

# Distinct Label Representations for Few-shot Text Classification

Sora Ohashi\*, Junya Takayama\*, Tomoyuki Kajiwara\*\*, Yuki Arase\* [ohashi.sora@ist.osaka-u.ac.jp](mailto:ohashi.sora@ist.osaka-u.ac.jp)

\*Osaka University \*\*Ehime University

## Background: Few-shot Text Classification

- Pre-trained models (e.g. BERT) easily overfit to train data when only a few examples are available
- Few-shot text classification aims to solve this problem in such a low-resource scenario

A few-shot text classification model is given two sets of texts: Support Set and Query Set. The model generates vector representations from Support Set and Query Set, and then it predicts the label of texts in Query Set.

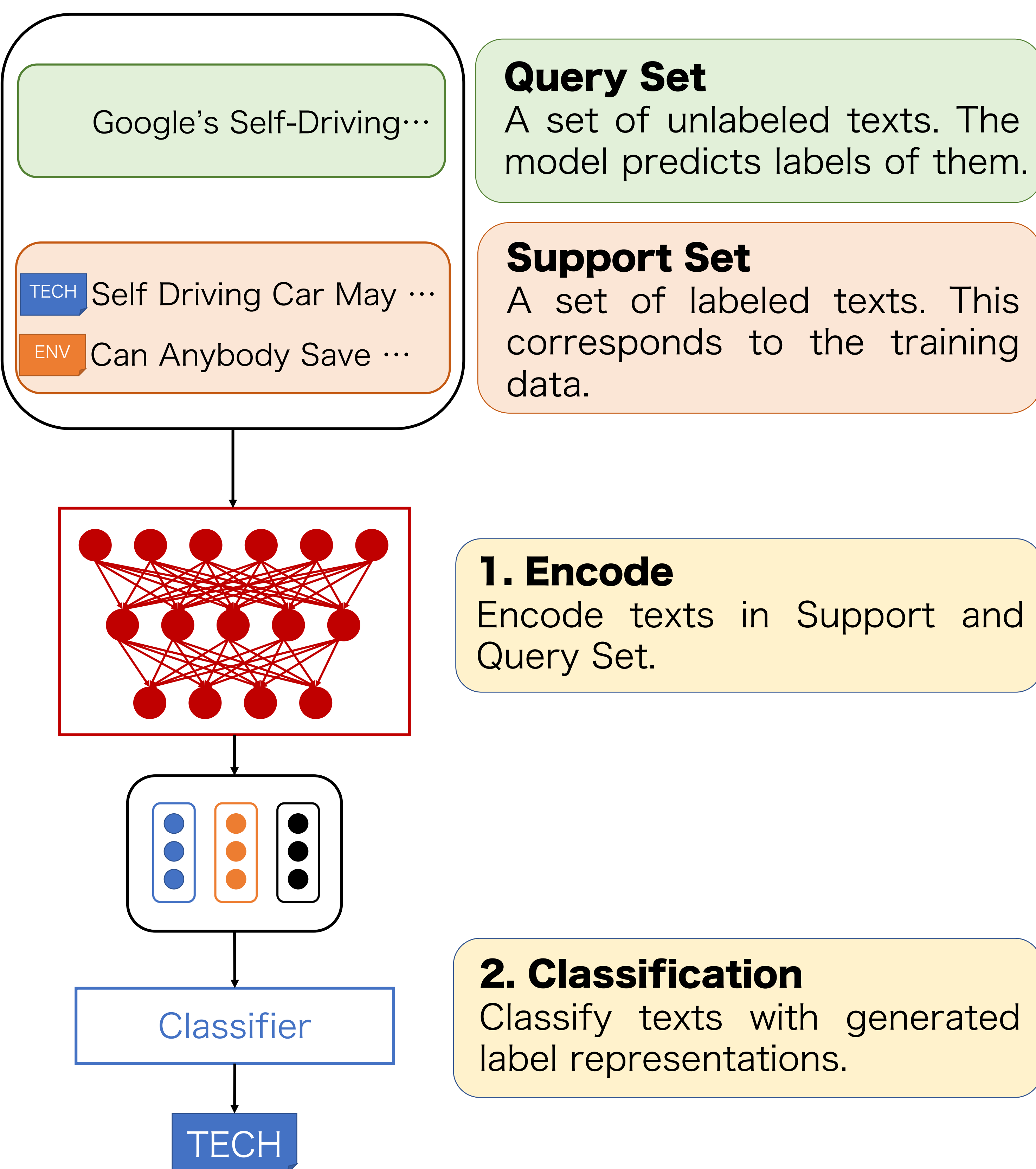


Fig 1: Few-shot Classification model

## Approach: Difference Extractor

- Existing models do not consider semantic relations among labels
- Our method extracts the difference among labels and classify texts accurately

We introduce the difference extractor, which extracts the information specific to each label and embeds them into label representations.

- We assume that label representations presenting specific information should have low mutual information
- We add a new loss function using mutual information based on the assumption

### Mutual Information-based Loss

$$L = L_c + \alpha L_{DE}$$

$$L_{DE} = \sum_{1 \leq i < j \leq N} \hat{I}(l_i; l_j)$$

$L_c$ : A loss function of classification

$l_i$ : A label representation

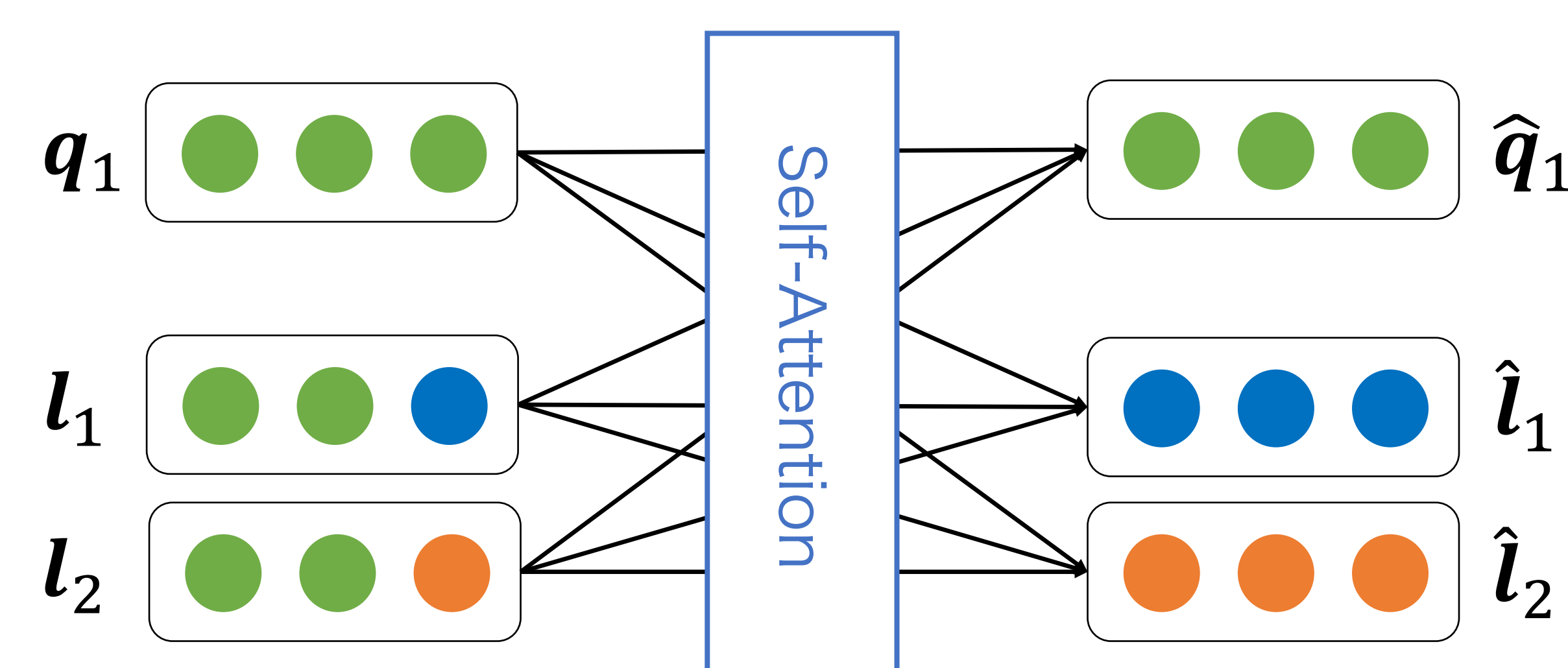


Fig 2: Difference Extractor

## Evaluation: 5-way 1-shot Classification

### Datasets

	Task
Huffpost	Classification of news category
FewRel	Classification of the relationship between entities

**The performance significantly improved across datasets and models**

	Huffpost	FewRel
ProtoNet	51.03	78.61
+ Difference Extractor	<b>51.76</b>	77.35
+ Difference Extractor + $\hat{L}$	<b>52.34*</b>	<b>79.52*</b>
MLMAN	47.08	73.61
+ Difference Extractor	<b>49.37</b>	<b>74.38</b>
+ Difference Extractor + $\hat{L}$	<b>48.98</b>	<b>78.21*</b>

Table 1: Experimental results

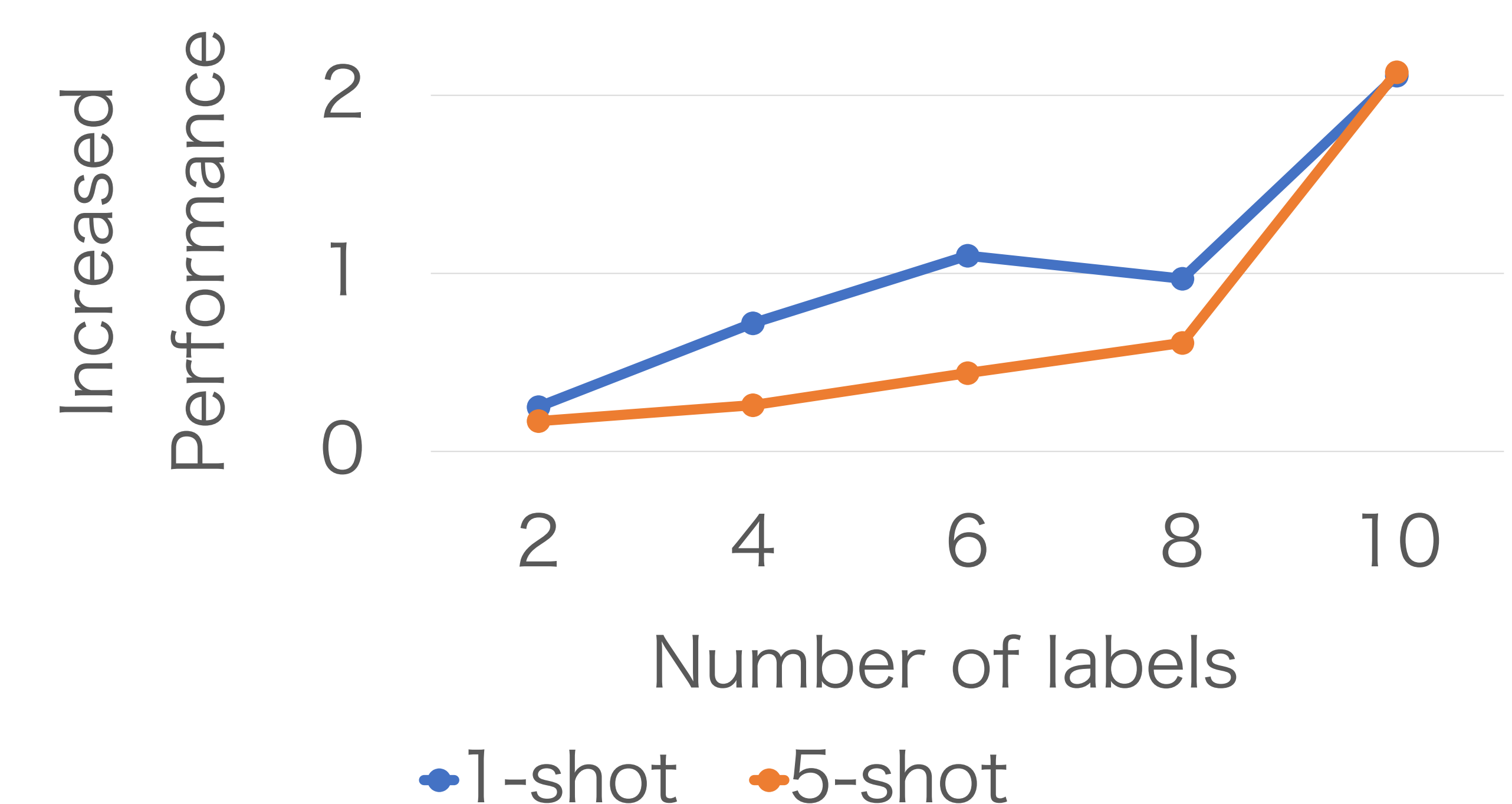


Fig 3: Change in performance improvement with number of labels and examples