# Hyperbolic Representations for Prompt Learning

Nan Chen[1,2,3], Xiangdong Su[1,2,3], Feilong Bao[1,2,3*]

[1] College of Computer Science, Inner Mongolia University, Hohhot, China

[2] National & Local Joint Engineering Research Center of Intelligent Information Processing Technology for Mongolian, Hohhot, China

[3] Inner Mongolia Key Laboratory of Mongolian Information Processing Technology

## Motivation

Continuous prompt tuning has gained significant attention for its ability to train only continuous prompts while freezing the language model. This approach greatly reduces the training time and storage for downstream tasks. In this work, we delve into the hierarchical relationship between the prompts and downstream text inputs. In prompt learning, the prefix prompt acts as a module to guide the downstream language model, establishing a hierarchical relationship between the prefix prompt and subsequent inputs. Furthermore, we explore the benefits of leveraging hyperbolic space for modeling hierarchical structures. We project representations of pre-trained models from Euclidean space into hyperbolic space using the Poincar\'{e} disk which effectively captures the hierarchical relationship between the prompt and input text. The experiments on natural language understanding (NLU) tasks illustrate that hyperbolic space can model the hierarchical relationship between prompt and text input. Overall, our main contributions can be summarized as follows: (1) We investigate the hierarchical structure between prompts and downstream task inputs, and propose the utilization of the Poincar\'{e} disk hyperbolic space to model and substantiate this relationship. (2) Experiments on sentence classification, question answering, and token classification tasks demonstrate the effectiveness of our proposed approach.

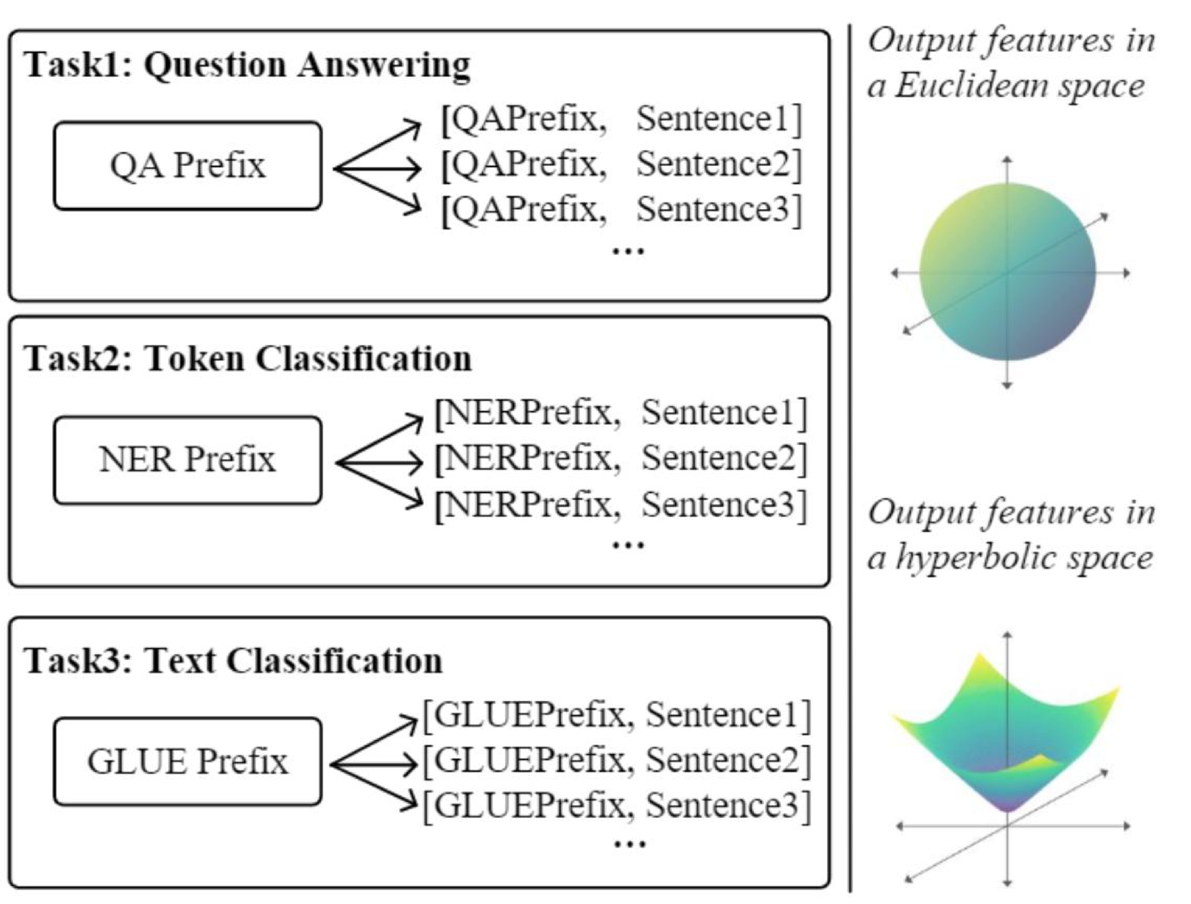


Figure 1: Hierarchical structure of prompt learning (Left). Euclidean space and hyperbolic space (Right).

## Methods

We first describe the Poincaré ball model projection. To map from the Euclidean tangent space to the hyperbolic space, networks operate on the Poincaré ball. The projection of a Euclidean vector **x** onto the Poincaré ball is given by the exponential map with anchor **v**:

$$\exp_v^c(\boldsymbol{x}) = v \oplus_c \left(\tanh\left(\sqrt{c}\frac{\lambda_v^c \parallel \boldsymbol{x} \parallel}{2}\right)\frac{\boldsymbol{x}}{\sqrt{c} \parallel \boldsymbol{x} \parallel}\right) \quad (1)$$

with ⊕ c the Möbius addition:

$$v \oplus_c w = \frac{(1 + 2c\langle v, w\rangle + c \parallel w \parallel^2)v + (1 - c \parallel v \parallel^2)w}{1 + 2c\langle v, w\rangle + c^2 \parallel v \parallel^2 \parallel w \parallel^2} \quad (2)$$

In practice, **v** is commonly set to the origin, simplifying the exponential map to:

$$\exp_0(\boldsymbol{x}) = \tanh((\sqrt{c} \parallel \boldsymbol{x} \parallel)(\boldsymbol{x}/(\sqrt{c} \parallel \boldsymbol{x})) \quad (3)$$

As shown in Fig. 2, we take $\mathrm{BERT}_{\mathrm{large}}$ model as an example. Given the trainable continuous embeddings $[p_1, p_2, \ldots, p_n\}]$ as prefix representations, the prompt representation and input text are fed into the $\mathrm{BERT}_{\mathrm{large}}$ model. Finally, according to per-task-specific settings, the outputs of hyperbolic space are fed into a linear classifier to get final logits:

$$\begin{aligned} \text{outputs}_{\text{ feature}} &= \mathrm{BERT}_{\text{large}}([\boldsymbol{P_e}; \boldsymbol{X_e}]) \\ \text{outputs}_{\text{ hy}} &= \exp_0(\text{outputs}_{\text{feature}}) \\ \text{logits} &= \mathrm{BERT}_{\text{linear}}(\text{outputs}_{\text{hy}}) \end{aligned} \quad (4)$$

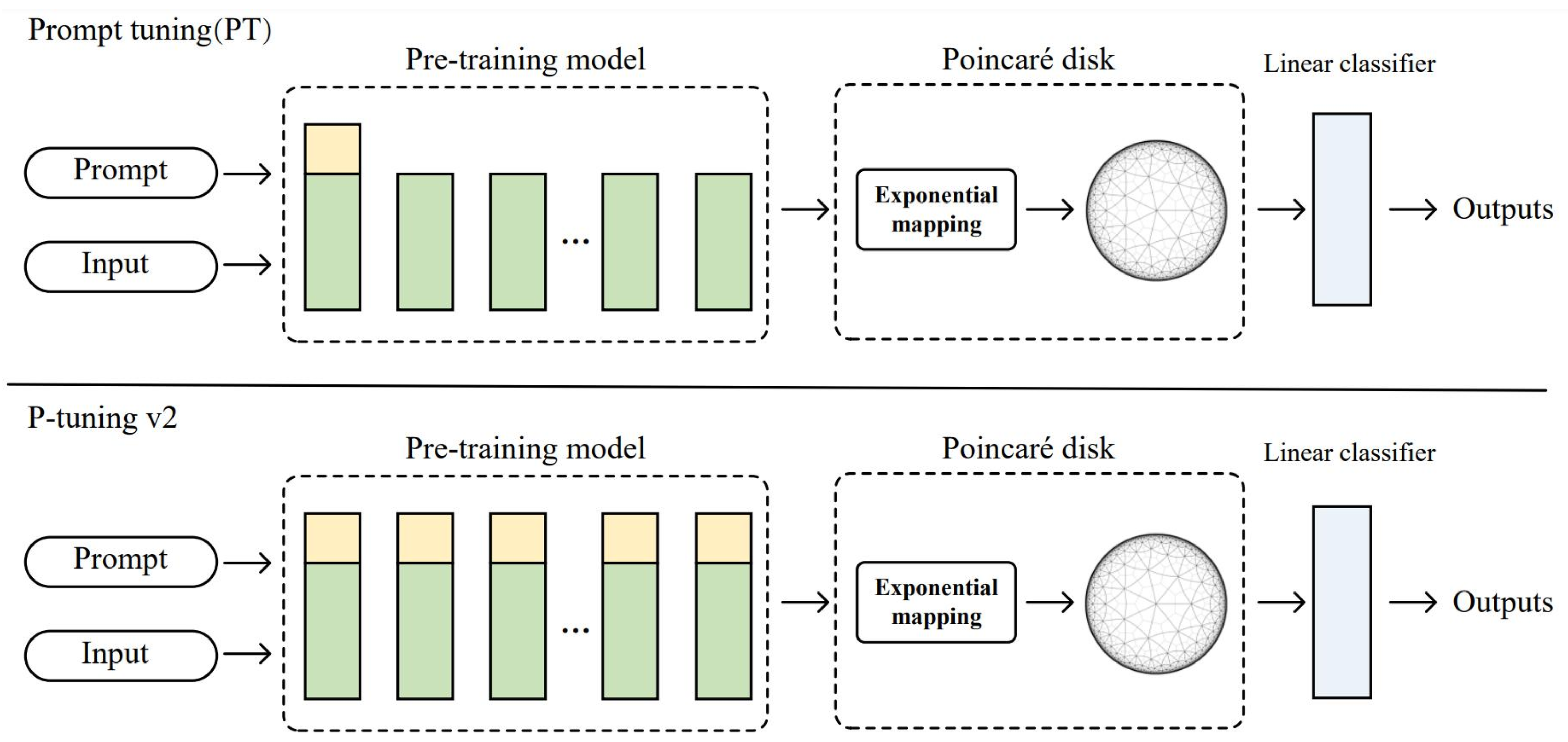


Figure 2: Overview of our proposed approach. Yellow blocks refer to trainable prompt embeddings. Green blocks are frozen pre-trained language models.

## Results

| | BoolQ | | CB | | COPA | | MultiRC (F1a) | |
|---|---|---|---|---|---|---|---|---|
| | PT | PT-2 | PT | PT-2 | PT | PT-2 | PT | PT-2 |
| $\mathrm{BERT}_{large}$ | 63.2 | 75.2 | 73.2 | 94.6 | 62.0 | 75.0 | **59.6** | **70.6** |
| $\mathrm{BERT}_{large}$ + Poincaré | **66.7** | **76.0** | **75.0** | **96.0** | **71.0** | **79.0** | 59.0 | 70.2 |
| $\mathrm{RoBERTa}_{large}$ | 62.2 | **84.5** | 69.6 | 94.6 | 63.0 | 88.2 | **59.9** | **82.5** |
| $\mathrm{RoBERTa}_{large}$ + Poincaré | **62.6** | **84.5** | **73.2** | **96.4** | **65.0** | **91.0** | 59.3 | 82.0 |
| | **ReCoRD (F1)** | | **RTE** | | **WiC** | | **WSC** | |
| | PT | PT-2 | PT | PT-2 | PT | PT-2 | PT | PT-2 |
| $\mathrm{BERT}_{large}$ | **44.2** | **72.8** | 53.5 | **78.3** | 56.9 | 71.0 | **63.5** | 66.3 |
| $\mathrm{BERT}_{large}$ + Poincaré | 44.0 | 72.6 | **65.0** | 77.6 | **65.1** | **73.2** | **63.5** | **67.3** |
| $\mathrm{RoBERTa}_{large}$ | **46.3** | **89.3** | 54.5 | 87.0 | 57.8 | 69.0 | **63.5** | **63.4** |
| $\mathrm{RoBERTa}_{large}$ + Poincaré | **46.3** | 89.0 | **57.4** | **88.4** | **71.3** | **71.0** | **63.5** | **63.4** |

Table 1: Results on SuperGLUE development set. (PT: Prompt tuning Lester et al. (2021); PT-2: P-tuningv2 Liu et al. (2021a); **bold**: the best

| | CoNLL03 | | CoNLL04 | | SQuAD 1.1 | | SQuAD 2.0 | |
|---|---|---|---|---|---|---|---|---|
| | PT | PT-2 | PT | PT-2 | PT | PT-2 | PT | PT-2 |
| BERT | 82.5 | 82.2 | 71.2 | 82.2 | 63.0/75.3 | **82.1/89.4** | 50.8/52.6 | 67.6/71.4 |
| + Poincaré | **84.2** | **83.4** | **72.8** | **84.1** | **62.8/75.0** | 82.0/89.3 | **51.0/53.0** | **68.2/73.0** |
| RoBERTa | 87.1 | 86.9 | 76.2 | 86.2 | 72.5/78.4 | 88.0/94.0 | 67.5/71.4 | 81.1/84.5 |
| + Poincaré | **88.8** | **92.1** | **78.2** | **89.0** | **73.0/78.6** | **88.0/94.2** | **68.6/72.3** | **81.5/85.0** |

Table 2: Results on named entity recognition (NER) and question answering (QA). PT: Prompt tuning Lester et al. (2021); PT-2: P-tuningv2 Liu et al. (2021a); **bold**: the best

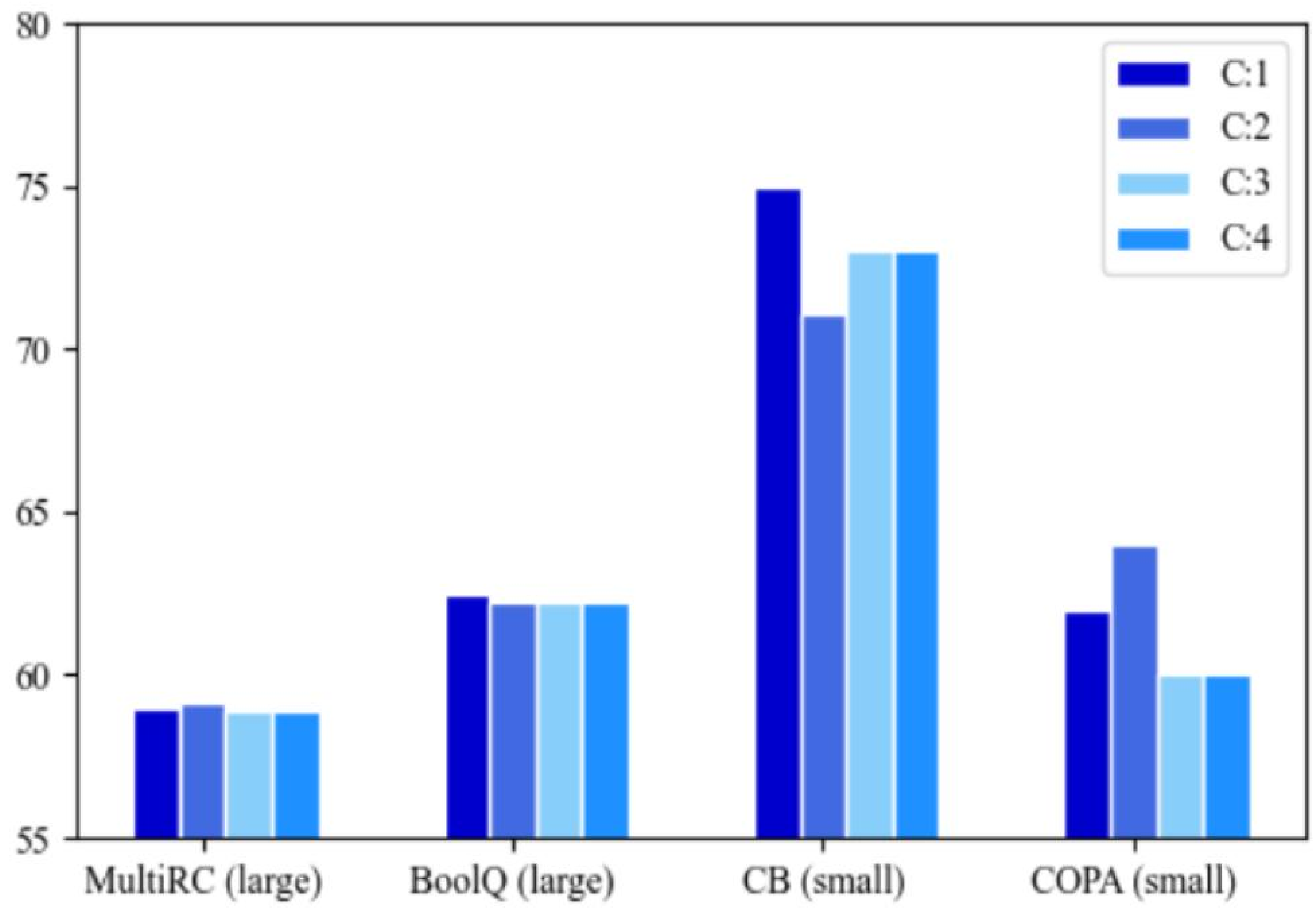


Figure 3: Comparison of curvature for high and low resources datasets. C:1 represents the curvature is set to 1.

## Conclusion

In this paper, we explore the Poincaré disk hyperbolic representations of pre-trained models in NLU tasks, projecting representations from Euclidean space into hyperbolic space to model the hierarchical relationship between the prompt and input text. With high accuracy and efficiency, hyperbolic representations can be an effective supplement to prompt learning.