

## Research Article

## Open Access

Linli Zhu, Gang Hua\*, and Adnan Aslam

# Ontology learning algorithm using weak functions

<https://doi.org/10.1515/phys-2018-0112>

Received November 4, 2018; accepted November 11, 2018

**Abstract:** Ontology is widely used in information retrieval, image processing and other various disciplines. This article discusses how to use machine learning approach to solve the most essential similarity calculation problem in multi-dividing ontology setting. The ontology function is regarded as a combination of several weak ontology functions, and the optimal ontology function is obtained by an iterative algorithm. In addition, the performance of the algorithm is analyzed from a theoretical point of view by statistical methods, and several results are obtained.

**Keywords:** ontology, similarity measuring, ontology mapping, machine learning

**PACS:** 05.10.-a, 02.50.-r

## 1 Introduction

As a structured concept representation model, the ontology has been applied to the field of artificial intelligence since its inception, and then applied to other areas of the computer, such as machine vision, parallel computing, information query extension, mathematical logic representation, and so on. In the past decade, as a useful tool, the research and application of ontology has expanded to the entire engineering science. Related applications of ontology models can be found in various fields such as chemistry, biology, pharmaceutical, material, medicine, neuroscience, and social sciences. In each special application field, a large number of professional ontology are constructed every year and applied to specific practices (for

instance, “GO” ontology in gene science and “PO” ontology in plant science). More related ontology work and engineering application can be referred to Biletskiy et al. [1], Benedikt et al. [2], Rajbabu et al. [3], Vidal et al. [4], Annane et al. [5], Adhikari et al. [6], Mili et al. [7], Ferreira et al. [8], Bayoudhi et al. [9], and Derguech et al. [10].

The core of various ontology algorithms is the similarity calculation between ontology concepts. For ontology mapping, the essence is to calculate the similarity of concepts between different ontology, so the ontology similarity calculation algorithm we designed is also applicable to ontology mapping. Some related studies on ontology mapping can be referred to Ding and Foo [11], Kalfoglou and Schorlemmer [12], Currie et al. [13], Wong et al. [14], Qazvinian et al. [15], Nagy and Vargas-Vera [16], Lukasiewicz et al. [17], Arch-int and Arch-int [18], Forsati and Shamsfard [19], and Sicilia et al. [20].

In recent years, as the amount of data processed by various types of applications has expanded the data stored and processed by the ontology has also been expanding. This has led to an increase in the requirements of ontology algorithms in the era of big data. Especially in biology and pharmacy where the ontology is responsible for handling large amounts of information. In order to meet the needs of practical engineering, learning algorithms are gradually applied to ontology similarity calculation and ontology mapping, and then applied to various subject areas. Several ontology learning algorithms and application to different engineering application can be found in Gao and Zhu [21], and Gao et al. [22], [23], [24], [25], and [27].

Several papers contributed to the theoretical analysis of machine learning based ontology algorithms. Gao et al. [28] studied the strong and weak stability of  $k$ -partite ranking based ontology algorithm. Gao and Xu [29] considered the uniform stability of learning algorithms for ontology similarity computation. Gao and Farahani [30] presented the generalization bounds and uniform bounds for multi-dividing ontology algorithms with convex ontology loss function. Gao et al. [31] proposed a partial multi-dividing ontology algorithm with the aim of obtaining an efficient trick to optimize the partial multi-dividing ontology learning framework, and several theoretical results from a statistical learning theory perspective are obtained.

**Linli Zhu:** School of Information and Control Engineering, China University of Mining and Technology, Xuzhou, China; School of Computer Engineering, Jiangsu University of Technology, Changzhou, China, E-mail: zhulinli@jsut.edu.cn

**\*Corresponding Author: Gang Hua:** School of Information and Control Engineering, China University of Mining and Technology, Xuzhou, China, E-mail: ghua@cumt.edu.cn

**Adnan Aslam:** Department of Natural Sciences and Humanities, University of Engineering and Technology, Lahore 54000, Pakistan, E-mail: adnanaslam15@yahoo.com

In this paper, we continue to study the theoretical analysis of ontology learning algorithm and focus on the multi-dividing ontology algorithm. The rest of the paper is organised as follows. First, we introduce the setting of multi-dividing ontology learning algorithm, and some notations are introduced. Then we manifest the main algorithm which is based on weak ontology functions. Finally, we give some results and detailed proofs from the perspective of statistical learning theory.

## 2 Setting

We use graph  $G = (V, E)$  to express the structure of the ontology and call it ontology graph in which each vertex represents a concept and each edge indicates a direct relationship (for example, belonging between two concepts). Assume  $S : V \times V \rightarrow \mathbb{R}^+ \cup \{0\}$  is the similarity function in ontology, and we usually unitize its value to  $[0, 1]$ . That is to say, similarity function  $S : V \times V \rightarrow [0, 1]$  maps each pair of vertices (concepts) to a real number belonging to interval from 0 to 1. Let  $v_1$  and  $v_2$  be two vertices in the ontology graph,  $S(v_1, v_2) = 1$  indicates that the concepts corresponding to  $v_1$  and  $v_2$  have the same meaning. Conversely,  $S(v_1, v_2) = 0$  means that there is no relationship between  $v_1$  and  $v_2$ . Fixed threshold  $M \in [0, 1]$  with the help of field experts, then for vertex  $v$ , we return a set of concepts  $\{v' | S(v, v') \geq M\}$  to the user as similarity vertices. In what follows, we always assume that  $n$  is the number of ontology samples, and it's called sample capacity.

Let  $S = \{v_1, \dots, v_n\}$  be the ontology sample set which is independent identically distributed according to an unknown distribution  $\mathcal{D}$  (we write  $v_i \sim \mathcal{D}$  for  $i \in \{1, \dots, n\}$ ), and  $l$  be the ontology loss function (we always assume it is convex, and can be express as  $l(f, v)$  with respect to ontology function  $f : V \rightarrow \mathbb{R}$  and ontology sample  $v$ ). The expected risk of ontology model is

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

However,  $er(f)$  can't be computed since we don't know  $\mathcal{D}$ . Instead, it is naturally to obtain optimal ontology vector via the ontology empirical framework as follows

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

When it comes to ontology learning setting, we aim to learn an ontology function  $f : V \rightarrow \mathbb{R}$  which maps each vertex to a real number. The similarity between ontology vertices  $v_1$  and  $v_2$  can be measured by means of  $|f(v_1) - f(v_2)|$ : the bigger of value  $|f(v_1) - f(v_2)|$  is, the

smaller similarity between  $v_1$  and  $v_2$  will be; the smaller of value  $|f(v_1) - f(v_2)|$  is, the larger similarity between  $v_1$  and  $v_2$  will become. In order to connect the statistic learning theory, all the information for a vertex  $v$  is package to a  $p$ -dimensional vector. To simplify expression, without confusion, we use  $v$  to express vertex, its corresponding vector and its corresponding ontology concept, and this mathematical symbol is no longer bolded in the following context.

Suppose that  $(v_i, y_i)$  are independent and identically distributed random variables to certain unknown distribution  $\mathcal{D}$ , where  $y_i \in \mathcal{Y}$  is the label of ontology vertex  $v_i$ . Fixed the ontology function  $f$ , denote  $l(f, v_i, y_i)$  as the ontology loss, and then the expected ontology risk can be stated as

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

Given ontology training sample set  $\{(v_i, y_i)\}_{i=1}^n$ , then the corresponding empirical ontology risk can be expressed as

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

When it comes to the pairwise setting, we also assume that  $(v_i, y_i)$  and  $(v_j, y_j)$  are independent and identically distributed random variables to certain unknown distribution  $\mathcal{D}$ , where  $y_i, y_j \in \mathcal{Y}$  are the label of ontology vertex  $v_i$  and  $v_j$ . Let  $l(f, v_i, v_j, y_i, y_j)$  be the ontology loss function, where  $y_{i,j}$  can be regarded as the function of  $y_i$  and  $y_j$ . Thus, the expected ontology risk becomes

$$er(f) = \int_{(V \times \mathcal{Y})^2} l(f, v_i, v_j, y_i, y_j) \mathcal{D}(dv_i, dy_i) \mathcal{D}(dv_j, dy_j).$$

With the ontology sample set  $\{(v_i, y_i)\}_{i=1}^n$ , the corresponding empirical ontology risk in the pairwise setting can be denoted by

$$\hat{er}(f) = \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{j=i+1}^n l(f, v_i, v_j, y_i, y_j).$$

### 2.1 Multi-dividing ontology setting

Since most ontology graph structures are trees (no cycle graph), multi-dividing ontology learning trick is popular in recent years. In this special ontology learning setting, all the vertices in the ontology graph are divided into  $k$  parts for  $k$  levels. The values of different levels are determined by domain experts in the specific application. For ontology function  $f$ , what we want is that the real number corresponding to the vertex in level  $a$  is greater than the real

number corresponding to the vertex in any level  $b$ , where  $1 \leq a < b \leq k$ . That is to say, in the ideal case,  $f(v) > f(v')$  if the level of vertex  $v$  is smaller than the level of vertex  $v'$ .

Formally, the learner is inferred to an ontology training set  $S = (S^1, S^2, \dots, S^k) \in V^{n_1} \times V^{n_2} \times \dots \times V^{n_k}$  which consists of a sequence of ontology training samples  $S^a = (v_1^a, \dots, v_{n_a}^a) \in V^{n_a}$  ( $1 \leq a \leq k$ ). By virtue of ontology sample  $S$ , a real-valued ontology function  $f : V \rightarrow \mathbb{R}$  is learned which allocates the future  $S^a$  vertices larger value than  $S^b$ , where  $a < b$ . Set  $\mathcal{D}_a$  as the conditional distributions for each rate  $1 \leq a \leq k$  and  $n = \sum_{i=1}^k n_i$  as the total size of ontology sample set, where  $n_i = |S_i|$  for  $i \in \{1, \dots, k\}$ .

The expected multi-dividing ontology expected risk on the ontology graph for an ontology function  $f : V \rightarrow \mathbb{R}$  associated with the ontology function  $f : V \rightarrow \mathbb{R}$  is defined as

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

The other expression for expected ontology risk can be formulated by

$$er(f) = \sum_{a=1}^{k-1} \sum_{b=a+1}^k \int l(f, v^a, v^b) \mathcal{D}_a(dv^a) \mathcal{D}_b(dv^b).$$

A large class of learning algorithms is generated by regularization schemes. They penalize an empirical error which is chosen here to be the multi-dividing empirical error on the ontology graph defined for a  $f : V \rightarrow \mathbb{R}$  associated with the sample  $S$  as

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

Thus, the optimal ontology function can be obtained by  $f^* = \operatorname{argmin}_{f \in \mathcal{F}} \hat{er}_{S,l}(f)$ . We simply write  $\hat{er}_{S,l}(f)$  as  $\hat{er}(f)$ .

The aim of this paper is to provide a new ontology learning algorithm in terms of weak ontology functions and give the theoretical analysis of in the multi-dividing setting.

### 3 New ontology learning algorithm and theoretical analysis

In this section, we first present our new ontology learning algorithm in multi-dividing setting based on weak ontology functions. Then, the theoretical analysis of the proposed ontology algorithm is derived.

#### 3.1 New ontology learning algorithm with weak ontology functions

Assume that the whole procedure can be separated by weak ontology functions in which it will be produced in each round. The new ontology learning algorithm keeps a distribution  $\mathcal{D}_t$  over  $V \times V$  which is passed on round  $t$  to the weak ontology function. In fact, it selects  $\mathcal{D}_t$  to emphasize different parts of the ontology training samples, and the large weight assigned to a pair of ontology vertices implies that it is very important for the weak ontology functions to map their order correctly.

We assume that the weak ontology functions keep the form  $f_t : V \rightarrow \mathbb{R}$ , and it provides order information in the same fashion as final ontology function. In the normal ontology setting ( $S = \{v_1, \dots, v_n\}$  and not to be divided into  $k$  parts), the procedure can be stated as follows:

Step 1: Given initial distribution  $\mathcal{D}$  over  $V \times V$  and set  $\mathcal{D}_1 = \mathcal{D}$ ;

Step 2: For  $t = 1, \dots, T$ , do the following actions: train weak ontology function by means of distribution  $\mathcal{D}_t$ ; obtain weak ontology function  $f_t : V \rightarrow \mathbb{R}$ ; select parameter  $\alpha_t \in \mathbb{R}$ ; calculate

$$\mathcal{D}_{t+1}(v_1, v_2) = \frac{\mathcal{D}_t(v_1, v_2) e^{\alpha_t(f_t(v_1) - f_t(v_2))}}{Z_t},$$

where  $Z_t$  is denoted as regularization parameter and thus  $\mathcal{D}_{t+1}$  will be determined;

Step 3: Return the final ontology function as the combination of weak ontology functions:

$$f(v) = \sum_{t=1}^T \alpha_t f_t(v).$$

In the step 2 above, one problem is how to get parameter  $\alpha_t$ . One method is to minimize  $Z_t$ , i.e.,

$$\begin{aligned} \alpha_t &= \arg \min \{Z_t\} \\ &= \arg \min \sum_{v_1, v_2} \mathcal{D}_t(v_1, v_2) e^{\alpha_t(f_t(v_1) - f_t(v_2))}. \end{aligned}$$

In the multi-dividing ontology setting, the above algorithm can be rewritten as follows.

Initialize: For each pair of  $(a, b)$  with  $1 \leq a < b \leq k$ , and  $v \in S^a \cup S^b$ , set  $\rho_1^{a,b}(v) = \frac{1}{n_a}$  if  $v \in S^a$  and  $\rho^{a,b}(v) = \frac{1}{n_b}$  if  $v \in S^b$ .

For  $t = 1, \dots, T$ :

- train the weak ontology function in terms of  $\mathcal{D}_t$  (if  $v_1 \in S^a$  and  $v_2 \in S^b$ , then  $\mathcal{D}_t(v_1, v_2) = \rho_t^{a,b}(v_1) \rho_t^{a,b}(v_2)$ ) and obtain weak ontology function  $f_t : V \rightarrow \mathbb{R}$ ;
- for each pair of  $(a, b)$  with  $1 \leq a < b \leq k$ , select  $\alpha_t^{a,b} \in \mathbb{R}$  and update

$$\rho_{t+1}^{a,b}(v) = \frac{\rho_t^{a,b}(v)e^{-\alpha_t^{a,b}f_t(v)}}{\sum_{v' \in S^a} \rho_t^{a,b}(v')e^{-\alpha_t^{a,b}f_t(v')}}.$$

if  $v \in S^a$  and

$$\rho_{t+1}^{a,b}(v) = \frac{\rho_t^{a,b}(v)e^{\alpha_t^{a,b}f_t(v)}}{\sum_{v' \in S^b} \rho_t^{a,b}(v')e^{-\alpha_t^{a,b}f_t(v')}}.$$

if  $v \in S^b$ .

Select balance parameter  $\alpha_t$  for each weak ontology function, and return the final ontology function  $f(v) = \sum_{t=1}^T \alpha_t f_t(v)$ .

### 3.2 Theoretical results

To ensure that the calculations for each iteration are valid, it must be confirmed that  $\mathcal{D}_t(v_1, v_2) = \rho_t^{a,b}(v_1)\rho_t^{a,b}(v_2)$  is true at each step (here  $v_1 \in S^a$  and  $v_2 \in S^b$ ). To show this, we use mathematical induction, assume that this is true for round  $t$ , and when it comes to round  $t+1$ , according to the produces, we have

$$\begin{aligned} & \mathcal{D}_t(v_1, v_2) \\ &= \frac{\rho_t^{a,b}(v_1)e^{-\alpha_t^{a,b}f_t(v_1)}}{\sum_{v' \in S^a} \rho_t^{a,b}(v')e^{-\alpha_t^{a,b}f_t(v')}} \times \frac{\rho_t^{a,b}(v_2)e^{\alpha_t^{a,b}f_t(v_2)}}{\sum_{v' \in S^b} \rho_t^{a,b}(v')e^{-\alpha_t^{a,b}f_t(v')}} \\ &= \rho_{t+1}^{a,b}(v_1)\rho_{t+1}^{a,b}(v_2). \end{aligned}$$

Note that our final ontology function has the form  $f(v) = \sum_{t=1}^T \alpha_t f_t(v)$ , and we can set  $\Theta : V \times V \rightarrow \{-1, 0, 1\}$  as

$$\Theta(v_1, v_2) = \text{sign}\left(\sum_{t=1}^T \alpha_t f_t(v_1) - \sum_{t=1}^T \alpha_t f_t(v_2)\right).$$

That is to say, if  $\sum_{t=1}^T \alpha_t f_t(v_1) - \sum_{t=1}^T \alpha_t f_t(v_2) > 0$  then  $\Theta(v_1, v_2) = 1$ ; if  $\sum_{t=1}^T \alpha_t f_t(v_1) = \sum_{t=1}^T \alpha_t f_t(v_2)$ , then  $\Theta(v_1, v_2) = 0$ ; and if  $\sum_{t=1}^T \alpha_t f_t(v_1) - \sum_{t=1}^T \alpha_t f_t(v_2) < 0$  then  $\Theta(v_1, v_2) = -1$ . For each pair of  $(a, b)$  with  $1 \leq a < b \leq k$ , if function  $\Theta(v^a, v^b) \neq 1$  where  $v^a \in S^a$  and  $v^b \in S^b$ , then it implies the error by the multi-dividing rule. Thus, the generalization error (expect risk) of  $\Theta$  in multi-dividing ontology setting is denoted as

$$\begin{aligned} \Delta(\Theta) &= \sum_{a=1}^{k-1} \sum_{b=a+1}^k \mathbb{P}_{v \sim \mathcal{D}_a, v' \sim \mathcal{D}_b} \{\Theta(v, v') \neq 1\} \\ &= \sum_{a=1}^{k-1} \sum_{b=a+1}^k \mathbb{E}_{\mathcal{D}_a, \mathcal{D}_b} \mathbb{I}(\Theta(v, v') \neq 1), \end{aligned}$$

where  $\mathbb{I}$  is the truth function, i.e.,  $\mathbb{I}(x) = 1$  if  $x$  is true, otherwise  $\mathbb{I}(x) = 0$ . Given the ontology training set  $S = (S^1, S^2, \dots, S^k) \in V^{n_1} \times V^{n_2} \times \dots \times V^{n_k}$  which consists of a sequence of ontology training samples  $S^a = (v_1^a, \dots, v_{n_a}^a) \in V^{n_a}$  ( $1 \leq a \leq k$ ), the expected empirical error of  $\Theta$  can be denoted as

$$\hat{\Delta}(\Theta) = \sum_{a=1}^{k-1} \sum_{b=a+1}^k \frac{1}{n_a n_b} \sum_{i=1}^{n_a} \sum_{j=1}^{n_b} \mathbb{I}(\Theta(v_i^a, v_j^b) \neq 1).$$

The results presented in our paper aim to show that the difference between  $\hat{\Delta}(\Theta)$  and  $\Delta(\Theta)$  is small with large possibility. Setting  $\Gamma$  as the function space for functions  $\Theta$  we have the following theorem.

**Theorem 1.** Suppose all the weak ontology functions belong to function space  $\mathcal{F}$  with a finite VC dimension  $K$ , the ontology functions  $f$  (as the weighted combinations of the weak ontology functions) belong to function space  $\mathcal{F}$ . Let  $S = (S^1, S^2, \dots, S^k) \in V^{n_1} \times V^{n_2} \times \dots \times V^{n_k}$  be ontology training set which consists of a sequence of ontology training samples  $S^a = (v_1^a, \dots, v_{n_a}^a) \in V^{n_a}$  and  $S^a \sim \mathcal{D}_a$  ( $1 \leq a \leq k$ ). We have with probability at least  $1 - \delta$  ( $0 < \delta < 1$ ), the following inequality holds for any  $f \in \mathcal{F}$ :

$$\begin{aligned} & |er(f) - \hat{er}(f)| \\ & \leq 2 \sum_{a=1}^{k-1} \sum_{b=a+1}^k \left\{ \sqrt{\frac{K'(\log \frac{2n_a}{K} + 1) + \log \frac{18}{\delta}}{n_a}} \right. \\ & \quad \left. + \sqrt{\frac{K'(\log \frac{2n_b}{K} + 1) + \log \frac{18}{\delta}}{n_b}} \right\}, \end{aligned}$$

where  $K' = 2(K+1)(T+1)\log_2(e(T+1))$ ,  $T$  is the number of weak ontology functions in ontology algorithm and  $e$  is the base of the natural logarithm.

**Proof of Theorem 1.** First, we show that for each pair of  $(a, b)$  with  $1 \leq a < b \leq k$ , and each  $\delta > 0$ , there is a small number  $\varepsilon$  satisfying

$$\begin{aligned} & \mathbb{P}\{\exists \Theta \in \Gamma : \left| \sum_{a=1}^{k-1} \sum_{b=a+1}^k \frac{1}{n_a n_b} \sum_{i=1}^{n_a} \sum_{j=1}^{n_b} \mathbb{I}(\Theta(v_i^a, v_j^b) \neq 1) \right. \\ & \quad \left. - \sum_{a=1}^{k-1} \sum_{b=a+1}^k \mathbb{E}_{v^a, v^b} \mathbb{I}(\Theta(v^a, v^b) \neq 1) \right| > \varepsilon\} \leq \delta, \end{aligned}$$

where the value of  $\varepsilon$  will be determined later.

Define  $\Xi : V \times V \rightarrow \{0, 1\}$  as  $\Xi(v^a, v^b) = \mathbb{I}(\Theta(v^a, v^b) \neq 1)$ . Clearly,  $\Xi$  indicates whether  $\Theta$  makes mistake or not for the ontology vertices pair  $(v^a, v^b)$  for  $v^a \in S^a$  and  $v^b \in S^b$  according to the multi-dividing rule. We infer

$$\sum_{a=1}^{k-1} \sum_{b=a+1}^k \frac{1}{n_a n_b} \sum_{i=1}^{n_a} \sum_{j=1}^{n_b} \mathbb{I}(\Theta(v_i^a, v_j^b) \neq 1)$$

$$\begin{aligned}
& - \sum_{a=1}^{k-1} \sum_{b=a+1}^k \mathbb{E}_{v^a, v^b} \mathbb{I}(\Theta(v^a, v^b) \neq 1) \\
& = \sum_{a=1}^{k-1} \sum_{b=a+1}^k \left\{ \frac{1}{n_a n_b} \sum_{i=1}^{n_a} \sum_{j=1}^{n_b} \Xi(v_i^a, v_j^b) \right. \\
& \quad \left. - \frac{1}{n_a} \sum_{i=1}^{n_a} \mathbb{E}_{v^b} \{\Xi(v_i^a, v^b)\} + \frac{1}{n_b} \sum_{j=1}^{n_b} \mathbb{E}_{v^a} \{\Xi(v^a, v_j^b)\} \right. \\
& \quad \left. - \mathbb{E}_{v^a, v^b} \{\Xi(v_i^a, v_j^b)\} \right\} \\
& = \sum_{a=1}^{k-1} \sum_{b=a+1}^k \left\{ \frac{1}{n_a} \sum_{i=1}^{n_a} \left( \frac{1}{n_b} \sum_{j=1}^{n_b} \Xi(v_i^a, v_j^b) \right) \right. \\
& \quad \left. - \mathbb{E}_{v^b} \Xi(v_i^a, v^b) \right\} + \mathbb{E}_{v^b} \left\{ \frac{1}{n_a} \sum_{i=1}^{n_a} \Xi(v_i^a, v^b) \right. \\
& \quad \left. - \mathbb{E}_{v^a} \Xi(v^a, v^b) \right\}.
\end{aligned}$$

Obviously, it is enough to show that there exist  $\varepsilon_1$  and  $\varepsilon_2$  with  $\varepsilon_1 + \varepsilon_2 = \varepsilon$  such that  $(\exists v^a \in V^a$  in each pair of  $(a, b)$  for (1) and  $\exists v^b \in V^b$  in each pair of  $(a, b)$  for (2))

$$\begin{aligned}
& \mathbb{P}\{\exists \Xi \in Y : \left| \sum_{a=1}^{k-1} \sum_{b=a+1}^k \frac{1}{n_b} \sum_{j=1}^{n_b} \Xi(v^a, v_j^b) \right. \\
& \quad \left. - \sum_{a=1}^{k-1} \sum_{b=a+1}^k \mathbb{E}_{v^b} \Xi(v^a, v^b) \right| \geq \varepsilon_1\} \leq \frac{\delta}{2}, \quad (1)
\end{aligned}$$

and

$$\begin{aligned}
& \mathbb{P}\{\exists \Xi \in Y : \left| \sum_{a=1}^{k-1} \sum_{b=a+1}^k \frac{1}{n_a} \sum_{i=1}^{n_a} \Xi(v_i^a, v^b) \right. \\
& \quad \left. - \sum_{a=1}^{k-1} \sum_{b=a+1}^k \mathbb{E}_{v^a} \Xi(v^a, v^b) \right| \geq \varepsilon_2\} \leq \frac{\delta}{2} \quad (2)
\end{aligned}$$

respectively, where  $Y$  is the function space of  $\Xi$ .

Now, we only prove (2) in light of standard results, and (1) can be yielded in the same fashion. Let  $Y_{v^b}$  be the set of all such functions  $\Xi$  for a given  $v^b$ , then the selection of  $\Xi$  in (2) is from function space  $\cup_{v^b} Y_{v^b}$ . In view of theorem of Vapnik [32] which provides a selection of  $\varepsilon_2$  in (2) relying on the size  $n_a$  of  $S^a$  for each pair of  $(a, b)$ , complexity  $K'$  of  $\cup_{v^b} Y_{v^b}$  (considered as VC Dimension), and the possibility  $\delta$ . Specifically, for any  $\delta > 0$ , set

$$\varepsilon_3 = 2 \sum_{a=1}^{k-1} \sum_{b=a+1}^k \sqrt{\frac{K'(\log \frac{2n_a}{K'} + 1) + \log \frac{18}{\delta}}{n_a}},$$

we have

$$\begin{aligned}
& \mathbb{P}\{\exists \Xi \in \cup_{v^b} Y_{v^b} : \left| \sum_{a=1}^{k-1} \sum_{b=a+1}^k \frac{1}{n_a} \sum_{i=1}^{n_a} \Xi(v_i^a, v^b) \right. \\
& \quad \left. - \sum_{a=1}^{k-1} \sum_{b=a+1}^k \mathbb{E}_{v^a} \Xi(v^a, v^b) \right| \geq \varepsilon_3\} \leq \delta.
\end{aligned}$$

Next, we need to determine the VC Dimension of  $\cup_{v^b} Y_{v^b} : K'$ . For a given  $v^b \in V^b$ , we obtain

$$\begin{aligned}
& \Xi(v^a, v^b) = \mathbb{I}(\Theta(v^a, v^b) \neq 1) \\
& = \mathbb{I}(\text{sign}(\sum_{t=1}^T \alpha_t f_t(v^a) - \sum_{t=1}^T \alpha_t f_t(v^b)) \neq 1) \\
& = \mathbb{I}(\sum_{t=1}^T \alpha_t f_t(v^a) - \sum_{t=1}^T \alpha_t f_t(v^b) \geq 0) \\
& = \mathbb{I}(\sum_{t=1}^T \alpha_t f_t(v^a) - c \geq 0)
\end{aligned}$$

where  $c = \sum_{t=1}^T \alpha_t f_t(v^b)$  is a constant since  $v^b$  is given. It reveals that the functions in the space  $\cup_{v^b} Y_{v^b}$  are the subset of all potential thresholds of all the linear combination of  $T$  ontology weak functions. Using the standard result on VC Dimension of weak functions, we yield that if the ontology weak function space has VC Dimension  $K$  bigger than two, then  $K'$  can't exceed to  $2(K+1)(T+1)\log_2(e(T+1))$ .

Therefore, we get the desired conclusion.  $\square$

According to Theorem 1 above, the generalization bound converges to zero at a rate of  $O(\sum_{a=1}^{k-1} \sum_{b=a+1}^k \max\{\sqrt{\frac{\log(n_a)}{n_a}}, \sqrt{\frac{\log(n_b)}{n_b}}\})$ .

For each pair of  $(a, b)$  with  $1 \leq a < b \leq k$ , the shatter coefficient is denoted as  $r^{a,b}(\mathcal{F}, n_a, n_b)$  (see Gao and Wang [33] for more details). Then we deduce the following result.

**Theorem 2.** Let  $\mathcal{F}$  be the real valued ontology function space on  $V$ , then with probability at least  $1 - \delta$  ( $0 < \delta < 1$ ), for any  $f \in \mathcal{F}$ , we have

$$|er(f) - \widehat{er}(f)| \leq$$

$$\sum_{a=1}^{k-1} \sum_{b=a+1}^k \sqrt{\frac{8(n_a + n_b)(\log \frac{4}{\delta} + \log r^{a,b}(\mathcal{F}, 2n_a, 2n_b))}{n_a n_b}}.$$

Theorem 2 implies that if the ontology function is a linear function in the one-dimensional function space, then for each pair of  $(a, b)$  with  $1 \leq a < b \leq k$ ,  $r^{a,b}(\mathcal{F}, n_a, n_b)$  are constants, regardless of the values of  $n_a$  and  $n_b$ , and thus the bound converges to zero at a rate of  $O(\sum_{a=1}^{k-1} \sum_{b=a+1}^k \max\{\frac{1}{\sqrt{n_a}}, \frac{1}{\sqrt{n_b}}\})$ . It further reveals that the bound yield in Theorem 2 is sharper than bound obtained in Theorem 1. However, if the ontology function is a linear function in the  $d$ -dimensional function space (where  $d \geq 2$ ), then  $r^{a,b}(\mathcal{F}, n_a, n_b)$  with order  $O((n_a n_b)^d)$ , and in this case the bound in Theorem 2 has convergence rate relying on VC dimension, i.e., still  $O(\sum_{a=1}^{k-1} \sum_{b=a+1}^k \max\{\sqrt{\frac{\log(n_a)}{n_a}}, \sqrt{\frac{\log(n_b)}{n_b}}\})$ .



## 4 Conclusion

Multi-dividing ontology learning algorithms have been proved to be effective in biology science, plant science, robot structure analysis, etc. It is necessary to give a deep theoretical analysis of this kind of algorithm. In this paper, we give a new ontology learning algorithm based on weak ontology functions, and we discuss the generation bound in this special setting. The obtained ontology algorithm and theoretical conclusions have potential engineering use in various fields.

**Conflict of Interests** The authors hereby declare that there is no conflict of interests regarding the publication of this paper.

**Acknowledgement:** We thank the reviewers for their constructive comments in improving the quality of this paper. This work was supported in part by the National Natural Science Foundation of China (51574232), the Open Research Fund by Jiangsu Key Laboratory of Recycling and Reuse Technology for Mechanical and Electronic Products (RRME-KF1612), the Industry-Academia Cooperation Innovation Fund Project of Jiangsu Province (BY2016030-06) and Six Talent Peaks Project in Jiangsu Province (2016-XYDXXJS-020).

## References

- [1] Biletskiy Y., Brown J.A., Ranganathan G.R., Bagheri E., Akbari I., Building a business domain meta-ontology for information pre-processing, *Inform. Process. Lett.*, 2018, 138, 81–88.
- [2] Benedikt M., Grau B.C., Kostylev E.V., Logical foundations of information disclosure in ontology-based data integration, *Artif. Int.*, 2018, 262, 52–95.
- [3] Rajbabu K., Srinivas H., Sudha S., Industrial information extraction through multi-phase classification using ontology for unstructured documents, *Comput. Ind.*, 2018, 2018, 137–147.
- [4] Vidal J.C., Rabelo T., Lama M., Amorim R., Ontology-based approach for the validation and conformance testing of xAPI events, *Know. Based Syst.*, 2018, 155, 22–34.
- [5] Annane A., Bellahsene Z., Azouaou F., Jonquet C., Building an effective and efficient background knowledge resource to enhance ontology matching, *J. Web Semant.*, 2018, 51, 51–68.
- [6] Adhikari A., Dutta B., Dutta A., Mondal D., Singh S., An intrinsic information content-based semantic similarity measure considering the disjoint common subsumers of concepts of an ontology, *J. Assoc. Inf. Syst. Tech.*, 2018, 69, 1023–1034.
- [7] Mili H., Valtchev P., Szathmary L., Boubaker A., Leshob A., Charif Y., Martin L., Ontology-based model-driven development of a destination management portal: Experience and lessons learned, *Software Pract. Exper.*, 2018, 48, 1438–1460.
- [8] Ferreira W., Baldassarre M.T., Soares S., Codex: A metamodel ontology to guide the execution of coding experiments, *Comput. Stand. Inter.*, 2018, 59, 35–44.
- [9] Bayoudhi L., Sassi N., Jaziri W., How to repair inconsistency in OWL 2 DL ontology versions?, *Data Knowl. Eng.*, 2018, 116, 138–158.
- [10] Derguech W., Bhiri S., Curry E., Using ontologies for business capability modelling: describing what services and processes achieve, *Comput. J.*, 2018, 61, 1075–1098.
- [11] Ding Y., Foo S., Ontology research and development. Part 2—a review of ontology mapping and evolving, *J. Inform. Sci.*, 2002, 28, 375–388.
- [12] Kalfoglou Y., Schorlemmer M., Ontology mapping: the state of the art, *Knowl. Eng. Rev.*, 2003, 18, 1–31.
- [13] Currie R.A., Bombail V., Oliver J.D., Moore D.J., Lim F.L., Gwilliam V., Kimber I., Chipman K., Moggs J.G., Orphanides G., Gene ontology mapping as an unbiased method for identifying molecular pathways and processes affected by toxicant exposure: Application to acute effects caused by the rodent non-genotoxic carcinogen diethylhexylphthalate, *Toxicol. Sci.*, 2005, 86, 453–469.
- [14] Wong A.K.Y., Ray P., Parameswaran N., Strassner J., Ontology mapping for the interoperability problem in network management, *IEEE J. Sel. Area. Comm.*, 2005, 23, 2058–2068.
- [15] Qazvinian V., Abolhassani H., Haeri S.H., Hariri B.B., Evolutionary coincidence-based ontology mapping extraction, *Expert Syst.*, 2008, 25, 221–236.
- [16] Nagy M., Vargas-Vera M., Multiagent ontology mapping framework for the semantic web, *IEEE T. Syst. Man. Cy. A*, 2011, 41, 693–704.
- [17] Lukasiewicz T., Predoiu L., Stuckenschmidt H., Tightly integrated probabilistic description logic programs for representing ontology mappings, *Ann. Math. Artif. Intel.*, 2011, 63, 385–425.
- [18] Arch-int N., Arch-int S., Semantic ontology mapping for interoperability of learning resource systems using a rule-based reasoning approach, *Expert Syst. Appl.*, 2013, 40, 7428–7443.
- [19] Forsati R., Shamsfard M., Symbiosis of evolutionary and combinatorial ontology mapping approaches, *Inform. Sciences*, 2016, 342, 53–80.
- [20] Sicilia A., Nemirovski G., Nolle A., Map-On: A web-based editor for visual ontology mapping, *Semantic Web*, 2017, 8, 969–980.
- [21] Gao W., Zhu L.L., Gradient learning algorithms for ontology computing, *Comput. Intell. Neurosci.*, 2014, <http://dx.doi.org/10.1155/2014/438291>.
- [22] Gao W., Zhu L.L., Guo Y., Wang K.Y., Ontology learning algorithm for similarity measuring and ontology mapping using linear programming, *J. Intell. Fuzzy Syst.*, 2017, 33, 3153–3163.
- [23] Gao W., Zhu L.L., Wang K.Y., Ontology sparse vector learning algorithm for ontology similarity measuring and ontology mapping via ADAL technology, *Int. J. Bifurcat. Chaos*, 2015, 25, DOI: 10.1142/S0218127415400349.
- [24] Gao W., Farahani M.R., Aslam A., Hosamani S., Distance learning techniques for ontology similarity measuring and ontology mapping, *Cluster Comput.*, 2017, 20, 959–968.
- [25] Gao W., Baig A.Q., Ali H., Sajjad W., Farahani M.R., Margin based ontology sparse vector learning algorithm and applied in biology science, *Saudi J. Biol. Sci.*, 2017, 24, 132–138.
- [26] Gao W., Zhu L.L., Wang K.Y., Ranking based ontology scheming using eigenpair computation, *J. Intell. Fuzzy Syst.*, 2016, 4,

- 2411–2419.
- [27] Gao W., Guo Y., Wang K.Y., Ontology algorithm using singular value decomposition and applied in multidisciplinary, *Cluster Comput.*, 2016, 19, 2201–2210.
- [28] Gao W., Gao Y., Zhang Y.G., Strong and weak stability of  $k$ -partite ranking algorithms, *Information*, 2012, 15, 4585–4590.
- [29] Gao W., Xu T.W., Stability analysis of learning algorithms for ontology similarity computation, *Abstr. Appl. Anal.*, 2013, <http://dx.doi.org/10.1155/2013/174802>.
- [30] Gao W., Farahani M.R., Generalization bounds and uniform bounds for multi-dividing ontology algorithms with convex ontology loss function, *Comput. J.*, 2017, 60, 1289–1299.
- [31] Gao W., Guirao J.L.G., Basavanagoud B., Wu J.Z., Partial multi-dividing ontology learning algorithm, *Inform. Sciences*, 2018, 467, 35–58.
- [32] Vapnik V.N., *Estimation of Dependences Based on Empirical Data*. Springer–Verlag, 1982.
- [33] Gao W., Wang W.F., Analysis of  $k$ -partite ranking algorithm in area under the receiver operating characteristic curve criterion, *Int. J. Comput. Math.*, 2018, 95, 1527–1547.