IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received October 25, 2020, accepted November 25, 2020, date of publication December 2, 2020,
date of current version December 18, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3041659

Theoretical Perspective of Multi-Dividing Ontology Learning Trick in Two-Sample Setting

LINLI ZHU[1,2] **AND GANG HUA**[1]
[1]School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116, China
[2]School of Computer Engineering, Jiangsu University of Technology, Changzhou 213001, China

Corresponding authors: Linli Zhu (zhulinli@jsut.edu.cn) and Gang Hua (ghua@cumt.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 51574232, in part by the Industry-Academia Cooperation Innovation Fund Project of Jiangsu Province under Grant BY2016030-06, and in part by the Modern Education Technology Research Project in Jiangsu Province under Grant 2019-R-75637.

ABSTRACT The multi-dividing ontology learning framework has been proven to have a higher efficiency for tree-structured ontology learning, and in this work, we consider a special setting of this learning framework in which ontology sample set for each rate is divided into two groups. This setting can be regarded as the classic two-sample learning problem associated with multi-dividing ontology framework. In this work, we mainly focus on the theoretical analysis of multi-dividing two-sample ontology learning algorithm, whose ontology objective function is proposed, and the generalization bounds in this setting is obtained in terms of U-statistics technique. The theoretical result given is of potential guiding significance in the field of ontology engineering applications.

INDEX TERMS Small ontology, multi-dividing ontology learning, similarity measure.

I. INTRODUCE OF ONTOLOGY

The concept of ontology originally belongs to the category of western philosophy, which refers to the expression and summary of the objective existence at the logical level. Ontology began to be introduced in the artificial intelligence in the 1980s, with 20 years of development, and it has been widely recognized in this century which was defined by a clear formal specification of a shared conceptual model. Ontology can specifically describe the complex conceptual relationships in a certain field. Human-machine can communicate and share data information between machines because the definition of domain information by ontology is unanimously recognized. In addition to the study of ontology in the field of philosophy, computers and other aspects have also made active research on ontology theory and applied it to corresponding fields (see Gao *et al.* [1] and [2]).

Artificial intelligence has its own definition of ontology, for example, the standard stipulate the concept of ontology can be described as follows: ''Ontology can define concepts in a specific field and give clear terms, describe the relationship between them in detail, and vocabulary extension rules can also describe and express based on defined terms.'' The main research problem of ontology in the field of engineering is how to construct ontology, which involves the use of ontology principles. Although the definitions of ontologies given by experts in various fields are different from all angles, researchers agree that ontologies can clearly define the information concepts in the field, and the use of ontologies in specific fields can make each subject accessible [3]. This is the essential connotation of the ontology concept. Ontology is defined in the field of library and information science as ''Ontology can use a specific domain vocabulary to describe a specific fact and infer the deep meaning of the vocabulary'', and it's also defined to be able to represent specific domain conceptual information from a specific perspective [4].

Ontology has been studied and applied in various engineering applications. Skalle and Aamodt [5] showed how tricks of knowledge modeling and drilling ontology have been employed to predict downhole failures during drilling. Sobral *et al.* [6] proposed an ontology-based modelling to support integration and visualization of data from ITS. Al-Sayed *et al.* [7] presented a comprehensive cloud ontology named CloudFNF. Tebes *et al.* [8] evaluated identify and synthesize the available primary studies on conceptualized software testing ontologies. Pradeep and Sundar [9] suggested retrieving the information with the design of QAOC architecture. Hema and Kuppusamy [10] raised a trust-based privacy preservation modelling for service handling by means of ontology service ranking. Messaoudi *et al.* [11] presented a review of medical ontologies. Mantovani *et al.* [12]

The associate editor coordinating the review of this manuscript and approving it for publication was Fabrizio Marozzo.

introduced an ontology based trick for the interpretation and the encoding of the map data. Abeysinghe *et al.* [13] developed a SSIF by leveraging a novel term-algebra on top of a sequence-based representation of gene ontology concepts. Kossmann *et al.* [14] presented an ontology based federative trick to managing the inherent complexity of CM in the context of SoS.

In this article, we don't consider the philosophical category of ontology, but regard it as a structured conceptual model. The so-called structured means the data in the ontology is not a single record, but the data are mutually related, and a graph can be used to represent the data structure. Vertices are used to represent concepts, and edges between vertices represent direct connections or relationship between concepts. Hence, the entire ontology is represented by a graph. In addition, after all the information related to the concept is numerically expressed, a multi-dimensional vector is used to encapsulate the representation, that is, each vertex is a fixed p-dimensional vector, and then a learning model can be used to learn various ontology graphs [15].

II. SETTING OF ONTOLOGY ALGORITHM

A. ONTOLOGY LEARNING ALGORITHM

Set $G = (V, E)$ as an ontology graph whose vertex set corresponds to concepts and edge set reveals the set of directly relationship between two concepts. Suppose $Sim : V^2 \to \mathbb{R}^+ \cup \{0\}$ as the similarity function and it always unitizes the value to interval [0,1] for convenience. Let $v_1, v_2 \in V(G)$ be two differ vertices, $Sim(v_1, v_2) = 1$ indicates the same meaning of concepts corresponding to v_1 and v_2. Conversely, $Sim(v_1, v_2) = 0$ reveals no relationship between two concepts corresponding to v_1 and v_2. Threshold parameter $M \in [0, 1]$ is determined in light of field experts, then for the given vertex v, $\{v'|Sim(v, v') \geq M\}$ is returned to the user as similarity vertices. In this whole article, suppose n is the sample capacity, i.e., the number of ontology samples.

Let $S = \{v_i\}_{i=1}^n$ be the ontology sample set with n ontology vertices which is independent identically distributed according to an unknown distribution $\mathcal{D}$ (written as $v_i \sim \mathcal{D}$ for $i \in \{1, \cdots, n\}$), $f : V \to \mathbb{R}$ be an ontology function which maps each ontology vertex to a real number (in this setting, the similarity between ontology vertices v_1 and v_2 is determined in terms of $|f(v_1) - f(v_2)|$ in which we desire a small number for a high similar pair (v_1, v_2), and on the contrary a large number for dissimilar pair) and $l(f, v)$ be the ontology loss function. The expected risk of ontology learning model can be formulated by

$$R(f) = \mathbb{E}_{v\sim\mathcal{D}} l(f, v).$$

Unfortunately, we can't directly calculate $R(f)$ since $\mathcal{D}$ is unknown. Instead, ontology empirical framework is applied in the specific ontology learning process which is denoted as follows

$$\widehat{R}_S(f) = \frac{1}{n}\sum_{i=1}^{n} l(f, v_i).$$

For the supervised ontology learning, assume that ontology samples are denoted by (v_i, y_i) where $y_i \in \mathcal{Y}$ is the label of v_i. For given $f : V \to \mathbb{R}$, ontology loss $l(f, v_i, y_i)$, and hence the expected ontology risk is denoted as

$$R(f) = \int_{V\times\mathcal{Y}} l(f, v_i, y_i)\mathcal{D}(dv_i, dy_i).$$

The corresponding empirical ontology risk with $S = \{(v_i, y_i)\}_{i=1}^n$ can be modified as

$$\widehat{R}(f) = \frac{1}{n}\sum_{i=1}^{n} l(f, v_i, y_i).$$

B. MULTI-DIVIDING ONTOLOGY ALGORITHM BY MAXIMIZING AUC MEASURE

Multi-dividing ontology learning framework has attracted the attention of scholars in the recent decade since it fits the ontology graph with tree structure. In this special ontology learning setting, the ontology vertex set is divided into k rates corresponding to k branches under the top vertex. The values of different branches are always obtained from domain experts. For ontology function f, we expect $f(v^a) > f(v^b)$ where v^a and v^b belong to branches a and b respectively, and a, b are positive integers with $1 \leq a < b \leq k$.

Specifically, the learner is inferred to a set of ontology sample $S = (S_1, S_2, \cdots, S_k) \in V^{n_1} \times V^{n_2} \times \cdots \times V^{n_k}$ where $S_a = (v_1^a, \cdots, v_{n_a}^a) \in V^{n_a}$ $(1 \leq a \leq k)$. Ontology function $f : V \to \mathbb{R}$ is learned in terms of S which assigns the S_a vertices larger value than S_b vertices, where $a < b$. Let $\mathcal{D}_a$ be the conditional distributions for $a \in \{1, \cdots, k\}$ and the sample capacity is denoted by $n = \sum_{i=1}^k n_i$ with $n_i = |S_i|$.

The expected multi-dividing ontology expected risk with the ontology function $f : V \to \mathbb{R}$ is formulated by

$$R(f) = \sum_{a=1}^{k-1}\sum_{b=a+1}^{k} E_{v\sim\mathcal{D}_a, v'\sim\mathcal{D}_b}\{l(f, v, v'))\}. \quad (1)$$

The transformation expression for expected ontology risk can be denoted as

$$R(f) = \sum_{a=1}^{k-1}\sum_{b=a+1}^{k}\int_{V^a\times V^b} l(f, v^a, v^b)\mathcal{D}_a(dv^a)\mathcal{D}_b(dv^b). \quad (2)$$

The first $R(f)$ expression is in expectation form, while the second $R(f)$ expression is displayed in integral form, and the two are equivalent. Furthermore, the multi-dividing empirical error is formulated as

$$\widehat{R}_{S,l}(f) = \sum_{a=1}^{k-1}\sum_{b=a+1}^{k}\frac{1}{n_a n_b}\sum_{i=1}^{n_a}\sum_{j=1}^{n_b} l(f, v_i^a, v_j^b). \quad (3)$$

Hence, the desired ontology function is deduced from ontology learning framework $f* = argmin\widehat{R}_{\mathcal{T},l}(f)$ in which $\widehat{R}_{S,l}(f)$ can be simply written as $\widehat{R}(f)$.

From another angle, the idea of multi-dividing ontology learning modelling can be explained in terms of maximizing AUC (Area Under the ROC (Receiver Operating Characteristic) Curve) criterion. Consider $k = 2$ or imagine it as a binary classification problem, ROC curve is relied on a series of different two classification methods (cut-off value or decision threshold), with true positive rate (sensitivity) as the vertical coordinate, false positive rate is the curve drawn on the abscissa. AUC is introduced by the area under the ROC curve and the coordinate axis. Clearly, the value of this area will not be larger than 1. Because the ROC curve is generally above the line $y = x$ in, the value range of AUC is between 1/2 and 1. The closer the AUC is to 1, the higher the authenticity of the detection trick; when it is equal to 1/2, the authenticity is the lowest and it has no application value. In multi-dividing ontology setting, there are k classes corresponding to k rate, and we consider the pairwise comparison. Hence, in this case, the AUC criterion in multi-dividing setting can be formulated by the accumulation of each pair of (a, b) with $1 \le a < b \le k$.

Specifically, let $H_{f,a}(t) = \mathbb{P}\{f(v) \le t|Y = a\}$ and $\widehat{H}_{f,a}(t) = \frac{1}{n_a}\sum_{i=1}^{n_a} \mathrm{I}(f(v_i^a) \le t)$ for $a \in \{1, \cdots, k\}$, $t \in \mathbb{R}$, and $\mathrm{I}(\cdot)$ is a binary function such that its value takes 1 if argument is true and 0 otherwise. For convenience, in what follows, we write $H_{f,a}$ and $\widehat{H}_{f,a}$ instead of $H_{f,a}(t)$ and $\widehat{H}_{f,a}(t)$ in the AUC expression. The expected AUC framework in multi-dividing setting is associated with $H_{f,a}$ which is denoted by

$$AUC_{H_{f,a},H_{f,b}}(f) = \sum_{a=1}^{k-1}\sum_{b=a+1}^{k}\{\mathbb{P}\{f(v^a) > f(v^b)\} + \frac{1}{2}\mathbb{P}\{f(v^a) = f(v^b)\}\}, \quad (4)$$

and the corresponding empirical multi-dividing ontology framework under AUC certerion is

$$\widehat{AUC}_{H_{f,a},H_{f,b}}(S,f) = \sum_{a=1}^{k-1}\sum_{b=a+1}^{k}\frac{1}{n_a n_b}\sum_{i=1}^{n_a}\sum_{j=1}^{n_b}\{\mathrm{I}(f(v_i^a) > f(v_j^b)) + \frac{1}{2}\mathrm{I}(f(v_i^a) = f(v_j^b))\}. \quad (5)$$

If we limit it to a specific pair (a, b), the AUC criterion can be expressed as

$$AUC^{a,b}_{H_{f,a},H_{f,b}}(f) = \{\mathbb{P}\{f(v^a) > f(v^b)\} + \frac{1}{2}\mathbb{P}\{f(v^a) = f(v^b)\}\}, \quad (6)$$

$$\widehat{AUC}^{a,b}_{H_{f,a},H_{f,b}}(S,f) = \frac{1}{n_a n_b}\sum_{i=1}^{n_a}\sum_{j=1}^{n_b}\{\mathrm{I}(f(v_i^a) > f(v_j^b)) + \frac{1}{2}\mathrm{I}(f(v_i^a) = f(v_j^b))\}. \quad (7)$$

Hence, we admit

$$AUC_{H_{f,a},H_{f,b}}(f) = \sum_{a=1}^{k-1}\sum_{b=a+1}^{k} AUC^{a,b}_{H_{f,a},H_{f,b}}(f), \quad (8)$$

$$\widehat{AUC}_{H_{f,a},H_{f,b}}(S,f) = \sum_{a=1}^{k-1}\sum_{b=a+1}^{k} \widehat{AUC}^{a,b}_{H_{f,a},H_{f,b}}(S,f). \quad (9)$$

The aim of this article is to propose the statistical characterization of multi-dividing ontology learning algorithm in the two-sample setting, where each S_a with $a \in \{1, \cdots, k\}$ is divided into two groups S_a^0 and S_a^1. In the classic two-sample learning problem, the first data set is used to obtain an ontology function on its function space and the second ontology data set is served to compute a pseudo-two-sample test statistic from the given ontology data. In our multi-dividing ontology setting, we can consider the two-sample ontology data as the similar meaning, for example, the first group of ontology data is applied to training, and the second group of ontology data is for testing, ect. More contents on two-sample learning problem in different setting and applications can be referred to Ma and Wong [16], Tang *et al.* [17], Chen *et al.* [18], Kim *et al.* [19], Rabin *et al.* [20], and Emura and Hsu [21].

The two-sample ontology setting can be explained from another angle. We require ontology learning algorithms to have generalization capabilities, i.e., ontology functions deduced from one ontology sample set can be well applied to other ontology data sets in the same type. In other words, for the same type of ontology data, the ontology functions obtained from different ontology sample sets should have similar characteristics, and should not be very different. In statistical learning theory, it can be understood that two ontology functions obtained from different ontology samples of the same type of ontology data are very close in the ontology function space and have similar statistical characteristics to each other.

The main result regarding to generalization bounds in this setting is manifested in next section, the proof of this result relies on the techniques of U-statistics and its applications which can be found in Fuchs *et al.* [22], Bouzebda and Nemouchi [23], Fuglsby *et al.* [24], Privault and Serafin [25], Bachmann and Reitzner [26], and Garg and Dewan [27], and we skip the details here.

III. THEORETICAL ANALYSIS IN TWO-SAMPLE SETTING

For $a \in \{1, \cdots, k\}$, we denote $z \in \{0, 1\}$ as the natation of two groups, i.e., ontology sub sample set S_a in new multi-dividing ontology setting is divided into two sets S_a^z : S_a^0 and S_a^1 (denote $n_a^0 = |S_a^0|$ and $n_a^1 = |S_a^1|$). Let $H^z_{f,a}(t) = \mathbb{P}\{f(v) \le t|Y = a, Z = z\}$ and $\widehat{H}^z_{f,a}(t) = \frac{1}{n_a^z}\sum_{i=1}^{n_a} \mathrm{I}(f(v_i^a) \le t, Z_i = z)$ for $a \in \{1, \cdots, k\}$, $t \in \mathbb{R}$. To simplify the symbol, we use $H^z_{f,a}$ and $\widehat{H}^z_{f,a}$ to replace $H^z_{f,a}(t)$ and $\widehat{H}^z_{f,a}(t)$. Hence, the AUC framework in multi-dividing ontology setting is reformulated by

$$AUC_{H^0_{f,a},H^0_{f,b}}(f) = \sum_{a=1}^{k-1}\sum_{b=a+1}^{k}\{\mathbb{P}\{f(v^a) > f(v^b)|z = 0\} + \frac{1}{2}\mathbb{P}\{f(v^a) = f(v^b)|z = 0\}\},$$

$$AUC_{H^1_{f,a},H^1_{f,b}}(f) = \sum_{a=1}^{k-1}\sum_{b=a+1}^{k}\{\mathbb{P}\{f(v^a) > f(v^b)|z = 1\}$$

$$+\frac{1}{2}\mathbb{P}\{f(v^a)=f(v^b)|z=1\}\},$$

$$\widehat{AUC}_{H^0_{f,a},H^0_{f,b}}(S,f)=\sum_{a=1}^{k-1}\sum_{b=a+1}^{k}\frac{1}{n^0_a n^0_b}\sum_{i=1}^{n_a}\sum_{j=1}^{n_b}\{\mathrm{I}(f(v^a_i)>f(v^b_j),z=0)+\frac{1}{2}\mathrm{I}(f(v^a_i)=f(v^b_j),z=0)\},$$

$$\widehat{AUC}_{H^1_{f,a},H^1_{f,b}}(S,f)=\sum_{a=1}^{k-1}\sum_{b=a+1}^{k}\frac{1}{n^1_a n^1_b}\sum_{i=1}^{n_a}\sum_{j=1}^{n_b}\{\mathrm{I}(f(v^a_i)>f(v^b_j),z=1)+\frac{1}{2}\mathrm{I}(f(v^a_i)=f(v^b_j),z=1)\}.$$

Similarly, if we restrict on a specific combination of (a, b) with $a, b \in \{1, \cdots, k\}$ and $a < b$, then

$$\mathrm{AUC}^{a,b}_{H^0_{f,a},H^0_{f,b}}(f)=\mathbb{P}\{f(v^a)>f(v^b)|z=0\}+\frac{1}{2}\mathbb{P}\{f(v^a)=f(v^b)|z=0\},$$

$$\mathrm{AUC}^{a,b}_{H^1_{f,a},H^1_{f,b}}(f)=\mathbb{P}\{f(v^a)>f(v^b)|z=1\}+\frac{1}{2}\mathbb{P}\{f(v^a)=f(v^b)|z=1\},$$

$$\widehat{\mathrm{AUC}}^{a,b}_{H^0_{f,a},H^0_{f,b}}(S,f)=\frac{1}{n^0_a n^0_b}\sum_{i=1}^{n_a}\sum_{j=1}^{n_b}\{\mathrm{I}(f(v^a_i)>f(v^b_j),z=0)+\frac{1}{2}\mathrm{I}(f(v^a_i)=f(v^b_j),z=0)\},$$

$$\widehat{\mathrm{AUC}}^{a,b}_{H^1_{f,a},H^1_{f,b}}(S,f)=\frac{1}{n^1_a n^1_b}\sum_{i=1}^{n_a}\sum_{j=1}^{n_b}\{\mathrm{I}(f(v^a_i)>f(v^b_j),z=1)+\frac{1}{2}\mathrm{I}(f(v^a_i)=f(v^b_j),z=1)\}.$$

Thus, we have

$$\mathrm{AUC}_{H^0_{f,a},H^0_{f,b}}(f)=\sum_{a=1}^{k-1}\sum_{b=a+1}^{k}\mathrm{AUC}^{a,b}_{H^0_{f,a},H^0_{f,b}}(f),$$

$$\mathrm{AUC}_{H^1_{f,a},H^1_{f,b}}(f)=\sum_{a=1}^{k-1}\sum_{b=a+1}^{k}\mathrm{AUC}^{a,b}_{H^1_{f,a},H^1_{f,b}}(f),$$

$$\widehat{\mathrm{AUC}}_{H^0_{f,a},H^0_{f,b}}(S,f)=\sum_{a=1}^{k-1}\sum_{b=a+1}^{k}\widehat{\mathrm{AUC}}^{a,b}_{H^0_{f,a},H^0_{f,b}}(S,f),$$

$$\widehat{\mathrm{AUC}}_{H^1_{f,a},H^1_{f,b}}(S,f)=\sum_{a=1}^{k-1}\sum_{b=a+1}^{k}\widehat{\mathrm{AUC}}^{a,b}_{H^1_{f,a},H^1_{f,b}}(S,f).$$

In two-sample ontology setting, the optimal ontology function is denoted by f^*_λ which is obtained by maximizing the following ontology objective function:

$$R_\lambda(f)=\mathrm{AUC}_{H_{f,a},H_{f,b}}(f)-\lambda|\mathrm{AUC}_{H^0_{f,a},H^0_{f,b}}(f)-\mathrm{AUC}_{H^1_{f,a},H^1_{f,b}}(f)|,$$

where $\lambda > 0$ is an offset variable. The corresponding empirical ontology version with positive offset parameter λ is

$$\widehat{R}_\lambda(f,S)=\widehat{\mathrm{AUC}}_{H_{f,a},H_{f,b}}(S,f)-\lambda|\widehat{\mathrm{AUC}}_{H^0_{f,a},H^0_{f,b}}(S,f)-\widehat{\mathrm{AUC}}_{H^1_{f,a},H^1_{f,b}}(S,f)|,$$

and we express its maximizer as $\widehat{f}_{\lambda,S}$.

We say an ontology function space $\mathcal{F}$ with ontology functions $f : V \to \mathbb{R}$ is VC-major if the major sets of the elements in $\mathcal{F}$ is a VC-class of sets in V. Specifically, $\mathcal{F}$ is a VC-major class if and only if $\{\{v \in V|\, f(v) > t\}|t \in \mathbb{R}, f \in \mathcal{F}\}$ is a VC-class of sets.

Our main conclusion reveals the learning rate of multi-dividing ontology problem in two-sample setting.

Theorem 1: Let $\mathcal{F}$ be a VC-major ontology function space, and Γ be its VC-dimension which is a finite number. Set $n_{\min} = \min\{n_1, \cdots, n_k\}$. Suppose there is a positive number ε satisfying

$$\min_{y\in\{1,\cdots,k\},z\in\{0,1\}}\mathbb{P}\{Y=y,Z=z\}\geq\varepsilon.$$

Then, with probability at least $1-\delta$ for any $\delta > 0$, we have

$$\varepsilon^2(R_\lambda(f^*_\lambda)-R_\lambda(\widehat{f}_{\lambda,S}))\leq C\left(\binom{k}{2}\sqrt{\frac{\Gamma}{2n_{\min}}}(4\lambda+\frac{1}{2})+\binom{k}{2}\sqrt{\frac{\log(\frac{13}{\delta})}{2n_{\min}-1}}(4\lambda+(4\lambda+2)\varepsilon)+O(\frac{\binom{k}{2}}{2n_{\min}}\right).$$

Proof of Theorem 1. Note that $R_\lambda(f^*_\lambda) - R_\lambda(\widehat{f}_{\lambda,S}) \leq 2\sup_{f\in\mathcal{F}}|\widehat{R}_\lambda(f,S)-R_\lambda(f)|$. Set

$$\widehat{\Psi}=\sup_{f\in\mathcal{F}}|\widehat{AUC}_{H_{f,a},H_{f,b}}(f,S)-\mathrm{AUC}_{H_{f,a},H_{f,b}}(f)|,$$

$$\widehat{\Psi}_0=\sup_{f\in\mathcal{F}}|\widehat{AUC}_{H^0_{f,a},H^0_{f,b}}(f,S)-\mathrm{AUC}_{H^0_{f,a},H^0_{f,b}}(f)|,$$

and

$$\widehat{\Psi}_1=\sup_{f\in\mathcal{F}}|\widehat{AUC}_{H^1_{f,a},H^1_{f,b}}(f,S)-\mathrm{AUC}_{H^1_{f,a},H^1_{f,b}}(f)|.$$

In light of triangular inequality, we acquire $\sup_{f\in\mathcal{F}}|\widehat{R}_\lambda(f,S)-R_\lambda(f)|\leq\widehat{\Psi}+\lambda(\widehat{\Psi}_0+\widehat{\Psi}_1)$.

Set

$$\begin{aligned}p^{a,b}&=\mathbb{P}\{Y=a|Y\in\{a,b\}\},\\p^{a,b}_0&=\mathbb{P}\{Y=a|Z=0,Y\in\{a,b\}\},\\p^{a,b}_1&=\mathbb{P}\{Y=a|Z=1,Y\in\{a,b\}\},\\q^{a,b}_0&=\mathbb{P}\{Z=0|Y\in\{a,b\}\},\\q^{a,b}_1&=\mathbb{P}\{Z=1|Y\in\{a,b\}\},\end{aligned}$$

for $a, b \in \{1, \cdots, k\}$ with $a < b$ and $Z \in \{0, 1\}$. Thus, we obtain $q^{a,b}_0 = 1 - q^{a,b}_1$.

Denote

$$\widehat{U}_{a,b}(f) = \frac{2}{(n_a+n_b)(n_a+n_b-1)}\sum_{i=1}^{n_a}\sum_{j=1}^{n_b}\{\mathrm{I}(f(v_i^a) > f(v_j^b)) + \frac{1}{2}\mathrm{I}(f(v_i^a) = f(v_j^b))\},$$

and by means of U-statistic tricks, we get

$$\widehat{AUC}_{H_{f,a},H_{f,b}}(f,S) = \sum_{a=1}^{k-1}\sum_{b=a+1}^{k}\frac{(n_a+n_b)(n_a+n_b-1)}{2n_an_b}\widehat{U}_{a,b}(f).$$

Furthermore, by means of characteristics of U-statistics, we infer $U_{a,b}(f) = \mathbb{E}[\widehat{U}_{a,b}(f)] = 2(1-p^{a,b})p^{a,b}AUC^{a,b}_{H_{f,a},H_{f,b}}(f)$ and

$$\sup_{f\in\mathcal{F}}|\widehat{U}_{a,b}(f) - U_{a,b}(f)| \le 2C\sqrt{\frac{\Gamma}{n_a+n_b}} + 2\sqrt{\frac{\log\frac{1}{\delta}}{n_a+n_b-1}} \tag{10}$$

holds with possibility at least $1-\delta$ where C is an universal constant. The details on the proof and related statement (10) can be found in Bousquet *et al.* [28] and Clémencon *et al.* [29]. Using $\sup_{f\in\mathcal{F}}|\widehat{U}_{a,b}(f)| \le \frac{2n_an_b}{(n_a+n_b)(n_a+n_b-1)}$, we deduce

$$\begin{aligned}\widehat{\Psi} &\le \sup_{f\in\mathcal{F}}\sum_{a=1}^{k-1}\sum_{b=a+1}^{k}\{|(\frac{(n_a+n_b)(n_a+n_b-1)}{2n_an_b} - \frac{1}{2(1-p^{a,b})p^{a,b}})\widehat{U}_{a,b}(f)| \\ &\quad + \frac{1}{2(1-p^{a,b})p^{a,b}}|\widehat{U}_{a,b}(f) - U_{a,b}(f)|\} \\ &\le \sum_{a=1}^{k-1}\sum_{b=a+1}^{k}\{\frac{1}{(1-p^{a,b})p^{a,b}}|\frac{n_a+n_b}{(n_a+n_b)^2} - p^{a,b}(1-p^{a,b}) + \frac{n_an_b}{(n_a+n_b)^2(n_a+n_b-1)}| \\ &\quad + \frac{1}{2(1-p^{a,b})p^{a,b}}\sup_{f\in\mathcal{F}}|\widehat{U}_{a,b}(f) - U_{a,b}(f)|\}.\end{aligned}$$

It follows from Hoeffding inequality that

$$\sum_{a=1}^{k-1}\sum_{b=a+1}^{k}|\frac{n_a}{n_a+n_b} - p^{a,b}| \le \sum_{a=1}^{k-1}\sum_{b=a+1}^{k}\sqrt{\frac{\log\frac{2}{\delta}}{2(n_a+n_b)}}. \tag{11}$$

Note that $\frac{n_a+n_b}{(n_a+n_b)^2} - p^{a,b}(1-p^{a,b}) = (1-2p^{a,b})(\frac{n_a}{n_a+n_b} - p) - (\frac{n_a}{n_a+n_b} - p^{a,b})^2$. By setting $\Xi^{a,b}(\delta) = \frac{1}{4(n_a+n_b-1)} + \frac{\log\frac{2}{\delta}}{2(n_a+n_b)} = O(\frac{1}{n_a+n_b})$, we yield

$$\begin{aligned}&\sum_{a=1}^{k-1}\sum_{b=a+1}^{k}|\frac{n_a+n_b}{(n_a+n_b)^2} - p^{a,b}(1-p^{a,b}) + \frac{n_an_b}{(n_a+n_b)^2(n_a+n_b-1)}| \\ &\le \sum_{a=1}^{k-1}\sum_{b=a+1}^{k}(|\frac{n_a+n_b}{(n_a+n_b)^2} - p^{a,b}(1-p^{a,b})| + \frac{1}{4(n_a+n_b-1)}) \\ &\le \sum_{a=1}^{k-1}\sum_{b=a+1}^{k}((1-2p^{a,b})\sqrt{\frac{\log\frac{2}{\delta}}{2(n_a+n_b)}} + \Xi^{a,b}(\delta)).\end{aligned}$$

In view of

$$\sum_{a=1}^{k-1}\sum_{b=a+1}^{k}\frac{1}{2(n_a+n_b)} \le \sum_{a=1}^{k-1}\sum_{b=a+1}^{k}\frac{1}{n_a+n_b-1},$$

the following inequality is established with possibility at least $1-\delta$:

$$\widehat{\Psi} \le \sum_{a=1}^{k-1}\sum_{b=a+1}^{k}\{(C\sqrt{\frac{\Gamma}{n_a+n_b}} \tag{12}$$

$$+ 2(1-p^{a,b})\sqrt{\frac{\log\frac{3}{\delta}}{n_a+n_b-1}} + \Xi^{a,b}(\frac{2\delta}{3}))((1-p^{a,b})p^{a,b})^{-1}\}. \tag{13}$$

Now, we consider $\widehat{\Psi}_0$. Since

$$\widehat{AUC}_{H^0_{f,a},H^0_{f,b}}(S,f) = \sum_{a=1}^{k-1}\sum_{b=a+1}^{k}\frac{(n_a+n_b)(n_a+n_b-1)}{2n^0_an^0_b}\widehat{U}^0_{a,b}(f),$$

where

$$\begin{aligned}\widehat{U}^0_{a,b}(f) &= \frac{2}{(n_a+n_b)(n_a+n_b-1)}\sum_{i=1}^{n_a}\sum_{j=1}^{n_b}\{\mathrm{I}(f(v_i^a) > f(v_j^b)|z_i^a = z_j^b = 0) \\ &\quad + \frac{1}{2}\mathrm{I}(f(v_i^a) = f(v_j^b)|z_i^a = z_j^b = 0)\}.\end{aligned}$$

Moreover, it's not difficult to verify the following fact

$$\begin{aligned}U^0_{a,b}(f) &= \mathbb{E}[\widehat{U}^0_{a,b}(f)] \\ &= 2(q_0^{a,b})^2p_0^{a,b}(1-p_0^{a,b})AUC^{a,b}_{H^0_{f,a},H^0_{f,b}}(f).\end{aligned}$$

In light of the similar trick as dealing with $\widehat{\Psi}$, we acquire

$$\begin{aligned}\widehat{\Psi}_0 &\le \sum_{a=1}^{k-1}\sum_{b=a+1}^{k}\{\frac{1}{(q_0^{a,b})^2p_0^{a,b}(1-p_0^{a,b})}|\frac{n^0_an^0_b}{(n_a+n_b)^2} - (q_0^{a,b})^2p_0^{a,b}(1-p_0^{a,b}) \\ &\quad + \frac{n^0_an^0_b}{(n_a+n_b)^2(n_a+n_b-1)}| \\ &\quad + \frac{1}{2(q_0^{a,b})^2p_0^{a,b}(1-p_0^{a,b})}\sup_{f\in\mathcal{F}}|\widehat{U}^0_{a,b}(f) - U^0_{a,b}(f)|\}.\end{aligned}$$

Via calculation and simplification, we confirm that for each pair of (a, b),

$$\begin{aligned}&\frac{n_a^0 n_b^0}{(n_a+n_b)^2}-(q_0^{a,b})^2 p_0^{a,b}(1-p_0^{a,b})\\&=(\frac{n_a^0+n_b^0}{n_a+n_b}-q_0^{a,b})q_0^{a,b}p_0^{a,b}\\&+(\frac{n_a^0}{n_a+n_b}-p_0^{a,b}q_0^{a,b})q_0^{a,b}(1-2p_0^{a,b})\\&+(\frac{n_a^0}{n_a+n_b}-p_0^{a,b}q_0^{a,b})(\frac{n_a^0+n_b^0}{n_a+n_b}-q_0^{a,b})\\&-(\frac{n_a^0}{n_a+n_b}-p_0^{a,b}q_0^{a,b})^2.\end{aligned}$$

According to Hoeffding inequalities again, we deduce that with possibility at least $1-\delta$, two inequalities stated as follows estabilished simultaneously:

$$\sum_{a=1}^{k-1}\sum_{b=a+1}^{k}|\frac{n_a^0+n_b^0}{n_a+n_b}-q_0^{a,b}| \le \sum_{a=1}^{k-1}\sum_{b=a+1}^{k}\sqrt{\frac{\log\frac{4}{\delta}}{2(n_a+n_b)}}, \quad (14)$$

$$\sum_{a=1}^{k-1}\sum_{b=a+1}^{k}|\frac{n_a^0}{n_a+n_b}-p_0^{a,b}q_0^{a,b}| \le \sum_{a=1}^{k-1}\sum_{b=a+1}^{k}\sqrt{\frac{\log\frac{4}{\delta}}{2(n_a+n_b)}}. \quad (15)$$

Set $\Upsilon^{a,b}(\delta)=\frac{1}{4(n_a+n_b-1)}+\frac{\log\frac{4}{\delta}}{n_a+n_b}$. We verify

$$\begin{aligned}&\sum_{a=1}^{k-1}\sum_{b=a+1}^{k}|\frac{n_a^0n_b^0}{(n_a+n_b)^2}-(q_0^{a,b})^2p_0^{a,b}(1-p_0^{a,b})\\&\quad+\frac{n_a^0n_b^0}{(n_a+n_b)^2(n_a+n_b-1)}|\\&\le\sum_{a=1}^{k-1}\sum_{b=a+1}^{k}\{|\frac{n_a^0n_b^0}{(n_a+n_b)^2}-(q_0^{a,b})^2p_0^{a,b}(1-p_0^{a,b})|\\&\quad+|\frac{(n_a^0+n_b^0)^2}{4(n_a+n_b)^2(n_a+n_b-1)}|\}\\&\le\sum_{a=1}^{k-1}\sum_{b=a+1}^{k}\{q_0^{a,b}(1-p_0^{a,b})\sqrt{\frac{\log\frac{4}{\delta}}{2(n_a+n_b)}}+\Upsilon^{a,b}(\delta)\}.\end{aligned}$$

Using $\sum_{a=1}^{k-1}\sum_{b=a+1}^{k}\frac{1}{2(n_a+n_b)}\le\sum_{a=1}^{k-1}\sum_{b=a+1}^{k}\frac{1}{n_a+n_b-1}$ again, we have the following holds with possibility at least $1-\delta$:

$$\begin{aligned}\widehat{\Psi}_0\le&\sum_{a=1}^{k-1}\sum_{b=a+1}^{k}\{(C\sqrt{\frac{\Gamma}{n_a+n_b}}+(1+q_0^{a,b}(1\\&-p_0^{a,b}))\sqrt{\frac{\log\frac{5}{\delta}}{n_a+n_b}}+\Upsilon^{a,b}(\frac{4\delta}{5}))((q_0^{a,b})^2(1\\&-p_0^{a,b})p_0^{a,b})^{-1}\}.\end{aligned} \quad (16)$$

When it comes to $\widehat{\Psi}_1$, we have the similar conclusion that with possibility at least $1-\delta$,

$$\begin{aligned}\widehat{\Psi}_1\le&\sum_{a=1}^{k-1}\sum_{b=a+1}^{k}\{(C\sqrt{\frac{\Gamma}{n_a+n_b}}+(1+q_1^{a,b}(1\\&-p_1^{a,b}))\sqrt{\frac{\log\frac{5}{\delta}}{n_a+n_b}}+\Upsilon^{a,b}(\frac{4\delta}{5}))((q_1^{a,b})^2(1\\&-p_1^{a,b})p_1^{a,b})^{-1}\}.\end{aligned} \quad (17)$$

Finally, put (12), (16) and (17) together, apply the assumption $\min_{y\in\{1,\cdots,k\},z\in\{0,1\}}\mathbb{P}\{Y=y,Z=z\}\ge\varepsilon$, and note the fact that $\min\{p^{a,b},1-p^{a,b}\}\ge 2\varepsilon$, we get

$$\begin{aligned}&R_\lambda(f_\lambda^*)-R_\lambda(\widehat{f}_{\lambda,S})\\&\le(C\binom{k}{2}\sqrt{\frac{\Gamma}{2n_{\min}}}(4\lambda+\frac{1}{2})+\binom{k}{2}\sqrt{\frac{\log(\frac{13}{\delta})}{2n_{\min}-1}}(4\lambda\\&\quad+(4\lambda+2)\varepsilon)+O(\frac{\binom{k}{2}}{2n_{\min}}))(\varepsilon^{-2})\end{aligned}$$

holds with possibility at least $1-\delta$. Hence we get the desired conclusion. □

IV. CONCLUSION AND DISCUSSION

In this article, a novel multi-dividing ontology learning setting is proposed where each rate of sub ontology sample is divided into two groups, the corresponding objective function is given and the generalization bound in this specific ontology setting is determined by means of U-statistics technique and presented in Theorem 1.

In the real ontology engineering applications, $\widehat{R}_\lambda$ is tough difficult to maximin since truth function I is a discrete undifferentiable function. In statistical learning, one commended trick is replacing truth function to logistic function $\sigma: x\to\frac{1}{1+e^{-x}}$. In this way, the surrogate relaxation of multi-dividing ontology AUC criterion $\widehat{AUC}_{H_{f,a},H_{f,b}}$ is denoted by

$$A\tilde{U}C_{H_{f,a},H_{f,b}}=\sum_{a=1}^{k-1}\sum_{b=a+1}^{k}\frac{1}{n_an_b}\sum_{i=1}^{n_a}\sum_{j=1}^{n_b}\sigma(f(v_i^a)-f(v_j^b)).$$

When it comes to two-sample multi-dividing ontology setting, for $z\in\{0,1\}$, the corresponding objective ontology function we expect is to maximize

$$\begin{aligned}&A\tilde{U}C_{H_{f,a},H_{f,b}}(S,f)-c\lambda|A\tilde{U}C_{H_{f,a}^0,H_{f,b}^0}(S,f)\\&\qquad\qquad-A\tilde{U}C_{H_{f,a}^1,H_{f,b}^1}(S,f)|,\end{aligned}$$

where

$$\begin{aligned}&A\tilde{U}C_{H_{f,a}^0,H_{f,b}^0}(S,f)\\&=\sum_{a=1}^{k-1}\sum_{b=a+1}^{k}\frac{1}{n_a^0n_b^0}\sum_{i=1,z_i^a=0}^{n_a}\sum_{j=1,z_j^b=0}^{n_b}\sigma(f(v_i^a)-f(v_j^b)),\end{aligned}$$

$$\begin{aligned}&A\tilde{U}C_{H_{f,a}^1,H_{f,b}^1}(S,f)\\&=\sum_{a=1}^{k-1}\sum_{b=a+1}^{k}\frac{1}{n_a^1n_b^1}\sum_{i=1,z_i^a=1}^{n_a}\sum_{j=1,z_j^b=1}^{n_b}\sigma(f(v_i^a)-f(v_j^b)).\end{aligned}$$

However, we don't know exactly the statistical characters of the above surrogate relaxation of multi-dividing ontology criterion in two-sample assumption, and therefore it deserves to be studied in the future.

So far, thousands of ontologies have been defined according to their specific needs, and they are distributed in various fields of natural science and social science. Due to different ontology applications, and even different application backgrounds of the same ontology, the ontology data will perform large differences, which leads to different ontology sample dividing in the two-sample setting. In other words, for each ontology application, specific problems must be analyzed in detail, and unified parameters and standards cannot be used for direct execution. Our work only stays at the theoretical stage. For the specific application of the multi-dividing two-sample ontology algorithm in the specific ontology application field, further research is needed in future works.

REFERENCES

[1] W. Gao, L. Zhu, and K. Wang, "Ontology sparse vector learning algorithm for ontology similarity measuring and ontology mapping via ADAL technology," *Int. J. Bifurcation Chaos*, vol. 25, no. 14, Dec. 2015, Art. no. 1540034.

[2] W. Gao, Y. Zhang, J. L. G. Guirao, and Y. Chen, "A discrete dynamics approach to sparse calculation and applied in ontology science," *J. Difference Equ. Appl.*, vol. 25, nos. 9–10, pp. 1239–1254, Oct. 2019.

[3] W. Gao, J. L. G. Guirao, B. Basavanagoud, and J. Wu, "Partial multi-dividing ontology learning algorithm," *Inf. Sci.*, vol. 467, pp. 35–58, Oct. 2018.

[4] W. Gao and M. R. Farahani, "Generalization bounds and uniform bounds for multi-dividing ontology algorithms with convex ontology loss function," *Comput. J.*, vol. 60, pp. 1289–1299, Jan. 2017.

[5] P. Skalle and A. Aamodt, "Petrol 18 946: Downhole failures revealed through ontology engineering," *J. Petroleum Sci. Eng.*, vol. 191, Aug. 2020, Art. no. 107188.

[6] T. Sobral, T. Galvão, and J. Borges, "An ontology-based approach to knowledge-assisted integration and visualization of urban mobility data," *Expert Syst. Appl.*, vol. 150, Jul. 2020, Art. no. 113260.

[7] M. M. Al-Sayed, H. A. Hassan, and F. A. Omara, "CloudFNF: An ontology structure for functional and non-functional features of cloud services," *J. Parallel Distrib. Comput.*, vol. 141, pp. 143–173, Jul. 2020.

[8] G. Tebes, D. Peppino, P. Becker, G. Matturro, M. Solari, and L. Olsina, "Analyzing and documenting the systematic review results of software testing ontologies," *Inf. Softw. Technol.*, vol. 123, Jul. 2020, Art. no. 106298.

[9] D. Pradeep and C. Sundar, "QAOC: Novel query analysis and ontology-based clustering for data management in Hadoop," *Future Gener. Comput. Syst.*, vol. 108, pp. 849–860, Jul. 2020.

[10] A. M. Hema and K. Kuppusamy, "A novel trust-based privacy preservation framework for service handling via ontology service ranking," *Wireless Pers. Commun.*, vol. 112, no. 3, pp. 1339–1354, Jun. 2020.

[11] R. Messaoudi, A. Mtibaa, A. Vacavant, F. Gargouri, and F. Jaziri, "Ontologies for liver diseases representation: A systematic literature review," *J. Digit. Imag.*, vol. 33, no. 3, pp. 563–573, Jun. 2020.

[12] A. Mantovani, F. Piana, and V. Lombardo, "Ontology-driven representation of knowledge for geological maps," *Comput. Geosci.*, vol. 139, Jun. 2020, Art. no. 104446.

[13] R. Abeysinghe, E. W. Hinderer, H. N. B. Moseley, and L. Cui, "SSIF: Subsumption-based sub-term inference framework to audit gene ontology," *Bioinformatics*, vol. 36, no. 10, pp. 3207–3214, May 2020.

[14] M. Kossmann, A. Samhan, M. Odeh, E. Qaddoumi, A. Tbakhi, and S. Watts, "Extending the scope of configuration management for the development and life cycle support of systems of systems—An ontology-driven framework applied to the enceladus submarine exploration lander," *Syst. Eng.*, vol. 23, no. 3, pp. 366–391, May 2020.

[15] W. Gao and Y. Chen, "Approximation analysis of ontology learning algorithm in linear combination setting," *J. Coloud Comput.*, vol. 9, pp. 1–10, Dec. 2020, doi: 10.1186/s13677-020-00173-y.

[16] L. Ma and W. H. Wong, "Coupling optional Pólya trees and the two sample problem," *J. Am. Stat. Assoc.*, vol. 106, no. 496, pp. 1553–1565, 2011.

[17] M. Tang, A. Athreya, D. L. Sussman, V. Lyzinski, Y. Park, and C. E. Priebe, "A semiparametric two-sample hypothesis testing problem for random graphs," *J. Comput. Graph. Stat.*, vol. 26, no. 2, pp. 344–354, 2017.

[18] Y. C. Chen, Y. S. Wang, and E. A. Erosheva, "On the use of bootstrap with variational inference: Theory, interpretation, and a two-sample test example," *Ann. Appl. Stat.*, vol. 12, no. 2, pp. 846–876, 2018.

[19] I. Kim, A. B. Lee, and J. Lei, "Global and local two-sample tests via regression," *Electron. J. Stat.*, vol. 13, no. 20, pp. 5253–5305, 2019.

[20] N. Rabin, M. Golan, G. Singer, and D. Kleper, "Modeling and analysis of students' performance trajectories using diffusion maps and kernel two-sample tests," *Eng. Appl. Artif. Intel.*, vol. 85, pp. 492–503, 2019.

[21] T. Emura and J. H. Hsu, "Estimation of the Mann–Whitney effect in the two-sample problem under dependent censoring," *Comput. Statist. Data. Anal.*, vol. 150, May 2020, Art. no. 106990.

[22] M. Fuchs, R. Hornung, A.-L. Boulesteix, and R. De Bin, "On the asymptotic behaviour of the variance estimator of a U-statistic," *J. Stat. Plan. Infer.*, vol. 209, pp. 101–111, Apr. 2020.

[23] S. Bouzebda and B. Nemouchi, "Uniform consistency and uniform in bandwidth consistency for nonparametric regression estimates and conditional U-statistics involving functional data," *J. Nonparametr. Stat.*, vol. 32, no. 2, pp. 452–509, 2020.

[24] C. Fuglsby, C. P. Saunders, and J. Buscaglia, "U-statistics for estimating performance metrics in forensic handwriting analysis," *J. Stat. Comput. Sim.*, vol. 90, no. 6, pp. 1082–1117, 2020.

[25] B. Privault and G. Serafin, "Normal approximation for sums of weighted U-statistics—Application to Kolmogorov bounds in random subgraph counting," *Bernoulli*, vol. 26, no. 1, pp. 587–615, 2020.

[26] S. Bachmann and M. Reitzner, "Concentration for Poisson U-statistics: Subgraph counts in random geometric graphs," *Stochastic Processes Appl.*, vol. 128, no. 10, pp. 3327–3352, 2018.

[27] M. Garg and I. Dewan, "On limiting distribution of U-statistics based on associated random variables," *Stat. Probabil. Lett.*, vol. 132, pp. 7–16, Jan. 2018.

[28] O. Bousquet, S. Boucheron, and G. Lugosi, "Introduction to statistical learning theory," *Advanced Lectures on Machine Learning* (Lecture Notes in Computer Science), vol. 3176. Berlin, Germany: Springer, 2004, pp. 169–207.

[29] S. Clémencon, G. Lugosi, and N. Vayatis, "Ranking and empirical minimization of U-statistics," *Ann. Stat.*, vol. 36, no. 2, pp. 844–874, 2008.

LINLI ZHU was born in Badong, Hubei, China, in 1975. He received the master's degree in computer software and theory from Yunnan Normal University in 2007. He is currently pursuing the Ph.D. degree with the School of Information and Control Engineering, China University of Mining and Technology. He has been an Associate Professor with the Department of Computer Engineering, Jiangsu University of Technology, since 2012. As a researcher in computer science, his interests are computer networks and artificial intelligence.

GANG HUA received the B.S. degree from Southeast University, Nanjing, and the M.S. and Ph.D. degrees from the China University of Mine and Technology, Xuzhou, China, in 2000 and 2004, respectively. He is currently a Professor with the Department of Information and Electronic Engineering, China University of Mine and Technology, Xuzhou, China. His research interests include the control and supervision of mining safety, signal processing, compressed sensing, and network technology.