

Partial Order-centered Hyperbolic Representation Learning for Few-shot Relation Extraction

Biao Hu^{1†}, Zhen Huang^{1†}, Minghao Hu^{2‡},
Pinglv Yang³, Peng Qiao¹, Yong Dou¹, Zhilin Wang¹

¹ National Key Laboratory of Parallel and Distributed Computing,
National University of Defense Technology

² Center of Information Research, Academy of Military Science

³ College of Meteorology and Oceanology, National University of Defense Technology
{hubiao, huangzhen}@nudt.edu.cn huminghao16@gmail.com

Abstract

Prototype network-based methods have made substantial progress in few-shot relation extraction (FSRE) by enhancing relation prototypes with relation descriptions. However, the distribution of relations and instances in distinct representation spaces isolates the constraints of relations on instances, making relation prototypes biased. In this paper, we propose an end-to-end partial order-centered hyperbolic representation learning (PO-HRL) framework, which imposes the constraints of relations on instances by modeling partial order in hyperbolic space, so as to effectively learn the distribution of instance representations. Specifically, we develop the hyperbolic supervised contrastive learning based on Lorentzian cosine similarity to align representations of relations and instances, and model the partial order by constraining instances to reside within the Lorentzian entailment cone of their respective relation. Experiments on three benchmark datasets show that PO-HRL outperforms the strong baselines, especially in 1-shot settings lacking relation descriptions.

1 Introduction

Relation extraction (RE) is a vital aspect of information extraction, focused on predicting the relation between two entities present in unstructured sentences. As the foundation for knowledge bases, RE is widely applied to various downstream natural language processing (NLP) tasks, including question answering (Han et al., 2020), knowledge graph completion (Shen et al., 2021; Wang et al., 2023) and rumor detection (Lu et al., 2022; Huang et al., 2022), etc. Conventional supervised RE methods (Xu et al., 2015; Miwa and Bansal, 2016) rely on substantial labeled instances and are limited

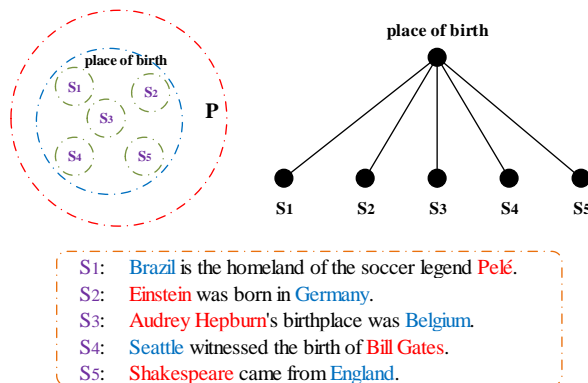


Figure 1: Illustration of the partial order. Left: The sets “ S_1, \dots, S_5 ” are subsets of instance set defined by the relation “*place of birth*”, each contain one specific instance. Let $P = \{S_1, \dots, S_5, \text{place of birth}\}$, thus (P, \subseteq) forms a poset (partially ordered set), with “ \subseteq ” being the partial order relation. Right: The Hasse diagram of the poset (P, \subseteq) . Below: Instances with subjects and objects are colored red and blue respectively.

to extracting pre-defined relations. However, data labeling is time-consuming and laborious, while new relations continuously emerge. Consequently, few-shot relation extraction (FSRE) (Han et al., 2021a; Dou et al., 2022; Li et al., 2023; Han et al., 2021b; Li et al., 2022) has attracted increasing attention, enabling models to generalize to new relations with a limited number of labeled instances.

In light of the efficacy of few-shot learning paradigm in the NLP community, Han et al. (2018) pioneered the incorporation of few-shot learning into RE task. Recently, meta-learning (Finn et al., 2017) based prototype networks (Snell et al., 2017) have emerged as the predominant approach for FSRE, aiming to learn an embedding space where query instances are classified by their proximity to relation prototypes. Specifically, works (Han et al., 2021a; Dou et al., 2022; Li et al., 2023) learn relation prototypes from episodes of an N -way- K -shot setup containing a few relation instances, with relation descriptions as auxiliary information.

[†]Co-first author

[‡]Corresponding author

While methods mentioned above achieve promising performance, they share a critical limitation. Relations and instances are encoded disparately and distributed in separate representation spaces. This disconnect severs the constraints that relations should exert on instances, resulting in a biased distribution of the learned instance representations. Particularly, this issue is acute in 1-shot settings, where the scarcity of samples further exacerbates the challenge of capturing the nuances of relations. In essence, relations, as generalized concepts, have inherent partial orders between the specific instances they encompass. For example, as illustrated in Fig. 1, the relation “*place of birth*” defines an instance set, containing sets “ S_1, \dots, S_5 ” that each encompass only one specific instance. In the set $P = \{S_1, \dots, S_5, \textit{place of birth}\}$, it is evident that the homogeneous relation “ \subseteq ” satisfies reflexivity, antisymmetry, and transitivity¹, clearly describing the partial order relation. Without partial order constraints, relation prototypes obtained by directly adding relation representations to instance representation centers are skewed. As a result, the learned representations of instances contained in sets S_1 and S_5 may incorrectly be closer to other relation prototypes, such as “*country of citizenship*” rather than “*place of birth*”.

To formalize the partial order, there are two main challenges. First, due to the polynomial growth of space capacity, the commonly used embedding space, Euclidean space, faces difficulties in capturing partial order under the premise of limited dimensions. Previous works (Nickel and Kiela, 2017, 2018) suggest that this can be accomplished in hyperbolic space, which has exponential volume growth², making it well-suited for modeling partial order. Second, since relations and instances are encoded separately, the alignment of relation and instance representations is necessary. However, the hyperbolic contrastive learning method based on negative geodesic distance or squared geodesic distance inadequately captures directional information, impeding partial order modeling.

To learn reliable representations of relations and instances in hyperbolic space, we propose

¹Forms of reflexivity, antisymmetry, and transitivity. Reflexivity: $\forall p_i \in P, p_i \subseteq p_i$. Antisymmetry: $\forall p_i, p_j \in P$, if $p_i \subseteq p_j$ and $p_j \subseteq p_i$, then $p_i = p_j$. Transitivity: $\forall p_i, p_j, p_k \in P$, if $p_i \subseteq p_j$ and $p_j \subseteq p_k$, then $p_i \subseteq p_k$.

²In a two-dimensional hyperbolic space with the curvature of k ($k < 0$), the circumference and area of a circle with radius r are $2\pi \sinh(\sqrt{|c|r})$ and $2\pi(\cosh(\sqrt{|c|r}) - 1)$ respectively, which grow exponentially with r .

an end-to-end **Partial Order-centered Hyperbolic Representation Learning (PO-HRL)** framework for FSRE. Specifically, relations and instances are initially encoded in Euclidean space by a pre-trained language model, and then projected into hyperbolic space. Subsequently, to align relation and instance representations, we develop a hyperbolic supervised contrastive learning method based on the derived cosine similarity in the Lorentz model³, which captures directional information effectively. With alignment, representations of relations and instances can be learned in the same embedding space. Meanwhile, we view the relations as generalized concepts and constrain the encompassed instances to reside within their Lorentzian entailment cones to model the inherent partial order, thereby facilitating the distinction of easily confused instances. Finally, query instances are classified by their similarity to the Lorentzian aggregation center of instances of each relation. The contributions of our work are summarized as follows:

- We theoretically derive the Lorentzian cosine similarity for representation alignment, and extend the Lorentzian entailment cones to the Lorentz model of arbitrary curvature to model the inherent partial order.
- Based on representation alignment and partial order modeling, we propose the PO-HRL framework for FSRE.
- Experiments on three public datasets show that PO-HRL achieves state-of-the-art performance in FSRE, especially outperforming in 1-shot settings lacking relation descriptions.

2 Preliminaries

2.1 Hyperbolic Geometry

Unlike Euclidean geometry, hyperbolic geometry is a Riemannian manifold with a constant negative curvature k ($k < 0$). Typical hyperbolic models include Poincaré ball model (Nickel and Kiela, 2017), Lorentz model (Nickel and Kiela, 2018), and Klein model (Gulcehre et al., 2019), etc. All these models are isometric, that is, they can be transformed into each other by mapping functions. Compared with the Poincaré ball model, the Lorentz model has better numerical stability and computational efficiency (Nickel and Kiela, 2018) due to the lack

³A typical hyperbolic geometric model.

of fractions in the distance function and calculation simplicity of exponential/logarithmic maps. Therefore, we select the Lorentz model as the foundational model. We give a brief introduction to some essential concepts in the Lorentz model of hyperbolic geometry in appendix A.

2.2 Problem Definition

We follow the typical N -way- K -shot few-shot task setting, which contains a support set S of and a query set Q . The support set $S = \{s_{i,k} | i = 1, \dots, N; k = 1, \dots, K\}$ consists of N classes, each with K labeled instances. The query set Q includes unlabeled instances of classes in S . For FSRE tasks, each instance is represented as (x, h, t, y) , where x denotes the given sentence, h and t indicate the head and tail entity respectively, y is the relation label. Additionally, the name and description of each relation provide auxiliary support. We randomly sample N relations and K instances per relation as the support set. Concurrently, the query set is constructed by sampling one instance per relation from the remaining samples. Note that relations are disjoint in the training and testing phases.

3 Methodology

In this section, we focus on the Lorentz cosine similarity for representation alignment and the Lorentz entailment cones for partial order modeling.

3.1 Lorentzian Cosine Similarity

In previous works (Han et al., 2021a; Dou et al., 2022), relations and instances are encoded separately, the alignment of their representations is crucial. A common approach is supervised contrastive learning (Khosla et al., 2020), which leverages cosine to measure the similarity of vectors. In the Lorentz model, existing measures like geodesic distance and its squared variant can be used for similarity assessment. However, they fail to fully capture the directional information, hindering subsequent partial order modeling based on angles. Therefore, we derive the Lorentzian cosine similarity.

Theorem 3.1. $\forall \mathbf{u}, \mathbf{v} \in \mathcal{L}_k^n \setminus \{\mathbf{o}\}$, let u_i and v_i denote the i -th dimension of \mathbf{u} and \mathbf{v} , the Lorentzian cosine similarity between them $\text{sim}(\mathbf{u}, \mathbf{v}) \in [0, 1]$ in the Lorentz model is calculated as:

$$\text{sim}(\mathbf{u}, \mathbf{v}) = \frac{-k \sum_{i=1}^n u_i v_i}{\sqrt{|k|u_0^2 - 1} \sqrt{|k|v_0^2 - 1}}$$

Proof. See appendix B. \square

3.2 Lorentzian Entailment Cones

Intuitively, relations capture generalized connections to encompassed instances. As such, an inherent partial order exists in any relation-instance pair, with the relation encompassing the broader conceptual tie and the instance exemplifying a specific manifestation. We introduce the Lorentzian entailment cones (Ganea et al., 2018a; Le et al., 2019) to formalize the partial order. Since Le et al. (2019) only provide the form of the entailment cones in the Lorentz model of curvature -1, inspired by Ganea et al. (2018a), we extend the half-aperture of entailment cones and exterior angles to the Lorentz model of any arbitrary negative curvature.

Lemma 3.1. $\forall \mathbf{u} \in \mathcal{L}_k^n \setminus \{\mathbf{o}\}$, let u_0 denotes the 0-th dimension of \mathbf{u} , $C > 0$ is a constant used to set the boundary conditions, the entailment cone of \mathbf{u} is defined by the half-aperture $\gamma(\mathbf{u}) \in [0, \frac{\pi}{2}]$:

$$\gamma(\mathbf{u}) = \sin^{-1}\left(\frac{C}{\sqrt{|k|u_0^2 - 1}}\right)$$

Proof. See appendix C. \square

Lemma 3.2. $\forall \mathbf{u}, \mathbf{v} \in \mathcal{L}_k^n \setminus \{\mathbf{o}\}$, let u_i and v_i denote the i -th dimension of \mathbf{u} and \mathbf{v} , the exterior angle $\phi(\mathbf{u}, \mathbf{v}) \in [0, \frac{\pi}{2}]$ between half-lines $(\mathbf{o}\mathbf{u}$ and $(\mathbf{u}\mathbf{v}$ can be calculated as:

$$\phi(\mathbf{u}, \mathbf{v}) = \cos^{-1}\left(\frac{v_0 - u_0 \cdot k \langle \mathbf{u}, \mathbf{v} \rangle_{\mathcal{L}}}{\sqrt{\sum_{i=1}^n u_i^2} \sqrt{(k \langle \mathbf{u}, \mathbf{v} \rangle_{\mathcal{L}})^2 - 1}}\right)$$

Proof. See appendix D. \square

4 End-to-end PO-HRL

With Lorentzian cosine similarity and Lorentzian entailment cones, we propose the end-to-end PO-HRL framework for FSRE, depicted in Fig. 2(a). It consists of four main components: an encoder to build representations, a representation alignment module to align the representations of relations and instances, a partial order modeling module to impose constraints of relations on instances and a relation classifier. A detailed explanation of each of these components is provided below.

Encoder. As in previous works (Han et al., 2021a; Dou et al., 2022), we adopt BERT (Devlin et al., 2019) as the base encoder. We introduce four special tokens “[E1][E1]” and “[E2][E2]” to augment the given sentence x to mark the start and end of each entity mention and concatenate the

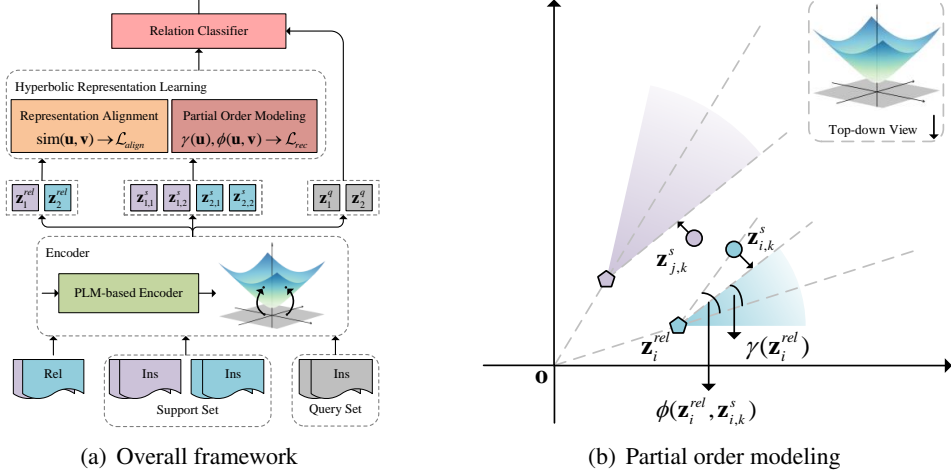


Figure 2: (a) The overall framework of PO-HRL. (b) Illustration of partial order modeling, the coordinate system is a top-down view of the Lorentz model. The blue and purple areas represent the Lorentzian entailment cones of relations i and j respectively. $\gamma(\mathbf{z}_i^{rel})$ is the half-aperture of relation i , $\phi(\mathbf{z}_i^{rel}, \mathbf{z}_{i,k}^s)$ is the exterior angle.

name and description to the augmented x . Then, we feed the concatenated sequence into BERT encoder. The instance representation is formulated as $\mathbf{x}^{ins} = [\mathbf{h}_{E1}; \mathbf{h}_{E2}]$ and the relation representation is denoted as $\mathbf{x}^{rel} = [\mathbf{h}_{CLS}; \mathbf{h}_{avg}]$, where \mathbf{h}_{E1} , \mathbf{h}_{E2} and \mathbf{h}_{CLS} are outputs of [E1], [E2] and [CLS] respectively, \mathbf{h}_{avg} is the average representations of tokens except [CLS]. Since Han et al. (2021a) and Liu et al. (2022) have proven that the addition of relation and instance representations enhances FSRE, we take a weighted sum of their representations as the representation of support instance $\mathbf{x}^s = \mathbf{x}^{ins} + \alpha \mathbf{x}^{rel}$, where $\alpha \in \{0, 1\}$ is a weight coefficient. As for the query instance, its representation $\mathbf{x}^q = \mathbf{x}^{ins}$, since it is unknown which class of relation it belongs to.

For $\mathbf{x} \in \mathbb{R}^n$ generated in Euclidean space, we need to map it to hyperbolic space. Generally, an element “0” is required to be added to \mathbf{x} , i.e., $\tilde{\mathbf{x}} = (0, \mathbf{x}) \in \mathbb{R}^{n+1}$. $\tilde{\mathbf{x}}$ is in the tangent space $\mathcal{T}_o \mathcal{L}_k^n$ of the Lorentz model \mathcal{L}_k^n at the origin $o(1/\sqrt{|k|}, \mathbf{0}^n)$ since $\langle o, \tilde{\mathbf{x}} \rangle_{\mathcal{L}} = 0^4$. Therefore, $\tilde{\mathbf{x}}$ can be mapped to the hyperbolic space by the exponential map \exp_o^k : $\mathbf{z} = \exp_o^k(\tilde{\mathbf{x}}) = \exp_o^k(0, \mathbf{x}) \in \mathbb{R}^{n+1}$. Due to the presence of the exponential operator in \exp_o , the numerical overflow of \mathbf{z} will occur after the exponential map operation. Therefore, we scale \mathbf{x} with the factor $\sqrt{1/n}$ before the exponential map.

$$\mathbf{z} = \exp_o^k(0, \mathbf{x}/\sqrt{n}) \in \mathbb{R}^{n+1} \quad (1)$$

⁴The n dimensional Lorentz model \mathcal{L}_k^n is embedded in the $n + 1$ dimensional Minkowski space, the definition of tangent space is provided in appendix A.

Representation Alignment. As the generated relation and instance representations are distributed in distinct representation spaces, we develop a hyperbolic supervised contrastive learning approach with the derived Lorentzian cosine similarity (described in Section 3.1) to align representations of relations and instances. This alignment is achieved by maximizing the cosine similarity between representations of relations and instances belonging to the same class, while minimizing the similarity between those from different classes. Specifically, in an episode of the N -way- K -shot task, let $Z_r = \{\mathbf{z}_i^{rel} | i = 1, \dots, N\}$ and $Z_s = \{\mathbf{z}_{i,k}^s | i = 1, \dots, N; k = 1, \dots, K\}$ denote the representations of relations and instances in S respectively, the alignment loss is formulated as:

$$\mathcal{L}_{align} = \sum_{i=1}^N -\log \frac{\sum_{k=1}^K e^{\text{sim}(\mathbf{z}_i^{rel}, \mathbf{z}_{i,k}^s)/\tau}}{\sum_{j=1}^N \sum_{k=1}^K e^{\text{sim}(\mathbf{z}_i^{rel}, \mathbf{z}_{j,k}^s)/\tau}} \quad (2)$$

where $\tau > 0$ is the temperature parameter.

Partial Order Modeling. With aligned relation and instance representations, we exploit the margin loss to impose constraints of relations on instances, which forces instances to reside in the Lorentzian entailment cone (described in Section 3.2) of their associated relation, thereby rectifying the representations of the given relation-instance pair. As illustrated in Fig. 2(b), given a relation \mathbf{z}_i^{rel} and two instances $\mathbf{z}_{i,k}^s, \mathbf{z}_{j,k}^s$, where $\mathbf{z}_{i,k}^s$ belongs to \mathbf{z}_i^{rel} but $\mathbf{z}_{j,k}^s$ is not. The objective of the partial order modeling is to enforce $\phi(\mathbf{z}_i^{rel}, \mathbf{z}_{i,k}^s) \leq \gamma(\mathbf{z}_i^{rel})$ while

ensuring $\phi(\mathbf{z}_i^{rel}, \mathbf{z}_{j,k}^s) > \gamma(\mathbf{z}_i^{rel})$. In an episode, it is implemented by the following loss function:

$$\mathcal{L}_{rec} = \sum_{i=1}^N \left(\sum_{k=1}^K l_{i,k}^{in} + \frac{1}{N-1} \sum_{j \neq i} \sum_{k=1}^K l_{i,j,k}^{out} \right) \quad (3)$$

where $l_{i,k}^{in} = E(\mathbf{z}_i^{rel}, \mathbf{z}_{i,k}^s)$, $l_{i,j,k}^{out} = \max(0, m - E(\mathbf{z}_i^{rel}, \mathbf{z}_{j,k}^s))$. $E(\mathbf{u}, \mathbf{v}) = \phi(\mathbf{u}, \mathbf{v}) - \gamma(\mathbf{u})$ captures the discrepancy between the exterior angle and half-aperture and $m > 0$ is the margin. In \mathcal{L}_{rec} , $l_{i,k}^{in}$ serves to pull instances outside of their corresponding relation’s entailment cone back into it. Conversely, $l_{i,j,k}^{out}$ works to push instances of other relations away from the entailment cone. The factor $\frac{1}{N-1}$ is included to balance the proportionality of the instances. Overall, \mathcal{L}_{rec} is designed to gather instances within their specific relation’s entailment cone while separating instances of other relations. **Relation Classifier.** Having aligned and rectified representations of relations and instances in S , the Lorentzian aggregation center of support instances for each relation is leveraged to classify query instances in Q . Zhang et al. (2021) have given the solution, the Lorentzian aggregation center is:

$$\mathbf{z}_i^c = \frac{\sum_{j=1}^K \omega_{ij} \mathbf{z}_{i,j}^s}{\sqrt{|k|} \left\| \sum_{j=1}^K \omega_{ij} \mathbf{z}_{i,j}^s \right\|_{\mathcal{L}}} \quad (4)$$

where $\omega_{ij} > 0$ is the aggregation weight, we calculate it by the Lorentzian cosine similarity:

$$\omega_{ij} = \text{softmax}(\text{sim}(\mathbf{z}_i^{rel}, \mathbf{z}_{i,j}^s)) \quad (5)$$

With the representation of query instance \mathbf{z}^q and the Lorentzian aggregation center \mathbf{z}_i^c in an episode, the probability of relations for the query instance

$$z(y = i | \mathbf{z}^q) = \frac{e^{\text{sim}(\mathbf{z}_i^c, \mathbf{z}^q)}}{\sum_{n=1}^N e^{\text{sim}(\mathbf{z}_n^c, \mathbf{z}^q)}} \quad (6)$$

Then, the loss function is written as:

$$\mathcal{L}_c = -\log(z_y) \quad (7)$$

The final objective function is the weighted sum of the above three loss functions:

$$\mathcal{L} = \mathcal{L}_c + \lambda_1 \mathcal{L}_{align} + \lambda_2 \mathcal{L}_{rec} \quad (8)$$

5 Experiments

5.1 Datasets

To assess the efficacy of our proposed PO-HRL, we conduct experiments on three established benchmark datasets FewRel 1.0 (Han et al., 2018), FewRel 2.0 (Gao et al., 2019) and Semeval (Hendrickx et al., 2009). For an introduction to the datasets, please see appendix E.

5.2 Implementation Details

The experimental platform is a 24 GB NVIDIA RTX 3090 GPU. Following previous works (Han et al., 2018; Gao et al., 2019), we evaluate PO-HRL by measuring its accuracy on the query set in four N -way- K -shot scenarios, where N is 5 or 10 while K is 1 or 5. We validate our model on 10,000 randomly sampled episodes in validation set. For FewRel 1.0 and FewRel 2.0, the test performance is achieved on the FewRel Leaderboard⁵. Hyperparameter settings see appendix F. We will release our code as open source for further research.

5.3 Main Results

Tables 1,2 and 3 present the comparative results on FewRel 1.0, FewRel 2.0, and Semeval, respectively. Introductions of baselines are listed in appendix G. For FewRel 1.0, results are bifurcated based on whether external information is incorporated. Additionally, we exhibit benchmarks with various approaches benefiting from post-training (Peng et al., 2020). On FewRel 2.0 and Semeval, only BERT encoder outcomes are displayed. The results uncovered the following insights:

(1) Our proposed PO-HRL method attains state-of-the-art performance across all three datasets, outperforming baselines with the same encoder and achieving substantial improvements in 1-shot settings. Notably, compared to the second-best method, our method boosts accuracy by 0.98 and 2.88 points on FewRel 2.0, and by 0.94 and 2.23 points on Semeval for 1-shot settings.

(2) Performance gains arise chiefly from enforcing partial order constraints. Relation and instance representations are collectively aligned to embed partial order, thereby instances are constrained by specificity. This constraint not only steers the conformity with generalization-specificity principles but also facilitates the transfer of relation knowledge to instances of unseen relations.

(3) PO-HRL achieves significantly higher accuracy improvements on validation set than on test set on FewRel 1.0. This divergence arises since relation descriptions serve as auxiliary inputs, increasing interpretability of relations and simplifying the task. In their absence on FewRel 2.0 and Semeval, test accuracy improvements are more evident, confirming the enhanced performance when relation descriptions are unavailable.

⁵<https://thunlp.github.io/fewrel.html>

	Model	5-way-1-shot	5-way-5-shot	10-way-1-shot	10-way-5-shot
w/ external information	REGRAB (Qu et al., 2020)	87.95/90.30	92.54/94.25	80.26/84.09	86.72/89.93
	ConceptFERE (Yang et al., 2021)	--/89.21	--/90.34	--/75.72	--/81.82
w/o external information	Proto-BERT (Snell et al., 2017)	82.92/80.68	91.32/89.60	73.24/71.48	83.68/82.89
	BERT-PAIR (Gao et al., 2019)	87.95/90.30	92.54/94.25	80.26/84.09	86.72/89.93
	HCRP (Han et al., 2021a)	90.90/93.76	93.22/95.66	84.11/89.95	87.79/92.10
	SimpleFSRE (Liu et al., 2022)	91.29/94.42	94.05/96.37	86.09/90.73	89.68/93.47
	FAEA (Dou et al., 2022)	90.81/95.10	94.24/96.48	84.22/90.12	88.74/92.72
	GM_GEN (Li and Qian, 2022)	92.65/94.89	95.62/96.96	86.81/91.23	91.27/94.30
	BMIPN (Li et al., 2023)	91.99/95.62	94.70/96.61	84.95/91.43	89.60/93.88
	HND (Zhang et al., 2023)	93.35/95.21	95.94/97.19	87.41/91.59	91.71/94.54
PO-HRL	93.38/95.51	95.73/97.28	88.65/91.71	91.97/94.59	
w/ post-training	MTB (Soares et al., 2019)	--/91.10	--/95.40	--/84.30	--/91.80
	CP (Peng et al., 2020)	--/95.10	--/97.10	--/91.20	--/94.70
	HCRP(CP) (Han et al., 2021a)	94.10/96.42	96.05/97.96	89.13/93.97	93.10/96.46
	SimpleFSRE(CP) (Liu et al., 2022)	96.21/96.63	97.07/97.93	93.38/94.94	95.11/96.39
	FAEA(CP) (Dou et al., 2022)	94.11/96.36	89.55/97.85	86.59/93.82	93.64/96.29
	GM_GEN(CP) (Li and Qian, 2022)	96.97/97.03	98.32/98.34	93.97/94.99	96.58/96.91
	PO-HRL(CP)	97.18/97.51	98.41/98.41	94.49/95.29	96.85/97.01

Table 1: Accuracy (%) of FSRE task on FewRel 1.0 validation/test set. The table is divided into three parts. The first two parts use BERT as the encoder, while the encoder of the third part is BERT with post-training.

Model	5-way 1-shot	5-way 5-shot	10-way 1-shot	10-way 5-shot
Proto-BERT	40.12	51.50	26.45	36.93
BERT-PAIR	67.41	78.57	54.89	66.85
HCRP	76.34	83.03	63.77	72.94
FAEA	73.58	90.10	62.98	80.51
GM_GEN	76.67	91.28	64.19	84.84
BMIPN	77.19	90.19	66.29	82.81
HND	78.37	91.41	66.54	84.92
DCFT w/o DTM	79.36	90.71	--	--
PO-HRL	80.34	91.44	69.36	85.26

Table 2: Accuracy (%) of FSRE task on FewRel 2.0 test set. Since DCFT (Liu et al., 2024) introduces additional unlabeled data from the target domain for domain-aware transformation, we only list the results of DCFT without DTM.

5.4 Ablation Study

To investigate the effectiveness of the main components of our proposed PO-HRL, we design four ablation experiments on FewRel 1.0 and FewRel 2.0 validation sets, and Semeval test set. The results are reported in Table 4, we delineate the specific variants and analyze the effects as follows:

(1) To verify the efficacy of the representation alignment module, we exclude it from the PO-HRL. On the three datasets, the average accuracy is reduced by 0.79, 1.34, and 1.48 points respectively, confirming that aligning relation and instance representations boosts the performance.

(2) To isolate the impact of partial order mod-

Model	5-way 1-shot	5-way 5-shot	10-way 1-shot	10-way 5-shot
Proto-BERT	48.54	78.48	36.43	68.15
BERT-PAIR	49.70	67.64	37.71	55.14
HCRP	56.98	73.65	43.75	62.64
FAEA	59.03	76.99	46.27	66.48
GM_GEN	51.48	79.02	44.96	69.86
PO-HRL	59.97	79.51	48.50	70.14

Table 3: Accuracy (%) of FSRE task on Semeval test set. Baseline results are reported by Liu et al. (2024).

eling, we directly feed the aligned relation and instance representations into the classifier. This scatters the distribution of instances, substantially degrading performance by 0.75, 2.44 and 1.35 points respectively. The collapse is most pronounced in the extremely sample-starved 1-shot settings, underscoring the indispensability of the partial order modeling module. Note that the decline in performance for FewRel 1.0 is smaller than that for FewRel 2.0 and Semeval. This can be attributed to the fact that FewRel 1.0 offers relation descriptions, which alleviates the collapse.

(3) We further remove the two modules of representation alignment and partial order modeling, reducing our proposed PO-HRL to the encoder and classifier only. The average accuracy decreases by 1.03 points on FewRel 1.0, 2.31 points on FewRel 2.0 and 1.46 on Semeval. This performance degradation confirms the essential role of

Model	FewRel 1.0				FewRel 2.0				Semeval			
	5-way 1-shot	5-way 5-shot	10-way 1-shot	10-way 5-shot	5-way 1-shot	5-way 5-shot	10-way 1-shot	10-way 5-shot	5-way 1-shot	5-way 5-shot	10-way 1-shot	10-way 5-shot
PO-HRL	93.38	95.73	88.65	91.97	82.39	92.95	73.82	87.24	59.97	79.51	48.50	70.14
w/o RA	92.94	94.69	88.05	90.90	80.26	91.59	73.63	85.57	57.90	79.24	45.66	69.41
w/o POM	92.79	95.38	87.41	91.15	80.06	92.60	68.12	85.86	58.69	78.90	46.84	68.31
w/o RSA+POM	92.40	94.86	87.63	90.74	80.35	91.68	69.92	85.23	56.52	79.18	47.23	69.34
LA w/o weight	--	94.98	--	91.36	--	91.92	--	86.39	--	78.82	--	69.65

Table 4: Accuracy (%) of ablation study on FewRel 1.0 and FewRel 2.0 validation sets, and Semeval test set. RA stands for representation alignment, POM indicates partial order modeling, LA denotes Lorentzian aggregation.

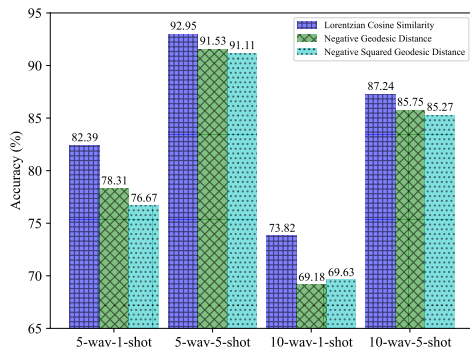


Figure 3: Accuracy (%) achieved by hyperbolic supervised contrastive learning when employing various similarity measures on FewRel 2.0 validation set.

representation alignment and partial order modeling in connecting relations and instances for FSRE.

(4) The Lorentzian aggregation center of each relation is obtained by a weighted sum of the support instances. To analyze the effect of this weighted summation, we perform an ablation by directly averaging the support instance representations instead. In 1-shot settings, this ablation is ineffective. Under 5-shot settings, the accuracy on FewRel 1.0 decreased by 0.75 and 0.61 points, on FewRel 2.0 fell by 1.03 and 0.85 points, while correspondingly, on Semeval, the performance degradations are 0.69 and 0.49. By equally averaging, PO-HRL fails to selectively amplify the most representative instances. In contrast, the learned weighted summation places greater emphasis on critical instances.

5.5 Analysis of Lorentzian Cosine Similarity

The partial order modeling is implemented by constraining the angle between the representations of relations and instances in the Lorentz model. Therefore, we exploit the Lorentzian cosine to quantify the similarity between vectors for contrastive learning. The Lorentzian cosine focuses on the angle between two vectors, assessing directional alignment while disregarding magnitude. Plausible alterna-

tives are negative geodesic distance and its squared variant (Law et al., 2019). However, these metrics capture both angular and norm differences, which is detrimental to partial order modeling. To assess the efficacy of Lorentzian cosine similarity, we conduct experiments involving the alignment loss on FewRel 2.0 validation set, substituting Lorentzian cosine similarity with negative geodesic distance and squared geodesic distance. The outcomes are presented in Fig. 3. Using negative geodesic distance and squared geodesic distance results in a 2.88 and 3.43 percentage point decrease in average accuracy, respectively. Notably, the accuracy drop is more significant in 1-shot settings. These findings suggest that angular alignment via Lorentzian cosine similarity is a more effective way to capture semantic similarity in FSRE. Although we cannot definitively assert that Lorentzian cosine similarity outperforms negative geodesic distance or negative squared geodesic distance in all cases, it seems more appropriate under the current model setting.

5.6 Efficacy of Partial Order Modeling

To validate the efficacy of partial order modeling, we present a 5-way-1-shot task drawn from the FewRel 2.0 validation set, as depicted in Fig.4. In the biomedical domain, the relationships “*is primary anatomic site of*” and “*occurs in*” are semantically akin. The scarcity of samples poses a significant challenge to the model’s classification capabilities. Through the application of partial order modeling, the model accurately categorizes the query instance as “*is primary anatomic site of*”. Conversely, without this modeling, the query instance is misclassified as “*occurs in*”. We visualize the similarity between the query instance and the supporting instance in both scenarios. The experimental outcomes corroborate that PO-HRL is superior in learning the instance distribution under conditions of limited sample availability.

Support Set
1.is primary anatomic site of Inactivation of GTPase at the pancreatic intraepithelial neoplasia stage promotes pancreatic tissue repair.
2.occurs in Neonatal and maternal outcomes were significantly improved in the study group compared with the control group (fetal distress [10 % vs 37 %]).
3.inheritance type of Arterial tortuosity syndrome is a rare autosomal recessive connective tissue disease.
4.gene found in organism Human endometriotic epithelial cells (11z and 12z) showed a high level of obr .
5.biological process involves gene product We investigated whether they interact with the 5-ht4(b) receptor , and are involved in the regulation of 5-ht4(b) receptor signaling .
Query Keloids are locally aggressive scars that typically invade into healthy surrounding skin and cause both physical and psychosocial distress to the patient.

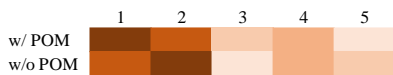


Figure 4: An example of a 5-way-1-shot task. We list the instances and their respective relation name, the subject and object of the instance are colored red and blue respectively. We visualize the similarity between query instance and support instances (w/ and w/o POM), with darker units representing greater similarity.

Model	5-way 1-shot	5-way 5-shot	10-way 1-shot	10-way 5-shot
HCRP	87.40	93.54	82.69	89.97
SimpleFSRE	89.23	94.38	82.90	89.35
FAEA	87.92	92.67	84.65	89.47
GM_GEN	89.77	95.14	83.62	91.29
PO-HRL	90.29	95.41	85.34	91.69

Table 5: Accuracy (%) of FSRE task without relation descriptions on FewRel 1.0 validation set.

5.7 Absence of Relation Descriptions

Existing FSRE works rely heavily on semantic information from relation descriptions. However, descriptions are often unavailable in real-world scenarios, as with FewRel 2.0. To compare the robustness of our approach and baselines when lacking relation descriptions, we conduct experiments on FewRel 1.0 where only relation names and instances are provided, without any descriptions. The results on the validation set are presented in Table 5. We observe performance degradation across all models when relation descriptions are removed, underscoring their importance. However, our proposed PO-HRL exhibited greater robustness compared to baselines under such low-resource conditions. This reflects the stronger generalizability of our model for tackling FSRE tasks, even when descriptive data is limited or unavailable. By leveraging representation alignment and partial order

modeling, PO-HRL is less dependent on relation descriptions and confers greater resilience and flexibility when facing incomplete real-world datasets.

5.8 Influence of Hyper-parameters

In PO-HRL, hyper-parameters C and m are crucial in partial order modeling. Through experiments on four distinct settings on the FewRel 2.0 dataset, we delve into how varying the values of C and m impacts outcomes. The results are shown in Fig. 5.

The constant C governs the half-aperture of the Lorentzian entailment cone. It is evident that the model’s efficacy peaks when C is dialed to 0.3 or 0.4. This optimal setting arises because an excessively narrow half-aperture may inadequately differentiate between relation representations, whereas an overly wide gap can lead to cone overlap, hampering instance aggregation.

Margin m , conversely, dictates the separation between Lorentzian entailment cones. Small m brings cones of disparate relations closer, while large m creates greater separation between them. The results indicate that PO-HRL achieves the best performance when m is set to 0.2. Too minute m could result in insufficient distinction between instances of various classes, whereas excessively large m may force instances out of their corresponding relation cones due to the influence of the loss function in Eq. 3.

5.9 Visualization of PO-HRL

To verify that our proposed PO-HRL model can learn reliable representations, we visualize the distribution of instance representations before and after training using t-SNE (Van der Maaten and Hinton, 2008). Specifically, we select the 5-way-5-shot trained PO-HRL model on FewRel 2.0 and perform inference on the validation set, randomly sampling 20 examples per relation. Since the learned representations reside in hyperbolic space, we apply the logarithmic map to project them to Euclidean space for visualization. As shown in Fig. 6, after training, instances of the same relation cluster together while those from different relations become clearly separated. This demonstrates PO-HRL’s ability to effectively learn discriminative representations that distinguish fine-grained semantic relations, validating its few-shot relation learning capacity. The clear separation and clustering of relation instances in the embedding space illustrate the model’s success in learning reliable representations.

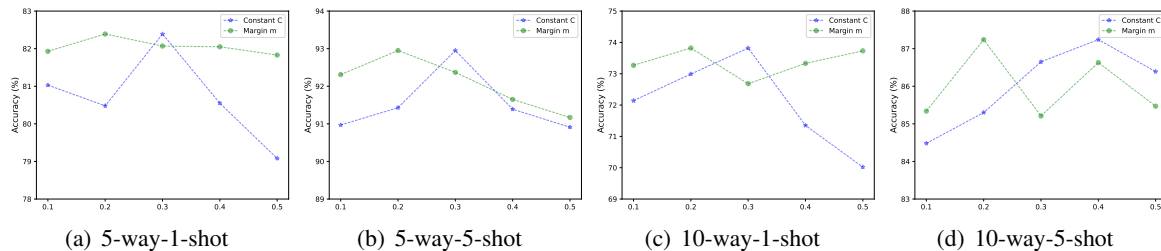


Figure 5: Accuracy (%) on FewRel 2.0 validation set with various constant C and margin m .

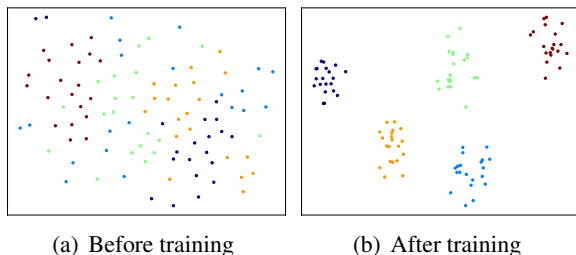


Figure 6: t-SNE plots of instance representations before/after training with 5 relations, 20 instances each.

6 Related Work

Few-shot relation extraction (FSRE) is intended to predict novel relations between entities mentioned in given sentences, using just a few labeled instances. Han et al. (2018) and Gao et al. (2019) propose the FewRel 1.0 and FewRel 2.0 benchmark datasets for FSRE, and provide effective baselines Proto-BERT and BERT-PAIR. Existing studies are mostly based on prototype networks. REGRAB (Qu et al., 2020) and Concept-FERE (Yang et al., 2021) enhance prototypes with external knowledge. However, utilizing external knowledge is laborious. Consequently, recent focus has turned to methods (Han et al., 2021a; Liu et al., 2022; Li et al., 2023) that only leverage the provided texts and relation descriptions. They directly add relation embeddings to instance representation centers to generate relation prototypes. Methods mentioned above follow a “one-for-all” scheme, to mine differences of each N -way- K -shot task, GM_GEN (Li and Qian, 2022) and HND (Zhang et al., 2023) optimize the model with generation modules. In this paper, we treat FSRE as a representation learning task that learns the representation distribution of relations and instances via intrinsic partial order constraints in hyperbolic space.

Hyperbolic neural networks have been widely investigated for representation learning due to the superiority of hyperbolic geometry in modeling

hierarchical structures. To embed the underlying hierarchy, works (Nickel and Kiela, 2017, 2018) first learn hierarchical representations in hyperbolic spaces like the Poincaré ball and Lorentz model. Since then, various neural models (Ganea et al., 2018b; Gulcehre et al., 2019; Liu et al., 2019; Chami et al., 2019) are extended to hyperbolic spaces. Further works (Shimizu et al., 2021; Chen et al., 2022) optimize hyperbolic neural components to maintain efficiency and stability. Among these, HCL (Ge et al., 2023) employs the negative geodesic distance to generalize contrastive learning (Chen et al., 2020; Khosla et al., 2020) to hyperbolic space, facilitating representation learning. Concurrently, several other studies (Ganea et al., 2018a; Le et al., 2019; Bai et al., 2021; Desai et al., 2023) utilize the hyperbolic entailment cones to capture partial order relations. Inspired by previous works on hyperbolic space, we derive Lorentzian cosine similarity and extend contrastive learning into hyperbolic space to align the representations of relations and instances. Meanwhile, we view relations as generic concepts and instances as specific exemplars, and model the inherent partial order using the Lorentzian entailment cones.

7 Conclusion

In this paper, we propose a novel partial order-centered hyperbolic representation learning (PO-HRL) framework to learn reliable representation of relations and instances and mitigate the problem of relation prototype bias present in existing works. The framework aligns relation and instance representations via hyperbolic contrastive learning based on Lorentzian cosine similarity and exploits Lorentzian entailment cones to model the partial order. Extensive experiments on three public datasets show that PO-HRL outperforms the strong baselines, with outstanding performance in low-resource 1-shot settings.

Limitations

The limitations of PO-HRL are primarily twofold: 1) Its efficacy as a representation learning method has only been confirmed in the FSRE task, and its generalizability to other few-shot learning scenarios that exhibit partial order relations has not been investigated. 2) Since PO-HRL is implemented in the Lorentz model, which involves numerical operations in non-Euclidean space, its computational efficiency is slightly insufficient.

Acknowledgement

We thank the anonymous reviewers for their helpful comments. This work was supported by the National Natural Science Foundation of China (No. 62376284, No. 62476283, No. 42305159).

References

- Yushi Bai, Zhitao Ying, Hongyu Ren, and Jure Leskovec. 2021. Modeling heterogeneous hierarchies with relation-specific hyperbolic cones. In *Advances in Neural Information Processing Systems*, pages 12316–12327.
- Ines Chami, Zhitao Ying, Christopher Ré, and Jure Leskovec. 2019. Hyperbolic graph convolutional neural networks. In *Advances in Neural Information Processing Systems*, pages 4869–4880.
- Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey E. Hinton. 2020. A simple framework for contrastive learning of visual representations. In *Proceedings of the 37th International Conference on Machine Learning*, pages 1597–1607.
- Weize Chen, Xu Han, Yankai Lin, Hexu Zhao, Zhiyuan Liu, Peng Li, Maosong Sun, and Jie Zhou. 2022. Fully hyperbolic neural networks. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*, pages 5672–5686.
- Karan Desai, Maximilian Nickel, Tanmay Rajpurohit, Justin Johnson, and Shanmukha Ramakrishna Vedantam. 2023. Hyperbolic image-text representations. In *Proceedings of the 40th International Conference on Machine Learning*, pages 7694–7731.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4171–4186.
- Chunliu Dou, Shaojuan Wu, Xiaowang Zhang, Zhiyong Feng, and Kewen Wang. 2022. Function-words adaptively enhanced attention networks for few-shot inverse relation classification. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence*, pages 2937–2943.
- Chelsea Finn, Pieter Abbeel, and Sergey Levine. 2017. Model-agnostic meta-learning for fast adaptation of deep networks. In *Proceedings of the 34th International Conference on Machine Learning*, pages 1126–1135.
- Octavian-Eugen Ganea, Gary Bécigneul, and Thomas Hofmann. 2018a. Hyperbolic entailment cones for learning hierarchical embeddings. In *Proceedings of the 35th International Conference on Machine Learning*, pages 1632–1641.
- Octavian-Eugen Ganea, Gary Bécigneul, and Thomas Hofmann. 2018b. Hyperbolic neural networks. In *Advances in Neural Information Processing Systems*, pages 5350–5360.
- Tianyu Gao, Xu Han, Hao Zhu, Zhiyuan Liu, Peng Li, Maosong Sun, and Jie Zhou. 2019. Fewrel 2.0: Towards more challenging few-shot relation classification. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, pages 6249–6254.
- Songwei Ge, Shlok Mishra, Simon Kornblith, Chun-Liang Li, and David Jacobs. 2023. Hyperbolic contrastive learning for visual representations beyond objects. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6840–6849.
- Caglar Gulcehre, Misha Denil, Mateusz Malinowski, Ali Razavi, Razvan Pascanu, Karl Moritz Hermann, Peter W. Battaglia, Victor Bapst, David Raposo, Adam Santoro, and Nando de Freitas. 2019. Hyperbolic attention networks. In *7th International Conference on Learning Representations*.
- Jiale Han, Bo Cheng, and Wei Lu. 2021a. Exploring task difficulty for few-shot relation extraction. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 2605–2616.
- Jiale Han, Bo Cheng, and Xu Wang. 2020. Two-phase hypergraph based reasoning with dynamic relations for multi-hop KBQA. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*, pages 3615–3621.
- Xu Han, Hao Zhu, Pengfei Yu, Ziyun Wang, Yuan Yao, Zhiyuan Liu, and Maosong Sun. 2018. Fewrel: A large-scale supervised few-shot relation classification dataset with state-of-the-art evaluation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4803–4809.
- Yi Han, Linbo Qiao, Jianming Zheng, Zhigang Kan, Linhui Feng, Yifu Gao, Yu Tang, Qi Zhai, Dongsheng Li, and Xiangke Liao. 2021b. Multi-view interaction learning for few-shot relation classification.

- In *CIKM '21: The 30th ACM International Conference on Information and Knowledge Management*, pages 649–658.
- Iris Hendrickx, Su Nam Kim, Zornitsa Kozareva, Preslav Nakov, Diarmuid Ó Séaghdha, Sebastian Padó, Marco Pennacchiotti, Lorenza Romano, and Stan Szpakowicz. 2009. Semeval-2010 task 8: Multi-way classification of semantic relations between pairs of nominals. In *Proceedings of the Workshop on Semantic Evaluations: Recent Achievements and Future Directions, SEW@NAACL-HLT*, pages 94–99.
- Zhen Huang, Zhilong Lv, Xiaoyun Han, Binyang Li, Menglong Lu, and Dongsheng Li. 2022. Social bot-aware graph neural network for early rumor detection. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 6680–6690.
- Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschiot, Ce Liu, and Dilip Krishnan. 2020. Supervised contrastive learning. In *Advances in Neural Information Processing Systems*, pages 18661–18673.
- Marc Teva Law, Renjie Liao, Jake Snell, and Richard S. Zemel. 2019. Lorentzian distance learning for hyperbolic representations. In *Proceedings of the 36th International Conference on Machine Learning*, pages 3672–3681.
- Matt Le, Stephen Roller, Laetitia Papaxanthos, Douwe Kiela, and Maximilian Nickel. 2019. Inferring concept hierarchies from text corpora via hyperbolic embeddings. In *Proceedings of the 57th Conference of the Association for Computational Linguistics*, pages 3231–3241.
- Wanli Li and Tiejun Qian. 2022. Graph-based model generation for few-shot relation extraction. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 62–71.
- Yile Li, Yinliang Yue, Xiaoyan Gu, Peng Fu, and Weiping Wang. 2023. BMIPN: A biased multi-granularity interaction prototype network for few-shot relation extraction. In *26th European Conference on Artificial Intelligence - Including 12th Conference on Prestigious Applications of Intelligent Systems*, pages 1455–1462.
- Zhenzhen Li, Yuyang Zhang, Jian-Yun Nie, and Dongsheng Li. 2022. Improving few-shot relation classification by prototypical representation learning with definition text. In *Findings of the Association for Computational Linguistics*, pages 454–464.
- Qi Liu, Maximilian Nickel, and Douwe Kiela. 2019. Hyperbolic graph neural networks. In *Advances in Neural Information Processing Systems*, pages 8228–8239.
- Yang Liu, Jinpeng Hu, Xiang Wan, and Tsung-Hui Chang. 2022. A simple yet effective relation information guided approach for few-shot relation extraction. In *Findings of the Association for Computational Linguistics*, pages 757–763.
- Yijun Liu, Feifei Dai, Xiaoyan Gu, Minghui Zhai, Bo Li, and Meiou Zhang. 2024. Domain-aware and co-adaptive feature transformation for domain adaptation few-shot relation extraction. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation*, pages 5275–5285.
- Menglong Lu, Zhen Huang, Binyang Li, Yunxiang Zhao, Zheng Qin, and Dong Sheng Li. 2022. SIFTER: A framework for robust rumor detection. *IEEE ACM Trans. Audio Speech Lang. Process.*, 30:429–442.
- Makoto Miwa and Mohit Bansal. 2016. End-to-end relation extraction using lstms on sequences and tree structures. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, pages 1105–1116.
- Maximilian Nickel and Douwe Kiela. 2017. Poincaré embeddings for learning hierarchical representations. In *Advances in Neural Information Processing Systems*, pages 6338–6347.
- Maximilian Nickel and Douwe Kiela. 2018. Learning continuous hierarchies in the lorentz model of hyperbolic geometry. In *Proceedings of the 35th International Conference on Machine Learning*, pages 3776–3785.
- Jouni Parkkonen. 2012. Hyperbolic geometry.
- Hao Peng, Tianyu Gao, Xu Han, Yankai Lin, Peng Li, Zhiyuan Liu, Maosong Sun, and Jie Zhou. 2020. Learning from context or names? an empirical study on neural relation extraction. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, pages 3661–3672.
- Meng Qu, Tianyu Gao, Louis-Pascal A. C. Xhonneux, and Jian Tang. 2020. Few-shot relation extraction via bayesian meta-learning on relation graphs. In *Proceedings of the 37th International Conference on Machine Learning*, pages 7867–7876.
- Ying Shen, Ning Ding, Hai-Tao Zheng, Yaliang Li, and Min Yang. 2021. Modeling relation paths for knowledge graph completion. *IEEE Trans. Knowl. Data Eng.*, 33(11):3607–3617.
- Ryohei Shimizu, Yusuke Mukuta, and Tatsuya Harada. 2021. Hyperbolic neural networks++. In *9th International Conference on Learning Representations*.
- Jake Snell, Kevin Swersky, and Richard S. Zemel. 2017. Prototypical networks for few-shot learning. In *Advances in Neural Information Processing Systems*, pages 4077–4087.

- Livio Baldini Soares, Nicholas FitzGerald, Jeffrey Ling, and Tom Kwiatkowski. 2019. Matching the blanks: Distributional similarity for relation learning. In *Proceedings of the 57th Conference of the Association for Computational Linguistics*, pages 2895–2905.
- Laurens Van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-sne. *Journal of machine learning research*, 9(11).
- Yilin Wang, Minghao Hu, Zhen Huang, Dongsheng Li, Wei Luo, Dong Yang, and Xicheng Lu. 2023. A canonicalization-enhanced known fact-aware framework for open knowledge graph link prediction. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, pages 2332–2342.
- Yan Xu, Lili Mou, Ge Li, Yunchuan Chen, Hao Peng, and Zhi Jin. 2015. Classifying relations via long short term memory networks along shortest dependency paths. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1785–1794.
- Shan Yang, Yongfei Zhang, Guanglin Niu, Qinghua Zhao, and Shiliang Pu. 2021. Entity concept-enhanced few-shot relation extraction. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*, pages 987–991.
- Liang Zhang, Chulun Zhou, Fandong Meng, Jinsong Su, Yidong Chen, and Jie Zhou. 2023. Hypernetwork-based decoupling to improve model generalization for few-shot relation extraction. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 6213–6223.
- Yiding Zhang, Xiao Wang, Chuan Shi, Nian Liu, and Guojie Song. 2021. Lorentzian graph convolutional networks. In *The Web Conference 2021*, pages 1249–1261.

A Introduction of the Lorentz Model

In this section, we provide an introduction to some necessary concepts in the Lorentz model that are relevant to this paper. Notations used in this paper are listed in Table 6.

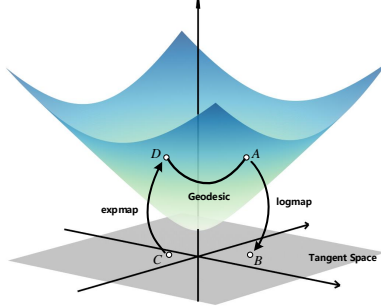


Figure 7: Illustration of the two dimensional Lorentz model \mathcal{L}_k^2 and tangent space $\mathcal{T}_o \mathcal{L}_k^2$ at the origin $\mathbf{o}(1/\sqrt{|k|}, 0, 0)$. A and D are points in \mathcal{L}_k^2 , B and C are points in $\mathcal{T}_o \mathcal{L}_k^2$. A maps to B via the logarithmic map, and C maps to D via the exponential map. The geodesic between A and D is the shortest curve joining them, its length is calculated by Eq. 12.

Definition. As illustrated in Fig. 7, With a constant negative curvature $k(k < 0)$, the n dimensional Lorentz model is defined as a Riemannian manifold $\mathcal{L}_k^n = (\mathcal{H}_k^n, g_x^{\mathcal{L}})$ embedded in the $n+1$ dimensional Minkowski space, in which

$$\mathcal{H}_k^n = \{\mathbf{x} \in \mathbb{R}^{n+1} | \langle \mathbf{x}, \mathbf{x} \rangle_{\mathcal{L}} = 1/k, x_0 > 0\} \quad (9)$$

represents the upper sheet of an n dimensional hyperboloid with the origin $\mathbf{o}(1/\sqrt{|k|}, \mathbf{0}^n)$, $g_x^{\mathcal{L}} = \text{diag}(-1, \mathbf{1}^n)$ is the Riemannian metric tensor. $\langle \cdot, \cdot \rangle_{\mathcal{L}}$ denotes the Lorentzian inner product:

$$\langle \mathbf{x}, \mathbf{y} \rangle_{\mathcal{L}} = \mathbf{x} g_x^{\mathcal{L}} \mathbf{y} = -x_0 y_0 + \sum_{i=1}^n x_i y_i \quad (10)$$

Tangent Space. $\forall \mathbf{x} \in \mathcal{L}_k^n$, the tangent space $\mathcal{T}_x \mathcal{L}_k^n$ of \mathcal{L}_k^n at \mathbf{x} is defined as an n dimensional vector space of the first-order approximation to \mathcal{L}_k^n around \mathbf{x} , which is the orthogonal space of \mathcal{L}_k^n at \mathbf{x} with respect to the Lorentzian inner product:

$$\mathcal{T}_x \mathcal{L}_k^n = \{\mathbf{v} \in \mathbb{R}^{n+1} | \langle \mathbf{x}, \mathbf{v} \rangle_{\mathcal{L}} = 0\} \quad (11)$$

Geodesics. Geodesics are the generalization of straight lines in Euclidean space to manifolds. In the Lorentz model, a geodesic between $\mathbf{x}, \mathbf{y} \in \mathcal{L}_k^n$ is the shortest curve joining \mathbf{x} to \mathbf{y} . Based on the Riemannian metric $g_x^{\mathcal{L}}$, the geodesic distance between \mathbf{x} and \mathbf{y} is given as:

$$d_{\mathcal{L}}^k(\mathbf{x}, \mathbf{y}) = \sqrt{1/|k|} \cdot \cosh^{-1}(k \langle \mathbf{x}, \mathbf{y} \rangle_{\mathcal{L}}) \quad (12)$$

Notation	Meaning
\mathcal{L}_k^n	n dimensional Lorentz model of curvature k
\mathbf{o}	Origin of the Lorentz model \mathcal{L}_k^n
$\langle \mathbf{x}, \mathbf{y} \rangle_{\mathcal{L}}$	Lorentzian inner product between \mathbf{x} and \mathbf{y}
$\mathcal{T}_x \mathcal{L}_k^n$	Tangent space of \mathcal{L}_k^n at \mathbf{x}
$d_{\mathcal{L}}^k(\mathbf{x}, \mathbf{y})$	Geodesic distance between \mathbf{x} and \mathbf{y}
\exp_x^k	Exponential map
$\gamma(\mathbf{u})$	Half-aperture of Lorentzian entailment cone of vector \mathbf{u}
$\phi(\mathbf{u}, \mathbf{v})$	Angle between half-lines $\langle \mathbf{o}\mathbf{u}$ and $\langle \mathbf{u}\mathbf{v}$
z_i^{rel}	Hyperbolic representation of i -th relation
$z_{j,k}^s$	Hyperbolic representation of k -th instance of i -th relation in support set
z^q	Hyperbolic representation of instance in query set
z_i^c	Lorentzian aggregation center of i -th relation

Table 6: Summary of notations.

Exponential and Logarithmic Maps. Mapping between Lorentz model and its tangent space is realized by exponential map and logarithmic map. Exponential map $\exp_x^k : \mathcal{T}_x \mathcal{L}_k^n \rightarrow \mathcal{L}_k^n$ projects a vector $\mathbf{v} \in \mathcal{T}_x \mathcal{L}_k^n$ to \mathcal{L}_k^n , and logarithmic map $\log_x^k : \mathcal{L}_k^n \rightarrow \mathcal{T}_x \mathcal{L}_k^n$ is the inverse of \exp_x^k . The exponential and logarithmic map are defined as:

$$\begin{aligned} \exp_x^k(\mathbf{v}) &= \cosh(\sqrt{|k|} \|\mathbf{v}\|_{\mathcal{L}}) \mathbf{x} + \frac{\sinh(\sqrt{|k|} \|\mathbf{v}\|_{\mathcal{L}})}{\sqrt{|k|} \|\mathbf{v}\|_{\mathcal{L}}} \mathbf{v} \\ \log_x^k(\mathbf{z}) &= \frac{\cosh^{-1}(k \langle \mathbf{x}, \mathbf{z} \rangle_{\mathcal{L}})}{\sqrt{(k \langle \mathbf{x}, \mathbf{z} \rangle_{\mathcal{L}})^2 - 1}} (\mathbf{z} - k \langle \mathbf{x}, \mathbf{z} \rangle_{\mathcal{L}} \mathbf{x}) \end{aligned} \quad (13)$$

where $\|\mathbf{v}\|_{\mathcal{L}} = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle_{\mathcal{L}}}$ denotes the Lorentzian norm of \mathbf{v} .

B Proof of Theorem 3.1

We first introduce the hyperbolic law of cosines. Then, we construct a hyperbolic triangle in the Lorentz model. Based on the hyperbolic law of cosines, we derive the Lorentzian cosine similarity.

Hyperbolic Law of Cosines. In hyperbolic space, angle is a generalization of angle in Euclidean space, which is defined as the angle formed by two geodesics at their intersection. As in Euclidean space, it is measured by the angle between the initial tangent vectors of these two geodesics. With the concepts of angle and geodesic, the triangle can be defined within hyperbolic space. For $A, B, C \in \mathcal{L}_k^n$, a hyperbolic triangle $\triangle ABC$ is constructed by joining any two points through geodesics. Let $a = d_{\mathcal{L}}^k(B, C)$ denotes the length of geodesic between points B and C (and others), the hyperbolic law of cosines (Parkkonen, 2012) is established as follows:

$$\cos(\angle C) = \frac{\cosh(a\sqrt{|k|}) \cosh(b\sqrt{|k|}) - \cosh(c\sqrt{|k|})}{\sinh(a\sqrt{|k|}) \sinh(b\sqrt{|k|})} \quad (14)$$

Lorentzian cosine similarity. As illustrated in Fig. 8, given $\mathbf{u}, \mathbf{v} \in \mathcal{L}_k^n \setminus \{\mathbf{o}\}$ and the origin

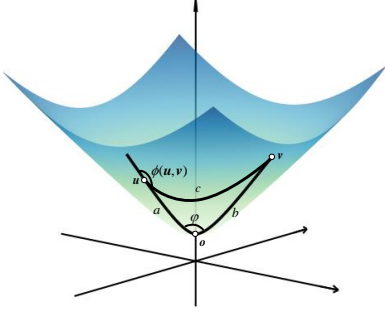


Figure 8: Given any two points \mathbf{u} and \mathbf{v} in \mathcal{L}_k^2 except the origin \mathbf{o} , they form a hyperbolic triangle with the origin. c represents the geodesic between \mathbf{u} and \mathbf{v} (similar for a and b), φ is the angle formed by a and b . The hyperbolic cosine of φ is utilized to quantify the similarity between \mathbf{u} and \mathbf{v} . $\phi(\mathbf{u}, \mathbf{v})$ denotes the exterior angle between half-lines $(\mathbf{o}\mathbf{u}$ and $(\mathbf{o}\mathbf{v}$.

$\mathbf{o}(1/\sqrt{|k|}, \mathbf{0}^n)$ in the Lorentz model, any two points are joined by geodesic to construct a hyperbolic triangle. We use φ to denote the angle formed by \mathbf{u} and \mathbf{v} , a , b , and c to represent the corresponding geodesics respectively. With Eq. 12, we have

$$\begin{aligned} a &= \sqrt{1/|k|} \cosh^{-1}(k\langle \mathbf{o}, \mathbf{u} \rangle_{\mathcal{L}}) = \sqrt{1/|k|} \cosh^{-1}(\sqrt{|k|}u_0) \\ b &= \sqrt{1/|k|} \cosh^{-1}(k\langle \mathbf{o}, \mathbf{v} \rangle_{\mathcal{L}}) = \sqrt{1/|k|} \cosh^{-1}(\sqrt{|k|}v_0) \\ c &= \sqrt{1/|k|} \cosh^{-1}(k\langle \mathbf{u}, \mathbf{v} \rangle_{\mathcal{L}}) \\ &= \sqrt{1/|k|} \cosh^{-1}(-k(u_0v_0 - \sum_{i=1}^n u_i v_i)) \end{aligned} \quad (15)$$

hence,

$$\begin{aligned} \cosh(a\sqrt{|k|}) &= \cosh(\sqrt{|k|} \cdot \sqrt{1/|k|} \cosh^{-1}(\sqrt{|k|}u_0)) \\ &= \sqrt{|k|}u_0 \\ \cosh(b\sqrt{|k|}) &= \cosh(\sqrt{|k|} \cdot \sqrt{1/|k|} \cosh^{-1}(\sqrt{|k|}v_0)) \\ &= \sqrt{|k|}v_0 \\ \cosh(c\sqrt{|k|}) &= -k(u_0v_0 - \sum_{i=1}^n u_i v_i) \\ \sinh(a\sqrt{|k|}) &= \sinh(\sqrt{|k|} \cdot \sqrt{1/|k|} \cosh^{-1}(\sqrt{|k|}u_0)) \\ &= \sinh(\cosh^{-1}(\sqrt{|k|}u_0)) \\ &= \sqrt{|k|u_0^2 - 1} \\ \sinh(b\sqrt{|k|}) &= \sinh(\sqrt{|k|} \cdot \sqrt{1/|k|} \cosh^{-1}(\sqrt{|k|}v_0)) \\ &= \sinh(\cosh^{-1}(\sqrt{|k|}v_0)) \\ &= \sqrt{|k|v_0^2 - 1} \end{aligned} \quad (16)$$

As φ corresponds to $\angle C$ in Eq. 14, we substitute Eq. 16 into Eq. 14, the Lorentzian cosine similarity

between \mathbf{u} and \mathbf{v} is calculated as:

$$\begin{aligned} \text{sim}(\mathbf{u}, \mathbf{v}) &= \cos \varphi \\ &= \frac{|k|u_0v_0 - (-k(u_0v_0 - \sum_{i=1}^n u_i v_i))}{\sinh(\cosh^{-1}(\sqrt{|k|}u_0)) \sinh(\cosh^{-1}(\sqrt{|k|}v_0))} \\ &= \frac{-k \sum_{i=1}^n u_i v_i}{\sqrt{|k|u_0^2 - 1} \sqrt{|k|v_0^2 - 1}} \end{aligned} \quad (17)$$

Thus, Theorem 3.1 is derived.

C Proof of Lemma 3.1

Lemma. $\forall \mathbf{u} \in \mathcal{L}_k^n \setminus \{\mathbf{o}\}$, let u_0 denotes the 0-th dimension of \mathbf{u} , $C > 0$ is a constant used to set the boundary conditions, the entailment cone of \mathbf{u} is defined by the half-aperture:

$$\gamma(\mathbf{u}) = \sin^{-1}\left(\frac{C}{\sqrt{|k|u_0^2 - 1}}\right)$$

Proof. Ganea et al. (2018a) propose the entailment cones in the Poincaré ball and provide a closed-form expression. Subsequently, Le et al. (2019) generalize it to the Lorentz model via the isometric property. In this section, we begin the derivation of the half-aperture of the Lorentzian entailment cone of arbitrary curvature from the convex cone.

The Lorentzian entailment cone is the generalization of the convex cone in the Lorentz model. The convex cone S is a closed set under non-negative linear combinations. For $\mathbf{v}_1, \mathbf{v}_2 \in S, \forall \alpha, \beta \geq 0$, the following formula is established:

$$\alpha \mathbf{v}_1 + \beta \mathbf{v}_2 \in S \quad (18)$$

Assume S is an arbitrary cone in the tangent space at point \mathbf{x} , i.e. $S \subseteq \mathcal{T}_{\mathbf{x}}\mathcal{L}_k^n$. Utilizing the exponential map, it can be mapped to the Lorentz model to obtain the Lorentzian entailment cone:

$$\mathfrak{S}_{\mathbf{x}} := \exp_{\mathbf{x}}^k(S) \quad (19)$$

Similar to Ganea et al. (2018a), we aspire for the Lorentzian entailment cones to embody four properties, i.e. axial symmetry, rotation invariance, continuous cone aperture functions, and transitivity of nested angular cones.

Axial symmetry. The Lorentzian entailment cone of point \mathbf{x} has a non-negative aperture $\gamma(\mathbf{x})$ as:

$$\begin{aligned} S_{\mathbf{x}}^{\gamma(\mathbf{x})} &:= \{\mathbf{v} \in \mathcal{T}_{\mathbf{x}}\mathcal{L}_k^n | \angle(\mathbf{v}, \bar{\mathbf{x}}) \leq \gamma(\mathbf{x})\} \\ \mathfrak{S}_{\mathbf{x}}^{\gamma(\mathbf{x})} &:= \exp_{\mathbf{x}}^k(S_{\mathbf{x}}^{\gamma(\mathbf{x})}) \end{aligned} \quad (20)$$

where $\bar{x} \in \mathcal{T}_x \mathcal{L}_k^n$ has the same direction as x . Further, the conic border is defined as:

$$\begin{aligned} \partial S_x^{\gamma(x)} &:= \{v \in \mathcal{T}_x \mathcal{L}_k^n \mid \angle(v, \bar{x}) = \gamma(x)\} \\ \partial \mathfrak{S}_x^{\gamma(x)} &:= \exp_x^k(\partial S_x^{\gamma(x)}) \end{aligned} \quad (21)$$

Rotation invariance. The Lorentzian entailment cone $\mathfrak{S}_x^{\gamma(x)}$ is contingent solely upon the geodesic distance between x and the origin $\mathfrak{o}(1/\sqrt{|k|}, \mathbf{0}^n)$, and entirely independent of the angular coordinate of the apex x . $\forall x, x' \in \mathcal{L}_k^n \setminus \{\mathfrak{o}\}$, if $d_{\mathcal{L}}^k(\mathfrak{o}, x) = d_{\mathcal{L}}^k(\mathfrak{o}, x')$, the following equation holds:

$$\gamma(x) = \gamma(x') \quad (22)$$

i.e., there exists a function $\tilde{\gamma} : (0, +\infty) \rightarrow [0, \pi)$, that is, $\forall x \in \mathcal{L}_k^n \setminus \{\mathfrak{o}\}$, $\gamma(x) = \tilde{\gamma}(d_{\mathcal{L}}^k(\mathfrak{o}, x))$.

Continuous cone aperture functions. Eq. 22 illustrates the continuity of $\tilde{\gamma}$, indicating that γ is continuous.

Transitivity of nested angular cones. Transitivity accounts for the partial order within the embedding space. For a nested structure $\forall x, x' \in \mathcal{L}_k^n \setminus \{\mathfrak{o}\}$, $x' \in \mathfrak{S}_x^{\gamma(x)}$, the transitivity is formulated as:

$$\mathfrak{S}_{x'}^{\gamma(x')} \subseteq \mathfrak{S}_x^{\gamma(x)} \quad (23)$$

Using the transitivity, Ganea et al. (2018a) prove the fact:

$$\forall x \in \text{Dom}(\gamma), \gamma(x) \leq \frac{\pi}{2} \quad (24)$$

With the properties mentioned above, we first introduce the hyperbolic law of sines (Parkkonen, 2012).

Hyperbolic Law of Sines. In the hyperbolic triangle $\triangle ABC$ constructed when the hyperbolic law of cosines is introduced, the hyperbolic law of sines is established as follows:

$$\frac{\sin(\angle A)}{\sinh(a\sqrt{|k|})} = \frac{\sin(\angle B)}{\sinh(b\sqrt{|k|})} = \frac{\sin(\angle C)}{\sinh(c\sqrt{|k|})} \quad (25)$$

Then, we prove the following fact:

Lemma C.1. $\forall x \in \mathcal{L}_k^n \setminus \{\mathfrak{o}\}$, and $\forall x' \in \partial \mathfrak{S}_x^{\gamma(x)}$,

$$\sin(\gamma(x')) \sinh(d_{\mathcal{L}}^k(\mathfrak{o}, x')) \leq \sin(\gamma(x)) \sinh(d_{\mathcal{L}}^k(\mathfrak{o}, x)) \quad (26)$$

Proof. As shown in Fig. 9, $x' \in \partial \mathfrak{S}_x^{\gamma(x)}$, therefore,

$$\angle yx'z \leq \gamma(x) \leq \frac{\pi}{2} \quad (27)$$

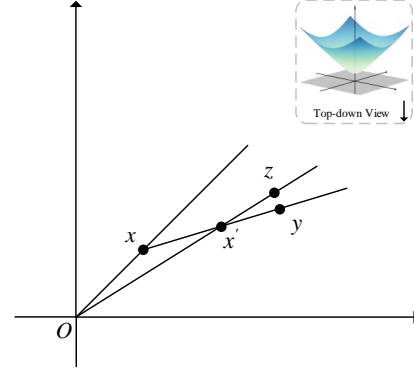


Figure 9: The coordinate system is a top-down view of the Lorentz model. $x' \in \partial \mathfrak{S}_x^{\gamma(x)}$ is any arbitrary point on the conic border of $\mathfrak{S}_x^{\gamma(x)}$, $y \in \partial \mathfrak{S}_x^{\gamma(x)}$ is any arbitrary point on the geodesic half-line (xx') starting from x' . z is any arbitrary point on the geodesic half-line (Ox') starting from x' , therefore, (Oz) is the axis of symmetry of $\mathfrak{S}_{x'}^{\gamma(x')}$.

With the transitivity, we have:

$$\angle yx'z \geq \gamma(x') \quad (28)$$

In addition, in hyperbolic triangle $\triangle Ox'x$,

$$\begin{aligned} \angle Ox'x &= \angle yx'z \\ \angle Oxx' &= \pi - \gamma(x) \end{aligned} \quad (29)$$

With the hyperbolic law of sines in Eq. 25, we have:

$$\frac{\sin(\angle Oxx')}{\sinh(\sqrt{|k|}d_{\mathcal{L}}^k(O, x'))} = \frac{\sin(\angle Ox'x)}{\sinh(\sqrt{|k|}d_{\mathcal{L}}^k(O, x))} \quad (30)$$

Utilizing the monotonically increasing property of $\sin(\cdot)$ over interval $[0, \frac{\pi}{2}]$, in conjunction with Eqs. 27, 28, 29, 30, Lemma C.1 is proved. \square

With Lemma C.1, we prove the following theorem:

Theorem C.1. *If transitivity holds, $k(k < 0)$ is the constant negative curvature of the Lorentz model, the function*

$$\begin{aligned} h &: (0, +\infty) \cap \text{Dom}(\tilde{\gamma}) \rightarrow \mathbb{R}_+ \\ h(r) &= \sinh(r\sqrt{|k|})\sin(\tilde{\gamma}(r)) \end{aligned} \quad (31)$$

is non-increasing.

Proof. As $d_{\mathcal{L}}^k(O, x) = \sqrt{1/|k|}\cosh^{-1}(\sqrt{|k|x_0})$ (described in Eq. 15), $d_{\mathcal{L}}^k(O, x)$ is a continuously monotonically increasing function related to the

0-th dimensional coordinate of x . Therefore, $\exists x \in \mathcal{L}_k^n \setminus \{\mathbf{o}\}, x' \in \mathfrak{S}_x^{\gamma(x)}$, s.t.

$$\begin{aligned} d_{\mathcal{L}}^k(O, x) &= r, & d_{\mathcal{L}}^k(O, x') &= r' \\ r < r' & \text{ and } & r, r' &\in (0, +\infty) \end{aligned} \quad (32)$$

Consequently, we simplify $\sinh(\sqrt{|k|}d_{\mathcal{L}}^k(O, x))$ as

$$\sinh(\sqrt{|k|}d_{\mathcal{L}}^k(O, x)) = \sinh(r\sqrt{|k|}) \quad (33)$$

and apply it to Lemma C.1, we have:

$$h(r) \geq h(r') \quad (34)$$

Combining Eqs. 32, 34, Theorem C.1 is proved. \square

Since $\lim_{r \rightarrow 0} h(r) = 0$ for any function $\tilde{\gamma}$, $\tilde{\gamma}$ can not be defined on the entire $(0, +\infty)$. Therefore, we restrict $Dom(\tilde{\gamma})$ to some $[\epsilon, +\infty)$, then Theorem C.1 implies that $\forall r \in [\epsilon, +\infty)$,

$$\sinh(r\sqrt{|k|})\sin(\tilde{\gamma}(r)) \leq \sinh(\epsilon\sqrt{|k|})\sin(\tilde{\gamma}(\epsilon)) \quad (35)$$

To make h constant, let $C > 0$ to be a constant, we set $h(r)$ equal to C :

$$\forall r \in [\epsilon, +\infty), \quad \sinh(r\sqrt{|k|})\sin(\tilde{\gamma}(r)) = C \quad (36)$$

which implies that:

$$C \leq \sinh(\epsilon\sqrt{|k|}), \quad \epsilon \geq \sinh^{-1}\left(\frac{C}{\sqrt{|k|}}\right) \quad (37)$$

also gives the definition of the half-aperture of the Lorentzian entailment cone of $\mathbf{u} \in \mathcal{L}_k^n \setminus \{\mathbf{o}\}$:

$$\begin{aligned} \gamma(\mathbf{u}) &= \sin^{-1}\left(\frac{C}{\sinh(\sqrt{|k|}d_{\mathcal{L}}^k(\mathbf{o}, \mathbf{u}))}\right) \\ &= \sin^{-1}\left(\frac{C}{\sinh(\sqrt{|k|} \cdot \frac{1}{\sqrt{|k|}} \cosh^{-1}(\sqrt{|k|}u_0))}\right) \\ &= \sin^{-1}\left(\frac{C}{\sqrt{|k|}u_0^2 - 1}\right) \end{aligned} \quad (38)$$

Thus, Lemma 3.1 is proved. \square

D Proof of Lemma 3.2

Lemma. $\forall \mathbf{u}, \mathbf{v} \in \mathcal{L}_k^n \setminus \{\mathbf{o}\}$, let u_i and v_i denote the i -th dimension of \mathbf{u} and \mathbf{v} , the exterior angle between half-lines $(\mathbf{o}\mathbf{u})$ and $(\mathbf{u}\mathbf{v})$ is calculated as:

$$\phi(\mathbf{u}, \mathbf{v}) = \cos^{-1}\left(\frac{v_0 - u_0 \cdot k\langle \mathbf{u}, \mathbf{v} \rangle_{\mathcal{L}}}{\sqrt{\sum_{i=1}^n u_i^2 \sqrt{(k\langle \mathbf{u}, \mathbf{v} \rangle_{\mathcal{L}})^2 - 1}}}\right)$$

Proof. We employ the hyperbolic triangle constructed in appendix B to derive Lemma 3.2. As illustrated in Fig. 8, the exterior angle between half-lines $(\mathbf{o}\mathbf{u})$ and $(\mathbf{u}\mathbf{v})$ is formulated as follows:

$$\begin{aligned} \phi(\mathbf{u}, \mathbf{v}) &= \pi - \angle \mathbf{o}\mathbf{u}\mathbf{v} \\ &= \pi - \cos^{-1}\left(\frac{\cosh(a\sqrt{|k|}) \cosh(c\sqrt{|k|}) - \cosh(b\sqrt{|k|})}{\sinh(a\sqrt{|k|}) \sinh(c\sqrt{|k|})}\right) \end{aligned} \quad (39)$$

Since $\pi - \cos^{-1}(x) = \cos^{-1}(-x)$, Eq. 39 can be simplified as:

$$\begin{aligned} \phi(\mathbf{u}, \mathbf{v}) &= \cos^{-1}\left(\frac{\cosh(b\sqrt{|k|}) - \cosh(a\sqrt{|k|}) \cosh(c\sqrt{|k|})}{\sinh(a\sqrt{|k|}) \sinh(c\sqrt{|k|})}\right) \end{aligned} \quad (40)$$

Substituting Eqs. 15, 16 into Eq. 40, we give the following derivation process:

$$\begin{aligned} \phi(\mathbf{u}, \mathbf{v}) &= \cos^{-1}\left(\frac{\sqrt{|k|}v_0 - k\langle \mathbf{u}, \mathbf{v} \rangle_{\mathcal{L}} \cdot \sqrt{|k|}u_0}{\sinh(\cosh^{-1}(\sqrt{|k|}u_0)) \sinh(\cosh^{-1}(k\langle \mathbf{u}, \mathbf{v} \rangle_{\mathcal{L}}))}\right) \\ &= \cos^{-1}\left(\frac{\sqrt{|k|}(v_0 - u_0 \cdot k\langle \mathbf{u}, \mathbf{v} \rangle_{\mathcal{L}})}{\sqrt{|k|}u_0^2 - 1 \sqrt{(k\langle \mathbf{u}, \mathbf{v} \rangle_{\mathcal{L}})^2 - 1}}\right) \end{aligned} \quad (41)$$

Leveraging the definition of the Lorentz model, we have:

$$-u_0^2 + \sum_{i=1}^n u_i^2 = \frac{1}{k} = -\frac{1}{|k|} \quad (42)$$

therefore,

$$\sqrt{|k|}u_0^2 - 1 = \sqrt{|k| \sum_{i=1}^n u_i^2} \quad (43)$$

Substituting Eq. 43 into Eq. 41:

$$\begin{aligned} \phi(\mathbf{u}, \mathbf{v}) &= \cos^{-1}\left(\frac{\sqrt{|k|}(v_0 - u_0 \cdot k\langle \mathbf{u}, \mathbf{v} \rangle_{\mathcal{L}})}{\sqrt{|k| \sum_{i=1}^n u_i^2 \sqrt{(k\langle \mathbf{u}, \mathbf{v} \rangle_{\mathcal{L}})^2 - 1}}}\right) \\ &= \cos^{-1}\left(\frac{v_0 - u_0 \cdot k\langle \mathbf{u}, \mathbf{v} \rangle_{\mathcal{L}}}{\sqrt{\sum_{i=1}^n u_i^2 \sqrt{(k\langle \mathbf{u}, \mathbf{v} \rangle_{\mathcal{L}})^2 - 1}}}\right) \end{aligned} \quad (44)$$

Thus, Lemma 3.2 is proved. \square

E Introduction of Datasets

FewRel 1.0. FewRel 1.0 comprises 100 relations with 70,000 instances extracted from Wikipedia, along with a name and description for each relation

Hyper-parameter	Range
Temperature τ	[0.01, 0.02]
Constant C	[0.2, 0.3, 0.4]
Margin m	[0.1, 0.2]
Weight coefficient λ_1	[0.2, 0.5, 1]
Weight coefficient λ_2	[0.02, 0.05, 0.1, 0.2, 0.5, 1]

Table 7: Searching ranges of hyper-parameters.

to increase interpretability. Relations are divided into 10 groups covering verbal, spatial, comparative relations, etc. The training and test sets cover the same Wikipedia domains.

FewRel 2.0. The training set of FewRel 2.0 is identical to FewRel 1.0, while validation and test sets are drawn from the biomedical field. Compared to FewRel 1.0, FewRel 2.0 is more challenging since only the names of relations are available.

Semeval Semeval consists of 19 relations with 10717 instances, of which the training set, validation set, and test set contain 6507, 1693, and 2717 instances, respectively.

Please be advised that in our experimental setup, we utilize the training set of FewRel 1.0 for model training. Additionally, the validation and test sets from the three respective datasets were employed for model validation and testing respectively.

F Implementation Details

We adopt the uncased model of BERT-base and CP (Peng et al., 2020) for sentence encoding, with respective learning rates of $1e-5$ and $5e-6$. The training and validation iterations are configured to 30,000 and 1,000, respectively. We utilize the AdamW optimizer to minimize the loss function, with a batch size of 2 for the 10-way-5-shot setting and 4 for other configurations. In the 1-shot settings, the weight coefficient α is assigned a value of 1, and in the 5-shot settings, it was set to 0. Random search is employed to determine the optimal values for remaining hyper-parameters, and the ranges are detailed in Table 7.

G Compared Baselines

We compare our proposed PO-HRL with the following baselines. Note that, **REGRAB**, **ConceptFERE** and **DCFT** are based on external knowledge, while others only leverage the provided texts and relation descriptions.

Proto-BERT (Snell et al., 2017): a prototypical network for few-shot learning.

BERT-PAIR (Gao et al., 2019): a metric-based method that estimates the similarity of query-support pairs, followed by a classifier to predict the label of each query instance.

REGRAB (Qu et al., 2020): a Bayesian meta-learning framework with an external global relation graph to investigate connections across discrete relations.

ConceptFERE (Yang et al., 2021): an attention-based approach introducing the inherent concepts of entities to provide clues for relation prediction.

HCRP (Han et al., 2021a): a hybrid contrastive relation-prototype model with task adaptive focal loss directed at advancements in extracting hard relations.

SimpleFSRE (Liu et al., 2022): a simple yet effective prototype-based method that directly adds embeddings of relation descriptions to prototypical representations.

FAEA (Dou et al., 2022): a function word augmented attention mechanism targeting inverse relation extraction via improvements to class-specific function word embeddings.

GM_GEN (Li and Qian, 2022): a graph-based model with a task-specific generation module to produce specialized models for individual FSRE tasks.

BMIPN (Li et al., 2023): a method based on bi-ased contrastive learning that incorporates explicit and adaptive interactions from both intra-class and inter-class perspectives.

HND (Zhang et al., 2023): a network generator-based approach that generates classifiers specialized in capturing relation-specific knowledge.

DCFT (Zhang et al., 2023): an approach to domain adaptation for FSRE introduces additional unlabeled data from the target domain to facilitate domain-aware transformation. Since DCFT employs additional unlabeled data, the comparison of PO-HRL is limited to DCFT-DTM without the additional unlabeled data.

MTB (Soares et al., 2019): a post-training task with matching the blanks for RE.

CP (Peng et al., 2020): an entity-masked contrastive post-training framework for RE.