Abstract

Taxonomy completion, a self-supervised task, aims to add new concepts to an existing taxonomy by attaching them to appropriate hypernym and hyponym pairs. Researchers have proposed several approaches to capture the essential relationships in taxonomy using semantic or structural information. However, they either construct training signals from a single view or simply use a random sampling strategy, making it insufficient to capture various relations in taxonomic structure and learn quality representations. To address this, we propose CoSTC, a contrastive learning framework that captures diverse relations and improves representations for taxonomy completion. It uses two contrasting views, namely intra-view and inter-view, to provide rich self-supervised signals. In intra-view contrasting, we exploit the correlations within queries and within positions by performing instance-level discrimination task. In inter-view contrasting, we use a sampling strategy that considers diversity and hardness to select representative pairs, enhancing the learning of fine-grained query-position relations. Experimental results on three datasets verify the effectiveness of our approach. Our code is available at https://github.com/nyh-a/CoSTC.

1 Introduction

Taxonomies, hierarchical knowledge graphs that organize concepts through “is-A” relations, have seen significant development due to their ability to structure knowledge in various scenarios, including product recommendation [Zhang et al., 2014] and information extraction [Karamanolakis et al., 2020]. Well-organized taxonomies are valuable resources for downstream tasks such as named entity recognition [Wang et al., 2021b], transfer learning [Luo et al., 2022] and extreme multi-label classification [Liu et al., 2017].

As new concepts are continuously emerging, how to maintain an up-to-date taxonomy becomes a fundamental issue. However, the manual maintenance of taxonomies by domain experts is laborious and time-consuming, especially with rapidly growing content on many platforms [Arous et al., 2023]. To tackle this, recent studies have focused on automatic taxonomy expansion (TE) [Shen et al., 2020; Liu et al., 2021] and completion (TC) [Zhang et al., 2021; Xu et al., 2023a], aiming to find a suitable position for new concepts (query) in the taxonomy tree. Both TE and TC tasks formulate it as a pair-wise matching problem for query-position. The central goal of both tasks is to find a suitable position in the taxonomy tree for the new concept (query), a position that in TE refers to a hypernym (parent), and in TC to a hypernym and hyponym pair (parent-child). As shown in Figure 1, the query “Relational Database” is inserted in position <DataBase, Relational Operator>; the query “Prompt Learning” is inserted in position <Machine Learning>.

As a self-supervised learning task, the goal is to capture the essential features or relationships in the data. For TC,
it is the relation between queries and positions that is the most essential. As a result, most existing TC methods construct training signals as “query-position” pairs for the matching objective, generating representations for each and using a classifier to predict a match [Arous et al., 2023]. The challenge lies in acquiring high-quality representations from existing taxonomies. To achieve this, leading solvers either use pre-trained language models (PLM) to encode concept descriptions for semantic information [Wang et al., 2022; Arous et al., 2023], or use a propagation module to aggregate node information for structural information [Shen et al., 2020; Jiang et al., 2022a].

While much effort has been devoted to the above technical components, these methods still suffer from two limitations. Firstly, little is known about how to mine relations between queries/nodes and between positions, which is essential for precise matching and capturing the taxonomic structure. TaxoComplete [Arous et al., 2023] model node relations by graph distance and learn it in a supervised manner, but they ignore the relations between positions. Secondly, in query-position relation modeling, negative pairs are often randomly or locally sampled (from a mini-path or its siblings), which may provide uninformative negatives [Xiong et al., 2021] due to concept diversity in a large graph. And as indicated in [Schroff et al., 2015; Wu et al., 2017], specific negative pairs demonstrate more useful than others at different stages of model training. Therefore, how to create meaningful training signals to exploit the self-supervised data remains a challenge.

To this end, we propose CoSTC, a Contrastive representation learning framework designed for self-supervised Taxonomy Completion. Contrastive learning is known to be effective in discriminative learning and capturing apparent similarity among instances [Wu et al., 2018]. Thus, CoSTC can effectively capture diverse relations within taxonomies and learn high-quality representations for queries and positions in a unified way. Specifically, two contrasting views are designed for modelling relations within taxonomy data. First, we conduct an intra-view contrastive task that distinguishes individual instances to capture the apparent similarities among queries and among positions. We utilize a simple unsupervised contrastive sampling approach to construct intra-view contrasting signals (<query-query> and <position-position>) from the well-structured taxonomy. Second, a negative mining-aware inter-view contrastive task is conducted to learn specific relations between query-position for TC matching. This involves a sampling technique regarding taxonomic structures, considering both diversity and hardness over <query-position>-training pairs to select representative pairs for inter-view contrasting.

We perform comprehensive experiments on three real-world datasets, comparing CoSTC with leading methods. We also create several framework variants for an ablation study to further evaluate our approach and its components. Results demonstrate CoSTC’s ability to accurately determine the correct position for a query node.

In summary, we make the following main contributions:
- We propose CoSTC, a contrastive learning framework to capture diverse relations and improve the quality of query/position representations for self-supervised taxonomy completion.
- To exploit the self-supervised data, we design two contrastive views of the taxonomic structure, namely intra-view and inter-view, and propose a negative mining strategy to construct representative training signals.
- Experimental results show that our CoSTC achieves performance comparable or superior to leading methods on three datasets, demonstrating its effectiveness.

2 Related Work

2.1 Taxonomy Expansion and Completion

Automatically updating taxonomies when new concepts emerge has been extensively studied. It can be categorized into two tasks based on the problem assumption: expansion (TE) and completion (TC). In taxonomy expansion, researchers [Shen et al., 2020; Yu et al., 2020; Wang et al., 2021a; 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], focus on attaching the new concepts to the bottom level of the taxonomy as leaf nodes, making it a one-to-one concept matching problem. Since the “leaf-only” assumption is inappropriate for real-world applications, TMN [Zhang et al., 2021] reformulated it to a one-to-one-pair matching problem and thus proposed TC task. In taxonomy completion, new concepts can be attached to any taxonomy position. GenTaxo [Zeng et al., 2021] proposed to first identify candidate positions that need new concepts and then generate appropriate names for them. QEN [Wang et al., 2022] and TaxoEnrich [Jiang et al., 2022a] proposed to consider the sibling relations by introducing horizontal structural information to enhance the positions’ representation. TaxoComplete [Arous et al., 2023] suggested that distant nodes can also provide fine-grained information and proposed a position-enhanced semantic matching method. TacoPrompt [Xu et al., 2023a] proposed a collaborative multi-task prompt-tuning framework to address the overfitting-to-leaf attachment problem.

Our Approach. Similar to most TC methods, we focus on representation learning. The difference is that we construct data augmentations and negative mining with the contrastive learning paradigm to exploit the self-supervised data.

2.2 Contrastive Representation Learning

Contrastive learning (CL) is among the most popular self-supervised representation learning methods in use today [Yang et al., 2022; Nie et al., 2023; Wu et al., 2023; Xu et al., 2023b]. It learns quality representations by clustering positive samples and dispersing negative samples in feature space [Hadsell et al., 2006]. Therefore, two key ingredients for this are notions of similar \((x, x^+\) ) and dissimilar \((x, x^-\) ) pairs of data points [Robinson et al., 2021]. As for computer vision, MoCo [He et al., 2020] and SimCLR [Chen et al., 2020] used data augmentations with geometric and appearance transformations. HCL [Ge et al., 2023] proposed a contrastive learning framework with Euclidean and hyperbolic loss. As for language, SimCSE [Gao et al., 2021] used a
contrasting objective for sentence representation with a simple data augmentation method. GenSE [Chen et al., 2022] suggested using the generative model to acquire synthesized data for contrastive training. As for graph, HomoGCL [Li et al., 2023] found that homophily is the most essential element of GCL and used this to expand the positive set. B*-Sampling [Liu et al., 2023] proposed a novel sampling approach for GCL to reduce training pair noise and improve their representativeness.

Our Approach. Contrastive learning emphasizes the similarity among positives and the dissimilarity among negatives, thereby enhancing the discriminative ability of the model. For representation-based TC tasks, the optimization objective of CL is compatible with the match of query and position: similar pairs are closer while the dissimilar pairs are farther. Motivated by this, we leverage it to learn more natural and promising representations for nodes and positions.

3 Methodology

In this section, we introduce the proposed framework CoSTC in detail. CoSTC focuses on capturing diverse relations and learning query/position representations from self-supervised data for better taxonomy completion. To achieve this, we first construct data augmentations on the taxonomy and apply the corresponding contrastive loss to pre-train the encoder for discriminative representations of queries and positions. Then we finetune the encoder with our sampling method in the matching task to learn fine-grained query-position relations. The architecture of CoSTC is shown in Figure 2.

3.1 Preliminaries

Problem Formulation

[Taxonomy] A taxonomy \( T = (N, E) \) is a directed acyclic graph, where nodes \( n \in N \) are connected by edges \( e \in E \). Each node \( n \in N \) represents a unique concept, with definitions derived from a supporting corpus \( D \). Each directed edge in \( (n_p, n_c) \in E \) represents a hypernym-hyponym relationship between a more general concept \( n_p \) (parent node) and a more specific concept \( n_c \) (child node).

[Problem Definition] Let \( T^0 = (N^0, E^0) \) represents the existing taxonomy which comprises nodes \( N^0 \) and edges \( E^0 \). With a set of new concepts \( C \) and a comprehensive corpus \( D \) defining existing nodes \( n \in N^0 \) and new concepts in \( C \), the objective is to complete \( T^0 \) to \( T \), by removing old edges and forming new ones, adding the new concept into a position, resulting \( T = (N^0 \cup C, E^1) \). To be more specific, for a new query concept \( q \in C \), the goal is to find one or more appropriate candidate positions \( \langle p, c \rangle \in T^0 \), (i.e. \( \langle \text{parent}, \text{child} \rangle \)), to append \( q \). Note that following [Shen et al., 2020; Zhang et al., 2021], we assume the concepts in \( C \) are independent so that the order of queries insertion need not be considered.

Prompt-based Representation

Given a query node \( q \) and a candidate position \( \langle p, c \rangle \), following [Wang et al., 2022; Jiang et al., 2022a], we add the important sibling \( s \) to the candidate position: \( \langle p, c, s \rangle \) by randomly selecting a child of \( p \). With their corresponding definitions \( d_q, d_p, d_c, d_s \in D \), we define their semantic description \( d_{in} \) as follows:

\[
d_{in} = \begin{cases} d_q, & \text{if it is a query} \\ d_p + d_c + d_s, & \text{if it is a position} \end{cases}
\]

where + represents string concatenation. Then following [Wang et al., 2022; Arous et al., 2023], we use a bi-encoder to generate their respective representations \( h \). Specifically, we use PromptBERT [Jiang et al., 2022b] to map the definitions to the representation space, which utilizes prompts to obtain sentence embeddings of high quality. Formally, we have a template function \( F \) that “The meaning of \([X]\)” is \([\text{MASK}]\)”, where \([X]\) stands for the sentence to put in and \([\text{MASK}]\) signifies the \([\text{MASK}]\) token. Given the definition sentence \( d_{in} \), we map \( d_{in} \) to \( d_{prompt} \) with the template function \( F \). Then we feed \( d_{prompt} \) to the BERT encoder to generate the representation \( h \). The hidden vector of the \([\text{MASK}]\) token is the exact representation:

\[
h = h_{[\text{MASK}]}
\]

we use \( h_q \) for query’s representation and \( h_a \) for position’s representation.

3.2 Intra-view Contrastive Pre-training

This section details our intra-view data generation strategies and contrastive pre-training. We use data augmentation approaches to create different views of the self-supervised data. Using these augmented pairs, we pre-train the model to obtain an effective encoder for query/position representation in taxonomy completion. The following sections describe the intra-view data generation strategies and training objectives.

Intra-view Data Generation

In this part, we introduce the function \( f_p() \) to generate intra-view training pairs for contrastive pre-training. Considering we have two elements to match: the query and the position, our goal is to learn structured relations within queries and within positions reducing the noise due to their similarity. Here, intra-view refers to the contrasting perspectives between query and query, position and position. Inspired by [Gao et al., 2021], we use node self-discrimination to generate positive and negative pairs to achieve this. Specifically, given a node \( n_q \) in the existing taxonomy \( T^0 = (N^0, E^0) \) as query, we consider views from the identical query as positive pairs (i.e. \( n_q, n_q^+, n_q^- \)). While views from distinct queries as negative pairs (i.e. \( n_q, n_q^- \)). For each node \( n_i \) in the existing taxonomy \( T^0 = (N^0, E^0) \), the situation will be slightly different since a position is determined by parent and child, excluding sibling. Therefore, its positive sample is defined as \( a^+ = \langle p, c, s \rangle \) if \( s \neq s' \), where the parent and child are the same, the sibling can be different. Its negative sample is defined as \( a^- = \langle p', c', s' \rangle \). Where one of the parent and child must be different, the sibling can be the same or different. This yields two types of intra-view training data.

Contrastive Pre-training Objective

With the generated intra-view data, we pre-train our encoder in the self-supervised contrastive learning paradigm. Similar
Contrasting performance. Following [Zhang et al., 2021], we learn fine-grained query-position relations, enhancing performance. We then fine-tune the pre-trained encoder with these pairs to learn representative training pairs considering diversity and hardness.

3.3 Inter-view Contrastive Matching

This section details our inter-view contrastive matching module. It utilizes a negative mining method to sample representative training pairs considering diversity and hardness. We then fine-tune the pre-trained encoder with these pairs to learn fine-grained query-position relations, enhancing performance. Following [Zhang et al., 2021; Wang et al., 2022], we perform query-position matching for a query to find the best fit among thousands of taxonomy positions.

Inter-view Data Sampling

In this part, we introduce our sampling strategy for our inter-view data contrasting. As suggested in these works [Schroff et al., 2015; Wu et al., 2017], certain negative pairs prove more useful than others at different training stages. Therefore, our goal is to sample the most representative pairs from the enormous negatives. We propose a two-phase sampling strategy: diversity first and then hardness, which are applied sequentially.

Diversity Sampling. Inspired by B²-Sampling [Liu et al., 2023], which uses a balanced sampling technique for graph contrastive learning considering topological and embedding diversity, we consider information diversity in taxonomy. Although taxonomy’s structure is a graph, B²-Sampling is not applicable because it samples nodes, whereas our focus is on positions. To this end, we propose another approach to consider diversity in taxonomy from two perspectives.

The first is the diversity of local and global structures. Specifically, positions in taxonomy are concept pairs \((p, c)\), with \(c\) being a descendant of \(p\). They are categorized based on the graph distance between \(p\) and \(c\) into local linked-edge (distance less than two-hop) and global jump-edge (distance greater than two-hop). When adding a query, the lined-edges are more likely to be a valid position for query. This is because there are many nodes between \(p\) and \(c\) in jump-edge, making it unlikely for the query node to be at the same level as both \(p\) and its descendants. This introduces confusion because the query node might align with multiple levels simultaneously. In linked-edge, the connection is direct, making it easier to determine where the query fits in the hierarchy. Another observation is that the two concepts in linked-edge are more similar than those in jump-edge [Arous et al., 2023]. The linked edges provide local and precise structural information, while the jump-edges model global and coarse structures to aid in position locating. To balance the local and global information, we set a sampling ratio \(r\) to select the pairs from the linked edge and jump edge. We empirically study the ratio of local and global information that can best fit model learning.

The second is the diversity in local structure. Here we follow the intuition in [Liu et al., 2023] that the sampled negative pairs are uniformly distributed, we define a function \(f_m(\cdot)\) that maps the node to positions in the taxonomy and

\[
\ell(h_i, h_i^+) = \log \frac{e^{\theta(h_i, h_i^+) / \tau}}{\sum_{j \neq i} e^{\theta(h_i, h_j) / \tau}}
\]

where \(\theta(\cdot, \cdot)\) is the cosine function to measure the similarity between two representations and \(\tau\) is the temperature parameter in softmax. In this manner, we can obtain a pre-trained robust encoder that can map the input definitions to the well-learned representation space.
adopt the topological diversity sampling method in [Liu et al., 2023]. Specifically, given a node \( n_i \in \mathcal{N}^0 \), we consider its connected linked-edges as corresponding positions (i.e. \( (n_i, n_j) \in \mathcal{E}^0 \mid p = i \lor c = j \)). Inspired by the balanced sampling in [Liu et al., 2023], we select positions based on the shortest path distance between two nodes \( d_s(n_i, n_j) \), which serves as an indicator of topological diversity. Formally, given a node \( n_i \), the probability to select position \( a \) is estimated by:

\[
p(n_i, a) = \frac{1}{|S(n_i)|} \times \frac{1}{|N(n_i, d_s(n_i, n_j))|} \times \frac{1}{|f_m(n_j)|}
\] (4)

where \( S(n_i) = \{d \mid d = d_s(n_i, n_j), n_j \in \mathcal{N}^0\} \), and the first term means to select a distance from \( S(n_i) \) with equal probability. \( N(n_i, d_s(n_i, n_j)) = \{n_j \mid \text{dist}(n_i, n_j) = d_s, n_j \in \mathcal{N}^0\} \), the second term refers to randomly sampling a node from the set where each node’s shortest path distance to \( n_i \) equals \( d_s \). \( f_m(n_j) = \{a \mid a = \langle n_p, n_c \rangle, p = j \lor c = j\} \), the third term means to select a linked-edge connected to it with equal probability. Finally, we obtain a position set \( A_{ds} \) through diversity sampling.

**Hard Negative Sampling.** Hard negative samples, proven to be beneficial for contrastive representation learning [Robinson et al., 2021], can efficiently guide a learning method towards quick mistake correction. We define informative negatives in two ways: representation learning and model training. For representation learning, they are pairs that should be far apart but are mapped nearby. For model training, they are samples that the model incorrectly predicted. Formally, given a node \( n_i \), we select informative negative positions \( A_{hs} \) that meet the above definition from the samples selected at the previous stage \( A_{ds} \) as follows:

\[
A_{hs} = \{a_i \mid ||h_{n_i} - h_{a_i}|| < \gamma, a_i \notin g(n_i), a_i \in A_{ds}\} \\
\cup \{a_j \mid ||h_{n_i} - h_{a_j}|| > \gamma, a_j \in g(n_i), a_j \in A_{ds}\}
\] (5)

where \( a \) represents the position, \( g(n_i) \) means the ground truth position of \( n_i \), and \( \gamma \) is the decision boundary.

**Contrastive Training Objective.** With these sampled inter-view data, we finetune the encoder using the pair-wise contrastive loss function proposed in [Hadsell et al., 2006]:

\[
\ell(y, h_n, h_a) = \frac{1}{2} (1 - y)(d(h_n, h_a))^2 + \frac{1}{2} y(\text{max}(0, m - d(h_n, h_a)))^2
\] (6)

where \( m > 0 \) is a margin and \( d(h_n, h_a) \) is the cosine distance between \( h_n \) and \( h_a \). The negative pairs contribute to the loss only if their distance falls within the margin. In this manner, our encoder further learns the fine-grained query-position relations, enhancing the representation learning.

**Curriculum Learning with CoSTC.** Using very hard negatives may lead to unstable model learning during the first few epochs of training [Harwood et al., 2017], so it is better to start with easier negatives. CoSTC can conveniently adjust the “hardness” of its negatives by changing the margin \( m \) (easy to hard) and the ratio \( r \) (global to local). This curriculum learning strategy can speed up convergence [Dahiya et al., 2023], by 30%-40% compared to training with a fixed \( m \) and \( r \) in our experiments. We found in experiments that dynamically reducing the margin \( m \) can help the model to learn the finer difference and therefore improve the Hit@1 metric.

### 4 Experiments

#### 4.1 Experimental Settings

**Evaluation Dataset.** We conduct experiments on three taxonomy completion datasets: SemEval-Food, MeSH and WordNet-Verb. These are hierarchy taxonomies from different domains. The statistics of these datasets are shown in Table 1.

- **SemEval-Food:** This dataset contains a food domain taxonomy, derived from SemEval-2015 Task 17 [Bordea et al., 2015]. It is the largest taxonomy of the SemEval-2015 task.
- **MeSH:** This dataset contains a widely used clinical domain taxonomy. It is a subgraph of the Medical Subject Headings (MeSH) [Lipscomb, 2000], which is a biomedical indexing hierarchy.
- **WordNet-Verb:** This dataset contains a verb taxonomy, derived from SemEval-2016 Task 14 [Jurgens and Pilehvar, 2016]. It is a hierarchy of verbs from WordNet 3.0.

**Evaluation Protocols.** We follow QEN [Wang et al., 2022] to split the datasets. For SemEval-Food and MeSH, we split them by 8:1:1 into train, validation and test sets. For WordNet-Verb, we randomly sample 1000 nodes respectively for validation and testing. We adopt the all-rank evaluation protocol, where all possible candidate positions are tested and ranked for each test query. We employ the commonly-used Mean Rank (MR), Mean Reciprocal Rank (MRR), Hit@k and Recall@k as evaluation metrics for matching performance evaluation.

**Compared Baseline Methods.** We evaluate our method by comparing it with the following Taxonomy Completion baselines:

- **TMN** [Zhang et al., 2021]: It is the first to propose Taxonomy Completion task by reformulating the matching problem from one-to-one to one-to-pair. It uses three auxiliary tasks which evaluate \( (q, p) \), \( (q, c) \) and \( (q, (p, c)) \) pairs to acquire matching scores.
- **TaxoEnrich** [Jiang et al., 2022a]: It leverages both semantic features and structural information to learn better position representations.

| Dataset            | \(|\mathcal{N}^1|/|\mathcal{N}^\text{train}|\) | \(|\mathcal{E}|\) | #depth | #candidates |
|--------------------|------------------------------------------|------------------|--------|-------------|
| SemEval-Food       | 1486/1190                               | 1,533            | 8      | 7313        |
| MeSH               | 9710/8072                               | 10,498           | 10     | 42970       |
| WordNet-Verb       | 13936/11936                             | 13,407           | 12     | 51159       |

Table 1: Dataset statistics. \(|\mathcal{N}|, |\mathcal{E}|\) are the total number of nodes and edges. #depth and #candidates indicate the depth of the taxonomy and the number of candidate positions.
Table 2: The experimental results of all three datasets. The best results are in bold, while the second-best ones are underlined.

- **QEN** [Wang et al., 2022]: This model leverages the important sibling information and uses term descriptions to enrich the semantic features.

- **TaxoComplete** [Arous et al., 2023]: This is a position-enhanced semantic match approach that emphasizes the importance of semantics and leverages the graph distance as supervising signal to train the nodes’ semantic similarity.

Following QEN [Wang et al., 2022], we adapt some taxonomy expansion baselines (BERT+MLP, TaxoExpan [Shen et al., 2020] and Arborist [Manzoor et al., 2020]) to taxonomy completion settings.

**Implementation Details**

CoSTC is implemented with PyTorch 1. For fair comparison, we use BERT 2 [Devlin et al., 2019] as the backbone model. We employ the AdamW [Loshchilov and Hutter, 2018] optimizer in both stages. For intra-view contrastive prompting, the batch size is set to 64, the training epoch to 5, and the learning rate to 5e-5. For inter-view contrastive matching, the batch size is set to 4, the negative size to 40, the training epoch to 50, and the learning rate to 5e-5. Training ends when the MRR score on the validation set doesn’t improve for 10 epochs, following the early stopping strategy. Initial hyperparameters $r$ and $m$ for our sampling method are 0.25 and 0.4, respectively. For curriculum learning, $r$ increases and $m$ decreases by 0.005 per epoch. All experiments are accelerated using NVIDIA RTX 4090 GPUs.

**4.2 Experimental Results**

Table 2 compares CoSTC’s performance with baseline models, revealing two key observations. Firstly, CoSTC significantly outperforms baselines across all datasets, with notable MR and MRR scores. This underscores the effectiveness of contrastive learning tasks in mining relations and enhancing representation quality in taxonomy completion. Secondly, CoSTC excels over models using random negative sampling. Specifically, it surpasses the second-best model, TaxoComplete, on Recall@10 by 16.3%, 6.4%, and 8.2% on the SemEval-Food, MeSH, and WordNet-Verb datasets, respectively. This is primarily due to our model’s representative negative sampling strategy in self-supervision, which selects meaningful training signals and fully leverages self-supervised signals in existing taxonomies.

**4.3 Ablation Studies**

CoSTC comprises two modules: *intra-view pertaining* and *inter-view sampling*. Their individual impacts on performance are evaluated across three datasets, which are shown in Table 2. To further evaluate the effectiveness of these modules’ choices, additional experiments were conducted on SemEval-Food, detailed in Table 3.
### Impact of Intra-view Pre-training

As shown in Table 2, our model outperforms the “w/o pre-train” variant, indicating the effectiveness of pre-training with intra-view data in improving matching representation. This is attributed to the augmented intra-view pairs providing more information about the taxonomy’s structured relations. We further explore the effects of contrastive learning by evaluating performance across various intra-view training pairs. The results in Table 3 consist of four variants: (1) “only query” uses just “query-query” pairs, (2) “only position” uses only “position-position” pairs, (3) “qry-pos hybrid” uses both, mixed within a mini-batch, and (4) “qry-pos separate” uses both, but a mini-batch contains only queries or positions. All four variants outperform the “w/o pre-train” model, demonstrating the value of contrastive pre-training. The “separate” variant performs better than the “hybrid”, as the latter may introduce noise due to the differing nature of queries and positions.

### Impact of Inter-view Data Sampling

Table 2 shows that CoSTC surpasses the “w/o sampling” variant, indicating our method’s ability to select representative pairs from a large pool to guide model training. We further evaluate our sampling method’s designs for diversity and hard negative mining in Table 3. Here, (1) “fixed sibling” refers to adding a fixed, rather than random, sibling to the position; (2) “only local/global view” implies using only linked or jump edges in training; (3) “w/o balance in div” means two types are randomly mixed instead of using a ratio for balance; (4) “w/o uniform in div” indicates that sampled linked edges are not uniformly distributed; (5) “w/o hard mining” means our hard negative sampling is not applied. The results show that each of these variants negatively affects the diversity or the hardness selection of our sampling strategy to some degree, suggesting the optimality of design choices made in CoSTC.

<table>
<thead>
<tr>
<th>Method</th>
<th>H@1</th>
<th>H@5</th>
<th>R@1</th>
<th>R@5</th>
<th>MRR</th>
</tr>
</thead>
<tbody>
<tr>
<td>w/o pre-train</td>
<td>31.8</td>
<td>66.2</td>
<td>15.1</td>
<td>37.9</td>
<td>0.604</td>
</tr>
<tr>
<td>only query</td>
<td>35.8</td>
<td>66.9</td>
<td>17.0</td>
<td>38.0</td>
<td>0.629</td>
</tr>
<tr>
<td>only position</td>
<td>33.8</td>
<td>69.6</td>
<td>16.1</td>
<td>38.9</td>
<td>0.627</td>
</tr>
<tr>
<td>qry-pos hybrid</td>
<td>33.8</td>
<td>70.3</td>
<td>16.1</td>
<td>40.5</td>
<td>0.616</td>
</tr>
<tr>
<td>qyr-pos separate</td>
<td>39.0</td>
<td>73.4</td>
<td>18.7</td>
<td>43.0</td>
<td>0.658</td>
</tr>
</tbody>
</table>

Table 3: Analyzing the impact of alternative design and algorithmic choices for CoSTC’s components on SemEval-Food.

### 4.4 Further Analysis

#### Parameter Study

The hyper-parameter $r$ balances local and global structures in position sampling. To make training samples progress from easy to difficult, we apply curriculum learning by gradually increasing $r$. An experiment comparing static and dynamic $r$ values was conducted, with results shown in Figure 3. These results indicate the need for a balance between local and global structure information, and the effectiveness of a dynamic increase in achieving this balance.

#### Qualitative Analysis

To intuitively delve into what the sampling strategy changed, we provide insight based on visualization. As shown in Figure 4, it is obvious that the distribution of different candidate positions in linked-edge is more balanced and the ratio between linked-edge and jump-edge is significantly changed. This indicates that our sampling strategy can select more diverse samples for training.

### 5 Conclusion

We propose CoSTC, a two-phase contrastive representation learning framework for taxonomy completion, designed to capture diverse relations and improve representation quality. It uses two contrastive views to mine relations within self-supervised taxonomic data. Phase one models query-query and position-position relations through intra-view contrasting, while phase two uses inter-view contrasting to mine query-position relations for finer representations. We propose a sampling strategy for effective pair selection in inter-view contrasting, considering diversity and hardness. Extensive experiments on three real-world datasets show CoSTC outperforms various baselines, with ablation experiments confirming the effectiveness of our two-view contrasting and sampling strategy.
Acknowledgments

We thank anonymous reviewers for their valuable comments. We also thank Yining Liang, Yuxiao Lii, Yuxun Qu, Xiangrui Cai for their insightful suggestions and kind help. This research is supported by National Engineering Research Center for Digital Construction and Evaluation Technology of Urban Rail Transit, National Natural Science Foundation of China (No. 62077031, 62372252, 72342017).

References


[Jiang et al., 2022a] Minhao Jiang, Xiangchen Song, Jieyu Zhang, and Jiawei Han. Taxoenrich: Self-supervised taxonomy completion via structure-semantic representations. In WWW, pages 925–934, 2022.


