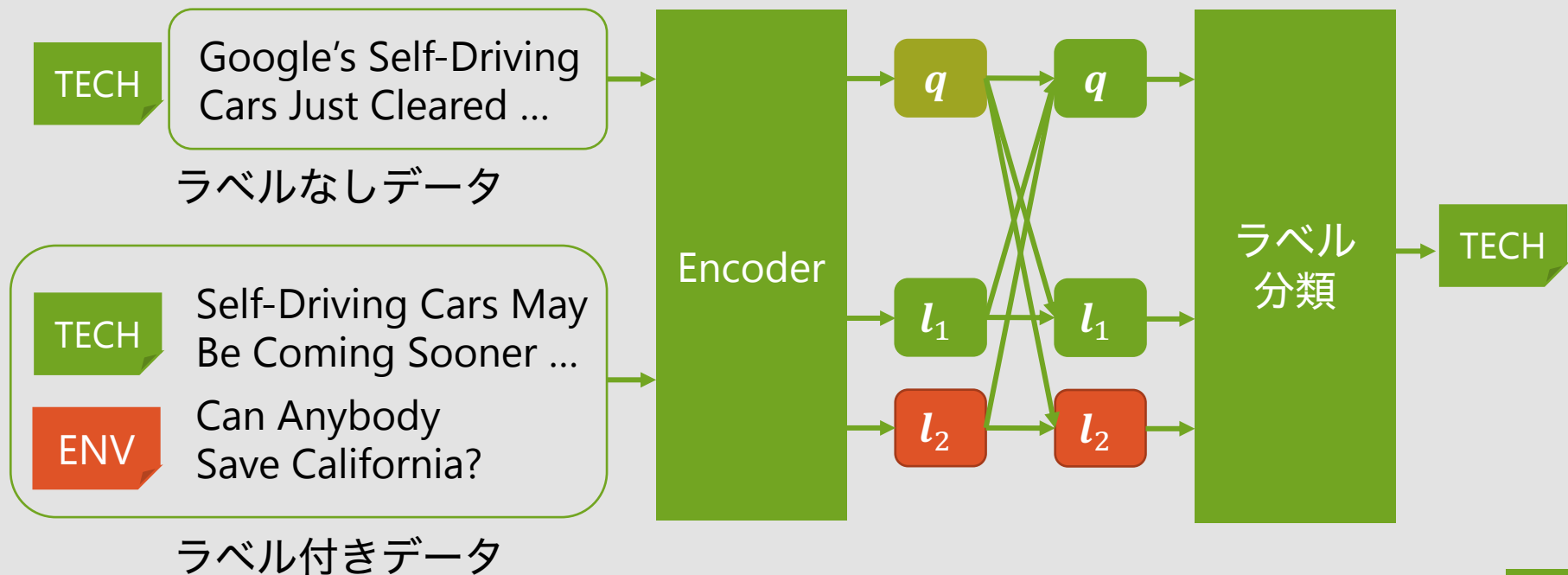
- ProtoNet<sup>[1]</sup>：シンプルなモデルで比較的高い分類精度を達成
- ラベル付きデータ、ラベルなしデータからベクトル表現を生成しラベルなしデータを分類



[1] Snell et al., Prototypical Networks for Few-shot Learning, NeurIPS, 2017

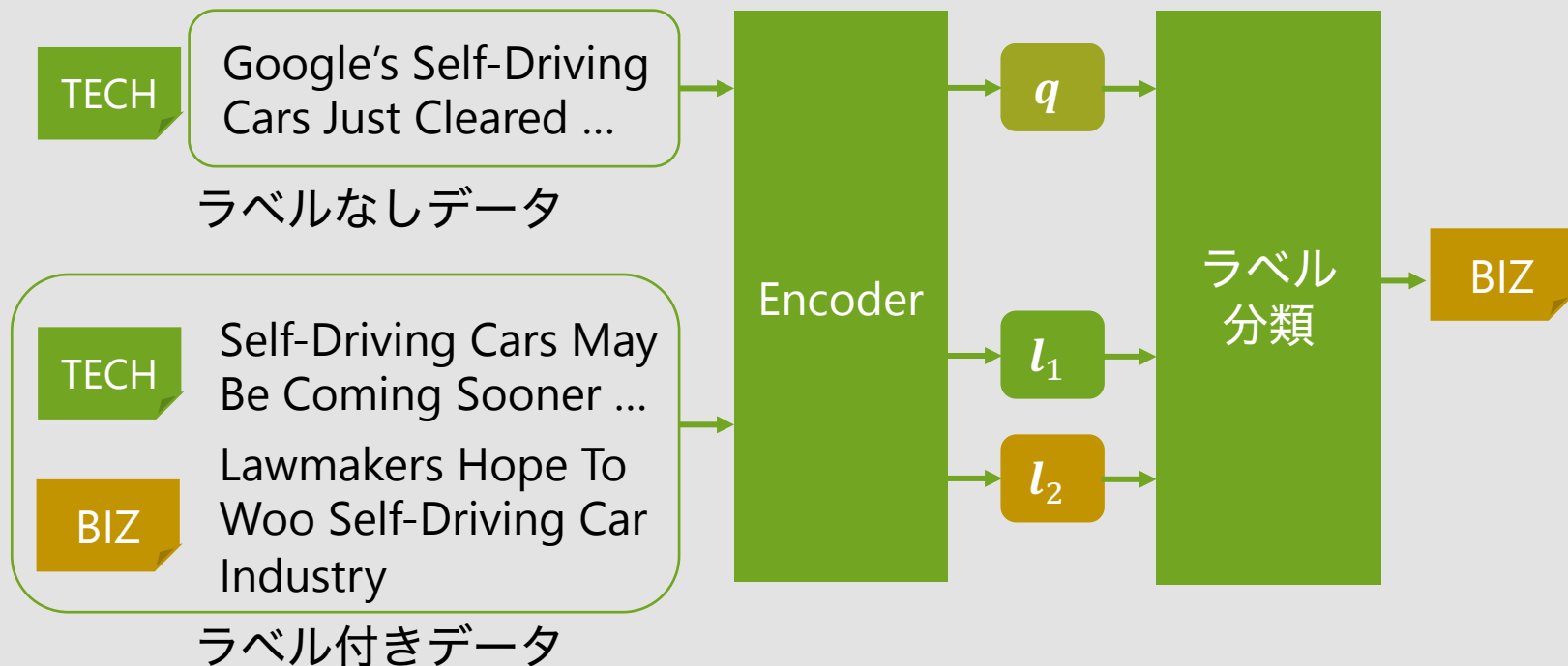
# 関連研究：Ye and Ling (2019)

- ラベル付きデータとラベルなしデータを比較しながらベクトル表現を生成
- ラベル付きデータ、ラベルなしデータ双方の情報をベクトル表現に織り込める



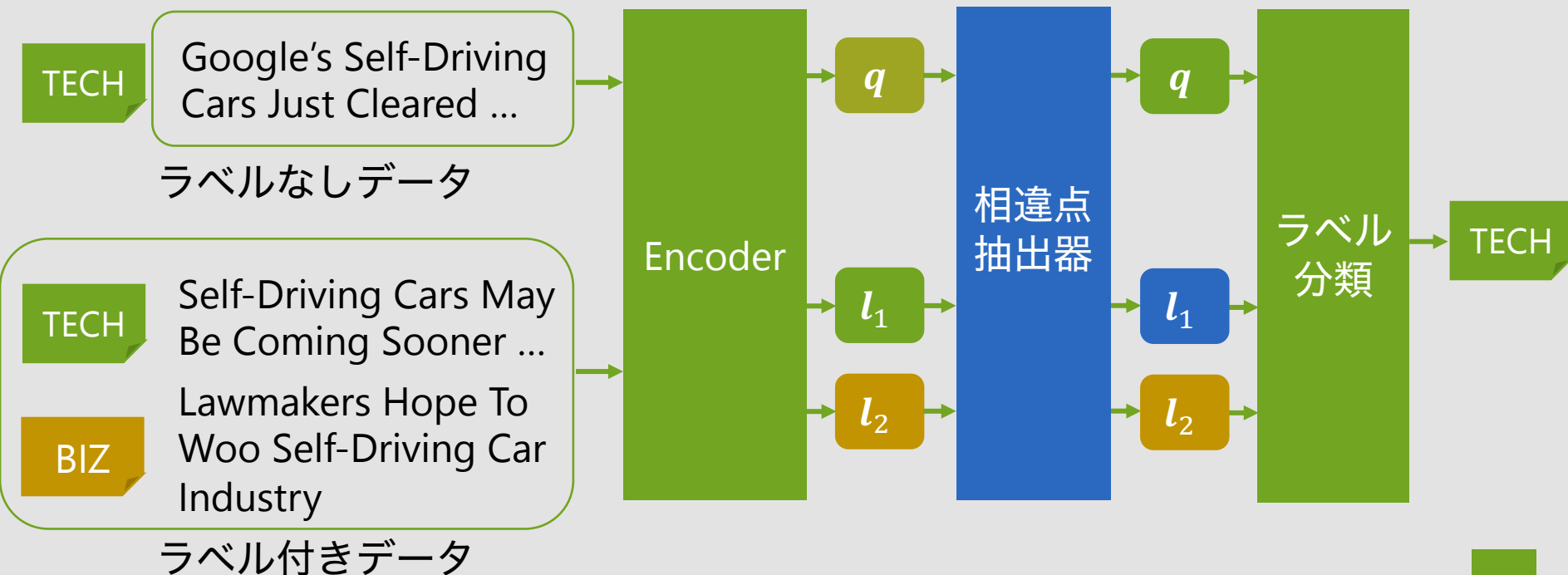
# 関連研究：問題

- 意味的に関連のあるラベルが存在する場合、分類が困難
  - ラベル付きデータのベクトル表現が類似してしまう



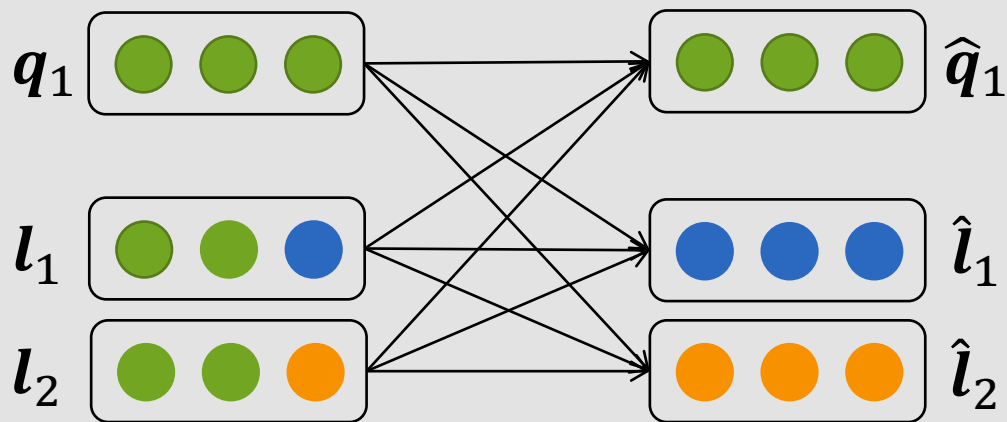
# 提案手法

- ラベル付きデータ間を比較するモデルを追加し、相違点を抽出
- ラベル間の違いを明確にすることで正確な分類を可能に



# 提案手法：相違点抽出器

- 自己注意機構を用いてラベル表現間の相違点を抽出
  - 全ラベル表現を参照し、相違点となっている箇所に着目
- ラベル表現間の相互情報量が低いラベル表現を出力
  - 特有の情報のみ持つラベル表現は他のラベル表現に依存しない



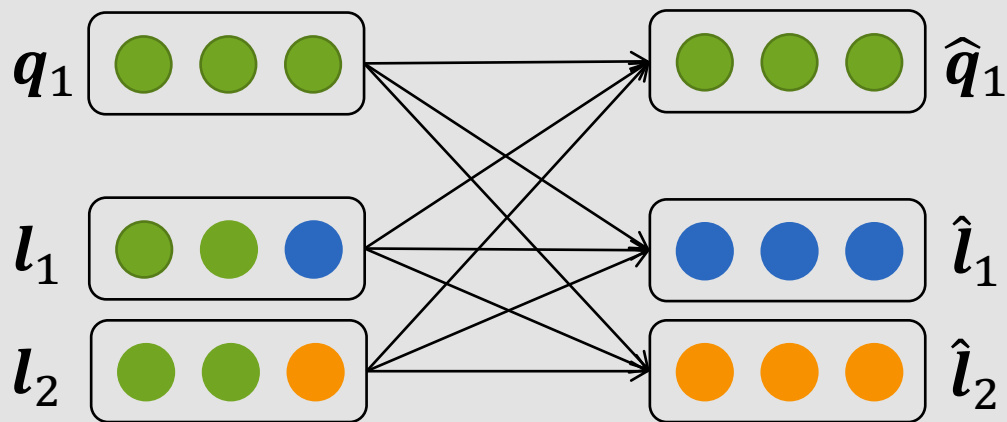


# 提案手法：相違点抽出器

■相違点抽出器の損失関数：

$$\hat{L} = \sum_{1 \leq i < j \leq N} I(\hat{l}_i; \hat{l}_j)$$

●  $\hat{l}_i$ ：相違点抽出器から出力されたラベル  $i$  のベクトル表現



# 相互情報量の最適化

- 一般に相互情報量  $I$  を直接求めるのは困難
- Cheng らによる相互情報量の上界  $\hat{I}$  を最小化
  - ベクトル表現から上界を推定可能

# 実験

- 相違点抽出器、損失関数  $\hat{L}$  の有効性を評価
- データセット

| データセット   | 概要                    |
|----------|-----------------------|
| Huffpost | ニュースの見出しからカテゴリを予測     |
| FewRel   | 指定された二つのエンティティ間の関係を予測 |

- 問題設定

- ラベル数：5
- ラベル毎の事例数：1

- ベースライン

- ProtoNet
- MLMAN<sub>[2]</sub>

# 結果

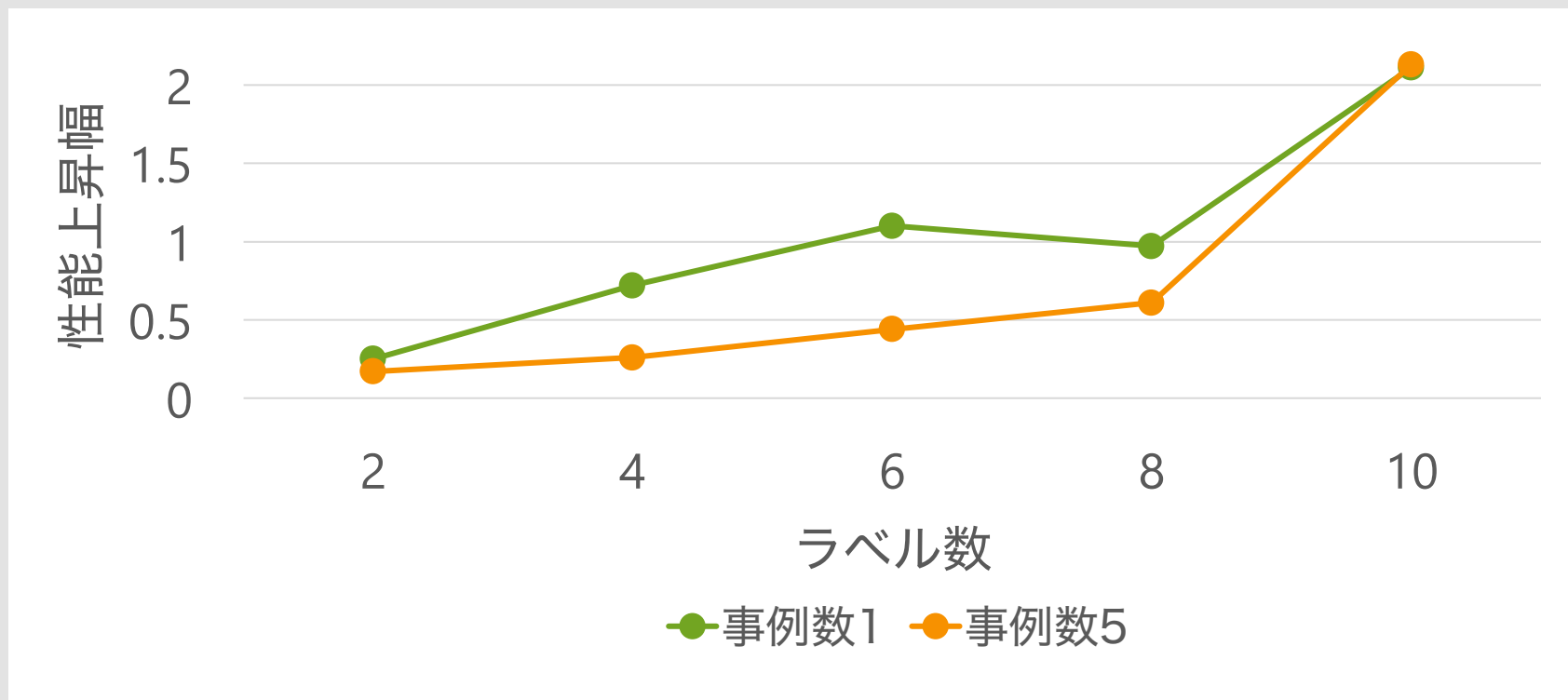
|                               | Huffpost | FewRel |
|-------------------------------|----------|--------|
| ProtoNet                      | 51.03    | 78.61  |
| ProtoNet + 相違点抽出器             | 51.76    | 77.35  |
| ProtoNet + 相違点抽出器 + $\hat{L}$ | 52.34*   | 79.52* |
| MLMAN                         | 47.08    | 73.61  |
| MLMAN + 相違点抽出器                | 49.37    | 74.38  |
| MLMAN + 相違点抽出器 + $\hat{L}$    | 48.98    | 78.21* |

- 双方のデータセットで性能が改善
- $\hat{L}$  をなくすと性能が低下
- ラベル間の相違点が Few-shot 分類性能の向上に貢献

# 分析

- 分類対象のラベル数や事例数による有効性の変化を調査
- ProtoNet と ProtoNet + 相違点抽出器 +  $\hat{L}$  間の性能上昇幅を測定
- 実験設定
  - ラベル数 : [2, 4, 6, 8, 10]
  - 事例数 : [1, 5]

# 分析：ラベル数や事例数への依存性



- 分類対象のラベル数が多いほど有効
- 事例数が少ないほど有効

## まとめ

- 意味的に関連したラベルの分類を容易にするため、ラベル間の相違点を抽出するモデルを提案
- 2つのデータセットにおいて、ベースラインより有意に性能が上昇し、有効性を確認
- 事例数が少ないほど、かつラベル数が多いほど有効性が高まる

## 予備実験

- 正解ラベルとの類似度が高いもの、低いものを全体の50%削除し、性能の違いを調査

### ■ 結果

| 高類似度50%削除 | 低類似度50%削除 |
|-----------|-----------|
| 94.09     | 67.12     |

- 類似度が高い不正解ラベルが分類精度を低下させている