DNG: Taxonomy Expansion by Exploring the Intrinsic Directed Structure on Non-gaussian Space

Songlin Zhai[1], Weiqing Wang[2*], Yuanfang Li[2], Yuan Meng[1]

[1] School of Computer Science and Engineering, Southeast University, China
[2] Faculty of Information Technology, Monash University, Australia
songlin_zhai@seu.edu.cn, {Teresa.Wang,YuanFang.Li}@monash.edu, yuan_meng@seu.edu.cn

Abstract

Taxonomy expansion is the process of incorporating a large number of additional nodes (*i.e.,* "queries") into an existing taxonomy (*i.e.,* "seed"), with the most important step being the selection of appropriate positions for each query. Enormous efforts have been made by exploring the seed's structure. However, existing approaches are deficient in their mining of structural information in two ways: poor modeling of the hierarchical semantics and failure to capture directionality of the *is-a* relation. This paper seeks to address these issues by explicitly denoting each node as the combination of inherited feature (*i.e.,* structural part) and incremental feature (*i.e.,* supplementary part). Specifically, the inherited feature originates from "parent" nodes and is weighted by an inheritance factor. With this node representation, the hierarchy of semantics in taxonomies (*i.e.,* the inheritance and accumulation of features from "parent" to "child") could be embodied. Additionally, based on this representation, the directionality of the *is-a* relation could be easily translated into the irreversible inheritance of features. Inspired by the Darmois-Skitovich Theorem, we implement this irreversibility by a non-Gaussian constraint on the supplementary feature. A log-likelihood learning objective is further utilized to optimize the proposed model (dubbed DNG), whereby the required non-Gaussianity is also theoretically ensured. Extensive experimental results on two real-world datasets verify the superiority of DNG relative to several strong baselines.

Introduction

Taxonomies provide the hierarchical information for various downstream applications, such as Query Understanding (Wang et al. 2015; Hua et al. 2017), Web Search (Wu et al. 2012; Liu et al. 2019a), and Personalized Recommender System (Wang et al. 2017; Hu et al. 2021), where the *is-a* relation connects the nodes (Hearst 1992; Bai et al. 2021). However, manually curating a taxonomy is labor-intensive, time-consuming, and difficult to scale, making the automatic taxonomy expansion draw considerable attention (Shen et al. 2020; Song et al. 2021; Jiang et al. 2022). As depicted in Figure 1, this task aims to insert new nodes (*i.e.,* "query nodes") into an existing taxonomy (*i.e.,* "seed taxonomy") by finding appropriate anchors without modifying the seed's
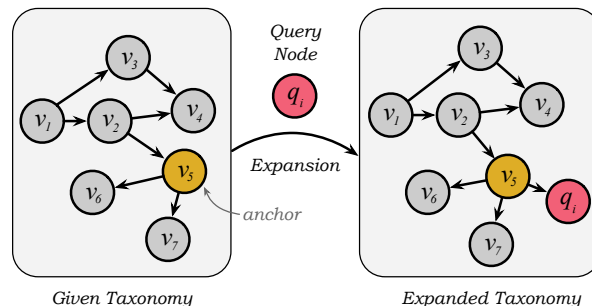

Figure 1: The "*Given Taxonomy*" is expanded by appending the "query" q_i into the anchor node v_5. Specifically, finding appropriate anchors (*e.g.,* v_5) for a query node (*e.g.,* q_i) is the crucial step for the taxonomy expansion task.

original structure for best preserving and utilizing the seed's design. In this paper, we will also follow this research line to improve taxonomy expansion.

Recent works primarily contribute to this task from two directions: capturing the seed's structural information via egonets (Shen et al. 2020), mini-paths (Yu et al. 2020), or vertical-horizontal views of a taxonomy (Jiang et al. 2022); and enhancing node embeddings by implicit edge semantics for a heterogeneous taxonomy (Manzoor et al. 2020), or hyperbolic space (Aly et al. 2019; Ma et al. 2021). It has been suggested that exploring the seed's structural information should be more important for this task (Wang et al. 2021). However, existing works naively incorporate this information by either GNNs (Yao et al. 2022) or the manually-designed natural language utterance (*e.g.,* "v_5 *is a type of* v_2" in Figure 1), leaving two aspects of the seed's structure under-explored, **Type-I**: the hierarchical semantics among "parent" and "child" nodes; and **Type-II**: the directionality of the *is-a* relation. We focus on these two issues in this paper and demonstrate how these two taxonomic structures can be applied to enhance the taxonomy expansion task.

For further analyzing these two issues, taking the *is-a* pair ("animal", "cat") as an example, the hypernym "animal" is semantically more general than its hyponym "cat" (Santus et al. 2014), implying two-fold information about these two nodes. (a) The hyponym "cat" contains all features of its hypernym "animal", indicating that a part of the feature
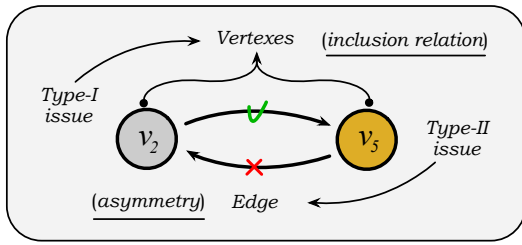
Figure 2: The difference between Type-I and Type-II issues. Type-I focuses on the inclusion relation of node semantics (*i.e.,* vertex level), whereas Type-II is designed for the asymmetric transformation of the *is-a* relation itself (*i.e.,* edge level).

for "cat" could be inherited from "animal". (b) The hypernym "animal" does NOT contain all features of its hyponym "cat", showing that "cat" owns extra features that could not be inherited from "animal". This two-fold information indicates the inclusion relation of semantics between the hypernym and hyponym (*i.e.,* "animal" is a proper subset of "cat" in semantics), reflecting the vertex-level character of a taxonomy. The relationship also shows the **inheritance** and **accumulation** of information from "parent" to "child". This variation in information reflects the hierarchical structure of taxonomies (*i.e.,* **Type-I**) and explains why different levels hold different information.

Besides the hierarchical characteristic, the directionality of the *is-a* relation (*i.e.,* "cat" is-a "animal", but "animal" is-not-a "cat") is another meaningful structural information in a taxonomy (*i.e.,* **Type-II**). This means that the feature inheritance is directed (*e.g.,* "cat" could inherit features from "animal", but not vice versa), which exactly matches the underlying DAG (Directed Acyclic Graph) structure of a taxonomy. To capture this edge-level character, we seek to derive an asymmetric transformation between hypernym and hyponym features since each relation in a graph could be treated as a mathematical mapping of features. However, existing approaches (Rimell 2014; Fu et al. 2014) focus heavily on manipulating hypernym and hyponym features (*e.g.,* resorting to subtraction), which cannot theoretically guarantee the directionality of the *is-a* relation.

In this paper, we propose the DNG model to address the two types of issues with two coherent techniques, and Figure 2 also illustrates the difference of these two issues. On one hand, each node is explicitly represented by a combination of the structural feature (inherited from "parent") and the supplementary feature (owned by "itself" and cannot be inherited from its "parent"), where the structural feature is weighted by an inheritance factor to measure the inheritability of the node. Under this representation, the hierarchical semantics is embodied in the feature transmission by inheritance strategy (*i.e.,* the feature is passed from one layer to the next) and the feature increment via supplementary mechanism (*i.e.,* some extra features will be supplemented in each layer). On the other hand, benefiting from the improved node representation, the directionality of the *is-a* relation could be easily converted to the irreversible inheritance of the supple-

mentary feature. Motivated by the Darmois-Skitovich Theorem, we propose a non-Gaussian constraint on the supplementary feature to implement this irreversibility, therefore theoretically guarantee the relation's directionality.

To optimize DNG, we adopt a log-likelihood learning objective and theoretically prove its effectiveness on modeling structural information. We conduct extensive experiments on two large real-world datasets from different domains to verify the efficacy of the DNG model. The experimental results demonstrate that DNG achieves state-of-the-art performance on the taxonomy expansion task. Additionally, for a better illustration, we will also give a detailed case study about the DNG's ability in modelling taxonomic structure. To summarize, our major contributions include:

- The node representation is improved by combining the structural and supplementary features; thereby, the hierarchical semantics in taxonomies could be explicitly reflected in the modelling process.

- A non-Gaussian constraint on the supplementary feature is adopted to theoretically guarantee the directionality of the *is-a* relation.

- A log-likelihood learning objective is employed to optimize the DNG model, ensuring that the required constraint is theoretically guaranteed.

- We conduct extensive experiments on two large datasets from different domains to validate the performance of DNG and the efficacy of modeling taxonomic structure.

Preliminaries

Task Definition

Taxonomy: A given taxonomy $\mathcal{T} = \{\mathcal{V}, \mathcal{E}\}$ can be treated as a directed acyclic graph composed of a vertex set ($\mathcal{V}$) and an edge set ($\mathcal{E}$) (Shen et al. 2020; Jiang et al. 2022). An edge $(v_i, v_j) \in \mathcal{E}$ between nodes v_i and v_j represents the *is-a* relation between them (*e.g.,* v_j *is-a* v_i), where v_i and v_j are the hypernym ("*parent*") and the hyponym ("*child*"), respectively.

Taxonomy Expansion: It refers to the process that appends a set of new nodes (query set $\mathcal{Q}$) into an existing taxonomy (seed taxonomy $\mathcal{T}^0$), whereby a larger taxonomy could be produced. Hence, this task requires: (1) a seed taxonomy $\mathcal{T}^0 = \{\mathcal{V}^0, \mathcal{E}^0\}$ with $|\mathcal{V}^0| = N$; and (2) the query set $\mathcal{Q} = \{q_i | 1 \le i \le M\}$. As illustrated in Figure 1, the crucial step is to find appropriate anchors in seed taxonomy $\mathcal{T}^0$ for each query q_i.

Shannon Entropy

Shannon entropy is the basic concept for measuring the degree of uncertainty of a random variable. It is defined for a discrete-valued random variable X as:

$$H(X) = -\sum p_x(X) \log p_x(X) = -\mathbb{E}\{\log p_x(X)\} \quad (1)$$

According to Eq. 1, entropy for the linear transformation $\boldsymbol{y} = \boldsymbol{B}\boldsymbol{x}$, can be derived as follows (Hyvärinen, Karhunen, and Oja 2001):

$$H(\boldsymbol{y}) = H(\boldsymbol{x}) + \ln|\det \boldsymbol{B}| \quad (2)$$

where det denotes the determinant; $\boldsymbol{B}$ and ln are the mapping matrix and napierian logarithm, respectively.

Methodology

Node Representation

In order to incorporate the taxonomic structure, each node (*e.g.*, v_i) is explicitly represented by the combination of structural feature (*inherited from its parents*) and supplementary feature (*could not be inherited from its parents*):

$$\boldsymbol{v}_i = \underbrace{\sum_{v_j \in \text{PA}_i} s_{ij} \boldsymbol{v}_j}_{\text{structural}} + \underbrace{\boldsymbol{u}_i}_{\text{supplementary}} = \boldsymbol{V} \boldsymbol{s}_i + \boldsymbol{u}_i \quad (3)$$

where $\boldsymbol{v}_i \in \mathbb{R}^d$ and PA_i are the d-dimensional representation and the parent set of node v_i, respectively. $s_{ij} \boldsymbol{v}_j$ denotes the feature inherited from the parent node v_j, where the inherited feature is weighted by an inheritance factor s_{ij}. $\boldsymbol{u}_i$ is the supplementary feature of node v_i which cannot be obtained from its parents. $\boldsymbol{V}$ indicates the feature matrix of seed's node set ($\mathcal{V}^0$), and $\boldsymbol{s}_i$ is the inheritance vector of v_i on $\mathcal{V}^0$.

Figure 3 illustrates the node representation in a taxonomy fragment with two *is-a* pairs (*i.e.*, v_5 *is-a* v_2, and v_2 *is-a* v_1). For the node v_1, its feature $\boldsymbol{v}_1$ only contains the supplementary part (*i.e.*, $\boldsymbol{v}_1 = \boldsymbol{u}_1$) since it is the root and does not inherit any feature from any node. Node v_2 has the parent v_1, making its representation $\boldsymbol{v}_2$ the combination of $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ (*i.e.*, $\boldsymbol{v}_2 = s_{21} \boldsymbol{u}_1 + \boldsymbol{u}_2$). It is observed that $\boldsymbol{u}_2$ is independent of $\boldsymbol{u}_1$ since neither $\boldsymbol{u}_1$ nor $\boldsymbol{u}_2$ contains any information about the other. Similarly, $\boldsymbol{v}_5$ can be represented by the inherited feature and its supplementary feature (*i.e.*, $\boldsymbol{v}_5 = s_{52}(s_{21} \boldsymbol{u}_1 + \boldsymbol{u}_2) + \boldsymbol{u}_5$), from which the independence among $\boldsymbol{u}_1$, $\boldsymbol{u}_2$, and $\boldsymbol{u}_5$ can be also inferred. Hence, based on Eq. 3, the hierarchical semantics in a taxonomy could be embodied: the inheritance and accumulation of features from one layer to the next layer. This example also shows that these supplementary features are independent of each other.

Additionally, the directionality is another crucial attribute of the *is-a* relation: "v_5 *is-a* v_2, but v_2 *is-not-a* v_5". How to capture this structural information is an issue to be resolved. In this paper, we attempt to translate this directionality into the irreversible inheritance of features: "v_5 could inherit feature from v_2, but v_2 could not inherit feature from v_5". Under this intuition, this attribute could be realized by the irreversible transformation between $\boldsymbol{v}_2$ and $\boldsymbol{v}_5$ in Eq. 3 (*i.e.*, $\boldsymbol{v}_2 \neq s_{25} \boldsymbol{v}_5 + \boldsymbol{u}_2$ under $\boldsymbol{v}_5 = s_{52} \boldsymbol{v}_2 + \boldsymbol{u}_5$). Inspired by the Darmois-Skitovich Theorem, ensuring the non-Gaussianity of the supplementary feature is a feasible implementation, as proven in the next section.

Directionality of *is-a* on Node Representation

Supposing the representation $\boldsymbol{v}_i = s_{ij} \boldsymbol{v}_j + \boldsymbol{u}_i$ admits the inverse version $\boldsymbol{v}_j = s_{ji} \boldsymbol{v}_i + \boldsymbol{u}_j$ under non-Gaussian $\boldsymbol{u}_i$, there will exist the following transformation:

$$\begin{aligned} \boldsymbol{u}_j &= \boldsymbol{v}_j - s_{ji} \boldsymbol{v}_i \\ &= \boldsymbol{v}_j - s_{ji}(s_{ij} \boldsymbol{v}_j + \boldsymbol{u}_i) \quad (4) \\ &= (1 - s_{ji} s_{ij}) \boldsymbol{v}_j - s_{ji} \boldsymbol{u}_i \end{aligned}$$

As such, the simultaneous equation will be obtained:

$$\begin{cases} \boldsymbol{v}_i = s_{ij} \boldsymbol{v}_j + \boldsymbol{u}_i \\ \boldsymbol{u}_j = (1 - s_{ji} s_{ij}) \boldsymbol{v}_j - s_{ji} \boldsymbol{u}_i \end{cases} \quad (5)$$
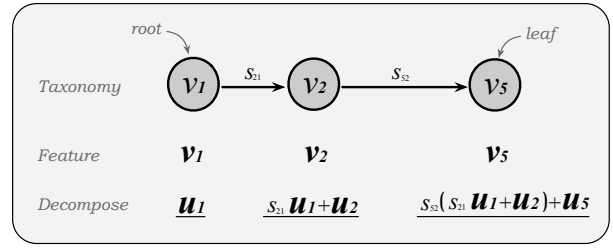

Figure 3: The hierarchical semantics in a taxonomy could be shown by the inheritance and accumulation of features from one layer to the next layer. For node v_i, s_{ij} is the inheritance factor to scale the inherited feature from node v_j.

Due to the non-Gaussian $\boldsymbol{u}_i$, it can be drawn from Eq.5 based on the corollary of Darmois-Skitovich Theorem[1] that $\boldsymbol{v}_i$ and $\boldsymbol{u}_j$ will be dependent (*i.e.*, $\boldsymbol{v}_i \not\perp\!\!\!\perp \boldsymbol{u}_j$). However, the fundamental condition of the inverse transformation $\boldsymbol{v}_j = s_{ji} \boldsymbol{v}_i + \boldsymbol{u}_j$ is $\boldsymbol{v}_i \perp\!\!\!\perp \boldsymbol{u}_j$. This contradiction between the assumption and the deduction verifies that the non-Gaussian supplementary feature will make the reversal node representation given in Eq. 3 false, indicating that the directionality of *is-a* is captured. In order to ensure this required constraint, we employ a log-likelihood learning target, which will be elaborated in the next section.

Learning Target

Based on the recursive node representation in Eq. 3, the node set $\mathcal{V}$ can be represented as [2]:

$$\boldsymbol{V} = \boldsymbol{U} \boldsymbol{W}^{-1} \quad (6)$$

where $\boldsymbol{W} = \boldsymbol{I} - \boldsymbol{S}$ is dubbed the transition matrix. $\boldsymbol{I}$ and $\boldsymbol{S}$ are an identity matrix and the inheritance factor matrix, respectively, and both of them have the size $N \times N$. $\boldsymbol{U}$ denotes the supplementary feature matrix with the shape of $d \times N$ (d is the dimension of this feature).

From Eq. 6, we can observe that the essential step is to estimate the supplementary feature set $\boldsymbol{U}$ and the inheritance factors $\boldsymbol{S}$ given the observational feature $\boldsymbol{V}$. For this inheritance factor matrix, the entry with position (i^{th}, j^{th}) reflects the probability of node v_j being the "parent" of node v_i.

In real-word datasets, the feature $\boldsymbol{V}$ could be easily obtained from the well-designed pre-trained models (Liu et al. 2019b; Devlin et al. 2019) by fusing sufficient context information[3]. However, how to learn $\boldsymbol{U}$ and $\boldsymbol{S}$ from $\boldsymbol{V}$ remains to be determined. To fulfill this goal, we will first analyze the supplementary features U as:

$$p_{\boldsymbol{U}}(\boldsymbol{U}) = \prod_{i=1}^{N} p_{\boldsymbol{U}}(\boldsymbol{u}_i) \quad (7)$$

---

[1] See appendix in the arXiv version.

[2] Derivation of Eq. 6: $\boldsymbol{V} = \boldsymbol{V}\boldsymbol{S} + \boldsymbol{U} \Rightarrow \boldsymbol{V}\boldsymbol{I} - \boldsymbol{V}\boldsymbol{S} = \boldsymbol{U} \Rightarrow \boldsymbol{V}(\boldsymbol{I} - \boldsymbol{S}) = \boldsymbol{U} \Rightarrow \boldsymbol{V} = \boldsymbol{U}(\boldsymbol{I} - \boldsymbol{S})^{-1}$

[3] DNG focuses on parsing $\boldsymbol{U}$ and $\boldsymbol{S}$ under the given $\boldsymbol{V}$ which can be extracted by other structural approaches with the structural measure, *e.g.*, Wu&P (K., Shet, and Acharya 2012).

where $p_U(\boldsymbol{u}_i)$ is the probability density of $\boldsymbol{u}_i$, implemented by $\tanh(\boldsymbol{u}_i) = (e^{\boldsymbol{u}_i} - e^{-\boldsymbol{u}_i})/(e^{\boldsymbol{u}_i} + e^{-\boldsymbol{u}_i})$ (Hyvärinen, Karhunen, and Oja 2001). Then, in order to build the relation between $p_U(\boldsymbol{U})$ and $p_V(\boldsymbol{V})$, the joint probability density of $\boldsymbol{V}$ could be derived from $p_U(\boldsymbol{U})$ by their probability distribution functions (PDFs) [4]:

$$p_V(\boldsymbol{V}) = |\det\boldsymbol{W}| \prod_{i=1}^{N} p_U(\boldsymbol{V}\boldsymbol{w}_i) \qquad (8)$$

where $\boldsymbol{W} = (\boldsymbol{w}_i, ..., \boldsymbol{w}_n)^{\top}$ indicates how to analyze the supplementary features from the observational representations.

With the definition of $p_V(\boldsymbol{V})$ in hand, the given taxonomy $\mathcal{T}$ could be treated as the joint distribution of the node set $\mathcal{V}$; hence, the learning target can be defined as maximizing the logarithm of the joint probability:

$$\begin{aligned} \jmath_\Theta(\mathcal{T}) &= -\ln(|\det\boldsymbol{W}| \prod_{i=1}^{N} p_U(\boldsymbol{V}\boldsymbol{w}_i)) \\ &= -\sum_{i=1}^{N} \ln(p_U(\boldsymbol{V}\boldsymbol{w}_i)) - \ln|\det\boldsymbol{W}| \end{aligned} \qquad (9)$$

where $\jmath$ denotes the loss to be minimized, and Θ is the parameters. For the DNG model, it is essential to guarantee that the non-Gaussian constraint is satisfied under this optimization, which will be proven in the next section.

The Non-Gaussian U

To demonstrate that the supplementary feature $\boldsymbol{U}$ is a non-Gaussian distribution under the learning objective, let us first derive the expectation of Eq. 9:

$$\begin{aligned} \mathbb{E}(\jmath) &= -\sum_{i=1}^{N} \mathbb{E}\{\ln(p_U(\boldsymbol{V}\boldsymbol{w}_i))\} - \ln|\det\boldsymbol{W}| \\ &= \sum_{i=1}^{N} H(\boldsymbol{V}\boldsymbol{w}_i) - \ln|\det\boldsymbol{W}| \\ &= \sum_{i=1}^{N} H(\boldsymbol{u}_i) - \ln|\det\boldsymbol{W}| \end{aligned} \qquad (10)$$

Additionally, in information theory, the entropy measures the uncertainty of a random variable; therefore, a variable with more randomness will have higher entropy. For all types of random variables with the same variance, the Gaussian variable has the highest entropy, leading us to borrow the Negentropy (Hyvärinen, Karhunen, and Oja 2001) as the measure of a variable's non-Gaussianity. With the help of

Datasets	$|\mathcal{V}|$	$|\mathcal{E}|$	#Depth
MAG-CS	24,754	42,329	6
WordNet-Verb	13,936	13,408	13

Table 1: Statistics of datasets, where $|\mathcal{V}|$ and $|\mathcal{E}|$ are the number of nodes and edges in taxonomies, respectively. #Depth indicates the depth of taxonomies.

Eq. 10, the non-Gaussianity of feature $\boldsymbol{U}$ can be derived as:

$$\begin{aligned} \mathrm{NG}(\boldsymbol{U}) &= H(\boldsymbol{U}_{Gaussian}) - H(\boldsymbol{U}) \\ &= H(\boldsymbol{U}_{Gaussian}) + \mathbb{E}\{\ln(p_U(\boldsymbol{V}\boldsymbol{W}))\} \\ &= \mathbb{E}\{\ln(p_U(\boldsymbol{V}\boldsymbol{W}))\} - \ln|\det\boldsymbol{W}| + \\ &\quad H(\boldsymbol{U}_{Gaussian}) + \ln|\det\boldsymbol{W}| \\ &= -\mathbb{E}(\jmath) + const \end{aligned} \qquad (11)$$

where NG denotes the non-Gaussianity of a variable. $\boldsymbol{U}_{Gaussian}$ is the feature with Gaussian distribution, and $H(\boldsymbol{U}_{Gaussian})$ could be treated as a constant (Tong, Inouye, and Liu 1993; Hyvärinen, Karhunen, and Oja 2001). Additionally, the term $\ln|\det\boldsymbol{W}|$ could be also considered as a constant since $\mathbb{E}(\boldsymbol{V}\boldsymbol{V}^{\top}) = \boldsymbol{W}\mathbb{E}(\boldsymbol{U}\boldsymbol{U}^{\top})\boldsymbol{W}^{\top} = \boldsymbol{I}$ [5] under the unit-variance $\boldsymbol{V}$ and $\det\boldsymbol{I} = 1 = \det\boldsymbol{W}\mathbb{E}(\boldsymbol{U}\boldsymbol{U}^{\top})\boldsymbol{W}^{\top} = (\det\boldsymbol{W})(\det\mathbb{E}(\boldsymbol{U}\boldsymbol{U}^{\top}))(\det\boldsymbol{W}^{\top})$. In this transformation, it can draw that the term $\det\mathbb{E}(\boldsymbol{U}\boldsymbol{U}^{\top})$ does not depend on $\boldsymbol{W}$, which implies $\det\boldsymbol{W}$ is a constant (Hyvärinen and Oja 2000). Consequently, based on Eq. 11, it can be also drawn that *maximizing the joint probability of $\boldsymbol{V}$ is also equivalent to maximizing the non-Gaussianity of $\boldsymbol{U}$*. This corollary exactly accords with the non-Gaussian condition of the proposed DNG model.

DNG Algorithm

In the DNG model, estimating the matrix $\boldsymbol{S}$ is the crucial step for the expansion task. To launch the learning procedure, the taxonomy with its vertex set, the feature matrix $\boldsymbol{V}$ and the edge set are taken as the inputs. Then, the proposed model DNG is used to analyze the supplementary features and estimate the inheritance factor matrix. Specially, the gradients of updating the parameters are:

$$\boldsymbol{W} \leftarrow \boldsymbol{W} + \alpha(\boldsymbol{G}\boldsymbol{V} + (\boldsymbol{W}^{-1})^{\top}) \qquad (12)$$

where $\boldsymbol{G} = (2/sinh(2\boldsymbol{u}_1), ..., 2/sinh(2\boldsymbol{u}_n))^{\top}$, and α is the learning rate. During the inference stage, the inheritance vector $\boldsymbol{s}_{q_i}$ for the query q_i is directly computed[6] based on the query feature $\boldsymbol{v}_{q_i}$, the learned supplementary feature matrix and the inheritance matrix. With $\boldsymbol{s}_{q_i}$ in hand, its appropriate positions could be easily found[7].

Experiments

Experimental Setup

Datasets We evaluate DNG and compare models on two large public real-world taxonomies:

---

[4] First, get the the relation between PDFs: $\mathcal{F}_V(V) = P_V(\boldsymbol{V} \leq V) = P_V(\boldsymbol{U}\boldsymbol{W}^{-1} \leq V) = P_U(\boldsymbol{U} \leq V\boldsymbol{W}) = \mathcal{F}_U(V\boldsymbol{W})$. Then, compute the first-order derivation of PDFs: $p_V(\boldsymbol{V}) = \mathcal{F}'_V(\boldsymbol{V}) = \mathcal{F}'_U(\boldsymbol{V}\boldsymbol{W}) = p_U(\boldsymbol{V}\boldsymbol{W})|\det\boldsymbol{W}|$

[5] V is pre-processed to unit variance (Hyvärinen and Oja 2000).

[6] For a more precise estimation, a new iteration is requisite on the whole node set.

[7] Code is available at https://github.com/SonglinZhai/DNG.

Models	MAG-CS			
	Recall	Precision	MR	MRR
TaxoExpan	0.100	0.163	197.776	0.562
ARBORIST	0.008	0.037	1142.335	0.133
TMN	0.174	0.283	118.963	0.689
GenTaxo	0.149	0.294	140.262	0.634
TaxoEnrich	0.182	0.304	67.947	0.721
DNG	**0.268**	**0.317**	**59.354**	**0.754**
Models	WordNet-Verb			
	Recall	Precision	MR	MRR
TaxoExpan	0.085	0.095	665.409	0.406
ARBORIST	0.016	0.067	608.668	0.380
TMN	0.110	0.124	615.021	0.423
GenTaxo	0.094	0.141	6046.363	0.155
TaxoEnrich	0.162	**0.294**	217.842	0.481
DNG	**0.199**	0.201	**43.723**	**0.513**

Table 2: Performance of all compared models, where "Recall" and "Precision" refer to the top-1 evaluation metrics. The results of baselines are from Jiang et al. (2022).

- **Microsoft Academic Graph (MAG)** is a public Field-of-Study (FoS) taxonomy (Sinha et al. 2015). This taxonomy contains over 660,000 scientific nodes and more than 700,000 *is-a* relations. Following (Shen et al. 2020; Zhang et al. 2021; Jiang et al. 2022), we re-construct a sub-taxonomy from MAG by picking the nodes of the computer science domain, named **MAG-CS**. Additionally, the node embeddings extracted via word2vec (Mikolov et al. 2013) based on the related paper abstracts corpus are used as the initial embeddings.

- **WordNet** is another large taxonomy, containing various concepts (Jurgens and Pilehvar 2016). Following (Zhang et al. 2021; Jiang et al. 2022), the concepts of the verb sub-taxonomy are selected based on WordNet 3.0, referred as **WordNet-Verb** since this sub-field is the part that has been fully-developed in WordNet. Analogously, the fasttext embeddings[8] are also generated as the initial feature vectors (Shen et al. 2020; Zhang et al. 2021). In addition, a pseudo root named "verb" is inserted as the root node to generate a complete taxonomy because there are isolated components in the constructed taxonomy.

Following the dataset splits used in (Zhang et al. 2021; Jiang et al. 2022), 1,000 nodes are randomly chosen to test our model in each dataset. The statistical information for these two datasets is listed in Table 1.

Compared Models Five recent methods are adopted to verify the performance of DNG model:

(1) **TaxoExpan** (Shen et al. 2020) employs a position-enhanced graph neural network to capture the taxonomic

---

[8] The wiki-news-300d-1M-subword.vec.zip on official website is used.

structure based on the InfoNCE (van den Oord, Li, and Vinyals 2018) for better model robustness.

(2) **ARBORIST** (Manzoor et al. 2020) aims to resolve the issues in heterogeneous taxonomies via a large margin ranking loss and a dynamic margin function.

(3) **TMN** (Zhang et al. 2021) performs the taxonomy completion task by computing the matching score of the triple (query, hypernym, hyponym). We employ the model for the expansion task in this paper.

(4) **GenTaxo** (Zeng et al. 2021) resorts to the sentence-based and subgraph-based node encodings to perform the matching of each query node. Following (Jiang et al. 2022), the GenTaxo++ is adopted in this paper since the original framework contains a name generation module which is not the focus of this paper.

(5) **TaxoEnrich** (Jiang et al. 2022) achieves state-of-the-art performance on the taxonomy expansion task. This approach leverages semantic and structural information to better represent nodes, which is a strong baseline in this paper.

Evaluation Metrics A ranked list of nodes is provided to choose appropriate anchors for each query. Following (Shen et al. 2020; Zhang et al. 2021; Jiang et al. 2022), four rank-based metrics are used to evaluate the performance of DNG and the compared methods:

- **Precision**@k is defined as

$$Precison@k = \#TP@k/k$$

where $\#TP@k$ is 1 if the ground-truth node is in the top-k predictions, otherwise it is 0. The results reported in Table 2 are the mean value of all test cases on Precision@1.

- **Recall**@k is computed by:

$$Recall@k = \#TP@k/\#GT$$

where $\#GT$ is the number of ground truth labels for one case. The results reported in Table 2 are also the mean value of all test cases on Recall@1.

- **Mean Rank** (MR) gives the average rank of true position among all candidate positions. Following (Zhang et al. 2021; Jiang et al. 2022), for multi-anchor queries, we also first compute the result of each individual triplet and then take the average.

- **Mean Reciprocal Rank** (MRR) defines the reciprocal rank of the true position for each query. In this paper, the scaled MRR is also adopted to amplify the difference among different approaches (Shen et al. 2020; Zhang et al. 2021; Jiang et al. 2022).

Performance Evaluation

Similar to (Shen et al. 2020; Zhang et al. 2021; Jiang et al. 2022), we also split the original taxonomy into separate sub-taxonomies, with each sub-taxonomy containing all paths from a source node (*e.g.,* the root node) to a destination node (*e.g.,* one leaf node). During inference, candidate anchors are first initially produced from these distinct sub-taxonomies. Then these candidate anchors are recombined
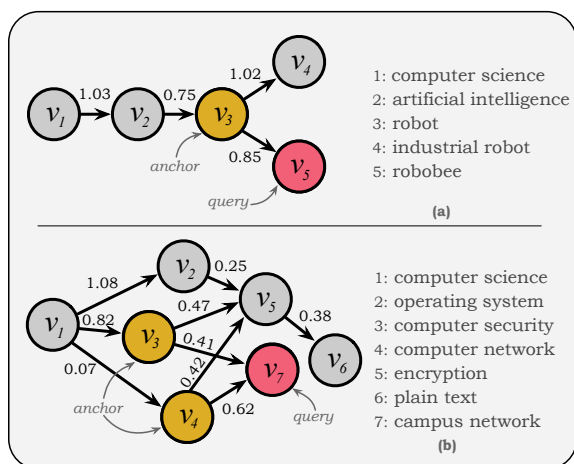
Figure 4: Two examples for validating DNG's ability to capture the hierarchical semantics. The upper figure depicts the expansion of a query into a taxonomy fragment with a chain structure. However, the bottom one is a slightly more complicated example with a graph-structure taxonomy.

by combining the anchors from the nearest sub-taxonomies to generate better anchors. After several iterations, the ultimate predictions could be generated.

Table 2 summarizes the performance of all compared models on Precision, Recall, MR and MRR evaluation metrics. Generally, the DNG model outperforms all baseline approaches on most metrics and especially gains a large margin improvement in Recall across all datasets, indicating the actual anchors being ranked at high positions. Additionally, it has been shown that not all models are capable of producing comparatively superior results for all evolution metrics. Despite the findings demonstrating the excellent performance of TaxoEnrich, the proposed DNG model still surpasses this strong baseline on the majority of criteria. The advantage of the DNG model on MR and MRR reveals its superior ranking performance, validating the effectiveness of the designed node representation and the non-Gaussian constraint presented in this paper.

Capturing Hierarchical Semantics

Section 4 presents a re-designed node representation to capture the hierarchical semantics reflected by the inheritance factor. By estimating S, appropriate anchors could be determined by locating the nodes with high factors, which will be illustrated in this section.

Figure 4 gives two real cases of the MAG-CS dataset for validating DNG's ability to capture taxonomic structure. In the top sub-figure, there are five nodes, each with a single parent in the taxonomy, where $v_1 \rightarrow v_2 \rightarrow v_3 \rightarrow v_4$ and v_5 are the candidate nodes and the query node, respectively. Hence, the DNG model needs to estimate the inheritance vector s_5 on these candidate nodes. After carrying out the inference, the estimated s_5 is $[0, 0, 0.8519, 0, 0]$, indicating v_3 is its only "parent" with the inheritance factor 0.8519 and it does not inherit any feature from any other
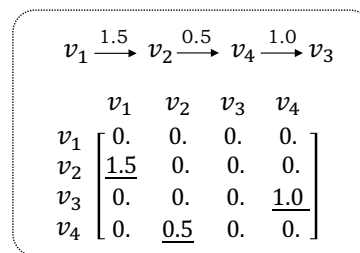

Figure 5: A toy taxonomy with four nodes and three edges. The bottom sub-figure depicts the actual inheritance matrix, where the entity in row i^{th} and column j^{th} denotes the inheritance factor used to weight the inherited feature for the node v_i. These underlined entries indicate that actual connections exist at the current locations. For instance, the underlined factor 1.5 with location $(2^{nd}, 1^{st})$ shows that the connection $v_1 \rightarrow v_2$ is a ground-truth one; however, the entity $0.$ with location $(1^{st}, 2^{nd})$ is the opposite.

nodes. Since the inheritance factor of v_3 is the largest and others are 0, the appropriate anchor of query v_5 is v_3. Figure 4 (b) depicts a more complicated example in which the convoluted relationships could more convincingly demonstrate DNG's ability. The matching process for the query node v_7 is to estimate s_7 on other nodes, and the result is $[0, 0, 0.4116, 0.6156, 0, 0, 0]$, indicating v_7 inherits features from both of v_3 and v_4. Similarly, appropriate anchors are v_3 and v_4 since all other factors are equal to zero. Another notable finding is that v_7 ("campus network") inherits more features from v_4 ("computer network") than from v_3 ("computer security"), despite the fact that both v_3 and v_4 are the "parents". This is consistent with their ground-truth relations in the MAG-CS dataset, which further validates the DNG's superiority in identifying appropriate anchors for queries in taxonomic structures.

Capturing Directionality and Taxonomic Structure

One fundamental contribution of this paper is capturing the directionality of the *is-a* relation under the non-Gaussian constraint; thereby, the structure of a taxonomy could be illustrated. Theoretically, to validate this, we need to solve the exact backward equation (*i.e.,* $v_j = f(v_i)$) under the non-Gaussian constraint to see if it obeys the reversed form of Eq. 3 (*i.e.,* $v_j = \sum_{v_i \in \mathrm{PA}_j} s_{ji} v_i + u_j$). However, despite the fact that the forward equation (see Eq. 3) has been theoretically proven correct, it is still difficult to determine the exact backward equation due to the lack of its prior knowledge. Consequently, we devise an alternative synthetic experiment to serve a similar purpose. Specifically, on the one hand, the distance between the actual and predicted features is also a measure of model's reliability, with a lower value indicating better performance. On the other hand, supposing the backward equation obeys the reversed Eq. 3, the difference between the distances computed by the forward and backward equations could reflect the "deviation degree" of the backward one. It is anticipated that this "deviation degree" will have a larger value for nodes with non-Gaussian supple-

$$
\begin{array}{c|cccc}
 & v_1 & v_2 & v_3 & v_4 \\
\hline
v_1 & 0. & -0.0245 & 0.0188 & -0.1137 \\
v_2 & \underline{1.4781} & 0. & -0.0995 & 0.1550 \\
v_3 & 0.0993 & -0.0413 & 0. & \underline{1.0304} \\
v_4 & 0.1243 & \underline{0.5271} & 0. & -0.0918
\end{array}
$$

$$
\begin{array}{c|cccc}
 & v_1 & v_2 & v_3 & v_4 \\
\hline
v_1 & 0. & 0.4802 & 0.1518 & 0.4498 \\
v_2 & \underline{-1.2770} & 0. & -0.3997 & 1.1127 \\
v_3 & -0.0546 & 0.3742 & 0. & \underline{0.3662} \\
v_4 & -0.7659 & \underline{0.5609} & 0.4163 & 0.
\end{array}
$$

Figure 6: The estimated inheritance matrices of the uniformly-distributed (the top sub-figure) and Gaussianly-distributed (the bottom one) supplementary feature matrices with respect to the nodes in Figure 5. It is observed that the value of inheritance matrix under the Gaussian distribution diverge significantly from the actual value (in Figure 5), and the structure under this distribution is also erroneous.

mentary features compared with that for Gaussian ones. We use a uniform distribution as a non-Gaussian distribution. Consequently, these distance differences of a uniformly-distributed and a Gaussianly-distributed supplementary features are computed to evaluate the reasonableness of the hypothetical backward equation in this section.

To conduct the experiment described above, two virtual taxonomies with four nodes and an identical structure (see Figure 5) are constructed using a uniform and a Gaussian distribution, respectively, where the edge set is $\{v_1 \rightarrow v_2 \rightarrow v_4 \rightarrow v_3\}$ with inheritance factors $\{1.5, 0.5, 1.0\}$. The DNG model is then employed to estimate the matrix S and the supplementary feature U under these two distributions. For the uniform distribution, the distance between the actual and the predicted features with forward equation is: $\{0.51, 0.58, 0.50, 0.59\}$, while the distance between the actual and the backward-predicted features is $\{0.58, 1.25, 1.63, 1.13\}$. An analogous procedure is followed to obtain the distances for the Gaussian distribution: $\{0.79, 0.84, 0.83, 0.83\}$ (forward) and $\{0.79, 1.47, 1.44, 1.11\}$ (backward). As a result, the mean ratios of the "deviation degree" for the uniform and Gaussian distributions are 107.19% and 45.56%, respectively. The result implies that the hypothetical backward equation incurs a higher loss for the uniform distribution, demonstrating the irreversibility of Eq. 3 under the non-Gaussian constraint.

Apart from demonstrating DNG's ability in capturing the directionality, we also conduct experiments to validate its ability in capturing the entire taxonomic structure. Figure 6 presents the estimated inheritance matrices S of the taxonomy (see Figure 5) under different distributions. As illustrated in the upper matrix, compared to the estimated S with Gaussian distribution, the one with non-Gaussian constraint (the top figure) more precisely capture the taxonomic structure, since the ground-truth "parents" have the larger inheritance factor, validating DNG's ability in modeling taxonomic structure.
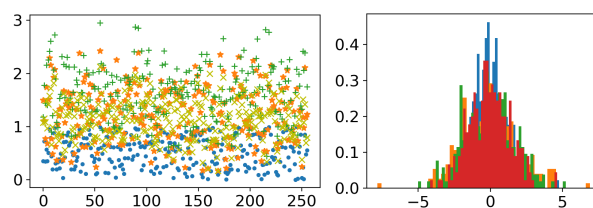

Figure 7: Visualizations of the nodes (in Figure 5) under the uniform (left one) and Gaussian (right one) distributions.

Related Work

To improve the taxonomy expansion task, enormous efforts have been made to incorporate the structural information of the seed. Manzoor et al. (2020) learns node embeddings with multiple linear maps to more effectively integrate the structural information of implicit edge semantics for heterogeneous taxonomies. For the specific modelling structural information, Shen et al. (2020) employs the egonet with a position-enhanced graph neural network, and a training scheme based on the InfoNCE loss is also employed to enhance the noise-robustness. However, Yu et al. (2020) converts the expansion task into the matching between query nodes and mini-paths with anchors to capture the locally structural information. For optimizing the taxonomic structure, Song et al. (2021) is the only one that considers the relationship among the different queries to ensure that the generated taxonomy has the optimal structure. Zhang et al. (2021) adds a new node by simultaneously considering the hypernym-hyponym pair, thereby overcoming the limitations of hyponym-only mode. Additionally, Wang et al. (2021) aims to maximize the coherence of the expanded taxonomy in order to take advantage of the hierarchical structure of the taxonomy. Jiang et al. (2022) combines the semantic features (extracted by surface names) and structural information (generated by the pseudo sentences of both vertical and horizontal views) to better represent each node, achieving state-of-the-art performance on this task.

Conclusion

In this paper, we sought to improve the taxonomy expansion task by investigating the taxonomic structure from hierarchical semantics and the directionality of the *is-a* relation, by which the DNG model was proposed. On the one hand, to model hierarchical semantics, each node is represented by the combination of the structure and the supplementary features, with the structure part weighted by an inheritance factor. On the other hand, based on the node representation, the *is-a*'s directionality is easily translated into the irreversible inheritance of features, implemented by the non-Gaussian distribution of the supplementary features. In order to optimize DNG, we employed the log-likelihood learning objective and demonstrated the equivalency between this learning target and the required constraint. Finally, extensive experiments were conducted on two large real-world datasets, and the results not only verified the DNG's superiority over other compared models on the expansion task but also confirmed its ability to capture the taxonomic structure.

Acknowledgements

References

Aly, R.; Acharya, S.; Ossa, A.; Köhn, A.; Biemann, C.; and Panchenko, A. 2019. Every Child Should Have Parents: A Taxonomy Refinement Algorithm Based on Hyperbolic Term Embeddings. In *Proceedings of the 57th Conference of the Association for Computational Linguistics (ACL) Volume 1: Long Papers*, 4811–4817. Association for Computational Linguistics.

Bai, Y.; Zhang, R.; Kong, F.; Chen, J.; and Mao, Y. 2021. Hypernym Discovery via a Recurrent Mapping Model. In *Findings of the Association for Computational Linguistics: ACL/IJCNLP*, volume ACL/IJCNLP 2021 of *Findings of ACL*, 2912–2921. Association for Computational Linguistics.

Devlin, J.; Chang, M.; Lee, K.; and Toutanova, K. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT Volume 1 (Long and Short Papers)*, 4171–4186. Association for Computational Linguistics.

Fu, R.; Guo, J.; Qin, B.; Che, W.; Wang, H.; and Liu, T. 2014. Learning Semantic Hierarchies via Word Embeddings. In *ACL Volume 1: Long Papers*, 1199–1209. Association for Computer Linguistics.

Hearst, M. A. 1992. Automatic Acquisition of Hyponyms from Large Text Corpora. In *14th International Conference on Computational Linguistics (COLING)*, 539–545.

Hu, X.; Xu, J.; Wang, W.; Li, Z.; and Liu, A. 2021. A Graph Embedding Based Model for Fine-grained POI Recommendation. *Neurocomputing*, 428: 376–384.

Hua, W.; Wang, Z.; Wang, H.; Zheng, K.; and Zhou, X. 2017. Understand Short Texts by Harvesting and Analyzing Semantic Knowledge. *IEEE Trans. Knowl. Data Eng.*, 29(3): 499–512.

Hyvärinen, A.; Karhunen, J.; and Oja, E. 2001. *Independent Component Analysis*. Wiley. ISBN 0-471-40540-X.

Hyvärinen, A.; and Oja, E. 2000. Independent component analysis: algorithms and applications. *Neural Networks*, 13(4-5): 411–430.

Jiang, M.; Song, X.; Zhang, J.; and Han, J. 2022. TaxoEnrich: Self-Supervised Taxonomy Completion via Structure-Semantic Representations. In *WWW'22: The ACM Web Conference 2022*, 925934. Association for Computing Machinery.

Jurgens, D.; and Pilehvar, M. T. 2016. SemEval-2016 Task 14: Semantic Taxonomy Enrichment. In *Proceedings of the 10th International Workshop on Semantic Evaluation, SemEval@NAACL-HLT*, 1092–1102. The Association for Computer Linguistics.

K., M. S.; Shet, K. C.; and Acharya, U. D. 2012. A New Similarity Measure for Taxonomy Based on Edge Counting. *CoRR*, arxiv.org/abs/1211.4709.

Liu, B.; Guo, W.; Niu, D.; Wang, C.; Xu, S.; Lin, J.; Lai, K.; and Xu, Y. 2019a. A User-Centered Concept Mining System for Query and Document Understanding at Tencent. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 1831–1841. Association for Computing Machinery.

Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; and Stoyanov, V. 2019b. RoBERTa: A Robustly Optimized BERT Pretraining Approach. *CoRR*, arxiv.org/abs/1907.11692.

Ma, M. D.; Chen, M.; Wu, T.; and Peng, N. 2021. HyperExpan: Taxonomy Expansion with Hyperbolic Representation Learning. In *Findings of the Association for Computational Linguistics: EMNLP*. Association for Computational Linguistics.

Manzoor, E. A.; Li, R.; Shrouty, D.; and Leskovec, J. 2020. Expanding Taxonomies with Implicit Edge Semantics. In *Proceedings of The Web Conference 2020*, 2044–2054. Association for Computing Machinery / IW3C2.

Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G. S.; and Dean, J. 2013. Distributed Representations of Words and Phrases and their Compositionality. In *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013 (NIPS)*, 3111–3119.

Rimell, L. 2014. Distributional Lexical Entailment by Topic Coherence. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, 511–519. The Association for Computer Linguistics.

Santus, E.; Lenci, A.; Lu, Q.; and im Walde, S. S. 2014. Chasing Hypernyms in Vector Spaces with Entropy. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics, EACL*, 38–42. The Association for Computer Linguistics.

Shen, J.; Shen, Z.; Xiong, C.; Wang, C.; Wang, K.; and Han, J. 2020. TaxoExpan: Self-supervised Taxonomy Expansion with Position-Enhanced Graph Neural Network. In *WWW'20: The Web Conference 2020*, 486–497. Association for Computing Machinery / IW3C2.

Sinha, A.; Shen, Z.; Song, Y.; Ma, H.; Eide, D.; Hsu, B. P.; and Wang, K. 2015. An Overview of Microsoft Academic Service (MAS) and Applications. In *Proceedings of the 24th International Conference on World Wide Web Companion, WWW*, 243–246. Association for Computing Machinery.

Song, X.; Shen, J.; Zhang, J.; and Han, J. 2021. Who Should Go First? A Self-Supervised Concept Sorting Model for Improving Taxonomy Expansion. In *Proceedings of the ACM*

Web Conference 2022, WWW'22. Association for Computing Machinery / IW3C2.

Tong, L.; Inouye, Y.; and Liu, R.-w. 1993. Waveform-Preserving Blind Estimation of Multiple Independent Sources. *IEEE Transactions on Signal Processing*, 41(7): 2461–2470.

van den Oord, A.; Li, Y.; and Vinyals, O. 2018. Representation Learning with Contrastive Predictive Coding. *CoRR*, arxiv.org/abs/1807.03748.

Wang, S.; Zhao, R.; Chen, X.; Zheng, Y.; and Liu, B. 2021. Enquire One's Parent and Child Before Decision: Fully Exploit Hierarchical Structure for Self-Supervised Taxonomy Expansion. In *Proceedings of the Web Conference 2021*, WWW'21, 3291–3304. Association for Computing Machinery / IW3C2.

Wang, W.; Yin, H.; Chen, L.; Sun, Y.; Sadiq, S.; and Zhou, X. 2017. ST-SAGE: A Spatial-Temporal Sparse Additive Generative Model for Spatial Item Recommendation. *ACM Transactions on Intelligent Systems and Technology*, 8(3).

Wang, Z.; Zhao, K.; Wang, H.; Meng, X.; and Wen, J. 2015. Query Understanding through Knowledge-Based Conceptualization. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI)*, 3264–3270. AAAI Press.

Wu, W.; Li, H.; Wang, H.; and Zhu, K. Q. 2012. Probase: a probabilistic taxonomy for text understanding. In *Proceedings of the ACM SIGMOD International Conference on Management of Data*, 481–492. Association for Computing Machinery.

Yao, K.; Liang, J.; Liang, J.; Li, M.; and Cao, F. 2022. Multiview Graph Convolutional Networks with Attention Mechanism. *Artif. Intell.*, 307(C): 103708.

Yu, Y.; Li, Y.; Shen, J.; Feng, H.; Sun, J.; and Zhang, C. 2020. STEAM: Self-Supervised Taxonomy Expansion with Mini-Paths. In *The 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 1026–1035. Association for Computing Machinery.

Zeng, Q.; Lin, J.; Yu, W.; Cleland-Huang, J.; and Jiang, M. 2021. Enhancing Taxonomy Completion with Concept Generation via Fusing Relational Representations. In *KDD'21: The 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2104–2113. Association for Computing Machinery.

Zhang, J.; Song, X.; Zeng, Y.; Chen, J.; Shen, J.; Mao, Y.; and Li, L. 2021. Taxonomy Completion via Triplet Matching Network. In *Proceeding of the Thirty-Fifth AAAI Conference on Artificial Intelligence*, 4662–4670. AAAI Press.