# Enhancing class understanding via prompt-tuning for zero-shot text classification

YuhaoDan[1]   Jie Zhou[2]   Qin Chen[1]   Qingchun Bai[3]   Liang He[1]

[1]School of Computer Science and Technology, East China Normal University, China
[2]School of Computer Science, Fudan University, China
[3]Shanghai Open University, China

# What is **zero-shot text classification**?

- Notation definition

  - $S$ denotes seen classes (samples can be seen during training)

  - $U$ denotes unseen classes (samples can not be seen during training)

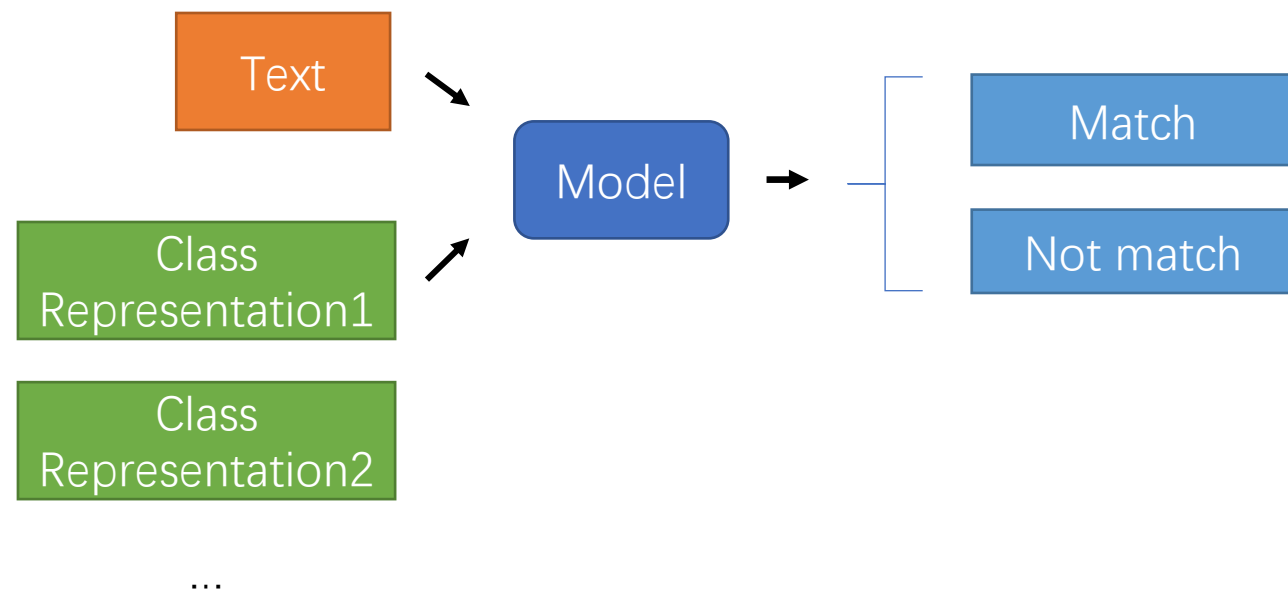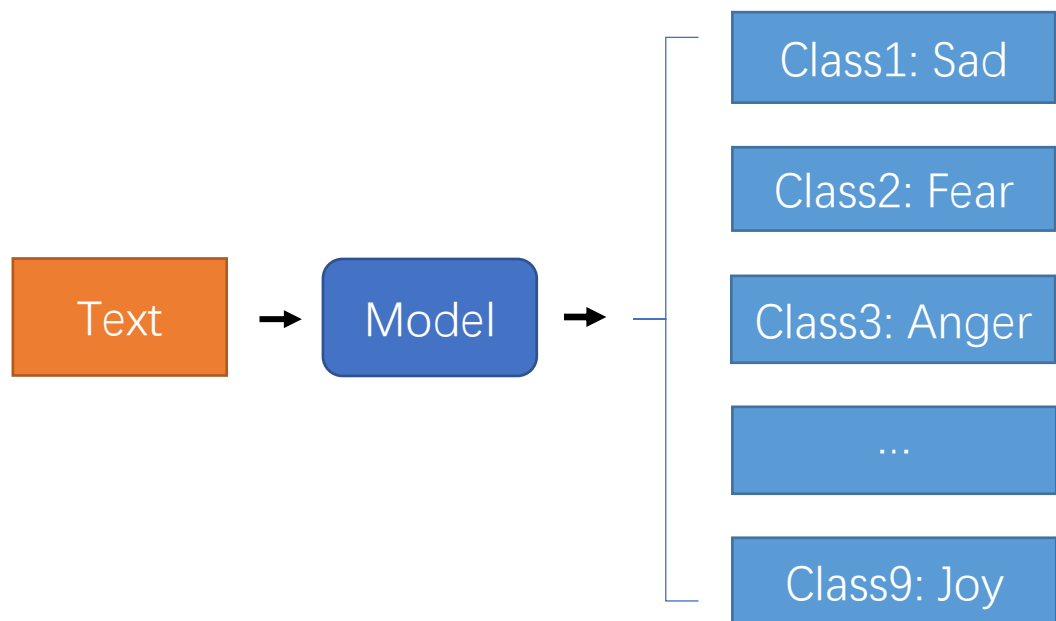- Given training data belonging to $S$, we aim to learn a model:

$$f(\cdot): X \rightarrow Y$$

Where: $Y = S \cup U$

# Classification Model VS Zero-shot Classification Model
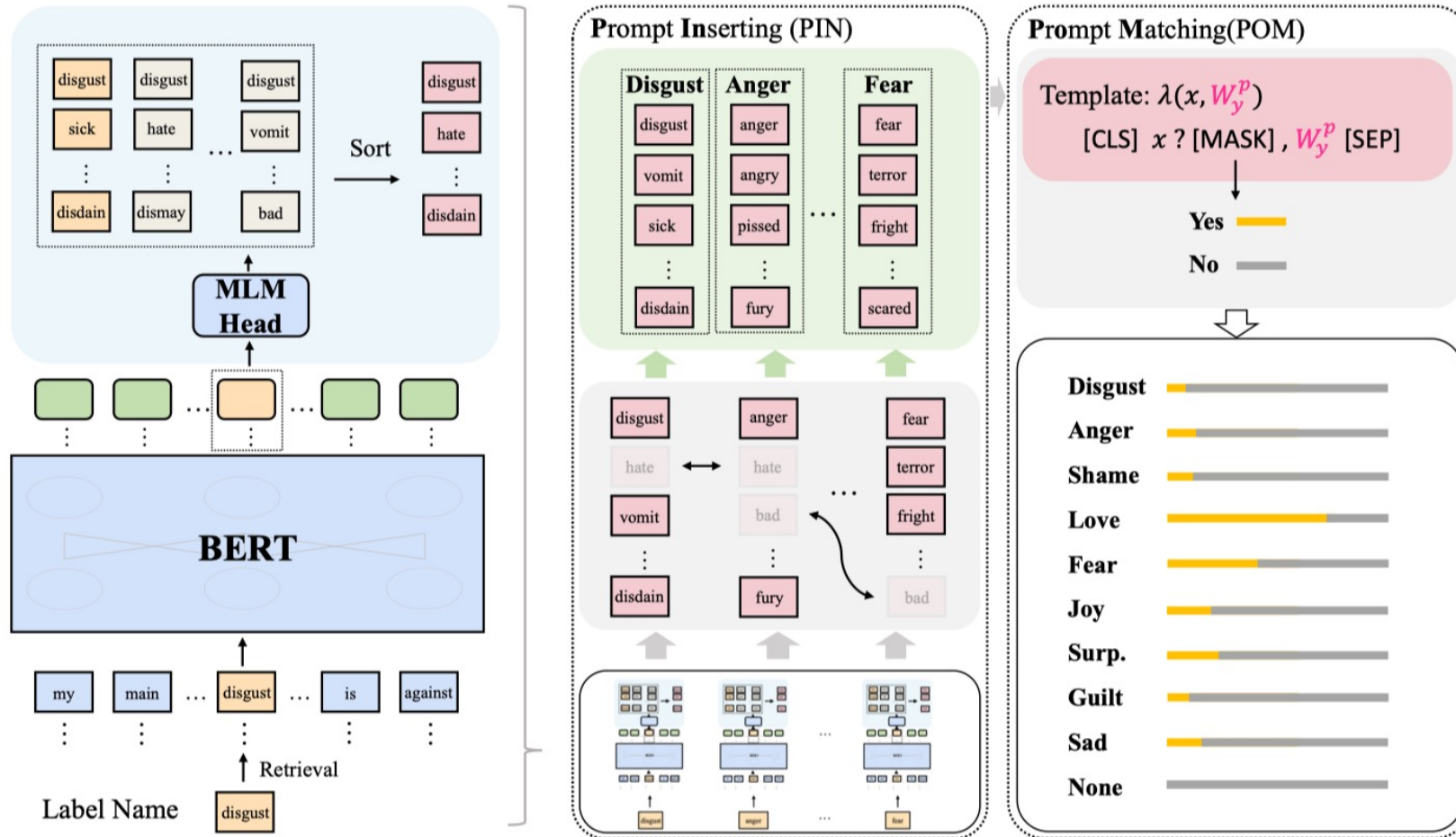
Text → Model →
- Class1: Sad
- Class2: Fear
- Class3: Anger
- ...
- Class9: Joy

Text
Class Representation1
Class Representation2
...
→ Model →
- Match
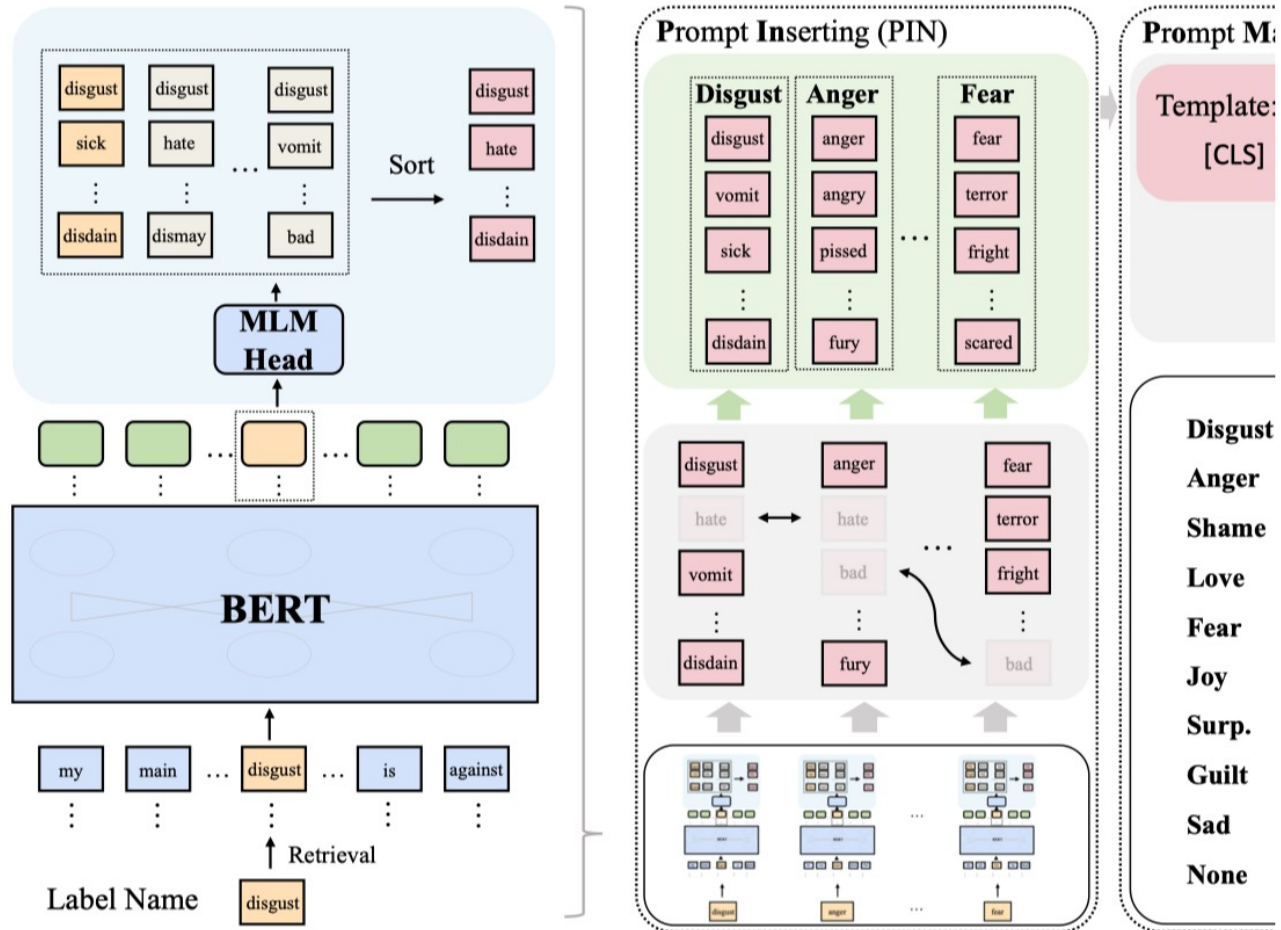- Not match

# Challenges

- How to represent a class with accurate and discriminative descriptions or knowledge?

- How to effectively learn the matching between texts and class representations on seen/labeled classes, and then extend to unseen/unlabeled classes.

# The Model



**Prompt Inserting (PIN)**

**Prompt Matching(POM)**

Template: $\lambda(x, W_y^p)$

[CLS] $x$ ? [MASK] , $W_y^p$ [SEP]

# Prompt Modeling



## Prompt Inserting (PIN)

| Disgust | Anger | | Fear |
|---------|-------|--|------|
| disgust | anger | | fear |
| vomit | angry | | terror |
| sick | pissed | ⋯ | fright |
| ⋮ | ⋮ | | ⋮ |
| disdain | fury | | scared |

Prompt Ma...

Template:

[CLS]

**Disgust**
**Anger**
**Shame**
**Love**
**Fear**
**Joy**
**Surp.**
**Guilt**
**Sad**
**None**

# Words generated by PIN

| Label | class-consistent Words |
|---|---|
| joy | fun, joy, happiness, delight, pleasure, excitement |
| sports | sporting, soccer, athletics, hockey, athletes |
| infra. | highways, lanes, routes, bridges, pathways |

# Prompt Matching
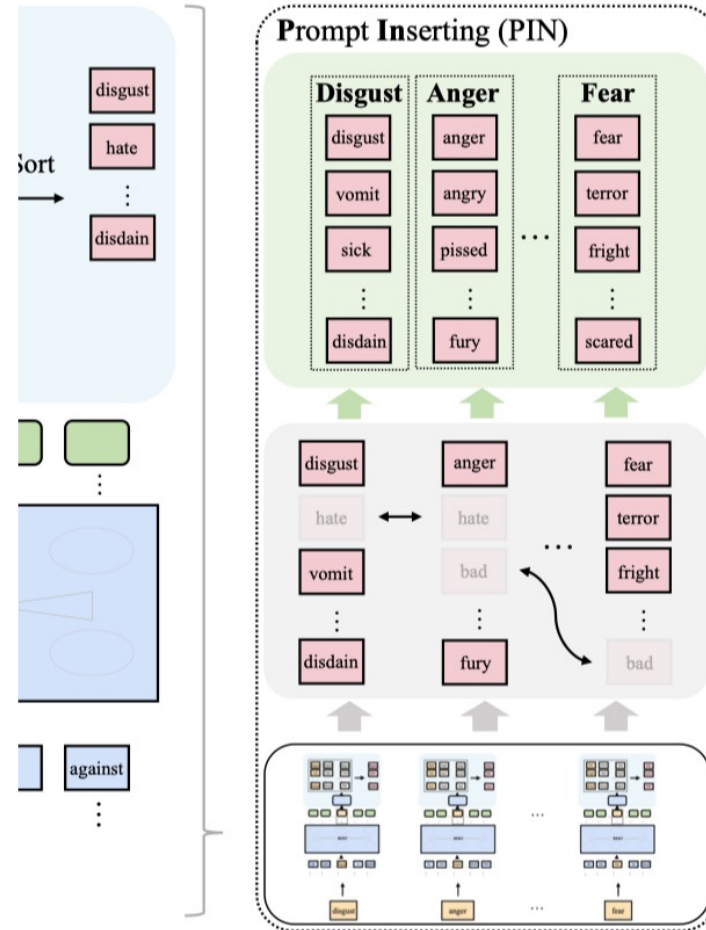
# Baselines

- Majority
- Word2Vec
- BERT(label)
- BERT(description)
- BERT-MNLI

# Main Results

| | News | | | | Emotion | | | | Event-Type | | | | Average | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | split0 | | split1 | | split0 | | split1 | | split0 | | split1 | | | | |
| | s | u | s | u | s | u | s | u | s | u | s | u | s | u | all |
| Majority* | 0.0 | 10.0 | 10.0 | 0.0 | 0.0 | 13.3 | 18.5 | 0.0 | 0.0 | 19.7 | 0.0 | 16.4 | 4.8 | 9.9 | 7.3 |
| Word2Vec* | 28.1 | 43.3 | 43.3 | 28.1 | 8.1 | 5.4 | 6.2 | 7.3 | 10.3 | 24.7 | 8.6 | 23.1 | 17.4 | 22.0 | 19.7 |
| BERT(label) | **73.8** | 46.4 | **82.6** | 29.3 | 32.2 | 17.2 | **38.1** | 17.9 | 68.6 | 49.0 | **68.7** | 49.4 | **60.7** | 34.9 | 47.8 |
| BERT(description)* | 72.6 | 44.3 | 80.6 | 34.9 | **35.6** | 17.5 | 37.1 | 14.2 | 72.4 | 48.4 | 63.8 | 42.9 | 60.4 | 33.7 | 47.0 |
| BERT+MNLI* | 70.9 | 52.1 | 77.3 | 45.3 | 33.4 | 26.6 | 33.9 | 21.4 | **74.8** | 53.4 | 68.4 | 33.7 | 59.8 | 38.8 | 49.3 |
| PINPOM | 69.6 | **54.7** | 77.0 | **52.1** | 32.4 | **29.5** | 37.6 | **23.3** | 73.7 | **56.6** | 66.6 | **52.5** | 59.5 | **44.8** | **52.1** |

# Ablation Studies

| | News | | | | Emotion | | | | Event-type | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | split0 | | split1 | | split0 | | split1 | | split0 | | split1 | |
| | seen | unseen | seen | unseen | seen | unseen | seen | unseen | seen | unseen | seen | unseen |
| PINPOM | 69.6 | **54.7** | 77.0 | **52.1** | **32.4** | **29.5** | **37.6** | **23.3** | **73.7** | **56.6** | **66.6** | **52.5** |
| -PIN | 72.3 | 49.9 | **79.9** | 38.1 | 31.6 | 19.8 | 37.0 | 18.7 | 68.9 | 52.6 | 66.1 | 49.8 |
| -POM | **73.9** | 46.8 | 79.0 | 44.1 | 31.2 | 27.1 | 37.1 | 21.4 | 68.9 | 52.8 | 65.4 | 50.5 |

# Contributions

(1) We make full use of the knowledge in the pre-trained language model to represent unseen classes, which obtains class representations with more accurate and richer semantics without relying on any human effort or external knowledge bases.

(2) We devise a new effective matching method, which well matches the texts with the corresponding classes with less labeled data during training.

(3) Experimental results on three benchmark datasets show the great superiority of our proposed method over the recent advanced methods under the zero-shot scenario.