PROMPT DA: Label-guided Data Augmentation

Eacl 2023,Advisor: JIA-LING KOH, Speaker: FAN-CHI-YU, Date:2023/05/30

for Prompt-based Few-shot Learners

Content

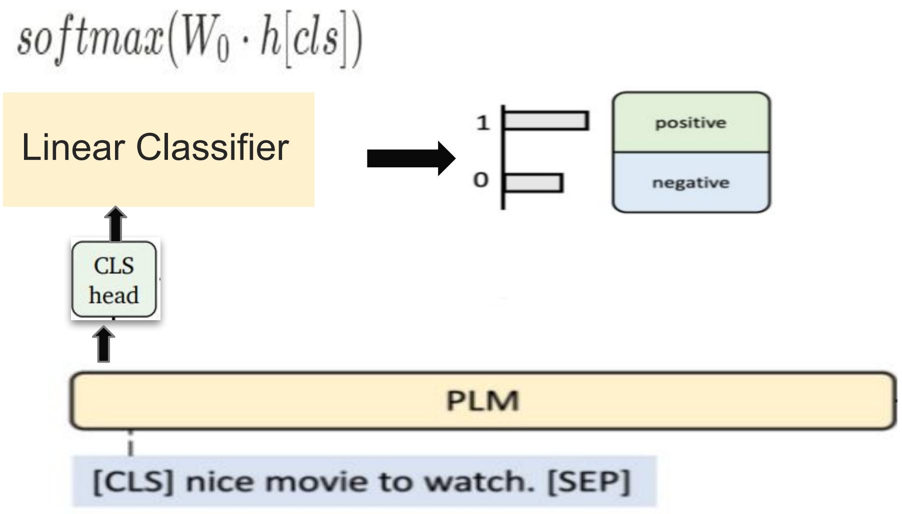
Introduction

Introduction

Introdction (Conventional Tuning)

- Compared with **Conventional Tuning, Prompt Tuning** has shown much more powerful in few-shot learning tasks

- Prompt-based tuning includes three key points: **Template design, Verbalizer design**

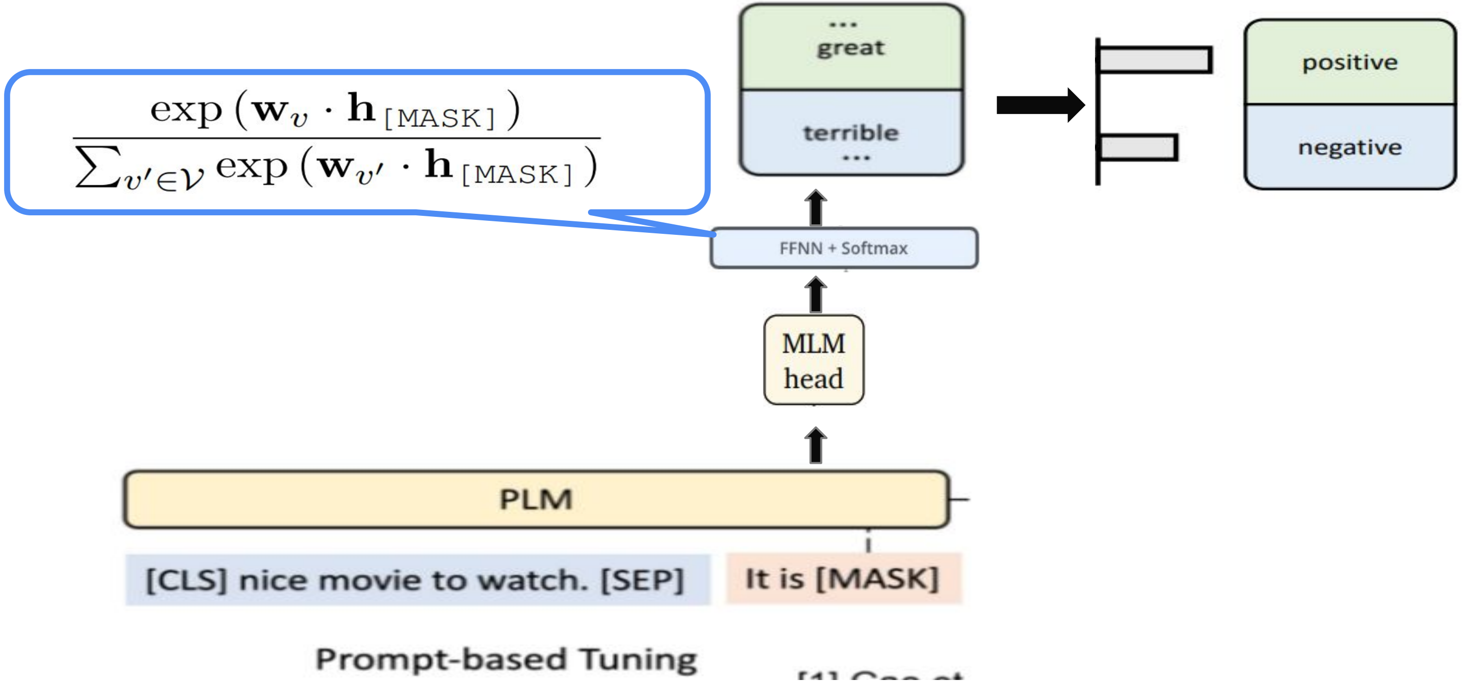


Conventional Fine Tuning

Introdction (Prompt-based tuning)

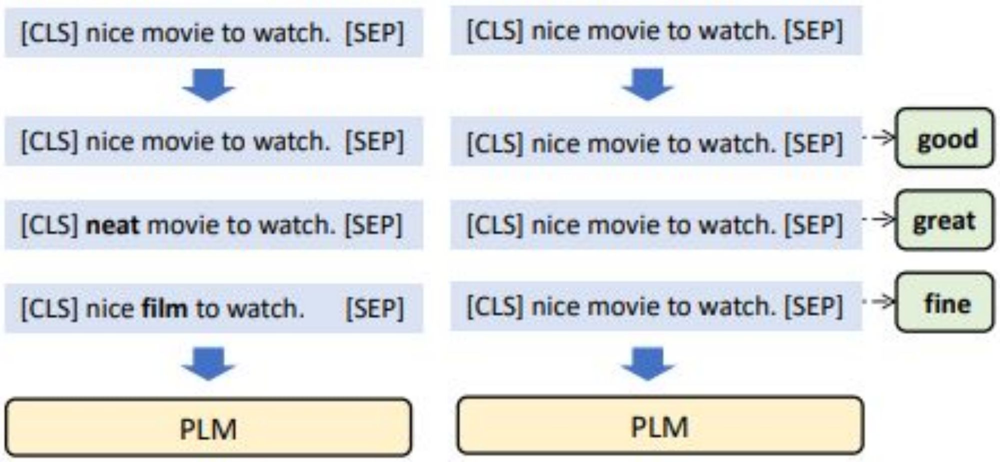
- Compared with **Conventional Tuning, Prompt Tuning** has shown much more powerful in few-shot learning tasks

- Prompt-based tuning includes three key points: **Template design, Verbalizer design**



Motivation

- **Conventional augmentation**: methods focus on constructing more **instance** .
- **PromptDA:** propose to construct **instance- label pair**(fuse **label semantics** **into augmentation)**

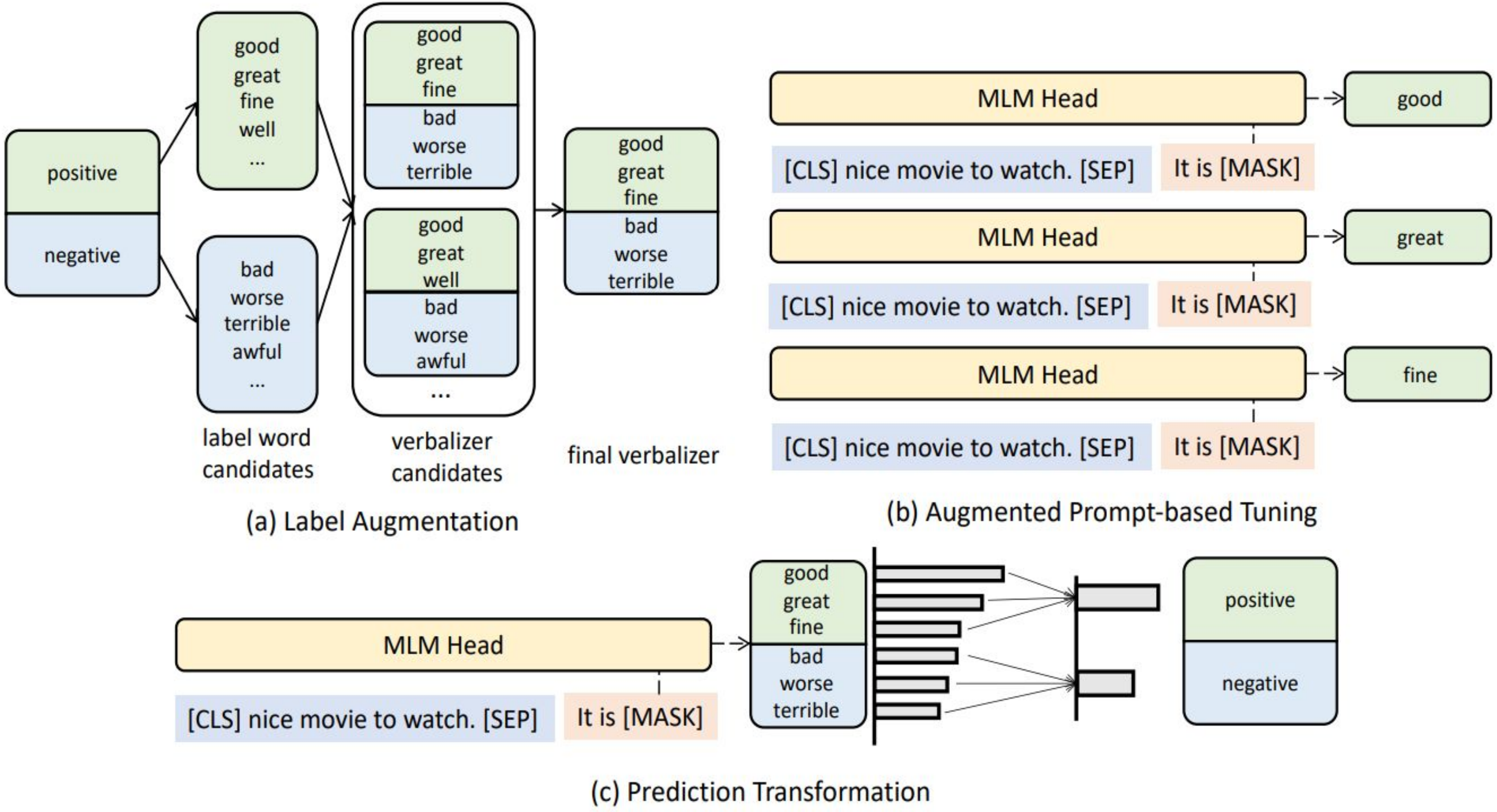


(a) Conventional augmentation method (Synonym Substitution)

(b) Our augmentation approach for Prompt Tuning

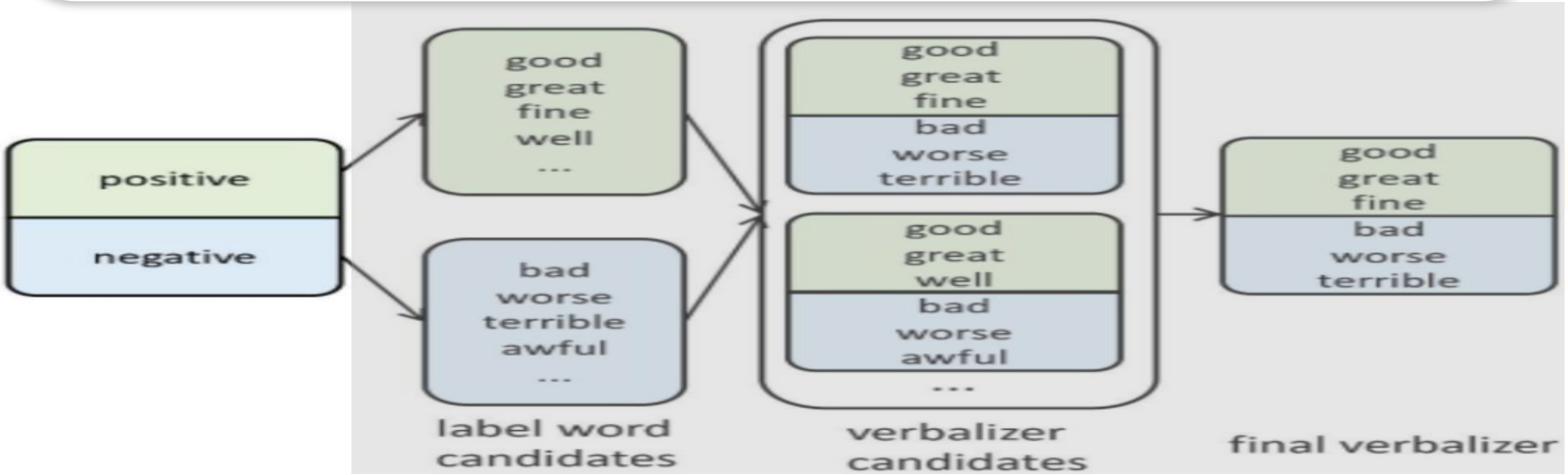
Method

Method



positive
negative
good
great
fine
well
...
bad
worse
terrible
awful
...
label word candidates
good
great
fine
bad
worse
terrible
good
great
well
bad
worse
awful
...
verbalizer candidates
good
great
fine
bad
worse
terrible
final verbalizer
(a) Label Augmentation
MLM Head
good
[CLS] nice movie to watch. [SEP]
It is [MASK]
MLM Head
great
[CLS] nice movie to watch. [SEP]
It is [MASK]
MLM Head
fine
[CLS] nice movie to watch. [SEP]
It is [MASK]
(b) Augmented Prompt-based Tuning
MLM Head
[CLS] nice movie to watch. [SEP]
It is [MASK]
good
great
fine
bad
worse
terrible
positive
negative
(c) Prediction Transformation

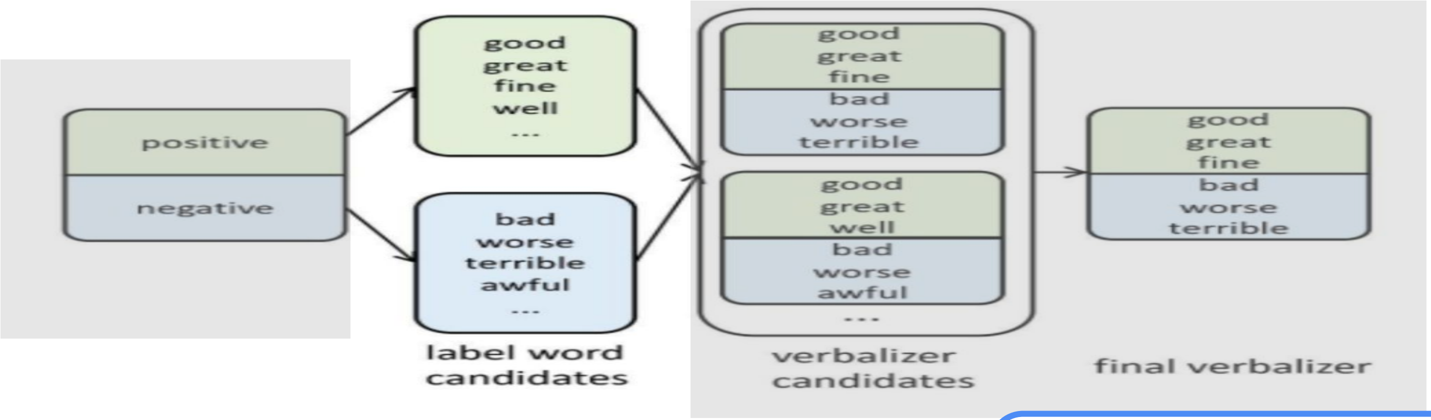
Method:Label Augmentation



A Set $V_y = \{V_y^1, V_y^2, ..., V_y^k\}$ of label word of $\mathcal{D}_{\text{train}}^{y}$

$\mathcal{Y} \longrightarrow \mathcal{V}_{\mathcal{Y}}$ denote the **one-to-multiple** verbalizer that maps each label category $y \in \mathcal{Y}$

Label Augmentation:label word candidates



Candidates label word set

$$\tilde{\mathcal{V}}_y = \underset{v \in \mathcal{V}}{\text{Top-}m} \left\{ \sum_{(x,y) \in \mathcal{D}_{\text{train}}^{y}} \Pr(v, \mathcal{T}(x)) \right\}$$

fixed template T

$$\Pr(v, \mathcal{T}(x)) = \Pr(\texttt{Po([mask])} = v \mid \mathcal{T}(x))$$

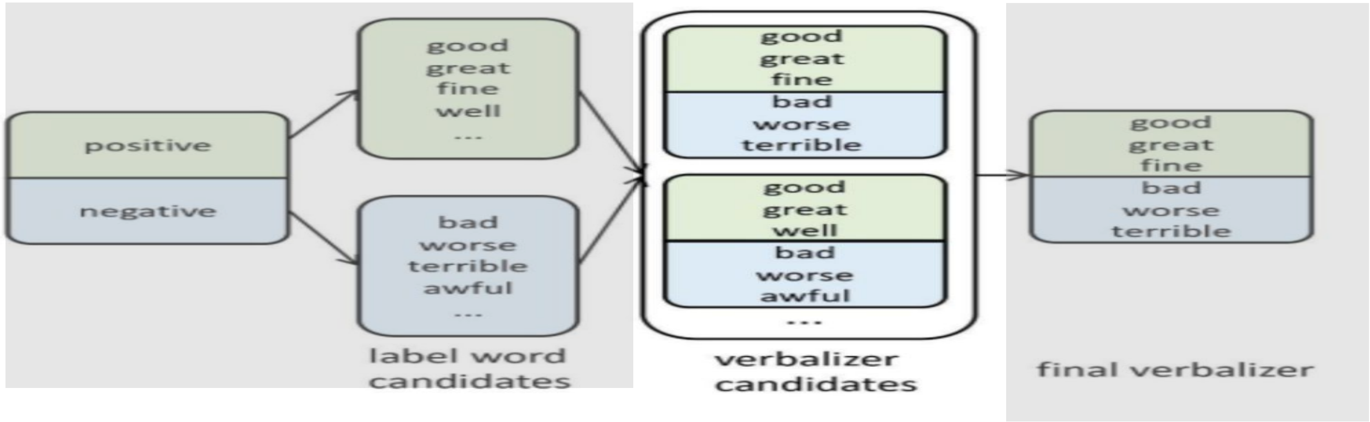
vocabulary → [CLS] nice movie to watch. [SEP] It is [MASK] → prob score

fixed template T

top m words of pos → Candidates set_{pos}

top m words of neg → Candidates set_{neg}

Label Augmentation:Verbalizer candidates



$$\binom{|\tilde{\mathcal{V}}_y|}{k_y}$$

label class num

$$|F| = \binom{|\tilde{\mathcal{V}}_y|}{k_y}^{|\mathcal{Y}|}$$

Candidates set$_{pos}$ ➡ **num of possible candidates**

Candidates set$_{neg}$ ➡ **num of possible candidates**

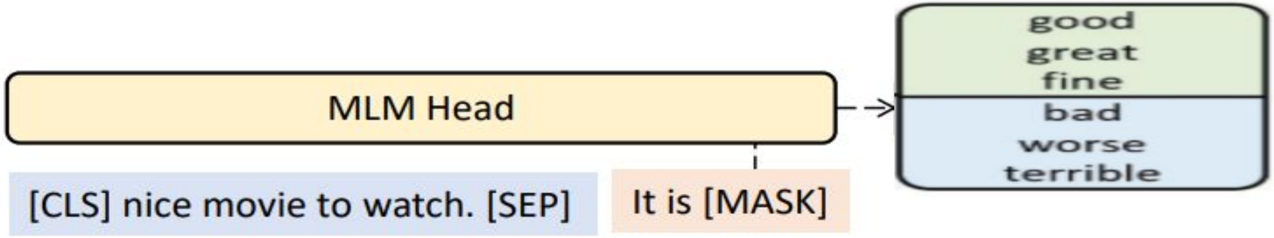
➡ **find top -n assignment using** D_{train}

Label Augmentation:Final Verbalizer

verbalizer candidate with the **highest accuracy score**

Find **top -n** assignment using $\mathbf{D}_{\mathrm{train}}$ ➡ Final **one-to-multiple** verbalizer

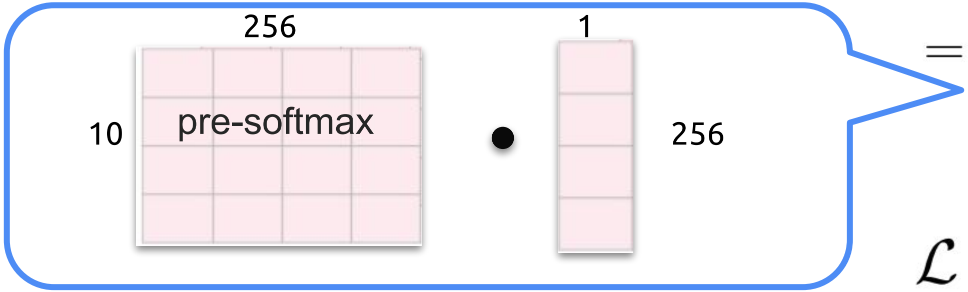
Augmented Prompt-based Tuning



$$\tilde{\mathcal{D}}_{\text{train}} = \bigcup_{(x,y)\in\mathcal{D}_{\text{train}}} \{(x, v_y^1), (x, v_y^2), ..., (x, v_y^{k_y})\}$$

Instance-label pairs → training label pairs

Augmented Prompt-based Tuning



$$\Pr(v \mid x) = \Pr(\texttt{Po}(\texttt{[mask]}) = v \mid x)$$

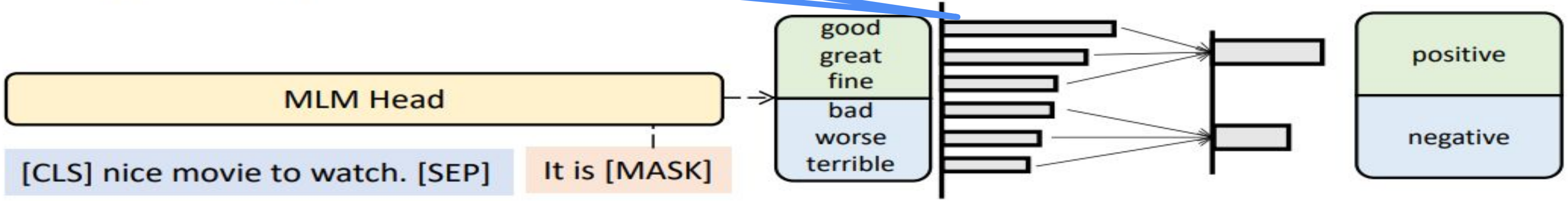
$$= \frac{\exp(\mathbf{w}_v \cdot \mathbf{h}_{\texttt{[MASK]}})}{\sum_{v' \in \mathcal{V}} \exp(\mathbf{w}_{v'} \cdot \mathbf{h}_{\texttt{[MASK]}})}$$

$$\mathcal{L} = \sum_{(x,v) \in \tilde{\mathcal{D}}_{\text{train}}} - \log \Pr(v \mid x)$$

Instance-label pairs → training label pairs

Prediction Transformation

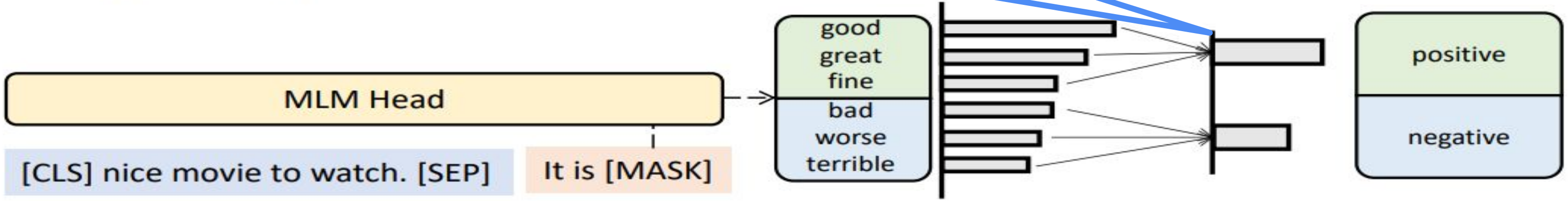
$$\mathrm{P}(v_y^i, x) = \Pr\left(\mathtt{Po}\,(\,[\mathtt{mask}]\,) = v_y^i \mid x\right)$$



Prediction Transformation

The **final probability of each class**:
h = max()

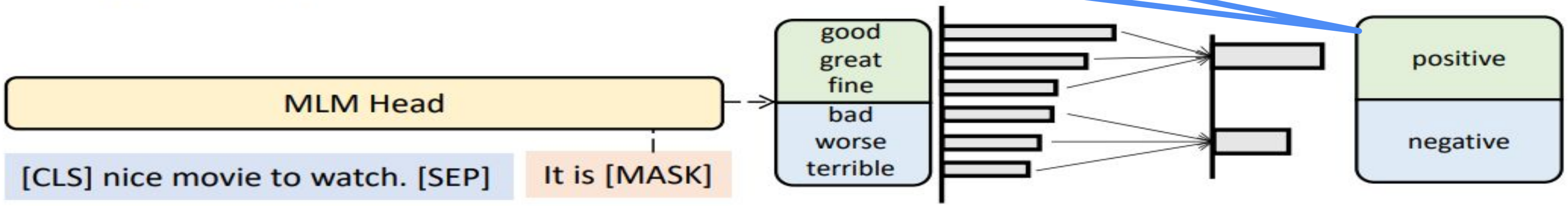
$$\Pr(y \mid x) = h(\mathrm{P}(v_y^1, x), \mathrm{P}(v_y^2, x), ..., \mathrm{P}(v_y^{k_y}, x))$$



Prediction Transformation

Final predicted class $\hat{y}$

$$\hat{y} = \text{argmax}_{y \in \mathcal{Y}} \Pr(y \mid x)$$



Exepriment

Exepriment

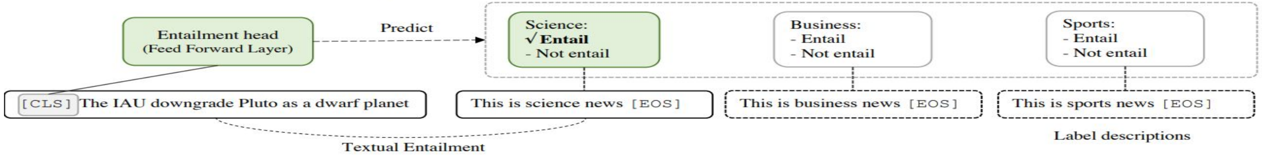
M
T
W
T
F

Exepriment:Research Question

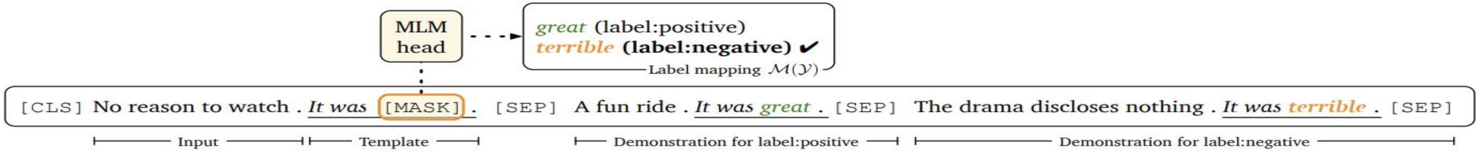
- **RQ1** Can PROMPTDA improve the performance of few-shot prompt-based tuning?
- **RQ2** Can the proposed Label Augmentation strategy help the target label prediction?
- **RQ3** Can the PROMPTDA make the promptbased tuning method more stable?

Exepriment:BaseLine Details

- Majority: The label is predicted by taking the majority class in the training set
- Fine-Tuning: Prediction is based on the pre-trained language model
- GPT-3:In-context tuning in the zero-shot setting
- EFL(Entailment as Few-Shot Learner): An entailmentbased prompt tuning framework



- LM-BFF:A prompt tuning model that automatically searches for demonstrations,



- Prompt Tuning:The standard Prompt-based Tuning augmented by a **simple template** or **template-free**

Exepriment:Research Question

- **RQ1** Can PROMPTDA improve the performance of few-shot prompt-based tuning?
- **RQ2** Can the proposed Label Augmentation strategy help the target label prediction?
- **RQ3** Can the PROMPTDA make the promptbased tuning method more stable?

Exepriment:Data Set Detail

- SST-2:**Sentiment Analysis on SST-2 Binary classification**
- MR: The Rotten Tomatoes **movie review** dataset
- CR: Mining-and-Summarising-**Customer-Review** **(give positive or negative opinions)**
- Subj: Movie review data set is **objective or subjective**
- Cola: The Corpus of Linguistic Acceptability(**include domain or out domain data**)
- MPQA: Data-set annotated for **opinions and other private states(i.e., beliefs, emotions, sentiments,** speculations, etc.)

Exepriment:Performance

- PROMPTDA can **improve** the performance of few-shot prompt-based tuning(**Q1**)

Method	SST-2 (Acc)	MR (Acc)	CR (Acc)	Subj (Acc)	CoLA (Acc)	MPQA (Acc)	SST-5 (Acc)
Majority (full)	50.9	50.0	50.0	50.0	69.1	50.0	23.1
Fine-Tuning (full)	95.0	90.8	89.4	97.0	86.2	89.4	58.7
Few-shot scenario with K=8							
Fine-Tuning	60.5 (3.1)	60.3 (7.5)	61.9 (5.1)	78.3 (8.2)	51.1 (8.4)	59.0 (3.4)	31.5 (7.5)
GPT-3 (Brown et al., 2020)	82.9 (3.4)	81.2 (2.5)	86.8 (1.5)	53.2 (1.5)	52.1 (6.2)	62.9 (3.5)	31.5 (4.3)
EFL (Wang et al., 2021)	67.5 (8.5)	69.8 (7.5)	75.3 (4.8)	78.9 (7.8)	54.3 (8.9)	68.4 (5.7)	35.2 (6.3)
LM-BFF (Gao et al., 2021)	89.1 (4.1)	83.6 (3.4)	87.8 (4.3)	81.6 (6.1)	53.5 (4.5)	73.9 (8.9)	41.2 (3.1)
PT + PROMPTDA (au.)†	**89.5 (2.9)**	83.7 (2.6)	**88.3 (4.1)**	**86.8 (3.1)**	**55.9 (7.1)**	**78.4 (9.2)**	**43.3 (1.6)**

Exepriment:Performance

- Automatic Label Augmentation outperforms the manual way.(**Q2**)

Method	SST-2 (Acc)	MR (Acc)	CR (Acc)	Subj (Acc)	CoLA (Acc)	MPQA (Acc)	SST-5 (Acc)
Majority (full)	50.9	50.0	50.0	50.0	69.1	50.0	23.1
Fine-Tuning (full)	95.0	90.8	89.4	97.0	86.2	89.4	58.7
Few-shot scenario with K=8							
Prompt Tuning†	85.8 (5.8)	79.3 (8.2)	86.1 (8.0)	81.2 (5.7)	52.7 (6.6)	75.1 (13.7)	38.4 (4.7)
PT + PROMPTDA(m.)†	88.9 (3.9)	**83.8 (1.9)**	84.9 (5.7)	82.4 (9.9)	51.3 (15.5)	78.1 (8.9)	42.7 (7.1)
PT + PROMPTDA(au.)†	**89.5 (2.9)**	83.7 (2.6)	**88.3 (4.1)**	**86.8 (3.1)**	**55.9 (7.1)**	**78.4 (9.2)**	**43.3 (1.6)**

- † template augmented:manually choose “Itis [MASK]"
- ‡ template-free: only append “[MASK]"

- (au.): automatic label augmentation,this paper proposed
- (m.): manual label augmentation,find the synonyms of label name from dictionary

Exepriment:Performance

- PROMPTDA makes the prompt-based tuning more stable (less variance).(**Q3**)

Method	SST-2 (Acc)	MR (Acc)	CR (Acc)	Subj (Acc)	CoLA (Acc)	MPQA (Acc)	SST-5 (Acc)
Majority (full)	50.9	50.0	50.0	50.0	69.1	50.0	23.1
Fine-Tuning (full)	95.0	90.8	89.4	97.0	86.2	89.4	58.7
Few-shot scenario with K=8							
Prompt Tuning‡	85.5 (5.2)	83.0 (3.7)	86.5 (3.0)	81.8 (5.6)	50.5 (10.3)	71.5 (9.8)	37.5 (5.5)
PT + PROMPTDA(m.)‡	87.3 (4.4)	82.5 (1.4)	88.1 (2.7)	81.3 (4.9)	51.2 (7.5)	72.9 (9.1)	39.4 (4.3)
PT + PROMPTDA(au.)‡	87.6 (4.1)	83.1 (3.1)	87.8 (1.2)	83.4 (2.5)	52.8 (8.1)	74.5 (7.8)	41.8 (3.9)
Prompt Tuning†	85.8 (5.8)	79.3 (8.2)	86.1 (8.0)	81.2 (5.7)	52.7 (6.6)	75.1 (13.7)	38.4 (4.7)
PT + PROMPTDA(m.)†	88.9 (3.9)	**83.8 (1.9)**	84.9 (5.7)	82.4 (9.9)	51.3 (15.5)	78.1 (8.9)	42.7 (7.1)
PT + PROMPTDA(au.)†	**89.5 (2.9)**	83.7 (2.6)	**88.3 (4.1)**	**86.8 (3.1)**	**55.9 (7.1)**	**78.4 (9.2)**	**43.3 (1.6)**

† template augmented:manually choose “Itis [MASK]"
‡ template-free: only append “[MASK]"

Expriment:Combine Conventional DA

Method	SST-2 (Acc)	MR (Acc)	CR (Acc)	Subj (Acc)	SST-5 (Acc)
PT	85.8 (5.8)	79.3 (8.2)	86.1 (8.0)	81.2 (5.7)	38.4 (4.7)
PT with Conventional DA	89.2 (1.3)	80.3 (3.1)	86.5 (4.5)	82.3 (8.0)	39.1 (4.5)
PT with PROMPTDA	89.5 (2.9)	83.7 (2.6)	88.3 (4.1)	86.8 (3.1)	43.3 (1.6)
PT with PROMPTDA & Conventional DA	**89.7 (1.6)**	**84.8 (1.5)**	**89.2 (1.3)**	**87.0 (3.1)**	**44.7 (1.1)**

- Combination with conventional DA can bring additional improvements, which suggests PROMPTDA can be regarded as orthogonal to conventional DA

- Conventional DA:Select synonym substitution method
  - Enlarge the training set by ×2
- PROMPT DA :This paper propose
  - Enlarge the training set by ×3

Exepriment:Label Word Selection

- **SST-2:** The label words automatically manually searched are **literally similar**.
- **Subj** :the label name {**objective/subjective**}, **not literally similar.**
  computer don't know what is the **object & subject** word
- **SST-5**:harder to select appropriate **label words(Performance bad)**

SST-2	label name	positive | negative
	label words (m.)	positive, great, good | negative terrible bad
	label words (au.)	wonderful brilliant fantastic | terrible done disappointing

Key word is **literally** same

Exepriment:Label Word Selection

- SST-2: The label words automatically manually searched are **literally similar**.
- **Subj** :the label name {**objective/subjective**}, **not literally similar.**
  computer don't know what is the **object & subject** word
- **SST-5**:harder to select appropriate **label words(Performance bad)**

Subj	label name	objective | subjective
	label words (m.)	good neutral fair | bad emotional personal
	label words (au.)	disturbing terrifying key | bad not nonsense

Hard to look **not literally similar** but **synonyms similar**

Exepriment:Label Word Selection

- SST-2: The label words automatically manually searched are **literally similar**.
- **Subj** :the label name {**objective/subjective**}, **not literally similar.**
  computer don't know what is the **object & subject** word
- **SST-5**: harder to select appropriate **label words(Performance bad)**

Hard to split out neutral ,**negative and very negative**

SST-5	label name	very positive | positive | neutral | negative | very negative
	label words (m.)	great perfect excellent | good, pretty, wonderful | neutral normal fine | bad worse not | terrible awful ridiculous
	label words (au.)	great brilliant fantastic | extraordinary remarkable fascinating | enough terrible funny | awful bad worse | boring done unnecessary

Conclusion

Conclusion

- A new problem of designing data augmentation strategies for prompt-based few-shot learners.
- Label-guided data augmentation framework PROMPTDA that exploits the rich label semantic information of one-to-multiple verbalizer for improving prompt tuning, which can be a plug-in module for any prompt-based method.
- Extensive experiments on real-world few-shot classification tasks demonstrate the effectiveness of the proposed framework