TaxoPro: A Plug-In LoRA-based Cross-Domain Method for Low-Resource Taxonomy Completion

Hongyuan Xu, Yuhang Niu, Ciyi Liu, Yanlong Wen*, and **Xiaojie Yuan**
TMCC & TBI Center, College of Computer Science, Nankai University, Tianjin, China
{xuhongyuan, niuyuhang, liuciyi}@dbis.nankai.edu.cn
{wenyl, yuanxj}@nankai.edu.cn

Abstract

Low-resource taxonomy completion aims to automatically insert new concepts into the existing taxonomy, in which only a few in-domain training samples are available. Recent studies have achieved considerable progress by incorporating prior knowledge from pre-trained language models (PLMs). However, these studies tend to overly rely on such knowledge and neglect the shareable knowledge across different taxonomies. In this paper, we propose TaxoPro, a plug-in LoRA-based cross-domain method, that captures shareable knowledge from the high-resource taxonomy to improve PLM-based low-resource taxonomy completion techniques. To prevent negative interference between domain-specific and domain-shared knowledge, TaxoPro decomposes cross-domain knowledge into domain-shared and domain-specific components, storing them using low-rank matrices (LoRA). Additionally, TaxoPro employs two auxiliary losses to regulate the flow of shareable knowledge. Experimental results demonstrate that TaxoPro improves PLM-based techniques, achieving state-of-the-art performance in completing low-resource taxonomies. Code is available at `https://github.com/cyclexu/TaxoPro`.

1 Introduction

Taxonomies are knowledge structures that hierarchically organize concepts through hypernym-hyponym (''*is-a*'') relations. They find extensive applications in fields such as natural language processing (Bai et al., 2022; Hu et al., 2022b), recommendation systems (Cheng et al., 2022), and information retrieval (Karamanolakis et al., 2020).

Most current taxonomies are manually curated by domain experts, which is both time-consuming and labour-intensive. With the constant emergence of new concepts, keeping taxonomies up-to-date for downstream applications has become a critical challenge (Shen et al., 2020; Zhang et al., 2021). To solve this problem, significant effort has been dedicated to the *taxonomy expansion task* (Shen et al., 2020; Liu et al., 2021; Xue et al., 2024). In this task, the existing taxonomy is expanded by inserting the new concept (*query*) to the most appropriate hypernym (*parent*) within the existing taxonomy as a leaf node. However, recent researchers contend that the ''leaf-only'' assumption is unsuitable (Zhang et al., 2021), leading to significant limitations in real-world scenarios (Wang et al., 2022a). Thus, they turn to the *taxonomy completion task* (Zhang et al., 2021; Xu et al., 2023; Niu et al., 2024), where the query is inserted between a pair of hypernym and hyponym (*child*). For example, the query ''wearable device'' is inserted between the parent ''electronic equipment'' and the child ''AR glass'' as shown in Figure 1.

In practical scenarios, the low-resource setting, where only a limited number of concepts exist in the existing taxonomy, is prevalent as most taxonomies typically comprise around a thousand concepts (Takeoka et al., 2021). Under such a setting, the early taxonomy expansion and completion methods (Shen et al., 2020; Zhang et al., 2021; Manzoor et al., 2020) suffer from performance degradation due to insufficient training samples (Takeoka et al., 2021). Several studies (Liu et al., 2021; Takeoka et al., 2021; Xu et al., 2023) have shown that this can be mitigated by incorporating prior knowledge from the pre-trained language model (PLM). However, such knowledge could be generic and irrelevant to the taxonomy tasks (Gururangan et al., 2020; Diao et al., 2023), thus limiting the performance of these PLM-based techniques. Meanwhile, taxonomies across different domains store the same type of knowledge, i.e., hierarchical relations between concepts. Consequently, the high-resource taxonomy can serve as an extra knowledge base for completing the low-resource taxonomy. In this

* Corresponding author.

Transactions of the Association for Computational Linguistics, vol. 13, pp. 557–576, 2025. https://doi.org/10.1162/tacl_a_00755
Action Editor: Tao Ge. Submission batch: 9/2024; Revision batch: 2/2025; Published 6/2025.
© 2025 Association for Computational Linguistics. Distributed under a CC-BY 4.0 license.

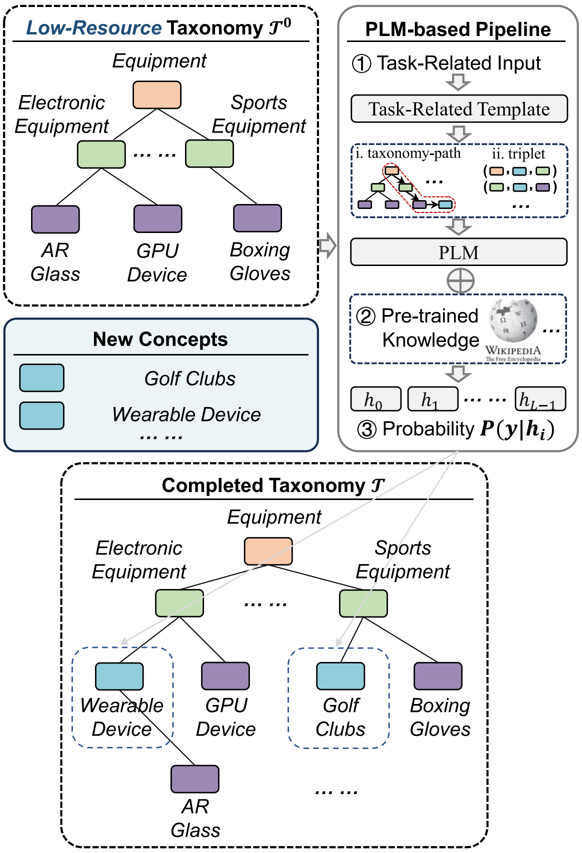


Figure 1: An example illustrating how pre-trained language models (PLMs) can complete the low-resource "Equipment" taxonomy by inserting new concepts into existing structures.

paper, we explore the research question of *how to capture knowledge from the high-resource taxonomy for the PLM-based low-resource taxonomy completion enhancement?*

Inspired by the recent studies (Diao et al., 2023; Zhang et al., 2023b; Wang et al., 2023c) that utilize the parameter-efficient fine-tuning (PEFT) techniques (Houlsby et al., 2019; Li and Liang, 2021) for knowledge storage and transfer, we utilize LoRA (Hu et al., 2022a), a widely used PEFT technique that leverages low-rank matrices as the knowledge update, to capture domain-shared knowledge from a high-resource source taxonomy and apply it to complete a low-resource target taxonomy. Specifically, we propose a LoRA-based cross-domain method that can be plugged into the PLM-based taxonomy completion techniques. Our method has two main modules: (i) knowledge decomposition and (ii) shareable knowledge flow control. In the first module, we decompose the knowledge of each taxonomy into domain-shared and domain-specific components to prevent negative interference between these two types of knowledge (Du et al., 2023). We store these components using separate low-rank matrices, which are updated through task-specific losses across various domains. In the second module, we employ two auxiliary losses to guide the flow of shareable knowledge. Specifically, we pull shareable knowledge into the domain-shared matrices and push it out from the domain-specific matrices. Our objective is to maximize the extraction of shareable knowledge from the high-resource taxonomy, thereby improving the completion of the low-resource taxonomy.

We plug the proposed method into two representative PLM-based taxonomy completion techniques and conduct extensive experiments on three low-resource taxonomy datasets. Experimental results show that our method improves the PLM-based techniques and achieves state-of-the-art performance in low-resource taxonomy completion.

In summary, our contributions include:

- We propose TaxoPro, a LoRA-based cross-domain framework that can be plugged into the PLM-based taxonomy completion techniques. To the best of our knowledge, it is the first work that captures shareable knowledge from the high-resource taxonomy to enhance low-resource taxonomy completion.
- We leverage the knowledge decomposition to prevent negative interference between domain-specific and domain-shared knowledge and employ two auxiliary losses to regulate the flow of shareable knowledge.
- We conduct extensive experiments to validate the effectiveness of the proposed method. Experimental results demonstrate that TaxoPro enhances PLM-based techniques and achieves state-of-the-art performance in completing low-resource taxonomies.

2 Related Work

2.1 Taxonomy Expansion and Completion

With regard to automatic taxonomy enrichment, there exist two lines of research: Taxonomy Expansion and Completion. To expand the existing taxonomy, researchers (Shen et al., 2020; Yu et al., 2020; Ma et al., 2021; Wang et al., 2021; Liu et al., 2021; Takeoka et al., 2021; Cheng et al., 2022; Phukon et al., 2022; Xu et al., 2022; Zhai et al., 2023; Jiang et al., 2023; Sun et al., 2024; Zhu et al., 2023; Mishra et al., 2024; Xu et al., 2024; Shen et al., 2024; Zeng et al., 2024; Qingkai et al., 2024; Meng et al., 2024; Moskvoretskii et al., 2024) attempted to insert the emergent concepts to the most appropriate leaf position.

In taxonomy completion, Zhang et al. (2021) extended the candidate insertion position from ''leaf-only'' to a pair of parent and child nodes. GenTaxo (Zeng et al., 2021) completed the taxonomy in a concept generation manner. TAXBOX (Xue et al., 2024) enhanced taxonomy completion by using specialized geometric scorers in box embedding space. TaxoEnrich (Jiang et al., 2022) and QEN (Wang et al., 2022a) incorporated sibling relations for semantic-rich concept representation. TaxoComplete (Arous et al., 2023) captured fine-grained information from distant nodes. CoSTC (Niu et al., 2024) captured diverse relations and improved representations through intra-view and inter-view contrastive learning. TEMP (Liu et al., 2021) and TacoPrompt (Xu et al., 2023) leveraged pre-trained language models as an implicit knowledge base and achieved remarkable performance. Additionally, researchers explored variant taxonomy completion settings. For instance, ATTEMPT (Xia et al., 2023) suggested initially identifying the parent and then locating all its children within the taxonomy. ICON (Shi et al., 2024) focused on generating new concepts based on the taxonomy's structure and existing concepts, which are then inserted into the taxonomy. These settings are beyond the scope of this paper.

In this paper, we explore the low-resource taxonomy completion scenario where little in-domain labelled data is available. Unlike Musubu (Takeoka et al., 2021), which solely relies on pre-trained knowledge, our focus lies in capturing shareable knowledge from the high-resource taxonomy for enhancing the completion of the low-resource taxonomy.

2.2 Parameter-Efficient Fine-Tuning for Knowledge Decomposition

One of the recently popular techniques in knowledge transfer involves decomposing input data into domain-specific and domain-shared knowledge (Sarafraz et al., 2024; Wang et al., 2023a,b; Wei et al., 2023; Ben-David et al., 2020). By setting distinct objectives for each, domain-specific information can be separated, enabling the use of domain-shared knowledge for predictions in new domains (Zhang and Gao, 2024). Early work, such as Daumé (2007), proposed expanding the feature space into common, source-specific, and target-specific components. Building on this, Bousmalis et al. (2016) introduced Domain Separation Networks (DSN), which utilize separate encoders to explicitly model domain-shared and domain-specific knowledge, hypothesizing that this separation enhances the extraction of transferable knowledge.

Lately, parameter-efficient fine-tuning (PEFT) methods, which adjust only a subset of model parameters (Li and Liang, 2021; Zhang et al., 2023a,c), have gained traction in NLP for adapting pre-trained language models to downstream tasks (Chen et al., 2022; Dettmers et al., 2023). These methods facilitate knowledge transfer and composition, supporting the integration of diverse knowledge sources (Ding et al., 2023; Wang et al., 2022b; Mao et al., 2022). Researchers have also begun exploring PEFT for knowledge decomposition. For instance, Zhang et al. (2023b) proposed a framework for Event Argument Extraction across datasets using Prompt Tuning and Adapters to manage overlapping and specific knowledge in sequential learning phases. Similarly, Wang et al. (2023c) decomposed knowledge across tasks into shared and task-specific prompt vectors, with shared vectors learned from multiple tasks for efficient adaptation to new tasks.

In our work, we advance this line of research by developing an end-to-end LoRA-based knowledge decomposition approach tailored for taxonomies across domains. Our focus is on (i) disentangling noisy domain-specific knowledge from domain-shared knowledge and (ii) regulating shareable knowledge flow through auxiliary loss functions.

3 Preliminaries

3.1 Problem Formulation

In this section, we provide a formal definition of the taxonomy and the taxonomy completion task.

Taxonomy. Building upon the formalism established by Shen et al. (2020), we formalize a taxonomy as a directed acyclic graph $\mathcal{T} = (\mathcal{N}, \mathcal{E})$, where nodes $\mathcal{N}$ correspond to concepts and edges $\mathcal{E}$ encode hypernym-hyponym relations through ordered pairs $\langle p, c\rangle$. This graph structure ensures that each parent concept p maintains maximal specificity while remaining semantically broader than its child concept c. Following Xu et al. (2023), a corpus $\mathcal{D}$ is provided, from which the concept descriptions are extracted using established retrieval methods.

Problem Definition. The taxonomy completion task operates on two primary components: (i) an existing taxonomic structure $\mathcal{T}^0 = \left(\mathcal{N}^0, \mathcal{E}^0\right)$ and (2) a collection of new concepts $\mathcal{C}$. The task aims to augment $\mathcal{T}^0$ by optimally placing each new concept $q \in \mathcal{C}$ at its conly appropriate position a. Following Zhang et al.'s (2021) formalization, valid insertion positions constitute ordered pairs $a = \langle p, c\rangle$, where $p \in \mathcal{N}^0$ serves as an ancestral node to descendant c in the original structure. Successful insertion of concept q triggers structural reorganization: The original $\langle p, c\rangle$ edge is replaced with hierarchical links $\langle p, q\rangle$ and $\langle q, c\rangle$, effectively inserting q between p and c. Consistent with the framework of Shen et al. (2020) and Zhang et al. (2021), we adopt the independence assumption among C elements, enabling decomposition of the global task into $|C|$ independent optimization subproblems (Xu et al., 2023):

$$a_i^* = \underset{a_i \in \mathcal{A}}{\arg\max} \log P\left(q_i \mid a_i, \Theta\right), \quad (1)$$

where $\forall i \in \{1, 2, \ldots, |C|\}$, Θ and $\mathcal{A}$ denote model parameters and the set of candidate positions, respectively.

In this paper, we focus on the taxonomy completion task in a low-resource setting, where $\mathcal{T}^0$ comprises only a limited number of training samples. We incorporate an external input, specifically a high-resource taxonomy, to supply additional training samples for completing the low-resource taxonomy. Our goal is to capture cross-domain shareable knowledge from the high-resource taxonomy (source domain) to enhance low-resource taxonomy (target domain) completion.

3.2 PLM-based Taxonomy Completion

Pre-trained language models have been superior techniques in expanding (Liu et al., 2021; Wang et al., 2021; Sun et al., 2024) or completing (Xu et al., 2023) taxonomies in a cross-encoder manner. Let $\mathcal{M}$ represent the PLM and $\mathcal{F}$ the template function. Given the task input $x = (q, a, \mathcal{T}, \mathcal{D})$, where $\mathcal{D}$ denotes the corpus where concept descriptions are extracted, the core idea is to convert it to the natural language form using the template function $\mathcal{F}(x) = (z_0, z_1, \ldots, z_{l-1})$ and input it to the $\mathcal{M}$ to perform self-attention encoding:

$$\mathcal{M}(\mathcal{F}(x)) = h_0, h_1, \ldots, h_{l-1}, \quad (2)$$

where h_i represents the i-th token's hidden vector output by the $\mathcal{M}$'s last layer. Then, the hidden vector of the special token, e.g., `[CLS]` or `[MASK]`, will be leveraged to represent the task-specific information. Lastly, a classification head g is utilized to decode the probability distribution corresponding to the task label y:

$$P(y \mid x) = g(h_{\text{special}}), \quad (3)$$

where $y = 1$ if a is a ground-truth position for the query q, and $y = 0$ otherwise. In this way, the insertion probability in Equation 1 is calculated using the above PLM-based pipeline.

4 Methodology

In this section, we propose a cross-domain method that can be plugged into the PLM-based taxonomy completion techniques (§3.2). The method comprises a knowledge decomposition module (§4.1) and a shareable knowledge flow control module (§4.2), as illustrated in Figure 2.

4.1 Cross-Domain Knowledge Decomposition

Taxonomies across domains embody two types of knowledge: (i) domain-shared knowledge, as they all serve as repositories for hierarchical relations between concepts, and (ii) domain-specific knowledge, characterized by their unique semantic distributions and structural granularity. To effectively capture this, we decompose the knowledge from cross-domain taxonomies into domain-shared and domain-specific components. This decomposition achieves two objectives: (i) enhancing the performance on the low-resource domain by leveraging shareable knowledge from the high-resource domain, and (ii) mitigating the negative interference between domain-specific and domain-shared knowledge (Du et al., 2023) by separating noisy domain-specific knowledge.

Parameter-efficient fine-tuning (PEFT) is a class of techniques (Houlsby et al., 2019; Li and Liang, 2021; Hu et al., 2022a) that aims at adapting PLMs to downstream tasks with few extra parameters. We leverage PEFT as the backbone

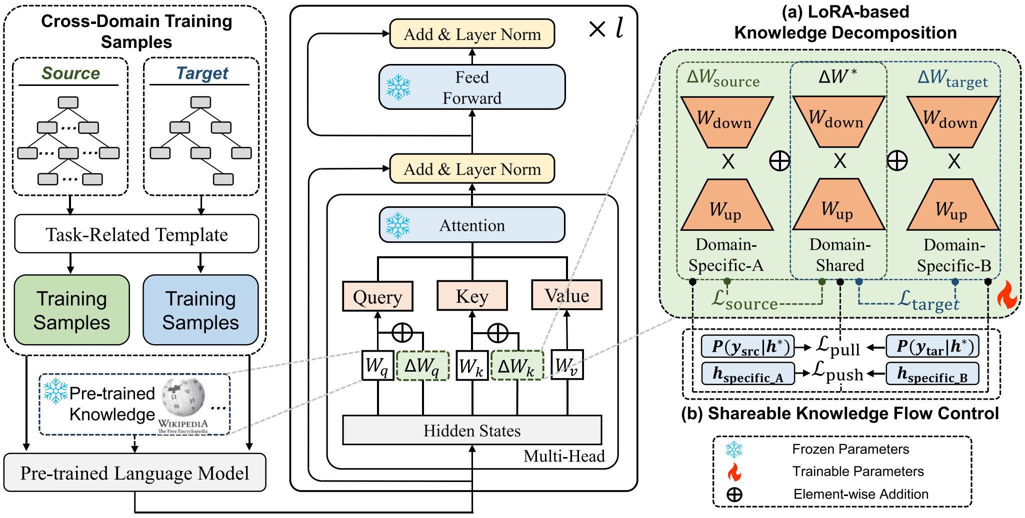


Figure 2: Illustration of the training pipeline of TaxoPro. $\mathcal{L}_{\text{source}}$ and $\mathcal{L}_{\text{target}}$ are the task-specific loss used by these techniques. We draw the parameter update corresponding to the loss with dotted lines.

technique of the cross-domain knowledge decomposition due to its effectiveness in knowledge storage (Zhang et al., 2023b) and combination (Pfeiffer et al., 2020; Wang et al., 2023c). Specifically, we utilize LoRA (Hu et al., 2022a), which is an effective and commonly used PEFT technique (He et al., 2022), in our method. Typically, LoRA represents the update of the pre-trained matrix $\mathbf{W} \in \mathbb{R}^{d\times k}$ with low-rank decomposition $\mathbf{W} + \Delta W = \mathbf{W} + \mathbf{BA}$, where $\mathbf{B} \in \mathbb{R}^{d\times r}$, $\mathbf{A} \in \mathbb{R}^{r\times k}$ are trainable low-rank matrices injected to the query and value projection matrices $(\mathbf{W}_q, \mathbf{W}_v)$ in the transformer layers (He et al., 2022). Thus, the forward output $\mathbf{h}$ given the specific input $\mathbf{x}$ yields:

$$\mathbf{h} \leftarrow \mathbf{Wx} + s \cdot \mathbf{BAx}, \tag{4}$$

where s is a scaling hyperparameter.

As depicted in Figure 2, we further decompose the updated knowledge ΔW into two components: (i) domain-shared knowledge across domains ΔW^*, and (ii) domain-specific knowledge ΔW_k. Here, the variable k can take values of 0 or 1, representing the source and target domain respectively. The forward output $\mathbf{h}$ for k-th domain is then formulated as:

$$\mathbf{h} \leftarrow \mathbf{Wx} + s \cdot (\mathbf{B}^*\mathbf{A}^* + \mathbf{B}_k\mathbf{A}_k)\mathbf{x}, \tag{5}$$

where $\mathbf{B}^*\mathbf{A}^*$ and $\mathbf{B}_k\mathbf{A}_k$ are low-rank matrices approximations for ΔW^* and ΔW_k, respectively.

Finally, we train the cross-domain knowledge decomposition (CDKD) module in the multi-task learning manner:

$$\mathcal{L}_{\text{CDKD}} = \mathcal{L}_{\text{source}}^{\mathcal{M}_{\text{KD}}} + \alpha \cdot \mathcal{L}_{\text{target}}^{\mathcal{M}_{\text{KD}}}, \tag{6}$$

where $\mathcal{M}_{\text{KD}}$ denotes the PLM injected with the CDKD module, and the hyperparameter α equilibrates the corresponding losses across different domains. The task-specific loss function for each domain $\mathcal{L}_{\text{domain}}^{\mathcal{M}_{\text{KD}}}$, e.g., BCELoss (Xu et al., 2023), is calculated after using $\mathcal{M}_{\text{KD}}$ to perform the pipeline described in Section 3.2 with training samples from the respective domain.

To facilitate cross-domain learning, we sample B instances from each domain per batch during training. The loss $\mathcal{L}_{\text{domain}}^{\mathcal{M}_{\text{KD}}}$ is then calculated as the average loss over all samples from the corresponding domain in a batch. Importantly, training samples for both source and target domain taxonomies are generated following the original sampling process of the plugged methods.

4.2 Shareable Knowledge Flow Control

As outlined in Section 4.1, the shared knowledge serves as a bridge between domains with low and high resources. The training objective $\mathcal{L}_{\text{CDKD}}$ regulates the flow of domain-specific knowledge from ΔW^* since it updates ΔW^* by training samples from both domains, thereby the noisy domain-specific knowledge will be separated by the task-specific loss of another domain. However, $\mathcal{L}_{\text{CDKD}}$ lacks an explicit constraint ensuring that shareable knowledge flows into ΔW^* rather than ΔW_k. To address this, our proposed method employs two auxiliary loss functions. The first one specifically pulls shareable knowledge from both domains into ΔW^*:

$$\mathcal{L}_{\text{pull}} = \mathcal{L}^{\mathcal{M}^*}_{\text{source}} + \alpha \cdot \mathcal{L}^{\mathcal{M}^*}_{\text{target}}, \tag{7}$$

which is the same as Equation 6 except that the PLM is only injected with the domain-shared matrices $\mathbf{B}^*\mathbf{A}^*$. The idea behind $\mathcal{L}_{\text{pull}}$ is straightforward: It assumes that knowledge across different domains is entirely shareable and leaves the domain-specific knowledge to be separated by $\mathcal{L}_{\text{CDKD}}$. The second auxiliary loss function aims to push shareable knowledge out of the ΔW_k:

$$\mathcal{L}_{\text{push}} = \frac{1}{B}\sum_{i=0}^{B} \max(0, \cos(h^{\mathcal{M}_{0,i}}_{\text{special}}, h^{\mathcal{M}_{1,i}}_{\text{special}})), \tag{8}$$

where B represents the number of training samples from the target domain in a batch and cos denotes the cosine similarity. $h^{\mathcal{M}_{k,i}}_{\text{special}}, k \in \{0, 1\}$ denotes the hidden vector utilized for the probability distribution decoding as described in Equation 3 for i-th target training sample in a batch. This vector is encoded by the PLM, which is only injected with the domain-specific matrices $\mathbf{B}_k\mathbf{A}_k$. The idea behind $\mathcal{L}_{\text{push}}$ is that the similar part in domain-specific knowledge of different domains could be the potentially shareable knowledge. Notably, $\mathcal{L}_{\text{push}}$ is calculated using training samples from the target domain, as our primary focus is enhancing performance in the target low-resource domain in this paper.

4.3 Overall Objective

The method is proposed to jointly minimize the task-related CDKD loss and auxiliary losses that control the shareable knowledge flow. The total objective function is formulated as:

$$\mathcal{L} = \mathcal{L}_{\text{CDKD}} + \lambda_1\mathcal{L}_{\text{push}} + \lambda_2\mathcal{L}_{\text{pull}}, \tag{9}$$

where λ_1 and λ_2 are trade-off hyperparameters to adjust the effect of different auxiliary losses.

4.4 Time Complexity Analysis

PLM-based taxonomy completion techniques solely rely on the backbone language model for computing probabilities at candidate positions. The computational complexity of this architecture follows $\mathcal{O}(\theta \times d \times l^2)$, with θ corresponding to model parameters, d indicating hidden dimension, and l quantifying average text sequence length. Assuming the number of training nodes is $|\mathcal{N}_{train}|$ and the number of negative samples is N, the training time complexity of PLM-based techniques is $\mathcal{O}(|\mathcal{N}_{train}| \times (1 + N) \times \theta \times d \times l^2)$. Due to our method's utilization of a high-resource taxonomy for extra training samples, it encompasses a larger $|\mathcal{N}|$ compared to the plugged technique. The auxiliary loss calculation also requires extra encoding by the PLM injected with corresponding LoRA modules. Correspondingly, the computational cost during inference is $\mathcal{O}(|\mathcal{C}| \times |\mathcal{A}| \times \theta \times d \times l^2)$, where $|\mathcal{C}|$ corresponds to query count and $|\mathcal{A}|$ signifies candidate positions. Our approach minimally impacts inference time, as all inference operations are conducted solely on the target dataset. The training and inference time is reported in Appendix A.1 and A.2, respectively.

5 Experimental Settings

In this section, we detail our experimental settings. **Implementation details** are in Appendix A.3.

5.1 Datasets

We leverage low-resource taxonomies from three different domains: **Science**, **Equipment**, and **Food** from SemEval-2015 Task 17 (Bordea et al., 2015) as the target taxonomies to evaluate the proposed TaxoPro. We construct their description corpus using the Wikipedia resource by the script provided by Wang et al. (2022a). Meanwhile, two high-resource taxonomies, **MeSH** and **WordNet-Verb**, are leveraged as the source taxonomies. MeSH is a widely used clinical domain taxonomy as a subgraph of the Medical Subject Headings (Lipscomb, 2000). WordNet-Verb is derived from SemEval-2016 Task 14 (Jurgens and Pilehvar, 2016) and it is the hierarchy of verbs from WordNet 3.0. We utilize their description corpus provided by Wang et al. (2022a). For our primary experiments, we use MeSH as the source dataset. The effects of source taxonomy choice will be discussed in Section 6.4.2, where we utilize WordNetVerb to test the impact of changing the source dataset. Following the typical experimental settings of previous taxonomy completion studies (Zhang et al., 2021; Wang et al., 2022a), we split the datasets into non-overlapping train,

Dataset	$\|\mathcal{N}\|/\|\mathcal{N}_{\text{train}}\|$	$\|\mathcal{E}\|$	#depth	#candidates
Science	429/345	441	7	2004
Equipment	475/381	485	7	1822
Food	1486/1190	1533	8	7313
MeSH	9710/8072	10498	10	42970
WordNet-Verb	13936/11936	13407	12	51159

Table 1: Detailed dataset statistics. $|\mathcal{N}|$ and $|\mathcal{E}|$ represent the total number of nodes and edges, respectively.

validation, and test nodes at a ratio of 8:1:1. Detailed dataset statistical information is shown in Table 1.

5.2 Evaluation Metrics

Following Zhang et al. (2021); Arous et al. (2023), we adopt the all-rank evaluation protocol, where a ranking list of all possible candidate positions is output and evaluated for each query concept. We employ Macro Mean Rank (**MR**), Mean Reciprocal Rank (**MRR**), **Recall**@k, and **Hit**@k as metrics for taxonomy completion performance evaluation. Notably, we utilize the original instead of the scaled version of the MRR (Shen et al., 2020).

5.3 Baseline Methods

We first reproduce two representative PLM-based taxonomy expansion and completion techniques and plug TaxoPro into them to verify the effectiveness of our method. **TEMP** (Liu et al., 2021) leverages the PLM to distinguish taxonomy-paths for *structure* information capture in taxonomy expansion. Following Xu et al. (2023), we adapt this method to the completion task by attaching child node c to the end of the taxonomy-path. To resolve non-unique paths in DAG taxonomies, we follow Xu et al. (2023) by sorting all root-to-parent paths in ascending order of length and selecting the shortest one, thereby ensuring TEMP's compatibility with general DAG structures. **TacoPrompt** (Xu et al., 2023) employs the PLM for triplet *semantic* matching in taxonomy completion. It only provides definitions for the Food dataset, and we follow Wang et al. (2022a) to obtain more accessible Wikipedia descriptions for all target datasets.

Secondly, we leverage the state-of-the-art taxonomy completion methods as baselines, including TMN (Zhang et al., 2021), TaxoEnrich (Jiang et al., 2022), QEN (Wang et al., 2022a), TaxoComplete (Arous et al., 2023), and CoSTC (Niu et al., 2024). Building on Zhang et al. (2021)'s framework, we reconfigure two established taxonomy expansion methods, TaxoExpan (Shen et al., 2020) and Arborist (Manzoor et al., 2020), as completion task methods. Lastly, we adapt the expansion method, **Musubu** (Takeoka et al., 2021), that utilizes the PLM as the implicit knowledge base to tackle the low-resource taxonomy expansion problem, to the completion task by averaging the scores of expanding query node q to parent node p and child node c to query node q.

Additionally, we train all original baselines with the additional high-resource taxonomy using the loss defined in Equation 6, while retaining their original models without any LoRA module. These models are termed **Baseline+Joint**.

6 Experimental Results

6.1 Impact of Cross-Domain Taxonomy on Baseline Performance

Table 2 systematically compares taxonomy completion performance between original baselines and their +Joint variants augmented with cross-domain taxonomic knowledge. Through this experiment, we address the core research question:

Q1. Can the cross-domain high-resource taxonomy enhance low-resource taxonomy completion through knowledge transfer? Yes. Empirical results demonstrate that integrating cross-domain knowledge considerably boosts baselines' performance on metrics like MRR and Recall@5, even without specialized algorithms. For instance, TacoPrompt+Joint achieves an absolute improvement of 5.5% over TacoPrompt in the Recall@5 metric. These results reinforce the central motivation behind our method: taxonomies across different domains contain shareable knowledge, which can compensate for data scarcity in low-resource settings. This finding can provide insights to future research, encouraging the exploration of cross-domain knowledge transfer in the taxonomy completion task.

6.2 Performance of TaxoPro

We integrate TaxoPro with two representative PLM-based baselines, TEMP and TacoPrompt, forming their +TaxoPro variants. Table 3 compares these variants with Baseline+Joint, enabling us to explore the key question below.

Target Dataset	Science			Equipment			Food		
Metric	MRR	Hit@1	Recall@5	MRR	Hit@1	Recall@5	MRR	Hit@1	Recall@5
TaxoExpan	$0.118_{\pm 0.005}$	$13.3_{\pm 1.9}$	$11.7_{\pm 0.8}$	$0.073_{\pm 0.003}$	$6.4_{\pm 0.0}$	$9.2_{\pm 1.1}$	$0.105_{\pm 0.013}$	$15.3_{\pm 1.6}$	$12.7_{\pm 2.2}$
TaxoExpan+Joint	$0.240_{\pm 0.032}$	$24.3_{\pm 4.1}$	$28.7_{\pm 4.1}$	$0.227_{\pm 0.030}$	$22.9_{\pm 3.1}$	$28.0_{\pm 2.8}$	$0.129_{\pm 0.014}$	$17.8_{\pm 3.1}$	$16.3_{\pm 1.7}$
Arborist	$0.254_{\pm 0.013}$	$29.1_{\pm 2.3}$	$26.4_{\pm 1.2}$	$0.258_{\pm 0.006}$	$31.5_{\pm 0.8}$	$27.1_{\pm 0.9}$	$0.142_{\pm 0.007}$	$20.8_{\pm 1.3}$	$16.8_{\pm 0.5}$
Arborist+Joint	$0.246_{\pm 0.015}$	$26.2_{\pm 2.1}$	$26.4_{\pm 0.0}$	$0.319_{\pm 0.017}$	$32.8_{\pm 2.9}$	$38.3_{\pm 3.3}$	$0.169_{\pm 0.006}$	$25.1_{\pm 1.6}$	$20.5_{\pm 0.9}$
TMN	$0.265_{\pm 0.020}$	$27.1_{\pm 4.9}$	$29.8_{\pm 2.2}$	$0.262_{\pm 0.011}$	$29.4_{\pm 1.6}$	$30.3_{\pm 1.7}$	$0.153_{\pm 0.006}$	$21.2_{\pm 1.9}$	$18.1_{\pm 0.9}$
TMN+Joint	$0.298_{\pm 0.018}$	$30.5_{\pm 2.8}$	$33.6_{\pm 2.2}$	$0.305_{\pm 0.017}$	$32.8_{\pm 3.7}$	$34.6_{\pm 1.0}$	$0.148_{\pm 0.010}$	$18.9_{\pm 1.6}$	$18.1_{\pm 1.9}$
TaxoEnrich	$0.355_{\pm 0.020}$	$36.7_{\pm 2.9}$	$41.9_{\pm 2.8}$	$0.264_{\pm 0.033}$	$27.6_{\pm 5.7}$	$34.3_{\pm 2.4}$	$0.169_{\pm 0.006}$	$20.8_{\pm 1.2}$	$22.9_{\pm 1.2}$
TaxoEnrich+Joint	$0.306_{\pm 0.019}$	$28.1_{\pm 2.4}$	$36.2_{\pm 3.2}$	$0.286_{\pm 0.019}$	$31.5_{\pm 3.1}$	$35.7_{\pm 1.8}$	$0.175_{\pm 0.008}$	$20.8_{\pm 1.8}$	$24.8_{\pm 0.9}$
QEN	$0.279_{\pm 0.024}$	$25.7_{\pm 3.8}$	$36.7_{\pm 4.0}$	$0.158_{\pm 0.033}$	$15.3_{\pm 6.2}$	$19.4_{\pm 3.4}$	$0.220_{\pm 0.013}$	$32.6_{\pm 2.8}$	$28.0_{\pm 1.6}$
QEN+Joint	$0.339_{\pm 0.037}$	$31.0_{\pm 4.5}$	$43.3_{\pm 5.3}$	$0.243_{\pm 0.014}$	$23.8_{\pm 3.1}$	$31.8_{\pm 3.7}$	$0.248_{\pm 0.021}$	$34.3_{\pm 3.7}$	$32.5_{\pm 2.3}$
TaxoComplete	$0.377_{\pm 0.017}$	$33.3_{\pm 2.1}$	$56.3_{\pm 1.9}$	$0.295_{\pm 0.005}$	$26.4_{\pm 1.0}$	$40.3_{\pm 0.7}$	$0.258_{\pm 0.005}$	$39.6_{\pm 1.4}$	$31.4_{\pm 0.4}$
TaxoComplete+Joint	$0.388_{\pm 0.037}$	$35.7_{\pm 5.0}$	$56.1_{\pm 5.1}$	$0.291_{\pm 0.021}$	$25.1_{\pm 3.4}$	$44.3_{\pm 4.9}$	$0.271_{\pm 0.019}$	$39.3_{\pm 3.1}$	$34.1_{\pm 1.4}$
Musubu	$0.337_{\pm 0.024}$	$28.1_{\pm 3.8}$	$48.9_{\pm 4.8}$	$0.301_{\pm 0.017}$	$26.4_{\pm 3.9}$	$43.4_{\pm 1.9}$	$0.213_{\pm 0.018}$	$27.2_{\pm 3.4}$	$28.0_{\pm 1.8}$
Musubu+Joint	$0.356_{\pm 0.023}$	$27.2_{\pm 2.4}$	$56.3_{\pm 4.8}$	$0.281_{\pm 0.062}$	$23.4_{\pm 9.1}$	$42.8_{\pm 8.1}$	$0.183_{\pm 0.023}$	$21.5_{\pm 4.2}$	$24.9_{\pm 3.2}$
CoSTC	$0.290_{\pm 0.003}$	$35.2_{\pm 1.0}$	$43.6_{\pm 1.2}$	$0.278_{\pm 0.014}$	$24.7_{\pm 1.0}$	$41.3_{\pm 2.9}$	$0.224_{\pm 0.024}$	$21.1_{\pm 4.0}$	$35.9_{\pm 2.8}$
CoSTC+Joint	$0.286_{\pm 0.013}$	$31.0_{\pm 4.5}$	$45.3_{\pm 2.1}$	$0.306_{\pm 0.021}$	$29.8_{\pm 4.7}$	$42.9_{\pm 1.8}$	$0.263_{\pm 0.011}$	$25.7_{\pm 2.8}$	$40.2_{\pm 1.0}$
TEMP	$0.425_{\pm 0.021}$	$37.6_{\pm 5.1}$	$57.8_{\pm 0.8}$	$0.290_{\pm 0.027}$	$25.1_{\pm 5.5}$	$42.5_{\pm 1.7}$	$0.288_{\pm 0.011}$	$41.6_{\pm 2.6}$	$36.7_{\pm 2.3}$
TEMP+Joint	$0.391_{\pm 0.039}$	$27.1_{\pm 6.3}$	$61.1_{\pm 2.3}$	$0.291_{\pm 0.038}$	$23.8_{\pm 7.4}$	$44.2_{\pm 3.3}$	$0.290_{\pm 0.004}$	$40.6_{\pm 2.0}$	$37.9_{\pm 0.9}$
TacoPrompt	$0.456_{\pm 0.027}$	$42.4_{\pm 4.9}$	$59.3_{\pm 3.1}$	$0.288_{\pm 0.008}$	$25.5_{\pm 3.0}$	$41.1_{\pm 3.1}$	$0.304_{\pm 0.006}$	$43.5_{\pm 1.6}$	$39.6_{\pm 0.9}$
TacoPrompt+Joint	$0.462_{\pm 0.030}$	$39.1_{\pm 7.4}$	$64.8_{\pm 3.5}$	$0.285_{\pm 0.016}$	$23.4_{\pm 3.0}$	$44.5_{\pm 1.1}$	$0.305_{\pm 0.011}$	$41.1_{\pm 2.7}$	$41.2_{\pm 1.8}$

Table 2: Impacts of the cross-domain high-resource taxonomy on baseline performance. Results are averaged over five independent runs. For the results of all metrics, please refer to Appendix A.4.

Q2. Can TaxoPro improve PLM-based taxonomy completion techniques in low-resource scenarios? Yes. By analyzing experimental results, we can draw several key observations. First, PLM-based methods, particularly TEMP and TacoPrompt, outperform others in low-resource scenarios, leading in the Recall@5 metric across all three datasets, as shown in Table 2. This indicates that PLMs can serve as an effective implicit knowledge base (Takeoka et al., 2021) for low-resource taxonomy completion. Second, the effectiveness of pre-trained knowledge varies by domain. For instance, TacoPrompt performs worse on the Equipment dataset than on Science or Food, confirming the limitations of relying solely on pre-trained knowledge for taxonomy completion.

Lastly, PLM-based methods' +TaxoPro variants consistently surpass their original counterparts. Specifically, TEMP+TaxoPro improves TEMP on MRR/Hit@1/Recall@5 by 0.060/9.1%/5.5%, 0.041/2.6%/8.0%, and 0.048/6.9%/3.8% on the Science, Equipment, and Food datasets, respectively. Similarly, TacoPrompt+TaxoPro achieves gains of 0.079/7.6%/10.7%, 0.061/8.1%/10.4%, and 0.033/6.2%/4.3% on these datasets. Additionally, TEMP+TaxoPro and TacoPrompt+TaxoPro outperform their +Joint variants in MRR and Recall@5 while improving Hit@1, which the +Joint variants decrease. Notably, TacoPrompt+TaxoPro surpasses all Baselines+Joint variants on most metrics. These results underscore TaxoPro's effectiveness in leveraging cross-domain knowledge to enhance PLM-based taxonomy completion in low-resource scenarios.

6.3 Ablation Studies

As indicated in Table 4, we study the performance of TaxoPro under different settings. Specifically, in the settings w/o CD (cross-domain) and w/o CDKD (cross-domain knowledge decomposition), we utilize vanilla LoRA (Hu et al., 2022a), where only a pair of low-rank matrices, namely, $\mathbf{B}$ and $\mathbf{A}$, are injected into the PLM to learn knowledge for taxonomy completion, as outlined in Equation 4. In the w/o CD setting, we use training samples solely from the target domain. In contrast, in the w/o CDKD setting, we use samples from both the target and source domains. Notably, Baseline+TaxoPro w/o CD corresponds to the vanilla LoRA-tuned Baseline, while Baseline+TaxoPro w/o CDKD represents the vanilla LoRA-tuned Baseline+Joint.

TacoPrompt is used as the backbone technique for ablation studies and further discussions since it achieves the most competitive performance. In this context, w/o CD refers to the vanilla LoRA-tuned version of TacoPrompt, while w/o CDKD denotes the vanilla LoRA-tuned version of

Method	MR↓	MRR	Recall@1	Recall@5	Recall@10	Hit@1	Hit@5	Hit@10
				Science				
TaxoExpan+Joint	126.5$_{\pm 28.5}$	0.240$_{\pm 0.032}$	19.3$_{\pm 3.2}$	28.7$_{\pm 4.1}$	34.7$_{\pm 3.9}$	24.3$_{\pm 4.1}$	36.2$_{\pm 5.1}$	43.3$_{\pm 4.1}$
Arborist+Joint	67.3$_{\pm 2.7}$	0.246$_{\pm 0.015}$	20.8$_{\pm 1.7}$	26.4$_{\pm 0.0}$	30.6$_{\pm 1.9}$	26.2$_{\pm 2.1}$	33.3$_{\pm 0.0}$	37.6$_{\pm 2.8}$
TMN+Joint	52.3$_{\pm 3.2}$	0.298$_{\pm 0.018}$	24.1$_{\pm 2.2}$	33.6$_{\pm 2.2}$	37.3$_{\pm 2.2}$	30.5$_{\pm 2.8}$	42.4$_{\pm 2.8}$	47.1$_{\pm 2.8}$
TaxoEnrich+Joint	31.4$_{\pm 4.2}$	0.306$_{\pm 0.019}$	22.2$_{\pm 1.8}$	36.2$_{\pm 3.2}$	45.7$_{\pm 4.4}$	28.1$_{\pm 2.4}$	45.2$_{\pm 4.0}$	56.2$_{\pm 3.9}$
QEN+Joint	58.4$_{\pm 23.1}$	0.339$_{\pm 0.037}$	24.1$_{\pm 3.5}$	43.3$_{\pm 5.3}$	50.0$_{\pm 4.8}$	31.0$_{\pm 4.5}$	53.3$_{\pm 6.3}$	57.6$_{\pm 4.4}$
TaxoComplete+Joint	46.7$_{\pm 14.9}$	0.388$_{\pm 0.037}$	27.8$_{\pm 3.9}$	56.1$_{\pm 5.1}$	65.6$_{\pm 4.3}$	35.7$_{\pm 5.0}$	62.8$_{\pm 7.1}$	72.4$_{\pm 3.9}$
Musubu+Joint	116.1$_{\pm 9.1}$	0.356$_{\pm 0.023}$	21.1$_{\pm 1.9}$	56.3$_{\pm 4.8}$	68.2$_{\pm 3.6}$	27.2$_{\pm 2.4}$	65.7$_{\pm 3.9}$	74.8$_{\pm 3.3}$
CoSTC+Joint	15.0$_{\pm 3.7}$	0.286$_{\pm 0.013}$	13.1$_{\pm 1.9}$	45.3$_{\pm 2.1}$	64.2$_{\pm 2.3}$	31.0$_{\pm 4.5}$	74.7$_{\pm 3.2}$	86.7$_{\pm 3.3}$
TEMP	19.9$_{\pm 4.8}$	0.425$_{\pm 0.021}$	29.2$_{\pm 4.0}$	57.8$_{\pm 0.8}$	66.7$_{\pm 2.6}$	37.6$_{\pm 5.1}$	74.3$_{\pm 1.0}$	84.8$_{\pm 2.4}$
TEMP+Joint	13.5$_{\pm 7.2}$	0.391$_{\pm 0.039}$	21.1$_{\pm 4.9}$	61.1$_{\pm 2.3}$	73.7$_{\pm 1.4}$	27.1$_{\pm 6.3}$	76.7$_{\pm 2.8}$	88.1$_{\pm 1.5}$
TEMP+TaxoPro	11.6$_{\pm 5.2}$ ↑	0.485$_{\pm 0.024}$ ↑	36.3$_{\pm 1.9}$ ↑	63.3$_{\pm 2.2}$ ↑	75.5$_{\pm 3.4}$ ↑	46.7$_{\pm 2.4}$ ↑	79.5$_{\pm 2.4}$ ↑	**90.9**$_{\pm 3.8}$ ↑
TacoPrompt	16.4$_{\pm 9.9}$	0.456$_{\pm 0.027}$	32.9$_{\pm 3.8}$	59.3$_{\pm 3.1}$	70.7$_{\pm 3.6}$	42.4$_{\pm 4.9}$	74.3$_{\pm 2.8}$	85.2$_{\pm 1.0}$
TacoPrompt+Joint	12.2$_{\pm 7.7}$	0.462$_{\pm 0.030}$	30.4$_{\pm 5.8}$	64.8$_{\pm 3.5}$	75.2$_{\pm 1.9}$	39.1$_{\pm 7.4}$	79.5$_{\pm 3.6}$	86.2$_{\pm 2.4}$
TacoPrompt+TaxoPro	**6.3**$_{\pm 1.1}$ ↑	**0.535**$_{\pm 0.013}$ ↑	**39.3**$_{\pm 2.4}$ ↑	**70.0**$_{\pm 1.4}$ ↑	**78.5**$_{\pm 1.9}$ ↑	**50.0**$_{\pm 3.0}$ ↑	**83.8**$_{\pm 1.8}$ ↑	90.0$_{\pm 3.2}$ ↑
				Equipment				
TaxoExpan+Joint	178.7$_{\pm 107.5}$	0.227$_{\pm 0.030}$	15.4$_{\pm 2.1}$	28.0$_{\pm 2.8}$	36.6$_{\pm 3.4}$	22.9$_{\pm 3.1}$	41.3$_{\pm 3.5}$	52.8$_{\pm 3.1}$
Arborist+Joint	38.3$_{\pm 3.4}$	0.319$_{\pm 0.017}$	22.0$_{\pm 1.9}$	38.3$_{\pm 3.3}$	41.7$_{\pm 4.5}$	32.8$_{\pm 2.9}$	49.8$_{\pm 1.1}$	53.2$_{\pm 3.8}$
TMN+Joint	40.5$_{\pm 6.6}$	0.305$_{\pm 0.017}$	22.0$_{\pm 2.5}$	34.6$_{\pm 1.0}$	42.3$_{\pm 1.9}$	32.8$_{\pm 3.7}$	47.2$_{\pm 1.6}$	54.0$_{\pm 3.5}$
TaxoEnrich+Joint	65.9$_{\pm 11.9}$	0.286$_{\pm 0.019}$	21.2$_{\pm 2.1}$	35.7$_{\pm 1.8}$	40.3$_{\pm 1.4}$	31.5$_{\pm 3.1}$	51.5$_{\pm 2.5}$	57.8$_{\pm 2.1}$
QEN+Joint	99.5$_{\pm 21.8}$	0.243$_{\pm 0.014}$	15.8$_{\pm 2.1}$	31.8$_{\pm 3.7}$	42.5$_{\pm 4.5}$	23.8$_{\pm 3.1}$	45.5$_{\pm 3.7}$	52.8$_{\pm 3.9}$
TaxoComplete+Joint	122.0$_{\pm 29.9}$	0.291$_{\pm 0.021}$	16.6$_{\pm 2.2}$	44.3$_{\pm 4.9}$	56.9$_{\pm 1.5}$	25.1$_{\pm 3.4}$	51.9$_{\pm 3.7}$	62.1$_{\pm 3.1}$
Musubu+Joint	117.8$_{\pm 11.3}$	0.281$_{\pm 0.062}$	15.5$_{\pm 6.0}$	42.8$_{\pm 8.1}$	58.3$_{\pm 5.3}$	23.4$_{\pm 9.1}$	51.9$_{\pm 8.6}$	64.7$_{\pm 4.2}$
CoSTC+Joint	41.3$_{\pm 8.2}$	0.306$_{\pm 0.021}$	18.7$_{\pm 2.9}$	42.9$_{\pm 1.8}$	59.1$_{\pm 2.8}$	29.8$_{\pm 4.7}$	58.7$_{\pm 4.2}$	69.8$_{\pm 2.5}$
TEMP	92.7$_{\pm 13.7}$	0.290$_{\pm 0.027}$	16.6$_{\pm 3.6}$	42.5$_{\pm 1.7}$	55.5$_{\pm 3.6}$	25.1$_{\pm 5.5}$	58.7$_{\pm 2.2}$	68.5$_{\pm 2.4}$
TEMP+Joint	72.9$_{\pm 6.6}$	0.291$_{\pm 0.038}$	15.8$_{\pm 4.9}$	44.2$_{\pm 3.3}$	57.2$_{\pm 1.9}$	23.8$_{\pm 7.4}$	60.4$_{\pm 4.0}$	69.4$_{\pm 2.1}$
TEMP+TaxoPro	68.4$_{\pm 4.1}$ ↑	0.331$_{\pm 0.020}$ ↑	18.3$_{\pm 2.7}$ ↑	50.5$_{\pm 1.6}$ ↑	62.3$_{\pm 3.5}$ ↑	27.7$_{\pm 4.0}$ ↑	63.8$_{\pm 1.9}$ ↑	71.1$_{\pm 3.9}$ ↑
TacoPrompt	65.3$_{\pm 38.0}$	0.288$_{\pm 0.008}$	16.9$_{\pm 2.0}$	41.1$_{\pm 3.1}$	57.7$_{\pm 3.1}$	25.5$_{\pm 3.0}$	56.6$_{\pm 3.9}$	67.7$_{\pm 4.1}$
TacoPrompt+Joint	69.4$_{\pm 11.8}$	0.285$_{\pm 0.016}$	15.5$_{\pm 2.0}$	44.5$_{\pm 1.1}$	59.7$_{\pm 1.5}$	23.4$_{\pm 3.0}$	60.4$_{\pm 2.6}$	68.9$_{\pm 2.1}$
TacoPrompt+TaxoPro	**34.7**$_{\pm 12.5}$ ↑	**0.349**$_{\pm 0.009}$ ↑	**22.2**$_{\pm 1.0}$ ↑	**51.5**$_{\pm 1.7}$ ↑	**63.1**$_{\pm 3.3}$ ↑	**33.6**$_{\pm 1.6}$ ↑	**66.0**$_{\pm 3.0}$ ↑	**72.8**$_{\pm 1.6}$ ↑
				Food				
TaxoExpan+Joint	403.0$_{\pm 171.4}$	0.129$_{\pm 0.014}$	8.8$_{\pm 1.5}$	16.3$_{\pm 1.7}$	20.1$_{\pm 1.9}$	17.8$_{\pm 3.1}$	31.8$_{\pm 3.5}$	38.2$_{\pm 3.0}$
Arborist+Joint	205.4$_{\pm 4.9}$	0.169$_{\pm 0.006}$	12.4$_{\pm 0.8}$	20.5$_{\pm 0.9}$	25.8$_{\pm 2.1}$	25.1$_{\pm 1.6}$	38.4$_{\pm 1.6}$	44.1$_{\pm 2.1}$
TMN+Joint	143.5$_{\pm 3.8}$	0.148$_{\pm 0.010}$	9.3$_{\pm 0.8}$	18.1$_{\pm 1.9}$	25.1$_{\pm 2.5}$	18.9$_{\pm 1.6}$	34.6$_{\pm 4.1}$	44.9$_{\pm 4.7}$
TaxoEnrich+Joint	198.8$_{\pm 22.7}$	0.175$_{\pm 0.008}$	10.3$_{\pm 0.9}$	24.8$_{\pm 0.9}$	30.9$_{\pm 0.8}$	20.8$_{\pm 1.8}$	45.7$_{\pm 1.6}$	55.8$_{\pm 0.7}$
QEN+Joint	173.7$_{\pm 25.9}$	0.248$_{\pm 0.021}$	16.3$_{\pm 1.8}$	32.5$_{\pm 2.3}$	41.4$_{\pm 2.6}$	34.3$_{\pm 3.7}$	59.0$_{\pm 3.7}$	68.9$_{\pm 3.0}$
TaxoComplete+Joint	385.0$_{\pm 31.2}$	0.271$_{\pm 0.019}$	18.7$_{\pm 1.5}$	34.1$_{\pm 1.4}$	42.9$_{\pm 2.0}$	39.3$_{\pm 3.1}$	60.7$_{\pm 2.0}$	66.8$_{\pm 1.8}$
Musubu+Joint	543.9$_{\pm 62.0}$	0.183$_{\pm 0.023}$	10.2$_{\pm 2.0}$	24.9$_{\pm 3.2}$	35.9$_{\pm 2.8}$	21.5$_{\pm 4.2}$	43.9$_{\pm 5.5}$	57.2$_{\pm 4.3}$
CoSTC+Joint	72.6$_{\pm 5.5}$	0.263$_{\pm 0.011}$	17.8$_{\pm 5.0}$	40.2$_{\pm 1.0}$	51.3$_{\pm 1.6}$	25.7$_{\pm 2.8}$	65.5$_{\pm 1.5}$	75.5$_{\pm 1.8}$
TEMP	66.7$_{\pm 12.4}$	0.288$_{\pm 0.011}$	19.8$_{\pm 1.2}$	36.7$_{\pm 2.3}$	46.1$_{\pm 1.8}$	41.6$_{\pm 2.6}$	69.6$_{\pm 3.5}$	78.9$_{\pm 2.1}$
TEMP+Joint	**53.3**$_{\pm 10.7}$	0.290$_{\pm 0.004}$	19.3$_{\pm 1.0}$	37.9$_{\pm 0.9}$	46.3$_{\pm 1.8}$	40.6$_{\pm 2.0}$	71.3$_{\pm 1.3}$	79.3$_{\pm 1.2}$
TEMP+TaxoPro	75.4$_{\pm 17.7}$ ↓	0.320$_{\pm 0.009}$ ↑	23.1$_{\pm 1.0}$ ↑	40.5$_{\pm 1.2}$ ↑	47.6$_{\pm 1.2}$ ↑	48.5$_{\pm 2.2}$ ↑	75.7$_{\pm 1.9}$ ↑	81.4$_{\pm 1.2}$ ↑
TacoPrompt	114.3$_{\pm 27.1}$	0.304$_{\pm 0.006}$	20.7$_{\pm 0.7}$	39.6$_{\pm 0.9}$	50.2$_{\pm 1.8}$	43.5$_{\pm 1.6}$	73.4$_{\pm 0.9}$	81.4$_{\pm 1.6}$
TacoPrompt+Joint	138.5$_{\pm 33.0}$	0.305$_{\pm 0.011}$	19.5$_{\pm 1.3}$	41.2$_{\pm 1.8}$	51.3$_{\pm 1.8}$	41.1$_{\pm 2.7}$	73.2$_{\pm 2.1}$	81.8$_{\pm 0.9}$
TacoPrompt+TaxoPro	78.0$_{\pm 26.6}$ ↑	**0.337**$_{\pm 0.017}$ ↑	**23.7**$_{\pm 1.8}$ ↑	**43.9**$_{\pm 2.0}$ ↑	**54.0**$_{\pm 2.3}$ ↑	**49.7**$_{\pm 3.7}$ ↑	**76.3**$_{\pm 3.1}$ ↑	**81.9**$_{\pm 2.1}$ ↑

Table 3: Performance comparison between TaxoPro and Baseline+Joint variants. Average results over five runs are reported. Please refer to Appendix A.4 for the comparison results between TaxoPro and Baselines.

TacoPrompt+Joint. Guided by the ablation results, we discuss the subsequent questions.

Q3. Is the cross-domain shareable knowledge effective in improving the low-resource taxonomy completion? Yes, the results reveal a performance degradation when removing the CD module (w/o CD). For instance, Hit@1/Recall@5 drops 11.9%/14.8%, 8.5%/10.6%, and 9.2%/7.3% on Science, Equipment, and Food datasets, respectively. This further demonstrates that shareable knowledge exists between taxonomies varying from domains and scales, and it helps to complete the target low-resource taxonomies, where such knowledge is inadequately learned from the limited training samples.

Q4. How does CDKD improve performance? It can prevent negative interference between domain-specific and domain-shared knowledge. Comparing the results under the settings w/o CD, w/o CDKD, and w/o $\mathcal{L}_{\text{pull}}$, $\mathcal{L}_{\text{push}}$, we can make the following two observations. First, a notable performance decline is observed when training data

Setting	Recall@1	Recall@5	Hit@1	Hit@5	MRR
			Science		
TacoPrompt+Joint	$30.4_{\pm5.8}$	$64.8_{\pm3.5}$	$39.1_{\pm7.4}$	$83.8_{\pm1.8}$	$0.462_{\pm0.030}$
TacoPrompt+TaxoPro	$\mathbf{39.3}_{\pm2.4}$	$\mathbf{70.0}_{\pm1.4}$	$\mathbf{50.0}_{\pm3.0}$	$\mathbf{83.8}_{\pm1.8}$	$\mathbf{0.535}_{\pm0.013}$
w/o CD	$29.6_{\pm4.5}$	$55.2_{\pm4.9}$	$38.1_{\pm5.8}$	$70.5_{\pm5.6}$	$0.415_{\pm0.032}$
w/o CDKD	$21.1_{\pm10.2}$	$61.9_{\pm3.4}$	$27.1_{\pm13.2}$	$76.7_{\pm3.5}$	$0.388_{\pm0.065}$
w/o $\mathcal{L}_{\text{pull}}, \mathcal{L}_{\text{push}}$	$31.1_{\pm4.7}$	$65.9_{\pm3.0}$	$40.0_{\pm6.1}$	$81.4_{\pm2.8}$	$0.464_{\pm0.032}$
w/o $\mathcal{L}_{\text{pull}}$	$36.3_{\pm3.2}$	$68.2_{\pm4.3}$	$46.7_{\pm4.2}$	$83.8_{\pm3.2}$	$0.501_{\pm0.017}$
w/o $\mathcal{L}_{\text{push}}$	$36.3_{\pm2.5}$	$66.7_{\pm4.8}$	$46.7_{\pm3.2}$	$81.0_{\pm2.6}$	$0.504_{\pm0.027}$
			Equipment		
TacoPrompt+Joint	$15.5_{\pm2.0}$	$44.5_{\pm1.1}$	$23.4_{\pm3.0}$	$60.4_{\pm2.6}$	$0.285_{\pm0.016}$
TacoPrompt+TaxoPro	$\mathbf{22.2}_{\pm1.0}$	$\mathbf{51.5}_{\pm1.7}$	$\mathbf{33.6}_{\pm1.6}$	$\mathbf{66.0}_{\pm3.0}$	$\mathbf{0.349}_{\pm0.009}$
w/o CD	$16.6_{\pm2.1}$	$40.9_{\pm2.5}$	$25.1_{\pm3.1}$	$56.2_{\pm3.5}$	$0.285_{\pm0.018}$
w/o CDKD	$13.3_{\pm2.1}$	$43.1_{\pm4.2}$	$20.0_{\pm3.2}$	$57.9_{\pm5.9}$	$0.274_{\pm0.017}$
w/o $\mathcal{L}_{\text{pull}}, \mathcal{L}_{\text{push}}$	$18.9_{\pm1.4}$	$45.7_{\pm1.9}$	$28.5_{\pm2.2}$	$60.4_{\pm2.9}$	$0.318_{\pm0.006}$
w/o $\mathcal{L}_{\text{pull}}$	$20.5_{\pm2.1}$	$47.1_{\pm2.9}$	$31.0_{\pm3.1}$	$60.4_{\pm3.5}$	$0.327_{\pm0.018}$
w/o $\mathcal{L}_{\text{push}}$	$21.4_{\pm4.3}$	$47.3_{\pm4.8}$	$32.3_{\pm6.5}$	$62.1_{\pm2.5}$	$0.334_{\pm0.041}$
			Food		
TacoPrompt+Joint	$19.5_{\pm1.3}$	$41.2_{\pm1.8}$	$41.1_{\pm2.7}$	$73.2_{\pm2.1}$	$0.305_{\pm0.011}$
TacoPrompt+TaxoPro	$\mathbf{23.7}_{\pm1.8}$	$\mathbf{43.9}_{\pm2.0}$	$\mathbf{49.7}_{\pm3.7}$	$\mathbf{76.3}_{\pm3.1}$	$\mathbf{0.337}_{\pm0.017}$
w/o CD	$19.3_{\pm1.4}$	$36.6_{\pm2.4}$	$40.5_{\pm2.9}$	$68.9_{\pm4.1}$	$0.286_{\pm0.013}$
w/o CDKD	$14.8_{\pm3.4}$	$35.8_{\pm2.5}$	$31.1_{\pm7.1}$	$68.1_{\pm3.7}$	$0.253_{\pm0.023}$
w/o $\mathcal{L}_{\text{pull}}, \mathcal{L}_{\text{push}}$	$20.8_{\pm0.9}$	$40.4_{\pm2.0}$	$43.7_{\pm1.9}$	$72.5_{\pm3.4}$	$0.307_{\pm0.011}$
w/o $\mathcal{L}_{\text{pull}}$	$21.7_{\pm1.6}$	$42.0_{\pm1.0}$	$45.5_{\pm3.3}$	$74.7_{\pm2.0}$	$0.316_{\pm0.015}$
w/o $\mathcal{L}_{\text{push}}$	$22.1_{\pm1.6}$	$41.6_{\pm1.5}$	$46.5_{\pm3.3}$	$74.3_{\pm2.1}$	$0.317_{\pm0.013}$

Table 4: Ablation studies on all three datasets. We report the average results of five runs.

Setting	MRR	R@1	R@5	R@10	H@1	H@5	H@10
Science (TaxoPro)	**0.529**	**38.9**	**70.4**	**79.6**	**50.0**	**83.3**	**90.5**
Reversed	0.317	9.3	59.3	74.1	11.9	73.8	88.1
Only Shared	0.476	31.5	63.0	75.9	40.5	78.6	88.1
Only Specific	0.025	0.0	3.7	5.6	0.0	4.8	7.1
Equipment (TaxoPro)	**0.345**	**22.5**	**49.3**	**59.2**	**34.0**	**63.8**	**70.2**
Reversed	0.181	4.2	32.4	50.7	6.4	44.7	66.0
Only Shared	0.308	18.3	40.8	57.7	27.7	57.4	70.2
Only Specific	0.061	4.2	5.6	11.3	6.4	8.5	12.9
Food (TaxoPro)	**0.350**	**24.8**	**45.3**	**56.3**	**52.0**	**77.7**	**81.8**
Reversed	0.187	4.8	32.5	44.1	10.1	62.2	73.0
Only Shared	0.320	21.2	40.5	52.7	44.6	71.6	80.4
Only Specific	0.044	1.6	5.1	8.7	3.4	10.8	18.2

Table 5: Impact of different combinations of the learned domain-shared and domain-specific knowledge. We leverage the trained model that performs best on the validation set among different runs for the study. For ‘‘Reversed’’, we replace the domain-specific knowledge of the target dataset with that of the source dataset.

from the target domain is used without knowledge decomposition. This drop is particularly pronounced in the Hit@1 metric. Specifically, the method w/o CDKD that learns from the extra target dataset performs even worse than that w/o CD that learns only from the source dataset. This illustrates that domain-specific knowledge will become noise when applied to a different domain.

Second, the method only w/t the CDKD module (w/o $\mathcal{L}_{\text{pull}}, \mathcal{L}_{\text{push}}$) completes low-resource taxonomies better than the method w/o CD. This improvement is due to the CDKD module's ability to separate noisy domain-specific knowledge from domain-shareable knowledge. Furthermore, it can be observed that the results of w/o CDKD (vanilla LoRA-tuned TacoPrompt+Joint) are worse than TacoPrompt+Joint, indicating that, in the absence of knowledge decomposition, LoRA-tuning is more strongly affected by noisy domain-specific knowledge compared to full fine-tuning. On the other hand, the method only w/t the CDKD module (w/o $\mathcal{L}_{\text{pull}}, \mathcal{L}_{\text{push}}$) outperforms TacoPrompt+Joint, further demonstrating the CDKD module's ability to isolate and mitigate the impact of noisy domain-specific knowledge.

Additionally, Figure 4 shows that simply adjusting the domain loss balance hyperparameter α in the w/o CDKD setting never outperforms TaxoPro. This underscores the necessity of the CDKD module when utilizing the external high-resource source taxonomy.

Q5. Can auxiliary losses control the flow of shareable knowledge? Yes. We find that both auxiliary losses, namely $\mathcal{L}_{\text{pull}}$ and $\mathcal{L}_{\text{push}}$, can individually improve the performance of the method w/o any of them. This demonstrates the effectiveness of both auxiliary losses in controlling the flow of shareable knowledge. More importantly, we observe that the method w/t both losses outperforms the method w/t either of them alone. For example, TacoPrompt+TaxoPro outperforms TacoPrompt+TaxoPro w/o $\mathcal{L}_{\text{pull}}$ by an average of 2.2% on the Recall@10 metric across three datasets. This observation indicates that the two losses control the flow of shareable knowledge from different perspectives, jointly improving performance.

6.4 Further Discussion

In this section, we discuss in two main ways: the additional perspective of learned knowledge, and the impact of key hyperparameters on TaxoPro.

6.4.1 Discussions of Learned Knowledge

After training converges, domain-shared and -specific knowledge are stored in their respective low-rank matrices (LoRA), as described in Equation 5. Firstly, we inject domain-shared and -specific LoRA into the PLM using different combinations during inference. As shown in Table 5, TaxoPro injects domain-shared LoRA and target-specific LoRA; Reversed injects domain-shared LoRA and source-specific LoRA; Only Shared injects only domain-shared LoRA; and Only Specific injects only target-specific LoRA. The subsequent research question is examined using evidence derived from the results.

Setting	MRR	R@1	R@5	R@10	H@1	H@5	H@10
				Science			
End-To-End	0.529	38.9	70.4	79.6	50.0	83.3	90.5
+ Specific LoRA	0.533 ↑	39.3 ↑	70.7 ↑	78.1 ↓	50.5 ↑	84.3 ↑	88.6 ↓
+ Fine-Tuning	0.528 ↓	37.8 ↓	70.4	76.7 ↓	48.6 ↓	84.3 ↑	86.7 ↓
+ Adapter	0.455 ↓	32.7 ↓	59.3 ↓	73.2 ↓	42.1 ↓	74.2 ↓	87.7 ↓
Only Fine-Tuning	0.456	32.9	59.3	70.7	42.4	74.3	85.2
Only Adapter	0.434	28.9	60.0	73.3	37.2	76.7	88.5
				Equipment			
End-To-End	0.345	22.5	49.3	59.2	34.0	63.8	70.2
+ Specific LoRA	0.355 ↑	25.1 ↑	47.1 ↓	58.0 ↓	37.9 ↑	61.7 ↓	67.3 ↓
+ Fine-Tuning	0.359 ↑	25.4 ↑	47.9 ↓	58.9 ↓	38.3 ↑	61.3 ↓	68.1 ↓
+ Adapter	0.267 ↓	16.6 ↓	34.9 ↓	47.6 ↓	25.1 ↓	51.5 ↓	65.1 ↓
Only Fine-Tuning	0.288	16.9	41.1	57.7	25.5	56.6	67.7
Only Adapter	0.237	14.1	31.8	44.8	21.3	48.1	62.6
				Food			
End-To-End	0.350	24.8	45.3	56.3	52.0	77.7	81.8
+ Specific LoRA	0.357 ↑	25.3 ↑	46.4 ↑	55.6 ↓	53.1 ↑	79.2 ↑	83.7 ↑
+ Fine-Tuning	0.357 ↑	25.7 ↑	46.0 ↑	55.9 ↓	53.9 ↑	77.8 ↑	83.2 ↑
+ Adapter	0.316 ↓	21.7 ↓	40.4 ↓	49.2 ↓	45.7 ↓	75.0 ↓	82.9 ↑
Only Fine-Tuning	0.304	20.7	39.6	50.2	43.5	73.4	81.4
Only Adapter	0.301	20.3	39.2	50.2	42.5	71.1	81.6

Table 6: Performance of different two-stage tuning strategies. We load and freeze the domain-shared knowledge from the trained model as the start point, and learn the domain-specific knowledge using the target training samples. For ‘‘Specific LoRA’’, we continually tune the domain-specific low-rank matrices. For ‘‘+ Adapter’’, we inject the Adapter (Houlsby et al., 2019) into the BERT that has loaded the learned knowledge. To study the impact of the knowledge stored in LoRA for other tuning techniques, we perform ‘‘Only’’ experiments without loading the learned knowledge. Empirically, we tune the Adapter, BERT, and Specific LoRA with the learning rate 3E-4, 5E-5, and 5E-4, respectively. We report average results of five runs.

Q6. What are the roles of shared and domain-specific knowledge in completing the target taxonomy? The domain-shared knowledge contains essential information necessary to complete the target low-resource taxonomy. When relying solely on domain-shared knowledge, it achieves comparable Hit@10 and Recall@10 performance to that obtained by combining both domain-shared and -specific knowledge. Meanwhile, domain-specific knowledge captures fine-grained distinctions related to the domain, which significantly impacts Hit@1 and Recall@1 performance. Additionally, the performance drops significantly in the Reversed setting. For instance, the MRR average drops by 0.180 compared to the original setting on three datasets. This observation highlights that domain-specific knowledge corresponds to fine-grained information that exhibits strong relevance to the specific domain.

Furthermore, we adopt two-stage tuning strategies, where the learned domain-shared knowledge is loaded and frozen, allowing us to focus on learning domain-specific knowledge by different techniques. The results are shown in Table 6. Additionally, we study the effectiveness of the learned domain-shared knowledge on transfer learning. The results are displayed in Table 7, from which we explore the following research questions.

Setting	Recall@1	Recall@5	Hit@1	Hit@5	MRR
			Science		
− CD	$29.6_{\pm 4.5}$	$55.2_{\pm 4.9}$	$38.1_{\pm 5.8}$	$70.5_{\pm 5.6}$	$0.415_{\pm 0.032}$
+ Transfer-ES	$37.4_{\pm 3.2}$	$64.1_{\pm 1.9}$	$48.1_{\pm 4.1}$	$78.6_{\pm 1.5}$	$0.501_{\pm 0.014}$
+ Transfer-FS	$36.3_{\pm 2.5}$	$61.5_{\pm 2.7}$	$46.7_{\pm 3.2}$	$76.2_{\pm 3.7}$	$0.476_{\pm 0.024}$
+ TaxoPro	$\mathbf{39.3}_{\pm 2.4}$	$\mathbf{70.0}_{\pm 1.4}$	$\mathbf{50.0}_{\pm 3.0}$	$\mathbf{83.8}_{\pm 1.8}$	$\mathbf{0.535}_{\pm 0.013}$
			Equipment		
− CD	$16.6_{\pm 2.1}$	$40.9_{\pm 2.5}$	$25.1_{\pm 3.1}$	$56.2_{\pm 3.5}$	$0.285_{\pm 0.018}$
+ Transfer-SE	$18.3_{\pm 3.1}$	$45.1_{\pm 4.6}$	$27.6_{\pm 4.6}$	$58.7_{\pm 4.2}$	$0.316_{\pm 0.026}$
+ Transfer-FE	$20.3_{\pm 1.4}$	$49.0_{\pm 2.7}$	$30.6_{\pm 2.4}$	$61.3_{\pm 3.9}$	$0.326_{\pm 0.008}$
+ TaxoPro	$\mathbf{22.2}_{\pm 1.0}$	$\mathbf{51.5}_{\pm 1.7}$	$\mathbf{33.6}_{\pm 1.6}$	$\mathbf{66.0}_{\pm 3.0}$	$\mathbf{0.349}_{\pm 0.009}$
			Food		
− CD	$19.3_{\pm 1.4}$	$36.6_{\pm 2.4}$	$40.5_{\pm 2.9}$	$68.9_{\pm 4.1}$	$0.286_{\pm 0.013}$
+ Transfer-SF	$21.4_{\pm 0.7}$	$41.8_{\pm 0.9}$	$45.0_{\pm 1.5}$	$74.6_{\pm 1.6}$	$0.313_{\pm 0.007}$
+ Transfer-EF	$20.9_{\pm 1.3}$	$40.8_{\pm 1.2}$	$43.9_{\pm 2.7}$	$73.5_{\pm 2.4}$	$0.308_{\pm 0.010}$
+ TaxoPro	$\mathbf{23.7}_{\pm 1.8}$	$\mathbf{43.9}_{\pm 2.0}$	$\mathbf{49.7}_{\pm 3.7}$	$\mathbf{76.3}_{\pm 3.1}$	$\mathbf{0.337}_{\pm 0.017}$

Table 7: Performance of transfer learning (TL) using the learned domain-shared knowledge. For example, in Transfer-ES, we load and freeze the domain-shared knowledge learned for the Equipment (E) dataset, then learn the domain-specific knowledge on the target dataset, Science (S). We compare the performance of TL with that of w/o CD and TaxoPro settings.

Q7. What is the impact of two-stage tuning? During training, we noticed that either tuning specific LoRA or fully fine-tuning can inherit the knowledge acquired during the initial end-to-end stage. Consequently, this leads to performance comparable to the first stage. However, although these methods enhance Hit@1, they also lead to a decrease in Recall@10, indicating a potential issue of overfitting domain-specific knowledge in the second stage. Additionally, we observed that incorporating the Adapter (Houlsby et al., 2019) in the second stage initially yields poor performance and ultimately leads to a significant drop in performance compared to the first stage. In conclusion, the end-to-end training strategy of TaxoPro proved to be more robust than two-stage tuning strategies.

Q8. What are potential applications of the learned domain-shared knowledge? Firstly, the learned domain-shared knowledge can enhance other tuning techniques for the task. As shown in Table 6, we compare the performance of different tuning techniques w/t (‘‘+ Tech’’) or w/o (‘‘Only Tech’’) the learned domain-shared knowledge. We find that incorporating domain-shared

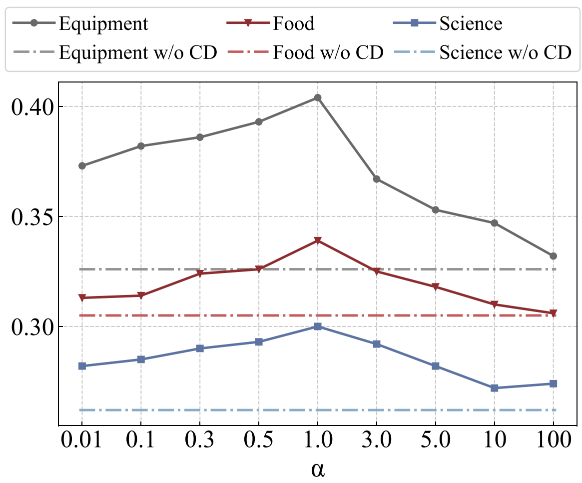


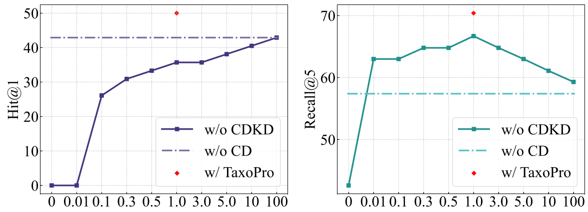
Figure 3: The results of TaxoPro using different domain loss balance hyperparameter α on the validation set. We report the MRR, which aligns with the monitoring metric used for early stopping.

knowledge consistently enhances the tuning techniques. Specifically, the average Hit@1 increases by 9.8% for Fine-Tuning and 4.0% for the Adapter-Tuning. Secondly, the domain-shared knowledge learned from one target domain dataset can be transferred to another. The transfer learning results shown in Table 7 indicate that all transfer settings outperform the single dataset setting (-CD), but not to a greater extent than TaxoPro. This demonstrates the potential for the efficient migration of learned domain-shared knowledge to another target dataset and validates the effectiveness of TaxoPro in augmenting the target dataset.

6.4.2 Discussions of Key Hyperparameters

In this section, we first calibrate the domain loss balance hyperparameter α on the validation set. Drawing from the results shown in Figure 3, we explore the following question.

Q9. What is the optimal domain loss balance hyperparameter α for TaxoPro? This hyperparameter modulates the impact of training samples from different domains on the shared matrices. Optimal performance is achieved at $\alpha = 1.0$, where equal contributions from both domains enhance the shared matrices' ability to retain shareable knowledge. When $\alpha > 1.0$, performance declines as the target domain's influence becomes too dominant, making the result tend to that of using the target dataset only (w/o CD). Conversely, when $\alpha < 1.0$, performance slightly drops within a certain range, but significantly deteriorates if the value is too small. This indicates that excessive influence from the source domain hampers the effective filtering of interfering knowledge by loss function of the target domain.



(a) Science

Figure 4: The results of vanilla LoRA-tuned TacoPrompt+Joint (TacoPrompt+TaxoPro w/o CDKD) using different domain loss balance hyperparameter α. Please refer to Appendix A.5 for results on all datasets.

We also examine the impact of the hyperparameter α in methods without knowledge decomposition, specifically the +Joint variants. Using vanilla LoRA-tuned TacoPrompt+Joint (TacoPrompt+TaxoPro w/o CDKD) as an example, we address the following question based on the results in Figure 4.

Q10. What is the impact of the hyperparameter α in +Joint variants? For variants without knowledge decomposition, as α increases beyond 1.0, the decline in Hit@1 and the improvement in Recall@5/10 brought by using additional high-resource taxonomies diminish, eventually converging to the performance of using only the target low-resource taxonomy (w/o CD). Conversely, when α decreases below 1.0, both metrics decline, ultimately converging to the performance of testing on the low-resource dataset after training with only the high-resource dataset. Similarly, we set $\alpha = 1.0$ for all +Joint variants, achieving the best overall performance.

Furthermore, we present the sensitive analysis of the auxiliary loss function hyperparameters, λ_1 for $\mathcal{L}_{\text{push}}$ and λ_2 for $\mathcal{L}_{\text{pull}}$, as shown in Figure 5, and analyze the related issue.

Q11. What is the sensitivity of TaxoPro to the auxiliary loss function weight hyperparameters λ_1 and λ_2? TaxoPro demonstrates robustness to λ_1 and λ_2 within a certain range. Excessively large values of λ_2 result in diminished performance, as

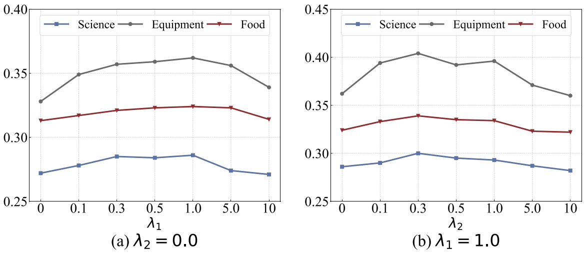


Figure 5: The MRR results of TaxoPro on validation sets, utilizing different auxiliary loss function weight hyperparameters: λ_1 for $\mathcal{L}_{\text{push}}$ and λ_2 for $\mathcal{L}_{\text{pull}}$.

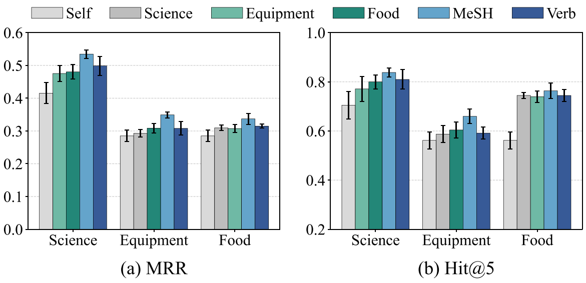


Figure 6: The results of TaxoPro using taxonomies varying in domains and scales as the source on three datasets. For "Self", we train the model only with the target dataset. The results are the average of five runs.

increasing λ_2 makes TaxoPro increasingly resemble the +Joint setting, where only the shared LoRA module is employed.

Then, we leverage datasets varying from domains and scales as the source domain dataset. Based on the results depicted in Figure 6, we investigate the following question.

Q12. What kind of taxonomy is the best choice for the source domain? Our preliminary analysis suggests two potential characteristics that may influence a taxonomy's suitability as a source domain. First, larger taxonomies may lead to performance improvements, as indicated by the observed gains from the large-scale MeSH and Verb compared to smaller taxonomies. Second, taxonomies with richer semantics could yield better performance. For instance, MeSH shows slightly better results than Verb, despite both being large-scale, which might be attributed to its richer semantic content. Based on our current findings, we hypothesize that a large-scale taxonomy rich in semantic information could be an ideal candidate for the source domain.

We further study the influence of the rank r of low-rank matrices in our framework. In addition, we replace LoRA with Prompt Tuning (Lester et al., 2021) to investigate the effect of the PEFT technique choice. Based on the results depicted in Figure 7, we discuss the questions below.

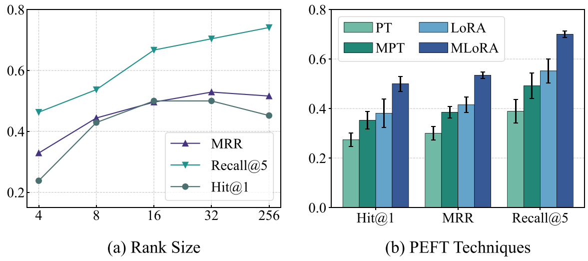


Figure 7: The results of TaxoPro using different PEFT-related hyperparameters on the Science dataset. We discuss the effect of LoRA's rank r in (a) and that of the PEFT technique choice in (b).

Q13. What is the effect of the rank r in the framework? Generally, a higher rank yields better results, as evidenced by the positive correlation between the Recall@5 metric and rank size. However, an increase in rank beyond a certain threshold results in a decrease in Hit@1. For instance, H@1 decreases when the rank increases from 32 to 256. This may be due to the insufficient training samples in the target low-resource dataset for domain-specific knowledge learning with high-rank matrices. Therefore, it is essential to choose an appropriate rank within a specific range. Our experiments indicate that a rank of 32 provides an optimal balance across all performance metrics.

Q14. What is the effect of the backbone PEFT technique? In line with previous research (He et al., 2022), LoRA outperforms Prompt Tuning in the task of taxonomy completion when only using the training samples from the target dataset. This pattern also applies to the proposed CDKD module, since LoRA surpasses Prompt Tuning as the knowledge decomposition technique. Hence, LoRA is a suitable PEFT choice for TaxoPro.

7 Conclusion

In this paper, we propose TaxoPro, a LoRA-based plug-in cross-domain method. It leverages shareable knowledge from the high-resource taxonomy to enhance PLM-based techniques in low-resource taxonomy completion. We decompose cross-domain knowledge into domain-shared and domain-specific parts, storing them with the low-rank matrices to avoid negative interference.

Two auxiliary losses direct the flow of shareable knowledge. Experiments prove TaxoPro's effectiveness. We believe our initial exploration of cross-domain taxonomy completion presents an interesting direction for the community.

8 Limitations and Future Work

Our method currently has two main limitations: (i) it relies on a single source taxonomy to enhance low-resource taxonomy completion, and (ii) training with all samples from a single high-resource taxonomy can be computationally expensive. We plan to extend TaxoPro to support multiple source taxonomies and investigate more efficient sampling techniques to alleviate the computational burden. Additionally, we aim to evaluate the effectiveness of TaxoPro on other tasks that require knowledge transfer.

Acknowledgments

We sincerely thank the anonymous reviewers for their rigorous and conscientious review, as well as their meticulous and insightful suggestions that greatly improved the quality of this work. We are also deeply grateful to the action editors, Hoifung Poon and Tao Ge, for their exacting editorial oversight, and constructive guidance throughout the review process that significantly strengthened the manuscript. We also thank Yuxun Qu and Yuxiao Liu for their helpful discussions during the research. Their questions and ideas during our meetings helped us clarify key points and solve several challenging problems. Additionally, I extend heartfelt appreciation to my close friend Ziheng Xiao for his unwavering support throughout this research endeavor. This research is supported by the National Natural Science Foundation of China (No. 62372252, 72342017), National Engineering Research Center for Digital Construction and Evaluation Technology of Urban Rail Transit, Development of a platform for quantity statistics and budget preparation of urban rail transit projects based on big data analysis (No. 2022A02158007).

References

Ines Arous, Ljiljana Dolamic, and Philippe Cudré-Mauroux. 2023. TaxoComplete: Self-supervised taxonomy completion leveraging position-enhanced semantic matching. In *WWW*, pages 2509–2518. https://doi.org/10.1145/3543507.3583342

He Bai, Tong Wang, Alessandro Sordoni, and Peng Shi. 2022. Better language model with hypernym class prediction. In *ACL*, pages 1352–1362. https://doi.org/10.18653/v1/2022.acl-long.96

Eyal Ben-David, Carmel Rabinovitz, and Roi Reichart. 2020. PERL: Pivot-based domain adaptation for pre-trained deep contextualized embedding models. *TACL*, 8:504–521. https://doi.org/10.1162/tacl_a_00328

Georgeta Bordea, Paul Buitelaar, Stefano Faralli, and Roberto Navigli. 2015. SemEval-2015 Task 17: Taxonomy extraction evaluation (texeval). In *SemEval@NAACL-HLT*, pages 902–910. https://doi.org/10.18653/v1/S15-2151

Konstantinos Bousmalis, George Trigeorgis, Nathan Silberman, Dilip Krishnan, and Dumitru Erhan. 2016. Domain separation networks. In *NeurIPS*, pages 343–351.

Shoufa Chen, Chongjian Ge, Zhan Tong, Jiangliu Wang, Yibing Song, Jue Wang, and Ping Luo. 2022. AdaptFormer: Adapting vision transformers for scalable visual recognition. In *NeurIPS*, pages 1–15.

Sijie Cheng, Zhouhong Gu, Bang Liu, Rui Xie, Wei Wu, and Yanghua Xiao. 2022. Learning what you need from what you did: Product taxonomy expansion with user behaviors supervision. In *ICDE*, pages 3280–3293. https://doi.org/10.1109/ICDE53745.2022.00310

Hal Daumé. 2007. Frustratingly easy domain adaptation. In *ACL*, pages 256–263.

Tim Dettmers, Artidoro Pagnoni, Ari Holtzman, and Luke Zettlemoyer. 2023. QLoRA: Efficient finetuning of quantized llms. In *NeurIPS*, pages 1–28.

Shizhe Diao, Tianyang Xu, Ruijia Xu, Jiawei Wang, and Tong Zhang. 2023. Mixture-of-Domain-Adapters: Decoupling and injecting domain knowledge to pre-trained language models' memories. In *ACL*, pages 5113–5129. https://doi.org/10.18653/v1/2023.acl-long.280

Ning Ding, Yujia Qin, Guang Yang, Fuchao Wei, Zonghan Yang, Yusheng Su, Shengding Hu, Yulin Chen, Chi-Min Chan, Weize Chen, Jing Yi, Weilin Zhao, Xiaozhi Wang, Zhiyuan Liu, Hai-Tao Zheng, Jianfei Chen, Yang Liu, Jie Tang, Juanzi Li, and Maosong Sun. 2023. Parameter-efficient fine-tuning of large-scale pre-trained language models. *NMI*, 5(3):220–235. https://doi.org/10.1038/S42256-023-00626-4

Qianjin Du, Shiji Zhou, Xiaohui Kuang, Gang Zhao, and Jidong Zhai. 2023. Joint geometrical and statistical domain adaptation for cross-domain code vulnerability detection. In *EMNLP*, pages 12791–12800. https://doi.org/10.18653/v1/2023.emnlp-main.788

Suchin Gururangan, Ana Marasovic, Swabha Swayamdipta, Kyle Lo, Iz Beltagy, Doug Downey, and Noah A. Smith. 2020. Don't stop pretraining: Adapt language models to domains and tasks. In *ACL*, pages 8342–8360. https://doi.org/10.18653/v1/2020.acl-main.740

Junxian He, Chunting Zhou, Xuezhe Ma, Taylor Berg-Kirkpatrick, and Graham Neubig. 2022. Towards a unified view of parameter-efficient transfer learning. In *ICLR*, pages 1–15.

Neil Houlsby, Andrei Giurgiu, Stanislaw Jastrzebski, Bruna Morrone, Quentin de Laroussilhe, Andrea Gesmundo, Mona Attariyan, and Sylvain Gelly. 2019. Parameter-efficient transfer learning for NLP. In *ICML*, pages 2790–2799.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2022a. LoRA: Low-rank adaptation of large language models. In *ICLR*, pages 1–26.

Shengding Hu, Ning Ding, Huadong Wang, Zhiyuan Liu, Jingang Wang, Juanzi Li, Wei Wu, and Maosong Sun. 2022b. Knowledgeable Prompt-tuning: Incorporating knowledge into prompt verbalizer for text classification. In *ACL*, pages 2225–2240.

Minhao Jiang, Xiangchen Song, Jieyu Zhang, and Jiawei Han. 2022. TaxoEnrich: Self-supervised taxonomy completion via structure-semantic representations. In *WWW*, pages 925–934. https://doi.org/10.1145/3485447.3511935

Song Jiang, Qiyue Yao, Qifan Wang, and Yizhou Sun. 2023. A single vector is not enough: Taxonomy expansion via box embeddings. In *WWW*, pages 2467–2476. https://doi.org/10.1145/3543507.3583310

David Jurgens and Mohammad Taher Pilehvar. 2016. SemEval-2016 Task 14: Semantic taxonomy enrichment. In *SemEval@NAACL-HLT 2016*, pages 1092–1102. https://doi.org/10.18653/v1/S16-1169

Giannis Karamanolakis, Jun Ma, and Xin Luna Dong. 2020. TXtract: Taxonomy-aware knowledge extraction for thousands of product categories. In *ACL*, pages 8489–8502. https://doi.org/10.18653/v1/2020.acl-main.751

Brian Lester, Rami Al-Rfou, and Noah Constant. 2021. The power of scale for parameter-efficient prompt tuning. In *EMNLP*, pages 3045–3059. https://doi.org/10.18653/v1/2021.emnlp-main.243

Xiang Lisa Li and Percy Liang. 2021. Prefix-Tuning: Optimizing continuous prompts for generation. In *ACL/IJCNLP*, pages 4582–4597.

Carolyn E. Lipscomb. 2000. Medical subject headings (MeSH). *Bulletin of the Medical Library Association*, page 265.

Zichen Liu, Hongyuan Xu, Yanlong Wen, Ning Jiang, Haiying Wu, and Xiaojie Yuan. 2021. TEMP: Taxonomy expansion with dynamic margin loss through taxonomy-paths. In *EMNLP*, pages 3854–3863. https://doi.org/10.18653/v1/2021.emnlp-main.313

Mingyu Derek Ma, Muhao Chen, Te-Lin Wu, and Nanyun Peng. 2021. HyperExpan: Taxonomy expansion with hyperbolic representation learning. In *EMNLP*, pages 4182–4194. https://doi.org/10.18653/v1/2021.findings-emnlp.353

Emaad Manzoor, Rui Li, Dhananjay Shrouty, and Jure Leskovec. 2020. Expanding taxonomies with implicit edge semantics. In *WWW*, pages 2044–2054. https://doi.org/10.1145/3366423.3380271

Yuning Mao, Lambert Mathias, Rui Hou, Amjad Almahairi, Hao Ma, Jiawei Han, Scott Yih, and Madian Khabsa. 2022.

UniPELT: A unified framework for parameter-efficient language model tuning. In *ACL*, pages 6253–6264. https://doi.org/10.18653/v1/2022.acl-long.433

Yuan Meng, Songlin Zhai, Zhihua Chai, Yuxin Zhang, Tianxing Wu, Guilin Qi, and Wei Song. 2024. Which is better? Taxonomy induction with learning the optimal structure via contrastive learning. *KBS*, 304:112405. https://doi.org/10.1016/j.knosys.2024.112405

Sahil Mishra, Ujjwal Sudev, and Tanmoy Chakraborty. 2024. FLAME: Self-supervised low-resource taxonomy expansion using large language models. *CoRR*, abs/2402.13623v1. https://doi.org/10.1145/3709007

Viktor Moskvoretskii, Ekaterina Neminova, Alina Lobanova, Alexander Panchenko, and Irina Nikishina. 2024. TaxoLLaMA: Wordnet-based model for solving multiple lexical sematic tasks. In *ACL*, pages 2331–2350. https://doi.org/10.18653/v1/2024.acl-long.127

Yuhang Niu, Hongyuan Xu, Ciyi Liu, Yanlong Wen, and Xiaojie Yuan. 2024. Contrastive representation learning for self-supervised taxonomy completion. In *IJCAI*, pages 6442–6450. https://doi.org/10.24963/ijcai.2024/712

Jonas Pfeiffer, Ivan Vulic, Iryna Gurevych, and Sebastian Ruder. 2020. MAD-X: An adapter-based framework for multi-task cross-lingual transfer. In *EMNLP*, pages 7654–7673. https://doi.org/10.18653/v1/2020.emnlp-main.617

Bornali Phukon, Anasua Mitra, Sanasam Ranbir Singh, and Priyankoo Sarmah. 2022. TEAM: A multitask learning based taxonomy expansion approach for attach and merge. In *NAACL Findings*, pages 366–378. https://doi.org/10.18653/v1/2022.findings-naacl.28

Zeng Qingkai, Bai Yuyang, Tan Zhaoxuan, Wu Zhenyu, Feng Shangbin, and Meng Jiang. 2024. CodeTaxo: Enhancing taxonomy expansion with limited examples via code language prompts. *CoRR*, abs/2408.09070v1. https://doi.org/10.48550/arXiv.2408.09070

Gita Sarafraz, Armin Behnamnia, Mehran Hosseinzadeh, Ali Balapour, Amin Meghrazi, and Hamid R. Rabiee. 2024. Domain adaptation and generalization of functional medical data: A systematic survey of brain data. *ACM Computing Surveys*, 56(10):255. https://doi.org/10.1145/3654664

Jiaming Shen, Zhihong Shen, Chenyan Xiong, Chi Wang, Kuansan Wang, and Jiawei Han. 2020. TaxoExpan: Self-supervised taxonomy expansion with position-enhanced graph neural network. In *WWW*, pages 486–497. https://doi.org/10.1145/3366423.3380132

Yanzhen Shen, Yu Zhang, Yunyi Zhang, and Jiawei Han. 2024. A unified taxonomy-guided instruction tuning framework for entity set expansion and taxonomy expansion. *CoRR*, abs/2402.13405v1. https://doi.org/10.48550/ARXIV.2402.13405

Jingchuan Shi, Hang Dong, Jiaoyan Chen, Zhe Wu, and Ian Horrocks. 2024. Taxonomy completion via implicit concept insertion. In *WWW*, pages 2159–2169. https://doi.org/10.1145/3589334.3645584

Kai Sun, Jifan Yu, Juanzi Li, and Lei Hou. 2024. Exploring sequence-to-sequence taxonomy expansion via language model probing. *ESWA*, 239:122321. https://doi.org/10.1016/J.ESWA.2023.122321

Kunihiro Takeoka, Kosuke Akimoto, and Masafumi Oyamada. 2021. Low-resource taxonomy enrichment with pretrained language models. In *EMNLP*, pages 2747–2758. https://doi.org/10.18653/v1/2021.emnlp-main.217

Jindong Wang, Cuiling Lan, Chang Liu, Yidong Ouyang, Tao Qin, Wang Lu, Yiqiang Chen, Wenjun Zeng, and Philip S. Yu. 2023a. Generalizing to unseen domains: A survey on domain generalization. *TKDE*, 35(8):8052–8072. https://doi.org/10.1109/TKDE.2022.3178128

Shanshan Wang, Yiyang Chen, Zhenwei He, Xun Yang, Mengzhu Wang, Quanzeng You, and Xingyi Zhang. 2023b. Disentangled representation learning with causality for unsupervised domain adaptation. In *ACM MM*, pages 2918–2926. https://doi.org/10.1145/3581783.3611725

Suyuchen Wang, Ruihui Zhao, Xi Chen, Yefeng Zheng, and Bang Liu. 2021. Enquire one's parent and child before decision: Fully exploit hierarchical structure for self-supervised taxonomy expansion. In *WWW*, pages 3291–3304. https://doi.org/10.1145/3442381.3449948

Suyuchen Wang, Ruihui Zhao, Yefeng Zheng, and Bang Liu. 2022a. QEN: Applicable taxonomy completion via evaluating full taxonomic relations. In *WWW*, pages 1008–1017. https://doi.org/10.1145/3485447.3511943

Yaqing Wang, Sahaj Agarwal, Subhabrata Mukherjee, Xiaodong Liu, Jing Gao, Ahmed Hassan Awadallah, and Jianfeng Gao. 2022b. AdaMix: Mixture-of-adaptations for parameter-efficient model tuning. In *EMNLP*, pages 5744–5760. https://doi.org/10.18653/v1/2022.emnlp-main.388

Zhen Wang, Rameswar Panda, Leonid Karlinsky, Rogério Feris, Huan Sun, and Yoon Kim. 2023c. Multitask prompt tuning enables parameter-efficient transfer learning. In *ICLR*, pages 1–16.

Pengfei Wei, Lingdong Kong, Xinghua Qu, Yi Ren, Zhiqiang Xu, Jing Jiang, and Xiang Yin. 2023. Unsupervised video domain adaptation for action recognition: A disentanglement perspective. In *NeurIPS*, pages 1–20.

Fei Xia, Yixuan Weng, Shizhu He, Kang Liu, and Jun Zhao. 2023. Find parent then label children: A two-stage taxonomy completion method with pre-trained language model. In *EACL*, pages 1032–1042. https://doi.org/10.18653/v1/2023.eacl-main.73

Fred Xu, Song Jiang, Zijie Huang, Xiao Luo, Shichang Zhang, Yuanzhou Chen, and Yizhou Sun. 2024. FUSE: Measure-theoretic compact fuzzy set representation for taxonomy expansion. In *ACL-Findings*, pages 2707–2720. https://doi.org/10.18653/v1/2024.findings-acl.158

Hongyuan Xu, Yunong Chen, Zichen Liu, Yanlong Wen, and Xiaojie Yuan. 2022. TaxoPrompt: A prompt-based generation method with taxonomic context for self-supervised taxonomy expansion. In *IJCAI*, pages 4432–4438. https://doi.org/10.24963/ijcai.2022/615

Hongyuan Xu, Ciyi Liu, Yuhang Niu, Yunong Chen, Xiangrui Cai, Yanlong Wen, and Xiaojie Yuan. 2023. TacoPrompt: A collaborative multi-task prompt learning method for self-supervised taxonomy completion. In *EMNLP*, pages 15804–15817. https://doi.org/10.18653/v1/2023.emnlp-main.979

Wei Xue, Yongliang Shen, Wenqi Ren, Jietian Guo, Shiliang Pu, and Weiming Lu. 2024. Insert or attach: Taxonomy completion via box embedding. In *ACL*, pages 3851–3863. https://doi.org/10.18653/v1/2024.acl-long.212

Yue Yu, Yinghao Li, Jiaming Shen, Hao Feng, Jimeng Sun, and Chao Zhang. 2020. STEAM: Self-supervised taxonomy expansion with mini-paths. In *SIGKDD*, pages 1026–1035. https://doi.org/10.1145/3394486.3403145

Qingkai Zeng, Yuyang Bai, Zhaoxuan Tan, Shangbin Feng, Zhenwen Liang, Zhihan Zhang, and Meng Jiang. 2024. Chain-of-Layer: Iteratively prompting large language models for taxonomy induction from limited examples. In *CIKM*, pages 3093–3102. https://doi.org/10.1145/3627673.3679608

Qingkai Zeng, Jinfeng Lin, Wenhao Yu, Jane Cleland-Huang, and Meng Jiang. 2021. Enhancing taxonomy completion with concept generation via fusing relational representations. In *SIGKDD*, pages 2104–2113. https://doi.org/10.1145/3447548.3467308

Songlin Zhai, Weiqing Wang, Yuan-Fang Li, and Yuan Meng. 2023. DNG: Taxonomy expansion by exploring the intrinsic directed structure on non-gaussian space. In *AAAI*, pages 6593–6601. https://doi.org/10.1609/aaai.v37i5.25810

Jinghan Zhang, Shiqi Chen, Junteng Liu, and Junxian He. 2023a. Composing parameter-efficient modules with arithmetic operation. In *NeurIPS*, pages 1–22.

Jieyu Zhang, Xiangchen Song, Ying Zeng, Jiaze Chen, Jiaming Shen, Yuning Mao, and Lei Li. 2021. Taxonomy completion via triplet matching network. In *AAAI*, pages 4662–4670. https://doi.org/10.1609/aaai.v35i5.16596

Kaihang Zhang, Kai Shuang, Xinyue Yang, Xuyang Yao, and Jinyu Guo. 2023b. What is overlap knowledge in event argument extraction? APE: A cross-datasets transfer learning model for EAE. In *ACL*, pages 393–409. https://doi.org/10.18653/v1/2023.acl-long.24

Lei Zhang and Xinbo Gao. 2024. Transfer Adaptation Learning: A decade survey. *TNNLS*, 35(1):23–44. https://doi.org/10.1109/TNNLS.2022.3183326, PubMed: 35727786

Qingru Zhang, Minshuo Chen, Alexander Bukharin, Pengcheng He, Yu Cheng, Weizhu Chen, and Tuo Zhao. 2023c. Adaptive budget allocation for parameter-efficient fine-tuning. In *ICLR*, pages 1–17.

Tinghui Zhu, Jingping Liu, Jiaqing Liang, Haiyun Jiang, Yanghua Xiao, Zongyu Wang, Rui Xie, and Yunsen Xian. 2023. Towards visual taxonomy expansion. In *ACM MM*, pages 6481–6490. https://doi.org/10.1145/3581783.3613845

A Appendix

A.1 Training Time Comparsion

Datasets	**Science**	**Equipment**	**Food**
TEMP	0.33	0.32	1.07
TEMP+Joint	14.38	14.13	14.90
TEMP+TaxoPro	36.20	34.70	36.92
w/o $\mathcal{L}_{\text{pull}}$, $\mathcal{L}_{\text{push}}$	15.97	15.70	15.98
w/o $\mathcal{L}_{\text{push}}$	26.45	25.78	26.62
w/o $\mathcal{L}_{\text{pull}}$	26.15	25.15	26.22
TacoPrompt	0.47	0.45	1.42
TacoPrompt+Joint	20.15	19.95	19.45
TacoPrompt+TaxoPro	60.05	51.05	56.47
w/o $\mathcal{L}_{\text{pull}}$, $\mathcal{L}_{\text{push}}$	21.42	20.70	20.98
w/o $\mathcal{L}_{\text{push}}$	40.12	37.68	38.98
w/o $\mathcal{L}_{\text{pull}}$	40.88	36.78	38.43

Table 8: Training time (minutes) per epoch using a single RTX 4090 GPU on three datasets.

A.2 Inference Time Comparsion

Datasets	**Science**	**Equipment**	**Food**
TEMP	0.52	0.50	6.20
TEMP+Joint	0.55	0.50	8.12
TEMP+TaxoPro	0.67	0.62	8.67
TacoPrompt	1.62	1.35	19.73
TacoPrompt+Joint	1.68	1.38	19.83
TacoPrompt+TaxoPro	2.03	1.7	24.82

Table 9: Total inference time (minutes) utilizing a single RTX 4090 GPU device. Note that auxiliary loss functions $\mathcal{L}_{\text{pull}}$ and $\mathcal{L}_{\text{push}}$ are only active during training and do not affect inference time.

A.3 Implementation Details

We use BERT[1] as the backbone language model for fair comparison with other methods. The model is trained using the AdamW optimizer, with a learning rate of 1e-4 and an accumulation step of 4. Hyperparameters λ_1, λ_2, rank, and scaling rate s are set to 1.0, 0.3, 32, and 1.0, respectively across all datasets. The domain loss balance hyperparameter α is set to 1.0 for all datasets. We sample 15 negative positions per training instance. The batch size $2B$ is set to 2. The high-resource taxonomy determines the batch steps per epoch. Model convergence is monitored through validation MRR trajectories, terminating training upon detecting five-epoch plateaus. Then the best checkpoint is deployed to the test set. For baselines, we follow the experimental settings provided by Xu et al. (2023).[2] In the Baseline+Joint experiments, we sample an equal number of training instances from both the source and target domains within each batch. All experiments were conducted using NVIDIA RTX 4090 GPU devices.

A.4 Complete Taxonomy Completion Performance Comparison

Table 10 provides comprehensive results comparing the taxonomy completion performance across three method categories: (1) baseline approaches, (2) their +Joint variants, and (3) the +TaxoPro variants of PLM-based techniques, namely TEMP and TacoPrompt.

[1] https://huggingface.co/bert-base-uncased.

[2] https://github.com/cyclexu/TacoPrompt/tree/main/Baselines.

Method	MR↓	MRR	Recall@1	Recall@5	Recall@10	Hit@1	Hit@5	Hit@10
				Science				
TaxoExpan	215.1±2.6	0.118±0.005	10.5±1.5	11.7±0.8	11.7±0.8	13.3±1.9	14.8±1.0	14.8±1.0
TaxoExpan+Joint	126.5±28.5	0.240±0.032	19.3±3.2	28.7±4.1	34.7±3.9	24.3±4.1	36.2±5.1	43.3±4.1
Arborist	81.4±2.0	0.254±0.013	23.0±1.8	26.4±1.2	26.8±1.4	29.1±2.3	33.3±1.5	33.8±1.8
Arborist+Joint	67.3±2.7	0.246±0.015	20.8±1.7	26.4±0.0	30.6±1.9	26.2±2.1	33.3±0.0	37.6±2.8
TMN	72.2±4.1	0.265±0.020	21.5±3.9	29.8±2.2	32.5±2.2	27.1±4.9	37.6±2.8	41.0±2.8
TMN+Joint	52.3±3.2	0.298±0.018	24.1±2.2	33.6±2.2	37.3±2.2	30.5±2.8	42.4±2.8	47.1±2.8
TaxoEnrich	36.1±4.6	0.355±0.020	29.1±2.3	41.9±2.8	47.6±2.2	36.7±2.9	52.8±3.5	59.0±2.8
TaxoEnrich+Joint	31.4±4.2	0.306±0.019	22.2±1.8	36.2±3.2	45.7±4.4	28.1±2.4	45.2±4.0	56.2±3.9
QEN	146.0±35.1	0.279±0.024	20.0±3.0	36.7±4.0	40.0±2.8	25.7±3.8	47.2±5.1	51.0±2.9
QEN+Joint	58.4±23.1	0.339±0.037	24.1±3.5	43.3±5.3	50.0±4.8	31.0±4.5	53.3±6.3	57.6±4.4
TaxoComplete	52.3±4.0	0.377±0.017	25.9±1.7	56.3±1.9	69.3±1.9	33.3±2.1	64.8±1.8	76.2±1.5
TaxoComplete+Joint	46.7±14.9	0.388±0.037	27.8±3.9	56.1±5.1	65.6±4.3	35.7±5.0	62.8±7.1	72.4±3.9
Musubu	16.4±9.9	0.337±0.024	21.8±2.9	48.9±4.8	62.3±3.6	28.1±3.8	61.4±6.3	75.3±4.4
Musubu+Joint	116.1±9.1	0.356±0.023	21.1±1.9	56.3±4.8	68.2±3.6	27.2±2.4	65.7±3.9	74.8±3.3
CoSTC	17.1±1.6	0.290±0.003	15.0±0.4	43.6±1.2	59.4±3.0	35.2±1.0	70.0±1.1	81.4±2.3
CoSTC+Joint	15.0±3.7	0.286±0.013	13.1±1.9	45.3±2.1	64.2±2.3	31.0±4.5	74.7±3.2	86.7±3.3
TEMP	19.9±4.8	0.425±0.021	29.2±4.0	57.8±0.8	66.7±2.6	37.6±5.1	74.3±1.0	84.8±2.4
TEMP+Joint	13.5±7.2	0.391±0.039	21.1±4.9	61.1±2.3	73.7±1.4	27.1±6.3	76.7±2.8	88.1±1.5
TEMP+TaxoPro	11.6±5.2 ↑	0.485±0.024 ↑	36.3±1.9 ↑	63.3±2.2 ↑	75.5±3.4 ↑	46.7±2.4 ↑	79.5±2.4 ↑	**90.9**±3.8 ↑
TacoPrompt	16.4±9.9	0.456±0.027	32.9±3.8	59.3±3.1	70.7±3.6	42.4±4.9	74.3±2.8	85.2±1.0
TacoPrompt+Joint	12.2±7.7	0.462±0.030	30.4±5.8	64.8±3.5	75.2±1.9	39.1±7.4	79.5±3.6	86.2±2.4
TacoPrompt+TaxoPro	**6.3**±1.1 ↑	**0.535**±0.013 ↑	**39.3**±2.4 ↑	**70.0**±1.4 ↑	**78.5**±1.9 ↑	**50.0**±3.0 ↑	**83.8**±1.8 ↑	90.0±3.2 ↑
				Equipment				
TaxoExpan	275.3±5.4	0.073±0.003	4.3±0.0	9.2±1.1	12.0±1.7	6.4±0.0	13.6±1.7	17.9±2.5
TaxoExpan+Joint	178.7±107.5	0.227±0.030	15.4±2.1	28.0±2.8	36.6±3.4	22.9±3.1	41.3±3.5	52.8±3.1
Arborist	50.5±1.5	0.258±0.006	21.1±0.5	27.1±0.9	29.2±0.7	31.5±0.8	38.3±1.3	41.3±1.1
Arborist+Joint	38.3±3.4	0.319±0.017	22.0±1.9	38.3±3.3	41.7±4.5	32.8±2.9	49.8±1.1	53.2±3.8
TMN	53.4±2.0	0.262±0.011	19.7±1.0	30.3±1.7	35.7±2.6	29.4±1.6	43.0±2.5	49.8±2.9
TMN+Joint	40.5±6.6	0.305±0.017	22.0±2.5	34.6±1.0	42.3±1.9	32.8±3.7	47.2±1.6	54.0±3.5
TaxoEnrich	74.0±8.6	0.264±0.033	18.6±1.0	34.3±2.4	39.4±2.6	27.6±5.7	51.1±3.5	57.0±3.1
TaxoEnrich+Joint	65.9±11.9	0.286±0.019	21.2±2.1	35.7±1.8	40.3±1.4	31.5±3.1	51.5±2.5	57.8±2.1
QEN	171.4±32.2	0.158±0.033	10.1±4.1	19.4±3.4	25.3±3.6	15.3±6.2	28.9±4.8	35.7±2.5
QEN+Joint	99.5±21.8	0.243±0.014	15.8±2.1	31.8±3.7	42.5±4.5	23.8±3.1	45.5±3.7	52.8±3.9
TaxoComplete	144.2±7.5	0.295±0.005	17.5±0.7	40.3±0.7	52.1±1.5	26.4±1.0	47.7±1.0	58.7±2.2
TaxoComplete+Joint	122.0±29.9	0.291±0.021	16.6±2.2	44.3±4.9	56.9±1.5	25.1±3.4	51.9±3.7	62.1±3.1
Musubu	130.6±14.1	0.301±0.017	17.5±2.6	43.4±1.9	57.5±2.9	26.4±3.9	53.6±2.1	63.0±4.0
Musubu+Joint	117.8±11.3	0.281±0.062	15.5±6.0	42.8±8.1	58.3±5.3	23.4±9.1	51.9±8.6	64.7±4.2
CoSTC	60.8±3.7	0.278±0.014	15.5±0.6	41.3±2.9	54.6±4.6	24.7±1.0	54.9±2.8	64.2±2.1
CoSTC+Joint	41.3±8.2	0.306±0.021	18.7±2.9	42.9±1.8	59.1±2.8	29.8±4.7	58.7±4.2	69.8±2.5
TEMP	92.7±13.7	0.290±0.027	16.6±3.6	42.5±1.7	55.5±3.6	25.1±5.5	58.7±2.2	68.5±2.4
TEMP+Joint	72.9±6.6	0.291±0.038	15.8±4.9	44.2±3.3	57.2±1.9	23.8±7.4	60.4±4.0	69.4±2.1
TEMP+TaxoPro	68.4±4.1 ↑	0.331±0.020 ↑	18.3±2.7 ↑	50.5±1.6 ↑	62.3±3.5 ↑	27.7±4.0 ↑	63.8±1.9 ↑	71.1±3.9 ↑
TacoPrompt	65.3±38.0	0.288±0.008	16.9±2.0	41.1±3.1	57.7±3.1	25.5±3.0	56.6±3.9	67.7±4.1
TacoPrompt+Joint	69.4±11.8	0.285±0.016	15.5±2.0	44.5±1.1	59.7±1.5	23.4±3.0	60.4±2.6	68.9±2.1
TacoPrompt+TaxoPro	**34.7**±12.5 ↑	**0.349**±0.009 ↑	**22.2**±1.0 ↑	**51.5**±1.7 ↑	**63.1**±3.3 ↑	**33.6**±1.6 ↑	**66.0**±3.0 ↑	**72.8**±1.6 ↑
				Food				
TaxoExpan	593.3±128.9	0.105±0.013	7.6±0.8	12.7±2.2	15.9±2.4	15.3±1.6	25.1±4.1	30.8±4.7
TaxoExpan+Joint	403.0±171.4	0.129±0.014	8.8±1.5	16.3±1.7	20.1±1.9	17.8±3.1	31.8±3.5	38.2±3.0
Arborist	247.9±7.6	0.142±0.007	10.2±0.6	16.8±0.5	21.3±0.8	20.8±1.3	32.6±1.1	38.4±1.4
Arborist+Joint	205.4±4.9	0.169±0.006	12.4±0.8	20.5±0.9	25.8±2.1	25.1±1.6	38.4±1.6	44.1±2.1
TMN	147.7±7.6	0.153±0.006	10.5±0.9	18.1±0.9	23.4±1.2	21.2±1.9	35.1±2.0	42.2±2.6
TMN+Joint	143.5±3.8	0.148±0.010	9.3±0.8	18.1±1.9	25.1±2.5	18.9±1.6	34.6±4.1	44.9±4.7
TaxoEnrich	216.5±23.6	0.169±0.006	10.3±0.6	22.9±1.2	29.3±1.2	20.8±1.2	42.7±2.2	54.9±1.5
TaxoEnrich+Joint	198.8±22.7	0.175±0.008	10.3±0.9	24.8±0.9	30.9±0.8	20.8±1.8	45.7±1.6	55.8±0.7
QEN	301.4±22.1	0.220±0.013	15.5±1.4	28.0±1.6	32.7±1.6	32.6±2.8	52.0±2.8	58.1±1.9
QEN+Joint	173.7±25.9	0.248±0.021	16.3±1.8	32.5±2.3	41.4±2.6	34.3±3.7	59.0±3.7	68.9±3.0
TaxoComplete	416.9±4.9	0.258±0.005	18.8±0.7	31.4±0.4	40.3±0.7	39.6±1.4	58.6±0.8	65.0±0.7
TaxoComplete+Joint	385.0±31.2	0.271±0.019	18.7±1.5	34.1±1.4	42.9±2.0	39.3±3.1	60.7±2.0	66.8±1.8
Musubu	504.9±52.9	0.213±0.018	12.9±1.6	28.0±1.8	38.8±2.5	27.2±3.4	48.6±2.5	61.1±3.5
Musubu+Joint	543.9±62.0	0.183±0.023	10.2±2.0	24.9±3.2	35.9±2.8	21.5±4.2	43.9±5.5	57.2±4.3
CoSTC	69.9±18.5	0.224±0.024	11.1±2.1	35.9±2.8	45.6±2.4	21.1±4.0	60.7±4.2	70.9±2.1
CoSTC+Joint	72.6±5.5	0.263±0.011	17.8±5.0	40.2±1.0	51.3±1.6	25.7±2.8	65.5±1.5	75.5±1.8
TEMP	66.7±12.4	0.288±0.011	19.8±1.2	36.7±2.3	46.1±1.8	41.6±2.6	69.6±3.5	78.9±2.1
TEMP+Joint	**53.3**±10.7	0.290±0.004	19.3±1.0	37.9±0.9	46.3±1.8	40.6±2.0	71.3±1.3	79.3±1.2
TEMP+TaxoPro	75.4±17.7 ↓	0.320±0.009 ↑	23.1±1.0 ↑	40.5±1.2 ↑	47.6±1.2 ↑	48.5±2.2 ↑	75.7±1.9 ↑	81.4±1.2 ↑
TacoPrompt	114.3±27.1	0.304±0.006	20.7±0.7	39.6±0.9	50.2±1.8	43.5±1.6	73.4±0.9	81.4±1.6
TacoPrompt+Joint	138.5±33.0	0.305±0.011	19.5±1.3	41.2±1.8	51.3±1.8	41.1±2.7	73.2±2.1	81.8±0.9
TacoPrompt+TaxoPro	78.0±26.6 ↑	**0.337**±0.017 ↑	**23.7**±1.8 ↑	**43.9**±2.0 ↑	**54.0**±2.3 ↑	**49.7**±3.7 ↑	**76.3**±3.1 ↑	**81.9**±2.1 ↑

Table 10: We present experimental results on three benchmark datasets, with five-run averaged outcomes from our reproduced baselines.

A.5 Impacts of Domain Balance Hyperparameters on Model+Joint Variants

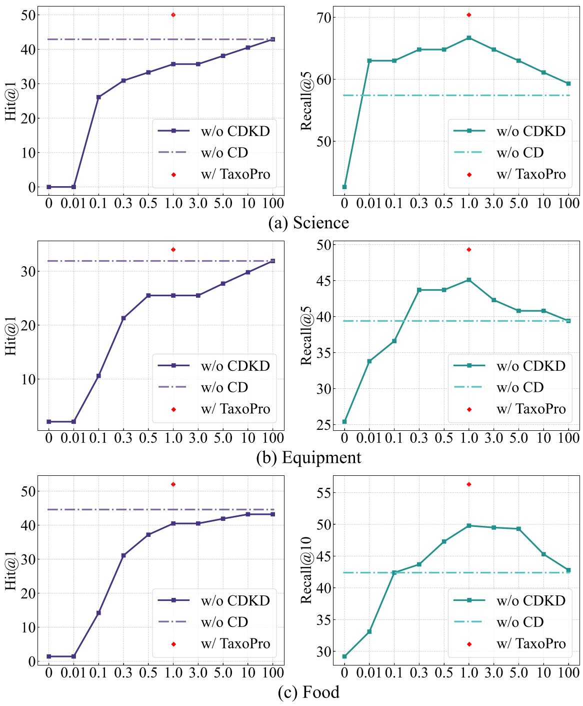


Figure 8: The results of vanilla LoRA-tuned TacoPrompt+Joint (TacoPrompt+TaxoPro w/o CDKD) using different domain loss balance hyperparameter α. For the Science, Equipment, and Food datasets, we report Hit@1, Recall@5, and Recall@10, respectively, as these metrics best capture the performance improvements from cross-domain knowledge.