Locally Ontology Relaxed Stability Analysis in Various Ontology Learning Settings

Shu Gong, Xinxin Huang, Caihua Qiu, and Wei Gao, *Member, IAENG,*

***Abstract*—Ontology is an effective tool for processing concept semantics, and in the ontology learning algorithm, all the semantic information of each vertex is expressed by a multi-dimensional vector. The essence of ontology learning algorithm is to obtain ontology function in terms of ontology data samples, so as to map each concept in ontology to a real number. Stability is the foundation of the ontology learning algorithm and the guarantee of its generalization ability. This article relaxes the original uniformly stable hypothesis and proposes the concept of locally ontology relaxed stability. And under the setting of reproducing kernel Hilbert space, the upper bound of stability is verified. Under the framework of random ontology algorithm, the original concept is redefined. The error bounds in general, the reproducing kernel Hilbert space and the stochastic ontology learning algorithm frameworks are obtained in terms of their respective stability definitions.**

***Index Terms*—ontology, similarity computing, stability, generalization bound.**

I. Introduction

As a semantic tool, ontology expresses the interrelationships between concepts in light of graph structures. It uses vertices to represent concepts, and the edges between vertices to represent the relationship between concepts. In the field of data representation, the information corresponding to a certain concept is denoted by a vector, so that the ontology function can be described as a dimensionality reduction operator that maps a high-dimensional space vector to a low-dimensional space. The so-called ontology learning is used to obtain the optimal ontology function in view of learning ontology sample points. Due to its powerful representation ability, ontology has a wide range of applications in various engineering fields (several related literatures can be referred to Stratogiannis et al. [1], Epstein et al. [2], Gheisari et al. [3], Selvalakshmi et al. [4], Mavracic et al. [5], Goncalves et al. [6], Maitra et al. [7], Zhu et al. [8] and [9], and Lan et al. [10] and [11]).

Manuscript received December 5, 2021; revised March 24, 2022.

This work is supported by 2021 Guangdong Basic and Applied Basic Youth Fund Project (No. 2021A1515110834), National Natural Science Foundation of China (No. 51574232), Guangdong University of Science and Technology University Major Scientific Research Achievement Cultivation Program Project 2020 (No. GKY-2020CQPY-2), Guangdong Provincial Department of Education Project (No. 2020KTSCX166), and Guangdong University of Science and Technology Project (No. GKY-2020KYZDK-10).

S. Gong is an associate professor of Department of Computer Science, Guangdong University Science and Technology, Dongguan 523083, China (email: gongshu_gk@126.com).

X. X. Huang is an associate professor of Department of Computer Science, Guangdong University Science and Technology, Dongguan 523083, China (email: 6280123@qq.com).

C. H. Qiu is an associate professor of Department of Computer Science, Guangdong University Science and Technology, Dongguan 523083, China (email: 32320062@qq.com).

W. Gao is a professor of School of Information and Technology, Yunnan Normal University, Kunming, 650500 China (e-mail: gaowei@ynnu.edu.cn).

In recent years, ontology algorithm and applications have raised great attention among scientists and engineers. Son and Lee [12] separated 3D geometry into several parts by means of PointNet and constructed local ontology in light of summarizing the features of each part. Patel et al. [13] provided ontology to represent semantic information about the impact of Covid-19 on the banking sector of India. Xue and Chen [14] established the semantic links between heterogenous biomedical concepts called biomedical ontology matching. Lakzaei and Shamsfard [15] suggested a new trick to automatically obtain an OWL ontology using a relational database. Ratnaike et al. [16] collated 26348 human phenotype ontology terms to build the MitoPhen database. Rahman and Hussain [17] introduced a light-weight dynamic ontology in view of the most important concepts and clustering approach. Perea-Romero et al. [18] improved clinical and molecular SRDs diagnosis in terms of structuring phenotypic ontology and next-generation sequencing (NGS)-based pipelines. Bao et al. [19] determined the ontology-based modeling trick for assembly resource and process. Chen et al. [20] developed an ontology-based Bayesian network framework to express causal relationships between design parameters or process parameters and structure properties or mechanical properties. Belabbes et al. [21] considered the problem of dealing with inconsistency in lightweight ontologies.

The ontology learning algorithms which combine ontology and machine learning techniques have attracted attention from both theoretical analysis and engineering applications, especially the stability analysis of ontology learning tricks. Wu et al. [22] raised disequilibrium multi dividing ontology learning algorithm. Gao et al. [23] determined generalization bounds and uniform bounds for multi-dividing ontology algorithms with convex ontology loss function. Gao et al. [24] suggested partial multi-dividing ontology learning approach. Gao et al. [25] introduced ontology geometry distance computation in terms of deep learning trick. Gao et al. [26] raised a new ontology learning algorithm using discrete dynamics sparse calculation method.

In the process of ontology learning, we usually divide the ontology dataset into training set and test set. The training set is the ontology sample set, which is used to obtain the optimal ontology function, and the test set is used to test the quality of the ontology function. In order to make the learned ontology function generalized, that is, to show excellent performance on the test set, the ontology learning algorithm must be required to have a certain stability. That is, changing a small number of ontology samples will lead to significant changes in the final learned ontology function.

Ontology algorithm stability is usually expressed in two ways: loss stability and error stability. However, this setting often requires some preconditions, such as each ontology sample point is equally important and treated without discrimination. And the weakness of this ontology stability framework is that it can only give global stability and cannot reflect the local stability of a certain part of the vertices. Based on the weakness of the original ontology theoretical framework, the main contribution of this work is to study the ontology stability in relaxed setting. New classes of ontology stability are defined and the corresponding generalization bounds are deduced.

The rest of the paper is organized as follows: the notations, terminologies and new definitions are explained in the next section, and then the main results and proofs are manifested in the third section. Finally, we give the conclusion and future work.

II. Setting

Let $S = \{z_1, z_2, \cdots, z_n\}$ be the ontology training set, where z_i is independent and identically distributed which is drawn from a distribution $\mathcal{D}$ on the ontology space Z. In supervised ontology learning setting, $Z = V \times Y$, where V and Y are input space and label space respectively; while in unsupervised ontology learning setting, $Z = V$. For an ontology function class $\mathcal{F}$, a learning algorithm $\mathcal{A} : Z^n \to \mathcal{F}$ gets an ontology function $\mathcal{A}_S \to \mathcal{F}$ by means of ontology training set S. For a given ontology training set S with n ontology vertices, let $S^i = \{z_1, \cdots, z_{i-1}, z_i^{'}, z_{i+1}, \cdots, z_n\}$ be the new ontology training set by replacing the i-th element from S where $1 \le i \le n$, and $S^{\backslash i,j} = \{z_1, \cdots, z_{i-1}, , z_{i+1}, \cdots, z_{j-1}, , z_{j+1}, \cdots, z_n\}$ be the new ontology training set by deleting the i-th and j-th elements from S where $1 \le i < j \le n$. For any ontology input z, we consider an ontology loss function $l(f, z)$ with notation $f = \mathcal{A}_S$.

Let $\beta_n(\cdot, \cdot, \cdot)$ be functions with $n \ge 3$ that each maps any $z_1, z_2, z_3 \in Z$ to a positive score. Now, we introduce PO ontology relaxed stability and LTO ontology relaxed stability as follows, where PO stands for "replace one" and LTO stands for "leave two out".

Definition 1: (Locally PO Ontology Relaxed Stability) An ontology algorithm $\mathcal{A}$ has locally PO ontology relaxed stability $\beta_n(\cdot, \cdot, \cdot)$ with respect to the ontology loss function l if, for any n, the inequality

$$|l(\mathcal{A}_S, z) - l(\mathcal{A}_{S^i}, z)| \le \beta_n(z_i, z_i^{'}, z)$$

establishes for any $S \in Z^n$, $1 \le i \le n$ and $z_i^{'}, z \in Z$.

Definition 2: (Locally LTO Ontology Relaxed Stability) An ontology algorithm $\mathcal{A}$ has locally LTO ontology relaxed stability $\beta_n(\cdot, \cdot, \cdot)$ with respect to the ontology loss function l if, for any n, the inequality

$$|l(\mathcal{A}_S, z) - l(\mathcal{A}_{S^{\backslash i,j}}, z)| \le \beta_n(z_i, z_j, z)$$

establishes for any $S \in Z^n$, $1 \le i < j \le n$ and $z \in Z$.
Compare Definition 2 with uniform LTO ontology stability introduced by Wu et al. [27] which is stated as

$$|l(\mathcal{A}_S, z) - l(\mathcal{A}_{S^{\backslash i,j}}, z)| \le \beta_n^{LTO}$$

holds for any $S \in Z^n$, $1 \le i < j \le n$ and $z \in Z$. The relationship between them is that $\beta_n^{LTO} = \sup_{z_i, z_j, z} \beta_n(z_i, z_j, z)$, and it also holds for the relationship between standard PO uniform ontology stability and locally PO ontology relaxed stability introduced in Definition 1.

Consider that the ontology function f which is parameterized by θ (i.e., $f = f_\theta$) and in this case the ontology loss $l(f, z)$ can be re-written as $l(\theta, z)$. The aim of ontology learning algorithm $\mathcal{A}$ is to output $f_{\widehat{\theta}}$ such that

$$\widehat{\theta} = \frac{1}{n} \mathrm{argmin}_{\theta \in \Theta} \sum_{i=1}^{n} l(\theta, z_i).$$

By setting $\widehat{\theta}^{\backslash i,j} = \frac{1}{n} \mathrm{argmin}_{\theta \in \Theta} \sum_{k \ne i, k \ne j} l(\theta, z_k)$, we have the following approximation:

$$\begin{aligned} \beta_n(z_i, z_j, z) &= |l(\widehat{\theta}, z) - \widehat{\theta}^{\backslash i,j}| \\ &\approx \frac{1}{n} |\bigtriangledown_\theta \, l(\widehat{\theta}, z) \bigtriangledown_\theta \, H_{\widehat{\theta}}^{-1} l(\widehat{\theta}, z_i)|, \end{aligned}$$

where $H_{\widehat{\theta}} = \frac{1}{n} \sum_{i=1}^{n} \bigtriangledown^2 l(\widehat{\theta}, z_i)$ is the Hessian matrix. In what follows, we always keep the following hypothesis: for the function $\beta_n(\cdot, \cdot, \cdot)$, for any $z_1, z_2, z_3 \in Z$, $\beta_n(z_1, z_2, z_3) = \frac{\beta(z_1, z_2, z_3)}{n}$ for several function $\beta(\cdot, \cdot, \cdot)$ which has nothing to do with n. Furthermore, $\beta(\cdot, \cdot, z)$ is the function of its first and second parameters which are L-Lipchitz continuous for arbitrary $z \in Z$ and the ontology loss function, and there is positive M_β such that $\beta(\cdot, \cdot, \cdot) \le M_\beta$.

In fact, the assumption $\beta_n(z_1, z_2, z_3) = \frac{\beta(z_1, z_2, z_3)}{n}$ equals to the assumption that $\sup_n n\beta_n(z_1, z_2, z_3) < +\infty$ for any $z_1, z_2, z_3 \in Z$. The boundedness hypothesis of $\beta(\cdot, \cdot, \cdot)$ establishes if β is a continuous function combined with the finiteness of Z. In what follows, we denote

$$\Xi(\mathcal{A}_S) = \mathbb{E}_{z \sim \mathcal{D}} l(\mathcal{A}_S, z) - \frac{1}{n} \sum_{i=1}^{n} l(\mathcal{A}_S, z_i),$$

where $\mathbb{E}_{z \sim \mathcal{D}} l(\mathcal{A}_S, z)$ heavily relies on $\mathcal{A}_S$.

III. Main Results and Proofs

In this section, we present our main theoretical analysis and specific proofs. The content of this section is organized as follows: first, we give some preparatory lemmas and conclusions to prepare for the proof of the main theorem; then, in the second subsection, we yield the main conclusions and proofs; in the following two parts, we discuss the local ontology relaxed stability and the generalized bound analysis of the ontology mathematical framework under the conditions of reproducing kernel Hilbert space and the specific ontology execution algorithm using stochastic gradient descent.

A. *Useful lemmas*

Set

$$R = \frac{1}{n} \sum_{i=1}^{n} l(\mathcal{A}_S, z_i),$$

$$R^i = \frac{1}{n} \sum_{j=1}^{n} l(\mathcal{A}_{S^i}, z_j),$$

$$R^{\backslash i,j} = \frac{1}{n} \sum_{k=1}^{n} l(\mathcal{A}_{S^{\backslash i,j}}, z_k).$$

In order to prove our main results in the next subsection, we need the following lemmas.

Lemma 3: Assume that an ontology algorithm $\mathcal{A}$ with locally PO ontology relaxed stability $\beta_n(\cdot,\cdot,\cdot)$ with respect to the ontology loss function l. For arbitrary $\iota > 0$, let

$$M = 2(M_\beta + \sup_{z\in Z}\mathbb{E}_{z_i,z_i'}\beta(z_i,z_i',z) + 2M_l)$$

and

$$\tilde{M} = 2(\sup_{z\in Z}\mathbb{E}_{z_i,z_i'}\beta(z_i,z_i',z) + \iota + 2M_l).$$

There exists a positive constant C' determined by the Lipchitz constant L and the dimension d of z, suppose ε is small enough and n is large enough to satisfy

$$\frac{\iota^2}{32M_\beta^2} - \frac{\log nC'}{n} \ge \frac{\varepsilon}{2\tilde{M}^2}(\frac{4\varepsilon M^2}{\tilde{M}^2} + 4M - \varepsilon), \qquad (1)$$

we have

$$\mathbb{P}(\mathbb{E}_{z\sim\mathcal{D}}[l(\mathcal{A}_S,z)] \ge \frac{1}{n}\sum_{i=1}^n l(\mathcal{A}_S,z_i) + \varepsilon + \frac{2\sup_{z\in Z}\mathbb{E}_{z_i,z_i'}\beta(z_i,z_i',z)}{n}) \le 2e^{-\frac{n\varepsilon^2}{2\tilde{M}^2}}.$$

Lemma 4: Assume that an ontology algorithm $\mathcal{A}$ with locally LTO ontology relaxed stability $\beta_n(\cdot,\cdot,\cdot)$ with respect to the ontology loss function l. For arbitrary $\iota > 0$, let

$$M = 2(M_\beta + \sup_{z\in Z}\mathbb{E}_{z_i,z_j}\beta(z_i,z_j,z) + 2M_l)$$

and

$$\tilde{M} = 2(\sup_{z\in Z}\mathbb{E}_{z_i,z_j}\beta(z_i,z_j,z) + \iota + 2M_l).$$

There exists a positive constant C' determined by the Lipchitz constant L and the dimension d of z, suppose ε is small enough and n is large enough to satisfy

$$\frac{\iota^2}{32M_\beta^2} - \frac{\log nC'}{n} \ge \frac{\varepsilon}{2\tilde{M}^2}(\frac{4\varepsilon M^2}{\tilde{M}^2} + 4M - \varepsilon),$$

we have

$$\mathbb{P}(\mathbb{E}_{z\sim\mathcal{D}}[l(\mathcal{A}_S,z)] \ge \frac{1}{n}\sum_{i=1}^n l(\mathcal{A}_S,z_i) + \varepsilon + \frac{2\sup_{z\in Z}\mathbb{E}_{z_i,z_j}\beta(z_i,z_j,z)}{n}) \le 2e^{-\frac{n\varepsilon^2}{2\tilde{M}^2}}.$$

Here, we only prove the LTO part, and the proof of Lemma 3 can be processed in the similar way.

Proof of Lemma 4. Note that

$$|R - R^{\setminus i,j}| \le \frac{1}{n}\sum_{k\ne i,k\ne j}\frac{\beta(z_i,z_j,z_k)}{n} + \frac{2M_l}{n},$$

$$\varrho(S) = \sum_{i=1}^n l(\mathcal{A}_S,z_i),$$

$$\varrho(S^{\setminus i,j}) = \sum_{k\ne i,k\ne j}^n l(\mathcal{A}_{S^{\setminus i,j}},z_k).$$

Let $\mathcal{F}_m$ be the σ-field obtained from $z_1,\cdots,z_m$, and set

$$\varsigma_m = \mathbb{E}[\varrho(S)|\mathcal{F}_m] - \mathbb{E}[\varrho(S)|\mathcal{F}_{m-2}],$$

$$\begin{aligned}\Phi_{-i,j} \quad = \quad & \{S|\sup_{z_i,z_j}|\sum_{k\ne i,k\ne j}\frac{\beta(z_i,z_j,z_k)}{n} \\ & -\mathbb{E}_{z\sim\mathcal{D}}\beta(z_i,z_j,z)| \le \iota\}.\end{aligned}$$

Let

$$\varsigma_m^1 = \mathbb{E}[\varrho(S)I_{\Phi_{-i,j}}|\mathcal{F}_m] - \mathbb{E}[\varrho(S)I_{\Phi_{-i,j}}|\mathcal{F}_{m-2}],$$

$$\varsigma_m^2 = \mathbb{E}[\varrho(S)I_{\Phi_{-i,j}^c}|\mathcal{F}_m] - \mathbb{E}[\varrho(S)I_{\Phi_{-i,j}^c}|\mathcal{F}_{m-2}].$$

Then, we get

$$\begin{aligned}\mathbb{E}[\exp\{\lambda\sum_{k=1}^n\varsigma_k\}] \quad \le \quad & \frac{1}{2}\mathbb{E}[\exp\{2\lambda\sum_{k=1}^n\varsigma_k^1\}] \\ & +\frac{1}{2}\mathbb{E}[\exp\{2\lambda\sum_{k=1}^n\varsigma_k^2\}].\end{aligned} \qquad (2)$$

For estimating the second part of the right hand of the above inequality, we set

$$\begin{aligned}\Gamma_k^2 \quad = \quad & \inf_x\mathbb{E}[\varrho(S)I_{\Phi_{-i,j}^c}|z_1,\cdots,z_{k-1},z_k=x] \\ & -\mathbb{E}[\varrho(S)I_{\Phi_{-i,j}^c}|z_1,\cdots,z_{k-1}],\end{aligned}$$

$$\begin{aligned}\Lambda_k^2 \quad = \quad & \sup_x\mathbb{E}[\varrho(S)I_{\Phi_{-i,j}^c}|z_1,\cdots,z_{k-1},z_k=x] \\ & -\mathbb{E}[\varrho(S)I_{\Phi_{-i,j}^c}|z_1,\cdots,z_{k-1}].\end{aligned}$$

Clearly, we infer

$$\Gamma_k^2 \le \varsigma_k^2 \le \Lambda_k^2.$$

Furthermore,

$$\begin{aligned}& \Lambda_k^2 - \Gamma_k^2 \\ = \quad & \sup_x\mathbb{E}[\varrho(S)I_{\Phi_{-i,j}^c}|z_1,\cdots,z_{k-1},z_k=x] \\ & -\inf_x\mathbb{E}[\varrho(S)I_{\Phi_{-i,j}^c}|z_1,\cdots,z_{k-1},z_k=x] \\ \le \quad & \sup_{x,y}\mathbb{E}[\varrho(S)I_{\Phi_{-i,j}^c}|z_1,\cdots,z_{k-1},z_k=x] \\ & -\mathbb{E}[\varrho(S)I_{\Phi_{-i,j}^c}|z_1,\cdots,z_{k-1},z_k=y] \\ = \quad & \sup_{x,y}\mathbb{E}[\varrho(S)I_{\Phi_{-i,j}^c}|z_1,\cdots,z_{k-1},z_k=x] \\ & -\mathbb{E}[\varrho(S^{\setminus i,j})I_{\Phi_{-i,j}^c}|z_1,\cdots,z_{k-1},z_k=x] \\ & +\mathbb{E}[\varrho(S^{\setminus i,j})I_{\Phi_{-i,j}^c}|z_1,\cdots,z_{k-1},z_k=x] \\ & -\mathbb{E}[\varrho(S^{\setminus i,j})I_{\Phi_{-i,j}^c}|z_1,\cdots,z_{k-1},z_k=y] \\ & +\mathbb{E}[\varrho(S^{\setminus i,j})I_{\Phi_{-i,j}^c}|z_1,\cdots,z_{k-1},z_k=y] \\ & -\mathbb{E}[\varrho(S)I_{\Phi_{-i,j}^c}|z_1,\cdots,z_{k-1},z_k=y].\end{aligned}$$

In terms of $|\beta(\cdot,\cdot,\cdot)| \le M_\beta$ and $l(\cdot,\cdot) \le M_l$, we get

$$\begin{aligned}& \mathbb{E}[(\varrho(S)-\varrho(S^{\setminus i,j}))I_{\Phi_{-i,j}^c}|z_1,\cdots,z_{k-1},z_k=x] \\ & +\mathbb{E}[(\varrho(S)-\varrho(S^{\setminus i,j}))I_{\Phi_{-i,j}^c}|z_1,\cdots,z_{k-1},z_k=y] \\ \le \quad & 2(M_\beta+M_l)\mathbb{P}(\Phi_{-i,j}^c|z_1,\cdots,z_{k-1}).\end{aligned}$$

In light of

$$\begin{aligned}& \mathbb{E}[\varrho(S^{\setminus i,j})I_{\Phi_{-i,j}^c}|z_1,\cdots,z_{k-1},z_k=x] \\ = \quad & \mathbb{E}[\varrho(S^{\setminus i,j})I_{\Phi_{-i,j}^c}|z_1,\cdots,z_{k-1},z_k=y],\end{aligned}$$

we acquire

$$\Lambda_k^2 - \Gamma_k^2 \le 2(M_\beta+M_l)\mathbb{P}(\Phi_{-i,j}^c|z_1,\cdots,z_{k-1}).$$

To simplify the notation, set

$$M = 2(M_\beta+M_l),$$

$$\Upsilon_k = \mathbb{P}(\Phi_{-i,j}^c|z_1,\cdots,z_{k-1}).$$

Hence,

$$\begin{aligned}&\mathbb{E}[\exp\{2\lambda\sum_{k=1}^{n}\varsigma_k^2\}]\\=&\mathbb{E}[\exp\{2\lambda\sum_{k=1}^{n-2}\varsigma_k^2\}\mathbb{E}[\exp\{2\lambda\varsigma_n^2\}|\mathcal{F}_{n-2}]]\\\le&\mathbb{E}[\exp\{2\lambda\sum_{k=1}^{n-2}\varsigma_k^2\}\exp\{\frac{1}{2}\lambda^2M^2\Upsilon_n^2\}].\end{aligned}$$

Assume that for certain constant ζ_n such that $\sup_{i,j}\Phi_{-i,j}^c\le\zeta_n$, then for any $k=1,\cdots,n$, we have

$$\mathbb{P}(\Upsilon_n\ge c)\le\frac{\zeta_n}{c}.$$

Therefore,

$$\begin{aligned}&\mathbb{E}[\exp\{2\lambda\sum_{k=1}^{n-2}\varsigma_k^2\}\exp\{\frac{1}{2}\lambda^2M^2\Upsilon_n^2\}]\\=&\mathbb{E}[\exp\{2\lambda\sum_{k=1}^{n-2}\varsigma_k^2\}\exp\{\frac{1}{2}\lambda^2M^2\Upsilon_n^2\}I_{\Upsilon_n\ge c}]\\&+\mathbb{E}[\exp\{2\lambda\sum_{k=1}^{n-2}\varsigma_k^2\}\exp\{\frac{1}{2}\lambda^2M^2\Upsilon_n^2\}I_{\Upsilon_n<c}]\\\le&\mathbb{E}[\exp\{2\lambda\sum_{k=1}^{n-2}\varsigma_k^2\}\exp\{\frac{1}{2}\lambda^2M^2\}I_{\Upsilon_n\ge c}]\\&+\mathbb{E}[\exp\{2\lambda\sum_{k=1}^{n-2}\varsigma_k^2\}\exp\{\frac{1}{2}\lambda^2M^2c^2\}I_{\Upsilon_n<c}].\end{aligned}$$

On the other hand, note that

$$\begin{aligned}&\mathbb{E}[\exp\{2\lambda\varsigma_{n-1}^2\}I_{\Upsilon_n\ge c}|\mathcal{F}_{n-2}]\\\le&\exp\{2\lambda M\}\mathbb{E}_{z_{n-1}}[I_{\Upsilon_n\ge c}].\end{aligned}$$

Hence, for given positive λ, and arbitrary $k\in\{1,\cdots,n\}$, we get

$$\begin{aligned}&\mathbb{E}[\exp\{2\lambda\sum_{k=1}^{n-2}\varsigma_k^2\}I_{\Upsilon_n\ge c}]\\\le&\exp\{2M(k-1)\lambda\}\mathbb{P}(\Upsilon_n\ge c)\\\le&\exp\{2M(k-1)\lambda\}\frac{\zeta_n}{c}.\end{aligned}$$

Note that

$$\begin{aligned}&\mathbb{E}[\exp\{2\lambda\sum_{k=1}^{n-2}\varsigma_k^2\}\exp\{\frac{1}{2}\lambda^2M^2\Upsilon_n^2\}]\\\le&\mathbb{E}[\exp\{2\lambda\sum_{k=1}^{n-2}\varsigma_k^2\}\exp\{\frac{1}{2}\lambda^2M^2\Upsilon_n^2\}I_{\Upsilon_n\ge c}]\\&+\mathbb{E}[\exp\{2\lambda\sum_{k=1}^{n-2}\varsigma_k^2\}\exp\{\frac{1}{2}\lambda^2M^2\Upsilon_n^2\}I_{\Upsilon_n<c}]\end{aligned}$$

and

$$\begin{aligned}&I_{\Upsilon_n<c}\\=&I_{\Upsilon_n<c}(I_{\Upsilon_n\ge c}+I_{\Upsilon_n<c})\\=&I_{\Upsilon_n<c}I_{\Upsilon_n\ge c}+I_{\Upsilon_n<c}I_{\Upsilon_n<c}(I_{\Upsilon_n\ge c}+I_{\Upsilon_n<c})\\=&\cdots.\end{aligned}$$

It implies that

$$\begin{aligned}&\mathbb{E}[\exp\{2\lambda\sum_{k=1}^{n-2}\varsigma_k^2\}]\\\le&\exp\{\frac{1}{2}\lambda^2M^2c^2\}+n\frac{\zeta_n}{c}\exp\{\frac{1}{2}\lambda M\max\{1,\lambda M\}\}.\end{aligned}$$

If $\iota>\frac{2M_\beta}{n}$, then

$$\begin{aligned}\cup_{i,j}\Phi_{-i,j}^c\subseteq&\{S|\sup_{z_i,z_j\in Z}|\sum_{k=1}^{n}\frac{\beta(z_i,z_j,z_k)}{n}\\&-\mathbb{E}_{z\sim\mathcal{D}}\beta(z_i,z_j,z)|\ge\frac{\iota}{2}\}\end{aligned}$$

and hence

$$\begin{aligned}\sup_k\mathbb{P}\{\Phi_{-i,j}^c\}\le&\mathbb{P}\{\{S|\sup_{z_i,z_j\in Z}|\sum_{k=1}^{n}\frac{\beta(z_i,z_j,z_k)}{n}\\&-\mathbb{E}_{z\sim\mathcal{D}}\beta(z_i,z_j,z)|\ge\frac{\iota}{2}\}\}.\end{aligned}$$

Using L-Lipschitz assumption, select $\varepsilon=\frac{\iota}{6L}$ and denote Π by the ε-net of Z. Then, the following inequality holds with large possibility:

$$\sup_{z_i,z_j\in\Pi}|\sup_{z_i,z_j\in Z}|\sum_{k=1}^{n}\frac{\beta(z_i,z_j,z_k)}{n}-\mathbb{E}_{z\sim\mathcal{D}}\beta(z_i,z_j,z)|\ge\frac{\iota}{6}.$$

For any $\{z_{i_1},z_{j_1},z_{i_2},z_{j_2}\}$ with $|z_{i_1}-z_{i_2}|\le\frac{\iota}{6L}$ and $|z_{j_1}-z_{j_2}|\le\frac{\iota}{6L}$, we get

$$\begin{aligned}&|\sup_{z_{i_1},z_{j_1}\in Z}|\sum_{k=1}^{n}\frac{\beta(z_{i_1},z_{j_1},z_k)}{n}-\mathbb{E}_{z\sim\mathcal{D}}\beta(z_i,z_j,z)|\\-&|\sup_{z_{i_2},z_{j_2}\in Z}|\sum_{k=1}^{n}\frac{\beta(z_{i_2},z_{j_2},z_k)}{n}-\mathbb{E}_{z\sim\mathcal{D}}\beta(z_i,z_j,z)|\\\le&\frac{\iota}{3}.\end{aligned}$$

Thus,

$$\begin{aligned}&\mathbb{P}\{\{S|\sup_{z_i,z_j\in Z}|\sum_{k=1}^{n}\frac{\beta(z_i,z_j,z_k)}{n}\\&-\mathbb{E}_{z\sim\mathcal{D}}\beta(z_i,z_j,z)|\ge\frac{\iota}{2}\}\}\\=&e^{\tilde{C}d\log(\frac{Ld}{\iota})}e^{-\frac{n\iota^2}{32M_\beta^2}}\end{aligned}$$

where $\tilde{C}$ is a positive constant, d is the dimension of ontology data and L is Lipschitz constant.

It concludes that

$$\sup_k\mathbb{P}\{\Phi_{-i,j}^c\}\le\zeta_n=Ce^{-\frac{n\iota^2}{32M_\beta^2}}$$

for some positive constants C. In fact, we further have

$$\begin{aligned}&\mathbb{E}[\exp\{2\lambda\sum_{k=1}^{n-2}\varsigma_k^2\}]\le\exp\{\frac{1}{2}\lambda^2M^2c^2\}\\+&\frac{nC}{c}e^{-\frac{n\iota^2}{32M_\beta^2}}\exp\{\frac{1}{2}\lambda M\max\{1,\lambda M\}\}.\end{aligned}$$

Next, let's focus on $\mathbb{E}[\exp\{2\lambda\sum_{k=1}^{n-2}\varsigma_k^1\}]$ part. Set

$$\begin{aligned}\Gamma_k^1=&\inf_x\mathbb{E}[\varrho(S)I_{\Phi_{-i,j}}|z_1,\cdots,z_{k-1},z_k=x]\\&-\mathbb{E}[\varrho(S)I_{\Phi_{-i,j}}|z_1,\cdots,z_{k-1}],\end{aligned}$$

$$\begin{aligned}\Lambda_k^1 &= \sup_x \mathbb{E}[\varrho(S)I_{\Phi_{-i,j}}|z_1,\cdots,z_{k-1},z_k=x] \\ &\quad -\mathbb{E}[\varrho(S)I_{\Phi_{-i,j}}|z_1,\cdots,z_{k-1}].\end{aligned}$$

Clearly, we infer

$$\Gamma_k^1 \le \varsigma_k^1 \le \Lambda_k^1.$$

As deduced in the first part, we get

$$\begin{aligned}&\Lambda_k^1-\Gamma_k^1 \\ \le\ &\sup_{x,y}\mathbb{E}[\varrho(S)I_{\Phi_{-i,j}}|z_1,\cdots,z_{k-1},z_k=x] \\ &-\mathbb{E}[\varrho(S^{\setminus i,j})I_{\Phi_{-i,j}}|z_1,\cdots,z_{k-1},z_k=x] \\ &+\mathbb{E}[\varrho(S^{\setminus i,j})I_{\Phi_{-i,j}}|z_1,\cdots,z_{k-1},z_k=x] \\ &-\mathbb{E}[\varrho(S^{\setminus i,j})I_{\Phi_{-i,j}}|z_1,\cdots,z_{k-1},z_k=y] \\ &+\mathbb{E}[\varrho(S^{\setminus i,j})I_{\Phi_{-i,j}}|z_1,\cdots,z_{k-1},z_k=y] \\ &-\mathbb{E}[\varrho(S)I_{\Phi_{-i,j}}|z_1,\cdots,z_{k-1},z_k=y].\end{aligned}$$

In view of $|\beta(\cdot,\cdot,\cdot)|\le M_\beta$ and $l(\cdot,\cdot)\le M_l$, we obtain

$$\begin{aligned}&\mathbb{E}[(\varrho(S)-\varrho(S^{\setminus i,j}))I_{\Phi_{-i,j}}|z_1,\cdots,z_{k-1},z_k=x] \\ &+\mathbb{E}[(\varrho(S)-\varrho(S^{\setminus i,j}))I_{\Phi_{-i,j}}|z_1,\cdots,z_{k-1},z_k=y] \\ \le\ & 2\sup_{z_i,z_j\in Z}\mathbb{E}_z\beta(z_i,z_j,z)+2\iota+2M_l.\end{aligned}$$

In light of

$$\begin{aligned}&\mathbb{E}[\varrho(S^{\setminus i,j})I_{\Phi_{-i,j}}|z_1,\cdots,z_{k-1},z_k=x] \\ =\ &\mathbb{E}[\varrho(S^{\setminus i,j})I_{\Phi_{-i,j}}|z_1,\cdots,z_{k-1},z_k=y],\end{aligned}$$

we get

$$\Lambda_k^2-\Gamma_k^2 \le 2\sup_{z_i,z_j\in Z}\mathbb{E}_z\beta(z_i,z_j,z)+2\iota+2M_l.$$

Let $\tilde{M}=2\sup_{z_i,z_j\in Z}\mathbb{E}_z\beta(z_i,z_j,z)+2\iota+2M_l$. We have

$$\mathbb{E}[\exp\{2\lambda\sum_{k=1}^{n-2}\varsigma_k^1\}] \le \exp\{\frac{1}{2}n\lambda^2\tilde{M}^2\}.$$

In light of (2), we infer

$$\begin{aligned}&\mathbb{E}[\exp\{\lambda\sum_{k=1}^{n}\varsigma_k\}] \\ \le\ &\frac{1}{2}\exp\{\frac{1}{2}n\lambda^2\tilde{M}^2\}+\frac{1}{2}\{\exp\{\frac{1}{2}\lambda^2M^2c^2\} \\ &+\frac{nC\zeta_n}{c}\exp\{2n\lambda M\max\{1,\lambda M\}\}\}.\end{aligned}$$

If we select

$$c=\frac{2\sup_{z_i,z_j\in Z}\mathbb{E}_z\beta(z_i,z_j,z)+2\iota+2M_l}{2M_\beta+2\iota+2M_l},$$

then we have $c<1$. Set

$$\gamma_n=\frac{nC\zeta_n}{c}\exp\{2n\lambda M\max\{1,\lambda M\}.$$

Then, for any positive λ, we yield

$$\mathbb{P}(\varrho(S)-\mathbb{E}[\varrho(S)]\ge n\varepsilon)\le\frac{\exp\{\frac{n}{2}\lambda^2\tilde{M}^2\}+\gamma_n}{\exp\{\lambda n\varepsilon\}}.$$

For small positive ε and large n, take $\lambda=\frac{\varepsilon}{\tilde{M}^2}$, and we want $\gamma_n\le\exp\{\frac{n}{2}\lambda^2\tilde{M}^2\}$, i.e.,

$$\frac{nC\zeta_n}{c}\le\exp\{\frac{n\lambda}{2}(\tilde{M}^2\lambda-4M\max\{M\lambda,1\})\}.$$

Plugging in $\zeta_n=Ce^{-\frac{n\iota^2}{32M_\beta^2}}$ and $\lambda=\frac{\varepsilon}{\tilde{M}^2}$, set $C'=\frac{C}{2c}$ which satisfies

$$C'ne^{-\frac{n\iota^2}{32M_\beta^2}}\le\exp\{(\varepsilon-\frac{4\varepsilon M^2}{\tilde{M}^2}-4M)\frac{n\varepsilon}{2\tilde{M}^2}\}.$$

It can be expressed by

$$\frac{\iota^2}{32M_\beta^2}-\frac{lognC'}{n}\ge(\varepsilon-\frac{4\varepsilon M^2}{\tilde{M}^2}-4M)\frac{n\varepsilon}{2\tilde{M}^2},$$

which leads to

$$\begin{aligned}&\mathbb{P}(\varrho(S)-\mathbb{E}[\varrho(S)]\ge n\varepsilon) \\ \le\ &\frac{\exp\{\frac{n}{2}\lambda^2\tilde{M}^2\}+\gamma_n}{\exp\{\lambda n\varepsilon\}} \\ \le\ &2e^{-\frac{n\varepsilon^2}{2\tilde{M}^2}}.\end{aligned}$$

Let

$$\tilde{l}(\mathcal{A}_S,z)=\mathbb{E}_{z\sim\mathcal{D}}[l(\mathcal{A}_S,z)]-l(\mathcal{A}_S,z),$$

$$\tilde{\varrho}(S)=\frac{1}{n}\sum_{i=1}^{n}\tilde{l}(\mathcal{A}_S,z_i),$$

$$\tilde{\varrho}(S)^{\setminus i,j}=\frac{1}{n}\sum_{k=1}^{n}\tilde{l}(\mathcal{A}_{S^{\setminus i,j}},z_k).$$

We obtain

$$\begin{aligned}&\mathbb{E}[(\tilde{\varrho}(S)-\tilde{\varrho}(S^{\setminus i,j}))I_{\Phi^c_{-i,j}}|z_1,\cdots,z_{k-1},z_k=x] \\ &+\mathbb{E}[(\tilde{\varrho}(S)-\tilde{\varrho}(S^{\setminus i,j}))I_{\Phi^c_{-i,j}}|z_1,\cdots,z_{k-1},z_k=y] \\ \le\ &2(M_\beta+2\sup_{z_i,z_j\in Z}\mathbb{E}_{z\sim\mathcal{D}}\beta(z_i,z_j,z) \\ &+2M_l)\mathbb{P}(\Phi^c_{-i,j}|z_1,\cdots,z_{k-1})\end{aligned}$$

and

$$\begin{aligned}&\mathbb{E}[(\tilde{\varrho}(S)-\tilde{\varrho}(S^{\setminus i,j}))I_{\Phi_{-i,j}}|z_1,\cdots,z_{k-1},z_k=x] \\ &+\mathbb{E}[(\tilde{\varrho}(S)-\tilde{\varrho}(S^{\setminus i,j}))I_{\Phi_{-i,j}}|z_1,\cdots,z_{k-1},z_k=y] \\ \le\ &2(\iota+2\sup_{z_i,z_j\in Z}\mathbb{E}_{z\sim\mathcal{D}}\beta(z_i,z_j,z) \\ &+2M_l)\mathbb{P}(\Phi_{-i,j}|z_1,\cdots,z_{k-1}).\end{aligned}$$

By selecting

$$M=2(M_\beta+2\sup_{z_i,z_j\in Z}\mathbb{E}_{z\sim\mathcal{D}}\beta(z_i,z_j,z)+2M_l)$$

and

$$\tilde{M}=2(\iota+2\sup_{z_i,z_j\in Z}\mathbb{E}_{z\sim\mathcal{D}}\beta(z_i,z_j,z)+2M_l),$$

we finally get

$$\begin{aligned}&\mathbb{E}_S[\mathbb{E}_{z\sim\mathcal{D}}[l(\mathcal{A}_S,z)]-\frac{1}{n}\sum_{i=1}^{n}l(\mathcal{A}_S,z_i)] \\ \le\ &\frac{2\sup_{z_i,z_j\in Z}\mathbb{E}_{z\sim\mathcal{D}}\beta(z_i,z_j,z)}{n}.\end{aligned}$$

Then the desired conclusion follows. □

B. Main conclusion

Theorem 5: Let $\mathcal{A}$ be an ontology algorithm with locally PO ontology relaxed stability $\beta_n(\cdot,\cdot,\cdot)$ with respect to the ontology loss function l. There is a positive constant M_l such that $0 \le l \le M_l$. Suppose that n is a large number. For any given ι and $\delta \in (0,1)$, the following inequality holds with probability at least $1-\delta$:

$$\begin{aligned} & \Xi(\mathcal{A}_S) \\ \le\ & \frac{2\sup_{z_1,z_2\in Z}\mathbb{E}_{z\sim\mathcal{D}}\beta(z_1,z_2,z)}{n} \\ & +2(2\sup_{z\in Z}\mathbb{E}_{z\sim\mathcal{D}}\beta(z_1,z_2,z)+\iota+2M_l)\sqrt{\frac{2\log\frac{2}{\delta}}{n}}. \end{aligned}$$

Theorem 6: Let $\mathcal{A}$ be an ontology algorithm with locally LTO ontology relaxed stability $\beta_n(\cdot,\cdot,\cdot)$ with respect to the ontology loss function l. There is a positive constant M_l such that $0 \le l \le M_l$. Suppose that n is a large number. For any given ι and $\delta \in (0,1)$, the following inequality holds with probability at least $1-\delta$:

$$\begin{aligned} & \Xi(\mathcal{A}_S) \\ \le\ & \frac{2\sup_{z_1,z_2\in Z}\mathbb{E}_{z\sim\mathcal{D}}\beta(z_1,z_2,z)}{n} \\ & +2(2\sup_{z\in Z}\mathbb{E}_{z\sim\mathcal{D}}\beta(z_1,z_2,z)+\iota+2M_l)\sqrt{\frac{2\log\frac{2}{\delta}}{n}}. \end{aligned}$$

Here, we only provide the detailed proof of Theorem 6, and the proof of Theorem 5 can be done by means of the same tricks.

Proof of Theorem 6. Let $\delta = 2e^{-\frac{n\varepsilon^2}{2\tilde{M}^2}}$, and then we infer $\varepsilon = \tilde{M}\sqrt{\frac{2\log\frac{2}{\delta}}{n}}$. Put it into (1) and we deduce

$$\begin{aligned} & \frac{\iota^2}{32M_\beta^2} - \frac{\log nC'}{n} \\ \ge\ & \frac{1}{2\tilde{M}^2}\sqrt{\frac{2\log\frac{2}{\delta}}{n}}(-\tilde{M}\sqrt{\frac{2\log\frac{2}{\delta}}{n}} \\ & +\frac{4M^2}{\tilde{M}}\sqrt{\frac{2\log\frac{2}{\delta}}{n}}+4M). \end{aligned}$$

Note that

$$\frac{\log nC'}{n} \le \frac{\iota^2}{64M_\beta^2},$$

$$\frac{2M^2\log(\frac{2}{\delta})}{n\tilde{M}^2} \le \frac{\iota^2}{128M_\beta^2},$$

$$\frac{2M}{\tilde{M}}\sqrt{\frac{2\log\frac{2}{\delta}}{n}} \le \frac{\iota^2}{128M_\beta^2},$$

and

$$\lim_{n\to\infty}\frac{\log nC'}{n} \to 0.$$

Hence, we yield

$$\frac{\iota^2}{32M_\beta^2} - \frac{\log nC'}{n} \ge \frac{2M^2\log(\frac{2}{\delta})}{n\tilde{M}^2} + \frac{2M}{\tilde{M}}\sqrt{\frac{2\log\frac{2}{\delta}}{n}}.$$

Therefore, the final conclusion is obtained by means of Lemma 4. □

C. Connect to reproducing kernel Hilbert space (RKHS)

In this subsection, we assume that the ontology loss function associates with an ontology cost function c such that $l(f,z) = c(f(x),y)$ for ontology samples in the supervised ontology learning setting. An ontology loss function l is σ-admissible with respect to $\mathcal{Y}^{\mathcal{X}}$ if the associated ontology cost function c is convex with respect to its first argument and for any $y_1, y_2, y_3 \in \mathcal{Y}$, we have

$$|c(y_1,y_3)-c(y_2,y_3)| \le \sigma\|y_1-y_2\|_{\mathcal{Y}},$$

where $\|\cdot\|_{\mathcal{Y}}$ is a certain norm defined on $\mathcal{Y}$.

A reproducing kernel Hilbert space (in short, RKHS) $\mathcal{H}$ is a Hilbert space of continuous linear functions, in which for arbitrary $h \in \mathcal{H}$ and $x \in X$, we have

$$h(x) = < h, K(x,\cdot) >,$$

where K is the kernel in $\mathcal{H}$. Specially, we have for arbitrary $h \in \mathcal{H}$ and $x \in X$,

$$|h(x)| \le \|h\|_K\sqrt{K(x,x)},$$

where $\sqrt{K(x,x)}$ is always denoted by $\kappa(x)$ and $\|\cdot\|_K$ is the norm induced by kernel K in RKHS. Remind that K should be positive semi-defined kernel and $\kappa(x) \ge 0$. The main result in this subsection is stated as follows which associates locally ontology relaxed stability with reproducing kernel Hilbert space.

Theorem 7: Let $\mathcal{H}$ be a reproducing kernel Hilbert space with kernel K, and for arbitrary $x \in X$ have $K(x,x) \le \kappa^2 < \infty$. The ontology loss function l is σ-admissible with respect to $\mathcal{H}$ and the ontology learning algorithm is formulated by

$$\mathcal{A}_S = \mathrm{argmin}_{h\in\mathcal{H}}\frac{1}{n}\sum_{i=1}^{n} l(h,z_i) + \lambda\|h\|_K^2,$$

where $\lambda > 0$ is a balance parameter. Then,
(i) $\mathcal{A}_S$ has locally PO ontology relaxed stability $\beta_n(z_i, z_i^{'}, z)$ such that

$$\beta_n(z_i, z_i^{'}, z) \le \frac{\sigma^2\kappa(x_i)\kappa(x)}{n\lambda}.$$

(ii) $\mathcal{A}_S$ has locally LTO ontology relaxed stability $\beta_n(z_i, z_j, z)$ such that

$$\beta_n(z_i, z_j, z) \le \frac{\sigma^2(\kappa(x_i)+\kappa(x_j))\kappa(x)}{2n\lambda}.$$

Brief Proof of Theorem 7. Let's briefly prove the (ii) part of theorem, and for (i) part, the discussion is similar. Let

$$R(h) = \frac{1}{n}\sum_{i=1}^{n} l(h,z_i) + \lambda\|h\|_K^2,$$

$$R^{\setminus i,j}(h) = \frac{1}{n}\sum_{k\ne i, k\ne j}^{n} l(h,z_k) + \lambda\|h\|_K^2,$$

$$f = \mathrm{argmin}_{h\in\mathcal{H}}R(h),$$

and

$$f^{\setminus i,j} = \mathrm{argmin}_{h\in\mathcal{H}}R^{\setminus i,j}(h).$$

By setting $\Delta f = f^{\setminus i,j} - f$, we get

$$2\|\Delta f\|_K^2 \le \frac{\sigma(|\Delta f(x_i)| + |\Delta f(x_j)|)}{n\lambda}.$$

By means of

$$|f(x_i)| \le \|f\|_K \sqrt{K(x_i, x_i)} \le \|f\|_K \kappa(x_i)$$

and

$$|f(x_j)| \le \|f\|_K \sqrt{K(x_j, x_j)} \le \|f\|_K \kappa(x_j),$$

we deduce

$$\|\Delta f\|_K \le \frac{\sigma(\kappa(x_i) + \kappa(x_j))}{2n\lambda}.$$

Since ontology loss l is σ-admissibility, we infer

$$\begin{aligned} & |l(f, z) - l(f^{\setminus i,j}, z)| \le \sigma|f(x) - f^{\setminus i,j}(x)| \\ = \;& \sigma|\Delta f(x)| \le \sigma\|\Delta f\|_K \kappa(x) \\ \le \;& \frac{\sigma^2(\kappa(x_i) + \kappa(x_j))\kappa(x)}{2n\lambda}. \end{aligned}$$

Hence, we prove the desired theorem. □

D. Connect to stochastic gradient descent

In stochastic gradient descent ontology learning setting, SGD ontology learning algorithm is composed of certain steps of stochastic gradient updates $\widehat{\theta}_{t+1} = \widehat{\theta}_t - \alpha_t \nabla_\theta (\widehat{\theta}_t, z_{it})$, where the learning rate α_t is allowed to change over time, $i_t \in \{1, \cdots, n\}$ is selected uniformly at random at time t. Let T be the total steps in SGD ontology iteration process. For a randomized ontology algorithm $\mathcal{A}$ such as SGD ontology algorithm, we introduce the locally ontology relaxed stability as follows.

Definition 8: (locally PO ontology relaxed stability in randomized ontology algorithm setting) An ontology randomized algorithm $\mathcal{A}$ is $\beta_n(\cdot, \cdot, \cdot)$-locally PO ontology relaxed stability if for any ontology training set $S \in Z^n$ and $z_i^{'} \in Z$, we have

$$|\mathbb{E}_{\mathcal{A}}[l(\mathcal{A}_S, z)] - \mathbb{E}_{\mathcal{A}}[l(\mathcal{A}_{S^i}, z)]| \le \beta_n(z_i, z_i^{'}, z),$$

where the expectation is over the randomness embedded in the ontology algorithm $\mathcal{A}$.

Definition 9: (locally LTO ontology relaxed stability in randomized ontology algorithm setting) An ontology randomized algorithm $\mathcal{A}$ is $\beta_n(\cdot, \cdot, \cdot)$-locally LTO ontology relaxed stability if for any ontology training set $S \in Z^n$, $1 \le i < j \le n$, we have

$$|\mathbb{E}_{\mathcal{A}}[l(\mathcal{A}_S, z)] - \mathbb{E}_{\mathcal{A}}[l(\mathcal{A}_{S^{\setminus i,j}}, z)]| \le \beta_n(z_i, z_j, z),$$

where the expectation is stated as the last definition.

For SGD based ontology algorithm $\mathcal{A}$, outputs functions $\mathcal{A}_S$, $\mathcal{A}_{S^i}$ and $\mathcal{A}_{S^{\setminus i,j}}$ are parameterized by $\widehat{\theta}_T$, $\widehat{\theta}_T^i$ and $\widehat{\theta}_T^{\setminus i,j}$ respectively, and the main conclusions in this subsection study whether there is a function $\beta_n(\cdot, \cdot, \cdot)$ satisfying

$$|\mathbb{E}[l(\widehat{\theta}_T, z)] - \mathbb{E}[l(\widehat{\theta}_T^i, z)]| \le \beta_n(z_i, z_i^{'}, z)$$

or

$$|\mathbb{E}[l(\widehat{\theta}_T, z)] - \mathbb{E}[l(\widehat{\theta}_T^{\setminus i,j}, z)]| \le \beta_n(z_i, z_j, z),$$

where the expectation is taken with respect to randomness coming from uniformly selecting the index in every iteration. Specifically, the main results consider convex ontology setting and non-convex ontology setting respectively.

Theorem 10: (Convex Ontology Optimization) Suppose that the ontology loss function $l(\cdot, z)$ is ϑ-smooth, $L(z)$-Lipschitz (here $L(z)$ is finite for any $z \in Z$: $|l(\theta_1, z) - l(\theta_2, z)| \le L(z)\|\theta_1 - \theta_2\|_2$ holds for any $\theta_1, \theta_2 \in \Theta$), and convex for any $z \in Z$. Moreover, $L = \sup_{z \in Z} L(z) < \infty$ and $\alpha_t \le \frac{2}{\vartheta}$ for $t \le T$. Then, we have

$$|\mathbb{E}[l(\widehat{\theta}_T, z)] - \mathbb{E}[l(\widehat{\theta}_T^i, z)]| \le \frac{2(L + L(z_i))L(z)}{n} \sum_{t=1}^{T} \alpha_t$$

and

$$\begin{aligned} & |\mathbb{E}[l(\widehat{\theta}_T, z)] - \mathbb{E}[l(\widehat{\theta}_T^{\setminus i,j}, z)]| \\ \le \;& \frac{(2L + L(z_i) + L(z_j))L(z)}{n} \sum_{t=1}^{T} \alpha_t. \end{aligned}$$

Theorem 11: (Non-convex Ontology Optimization) Suppose that the ontology loss function $l(\cdot, z)$ satisfies $0 \le l(\cdot, z) \le \Upsilon$ (without loss of generality, we assume that $\Upsilon = 1$), ϑ-smooth, $L(z)$-Lipschitz with finite $L(z)$ for any $z \in Z$, and $L = \sup_{z \in Z} L(z) < \infty$. Suppose that $\alpha_t \le \frac{c}{t}$ is a non-increasing sequence for $t \le T$ and c is a positive constant. Set

$$\psi_\vartheta(n, T, z_i, z_i^{'}, z) = (2c(L(z_i) + L)L(z)T^{\vartheta c})^{\frac{1}{\vartheta c + 1}},$$

$$\psi_\vartheta(n, T, z_i, z_j, z) = (c(L(z_i) + L(z_j) + 2L)L(z)T^{\vartheta c})^{\frac{1}{\vartheta c + 1}}$$

and

$$\mu_n = \frac{1 + \frac{1}{\vartheta c}}{n - 1}.$$

Then, we have

$$|\mathbb{E}[l(\widehat{\theta}_T, z)] - \mathbb{E}[l(\widehat{\theta}_T^i, z)]| \le \mu_n \psi_\vartheta(n, T, z_i, z_i^{'}, z)$$

and

$$|\mathbb{E}[l(\widehat{\theta}_T, z)] - \mathbb{E}[l(\widehat{\theta}_T^{\setminus i,j}, z)]| \le \mu_n \psi_\vartheta(n, T, z_i, z_j, z).$$

The proof of Theorem 10 and Theorem 11 follows from standard statistical learning theory approaches, and we skip the specific proofs here.

IV. Conclusion

In this work, we mainly study the special class of ontology stability which can be regraded as the relaxed version of uniform ontology stability. We provide the formal definition in PO and LTO setting, and the upper error bounds of ontology learning algorithm are derived in terms of locally PO ontology relaxed stability and locally LTO ontology relaxed stability, respectively. Then, we further discuss the setting when the ontology function space is associated with the reproduction of kernel Hilbert space, and it is determined that locally ontology relaxed stability exists if the ontology loss function is σ-admissible. Finally, we introduce the locally PO ontology relaxed stability and locally LTO ontology relaxed stability in randomized ontology learning algorithm, and the generalization bound is deduced when the key iterative part of the algorithm is based on stochastic gradient descent.

The following topics can be used as the content of continued research:

- Although the formal locally ontology relaxed stability is defined in PO and LTO setting is applied to reproducing kernel Hilbert space and randomized ontology learning setting, the specific expressed and statistical error bound in many other ontology learning settings are still open. The locally ontology relaxed stability in other ontology learning framework should be discussed in the future.

• The existence of stability under the framework of regenerative nuclear Hilbert space is proved in our contribution, but the theoretical condition of this result is that the loss function of the ontology satisfies certain conditions. What needs to be studied later is whether the mathematical conditions in these assumptions can be achieved in the actual ontology algorithm design and applications. In other words, there are gaps between theory and practical applications, and it is necessary to study how to bridge these gaps so that the theoretical results obtained in this paper can be truly applied to the field of ontology engineering.

• The multi-dividing ontology learning algorithm has been verified to be well used in the ontology graph learning of tree structure. The optimal ontology function is obtained by means of compared learning tricks. The definition and corresponding properties of locally ontology relaxed stability in multi-dividing ontology learning setting are awaiting for further studies.

REFERENCES

[1] G. Stratogiannis, P. Kouris, G. Alexandridis, et al., "Semantic enrichment of documents: a classification perspective for ontology-based imbalanced semantic descriptions," *Knowledge and Information Systems*, vol. 63, no. 11, pp. 3001-3039, 2021.

[2] R. H. Epstein, Y. K. Jean, R. Dudaryk, et al., "Natural language mapping of electrocardiogram interpretations to a standardized ontology," *Methods of Information in Medicine*, vol. 60, no. 3-4, pp. 104-109, 2021.

[3] M. Gheisari, H. E. Najafabadi, J. A. Alzubi, et al., "OBPP: An ontology-based framework for privacy-preserving in IoT-based smart city," *Future Generation Computer Systems-The International Journal of Escience*, vol. 123, pp. 1-13, 2021.

[4] B. Selvalakshmi, M. Subramaniam, and K. Sathiyasekar, "Semantic conceptual relational similarity based web document clustering for efficient information retrieval using semantic ontology," KSII Transactions on Internet and Information Systems, vol. 15, no. 9, pp. 3102-3119, 2021.

[5] J. Mavracic, C. J. Court, T. Isazawa, et al., "ChemDataExtractor 2.0: Autopopulated ontologies for materials science," *Journal of Chemical Information and Modeling*, vol. 61, no. 9, pp. 4280-4289, 2021.

[6] P. J. S. Goncalves, A. Olivares-Alarcos, J. Bermejo-Alonso, et al., "IEEE standard for autonomous robotics ontology," *IEEE Robotics & Automation Magazine*, vol. 28, no. 3, pp. 171-173, 2021.

[7] A. Maitra, M. R. Kamdar, D. M. Zulman, et al., "Using ethnographic methods to classify the human experience in medicine: a case study of the presence ontology," *Journal of The American Medical Informatics Association*, vol. 28, no. 9, pp. 1900-1909, 2021.

[8] L. L. Zhu, Y. Pan, M. K. Jamil, et al., "Boosting based ontology sparse vector computation approach," *Engineering Letters*, vol. 25, no. 4, pp. 406-415, 2017.

[9] L. L. Zhu, W. G. Tao, X. Z. Min, et al., "Theoretical characteristics of ontology learning algorithm in multi-dividing setting," *IAENG International Journal of Computer Science*, vol. 43, no. 2, pp. 184-191, 2016.

[10] M. H. Lan, J. Xu, and W. Gao, "Ontology feature extraction via vector learning algorithm and applied to similarity measuring and ontology mapping," *IAENG International Journal of Computer Science*, vol. 43, no. 1, pp. 10-19, 2016.

[11] M. H. Lan, J. Xu, and W. Gao, "Ontology similarity computation and ontology mapping using distance matrix learning approach," *IAENG International Journal of Computer Science*, vol. 45, no. 1, pp. 164-176, 2018.

[12] H. Son and S. H. Lee, "Three-dimensional model retrieval in single category geometry using local ontology created by object part segmentation through deep neural network," *Journal of Mechanical Science and Technology*, vol. 35, no. 11, pp. 5071-5079, 2021.

[13] A. Patel, N. C. Debnath, A. K. Mishra, et al., "Covid19-IBO: A covid-19 impact on indian banking ontology along with an efficient schema matching approach," *New Generation Computing*, vol. 39, no. 3-4, pp. 647-676, 2021.

[14] X. S. Xue and J. F. Chen, "Matching biomedical ontologies through Compact Differential Evolution algorithm with compact adaption schemes on control parameters," *Neurocomputing*, vol. 458, pp. 526-534, 2021.

[15] B. Lakzaei and M. Shamsfard, "Ontology learning from relational databases," *Information Sciences*, vol. 577, pp. 280-297, 2021.

[16] T. E. Ratnaike, D. Greene, W. Wei, et al., "MitoPhen database: a human phenotype ontology-based approach to identify mitochondria DNA diseases," *Nucleic Acids Research*, vol. 49, no. 17, pp. 9686-9695, 2021.

[17] H. Rahman and M. I. Hussain, "A light-weight dynamic ontology for Internet of Things using machine learning technique," *ICT Express*, vol. 7, no. 3, pp. 355-360, 2021.

[18] I. Perea-Romero, F. Blanco-Kelly, I Sanchez-Navarro, et al., "NGS and phenotypic ontology-based approaches increase the diagnostic yield in syndromic retinal diseases," *Human Genetics*, vol. 140, no. 12, pp. 1665-1678, 2021.

[19] Q. W. Bao, G. Zhao, Y. Yu, et al., "The ontology-based modeling and evolution of digital twin for assembly workshop," *International Journal of Advanced Manufacturing Technology*, vol. 117, no. 1-2, pp. 395-411, 2021.

[20] R. M. Chen, Y. Lu, P. Witherell, et al., "Ontology-driven learning of bayesian network for causal inference and quality assurance in additive manufacturing," *IEEE Robotics and Automation Letters*, vol. 6, no. 3, pp. 6032-6038, 2021.

[21] S. Belabbes, S. Benferhat, J. Chomicki, "Handling inconsistency in partially preordered ontologies: the Elect method," *Journal of Logic and Computation*, vol. 31, no. 5 pp. 1356-1388, 2021.

[22] J. Wu, X. Yu, and W. Gao, "Disequilibrium multi dividing ontology learning algorithm," *Communications in Statistics-Theory and Methods*, vol. 46, no. 18, pp. 8925-8942, 2017.

[23] W. Gao and M. R. Farahani, "Generalization bounds and uniform bounds for multi-dividing ontology algorithms with convex ontology loss function," *The Computer Journal*, 2017, vol. 60, no. 9, pp. 1289-1299, 2017.

[24] W. Gao, J. L. G. Guirao, B. Basavanagoud, et al., "Partial multi-dividing ontology learning algorithm," *Information Sciences*, vol. 467, pp. 35-58, 2018.

[25] W. Gao, Y. Chen, A. Q. Baig, et al., "Ontology geometry distance computation using deep learning technology," *Journal of Intelligent & Fuzzy Systems*, 2018, vol. 35, no. 4, pp. 4517-4524, 2018.

[26] W. Gao, Y. Q. Zhang, J. L. G. Guirao, et al., "A discrete dynamics approach to sparse calculation and applied in ontology science," *Journal of Difference Equations and Applications*, 2019, vol. 25, no. 9-10, pp. 1239-1254, 2019.

[27] J. Z. Wu, X. Yu, L. L. Zhu, et al., "Leave-two-out stability of ontology learning algorithm," *Chaos, Solitons and Fractals*, vol. 89, pp. 322-327, 2016.

Shu Gong is an associate professor in Department of Computer Science, Guangdong University Science and Technology, City of Dongguan, China. She got bachelor degree on computer science from Nanchang University in 2006. Then, she enrolled in Department of Computer Science and Information Technology, Yunnan Normal University, and got Master degree there in 2009. She is currently a PhD student in School of Information and Control Engineering, China University of Mining and Technology, City of Xuzhou, China. During the years, as a researcher in computer science, her interests are covered two areas: Information Retrieval and Artificial Intelligence.

Xinxin Huang is an associate professor in Department of Computer Science, Guangdong University Science and Technology, City of Dongguan, China.

Caihua Qiu is an associate professor in Department of Computer Science, Guangdong University Science and Technology, City of Dongguan, China.

Wei Gao was born in the city of Shaoxing, Zhejiang Province, China on Feb. 13, 1981. He got two bachelor degrees on computer science from Zhejiang industrial university in 2004 and mathematics education from College of Zhejiang education in 2006. Then, he was enrolled in department of computer science and information technology, Yunnan normal university, and got Master degree there in 2009. In 2012, he got PhD degree in department of Mathematics, Soochow University, China.

Now, he acts as professor in the School of Information Science and Technology, Yunnan Normal University. As a researcher in computer science and mathematics, his interests are covering two disciplines: Graph theory, Statistical learning theory, Information retrieval, and Artificial Intelligence.