J. Math. Computer Sci., 39 (2025), 325–337

Online: ISSN 2008-949X

Journal of Mathematics and Computer Science

Journal Homepage: www.isr-publications.com/jmcs

Theoretical analysis of perturbation multi-dividing ontology learning algorithm

Wei Gao[a,*], Juxiang Zhou[b]

[a] *School of Information Science and Technology, Yunnan Normal University, Kunming 650500, China.*

[b] *Key Laboratory of Education Informatization for Nationalities, Ministry of Education, Yunnan Normal University, Kunming 650500, China.*

Abstract

The multi-dividing ontology learning algorithm is specially designed for tree-structured ontology graphs, and has become a paradigm of graph-based ontology learning. In view of the disturbance of ontology data, this paper proposes perturbation multi-dividing ontology learning approach. Assuming that the perturbed ontology data are drawn from the same distribution as before, the error bound of perturbation multi-dividing ontology learning is given in such hypothesis. Finally, we analyze flaws in theoretical results and gaps with practical applications, and raise the open problem for future study.

Keywords: Ontology, similarity measuring, perturbation multi-dividing ontology learning algorithm.

2020 MSC: 68Q32, 68T05.

1. Introduction

Ontology is an efficient tool for conceptual structured management and semantic computing, which has been a salient field in artificial intelligence (See Gao et al. [9, 11, 12]). In terms of graph structure, ontology describes the relationship between concepts, and is used in medicine, biology and other fields due to its powerful efficiencies. Specifically, the structured data in ontology is formulated by a (directed) graph, where each concept represented by a vertex and each edge represents the direct correlation between two concepts. Let $G = (V(G), E(G))$ be a graph corresponding to a specific ontology O. The target of ontology learning is to get an optimal ontology function $f : V(G) \to \mathbb{R}$ with the help of the ontology sample set and optimizer. In classification setting, f is a classifier which assigns each vertex a label in ontology graph. While in information retrieval setting, f is a score function, and the similar ones between ontology concepts (their corresponding vertices denoted by v and v') are measured by $|f(v) - f(v')|$.

Ontology and related derivative algorithms are applied in various fields of artificial intelligence and introduced into various engineering applications. Bozic [4] conceptualized ontologies for input generation and output processing by a metamorphic testing trick. Dubslaff et al. [8] proposed and implemented an approach for the quantitative analysis of ontologized procedures in terms of standard description

*Corresponding author
Email address: gaowei@ynnu.edu.cn (Wei Gao)

doi: 10.22436/jmcs.039.03.02

Received: 2024-05-08 *Revised:* 2024-07-29 *Accepted:* 2025-02-28

logic reasoning and probabilistic model checking tricks. Osman et al. [14] focused on the heterogeneity problem, and conducted a comprehensive survey from all ontology integration aspects. Sinha et al. [17] provided a comprehensive overview of the extant data mining ontologies. Anand and Kumar [2] identified various uncertainties in ontology by means of different classification of ontology. Abbes et al. [1] proposed an ontology-based fuzzy melanoma diagnosis system, where the qualitative characteristics are obtained by fuzzy classifier. Xue et al. [20] introduced an interactive compact memetic algorithm based semiautomatic ontology matching approach. Buoncompagni et al. [5] determined a new model to be tested on an ontology network representing knowledge to enable smart homes to perform human activity recognition online. In Yang et al. [21], an ontology learning technology was introduced to extract systems engineering ontology from existing standards. Lu et al. [13] introduced the GOPPRRE ontologies to create the model-based systems engineering formalisms in a domain-specific modeling tool. More related works on ontology and data representation can be referred to [18, 19].

Most of the existing ontology algorithms are designed according to the specific application background, or in light of the characteristics of semantic computing. However, due to the need for structured storage, ontologies are stored in ontology graphs. The graph structure depicts the internal relationship between ontology concepts, and its topological features contain the structured information of the entire ontology concept distribution. In order to learn ontology functions from structural information, a multi-dividing ontology learning algorithm, an ontology learning strategy dedicated to tree structure, is proposed, and its deformation under various frameworks is studied in depth. In details, ontology concepts are divided into k parts according to the structural features of a particular tree-shaped ontology graph, and the classes levels are determined by the analysis of the relationship between the various classes. The advantage and merit lie in that concepts are subdivided and sorted according to their classification. Similar classes are arranged closed. The greater the gap between concept categories, the difference between the corresponding categories is even bigger.

Simply speaking, the multi-dividing ontology learning algorithm divides all vertices into k classes (rates), marked as $1, 2, \dots, k$. Under the ontology function f, the vertices with small rate labels have higher corresponding values than those with larger rate labels. That is to say, $f(v^a) > f(v^b)$ if v^a belongs to rate a and v^b belongs to rate b with $1 \leqslant a < b \leqslant k$. Correspondingly, the ontology sample set is also divided into k classes for learning the ontology function. This approach has been verified to have high efficiency for ontology learning algorithms on tree-structured ontology graphs. In recent years, the theoretical analysis of multi-dividing ontology learning algorithms have become one of the main streams in ontology studies. Gao and Farahani [10] determined the generalization bounds and uniform bounds for convex ontology loss function multi-dividing ontology algorithms. Gao et al. [12] analyzed the partial multi-dividing ontology learning algorithm from the perspective of statistical learning theory. Zhu and Hua [22] proposed statistical analysis of multi-dividing ontology learning algorithm in two-sample setting.

Since only a small number of papers have analyzed the multi-dividing ontology learning algorithm, the theoretical results in most of the settings are still unknown, which inspires us to study the statistical characteristics of the multi-dividing ontology learning algorithm under more specific frameworks. In this paper, we consider the multi-dividing ontology learning algorithm in the novel setting such that the ontology data in each pair of rates are perturbation (due to various factors such as collection errors, subjective labeling, and malicious poisoning in actual datasets, perturbation is inevitably present in the constructed dataset). The rest parts of paper are organized as follows. The notations and setting of perturbation multi-dividing ontology learning are presented in the next section. Next, the theoretical analysis of perturbation multi-dividing ontology setting is determined. Finally, we discuss the defects and the open problem in this new multi-dividing ontology learning setting.

2. Setting

The purpose of this section is to give a framework for perturbation multi-dividing ontology learning algorithm, including model descriptions and the related mathematical terminologies.

2.1. Standard multi-dividing ontology learning

Assume that each concept in ontology is a vertex in ontology graph, and all its information is extracted and represented in a d dimension vector. The ontology function $f : V \to \mathbb{R}$ is essentially a dimensional reduction map $f : \mathbb{R}^d \to \mathbb{R}$ ($d \in \mathbb{N}$). Let $V \subseteq \mathbb{R}^d$ be a vertex space (instance space) for ontology graph G, and the vertices in V are drawn independently and randomly according to certain unknown distribution $\mathcal{D}$. Assume $k \geqslant 2$ is an integer. In the multi-dividing ontology setting, ontology vertices are divided into k classes (corresponding to k rates) and the order (rank) of these k rates is determined by the experts.

Formally, the learner is inferred to an ontology training set $S = (S^1, S^2, \cdots, S^k) \in V^{n_1} \times V^{n_2} \times \cdots \times V^{n_k}$, which consists of a sequence of ontology training samples $S^a = (v_1^a, \ldots, v_{n_a}^a) \in V^{n_a}$ ($1 \leqslant a \leqslant k$). Here, $n_1, \ldots, n_k \in \mathbb{N} \cup \{0\}$, and $\sum_{i=1}^k n_i$ is called the total multi-dividing ontology sample size. In light of ontology sample S, a real-valued ontology function $f : \mathbb{R}^p \to \mathbb{R}$ is learned, which satisfies $f(v^a) > f(v^b)$ for any pair of (a, b), where $1 \leqslant a < b \leqslant k$.

The standard multi-dividing ontology expected framework can be formulated as

$$f^* = \arg\min_{f \in \mathcal{F}} \sum_{a=1}^{k-1} \sum_{b=a+1}^{k} \mathbb{P}_{(v^a, v^b) \sim \mathcal{D}^a \times \mathcal{D}^b}(f(v^a) > f(v^b)),$$

where $\mathcal{F}$ is an ontology function space (such as Reproducing Kernel Hilbert Space) and $\mathcal{D}^a$ is underlying conditional distributions for each rate. The corresponding empirical version with multi-dividing ontology sample set $S = (S^1, S^2, \ldots, S^k) \in V^{n_1} \times V^{n_2} \times \cdots \times V^{n_k}$ is given by

$$f^* = \arg\min_{f \in \mathcal{F}} \sum_{a=1}^{k-1} \sum_{b=a+1}^{k} \frac{1}{n_a n_b} \sum_{i=1}^{n_a} \sum_{j=1}^{n_b} \mathbf{I}(f(v_i^a) > f(v_j^b)).$$

Since $\mathbf{I}$ is a binary discrete function, which is non-derivative, makes it very difficult to directly optimize the aforementioned multi-dividing ontology learning model. The common trick to deal with this problem is to replace the binary function with a continuous ontology loss function, and for the purpose of theoretical analysis, such ontological loss function is always assumed to be a smooth convex function satisfying the Lipschitz condition from a theoretical point of view. Formally, a multi-dividing ontology loss function is a function $l : \mathbb{R}^V \times V \times V \to \mathbb{R}_+ \cup \{0\}$ that assigns, for ontology function $f : \mathbb{R}^d \to \mathbb{R}$ and v^a, v^b in rate pair (a, b) with $1 \leqslant a < b \leqslant k$, a non-negative real number $l(f, v^a, v^b)$ interpreted as the loss of multi-dividing ontology function f in its relative order of v^a and v^b. The expected multi-dividing ontology error (risk) on the tree-shaped ontology graph G for an ontology function $f : \mathbb{R}^d \to \mathbb{R}$ associated with the loss function l is formulated by

$$R_{l,\mathcal{D}}(f) = \sum_{a=1}^{k-1} \sum_{b=a+1}^{k} \mathbf{E}_{v^a \sim \mathcal{D}^a, v^b \sim \mathcal{D}^b}\{l(f, v^a, v^b)\},$$

where $\mathcal{D}^a$ is the conditional distribution of $\mathcal{D}$ on V^a. Its corresponding empirical multi-dividing ontology error (risk) on the tree-shaped ontology graph G for an ontology function $f : \mathbb{R}^d \to \mathbb{R}$ associated with the loss function l and multi-dividing ontology sample $S = (S^1, S^2, \ldots, S^k) \in V^{n_1} \times V^{n_2} \times \cdots \times V^{n_k}$ is defined by

$$\widehat{R}_{l,S}(f) = \sum_{a=1}^{k-1} \sum_{b=a+1}^{k} \frac{1}{n_a n_b} \sum_{i=1}^{n_a} \sum_{j=1}^{n_b} l(f, v_i^a, v_j^b).$$

2.2. Perturbation multi-dividing ontology learning

In this paper, we consider the perturbation multi-dividing ontology expected framework and its corresponding empirical version, where for each pair of (a, b) with $1 \leqslant a < b \leqslant k$, the ontology samples in rate a and rate b are perturbation. Our setting is simplified in, which only ontology samples in rate a are assumed to be perturbed for each pair of (a, b), and this hypothesis can be reversed (only consider the ontology samples in lower rate b are perturbed in each pair of (a, b) with $1 \leqslant a < b \leqslant k$). For each v^a in rate $a \in \{1, \dots, k-1\}$, its neighborhood is denoted by $N(v^a) = \{v^{a'} : v^{a'} - v^a \in \mathcal{B}\}$, where $\mathcal{B}$ is a symmetric, convex, and closed set, which is usually defined as a ball with certain kind of norm, for example, l_q-ball with $q \geqslant 1$.

Let $\mathcal{F}$ be the space of multi-dividing ontology functions and $l(f, v^a, v^b)$ be the ontology loss function for any $f \in \mathcal{F}$, $v^a \in V^a$, $v^b \in V^b$, where $a, b \in \mathbb{N}$ and $1 \leqslant a < b \leqslant k$. The perturbation multi-dividing ontology expected risk of ontology function $f \in \mathcal{F}$ is defined as

$$R^{per}_{l,\mathcal{D}}(f, \mathcal{B}) = \sum_{a=1}^{k-1} \sum_{b=a+1}^{k} \mathbf{E}_{v^a \sim \mathcal{D}^a, v^b \sim \mathcal{D}^b} \{ \max_{v^{a'} \in N(v^a)} l(f, v^{a'}, v^b) \}. \tag{2.1}$$

Associated with multi-dividing ontology sample set $S = (S^1, S^2, \dots, S^k) \in V^{n_1} \times V^{n_2} \times \cdots \times V^{n_k}$, the perturbation multi-dividing ontology empirical risk of ontology function $f \in \mathcal{F}$ is defined as

$$\widehat{R}^{per}_{l,S}(f, \mathcal{B}) = \sum_{a=1}^{k-1} \sum_{b=a+1}^{k} \frac{1}{n_a n_b} \sum_{i=1}^{n_a} \sum_{j=1}^{n_b} \max_{v^{a'} \in N(v_i^a)} l(f, v_i^{a'}, v_j^b). \tag{2.2}$$

For convenience, assume that for each perturbation pair (a, b) with $1 \leqslant a < b \leqslant k$, v_i^a and v_j^b are independently drawn according to the underlying discrete distributions $\mathcal{D}^a_{n_a}$ and $\mathcal{D}^b_{n_b}$, respectively. Hence, the perturbation multi-dividing ontology empirical risk of ontology function $f \in \mathcal{F}$ (2.2) can be re-written by

$$\widehat{R}^{per}_{l,\mathcal{D}_{n_1,\dots,n_k}}(f, \mathcal{B}) = \sum_{a=1}^{k-1} \sum_{b=a+1}^{k} \mathbf{E}_{v^a \sim \mathcal{D}^a_{n_a}, v^b \sim \mathcal{D}^b_{n_b}} \{ \max_{v^{a'} \in N(v_i^a)} l(f, v^{a'}, v_j^b) \}.$$

When we focus on one pair (a, b), then the partial version of above definitions are defined by

$$R^{per,a,b}_{l,\mathcal{D}}(f, \mathcal{B}) = \mathbf{E}_{v^a \sim \mathcal{D}^a, v^b \sim \mathcal{D}^b} \{ \max_{v^{a'} \in N(v^a)} l(f, v^{a'}, v^b) \},$$

$$\widehat{R}^{per,a,b}_{l,S^a,S^b}(f, \mathcal{B}) = \frac{1}{n_a n_b} \sum_{i=1}^{n_a} \sum_{j=1}^{n_b} \max_{v^{a'} \in N(v_i^a)} l(f, v_i^{a'}, v_j^b),$$

and

$$\widehat{R}^{per,a,b}_{l,\mathcal{D}_{n_a,n_b}}(f, \mathcal{B}) = \mathbf{E}_{v^a \sim \mathcal{D}^a_{n_a}, v^b \sim \mathcal{D}^b_{n_b}} \{ \max_{v^{a'} \in N(v_i^a)} l(f, v^{a'}, v_j^b) \},$$

respectively.

3. Preliminaries

To get our main result in perturbation multi-dividing ontology learning setting, we need to prepare some new notations and observations.

3.1. New concepts and remarks

Define a mapping for each perturbation pair (a, b),

$$\Psi^{a,b}_{f,v^b}(v^a) = v^{a,b,*} = \arg\max_{v^{a'} \in N(v^a)} l(f, v^{a'}, v^b),$$

where $v^a \in V^a$. Acted as a perturbation multi-dividing ontology sample, the rate of $v^{a,b,*}$ keeps a, i.e., the label of $v^{a,b,*}$ is the same as v^a since it's from the neighborhood v^a. We have

$$\begin{aligned} R^{per}_{l,\mathcal{D}}(f,\mathcal{B}) &= \sum_{a=1}^{k-1}\sum_{b=a+1}^{k} \mathbf{E}_{v^a\sim\mathcal{D}^a, v^b\sim\mathcal{D}^b}\{\max_{v^{a'}\in N(v^a)} l(f, v^{a'}, v^b)\} \\ &= \sum_{a=1}^{k-1}\sum_{b=a+1}^{k} \mathbf{E}_{v^a\sim\mathcal{D}^a, v^b\sim\mathcal{D}^b}\{l(f, v^{a,b,*}, v^b)\} \\ &= \sum_{a=1}^{k-1}\sum_{b=a+1}^{k} \mathbf{E}_{v^{a,b,*}\sim\mathcal{D}^{a'}, v^b\sim\mathcal{D}^b}\{l(f, v^{a,b,*}, v^b)\} \\ &= \sum_{a=1}^{k-1}\sum_{b=a+1}^{k} \mathbf{E}_{v^a\sim\mathcal{D}^{a'}, v^b\sim\mathcal{D}^b}\{l(f, v^a, v^b)\} = R_{l,\mathcal{D}'}(f), \end{aligned} \tag{3.1}$$

where $\mathcal{D}^{a'} = \Psi^{a,b}_{f,v^b}\#\mathcal{D}^a$ is the pushforward of $\mathcal{D}^a$ by $\Psi^{a,b}_{f,v^b}$ for the given pair (a,b), and $\mathcal{D}'$ is regarded as the overall pushforward of $\mathcal{D}$ in which $\mathcal{D}^a$ is replaced by $\mathcal{D}^{a'}$ for each pair of (a,b). The expression of (3.1) implies that the perturbation multi-dividing ontology learning risk (2.1) can be estimated by the standard multi-dividing ontology learning risk with respect to $\mathcal{D}^{a'}$ and $\mathcal{D}^b$ for each pair of (a,b) with $1 \leqslant a < b \leqslant k$.

To characterize $\mathcal{D}^{a'}$, we show that $\mathcal{D}^{a'}$ is located in a certain ball, which centers at $\mathcal{D}^a$. We apply the tricks proposed by Celik et al. [6], Panaretos and Zemel [15], Chen and Niles-Weed [7], Shi and Wang [16], and Assa and Plataniotis [3] to deal with it. Let $P(Z)$ be the Borel probability measure space on $Z = V \times Y$, where $Y = \{1, \ldots, k\}$ indicates the rate of each ontology data except the top root vertex (such vertex is an artificially added virtual vertex, for instance, in famous gene ontology, the top root vertex is "GO" and the top root vertex in plant ontology is "PO"). Let

$$d^p_Z(z,z') = d^p_Z((v,y),(v',y')) = d^p_V(v,v') + d^p_Y(y,y'),$$

where d_V is the metric in d dimensional vector space and d_Y is the direct distance between labels. For instance, if $v^a \in V^a$ and $v^b \in V^b$, then $d_Y(v^a, v^b) = |a-b|$. We assume that d_V satisfies the following translation property: $d_V(v,v') = d_V(v-v',0)$. Let

$$P_p(Z) = \{\mathcal{D} \in P(Z) : \mathbb{E}_{z\sim\mathcal{D}}[d^p_Z(z,z')] < \infty, z' \in Z, p \geqslant 1\}.$$

For any two probability measures $\mathcal{D}_1, \mathcal{D}_2 \in P_p(Z)$, let $\Upsilon(\mathcal{D}_1,\mathcal{D}_2)$ be the set of all measures on $V \times Y$ with marginal distributions $\mathcal{D}_1$ and $\mathcal{D}_2$ on the first two factors. Then, the p-th Wasserstein distance between $\mathcal{D}_1$ and $\mathcal{D}_2$ is defined by

$$\mathcal{WD}_p(\mathcal{D}_1,\mathcal{D}_2) = \inf_{\mathcal{D}\in\Upsilon(\mathcal{D}_1,\mathcal{D}_2)} (\mathbb{E}_{Z\times Z'\sim\mathcal{D}}[d^p_Z(z,z')])^{\frac{1}{p}}.$$

Hence, when rate a is located in small class for pair (a,b), the Wasserstein distance between $\mathcal{D}^a$ and $\mathcal{D}^{a'}$ is calculated by

$$\mathcal{WD}_p(\mathcal{D}^a,\mathcal{D}^{a'}) = \inf_{\mathcal{D}\in\Upsilon(\mathcal{D}^a,\mathcal{D}^{a'})} (\mathbb{E}_{V^a\times V^{a'}\sim\mathcal{D}}[d^p_Z(v^a,v^{a'})]^{\frac{1}{p}}, \mathcal{D}^a, \mathcal{D}^{a'} \in P_p(V^a)),$$

where $\Upsilon(\mathcal{D}^a,\mathcal{D}^{a'})$ is the set of all measures on $V^a \times V^a$ with marginal distributions $\mathcal{D}^a$ and $\mathcal{D}^{a'}$, and

$$P_p(V^a) = \{\mathcal{D}^a \in P(V^a) : \mathbb{E}_{v\sim\mathcal{D}^a}[d^p_V(v^a,v^{a'})] < \infty, v^{a'} \in V^a\}.$$

In addition, the Wasserstein ball of $\mathcal{D}^a$ with radius ρ is

$$\mathcal{B}_{\rho,p}(\mathcal{D}^a) = \{\mathcal{D}^{a'} \in P_p(V^a) : \mathcal{WD}_p(\mathcal{D}^a,\mathcal{D}^{a'}) \leqslant \rho\}.$$

Then, the local worse-case multi-dividing ontology learning risk of ontology function f can be defined by

$$R^{per}_{l,\rho,p}(f)=\sum_{a=1}^{k-1}\sum_{b=a+1}^{k}\sup_{\mathcal{D}^{a'}\in\mathcal{B}_{\rho,p}(\mathcal{D}^a)}R_{l,\mathcal{D}^a,\mathcal{D}^b}(f).$$

It's partial version for the fixed pair (a,b) is defined by

$$R^{per}_{l,\rho,p,\mathcal{D}^b}(\mathcal{D}^a,f)=\sup_{\mathcal{D}^{a'}\in\mathcal{B}_{\rho,p}(\mathcal{D}^a)}R_{l,\mathcal{D}^a,\mathcal{D}^b}(f).$$

Denote $\rho_{\mathcal{B}}=\sup_{v\in\mathcal{B}}d_V(v,0)$ by the radius of the adversary set $\mathcal{B}$. For any $f\in\mathcal{F}$, $\mathcal{D}'$ (totally pushforward by $\mathcal{D}^{a'}=\Psi^{a,b}_{f,v^b}\#\mathcal{D}^a$ forward rate a in each pair of (a,b)), in light of translation property, the definition of $\Psi^{a,b}_{f,v^b}$ and $\rho_{\mathcal{B}}$, and the definition of Wasserstein distance, we have

$$\mathcal{WD}^p_p(\mathcal{D}^a,\mathcal{D}^{a'})\leqslant\mathbb{E}_{v^a\sim\mathcal{D}^a}[d^p_V(v^a,\Psi^{a,b}_{f,v^b}(v^a))]=\mathbb{E}_{v^a\sim\mathcal{D}^a}[d^p_V(v^a,v^{a,b,*})]\leqslant\rho^p_{\mathcal{B}},\tag{3.2}$$

which implies $\mathcal{WD}_p(\mathcal{D}^a,\mathcal{D}^{a'})\leqslant\rho_{\mathcal{B}}$.

To further simplify the notation, we only consider $p=1$ in our multi-dividing ontology learning setting, and denote the partial expected version by $R^{per,a,b}_{\rho_{\mathcal{B}},\mathcal{D}^b}(\mathcal{D}^a,f)=R^{per,a,b}_{\rho_{\mathcal{B}},1,\mathcal{D}^b}(\mathcal{D}^a,f)$. In this way, the expression (3.2) reveals $R^{per,a,b}_{l,\mathcal{D}^a,\mathcal{D}^b}(f,\mathcal{B})\leqslant R^{per,a,b}_{l,\rho_{\mathcal{B}},\mathcal{D}^b}(\mathcal{D}^a,f)$ for any $f\in\mathcal{F}$, and hence $R^{per}_{l,\mathcal{D}}(f,\mathcal{B})\leqslant R^{per}_{l,\rho_{\mathcal{B}}}(f)$.

The idea is to get the upper bound of $R^{per}_{l,\mathcal{D}}(f,\mathcal{B})$ via obtaining the upper bound of $R^{per}_{l,\rho_{\mathcal{B}}}(f)$. For this purpose, we present several useful intermediate conclusions in the next subsection.

3.2. *Some intermediate derivation results*

Set

$$\begin{aligned}\mathcal{G}^{a,b}&=\{g^{a,b}:g^{a,b}(v^a,v^b)=l(f,v^a,v^b),f\in\mathcal{F},v^a\in V^a,v^b\in V^b\},\\ \Omega_{\lambda,g^{a,b},\mathcal{D}^b}(v^a)&=\sup_{v^{a'}\in V^a}\{\mathbb{E}_{v^b\sim\mathcal{D}^b}(g^{a,b}(v^{a'},v^b)-\lambda d_V(v^{a'},v^a))\},\end{aligned}\tag{3.3}$$

where λ is a non-negative real number. Let $g^{a,b}\in\mathcal{G}$ be a semi-continuous function, then for $\lambda\geqslant 0$, we get

$$R^{per,a,b}_{l,\rho_{\mathcal{B}},\mathcal{D}^b}(\mathcal{D}^a,g^{a,b})=\min_{\lambda}\{\lambda\rho_{\mathcal{B}}+\mathbb{E}_{v^a\sim\mathcal{D}^a}[\Omega_{\lambda,g^{a,b},\mathcal{D}^b}(v^a)]\}.\tag{3.4}$$

In fact, for each pair of (a,b), set $\Xi^{a,b}(v^a)=\int_{V^b}g^{a,b}(v^a,v^b)\mathcal{D}^b dv^b$,

$$\begin{aligned}\zeta^{a,b}_P&=\sup_{\mathcal{D}^{a'}\in P(V^a)}\left\{\int_{V^a}\Xi^{a,b}(v^a)\mathcal{D}^{a'}dv^a:\mathcal{WD}_p(\mathcal{D}^a,\mathcal{D}^{a'})\leqslant\xi\right\},\\ \zeta^{a,b}_D&=\inf_{\lambda\geqslant0}\left\{\lambda\xi^p-\int_{V^a}[\inf_{V^{a'}\in V^a}(\lambda d^p(v^a,v^{a'})-\Xi^{a,b}(v^a))]d\mathcal{D}^a(v^a)\right\}.\end{aligned}$$

Since $g^{a,b}(v^a,v^b)$ is an upper semi-continuous function for each pair (a,b), the growth rate of $\Xi^{a,b}(v^a)$ is zero if V^a is bounded, and otherwise $\limsup_{d(v^a,v')\to\infty}\frac{\Xi^{a,b}(v^a)-\Xi^{a,b}(v')}{d(v^a,v')}$, which is bounded, where $v'\in V^a$ is given. It follows that $\zeta^{a,b}_P=\zeta^{a,b}_D$ by the weak duality and strong duality theory, and thus (3.4) holds. Set $g\in\mathcal{G}$ as the generalized extension of $g^{a,b}$ such that $g|_{(a,b)}=g^{a,b}$, i.e., its restriction on pair (a,b) is $g^{a,b}$, and $\mathcal{G}$ is the conditional distribution of $\mathcal{G}$.

In order to obtain the desired theoretical results using statistical learning theory, we need to make some theoretical assumptions mathematically. First, the ontology instance space should be bounded, i.e., the diameter of V is finite. Second, for each pair (a,b), the upper semi-continue function $g^{a,b}$ is bounded, that is, for any $v^a \in V^a$ and $v^b \in V^b$, we have $0 \leqslant g^{a,b} \leqslant M^{a,b} < \infty$. Next, for any pair (a, b), function $g^{a,b}$, $v^a \in V^a$ and $v^b \in V^b$, $v^{a'} \in V^a$, there exists a constant $\lambda^{a,b}$ satisfying

$$g^{a,b}(v^{a'}, v^b) - g^{a,b}(v^a, v^b) \leqslant \lambda^{a,b} d_V(v^{a'}, v^a). \tag{3.5}$$

Note that parameter $\lambda^{a,b}$ depends on the pairwise perturbation ontology data v^a, v^b and the function $g^{a,b}$. For each pair of (a, b), let

$$\Phi_{g^{a,b},\mathcal{D}^a_{n_a},\mathcal{D}^b_{n_b}}(\lambda^{a,b}) = \mathbb{E}_{v^a_i \sim \mathcal{D}^a_{n_1}} \left\{ \sup_{v^{a'} \in V^a} \mathbb{E}_{v^b_j \sim \mathcal{D}^b_{n_b}} [g^{a,b}(v^{a'}, v^b_j) - \lambda^{a,b} d_{V^a}(v^a_i, v^{a'}) - g^{a,b}(v^a_i, v^b_j)] \right\}.$$

Then the assumption (3.5) holds if and only if for any $g^{a,b} \in \mathcal{G}$ and any S^a, S^b, we have $\{\lambda^{a,b} : \Phi_{g^{a,b},\mathcal{D}^a_{n_a},\mathcal{D}^b_{n_b}}(\lambda^{a,b}) = 0\} \neq \emptyset$. For each pair of (a, b), set

$$\lambda^{a*,b}_{g^{a,b},\mathcal{D}^a_{n_a},\mathcal{D}^b_{n_b}} = \inf\{\lambda^{a,b} : \Phi_{g^{a,b},\mathcal{D}^a_{n_a},\mathcal{D}^b_{n_b}}(\lambda^{a,b}) = 0\}.$$

As the function $g^{a,b}$ is determined by multi-dividing ontology function f, $\Phi_{g^{a,b},\mathcal{D}^a_{n_a},\mathcal{D}^b_{n_b}}(\lambda^{a,b})$ and $\lambda^{a*,b}_{g^{a,b},\mathcal{D}^a_{n_a},\mathcal{D}^b_{n_b}}$ defined above can be further denoted by $\Phi_{f,\mathcal{D}^a_{n_a},\mathcal{D}^b_{n_b}}(\lambda^{a,b})$ and $\lambda^{a*,b}_{f,\mathcal{D}^a_{n_a},\mathcal{D}^b_{n_b}}$. For any pair of (a, b) and $g^{a,b} \in \mathcal{G}^{a,b}$, set

$$\overline{\lambda}^{a,b} = \arg\min_{\lambda^{a,b} \geqslant 0} \{\lambda^{a,b} \rho_{\mathcal{B}} + \mathbb{E}_{v^a_i \sim \mathcal{D}^a_{n_a}} \Omega_{\lambda^{a,b}, g^{a,b}, \mathcal{D}^b_{n_b}}(v^a_i)\}$$

and

$$\bar{\lambda} = \arg\min_{\lambda \geqslant 0} \sum_{a=1}^{k-1} \sum_{b=a+1}^{k} \{\lambda \rho_{\mathcal{B}} + \mathbb{E}_{v^a_i \sim \mathcal{D}^a_{n_a}} \Omega_{\lambda, g^{a,b}, \mathcal{D}^b_{n_b}}(v^a_i)\}.$$

Moreover, we set

$$\lambda_{\max} = \max_{(a,b), 1 \leqslant a < b \leqslant k} \{\lambda^{a*,b}_{g^{a,b},\mathcal{D}^a_{n_a},\mathcal{D}^b_{n_b}}\}, \quad M_{\max} = \max_{(a,b), 1 \leqslant a < b \leqslant k} \{M^{a,b}\}.$$

Hence, $\overline{\lambda} \in [0, \frac{k(k-1)M_{\max}}{2}]$ if $\rho_{\mathcal{B}} \lambda^{a*,b}_{g^{a,b},\mathcal{D}^a_{n_a},\mathcal{D}^b_{n_b}} \geqslant M^{a,b}$ for each pair of (a, b). Otherwise, $\overline{\lambda} \in [0, \frac{k(k-1)\lambda_{max}}{2}]$. To simplify notation, we set

$$\overline{\lambda}^{a,b} \in [\iota^b_{g^{a,b},\mathcal{D}^a_{n_a},\mathcal{D}^b_{n_b}}, \iota^a_{g^{a,b},\mathcal{D}^a_{n_a},\mathcal{D}_{n_b}}] \quad \text{and} \quad \overline{\lambda} \in [\iota_{g,\mathcal{D}^b,\mathcal{D}^b}, \iota_{g,\mathcal{D}^b,\mathcal{D}^a}].$$

For a given standard ontology sample set $S = \{v_1, v_2, \ldots, v_n\}$ and a real-valued ontology function space $\mathcal{F}$, the empirical Rademacher complexity of ontology function space $\mathcal{F}$ with respect to otology sample set S is denoted by

$$\mathcal{R}_n(\mathcal{F}) = \frac{1}{n} \mathbb{E}[\sup_{f \in \mathcal{F}} \sum_{i=1}^{n} \sigma_i t(v_i)],$$

where $\sigma_1, \ldots, \sigma_n$ are independent random variables uniformly selected from $\{-1, 1\}$. For each pair of (a, b), set $0 \leqslant \lambda^{a,b}_1 \leqslant \lambda^{a,b}_2$, define

$$\Lambda^{a,b} = \{\Omega_{\lambda^{a,b}, g^{a,b} \mathcal{D}^b} : \lambda^{a,b} \in [\lambda^{a,b}_1, \lambda^{a,b}_2], g^{a,b} \in \mathcal{G}^{a,b}, \mathcal{D}^b \in P(V^b)\}, \quad D(V^a) = \max_{v^a_i, v^a_j} d(v^a_i, v^a_j),$$

and

$$\mathcal{C}(\mathcal{G}^{a,b}) = \int_0^{\infty} \sqrt{\log \mathcal{N}(\mathcal{G}^{a,b}, \|\cdot\|_{\infty}, \frac{x}{2})} dx,$$

where $\mathcal{N}(\mathcal{G}^{a,b}, \|\cdot\|_\infty, \frac{x}{2})$ is the cover number of function space $\mathcal{G}^{a,b}$ with respect to l_∞-norm and radius $\frac{x}{2}$. Set

$$\mathcal{C}(\mathcal{G}) = \max_{(a,b):1\leqslant a<b\leqslant k} \mathcal{C}(\mathcal{G}^{a,b}), \quad \Delta = \max_{(a,b):1\leqslant a<b\leqslant k} \{\lambda_2^{a,b} - \lambda_1^{a,b}\}.$$

Let Λ be a general function space in multi-dividing ontology learning setting, where the restriction on a specific pair (a,b) is $\Lambda|_{(a,b)} = \Lambda^{a,b}$. Then, the empirical Rademacher complexity of Λ for rate a ($a \in \{1,\cdot,k-1\}$) in multi-dividing ontology learning setting (given multi-dividing ontology sample set $S = (S^1, S^2, \dots, S^k) \in V^{n_1} \times V^{n_2} \times \cdots \times V^{n_k}$) is described by

$$\mathcal{R}_{n_a}(\Lambda) \leqslant \frac{12\mathcal{C}(\mathcal{G})}{\sqrt{n_a}} + \frac{12\Delta D(V^a)}{\sqrt{n_a}}. \tag{3.6}$$

For any $v^{a,*} \in V^a$, let

$$\mathcal{G}^{a,*,b} = \{g^{a,b} : g^{a,b}(v^{a,*}, v^b), g^{a,b} \in \mathcal{G}^{a,b}, v^b \in V^b\},$$

where $\mathcal{G}^{a,b}$ is defined in (3.3). Let $\mathcal{R}_{n_b}(\mathcal{G}^{a,*,b})$ be the expected Rademacher complexity of $\mathcal{G}^{a,*,b}$ in multi-dividing ontology setting. Then

$$\mathcal{R}_{n_b}(\mathcal{G}^{a,*,b}) \leqslant \frac{12\mathcal{C}(\mathcal{G}^{a,b})}{\sqrt{n_b}}. \tag{3.7}$$

4. Main result and proof

Our first main result manifested as follows, which focuses on the pairwise function g for every rate pair in perturbation multi-dividing ontology learning algorithm setting.

Theorem 4.1. *Assume the hypothesis defined in the last section are satisfied. Set*

$$[\iota_1^{a,b}, \iota_2^{a,b}] = \cup_{g^{a,b},\mathcal{D}^a_{n_a},\mathcal{D}^b_{n_b}} [\iota^b_{g^{a,b},\mathcal{D}^a_{n_a},\mathcal{D}^b_{n_b}}, \iota^a_{g^{a,b},\mathcal{D}^a_{n_a},\mathcal{D}_{n_b}}], \quad \Theta^{a,b}_{\rho_{\mathcal{B}}} = \iota_2^{a,b} - \iota_1^{a,b}.$$

Then, for any $g \in \mathcal{G}$, then the following inequality hold with possibility at least $1-\delta$,

$$R^{per}_{l,\mathcal{D}}(g,\mathcal{B}) = \sum_{a=1}^{k-1} \sum_{b=a+1}^{k} \frac{1}{n_a n_b} \sum_{i=1}^{n_a} \sum_{j=1}^{n_b} l(f, v_i^a, v_j^b) + \sum_{i=1}^{4} \chi_i,$$

where

$$\chi_1 = \sum_{a=1}^{k-1} \sum_{b=a+1}^{k} \lambda^{a*,b}_{g^{a,b},\mathcal{D}^a_{n_a},\mathcal{D}^b_{n_b}} \rho_{\mathcal{B}},$$

$$\chi_2 = \sum_{a=1}^{k-1} \sum_{b=a+1}^{k} 24\mathcal{C}(\mathcal{G}^{a,b})\left(\frac{1}{\sqrt{n_a}} + \frac{1}{\sqrt{n_b}}\right),$$

$$\chi_3 = \sum_{a=1}^{k-1} \sum_{b=a+1}^{k} M^{a,b}\left(\sqrt{\frac{\log\frac{1}{\delta}}{2n_a}} + \sqrt{\frac{\log\frac{1}{\delta}}{2n_b}}\right),$$

and

$$\chi_4 = \sum_{a=1}^{k-1} \sum_{b=a+1}^{k} \frac{24}{n_a} \Theta^{a,b}_{\rho_{\mathcal{B}}} D(V^a).$$

Proof of Theorem 4.1. For each pair of (a,b) with $1 \leqslant a < b \leqslant k$, the partial term of χ_i $(i \in \{1,2,3,4\})$ are denoted by $\chi_1^{a,b}$, $\chi_2^{a,b}$, $\chi_3^{a,b}$, and $\chi_4^{a,b}$, respectively. Hence,

$$\chi_1 = \sum_{a=1}^{k-1}\sum_{b=a+1}^{k}\chi_1^{a,b}, \qquad \chi_2 = \sum_{a=1}^{k-1}\sum_{b=a+1}^{k}\chi_2^{a,b}, \qquad \chi_3 = \sum_{a=1}^{k-1}\sum_{b=a+1}^{k}\chi_3^{a,b}, \qquad \chi_4 = \sum_{a=1}^{k-1}\sum_{b=a+1}^{k}\chi_4^{a,b}.$$

Note that

$$\begin{aligned} &R^{per,a,b}_{\rho_{\mathcal{B}},\mathcal{D}^b}(\mathcal{D}^a, g^{a,b}) - \widehat{R}^{per,a,b}_{\rho_{\mathcal{B}},\mathcal{D}^b_{n_2}}(\mathcal{D}^a_{n_1}, g^{a,b}) \\ &\qquad = R^{per,a,b}_{\rho_{\mathcal{B}},\mathcal{D}^b}(\mathcal{D}^a, g^{a,b}) - \widehat{R}^{per,a,b}_{\rho_{\mathcal{B}},\mathcal{D}^b_{n_2}}(\mathcal{D}^a, g^{a,b}) + \widehat{R}^{per,a,b}_{\rho_{\mathcal{B}},\mathcal{D}^b_{n_2}}(\mathcal{D}^a, g^{a,b}) - \widehat{R}^{per,a,b}_{\rho_{\mathcal{B}},\mathcal{D}^b_{n_2}}(\mathcal{D}^a_{n_1}, g^{a,b}). \end{aligned}$$

Re-define

$$\Lambda^{a,b} = \{\Omega_{\lambda^{a,b},g^{a,b}\mathcal{D}^b_{n_b}} : \lambda^{a,b} \in [\lambda_1^{a,b}, \lambda_2^{a,b}], g^{a,b} \in \mathcal{G}^{a,b}, \mathcal{D}^b_{n_b} \in P(V^b)\}.$$

In terms of $[\iota_1^{a,b}, \iota_2^{a,b}] \subset [0, \frac{M^{a,b}}{\rho_{\mathcal{B}}}]$, (3.6), and $|\Omega| \leqslant M^{a,b}$, we acquire the following inequality with possibility at least $1-\delta$,

$$\begin{aligned} &\widehat{R}^{per}_{l,\rho_{\mathcal{B}},\mathcal{D}_{n_1,n_2,\dots,n_k}}(\mathcal{D}, g) - \widehat{R}^{per}_{l,\rho_{\mathcal{B}},\mathcal{D}_{n_1,n_2,\dots,n_k}}(\mathcal{D}_{n_1,n_2,\dots,n_k}, g) \\ &\qquad = \sum_{a=1}^{k-1}\sum_{b=a+1}^{k}\left\{\widehat{R}^{per,a,b}_{l,\rho_{\mathcal{B}},\mathcal{D}^b_{n_2}}(\mathcal{D}^a, g^{a,b}) - \widehat{R}^{per,a,b}_{l,\rho_{\mathcal{B}},\mathcal{D}^b_{n_2}}(\mathcal{D}^a_{n_1}, g^{a,b})\right\} \\ &\qquad = \min_{\lambda}\sum_{a=1}^{k-1}\sum_{b=a+1}^{k}\left\{\lambda\rho_{\mathcal{B}} + \int_{V^a}\Omega_{\lambda,g^{a,b},\mathcal{D}^b}(v^a)d\mathcal{D}^a(v^a)\right\} - \left\{\bar{\lambda}\rho_{\mathcal{B}} + \frac{1}{n_a}\sum_{i=1}^{n_a}\Omega_{\lambda,g^{a,b},\mathcal{D}^b}(v_i^a)\right\} \\ &\qquad \leqslant \sum_{a=1}^{k-1}\sum_{b=a+1}^{k}\left\{\int_{V^a}\Omega_{\lambda,g^{a,b},\mathcal{D}^b}(v^a)d\mathcal{D}^a(v^a) - \frac{1}{n_a}\sum_{i=1}^{n_a}\Omega_{\lambda,g^{a,b},\mathcal{D}^b}(v_i^a)\right\} \\ &\qquad \leqslant \sum_{a=1}^{k-1}\sum_{b=a+1}^{k}\sup_{\Omega\in\Lambda^{a,b}}\left\{\int_{V^a}\Omega(v^a)d\mathcal{D}^a(v^a) - \frac{1}{n_a}\sum_{i=1}^{n_a}\Omega(v_i^a)\right\} \\ &\qquad \leqslant \sum_{a=1}^{k-1}\sum_{b=a+1}^{k}\left\{2\widehat{R}_{n_a}(\Lambda^{a,b}) + M^{a,b}\sqrt{\frac{\log\frac{1}{\delta}}{2n_a}}\right\} \\ &\qquad \leqslant \sum_{a=1}^{k-1}\sum_{b=a+1}^{k}\left\{\frac{24\mathcal{C}(\mathcal{G})}{\sqrt{n_a}} + M^{a,b}\sqrt{\frac{\log\frac{1}{\delta}}{2n_a}} + \chi_4^{a,b}\right\}. \end{aligned}$$

Set

$$\begin{aligned} \widehat{\lambda}^{a,b} &= \arg\min_{\lambda^{a,b}\geqslant 0}\left\{\lambda^{a,b}\rho_{\mathcal{B}} + \mathbb{E}_{v^a\sim\mathcal{D}^a}[\Omega_{\lambda^{a,b},g^{a,b},\mathcal{D}^b_{n_b}}(v_i^a)]\right\}, \\ v^{a,b,*} &= \arg\sup_{v^{a\prime}\in V^a}[\mathbb{E}_{v^b\sim\mathcal{D}^b}g^{a,b}(v^{a\prime}, v^b) - \widehat{\lambda}^{a,b}d_V(v^{a\prime}, v^a)]. \end{aligned}$$

Using (3.7), we deduce the following inequality with probability at least $1-\delta$:

$$\begin{aligned} &R^{per}_{l,\rho_{\mathcal{B}},\mathcal{D}}(\mathcal{D}, g) - \widehat{R}^{per}_{l,\rho_{\mathcal{B}},\mathcal{D}_{n_1,n_2,\dots,n_k}}(\mathcal{D}, g) \\ &\qquad = \sum_{a=1}^{k-1}\sum_{b=a+1}^{k}\left\{R^{per,a,b}_{l,\rho_{\mathcal{B}},\mathcal{D}^b}(\mathcal{D}^a, g^{a,b}) - \widehat{R}^{per,a,b}_{l,\rho_{\mathcal{B}},\mathcal{D}^b_{n_2}}(\mathcal{D}^a, g^{a,b})\right\} \end{aligned}$$

$$
\begin{aligned}
&= \min_{\lambda} \sum_{a=1}^{k-1} \sum_{b=a+1}^{k} \left\{ \lambda \rho_{\mathcal{B}} + \int_{V^a} \Omega_{\lambda, g^{a,b}, \mathcal{D}^b}(v^a) d\mathcal{D}^a(v^a) \right\} - \left\{ \widehat{\lambda} \rho_{\mathcal{B}} + \int_{V^a} \Omega_{\widehat{\lambda}, g^{a,b}, \mathcal{D}^b_{n_b}}(v^a) d\mathcal{D}^a(v^a) \right\} \\
&\leqslant \sum_{a=1}^{k-1} \sum_{b=a+1}^{k} \left\{ \left[\int_{V^a} \Omega_{\lambda, g^{a,b}, \mathcal{D}^b}(v^a) d\mathcal{D}^a(v^a) - \int_{V^a} \Omega_{\widehat{\lambda}, g^{a,b}, \mathcal{D}^b_{n_b}}(v^a) d\mathcal{D}^a(v^a) \right\} \right. \\
&= \sum_{a=1}^{k-1} \sum_{b=a+1}^{k} \left\{ \left[\int_{V^a} \left\{ \sup_{v^{a\prime} \sim \mathcal{D}^a} [\mathbb{E}_{v^b \sim \mathcal{D}^b} g^{a,b}(v^{a\prime}, v^b) - \widehat{\lambda} d_V(v^{a\prime}, v^a)] \right. \right. \right. \\
&\quad - \left. \left. \left. \sup_{v^{a\prime} \sim \mathcal{D}^a} [\mathbb{E}_{v^b_j \sim \mathcal{D}^b_{n_b}} g^{a,b}(v^{a\prime}, v^b_j) - \widehat{\lambda} d_V(v^{a\prime}, v^a)] \right\} d\mathcal{D}^a(v^a) \right\} \right. \\
&\leqslant \sum_{a=1}^{k-1} \sum_{b=a+1}^{k} \left\{ \sup_{g^{a,b} \in \mathcal{G}^{a,b}} \left[\int_{V^a} g^{a,b}(v^{a,b,*}, v^b) d\mathcal{D}^b(v^b) - \frac{1}{n_2} \sum_{j=1}^{n_2} (v^{a,b,*}, v^b_j) \right] \right\} \\
&\leqslant \sum_{a=1}^{k-1} \sum_{b=a+1}^{k} \left\{ 2\widehat{R}_{n_2}(\mathcal{D}^{a,b,*}) + M^{a,b} \sqrt{\frac{\log \frac{1}{\delta}}{2n_b}} \right\}.
\end{aligned}
$$

Combining two parts together, we infer

$$
R^{per,a,b}_{\rho_{\mathcal{B}}, \mathcal{D}^b}(\mathcal{D}^a, g^{a,b}) - \widehat{R}^{per,a,b}_{\rho_{\mathcal{B}}, \mathcal{D}^b_{n_2}}(\mathcal{D}^a_{n_1}, g^{a,b}) \leqslant \chi_2^{a,b} + \chi_3^{a,b} + \chi_4^{a,b}
$$

and

$$
\widehat{R}^{per}_{l, \rho_{\mathcal{B}}, \mathcal{D}_{n_1, n_2, \dots, n_k}}(\mathcal{D}, g) - \widehat{R}^{per}_{l, \rho_{\mathcal{B}}, \mathcal{D}_{n_1, n_2, \dots, n_k}}(\mathcal{D}, g) \leqslant \chi_2 + \chi_3 + \chi_4,
$$

holds with probability at least $1 - \delta$. On the other hand, in light of (3.4), we obtain

$$
\begin{aligned}
&\widehat{R}^{per}_{l, \rho_{\mathcal{B}}, \mathcal{D}_{n_1, n_2, \dots, n_k}}(\mathcal{D}, g) \\
&\quad = \sum_{a=1}^{k-1} \sum_{b=a+1}^{k} \widehat{R}^{per,a,b}_{\rho_{\mathcal{B}}, \mathcal{D}^b_{n_2}}(\mathcal{D}^a_{n_a}, g^{a,b}) \\
&\quad = \min_{\lambda} \sum_{a=1}^{k-1} \sum_{b=a+1}^{k} \{\lambda \rho_{\mathcal{B}} + \mathbb{E}_{v^a_i \sim \mathcal{D}^a_{n_a}} [\Omega_{\lambda, g^{a,b}, \mathcal{D}^b_{n_b}}(v^a_i)]\} \\
&\quad = \min_{\lambda} \sum_{a=1}^{k-1} \sum_{b=a+1}^{k} \{\lambda \rho_{\mathcal{B}} + \mathbb{E}_{v^a_i \sim \mathcal{D}^a_{n_a}} [\Omega_{\lambda, g^{a,b}, \mathcal{D}^b_{n_b}}(v^a_i) - E_{v^b_j \sim \mathcal{D}^b_{n_b}} [g^{a,b}(v^a_i, v^b_j)]]\} \\
&\qquad + \sum_{a=1}^{k-1} \sum_{b=a+1}^{k} \{E_{v^a_i \sim \mathcal{D}^a_{n_a}} [E_{v^b_j \sim \mathcal{D}^b_{n_b}} [g^{a,b}(v^a_i, v^b_j)]]\} \\
&\quad = \min_{\lambda} \sum_{a=1}^{k-1} \sum_{b=a+1}^{k} \{\lambda \rho_{\mathcal{B}} + \Phi_{g^{a,b}, \mathcal{D}^a_{n_a}, \mathcal{D}^b_{n_b}}(\lambda)\} + \sum_{a=1}^{k-1} \sum_{b=a+1}^{k} \{\frac{1}{n_a n_b} \sum_{i=1}^{n_a} \sum_{j=1}^{n_b} g^{a,b}(v^a_i, v^b_j)\} \\
&\quad \leqslant \sum_{a=1}^{k-1} \sum_{b=a+1}^{k} \{\lambda^a_{g^{a,b}, \mathcal{D}^a_{n_a}, \mathcal{D}^b_{n_b}} \rho_{\mathcal{B}} + \Phi_{g^{a,b}, \mathcal{D}^a_{n_a}, \mathcal{D}^b_{n_b}}(\lambda^a_{g^{a,b}, \mathcal{D}^a_{n_a}, \mathcal{D}^b_{n_b}}) + \frac{1}{n_a n_b} \sum_{i=1}^{n_a} \sum_{j=1}^{n_b} g^{a,b}(v^a_i, v^b_j)\} \\
&\quad = \sum_{a=1}^{k-1} \sum_{b=a+1}^{k} \{\lambda^a_{g^{a,b}, \mathcal{D}^a_{n_a}, \mathcal{D}^b_{n_b}} \rho_{\mathcal{B}} + \frac{1}{n_a n_b} \sum_{i=1}^{n_a} \sum_{j=1}^{n_b} g^{a,b}(v^a_i, v^b_j)\} \\
&\quad = \sum_{a=1}^{k-1} \sum_{b=a+1}^{k} \{\chi_1^{a,b} + \frac{1}{n_a n_b} \sum_{i=1}^{n_a} \sum_{j=1}^{n_b} g^{a,b}(v^a_i, v^b_j)\}.
\end{aligned}
$$

Therefore

$$R^{per,a,b}_{\rho_{\mathcal{B}},\mathcal{D}^b}(\mathcal{D}^a, g^{a,b}) \leqslant \frac{1}{n_a n_b}\sum_{i=1}^{n_a}\sum_{j=1}^{n_b} g^{a,b}(v_i^a, v_j^b) + \sum_{i=1}^{4}\chi_1^{a,b},$$

and thus

$$R^{per}_{\rho_{\mathcal{B}},\mathcal{D}}(\mathcal{D}, g) \leqslant \sum_{a=1}^{k-1}\sum_{b=a+1}^{k}\frac{1}{n_a n_b}\sum_{i=1}^{n_a}\sum_{j=1}^{n_b} g^{a,b}(v_i^a, v_j^b) + \sum_{i=1}^{4}\chi_1.$$

Therefore, the desired conclusion is derived. □

The main result on perturbation multi-dividing ontology learning algorithm is stated in the following theorem.

Theorem 4.2. *Suppose for any* $g, g' \in \mathcal{G}$, *there is a constant* L *satisfying* $\|g - g'\|_\infty \leqslant L\|f - f'\|_\infty$. *Under the condition in Theorem 4.1, we have*

$$R^{per}_{l,\mathcal{D}}(f, \mathcal{B}) = \sum_{a=1}^{k-1}\sum_{b=a+1}^{k}\frac{1}{n_a n_b}\sum_{i=1}^{n_a}\sum_{j=1}^{n_b} l(f, v_i^a, v_j^b) + \sum_{i=1}^{4}\chi_i,$$

with possibility at least $1 - \delta$.

Proof of Theorem 4.2. Let $\{f_1, \dots, f_q\}$ be the center of balls with radius $\frac{x}{2L}$, which covers ontology function space $\mathcal{F}$. For any multi-dividing ontology function $f \in \mathcal{F}$, there exists $f' \in \{f_1, \dots, f_q\}$ such that $\|f - f'\|_\infty \leqslant \frac{x}{2L}$, and hence $\|g' - g\|_\infty \leqslant L\|f - f'\|_\infty \leqslant \frac{x}{2}$, $\mathcal{N}(\mathcal{G}, \|\cdot\|_\infty, \frac{x}{2}) \leqslant \mathcal{N}(\mathcal{F}, \|\cdot\|_\infty, \frac{x}{2L})$, and $\mathcal{C}(\mathcal{G}) \leqslant \int_0^\infty \sqrt{\log \mathcal{N}(\mathcal{F}, \|\cdot\|_\infty, \frac{x}{2L})}dx = L\mathcal{C}(\mathcal{F})$. Then the desired result follows from the conclusion in Theorem 4.1. □

5. Conclusion, discussion, and open problem

In practical applications, most of the ontology graph structures are trees or approximate tree structures due to concept classification. The multi-dividing ontology learning algorithm is specially designed for the tree structure, and each branch below the top-level vertex becomes a class. This is the reason for the success of the multi-dividing ontology learning algorithm.

In our article, we proposed the perturbation multi-dividing ontology learning algorithm and determined its generalized risk bounds under certain conditions in terms of statistical learning theory approach. Since multi-dividing ontology learning algorithms are widely used in tree-structured ontology graphs, the algorithms and theoretical analysis presented in this paper have potential guiding roles for practical applications.

The biggest flaw in the theoretical analysis of this paper is the assumption that for each pair of vertex classes, the perturbation of v^a still obeys the same distribution $\mathcal{D}^a$, the entire framework is discussed under the assumption of independent and identical distribution (in short, IID), and the non-IID case is not considered. However, in practical applications, the heterogeneity of ontology data is a common problem and becomes the most crucial challenge in ontology learning. For the same concept set, different users build ontologies according to their own needs, and these constructed ontologies are often heterogeneous due to the difference between expert cognition and model sizes. Even between different branches within the same ontology, the corresponding ontology data are likely to be heterogeneous. For instance, the "live" ontology is depicted in Figure 1.

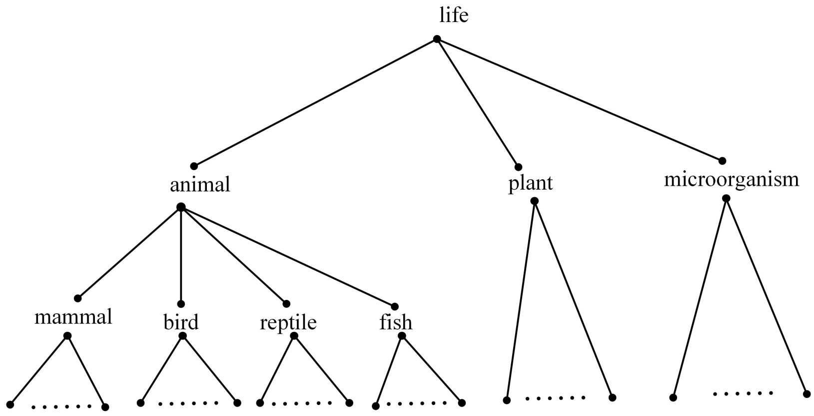

Figure 1: The "live" ontology.

To apply multi-dividing ontology learning approach, set $k = 3$ since there are three components below top vertex "live", which correspond to "animal", "plant" and "microorganism", respectively. Clearly, the ontology data in these three rates are heterogeneous due to huge differences between these three organisms. Even more, the corresponding ontology data of mammal, bird, reptile, and fish in the second-level division may also be heterogeneous, because the gap between each species is obvious.

Finally, perturbation often results in heterogeneous data. For example, in image processing, a new image dataset is obtained by rotating the image by a fixed angle, which is heterogeneous with the original data. Another instance, the perturbation is generalized by affine transformation (ϑ^a, ϖ^a), where the ontology data v_i^a is changed by $\vartheta^a v_i^a + \varpi^a$. However, the new ontology data is heterogeneous with the original ontology data.

In our perturbation multi-dividing ontology learning setting, the whole ontology data has a underlying distribution $\mathcal{D}$, and ontology data in each rate drawn according to $\mathcal{D}^a$, which is a conditional distribution of $\mathcal{D}$. In addition, our theoretical analysis is based on a basic assumption that the perturbed data follows the same distribution as the original data. Therefore, there is a certain degree of defects in our analysis, and it's still an open problem of perturbation multi-dividing ontology learning algorithm in heterogeneity hypothesis. Further study is needed on the non-IID ontology learning algorithm and specifically in (perturbation) multi-dividing ontology learning.

Acknowledgment

Many thanks to the reviewers for their constructive comments on revisions to the article.

References

[1] W. Abbes, D. Sellami, S. Marc-Zwecker, C. Zanni-Merk, *Fuzzy ontology for automatic skin lesion classification*, J. Test. Eval., **49** (2021), 2994–3014. 1

[2] S. K. Anand, S. Kumar, *Uncertainty analysis in ontology-based knowledge representation*, New Gener. Comput., **40** (2022), 339–376. 1

[3] A. Assa, K. N. Plataniotis, *Wasserstein-distance-based Gaussian mixture reduction*, IEEE Signal Process. Lett., **25** (2018), 1465–1469. 3.1

[4] J. Božić, *Ontology-based metamorphic testing for chatbots*, Softw. Qual. J., **30** (2022), 227–251. 1

[5] L. Buoncompagni, S. Y. Kareem, F. Mastrogiovanni, *Human activity recognition models in ontology networks*, IEEE Trans. Cybern., **52** (2022), 5587–5606. 1

[6] T. Ö. Çelik, A. Jamneshan, G. Montúfar, B. Sturmfels, L. Venturello, *Wasserstein distance to independence models*, J. Symb. Comput., **104** (2021), 855–873. 3.1

[7] H.-B. Chen, J. Niles-Weed, *Asymptotics of smoothed Wasserstein distances*, Potential Anal., **56** (2022), 571–595. 3.1

[8] C. Dubslaff, P. Koopmann, A.-Y. Turhan, *Enhancing probabilistic model checking with ontologies*, Form. Asp. Comput., **33** (2021), 885–921. 1

[9] W. Gao, Y. Chen, *Approximation analysis of ontology learning algorithm in linear combination setting*, J. Cloud Comput.-Adv. S., **9** (2020), 10 pages. 1

[10] W. Gao, M. R. Farahani, *Generalization bounds and uniform bounds for multi-dividing ontology algorithms with convex ontology loss function*, Comput. J., **60** (2017), 1289–1299. 1

[11] W. Gao, J. L. G. Guirao, B. Basavanagoud, J. Wu, *Partial multi-dividing ontology learning algorithm*, Inform. Sci., **467** (2018), 35–58. 1

[12] W. Gao, Y. Zhang, J. L. G. Guirao, Y. Chen, *A discrete dynamics approach to sparse calculation and applied in ontology science*, J. Differ. Equ. Appl., **25** (2019), 1239–1254. 1

[13] J. Lu, J. Ma, X. Zheng, G. Wang, H. Li, D. Kiritsis, *Design ontology supporting model-based systems engineering formalisms*, IEEE Syst. J., **16** (2022), 5465–5476. 1

[14] I. Osman, S. B. Yahia, G. Diallo, *Ontology integration: approaches and challenging issues*, Inf. Fusion, **71** (2021), 38–63. 1

[15] V. M. Panaretos, Y. Zemel, *Statistical aspects of Wasserstein distances*, Annu. Rev. Stat. Appl., **6** (2019), 405–431. 3.1

[16] J. Shi, Y. Wang, *Hyperbolic Wasserstein distance for shape indexing*, IEEE Trans. Pattern Anal. Mach. Intell., **42** (2020), 1362–1376. 3.1

[17] P. K. Sinha, S. B. Gajbe, S. Debnath, S. Sahoo, K. Chakraborty, S. S. Mahato, *A review of data mining ontologies*, Data Technol. Appl., **56** (2022), 172–204. 1

[18] C. Sirichanya, K. Kraisak, *Semantic data mining in the information age: A systematic review*, Int. J. Intell. Syst., **36** (2021), 3880–3916. 1

[19] L. Xiong, M. Ye, D. Zhang, Y. Gan, X. Li, Y. Zhu. *Source data-free domain adaptation of object detector through domain-specific perturbation*, Int. J. Intell. Syst., **36** (2021), 21 pages. 1

[20] X. Xue, J. Chen, X. Yao, *Efficient user involvement in semiautomatic ontology matching*, IEEE Trans. Emerg. Top. Comput. Intell., **5** (2021), 214–224. 1

[21] L. Yang, K. Cormican, M. Yu, *Ontology learning for systems engineering body of knowledge*, IEEE Trans. Ind. Inform., **17** (2021), 1039–1047. 1

[22] L. Zhu, G. Hua, *Theoretical perspective of multi-dividing ontology learning trick in two-sample setting*, IEEE Access, **8** (2020), 220703–220709. 1