

# New Multi-dividing Ontology Learning Algorithm Using Special Loss Functions

Wei Gao<sup>\*</sup>, Li Liang<sup>1</sup> and Tianwei Xu

School of Information, Yunnan Normal University, Kunming 650500, China

**Abstract:** As an important data structure model, ontology has become one of the core contents in information science. Multi-dividing ontology algorithm combines the advantages of graph structure and learning algorithms proved to have high efficiency. In this paper, in terms of multi-dividing proper loss functions, we propose new multi-dividing ontology learning algorithms for similarity measure and ontology mapping construction. Several theoretical statistical characteristics supporting the new learning model are given. Finally, four experiments on different scientific fields verify that our multi-dividing ontology algorithm has high accuracy and efficiency in application implements.

**Keywords:** Multi-dividing, ontology, ontology mapping, similarity measure, strong proper loss.

## 1. INTRODUCTION

As a knowledge representation and conceptual shared model, ontology has been applied in image retrieval, knowledge management and information retrieval search extension. Acting as an effective concept semantic model, ontology also employed in disciplines beyond computer science, such as social science (for instance, see Bouzeghoub and Elbyed [1]), biology science (for instance, see Hu *et al.*, [2]) and geography science (for instance, see Fonseca *et al.*, [3]).

The ontology model is actually a graph  $G=(V,E)$ , each vertex  $v$  in an ontology graph  $G$  represents a concept and each edge  $e=v_i v_j$  on an ontology graph  $G$  represents a relationship between concepts  $v_i$  and  $v_j$ . The target of ontology similarity measure is to find a similarity function  $Sim: V \times V \rightarrow \mathbb{R}^+ \cup \{0\}$  which maps each pair of vertices to a real number. The aim of ontology mapping is to bridge the link between two or more ontologies. Let  $G_1$  and  $G_2$  be two ontology graphs corresponding ontology  $O_1$  and  $O_2$ , respectively. For each  $v \in G_1$ , find a set  $S_v \subseteq V(G_2)$  where the concepts correspond to vertices in  $S_v$  are semantically close to the concept correspond to  $v$ . One method to get such mapping is, for each  $v \in G_1$ , computing the similarity  $S(v, v_j)$  where  $v_j \in V(G_2)$  and choose a parameter  $0 < M < 1$ . Then  $S_v$  is a collection such that the element in  $S_v$  satisfies  $S(v, v_j) \geq M$ . In this point of view, the essence of ontology mapping is to obtain a similarity function  $S$  and select a suitable parameter  $M$ . In our article, we focus on the technologies to yield a similarity matrix by virtue of distance learning.

For ontology similarity measure and ontology mapping, there are several effective learning tricks. Wang *et al.*, [4]

proposed to learn a score function which maps each vertex to a real number, and the similarity between two vertices can be measured according to the difference of real number they correspond to. Huang *et al.*, [5] presented fast ontology algorithm for calculating the ontology similarity in a short time. Gao and Liang [6] raised that the optimal ontology function can be determined by optimizing NDCG measure, and applied such idea in physics education. Gao and Gao [7] deduced the ontology function using the regression approach. Huang *et al.*, [8] obtained ontology similarity function based on half transductive learning. Gao and Xu [9] explored the learning theory approach for ontology similarity computation using  $k$ -partite ranking method. Zhu *et al.*, [10] proposed new criterion for ontology computation from AUC and multi-dividing standpoint. Gao *et al.*, [11] presented new ontology mapping algorithm using harmonic analysis and diffusion regularization on hypergraph.

In this paper, we consider the new multi-dividing ontology learning model using strong proper loss function. In next section, we present the setting of multi-dividing ontology algorithm and give our main computation model. Then, we obtain some probability conclusion on multi-dividing strongly proper loss. At last, the experiments are designed to show the efficiency of the algorithm.

## 2. SETTING AND NEW MULTI-DIVIDING ALGORITHM

First, for each vertex in ontology graph, we use a vector to express all the information for such vertex. Then, we say  $V$  takes its value in a high dimension feature space. The elements in  $V$  are drawn independently and randomly according to some unknown distribution  $D$ . Given a training set  $S = \{v_1, \dots, v_n\} \subset V$  with size  $n$ , the aim of ontology learning algorithms is to get an optimal score function  $f: V \rightarrow \mathbb{R}^+$ , the similarity between two vertices is judged by the difference

between two real numbers which they correspond to. The multi-dividing method is a special kind of ontology learning approach in which vertices come from  $k$  categories and the learner is given examples of vertices labeled as the  $k$  classes.

Formally, the settings of standard multi-dividing ontology problems can be described as follows. There is an instance space  $V$  from which vertices are drawn, and the learner is given a training sample  $(S_1, S_2, \dots, S_k) \in V^{n_1} \times V^{n_2} \times \dots \times V^{n_k}$  consisting of a sequence of training sample  $S_a = (v_1^a, \dots, v_{n_a}^a)$  ( $1 \leq a \leq k$ ). The goal is to learn from these samples a real-valued ontology score function  $f: V \rightarrow \circ$  that orders the future  $S_a$  vertices to have higher scores than  $S_b$ , where  $a < b$ . We assume that vertices in each  $S_a$  are drawn randomly and independently according to some unknown distribution  $D^a$  on the vertex space  $V$ , respectively.

In this paper, we assume that there is an instance space  $V$  and labels  $Y = \{1, 2, \dots, k\}$ , with an unknown distribution  $D$  on  $V \times \{1, 2, \dots, k\}$ . For  $(V, Y): D$ ,  $v \in V$ , we denote  $\eta^{a,b}(v) = P\{Y = a | V = v, Y \in \{a, b\}\}$  and  $p^{a,b} = P\{Y = a | Y \in \{a, b\}\}$  for each pair of  $(a, b)$  with  $1 \leq a < b \leq k$ . Given a multi-dividing sample set  $S = (S_1, S_2, \dots, S_k)$ , the aim is to learn a multi-dividing ontology function  $f: V \rightarrow \circ$  that maps the ontology graph into a real line and maps each vertex into a real number. Specifically, the goal is to learn a multi-dividing ontology function  $f: V \rightarrow \circ$  with small error defined as

$$R_D^k(f) = \sum_{a=1}^{k-1} \sum_{b=a+1}^k E[\Pi_{(Y^a - Y^b)(f(V^a) - f(V^b)) < 0} + \frac{1}{2} \Pi_{f(V^a) = f(V^b)}], \tag{1}$$

where,  $(X, Y), (X', Y')$  are assumed to be drawn independent and identically distributed from  $D$ , and  $\Pi$  is denoted as truth function with  $\Pi(\cdot) = 1$  if its argument is true and 0 otherwise. Hence, the error (1) of multi-dividing ontology function  $f$  is the probability that a randomly drawn rate  $a$  vertex receives a larger score under  $f$  than a randomly drawn rate  $b$  vertex, with ties broken uniformly at random. Now we define the Bayes multi-dividing ontology risk as

$$R_D^{k,*} = \inf_{f: V \rightarrow \circ} R_D^k(f) = \sum_{a=1}^{k-1} \sum_{b=a+1}^k \frac{1}{2p^{a,b}(1-p^{a,b})} E_{V^a, V^b}[\min(\eta^{a,b}(V^a)(1-\eta^{a,b}(V^b)), \eta^{a,b}(V^b)(1-\eta^{a,b}(V^a)))]$$

Then, the multi-dividing ontology regret of a multi-dividing ontology function  $f: V \rightarrow \circ$  is defined as

$$\text{regret}_D^k(f) = R_D^k(f) - R_D^{k,*}$$

Given a prediction space  $\hat{Y} \subseteq \circ$ , a loss function  $l: \{1, 2, \dots, k\} \times \hat{Y} \rightarrow \circ^+$  assigns a penalty  $l(y, \hat{y})$  for predicting  $\hat{y} \in \hat{Y}$ . For each pair of  $(a, b)$ , loss function restrictions on rate  $a$  and rate  $b$  is defined as  $l^{a,b}: \{a, b\} \times \hat{Y} \rightarrow \circ^+$ . For any such loss  $l$ , the  $l$ -error of a multi-dividing ontology function  $f: V \rightarrow \hat{Y}$  is defined as

$$R_D^{l,k}(f) = \sum_{a=1}^{k-1} \sum_{b=a+1}^k E_{(V, Y): D^{a,b}}[l^{a,b}(Y, f(V))],$$

and the Bayes multi-dividing ontology  $l$ -risk is defined as

$$R_D^{l,k,*} = \inf_{f: V \rightarrow \hat{Y}} R_D^{l,k}(f).$$

We define the conditional Bayes multi-dividing ontology  $l$ -risk  $L_l^k: [0, 1] \times \hat{Y} \rightarrow \circ^+$  as

$$L_l^k(\eta, \hat{y}) = \sum_{a=1}^{k-1} \sum_{b=a+1}^k E_{Y: \eta^{a,b}}[l^{a,b}(Y, \hat{y})] = \sum_{a=1}^{k-1} \sum_{b=a+1}^k \eta^{a,b} l^{a,b}(a, \hat{y}) + (1 - \eta^{a,b}) l^{a,b}(b, \hat{y}),$$

where,  $Y: \eta^{a,b}$  denotes a  $\{a, b\}$ -rates random variable taking in rate  $a$  with probability  $\eta^{a,b}$ . We define the conditional multi-dividing ontology Bayes  $l$ -risk  $H_l^k: [0, 1] \rightarrow \circ^+$  as

$$H_l^k = \inf_{\hat{y} \in \hat{Y}} L_l^k(\eta, \hat{y}).$$

The conditional  $l$ -regret  $R_l^k: [0, 1] \times \hat{Y} \rightarrow \circ^+$  is then given by

$$R_l^k(\eta, \hat{y}) = L_l^k(\eta, \hat{y}) - H_l^k.$$

For each multi-dividing ontology function  $f: V \rightarrow \hat{Y}$ ,

$$R_D^{l,k}(f) = E_V[L_l^k(\eta(V), f(V))],$$

$$R_D^{l,k,*} = E_V[H_l^k(\eta(V))],$$

And any  $\hat{Y} \subseteq \circ$ , a multi-dividing loss function  $l: \{1, 2, \dots, k\} \times \hat{Y} \rightarrow \circ^+$ , the conditional multi-dividing ontology Bayes  $l$ -risk  $H_l^k$  is a concave function on  $[0, 1]$ .

Let  $\hat{Y} = [0, 1]$ . A multi-dividing ontology loss function  $l: \{1, 2, \dots, k\} \times [0, 1] \rightarrow \circ^+$  is said to be proper if for any  $\eta \in [0, 1]$ ,

$$\eta \in \arg \min_{\hat{\eta} \in [0, 1]} L_l^k(\eta, \hat{\eta}),$$

And strictly proper if the minimizer is unique for any  $\eta \in [0,1]$ . This implies that  $l$  is proper if for any  $\eta \in [0,1]$ ,  $H_i^k = L_i^k(\eta, \hat{\eta})$  and strictly proper if  $H_i^k < L_i^k(\eta, \hat{\eta})$  for all  $\eta \neq \hat{\eta}$ .

The terminology of properness can be applied to multi-dividing loss function operating on prediction spaces  $\hat{Y}$  beyond  $[0,1]$  by virtue of compositing with a link function  $\Psi : [0,1] \rightarrow \hat{Y}$ . Specifically, for any  $\hat{Y} \subseteq \circ$ , a multi-dividing ontology loss function  $l : \{1, 2, \dots, k\} \times \hat{Y} \rightarrow \circ^+$  is said to be proper composite if it can be expressed as

$$l(y, \hat{y}) = l'(y, \Psi^{-1}(\hat{y}))$$

for some proper multi-dividing loss  $l' : \{1, 2, \dots, k\} \times [0,1] \rightarrow \circ^+$  and strictly increasing, invertible link function  $\Psi : [0,1] \rightarrow \hat{Y}$ .

Let  $l : \{1, 2, \dots, k\} \times [0,1] \rightarrow \circ^+$  be a strictly proper multi-dividing loss. Let  $\lambda > 0$ . We say  $l$  is  $\lambda$ -strong proper multi-dividing loss if for all  $\eta, \hat{\eta} \in [0,1]$ , we have

$$L_i^k(\eta, \hat{\eta}) - H_i^k(\eta) \geq \frac{\lambda}{2}(\eta - \hat{\eta})^2.$$

Thus, the ontology learning model in this paper is based on the Bayes multi-dividing ontology  $l$ -risk and conditional Bayes multi-dividing ontology  $l$ -risk. The main trick in our model is to use the proper multi-dividing loss.

### 3. MATHEMATICAL ANALYSIS

In this section, we present some theoretical analysis on multi-dividing proper loss function. Our first result stated as follows is a characterization of strict properness of a multi-dividing proper loss  $l$  in terms of its conditional multi-dividing ontology Bayes risk  $H_i^k$ :

**Lemma 1.** A proper multi-dividing loss  $l : \{1, 2, \dots, k\} \times [0,1] \rightarrow \circ^+$  is strictly proper if and only if  $H_i^k$  is strictly concave.

**Proof.** Let  $l : \{1, 2, \dots, k\} \times [0,1] \rightarrow \circ^+$  be a strictly proper multi-dividing ontology loss. Assume  $H_i^k$  is strictly concave. For  $\eta, \hat{\eta} \in [0,1]$  satisfy  $\eta \neq \hat{\eta}$ . Then we deduce

$$\begin{aligned} & L_i^k(\eta, \hat{\eta}) - H_i^k(\eta) \\ &= L_i^k(\eta, \hat{\eta}) + H_i^k(\hat{\eta}) - H_i^k(\hat{\eta}) - H_i^k(\eta) \\ &= L_i^k(\eta, \hat{\eta}) + H_i^k(\hat{\eta}) - 2\left(\frac{H_i^k(\hat{\eta})}{2} + \frac{H_i^k(\eta)}{2}\right) \\ &> L_i^k(\eta, \hat{\eta}) + H_i^k(\hat{\eta}) - 2H_i^k\left(\frac{\hat{\eta} + \eta}{2}\right) \end{aligned}$$

$$\begin{aligned} &= 2 \sum_{a=1}^{k-1} \sum_{b=a+1}^k \left( \frac{\eta^{a,b} + \hat{\eta}^{a,b}}{2} l(a, \hat{\eta}^{a,b}) \right) \\ &+ \left(1 - \frac{\eta^{a,b} + \hat{\eta}^{a,b}}{2}\right) l(b, \hat{\eta}^{a,b}) \\ &= 2\left(L_i^k\left(\frac{\eta + \hat{\eta}}{2}, \hat{\eta}\right) - H_i^k\left(\frac{\eta + \hat{\eta}}{2}\right)\right) \\ &\geq 0. \end{aligned}$$

This reveals that  $l$  is strictly proper multi-dividing loss.

On the other hand, we assume that  $l$  is strictly proper multi-dividing loss. Let  $\eta_1, \eta_2 \in [0,1]$  satisfy  $\eta_1 \neq \eta_2$  and let  $t \in (0,1)$ . Then, we infer

$$\begin{aligned} & H_i^k(t\eta_1 + (1-t)\eta_2) \\ &= L_i^k(t\eta_1 + (1-t)\eta_2, t\eta_1 + (1-t)\eta_2) \\ &= tL_i^k(\eta_1, t\eta_1 + (1-t)\eta_2) + (1-t)L_i^k(\eta_2, t\eta_1 + (1-t)\eta_2) \\ &> tH_i^k(\eta_1) + (1-t)H_i^k(\eta_2). \end{aligned}$$

Hence,  $H_i^k$  is strictly concave.

We have the following necessary and sufficient conditions for strong properness:

**Lemma 2.** Let  $\lambda > 0$ . If  $l : \{1, 2, \dots, k\} \times [0,1] \rightarrow \circ^+$  is  $\lambda$ -strong proper multi-dividing loss, then  $H_i^k$  is  $\lambda$ -strongly concave.

**Proof.** The proof is similar to the Lemma 1. Assume that  $l$  is  $\lambda$ -strongly proper multi-dividing loss. Let  $\eta_1, \eta_2 \in [0,1]$  satisfy  $\eta_1 \neq \eta_2$  and let  $t \in (0,1)$ . Then, we obtain

$$\begin{aligned} & H_i^k(t\eta_1 + (1-t)\eta_2) \\ &= L_i^k(t\eta_1 + (1-t)\eta_2, t\eta_1 + (1-t)\eta_2) \\ &= tL_i^k(\eta_1, t\eta_1 + (1-t)\eta_2) + (1-t)L_i^k(\eta_2, t\eta_1 + (1-t)\eta_2) \\ &\geq t(H_i^k(\eta_1) + \frac{\lambda}{2}(1-t)^2(\eta_1 - \eta_2)^2) \\ &+ (1-t)(H_i^k(\eta_2) + \frac{\lambda}{2}t^2(\eta_1 - \eta_2)^2) \\ &= tH_i^k(\eta_1) + (1-t)H_i^k(\eta_2) + \frac{\lambda}{2}t(1-t)(\eta_1 - \eta_2)^2 \end{aligned}$$

Hence,  $H_i^k$  is  $\lambda$ -strictly concave.

**Lemma 3.** Let  $\lambda > 0$  and  $l : \{1, 2, \dots, k\} \times [0,1] \rightarrow \circ^+$  be a regular proper multi-dividing loss. If  $H_i^k$  is  $\lambda$ -strongly concave, then  $l$  is  $\lambda$ -strong proper multi-dividing loss.

**Proof.** Let  $\eta, \hat{\eta} \in [0,1]$ . Then there exists a super-derivative  $H'_i(\hat{\eta})$  of  $H_i^k$  at  $\hat{\eta}$  such that

$$L_i^k(\eta, \hat{\eta}) = H_i^k(\hat{\eta}) + (\eta - \hat{\eta})H'_i(\hat{\eta}).$$

This implies

$$\begin{aligned} &L_i^k(\eta, \hat{\eta}) - H_i^k(\eta) \\ &= H_i^k(\hat{\eta}) - H_i^k(\eta) + (\eta - \hat{\eta})H'_i(\hat{\eta}) \\ &\geq \frac{\lambda}{2}(\hat{\eta} - \eta)^2. \end{aligned}$$

Hence,  $l$  is  $\lambda$ -strong proper multi-dividing loss function.

Combining Lemma 2 and Lemma 3, we get the character of strong properness for regular proper multi-dividing losses stated in Theorem 1.

**Theorem 1.** Let  $\lambda > 0$  and  $l: \{1, 2, \dots, k\} \times [0,1] \rightarrow \circ^+$  be a regular proper multi-dividing loss. Then  $l$  is  $\lambda$ -strong proper if and only if  $H_i^k$  is  $\lambda$ -strongly concave.

Next, we obtain the regret bound using strongly proper multi-dividing ontology losses. We begin with the following lemma.

**Lemma 4.** For any  $\hat{\eta}: V \rightarrow [0,1]$ , we have

$$\text{regret}_D^k[\hat{\eta}] \leq \sum_{a=1}^{k-1} \sum_{b=a+1}^k \frac{1}{p^{a,b}(1-p^{a,b})} E_V[|\hat{\eta}(V^a) - \eta(V^a)|].$$

**Proof.** Let  $\hat{\eta}: V \rightarrow [0,1]$ , we deduce

$$\begin{aligned} &\text{regret}_D^k[\hat{\eta}] \\ &\leq \sum_{a=1}^{k-1} \sum_{b=a+1}^k \frac{1}{2p^{a,b}(1-p^{a,b})} E_{V^a, V^b}[|\eta(V^a) - \eta(V^b)| \\ &\cdot \mathbb{I}_{(\hat{\eta}(V^a) - \hat{\eta}(V^b))(\eta(V^a) - \eta(V^b)) \leq 0}]. \end{aligned}$$

Note that for each pair of  $(a, b)$ , if  $\eta^{a,b}(v^a) > \eta^{a,b}(v^b)$ , we get

$$\begin{aligned} &(\hat{\eta}^{a,b}(v^a) - \hat{\eta}^{a,b}(v^b))(\eta^{a,b}(v^a) - \eta^{a,b}(v^b)) \leq 0 \Rightarrow \\ &\hat{\eta}^{a,b}(v^a) \leq \hat{\eta}^{a,b}(v^b) \\ &\Rightarrow \eta^{a,b}(v^a) - \eta^{a,b}(v^b) \leq (\eta^{a,b}(v^a) - \hat{\eta}^{a,b}(v^a)) \\ &+ (\hat{\eta}^{a,b}(v^b) - \eta^{a,b}(v^b)) \\ &\Rightarrow \eta^{a,b}(v^a) - \eta^{a,b}(v^b) \leq |\eta^{a,b}(v^a) - \hat{\eta}^{a,b}(v^a)| \\ &+ |\hat{\eta}^{a,b}(v^b) - \eta^{a,b}(v^b)| \\ &\Rightarrow |\eta^{a,b}(v^a) - \eta^{a,b}(v^b)| \leq |\hat{\eta}^{a,b}(v^a) - \eta^{a,b}(v^a)| \\ &+ |\hat{\eta}^{a,b}(v^b) - \eta^{a,b}(v^b)|. \end{aligned}$$

If  $\eta^{a,b}(v^a) < \eta^{a,b}(v^b)$ , by using the similar way, we have

$$\begin{aligned} &(\hat{\eta}^{a,b}(v^a) - \hat{\eta}^{a,b}(v^b))(\eta^{a,b}(v^a) - \eta^{a,b}(v^b)) \leq 0 \\ &\Rightarrow |\eta^{a,b}(v^a) - \eta^{a,b}(v^b)| \leq |\hat{\eta}^{a,b}(v^a) - \eta^{a,b}(v^a)| \\ &+ |\hat{\eta}^{a,b}(v^b) - \eta^{a,b}(v^b)|. \end{aligned}$$

And, this result is trivially true if  $\eta^{a,b}(v^a) = \eta^{a,b}(v^b)$ . Thus, we get

$$\begin{aligned} &\text{regret}_D^k[\hat{\eta}] \\ &\leq \sum_{a=1}^{k-1} \sum_{b=a+1}^k \frac{1}{2p^{a,b}(1-p^{a,b})} E_{V^a, V^b}[|\hat{\eta}^{a,b}(V^a) - \eta^{a,b}(V^a)| \\ &+ |\hat{\eta}^{a,b}(V^b) - \eta^{a,b}(V^b)|] \\ &= \sum_{a=1}^{k-1} \sum_{b=a+1}^k \frac{1}{p^{a,b}(1-p^{a,b})} E_V[|\hat{\eta}(V) - \eta(V)|] \end{aligned}$$

**Theorem 2.** Let  $\hat{Y} \subseteq \circ$  and  $\lambda > 0$ . Let  $l: \{1, 2, \dots, k\} \times \hat{Y} \rightarrow \circ^+$  be a  $\lambda$ -strongly proper composite multi-dividing loss. Then for any multi-dividing ontology function  $f: V \rightarrow \hat{Y}$ ,

$$\text{regret}_D^k(f) \leq \sqrt{\text{regret}_D^{k,l}(f)} \sum_{a=1}^{k-1} \sum_{b=a+1}^k \frac{\sqrt{2}}{p^{a,b}(1-p^{a,b})\sqrt{\lambda}}.$$

**Proof.** Let  $l: \{1, 2, \dots, k\} \times [0,1] \rightarrow \circ^+$  be a  $\lambda$ -strongly proper multi-dividing loss and  $\psi: [0,1] \rightarrow \hat{Y}$  be a strictly increasing link function such that  $l(y, \hat{y}) = l(y, \psi^{-1}(\hat{y}))$ .

Let  $f: V \rightarrow \hat{Y}$ . Then we have

$$\begin{aligned} &\text{regret}_D^k(f) = \text{regret}_D^k(\psi^{-1} \circ f) \\ &\leq \sum_{a=1}^{k-1} \sum_{b=a+1}^k \frac{1}{p^{a,b}(1-p^{a,b})} E_V[|\psi^{-1}(f(V)) - \eta(V)|] \\ &= \sum_{a=1}^{k-1} \sum_{b=a+1}^k \frac{1}{p^{a,b}(1-p^{a,b})} \sqrt{(E_V[|\psi^{-1}(f(V)) - \eta(V)|])^2} \\ &\leq \sum_{a=1}^{k-1} \sum_{b=a+1}^k \frac{1}{p^{a,b}(1-p^{a,b})} \sqrt{E_V[(\psi^{-1}(f(V)) - \eta(V))^2]} \\ &\leq \sum_{a=1}^{k-1} \sum_{b=a+1}^k \frac{1}{p^{a,b}(1-p^{a,b})} \sqrt{\frac{2}{\lambda} E_V[R_c^k(\eta(V), \psi^{-1}(f(V)))]} \\ &= \sum_{a=1}^{k-1} \sum_{b=a+1}^k \frac{1}{p^{a,b}(1-p^{a,b})} \sqrt{\frac{2}{\lambda} E_V[R_c^k(\eta(V), f(V))]} \\ &= \sqrt{\text{regret}_D^{k,l}(f)} \sum_{a=1}^{k-1} \sum_{b=a+1}^k \frac{\sqrt{2}}{p^{a,b}(1-p^{a,b})\sqrt{\lambda}}. \end{aligned}$$

The first equality holds since  $\psi$  is strictly increasing; the first inequality holds by Lemma 4; the second inequality established by convexity of  $\phi(u) = u^2$  and Jensen's inequality; the last inequality holds since  $l'$  is  $\lambda$ -strongly proper multi-dividing loss.

Theorem 2 implies that for any strong proper composite multi-dividing loss  $l: \{1, 2, \dots, k\} \times \hat{Y} \rightarrow \circ^+$ , a multi-dividing ontology function  $f: V \rightarrow \hat{Y}$  with low  $l$ -regret will also have low multi-dividing ontology regret.

In the last part of this section, we show that our result is stable under low-noise conditions. Let  $\alpha \in [0, 1]$ . A distribution  $D$  on  $V \times \{1, 2, \dots, k\}$  satisfies assumption that if there exists a positive constant  $C$  such that for all  $v \in V$ , each pair of  $(a, b)$  and any  $t \in [0, 1]$ ,

$$P_v(|\eta^{a,b}(V) - \eta^{a,b}(v)| \leq t) \leq Ct^\alpha$$

Note that  $\alpha = 0$  imposes restriction on  $D$ , while smaller value of  $\alpha$  impose smaller restrictions.

**Theorem 3.** Let  $\hat{Y} \subseteq \circ, \lambda > 0$  and  $\alpha \in [0, 1]$ . Let  $l: \{1, 2, \dots, k\} \times \hat{Y} \rightarrow \circ^+$  be a  $\lambda$ -strongly proper composite multi-dividing loss. Then there exists a positive constant  $C_\alpha$  such that for any distribution  $D$  on  $V \times \{1, 2, \dots, k\}$  satisfying above low-noise condition and for any multi-dividing ontology function  $f: V \rightarrow \hat{Y}$ , we have

$$\begin{aligned} & \text{regret}_D^k(f) \\ & \leq \left(\frac{2}{\lambda}\right)^{\alpha+2} (\text{regret}_D^{k,l}(f))^{\alpha+2} \sum_{a=1}^{\alpha+1} \sum_{b=a+1}^k \frac{C_\alpha}{p^{a,b}(1-p^{a,b})} \end{aligned}$$

**Proof.** Let  $l: \{1, 2, \dots, k\} \times [0, 1] \rightarrow \circ^+$  be a  $\lambda$ -strongly proper multi-dividing loss and  $\psi: [0, 1] \rightarrow \hat{Y}$  be a strictly increasing link function such that  $l(y, \hat{y}) = l'(y, \psi^{-1}(\hat{y}))$ . Let  $D$  be a distribution on  $V \times \{1, 2, \dots, k\}$  satisfying above low-noise condition. Let  $f: V \rightarrow \hat{Y}$ . Then we infer

$$\begin{aligned} & \text{regret}_D^k(f) = \text{regret}_D^k(\psi^{-1} \circ f) \\ & \leq \sum_{a=1}^{k-1} \sum_{b=a+1}^k \frac{C^{a,b}}{p^{a,b}(1-p^{a,b})} (E_V[(\psi^{-1}(f(V)) - \eta(V))^2])^{\frac{\alpha+1}{\alpha+2}} \\ & = \sum_{a=1}^{k-1} \sum_{b=a+1}^k \frac{C^{a,b}}{p^{a,b}(1-p^{a,b})} \left(\frac{2}{\lambda} E_V[R_l^k(\eta(V), \psi^{-1}(f(V)))]\right)^{\frac{\alpha+1}{\alpha+2}} \\ & = \sum_{a=1}^{k-1} \sum_{b=a+1}^k \frac{C^{a,b}}{p^{a,b}(1-p^{a,b})} \left(\frac{2}{\lambda} E_V[R_l^k(\eta(V), f(V))]\right)^{\frac{\alpha+1}{\alpha+2}} \end{aligned}$$

$$= \left(\frac{2}{\lambda}\right)^{\alpha+1} (\text{regret}_D^{k,l}(f))^{\alpha+2} \sum_{a=1}^{\alpha+1} \sum_{b=a+1}^k \frac{C^{a,b}}{p^{a,b}(1-p^{a,b})}$$

The first equality holds since  $\psi$  is strictly increasing; the first inequality holds by the fact that for  $q \in [1, \infty]$ , there exists a positive constant  $C_{\alpha,q}$  such that for any distribution  $D$  on  $V \times \{1, 2, \dots, k\}$  satisfying low-noise condition and any  $\hat{\eta}: V \rightarrow [0, 1]$

$$\begin{aligned} & \text{regret}_D^k(\hat{\eta}) \\ & \leq \sum_{a=1}^{k-1} \sum_{b=a+1}^k \frac{C_{\alpha,q}}{p^{a,b}(1-p^{a,b})} (E_V[|\hat{\eta}(V) - \eta(V)|^q])^{\frac{\alpha+1}{\alpha+q}}, \end{aligned}$$

and then taking  $q=2$ ; the second inequality established since  $l'$  is  $\lambda$ -strongly proper multi-dividing loss.

At last, the desired result is obtained by setting  $C_\alpha = \max_{a,b} \{C_{\alpha}^{a,b}\}$ .

#### 4. DESCRIPTION OF ONTOLOGY ALGORITHMS

The new multi-dividing ontology learning algorithm can be used in ontology concepts similarity measurement and ontology mapping. The basic idea is: *via* the empirical version of new model, the ontology graph is mapped into a line consisting of real numbers. The similarity between two concepts then can be measured by comparing the difference between their corresponding real numbers.

**Algorithm 1.** Ontology similarity measuring based on multi-dividing model using proper loss:

Step 1. Mathematizing ontology information. For each vertex in ontology graph, we use a vector to express all its information.

Step 2. By minimizing empirical version of Bayes multi-dividing ontology  $l$ -risk or conditional Bayes multi-dividing ontology  $l$ -risk, we map the ontology graph to the real line and map the vertices of ontology graph to real numbers.

Step 3. For each  $v \in V(G)$ , we use one of the following methods to obtain the similar vertices and return the outcomes to the users.

Method 1: Choosing a parameter  $M$ , return set  $\{v' \in V(G), |f(v) - f(v')| \leq M\}$ .

Method 2: Selecting an integer  $N$ , return the closest  $N$  concepts according to the value list in  $V(G)$ .

Clearly, method 1 looks fairer and method 2 can control the number of vertices that return to the users.

**Algorithm 2.** Ontology mapping based on multi-dividing model using proper loss:

Let  $G_1, G_2, \dots, G_m$  be ontology graph correspond to ontology  $O_1, O_2, \dots, O_m$ .

Step 1. Mathematizing ontology information. For each vertex in ontology graph, we use a vector to express all its information.

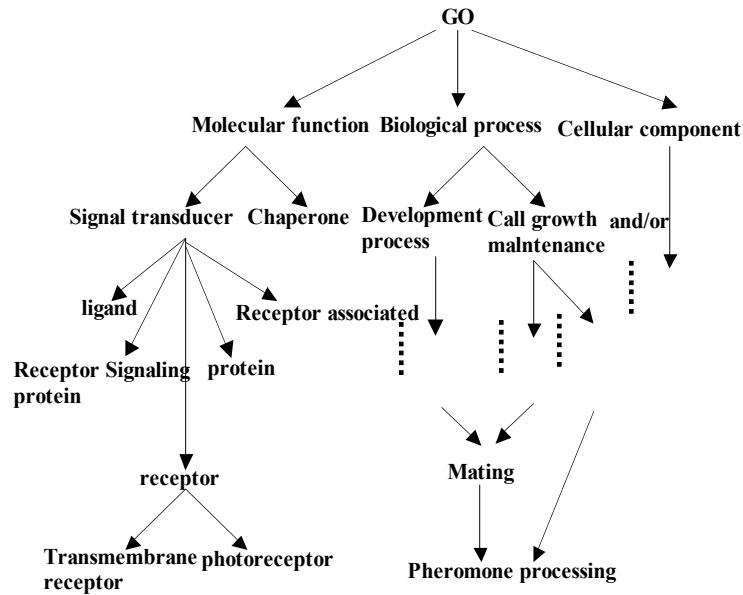


Fig. (1). “Go” ontology.

Table 1. The experiment data on biology ontology.

	<i>P@3</i> Average Precision Ratio	<i>P@5</i> Average Precision Ratio	<i>P@10</i> Average Precision Ratio	<i>P@20</i> Average Precision Ratio
Algorithm 1	46.76%	55.76%	65.91%	78.15%
Algorithm in Huang <i>et al.</i> , [5]	46.38%	53.48%	62.34%	74.59%
Algorithm in Gao and Liang [6]	43.56%	49.38%	56.47%	71.94%
Algorithm in Gao and Gao [7]	42.13%	51.83%	60.19%	72.39%

Step 2. By minimizing empirical version of Bayes multi-dividing ontology *l*-risk or conditional Bayes multi-dividing ontology *l*-risk, we map the ontology graph to the real line and map the vertices of ontology graph to real numbers.

Step 3. For  $v \in V(G_i)$ , where  $1 \leq i \leq m$ , we use one of following methods to obtain the similar vertices and return the outcome to the users.

Method 1: Choose a parameter *M*, return set  $\{v' \in V(G-G_i), |f(v) - f(v')| \leq M\}$ .

Method 2: Choose an integer *N*, return the closest *N* concepts on the list in  $V(G-G_i)$ .

Also, method 1 looks fairer but method 2 can control the number of vertices that return to the users.

### 5. ENTEXPERIMENTS

In this section, four simulation experiments relevance ontology similarity measure and ontology mapping are designed below. In order to adjacent to the setting of ontology algorithm, we use a vector with same dimension to express each vertex’s information. Such vector contains the information of name, instance, attribute and structure of vertex. Here the instance of vertex refers to the set of its reachable vertex in the directed ontology graph.

### 5.1. Experiment on Biology Data

We use “Go” ontology  $O_1$  which was constructed in <http://www.geneontology.org>. (Fig. 1 shows the basic structure of  $O_1$ ) for our experiment. *P@N* (Precision Ratio, see Craswell and Hawking [12] for more detail) is used to measure the equality of the experiment. We first give the closest *N* concepts for every vertex on the ontology graph by experts, and then we obtain the first *N* concepts for every vertex on ontology graph by the algorithm and compute the precision ratio. In this experiment, we use canonical exponential

$$l_{\text{exp,can}}(y, \hat{y}) = \sqrt{1 + \left(\frac{\hat{y}}{2}\right)^2} - \frac{|y - \hat{y}|}{2}$$

algorithms in Huang *et al.*, [5], Gao and Liang [6] and Gao and Gao [7] are employed to “Go” ontology, and we compare the precision ratio which we get from the four methods. Several experiment results refer to (Table 1).

When *N*= 3, 5, 10 or 20, the precision ratio by virtue of our algorithm is higher than the precision ratio determined by algorithms proposed in Huang *et al.*, [5], Gao and Liang [6] and Gao and Gao [7]. In particular, when *N* increases, such precision ratios are increasing apparently. Therefore, the algorithm described in our paper is superior to the method proposed by Huang *et al.*, [5], Gao and Liang [6] and Gao and Gao [7].

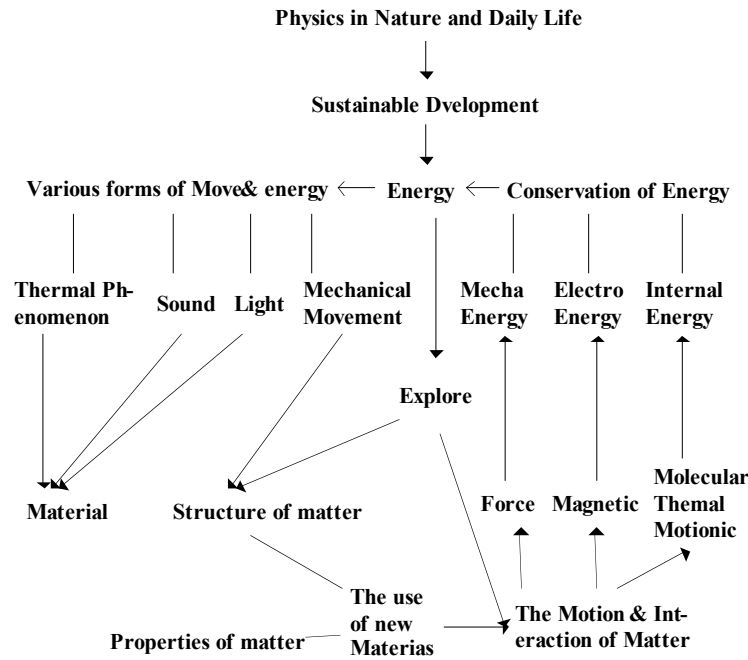


Fig. (2). “Physical Education” Ontology  $O_2$ .

Table 2. The experiment data on physical education ontology.

	$P@1$ Average Precision Ratio	$P@3$ Average Precision Ratio	$P@5$ Average Precision Ratio
Algorithm 2	70.97%	78.49%	90.32%
Algorithm in Huang <i>et al.</i> , [5]	61.29%	73.12%	79.35%
Algorithm in Gao and Liang [6]	69.13%	75.56%	84.52%
Algorithm in Gao <i>et al.</i> , [11]	67.74%	77.42%	89.68%

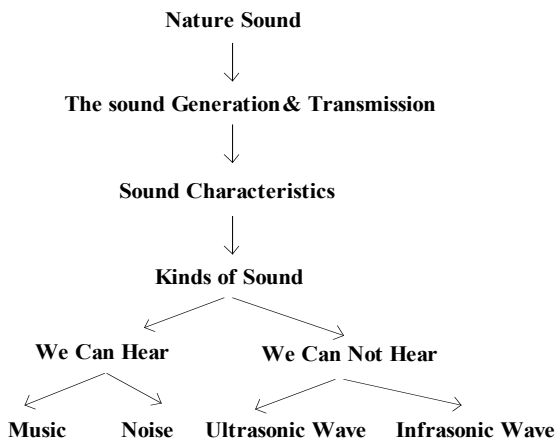


Fig. (3). “Physical Education” Ontology  $O_3$ .

5.2. Experiment on Physical Education Data

We use physical education ontologies  $O_2$  and  $O_3$  (the structures of  $O_2$  and  $O_3$  are presented in Figs. (2) and (3) respectively) for our second experiment. The goal of this experiment is determining the ontology mapping between  $O_2$  and  $O_3$  via similarity matrix which is deduced by Algorithm

1.  $P@N$  criterion is applied to measure the equality of the experiment. We first give the closest  $N$  concepts for each vertex on the ontology graph with the help of experts, and then we obtain the first  $N$  concepts for every vertex on ontology graph by the algorithm and compute the precision ratio. The loss function used in this experiment is squared loss function  $l_{sq}(y, \hat{y}) = (1 - |y - \hat{y}|)^2$  and canonical squared loss function  $l_{sq.can}(y, \hat{y}) = (1 - \frac{|y - \hat{y}|}{4})^2$ . Also, ontology algorithms in Huang *et al.*, [5], Gao and Liang [6] and Gao *et al.*, [11] are employed to “physical education” ontology, and we compare the precision ratio which we get from four methods. Several experiment results refer to (Table 2).

The experiment results in Table 2 reveal that our algorithm is more efficient than algorithms raised in Huang *et al.*, [5], Gao and Liang [6] and Gao *et al.*, [11] especially when  $N$  is sufficiently large.

5.3. Experiment on Plant Data

In this subsection, “PO” ontology  $O_4$  which was constructed in <http://www.plantontology.org>. (Fig. 4 shows the basic structure of  $O_4$ ) is used to test the efficiency of our new

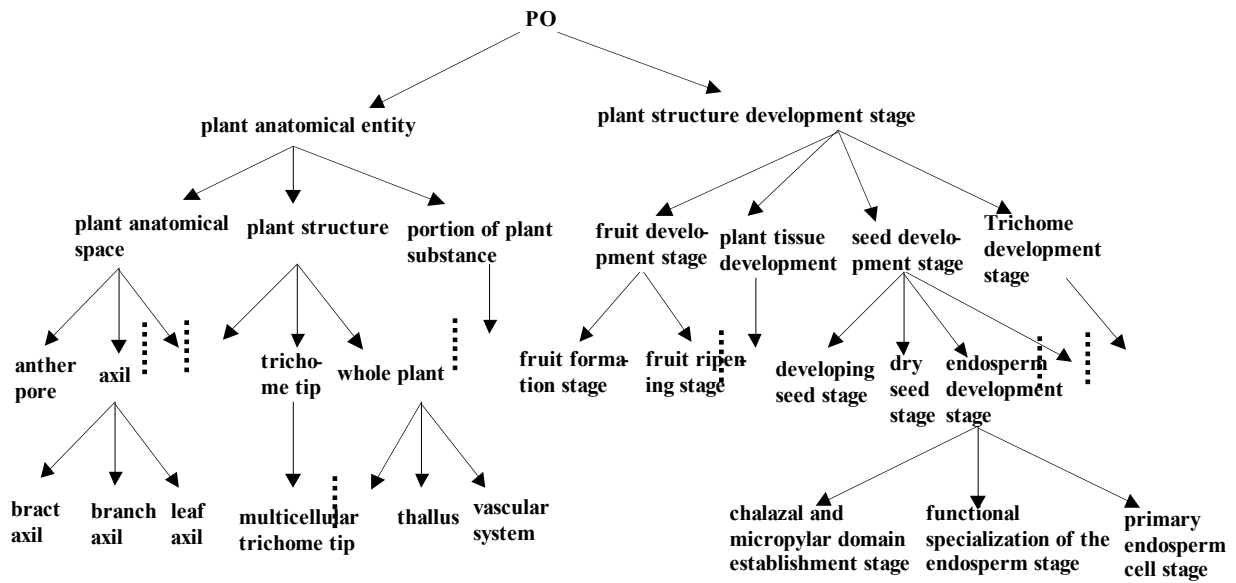


Fig. (4). “PO” ontology  $O_4$ .

Table 3. The experiment data on plant ontology.

	$P@3$ Average Precision Ratio	$P@5$ Average Precision Ratio	$P@10$ Average Precision Ratio
Algorithm 1	48.85%	57.02%	72.63%
Algorithm in Wang <i>et al.</i> , [4]	45.49%	51.17%	58.59%
Algorithm in Huang <i>et al.</i> , [5]	42.82%	48.49%	56.32%
Algorithm in Gao and Liang [6]	48.31%	56.35%	68.71%

algorithm for ontology similarity measuring. The  $P@N$  standard is used again for this experiment. The loss function we use logistic loss  $l_{\log} : \{1, 2, \dots, k\} \times \circ \rightarrow \circ +$  defined as

$$l_{\log}(y, \hat{y}) = \ln(1 + e^{-|y - \hat{y}|^2})$$

is a proper composite loss. Furthermore, we apply ontology method in Wang *et al.*, [4], Huang *et al.*, [5] and Gao and Liang [6] to the “PO” ontology. Calculating the accuracy by these three algorithms and compare the result to algorithm in our paper, part of the data refers to (Table 3).

When  $N=3, 5$ , or  $10$ , the precision ratio in terms of our algorithm is higher than the precision ratio determined by algorithms proposed in Wang *et al.*, [4], Huang *et al.*, [5] and Gao and Liang [6]. In particular, when  $N$  increases, such precision ratios are increasing apparently. Therefore, the algorithm described in our paper is superior to the method proposed by Wang *et al.*, [4], Huang *et al.*, [5] and Gao and Liang [6].

### 5.4. Experiment on Humanoid Robotics Data

We use humanoid robotics ontologies  $O_5$  and  $O_6$  (constructed by Gao and Zhu [13], and the structures of  $O_5$  and  $O_6$  are presented in Figs. (5) and (6) respectively) for our last experiment. The goal of this experiment is to determine ontology mapping between  $O_5$  and  $O_6$  via similarity matrix

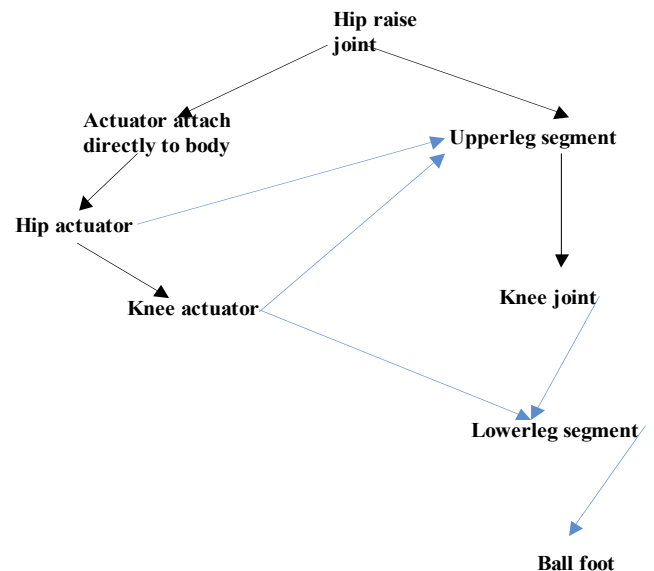


Fig. (5). “Humanoid Robotics” Ontology  $O_5$ .

which are deduced by Algorithm 1.  $P@N$  criterion is applied to measure the equality of the experiment. Ontology algorithms in Gao and Lan [14], Gao and Liang [6] and Gao *et al.*, [11] are employed to humanoid robotics ontologies, and we compare the precision ratio which we get from four methods. Several experiment results refer to (Table 4).



Table 4. The experiment data on humanoid robotics ontology.

	P@1 Average Precision Ratio	P@3 Average Precision Ratio	P@5 Average Precision Ratio
Algorithm 2	27.78%	50.00%	58.89%
Algorithm in Gao and Lan [14]	27.78%	48.15%	54.44%
Algorithm in Gao and Liang [6]	22.22%	40.74%	48.89%
Algorithm in Gao <i>et al.</i> , [11]	27.78%	46.30%	53.33%

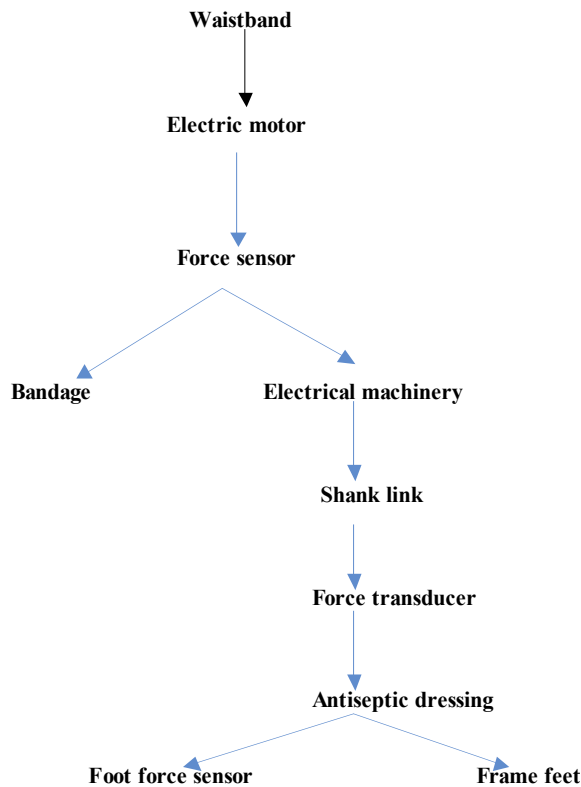


Fig. (6). “Humanoid Robotics” Ontology  $O_6$ .

In this experiment,  $k=2$ . The loss function we use exponential loss  $l_{exp} : \{1, 2, \dots, k\} \times \circ \rightarrow \circ +$  defined as

$$l_{exp}(y, \hat{y}) = e^{-\frac{|y-\hat{y}|^2}{2}}$$

is a proper composite loss.

The experiment results in Table 4 reveal that our algorithm is more efficient than algorithms raised in Gao and Lan [14], Gao and Liang [6] and Gao *et al.*, [11] especially when  $N$  is sufficiently large.

### 6. CONCLUSION

In this paper, we propose a new idea for multi-dividing ontology similarity measure and ontology mapping application. The tricks are based on the proper multi-dividing loss function and some new fashions and theoretical characteristics are employed for using such technologies. At last, simulation data shows that multi-dividing ontology algorithms using proper loss functions have high efficiency in biology, physics education, plant science and humanoid robotics.

### CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

### ACKNOWLEDGEMENTS

First we thank the reviewers for their constructive comments in improving the quality of this paper. This work was supported in part by the PHD initial funding of the first author. We also would like to thank the anonymous referees for providing us with constructive comments and suggestions.

### REFERENCES

- [1] A. Bouzeghoub and A. Elbyed, “Ontology mapping for web-based educational systems interoperability,” *Interoperability in Business Information Systems*, vol. 1, pp. 73-84, 2006.
- [2] B. Hu, S. Dasmahapatra, P. Lewis, and N. Shadbolt, “Ontology-based medical image annotation with description logics,” *In 15th IEEE International Conference on Tools with Artificial Intelligence*, California, Sacramento, USA, pp. 77-82, 2003.
- [3] F. Fonseca, E. Egenhofer, C. Davis, and G. Camara, “Semantic granularity in ontology-driven geographic information systems,” *AMAI Annals of Mathematics and Artificial Intelligence - Special Issue on Spatial and Temporal Granularity*, vol. 36, pp. 121-151, 2001.
- [4] Y. Wang, W. Gao Y. Zhang, and Y. Gao, “Ontology similarity computation use ranking learning Method,” *The 3rd International Conference on Computational Intelligence and Industrial Application*, Wuhan, China, pp. 20-22, 2010.
- [5] X. Huang, T. Xu, W. Gao and Z. Jia, “Ontology similarity measure and ontology mapping via fast ranking method,” *International Journal of Applied Physics and Mathematics*, vol. 1, pp. 54-59, 2011.
- [6] W. Gao and L. Liang, “Ontology similarity measure by optimizing NDCG measure and application in physics education,” *Future Communication, Computing, Control and Management*, vol. 142, pp. 415-421, 2011.
- [7] Y. Gao and W. Gao, “Ontology similarity measure and ontology mapping via learning optimization similarity function,” *International Journal of Machine Learning and Computing*, vol. 2, pp. 107-112, 2012.
- [8] X. Huang, T. Xu, W. Gao, and S. Gong, “Ontology similarity measure and ontology mapping using half transductive ranking,” *In Proceedings of 2011 4th IEEE international conference on computer science and Information technology*, Chengdu, China, pp. 571-574, 2011.
- [9] W. Gao and T. Xu, “Stability analysis of learning algorithms for ontology similarity computation,” *Abstract and Applied Analysis*, Vol. 2013, p. 9, 2013.
- [10] L. Zhu, F. Wu, and W. Gao, “Theoretical analysis for new multi-dividing ontology algorithm based on AUC criterion,” *Computing and Informatics*, In press.
- [11] W. Gao, Y. Gao, and L. Liang, “Diffusion and harmonic analysis on hypergraph and application in ontology similarity measure and ontology mapping,” *Journal of Chemical and Pharmaceutical Re-*

- search, vol. 5, no. 9, pp. 592-598, 2013.
- [12] N. Craswell and D. Hawking, "Overview of the TREC 2003 web track," *Proceeding of the Twelfth Text Retrieval Conference*, Gaithersburg, Maryland, NIST Special Publication, pp. 78-92, 2003.
- [13] W. Gao and L. Zhu, L, "Gradient learning algorithms for ontology computing," *Computational Intelligence and Neuroscience*, Vol. 2014, p. 12, 2014.
- [14] W. Gao and M. Lan, M, "Ontology mapping algorithm based on ranking learning method," *Microelectronics & Computer*, vol. 9, pp. 59-61, 2011.

---

Received: September 22, 2014

Revised: November 30, 2014

Accepted: December 02, 2014

© Gao et al.; Licensee Bentham Open.

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.